ACM 113

10/2 L1

Optimization is "the science of selecting the best element from a collection of elements".

Every optimization problem has two pieces:
- an objective function
- a collection of feasable elements

Origins of optimization:
- Variational principles (priciple of least action...)
- "Operations Research"
  ↳ Engineering, Medicine, Finance, Statistics, ML

Examples:
- Optimize returns to a constraint on the risk in allocating a dollar over a collection of assets
- Design the wings of an aircraft to minimize drag subject to constraints on weight, material costs, etc.
- Given some data, find the best model from some class that fits the data

__Mathematical Optimization:__

$$\underbrace{p^*}_{\text{optimal value}} = \inf_{x \in \underbrace{\mathcal{E}}_{\text{ground set}}} \underbrace{f(x)}_{\text{objective function}}$$

$$\underbrace{\text{s.t.}}_{\text{such that/subject to}} \ \underbrace{x}_{\text{variable/decision variable}} \in \underbrace{S}_{\text{constrained set}}$$

Ground Set: the domain of definition of the function f

Ex: $\mathcal{E} = \mathbb{R}^n$; $\mathcal{E} = \mathbb{Z}^n$; $\mathcal{E}$ = the set of graphs on n nodes; $\mathcal{E}$ = the set of functions from $[0,1] \mapsto \mathbb{R}$

Objective function - $f: \mathcal{E} \mapsto \mathbb{R}$

Constraint set - $S \subseteq \mathcal{E}$

__Definition:__ Let $G \subseteq \mathbb{R}$ be a subset of the reals. The __infimum__ of G is the largest lower bound of G.

Ex: $G = (0,5]$, $\inf(G) = 0$ $\quad \min(G) = \text{und.}$

$G = [0,5]$, $\inf(G) = 0$ $\quad \min(G) = 0$

~ __supremum__: smallest upper bound

Optimal value - p^* is the largest lower bound of $f(x)$ over $x \in S$. $\rightarrow p^* = \inf\{f(x) \mid x \in S\}$

Feasible element/point - any $x \in S$ (S is also sometimes called the feasible set)

Optimal solution - any $\tilde{x} \in S$ such that $f(\tilde{x}) = p^*$

Note: "optimization" $\equiv$ "programming"

Depending on the ground set, the resulting class of optimization problems usually has a name associated to it!
- $\mathcal{E} = \mathbb{R}^n \rightarrow$ continuous optimization
- $\mathcal{E} = \mathbb{Z}^n \rightarrow$ integer programming
- $\mathcal{E}$ = some discrete set (e.g. graphs) $\rightarrow$ network optimization
  ↳ combinatorial optimization
- $\mathcal{E}$ = some collection of functions (e.g.: $[0,1] \rightarrow \mathbb{R}$)
  ↳ calculus of variations

There are many other examples that do not fall into our abstraction that commonly arise in practice:
- Stochastic optimization - The objective function depends on additional random parameters: $f(x; \underbrace{u}_{\text{random variable}})$
- Dynamic programming - The objective function evolves over time: $f_1(x), f_2(x), \ldots$
- Multi-criterion (objective) optimization - Here $f: \mathcal{E} \rightarrow \mathbb{R}^n$ (for example)

Going back to our general form:

Questions we could ask:
- Is the feasible set empty or not?
  ↳ By convention, if $S = \emptyset$, we set $p^* = +\infty$ (in a minimization problem)
- Is the objective function bounded over the constraint set?
  ↳ If the objective function is unbounded below over the constraint set, we set $p^* = -\infty$ (in minimization problems)
- Is the optimal value attained? That is, is there an optimal solution? Is the optimal solution unique?
- How do we produce upper bounds on p^*? How do we produce lower bounds on p^*?
- How do we represent/solve an optimization problem on a computer?

10/4 L2

Focus for coming weeks: $\mathcal{E} = \mathbb{R}^n$

$$p^* = \inf_{x \in \mathbb{R}^n} f(x) \quad \text{s.t. } x \in S \qquad f: \mathbb{R}^n \rightarrow \mathbb{R}, \ S \subseteq \mathbb{R}^n$$

If $S = \mathbb{R}^n$, the resulting problem is sometimes called an unconstrained optimization problem

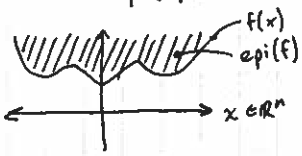
Q: How do we obtain an upper bound on p^*?

$$p^* = \inf_{\substack{x\in\mathbb{R}^n \\ t\in\mathbb{R}}} t \quad \text{s.t. } f(x) \le t, \quad x \in S$$

<u>Definition:</u> Let $f: \mathbb{R}^n \to \mathbb{R}$ be a function. Then the <u>epigraph</u> of f is the following subset of $\mathbb{R}^{n+1}$:

$$\text{epi}(f) = \{(x,t) \in \mathbb{R}^{n+1} \mid f(x) \le t\}$$

Ex:



Now, we have

$$p^* = \inf_{\substack{x\in\mathbb{R}^n \\ t\in\mathbb{R}}} t \quad \text{s.t. } (x,t) \in \text{epi}(f), \quad x \in S \leadsto \text{or} \leadsto (x,t) \in S \times \mathbb{R}$$

We can consider any point in $\text{epi}(f) \cap S \times \mathbb{R}$, read off the last coordinate of the point (i.e. the $n+1$ th coordinate) and obtain an upper bound on p^*.

Note: We have reformulated our original problem as one in which the objective function is linear

Q: How do we obtain lower bounds on p^*?

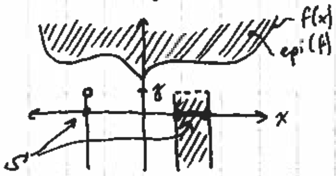
Proposition: Given an optimization problem

$$p^* = \inf_{x\in\mathbb{R}^n} f(x) \quad \text{s.t. } x \in S$$

we can reformulate it as:

$$p^* = \sup_{\gamma\in\mathbb{R}} \gamma \quad \text{s.t. } \text{epi}(f) \cap \{(x,t) \mid x \in S, t < \gamma\} = \emptyset$$

Ex:



Proof: Fix any $\gamma \le p^*$. Then we have that $\text{epi}(f) \cap \{(x,t) \mid x \in S, t < \gamma\} = \emptyset$. To prove this assertion, suppose for the sake of contradiction that $\text{epi}(f) \cap \{(x,t) \mid x \in S, t < \gamma\} \ni (\tilde{x}, \tilde{t})$. Then we conclude that $f(\tilde{x}) \le \tilde{t} < \gamma \le p^*$. This contradicts the fact that p^* is the optimal value of the original problem. In the other direction, fix any $\gamma > p^*$. Then we have that there exists $\tilde{x} \in S$ such that $f(\tilde{x}) < \gamma$ (based on the definition of p^* in our original problem), & consequently that $\text{epi}(f) \cap \{(x,t) \mid x \in S, t < \gamma\} \ne \emptyset$. There remains the situation in which $p^* = -\infty$. This arises if the objective is unbounded below over the constraint set. In this case, there is no $\gamma \in \mathbb{R}$ such that $\text{epi}(f) \cap \{(x,t) \mid x \in S, t < \gamma\} = \emptyset$. Consequently, $p^* = -\infty$ in the reformulation by convention. ■

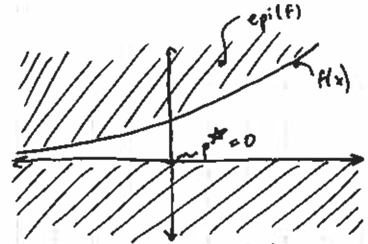
Remark: • The reformulation

$$p^* = \sup_{\gamma\in\mathbb{R}} \gamma \quad \text{s.t. } \text{epi}(f) \cap \{(x,t) \mid x \in S, t < \gamma\} = \emptyset$$

is called the dual problem.

• The dual problem has the feature that the optimal value is always attained (even if not in the original problem).

Ex:



$p^* = \inf_{x\in\mathbb{R}} \exp(x)$
↳ p^* never achieved

$p^* = \sup_{\gamma\in\mathbb{R}} \gamma$ s.t. $\text{epi}(\exp) \cap \{(x,t) \mid t < \gamma\} = 0$
↳ $p^* = 0$

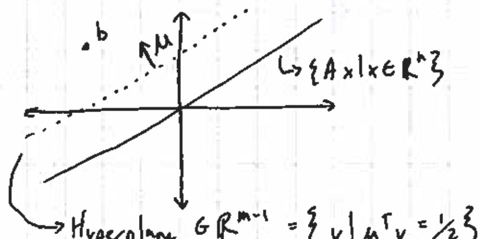
Q: How do we show that $C_1 \cap C_2 = \emptyset$ for some sets C_1, C_2?

Let's consider a simple example:
How do we show that $Ax = b$ has no solution, or equivalently that $\{b\} \cap \{Ax \mid x \in \mathbb{R}^n\} = \emptyset$?
Here $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$.

One way to certify that $Ax = b$ has no solution is to provide a $\mu \in \mathbb{R}^m$ such that: $A^T\mu = 0$; $b^T\mu = 1$

Such a $\mu \in \mathbb{R}^m$ certifies infeasibility because

$$\mu^T(Ax) = (A^T\mu)^T x = 0 \quad \forall x \in \mathbb{R}^n$$
$$\mu^T b = 1$$



Q: If $Ax = b$ has no solution, can I always find a μ s.t. $A^T\mu = 0$, $b^T\mu = 1$?

A: yes! (Proof is HW) [Fredholm's alternative]

10/7 <u>L3:</u>

Last time:

$$p^* = \inf_{x\in\mathbb{R}^n} f(x) \quad \text{s.t. } x \in S \quad \xRightarrow{\text{reformulation}} \quad p^* = \inf_{\substack{x\in\mathbb{R}^n \\ t\in\mathbb{R}}} t \quad \text{s.t. } (x,t) \in \text{epi}(f) \cap \{S \times \mathbb{R}\}$$

$f: \mathbb{R}^n \to \mathbb{R}$, $S \subseteq \mathbb{R}^n$

$$\text{epi}(f) = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \le t\}$$

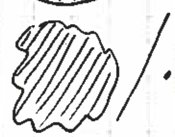
$$\therefore p^* = \sup_{\gamma\in\mathbb{R}} \gamma \quad \text{s.t. } \text{epi}(f) \cap \{(x,t) \mid x \in S, t < \gamma\} = \emptyset$$

Main Q: When can we <u>certify</u> that 2 sets have an empty intersection?

Goal: Generalize the linear algebra case of finding μ above to more general cases. If we want to certify that a point does not belong to a set (more general than a subspace), how do we go about this?

Ex:

Cannot draw a hyperplane that <u>separates</u> the point & the set



Can draw —/— because the point is not contained within the convex hull.

A: We have a relatively complete answer when the set is a <u>convex</u> set.

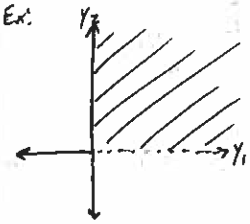
10/7 L3 cont'd:

<u>Definition</u>: A set $C \subseteq \mathbb{R}^n$ is a <u>convex set</u> if $\forall x^{(1)}, x^{(2)} \in C$ & $\forall \eta \in [0,1]$, we have that
$$\eta x^{(1)} + (1-\eta) x^{(2)} \in C$$

Ex:

convex. not convex.

Ex:

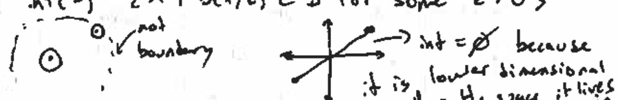


$\mathcal{Y} = \{(y_1, y_2) \mid y_1, y_2 > 0\}$
$\cup \{(y_1, y_2) \mid y_1 = 0, y_2 \geq 0\}$
Neither open nor closed but still convex.

Remarks:
- If $C_1 \subseteq \mathbb{R}^n$, $C_2 \subseteq \mathbb{R}^m$ are convex sets, then $C_1 \times C_2 = \{(x,y) \mid x \in C_1, y \in C_2\}$ is a convex set in $\mathbb{R}^{n \times m}$
- If $C_1, C_2 \subseteq \mathbb{R}^n$ are convex sets, then $C_1 + C_2 = \{x + y \mid x \in C_1, y \in C_2\}$ is a convex set in $\mathbb{R}^n$ [This operation is called the Minkowski sum]
- If $C \subseteq \mathbb{R}^n$ is a convex set & $f: \mathbb{R}^n \mapsto \mathbb{R}^m$ is an affine function, then $f(C)$ is a convex set in $\mathbb{R}^m$
- If $\{C_i\}_{i \in I}$ is any collection of convex subsets of $\mathbb{R}^n$, then $\bigcap_{i \in I} C_i$ is a convex set in $\mathbb{R}^n$

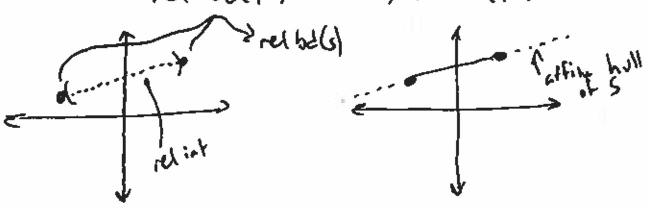
Note: The empty set is taken to be convex by convention

<u>Definition</u>: Let $S \subseteq \mathbb{R}^n$ be any set. The <u>interior</u> of S is the following subset:
$$\text{int}(S) = \{x \mid B(x,\epsilon) \subset S \text{ for some } \epsilon > 0\}$$



<u>Definition</u>: Let $S \subseteq \mathbb{R}^n$ be any set. The <u>relative interior</u> of S is the following subset:
$$\text{relint}(S) = \{x \mid B(x,\epsilon) \cap \text{aff}(S) \subset S \text{ for some } \epsilon > 0\}$$
where aff(S) is the affine hull of S, or the smallest affine space containing S (subspace + translation)

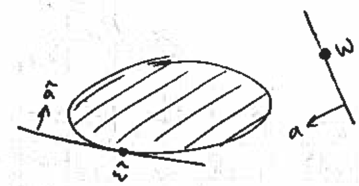
<u>Definition</u>: Let $S \subseteq \mathbb{R}^n$ be any set. The <u>relative boundary</u> of S is defined as:
$$\text{rel bd}(S) = \text{cl}(S) / \text{relint}(S)$$



<u>Definition</u>: A <u>hyperplane</u> in $\mathbb{R}^n$ is a set of the following form for $a \in \mathbb{R}^n / \{0\}$, $b \in \mathbb{R}$:
$$\{x \mid a^T x = b\}$$
normal vector at origin ↑ how much to shift in the direction of a

$\mathcal{H}$ is an affine space in $\mathbb{R}^{n-1}$

<u>Supporting Hyperplane Theorem</u>:
Let $C \subset \mathbb{R}^n$ be a convex set & let $w \notin \text{rel int}(C)$ ($w \in \mathbb{R}^n$). Then there exists $a \in \mathbb{R}^n / \{0\}$ & $b \in \mathbb{R}$ such that
- $a^T w = b$
- $\inf_{x \in C} a^T x \geq b$



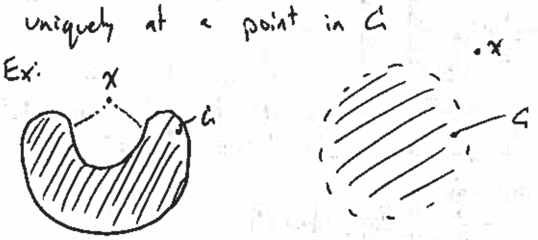
The set C is "bounded below" by b.

Remarks:
- If $w \in \text{rel bd}(C)$, then the hyperplane that we obtain as a consequence of the supporting hyperplane theorem is called a hyperplane supporting C at w.

10/9 L4:

<u>Theorem</u>: Projection onto Convex Sets
Let $C \subseteq \mathbb{R}^n$ be a closed, convex set. For any $x \in \mathbb{R}^n$, consider the following optimization problem:
$$\inf_{z \in \mathbb{R}^n} \|x - z\|_2 \quad \text{s.t. } z \in C$$
The optimal value of this problem is attained uniquely at a point in C

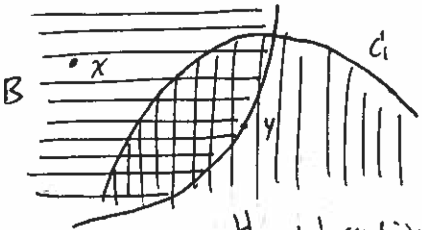
Ex:



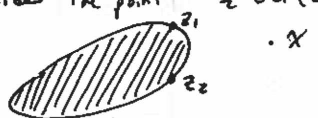
Uniqueness might not hold if C isn't convex

Attainment of optimal values could be problematic if C isn't closed

<u>Proof</u>: Fix any $y \in C$. Then we have that
$$\inf_{z \in \mathbb{R}^n} \|x - z\| \quad \text{s.t. } z \in C$$
is equivalent to
$$\inf_{z \in \mathbb{R}^n} \|x - z\| \quad \text{s.t. } z \in C \cap B(x, \|x - y\|)$$



The set $C \cap B(x, \|x-y\|)$ is closed (because the intersection of two closed sets is closed) and is convex (because the intersection of two convex sets is convex). Further, $C \cap B(x, \|x-y\|) \subseteq B(x, \|x-y\|)$ and $\therefore$ $C \cap B(x,\|x-y\|)$ is bounded. As the function $\|x-z\|$ is continuous in z, we have that the optimal value is attained. Concerning the question of uniqueness, suppose for the sake of contradiction that $\exists\, z_1, z_2 \in C$ with $z_1 \neq z_2$ such that both z_1 and z_2 are optimal solutions. Consider the point $\frac{z_1+z_2}{2} \in C$ (due to convexity of C)



$$\left\|\frac{z_1+z_2}{2} - x\right\|^2 = \left\|\left(\frac{z_1-x}{2}\right) + \left(\frac{z_2-x}{2}\right)\right\|^2 = \frac{1}{2}\|z_1-x\|^2 + \frac{1}{2}\|z_2-x\|^2 - \left\|\left(\frac{z_1-x}{2}\right) - \left(\frac{z_2-x}{2}\right)\right\|^2$$

~ $\boxed{\|u+v\|^2 + \|u-v\|^2 = 2\|u\|^2 + 2\|v\|^2}$

$$= \frac{1}{2}\|z_1-x\|^2 + \frac{1}{2}\|z_2-x\|^2 - \left\|\frac{z_1-z_2}{2}\right\|^2$$

This gives us the desired contradiction, as

$$\left\|\frac{z_1+z_2}{2} - x\right\| < \|z_1 - x\| \quad \text{unless } z_1 = z_2$$ ∎

<u>Proposition:</u> Let $C \subseteq \mathbb{R}^n$ be a convex set & let $x \in \mathbb{R}^n$ be any point. ~~Then we have that~~ Further, let $z^* \in cl(C)$ be the closest point to x in $cl(C)$. Then we have that

$$(x-z^*)^T(z-z^*) \leq 0 \quad \forall z \in C$$

~ must form angle larger than 90°

Proof: For any $z \in C$, we have that

$$\|x-z\|^2 \geq \|x-z^*\|^2$$

$$\Rightarrow \|x-z^*\|^2 \leq \|x-z\|^2$$
$$= \|(x-z^*) + (z^*-z)\|^2$$
$$= \|x-z^*\|^2 + \|z-z^*\|^2 + 2\langle x-z^*, z^*-z\rangle$$
$$\Rightarrow 2(x-z^*)^T(z-z^*) \leq \|z^*-z\|^2$$
$$\Rightarrow (x-z^*)^T(z-z^*) \leq 0 \quad \rightarrow \text{result from analysis}$$

[Suppose $z = z^*$. Then we are done. Suppose $z \neq z^*$. Then the inequality $2(x-z^*)^T(z-z^*) \leq \|z^*-z\|^2$ can be rewritten as

$$2(x-z^*)\left(\frac{z-z^*}{\|z-z^*\|}\right) \leq \|z-z^*\|.$$

As we take $z \to z^*$, we have that the LHS doesn't change but the RHS goes to 0. Hence,

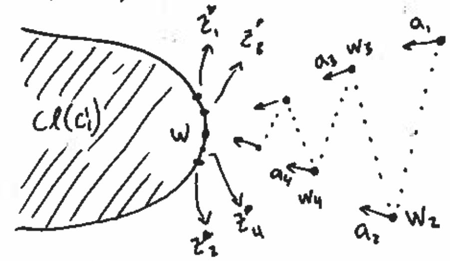
$$(x-z^*)^T\left(\frac{z-z^*}{\|z-z^*\|}\right) \leq 0$$

~~hence~~ which implies that

$$(x-z^*)^T(z-z^*) \leq 0.$$ ∎

<u>Proof of the Supporting Hyperplane Theorem:</u>

As $w \notin \text{rel Int}(C)$, we have that there exists a sequence $w_k \to w$ such that $w_k \notin cl(C)$ for each k. For each k, let $a_k = \frac{z_k^* - w_k}{\|z_k^* - w_k\|}$ where z_k^* is the closest point to w_k in $cl(C)$. (As $w_k \notin cl(C)$, we have that $\|z_k^* - w_k\| \neq 0$). As the sequence $\{a_k\}$ is bounded, $\exists$ a convergent subsequence $\{a_i\}_{i \in I}$ with $a_i \to a^*$.



For each $i \in I$, we have that

$$(w_i - z_i^*)^T(z - z_i^*) \leq 0 \quad \forall z \in C \quad \sim \text{from before}$$

In other words,

$$-a_i^T(z - z_i^*) \leq 0 \quad \forall z \in C$$
$$\Longrightarrow a_i^T z \geq a_i^T z_i^* \quad \forall z \in C$$
$$a_i^T z_i^* = a_i^T(z_i^* - w) + a_i^T w_i$$
$$\Rightarrow a_i^T w_i < a_i^T z_i^* \quad \text{as } a_i^T(z_i^* - w_i) > 0.$$

Hence, we have that

$$a_i^T z > a_i^T w_i \quad \forall z \in C$$

Letting $a_i \to a^*$ & $w_i \to w$, we have that

$$a^{*T} z \geq a^{*T} w \quad \forall z \in C$$

We have that $a^* \neq 0$ as it is the limit of a sequence of unit vectors. Setting $b = a^{*T}w$, we are done ∎

10/14 <u>L5:</u>

Recap:
- $p^* = \inf_{x \in S} f(x)$ s.t. $x \in S$
- Focus on $\mathcal{E} = \mathbb{R}^n$ [continuous optimization problems]
- Reformulate computing p^* as minimizing upper bound on p^*
- $p^* = \inf_{x \in \mathbb{R}^n,\, t \in \mathbb{R}} t$ s.t. $(x,t) \in epi(f)$, $(x,t) \in S \times \mathbb{R}$
- Reformulate computing p^* as maximizing lower bound of p^*
- $p^* = \sup_{\gamma \in \mathbb{R}} \gamma$ s.t. $epi(f) \cap \{(x,t) \mid x \in S, t < \gamma\} = \emptyset$

  ↳ This is the dual problem.

Q: When can we certify that two sets have an empty intersection?

A: Use hyperplanes!

Along the way, we defined convex sets

We will now build on projection onto convex sets theorem and supporting hyperplane theorem to completely answer our question.

10/14 L5 cont'd:

Theorem: [Separating Hyperplane]

Let $C_1, C_2 \subseteq \mathbb{R}^n$ be two non-empty convex sets such that $ri(C_1)\cap ri(C_2)=\emptyset$. Then, there exists $a\in\mathbb{R}^n/\{0\}$ such that:

$$\inf_{x\in C_1} a^Tx \ge \sup_{y\in C_2} a^Ty$$

Proof: Consider the set $C=C_1-C_2=\{x-y\,|\,x\in C_1, y\in C_2\}$. C is convex because $-C_2$ is convex (affine map applied to a convex set) & $C_1+(-C_2)$ is convex (as it is the Minkowski sum of convex sets). Finally, $ri(C)=ri(C_1)-ri(C_2)$. As $ri(C_1)\cap ri(C_2)=\emptyset$, we have that $0\notin ri(C)$. From the supporting hyperplane theorem, we have that there exists $a\in\mathbb{R}^n/\{0\}$ such that: $\inf_{z\in C} a^Tz \ge a^T0=0$. a^Tz for $z\in C$ can be decomposed as a^Tx-a^Ty for $x\in C_1$ & $y\in C_2$, which implies that

$$\inf_{x\in C_1, y\in C_2} a^T(x-y)\ge 0 \Leftrightarrow \inf_{x\in C_1} a^Tx + \inf_{y\in C_2} -a^Ty \ge 0$$

$$\Leftrightarrow \inf_{x\in C_1} a^Tx \ge \sup_{y\in C_2} a^Ty \quad \blacksquare$$

Remark: A hyperplane satisfying the conclusion of this theorem is called a separating hyperplane

Returning to our dual formulation of an optimization problem:

$$p^* = \sup_{\gamma\in\mathbb{R}} \gamma \quad \text{s.t. } epi(f)\cap\{(x,t)\,|\,x\in S, t<\gamma\}=\emptyset$$

In order to apply the separating hyperplane theorem in this context, we need to understand when

(i) $epi(f)$ is a convex set

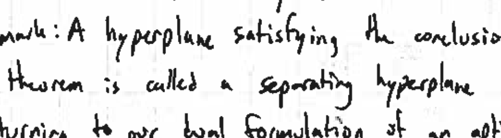
(ii) $\{(x,t)\,|\,x\in S, t<\gamma\}$ is a convex set

For (ii), one can check that $\{(x,t)\,|\,x\in S, t<\gamma\}$ is a convex set for a fixed $\gamma\in\mathbb{R}$ if & only if $S\subseteq\mathbb{R}^n$ is convex.

(i) leads to a central idea in continuous optimization

Convexity of a function:

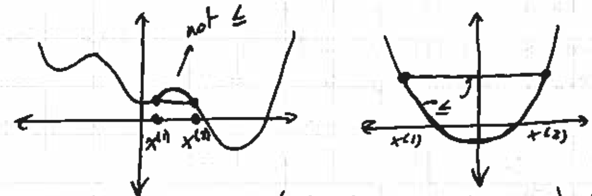
Definition: A function $f:\mathbb{R}^n\to\mathbb{R}$ is a convex function if the epigraph, $epi(f)$, is a convex set in $\mathbb{R}^{n+1}$



Proposition: A function $f:\mathbb{R}^n\to\mathbb{R}$ is convex if & only if $\forall\, x^{(1)}, x^{(2)}\in\mathbb{R}^n$ & $\forall\, \eta\in[0,1]$ we have that

$$f(\eta x^{(1)}+(1-\eta)x^{(2)}) \le \eta f(x^{(1)})+(1-\eta)f(x^{(2)})$$

Example:



Proof: $(\Leftarrow)$ Fix any $(x^{(1)},t_1),(x^{(2)},t_2)\in epi(f)$ & any $\eta\in[0,1]$. We need to show that $\eta(x^{(1)},t_1)+(1-\eta)(x^{(2)},t_2)\in epi(f)$. We are given that $f(\eta x^{(1)}+(1-\eta)x^{(2)}) \le \eta f(x^{(1)})+(1-\eta)f(x^{(2)})$. Further, $f(x^{(1)})\le t_1$ & $f(x^{(2)})\le t_2$ as $(x^{(1)},t_1),(x^{(2)},t_2)\in epi(f)$. As $\eta\in[0,1]$, we can conclude that

$$f(\eta x^{(1)}+(1-\eta)x^{(2)}) \le \eta t_1+(1-\eta)t_2$$
$$\Rightarrow (\eta x^{(1)}+(1-\eta)x^{(2)}, \eta t_1+(1-\eta)t_2)\in epi(f).$$

$(\Rightarrow)$ Fix any $x^{(1)},x^{(2)}\in\mathbb{R}^n$ & $\eta\in[0,1]$. We have that $(x^{(1)},f(x^{(1)})),(x^{(2)},f(x^{(2)}))\in epi(f)$ by definition, and that $epi(f)$ is convex. Therefore,

$$(\eta x^{(1)}+(1-\eta)x^{(2)}, \eta f(x^{(1)})+(1-\eta)f(x^{(2)}))\in epi(f)$$
$$\Rightarrow f(\eta x^{(1)}+(1-\eta)x^{(2)}) \le \eta f(x^{(1)})+(1-\eta)f(x^{(2)}) \quad \blacksquare$$

Some properties of convex functions:

- Let $f:\mathbb{R}^n\to\mathbb{R}$ be a convex function. For any $\alpha\ge 0$ the function $g(x)=\alpha f(x)$ is a convex function.
- Let $f_1,f_2:\mathbb{R}^n\to\mathbb{R}$ be convex functions. Then the function $g(x)=f_1(x)+f_2(x)$ is a convex function
- Let $f:\mathbb{R}^n\to\mathbb{R}$ be a convex function & let $g:\mathbb{R}^m\to\mathbb{R}^n$ be an affine function. Then the function $h(y)=f(g(y))$ from $\mathbb{R}^m$ to $\mathbb{R}$ is a convex function.
- Let $\{f_i\}_{i\in I}$ be a collection such that $f_i:\mathbb{R}^n\to\mathbb{R}$ is convex for each $i\in I$. Then we have that $g(x)=\sup_{i\in I} f_i(x)$ is a convex function.
  ↳ pointwise supremum of all f_i's.
  ↳ intersection of epigraphs.

Examples:
- $f(x)=x^2$, $x\in\mathbb{R}$ is convex
- $f(x)=x^TQx+b^Tx$, $Q\succeq 0$, $b\in\mathbb{R}^n$, $x\in\mathbb{R}^n$

As $Q\succeq 0$, we have that $Q=M^TM$ for some $M\in\mathbb{R}^{n\times n}$
$\Rightarrow x^TQx=(Mx)^TMx$. This is the affine function $x\mapsto Mx$ composed with the function $x\mapsto x^Tx$

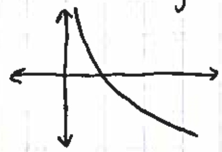
As the function $x_1^2+\dots+x_n^2$ is convex (x^2 is convex & sum of convex functions is convex), we can conclude that $f(x)=x^TQx+b^Tx$ is convex.

10/14 L5 cont'd:

Remarks: · The negative of a convex function is called a concave function

· On some occasions, we'll consider functions with domain being a subset of $\mathbb{R}^n$. All of our proceeding development goes through in largely unchanged manner.

Ex: $f(x) = -\log(x)$, $x \in \mathbb{R}$ $x > 0$



In such cases, we require the domain to be a convex set in $\mathbb{R}^n$.

10/16 L6:

Definition: An optimization problem with objective function $f: \mathbb{R}^n \to \mathbb{R}$ & constraint set $S \subseteq \mathbb{R}^n$ is a convex optimization problem if f is a convex function & S is a convex set

Remarks: Consider the dual problem where we require ways to certify that the intersection of two sets is empty. Our development thus far with the supporting/separating hyperplane theorems concerns convex sets. Therefore it is natural to consider convex optimization problems as we've just defined.

Theorem: [Duality for Convex Optimization]

Consider an optimization problem with objective function $f: \mathbb{R}^n \to \mathbb{R}$ & constraint set $S \subseteq \mathbb{R}^n$. If f is a convex function & S is a convex set, then the dual problem can be reformulated as follows:

$$p^* = \sup_{\substack{\gamma \in \mathbb{R} \\ g \in \mathbb{R}^n \\ \delta \in \mathbb{R}}} \gamma \quad \text{s.t. } \operatorname{epi}(f) \subseteq \{(x,t) \mid g^T x - t \le \delta\}$$
$$\{(x,t) \mid x \in S, t < \gamma\} \subseteq \{(x,t) \mid g^T x - t > \delta\}$$

*Note: This is only a valid reformulation if f and S are convex. Otherwise a hyperplane might intersect $\operatorname{epi}(f)$.

Proof: Fix $\gamma \in \mathbb{R}$ & consider the condition:

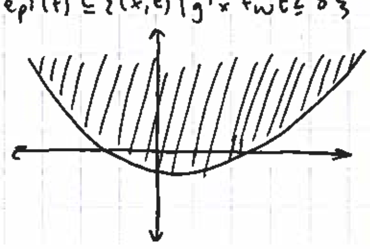
$$\operatorname{epi}(f) \cap \{(x,t) \mid x \in S, t < \gamma\} = \emptyset \quad (*)$$

As f is a convex function, the set $\operatorname{epi}(f)$ is a convex set. As $S \subseteq \mathbb{R}^n$ is a convex set & as $\{t \mid t < \gamma\}$ is a convex set, we have that $S \times \{t \mid t < \gamma\} = \{(x,t) \mid x \in S, t < \gamma\}$ is also a convex set. As these two sets have an empty intersection based on the condition $(*)$, we can appeal to the separating hyperplane theorem to conclude that there exists $(g, w) \in \mathbb{R}^n \times \mathbb{R}$ with $(g,w) \ne 0$ & $\delta \in \mathbb{R}$ such that

$$\operatorname{epi}(f) \subseteq \{(x,t) \mid g^T x + wt \le \delta\} \text{ and}$$
$$\{(x,t) \mid x \in S, t < \gamma\} \subseteq \{(x,t) \mid g^T x + wt \ge \delta\}.$$

Let's first consider the condition

$$\operatorname{epi}(f) \subseteq \{(x,t) \mid g^T x + wt \le \delta\}$$



Suppose for the sake of contradiction that $w = 0$. Then $g \ne 0$ & the set $\{x \mid g^T x \le \delta\}$ defines a halfspace in $\mathbb{R}^n$. But the domain of f is all of $\mathbb{R}^n$, which gives us a contradiction. Further, if $w > 0$, we have that wt can be arbitrarily large for $(x,t) \in \operatorname{epi}(f)$, and hence violate the inequality $g^T x + wt \le \delta$. Consequently, $w < 0$ & without loss of generality we can rescale it to be equal to -1. Thus, we have that

$$\operatorname{epi}(f) \subseteq \{(x,t) \mid g^T x - t \le \delta\}$$
$$\{(x,t) \mid x \in S, t < \gamma\} \subseteq \{(x,t) \mid g^T x - t \ge \delta\}$$

Considering the second condition, we need to show that the inequality $g^T x - t \ge \delta$ is actually a strict one. Suppose for the sake of contradiction that $\exists (\tilde{x}, \tilde{t})$ such that $\tilde{x} \in S$ & $\tilde{t} < \gamma$ with $g^T \tilde{x} - \tilde{t} = \delta$. As $\tilde{t} < \gamma$, we have that there exists $t' \in (\tilde{t}, \gamma)$ for which $g^T \tilde{x} - t' < \delta$. But $(\tilde{x}, t') \in \{(x,t) \mid x \in S, t < \gamma\}$ which gives the required contradiction. To conclude, we have that

$$\operatorname{epi}(f) \subseteq \{(x,t) \mid g^T x - t \le \delta\}$$
$$\{(x,t) \mid x \in S, t < \gamma\} \subseteq \{(x,t) \mid g^T x - t > \delta\}$$

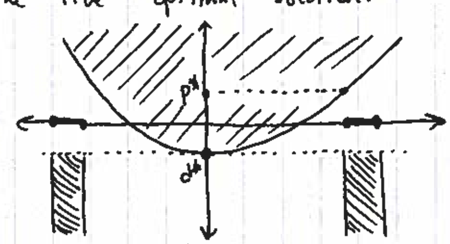
Thus, if $\operatorname{epi}(f) \cap \{(x,t) \mid x \in S, t < \gamma\} = \emptyset$, we have that

$$\operatorname{epi}(f) \subseteq \{(x,t) \mid g^T x - t \le \delta\}$$
$$\{(x,t) \mid x \in S, t < \gamma\} \subseteq \{(x,t) \mid g^T x - t > \delta\}$$ for some $g \in \mathbb{R}^n$ and $\delta \in \mathbb{R}$. In the reverse direction, whenever the latter condition holds, we have that $\operatorname{epi}(f) \cap \{(x,t) \mid x \in S, t < \gamma\} = \emptyset$. Hence, the dual problem can be rewritten as:

$$p^* = \sup_{\substack{\gamma \in \mathbb{R} \\ g \in \mathbb{R}^n \\ \delta \in \mathbb{R}}} \gamma \quad \text{s.t. } \operatorname{epi}(f) \subseteq \{(x,t) \mid g^T x - t \le \delta\}$$
$$\{(x,t) \mid x \in S, t < \gamma\} \subseteq \{(x,t) \mid g^T x - t > \delta\} \quad \blacksquare$$

Remarks:

· Suppose our optimization problem is not convex. Then we have that d^*, the solution to the upper dual problem with non convex problem statement acts as a lower bound of p^*, the true optimal solution.



As we search over those values of γ for which $\operatorname{epi}(f)$ & $\{(x,t) \mid x \in S, t < \gamma\}$ have an empty intersection that can be certified via a separating hyperplane, we have that $d^* \le p^*$. This fact is known as weak duality. When $d^* = p^*$ (such as in convex optimization problems), we say that strong duality holds.

[Recall that $p^* = \inf_{x \in \mathbb{R}^n} f(x)$ s.t. $x \in S$]

10/16 L6 cont'd:

Proposition: Consider an optimization problem with a convex objective function $f: \mathbb{R}^n \to \mathbb{R}$ & a convex constraint set $S \subseteq \mathbb{R}^n$: $p^* = \inf_{x \in \mathbb{R}^n} f(x)$ s.t. $x \in S$.

If there exists an optimal solution x^*, i.e. $x^* \in S$ and $p^* = f(x^*)$. Then there exists $g \in \mathbb{R}^n$ & $\delta \in \mathbb{R}$ such that $g^T x - t = \delta$ is a supporting hyperplane of epi(f) & of the set $\{(x,t) \mid x \in S, t < p^*\}$ at the point (x^*, p^*) $\left(= (x^*, f(x^*))\right)$.

10/21 L7:

- Specializing things when the optimal solution exists
- Obtain consequences in terms of functions from $\mathbb{R}^n$ to $\mathbb{R}$ and of sets in $\mathbb{R}^n$

Proposition: Suppose we have an optimization problem with objective function $f: \mathbb{R}^n \to \mathbb{R}$ and a constraint set $S \subseteq \mathbb{R}^n$. [This problem is not necessarily convex.] If there exists $g \in \mathbb{R}^n$, $\delta \in \mathbb{R}$ & $x^* \in S$ so that:

- $\text{epi}(f) \subseteq \{(x,t) \mid g^T x - t \leq \delta\}$
- $\{(x,t) \mid x \in S, t < f(x^*)\} \subseteq \{(x,t) \mid g^T x - t > \delta\}$

Then x^* is an optimal solution to our optimization problem.

Proof: As $x^* \in S$, we only need to show that $f(x^*) = p^*$ where p^* is the optimal value of our optimization problem:

$$p^* = \inf_{x \in \mathbb{R}^n} f(x) \quad \text{s.t. } x \in S$$

For any $\gamma \in \mathbb{R}$ such that

$$\text{epi}(f) \cap \{(x,t) \mid x \in S, t < \gamma\} = \emptyset$$

we have that $\gamma \leq p^*$. From the conditions given, we have that $f(x^*) \leq p^*$. As a result, we have that x^* is an optimal solution because $f(x) \geq p^*$ for all $x \in S$ ■

Proposition: Suppose we have an optimization problem with a convex ~~optimal~~ objective function $f: \mathbb{R}^n \to \mathbb{R}$ & a convex constraint set $S \subseteq \mathbb{R}^n$ such that x^* is an optimal solution, i.e., $x^* \in S$ & $f(x^*) = p^*$. Then there exists $g \in \mathbb{R}^n$, $\delta \in \mathbb{R}$ such that:

- $\text{epi}(f) \subseteq \{(x,t) \mid g^T x - t \leq \delta\}$
- $\{(x,t) \mid x \in S, t < p^*\} \subseteq \{(x,t) \mid g^T x - t > \delta\}$
- $g^T x^* - f(x^*) = g^T x^* - p^* = \delta$

Proof: As epi(f) & $\{(x,t) \mid x \in S, t < p^*\}$ are convex sets with an empty intersection, we have from the separating hyperplane theorem that $\exists$ $g \in \mathbb{R}^n$, $\delta \in \mathbb{R}$ such that:

- $\text{epi}(f) \subseteq \{(x,t) \mid g^T x - t \leq \delta\}$
- $\{(x,t) \mid x \in S, t < p^*\} \subseteq \{(x,t) \mid g^T x - t > \delta\}$

[To obtain this conclusion from the supporting hyperplane theorem, we need to follow some steps from the strong duality theorem for convex problems]

As $(x^*, f(x^*)) \in \text{epi}(f)$, we have that $g^T x^* - f(x^*) \leq \delta$. Further, for any $t < p^*$, we have that $g^T x^* - t > \delta$, which implies that $g^T x^* - p^* \geq \delta$. But $f(x^*) = p^*$, and we can conclude that $g^T x^* - f(x^*) = g^T x^* - p^* = \delta$ ■

Remark: The hyperplane $g^T x - t = \delta$ is a supporting hyperplane to epi(f) & to $\{(x,t) \mid x \in S, t < p^*\}$

Minimization of a convex function:

Q: Suppose we have a convex function $f: \mathbb{R}^n \to \mathbb{R}$ and we want to minimize it:

$$p^* = \inf_{x \in \mathbb{R}^n} f(x)$$

What properties must hold at an optimal solution?

A: To address this question, we'll try to understand supporting hyperplanes to the epigraph of f.

Let $g \in \mathbb{R}^n$ & $\delta \in \mathbb{R}$ define a supporting hyperplane to epi(f) for a convex function f:

- $\text{epi}(f) \subseteq \{(x,t) \mid g^T x - t \leq \delta\}$
- $g^T x^* - f(x^*) = \delta$

That is $(x^*, f(x^*))$ lies in the hyperplane.

Thus, we have that $g^T x - t \leq g^T x^* - f(x^*)$ $\forall (x,t) \in \text{epi}(f)$

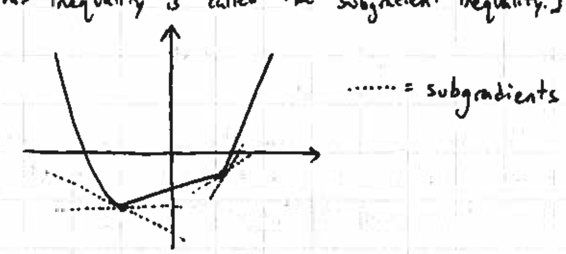
$\iff g^T x - f(x) \leq g^T x^* - f(x^*)$ $\forall x \in \mathbb{R}^n$

$\iff f(x) \geq f(x^*) + g^T (x - x^*)$ $\forall x \in \mathbb{R}^n$

Definition: Let $f: \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then $g \in \mathbb{R}^n$ is a subgradient of f at $x^* \in \mathbb{R}^n$ if

$$f(x) \geq f(x^*) + g^T (x - x^*) \quad \forall x \in \mathbb{R}^n$$

[This inequality is called the subgradient inequality.]



Definition: Let $f: \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then the subdifferential of f at $x^* \in \mathbb{R}^n$ is the set of all subgradients of f at x^* & is denoted as $\partial f(x^*)$

Remark: As there always exists supporting hyperplanes to epi(f) at $(x^*, f(x^*))$ for any $x^* \in \mathbb{R}^n$ from the supporting hyperplane theorem, the subdifferential is non-empty at any $x^* \in \mathbb{R}^n$.

10/21 L7 cont'd:

Proposition: Let $f:\mathbb{R}^n \to \mathbb{R}$ be a convex function. Then $\hat{x} \in \mathbb{R}^n$ is a minimizer of f if & only if $0 \in \partial f(\hat{x})$

Proof: Suppose 0 (zero) $\in \partial f(\hat{x})$. Then

$$f(x) \geq f(\hat{x}) + 0^T(x-\hat{x}) \quad \forall\, x \in \mathbb{R}^n$$

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

$\Rightarrow \hat{x}$ is a minimizer of f.

Conversely, suppose $\hat{x}$ is a minimizer of f. Then,

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

$\Rightarrow f(x) \geq f(\hat{x}) + 0^T(x-\hat{x}) \quad \forall\, x \in \mathbb{R}^n$

$\Rightarrow 0 \in \partial f(\hat{x})$ ∎

Proposition: Let $f:\mathbb{R}^n \to \mathbb{R}$ be a convex function. If f is differentiable at some $x^* \in \mathbb{R}^n$, then

$$\partial f(x^*) = \{\nabla f(x^*)\}$$

10/23 **L8: Continuing with subgradients:**

Proof: Let $g \in \partial f(x^*)$. We have that

$$f(x) \geq f(x^*) + g^T(x-x^*) \text{ for all } x \in \mathbb{R}^n$$

Set $x = x^* + hv$ for a fixed $v \in \mathbb{R}^n$ and any $h \in \mathbb{R}$. For $h>0$, we have that $\frac{f(x^*+hv)-f(x^*)}{h} \geq g^T v$

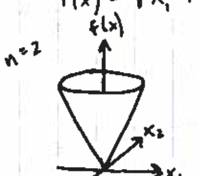
Consequently, we can conclude that $\lim_{h \searrow 0} \frac{f(x^*+hv)-f(x^*)}{h} \geq g^T v$ as f is differentiable at x^*. In the other direction, suppose $h<0$. Then we have that $\frac{f(x^*+hv)-f(x^*)}{h} \leq g^T v$. Again, we have that $\lim_{h \nearrow 0} \frac{f(x^*+hv)-f(x^*)}{h} \leq g^T v$. As f is differentiable at x^*, we have that $\lim_{h\to 0} \frac{f(x^*+hv)-f(x^*)}{h} = \nabla f(x^*)^T v$. Putting things together, we have that $\nabla f(x^*)^T v = g^T v$ for any $v \in \mathbb{R}^n \Rightarrow g = \nabla f(x^*)$ ∎

Ex: • $f(x) = |x|$, $x \in \mathbb{R}$

$$\partial f(x) = \begin{cases} 1 & x>0 \\ -1 & x<0 \\ [-1,1] & x=0 \end{cases}$$

At $x=0$, $g \in \mathbb{R}$ is a subgradient if it satisfies $|y| \geq g \cdot y \quad \forall y \in \mathbb{R}$

• $f(x) = \sqrt{x_1^2 + \cdots + x_n^2} = \|x\|$, $x \in \mathbb{R}^n$



$$\partial f(x) = \begin{cases} \frac{x}{\|x\|}, & x \neq 0 \\ \{g \mid \|g\| \leq 1\} & x = 0 \end{cases}$$

Remarks: $f:\mathbb{R}^n \to \mathbb{R}$ convex, $g \in \partial f(x^*)$

$\Rightarrow f(x) \geq f(x^*) + g^T(x-x^*)$

$\to f(x^*) + g^T(x-x^*)$ is an affine function that globally underestimates $f(x)$. If f is differentiable at x^*, we have that $\underbrace{f(x) \geq f(x^*) + \nabla f(x^*)^T(x-x^*)}$,

First order Taylor expansion of f at x^*

Supporting Hyperplanes to convex sets:

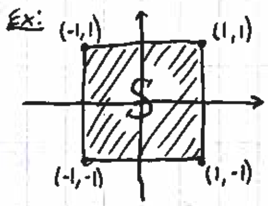
Let $S \subseteq \mathbb{R}^n$ be a convex set, and suppose $S \subseteq \{x \mid a^T x \geq b\}$, $a^T x^* = b$ for some $x^* \in S$

Therefore, $S \subseteq \{x \mid a^T x \geq a^T x^*\}$ —//—

In other words the minimum of the function $a^T x$ over $x \in S$ is attained at x^*. If the other inequality ($\leq$) were used, a maximum would be attained.

Definition: Let $S \subseteq \mathbb{R}^n$ & let $x^* \in S$. The normal cone of S at x^* is defined as the collection of linear functionals that attain their maximum over S at x^*, and this is denoted as $N_S(x^*)$

Ex:



$N_S((0,1)) = \{(0,a) \mid a \geq 0\}$

$N_S((0,0)) = \{(0,0)\}$

$N_S((1,1)) = \{(a,b) \mid a \geq 0 \; b \geq 0\}$

Proposition: Let $f:\mathbb{R}^n \to \mathbb{R}$ be a convex function & let $S \subseteq \mathbb{R}^n$ be a convex set. If f attains its minimum over S at $x^* \in S$, then there exists $g \in \mathbb{R}^n$ such that:

- $g \in \partial f(x^*)$
- $-g \in N_S(x^*)$

Proof: From the proposition proved previously (last lecture), we have that there exists $g \in \mathbb{R}^n$, $\delta \in \mathbb{R}$ such that:

- $\text{epi}(f) \subseteq \{(x,t) \mid g^T x - t \leq \delta\}$
- $\{(x,t) \mid x \in S \; t < p^*\} \subseteq \{(x,t) \mid g^T x - t > \delta\}$
- $g^T x^* - f(x^*) = g^T x - p^* = \delta$

From the first and third conclusions, we have that the hyperplane $g^T x - t = \delta$ is a supporting hyperplane of $\text{epi}(f)$ at the point $(x^*, f(x^*))$. Hence, $g \in \partial f(x^*)$. Similarly, we have from the second and third points that the hyperplane $g^T x - t = \delta$ is a supporting hyperplane to $\{(x,t) \mid x \in S, t < p^*\}$ at the point (x^*, p^*). From this we have that $g^T x - t > g^T x^* - p^*$ for all $x \in S$ and $t < p^*$. Taking $t \to p^*$, we have that $g^T x - p^* \geq g^T x^* - p^*$ for all $x \in S$. In other words, $-g^T x \leq -g^T x^* \quad \forall\, x \in S \Rightarrow -g \in N_S(x^*)$ ∎

Remark: Sometimes this condition is written as

$$0 \in \partial f(x^*) + N_S(x^*)$$

10/28 **L9: "All optimization problems are convex"**

Definition: Let $C \subseteq \mathbb{R}^n$ be any set. The convex hull of C is the set formed by taking convex combinations of elements of C:

$$\text{conv}(C) = \left\{ \sum_{i=1}^{k} \lambda_i x_i \;\middle|\; \sum_{i=1}^{k} \lambda_i = 1, \lambda_i \geq 0 \; \forall i, \; x_i \in C \; \forall i \right\}$$

Ex:

Remark: One can check that $\text{conv}(C)$ is indeed a convex set.

Proposition: Fix any set $S \subseteq \mathbb{R}^n$ & any vector $c \in \mathbb{R}^n$. Then the following optimization problem (not necessarily convex):

$$p^* = \inf_{x \in \mathbb{R}^n} c^T x \quad \text{s.t. } x \in S$$

can be reformulated as follows:

$$p^* = \inf_{x \in \mathbb{R}^n} c^T x \quad \text{s.t. } x \in \text{conv}(S)$$

Note: The reformulated problem is a convex optimization problem.

Proof: Let w^* be the optimal value of the reformulated problem:

$$w^* = \inf_{x \in \mathbb{R}^n} c^T x \quad \text{s.t. } x \in \text{conv}(S)$$

We need to show that $w^* = p^*$. In one direction, we have that $w^* \leq p^*$ because the feasible set in the reformulated problem contains the feasible set of the original problem.

...

10/28 L9 cont'd:

In the other direction, consider any $\tilde{x} \in \mathrm{conv}(S)$. We have from the definition of the convex hull that there exist:

- $\tilde{x}_i,\ i=1,\dots,k$ with $\tilde{x}_i \in S\ \forall i$
- $\lambda_i,\ i=1,\dots,k$ with $\sum_{i=1}^{k}\lambda_i = 1$, $\lambda_i \geq 0\ \forall i$

such that $\tilde{x} = \sum_{i=1}^{k}\lambda_i x_i$. We then have ~~that~~ the following sequence of inequalities/equalities:

$$c^T\tilde{x} = c^T\left[\sum_{i=1}^{k}\lambda_i\tilde{x}_i\right] = \sum_{i=1}^{k}\lambda_i(c^T\tilde{x}_i) \quad \text{[by linearity]}$$
$$\geq \min\{c^T\tilde{x}_1, \dots, c^T\tilde{x}_k\}$$

Hence, for any $\tilde{x} \in \mathrm{conv}(S)$ there exists $x' \in S$ such that $c^T\tilde{x} \geq c^T x'$. From this, we conclude that $w^* \geq p^*$. ■

Theorem: Let $S \subseteq \mathbb{R}^n$ be any set & $f: \mathbb{R}^n \to \mathbb{R}$. Then the following optimization problem: $p^* = \inf_{x\in\mathbb{R}^n} f(x)$ s.t. $x \in S$ can be reformulated as an optimization problem in $\mathbb{R}^{n+1}$ with a linear objective function and a convex constraint set, i.e. a convex optimization problem.

Proof: We can reformulate our optimization problem as:

$$p^* = \inf_{\substack{x\in\mathbb{R}^n \\ t\in\mathbb{R}}} t \quad \text{s.t. } (x,t) \in \mathrm{epi}(f),\ x \in S.$$

As the objective is a linear function in the decision variables (x,t), we can appeal to the preceding proposition to obtain a convex reformulation ■

Detour (brief): Complexity Theory

Q: What is a problem?

We will consider decision problems in which the answer is YES/NO as well as optimization problems (as we've discussed thus far).

Example: Knapsack Problem

Here we are given n items with associated values $v_i,\ i=1,\dots,n$ & costs $c_i,\ i=1,\dots,n$. We are given the total budget B and the goal is to find items that maximize overall value subject to the cost being $\leq B$:

$$p^* = \sup_{x\in\mathbb{R}^n} \sum_{i=1}^{n} v_i x_i \quad \text{s.t. } \sum_{i=1}^{n} c_i x_i \leq B,\ x_i \in \{0,1\}$$

This is the optimization version of the knapsack problem. The decision version is the following: "Does there exist a solution of items with $\sum_{i=1}^{n} c_i x_i \leq B$ & $\sum_{i=1}^{n} v_i x_i \geq \gamma$?" [$\gamma$ is an additional parameter]

"Definition": A "decision problem" is in the complexity class P if the problem can be solved using a number of operations that is at most a polynomial function of the input size.

Remark: In the knapsack problem, the input size is the number of items n.

- The natural brute-force approach to solve the problem is not a polynomial function of the input size (2^n)

"Definition": A "decision problem" is in the complexity class NP if a certificate of an affirmative solution can be verified in a number of operations that is at most a polynomial function of the input size

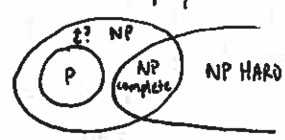
Remarks:
- Here an affirmative answer refers to an instance of a problem for which the answer is YES.
- In the knapsack example, a certificate of an affirmative answer is a selection of items with cost within the specifies budget & value greater than or equal to γ. Checking that this certificate actually certifies the answer YES can be accomplished in a linear number of operations. Hence, the decision version of the knapsack problem is in NP.

"Proposition": $P \subseteq NP$

Proof: Follows from the definitions ■

Checking whether this inclusion is strict is an open question [Most people believe that $P \neq NP$.]

"Definition": A "decision problem" is said to be NP-complete if it lies in NP and any other problem in NP can be reduced to it in a number of operations that is at most a polynomial function of input size.



10/30 L10:

- As yet, there is no algorithm that can solve an NP-complete problem in a polynomial # of operations. [This is the $P \overset{?}{=} NP$ question.]
- Q: Given that any NP-complete problem can be reformulated as a convex optimization problem (using the results from the last lecture), what does this mean for general optimization problems?

10/30 L10 cont'd:

Example: Knapsack (from last lecture)

- Knapsack is an NP-complete problem [Karp, 70's]
- We can reformulate Knapsack as a convex optimization:

$$p^* = \sup_{x\in\mathbb{R}^n} v^T x \quad \text{s.t. } x \in \text{conv}\{\tilde{x} \mid c^T\tilde{x} \le B,\ \tilde{x}_i \in \{0,1\}\ \forall i\}$$

Based on the preceding discussion, there is [likely] no procedure to solve this convex program in a polynomial (in n) number of operations [unless P=NP]. More generally, even though convex programs have nice mathematical properties (such as strong duality), it is in general not possible to solve them in a computationally efficient manner (i.e. via a procedure that requires at most a polynomial # of operations) [unless P=NP].

Q: Which convex problems can be solved efficiently?
[Without loss of generality, we can consider convex problems with a linear objective function.]

The preceding question reduces to understanding what makes a convex set easy or difficult to deal with.

Q: How do we describe a convex set?

Membership Oracle: $C \subseteq \mathbb{R}^n$ is a convex set

Input: $x \in \mathbb{R}^n$

Output: Yes if $x \in C$, No otherwise.

A stronger oracle than a membership oracle is a separation oracle: $C \subseteq \mathbb{R}^n$ is a convex set

Input: $x \in \mathbb{R}^n$

Output $\begin{cases} \text{Yes if } x \in C \\ a \in \mathbb{R}^n/\{0\} \text{ s.t. } a^T x \le \inf_{y \in C} a^T y \text{ if } x \notin C \end{cases}$

Theorem: [Khachiyan, 1970's]

Given a closed set $C \subseteq \mathbb{R}^n$ such that:

- there exists $r > 0$ with $B(x,r) \subseteq C$ for some x
- there exists $R > 0$ with $B(0,R) \supseteq C$

there is a procedure called the Ellipsoid method that can optimize a linear function over C using a number of calls to a separating oracle of C that is polynomial in $n, \frac{1}{r}, R$, & $\log\frac{1}{\epsilon}$ (ϵ is the desired tolerance to which we wish to compute the optimal value).

This result reduces the complexity of convex optimization to being able to come up with an efficient separation oracle for C. In particular, as general convex programs are not computationally efficient to solve, we should not expect to obtain efficient separation oracles for general convex sets.

We'll return to the question of how to describe a convex set keeping in mind that we want to ultimately obtain an efficient separating oracle.

Representation of Convex Sets:

Proposition: Let $C \subseteq \mathbb{R}^n$ be a closed convex set. Then C can be specified as an intersection of halfspaces:

$$C = \bigcap_{i \in \mathcal{I}} \{x \mid a_i^T x \le b_i\}$$

Proof: Suppose C has a non-empty interior (i.e. C is full-dimensional). Fix any $y \notin C$. Let $P_C(y)$ be the projection of y onto C [this projection is unique and moreover $\|P_C(y) - y\| > 0$]. & $P_C(y)$ lies on the boundary of C. By the supporting hyperplane theorem, there exists a hyperplane given by a normal vector $a_y \in \mathbb{R}^n$ s.t.

$$\sup_{x \in C} a_y^T x \le a_y^T P_C(y) < a_y^T y.$$

Thus $\bigcap_{y \notin C} \{x \mid a_y^T x \le a_y^T P_C(y)\}$ is equal to C. If C is not full-dimensional, we can add hyperplanes that restrict to the affine hull of C and repeat the above argument. ■

11/4 L11:

Recap: $p^* = \inf_{x \in \mathbb{R}^n} f(x)$ s.t. $x \in S$

Here $f: \mathbb{R}^n \to \mathbb{R}$, $S \subseteq \mathbb{R}^n$

can be reformulated as:

$$p^* = \inf_{x \in \mathbb{R}^n,\ t \in \mathbb{R}} t \quad \text{s.t. } (x,t) \in \text{epi}(f) \cap \{S \times \mathbb{R}\}$$

} This perspective corresponds to minimizing upper bounds on p^*

There is an alternative approach to computing p^* by maximizing lower bounds on p^*:

$$p^* = \sup_{\gamma \in \mathbb{R}} \gamma \quad \text{s.t. } \text{epi}(f) \cap \{(x,t) \mid x \in S, t < \gamma\} = \emptyset$$

Q: How can we certify that two sets have an empty intersection?

A: Separate using hyperplanes! This leads to the following dual problem:

$$d^* = \sup_{\substack{\gamma \in \mathbb{R} \\ g \in \mathbb{R}^n \\ \delta \in \mathbb{R}}} \gamma \quad \text{s.t. } \begin{array}{l} \text{epi}(f) \subseteq \{(x,t) \mid g^T x - t \le \delta\} \\ S \times (-\infty, \gamma) \subseteq \{(x,t) \mid g^T x - t > \delta\} \end{array}$$

In general, $d^* \le p^*$ and this inequality is known as weak duality. For convex optimization problems (f is a convex function, S is a convex set), strong duality holds based on the separation theorem i.e. $d^* = p^*$

Last week:

- How is a convex set described? (Since all optimization problems can be reformulated as convex ones.)
- A general answer is that any closed convex set can be expressed as an intersection of [a possibly infinite collection of] halfspaces.

Today (and the coming few lectures):

Q: How do we deal with an optimization problem as given? How do we think about duality in such contexts? In particular, what if we aren't allowed to perform (significant) reformulatio

To begin with let's consider the following problem:

$$p^* = \inf_{x \in \mathbb{R}^n} c^T x \quad \text{s.t. } x \in S_1,\ x \in S_2$$

Here the constraint set is $S = S_1 \cap S_2$. Let's consider the dual problem in terms of separating hyperplanes:

$$d^* = \sup_{\substack{\gamma \in \mathbb{R} \\ g \in \mathbb{R}^n \\ \delta \in \mathbb{R}}} \gamma \quad \text{s.t. } \begin{array}{l} \text{epi}(c^T x) \subseteq \{(x,t) \mid g^T x - t \le \delta\} \\ (S_1 \cap S_2) \times (-\infty, \gamma) \subseteq \{(x,t) \mid g^T x - t > \delta\} \end{array}$$

Note: If $S_1 \cap S_2$ is convex, then strong duality holds ($d^* = p^*$).

11/4 **L 11 cont'd:**

If we have a handle on the separating hyperplanes of S_1 & S_2 individually, how do we think about separating hyperplanes for $S_1 \cap S_2$?

Suppose $S_1 \subseteq \{x \mid g_1^T x \geq \delta_1\}$
$S_2 \subseteq \{x \mid g_2^T x \geq \delta_2\}$

Consider the set $\{x \mid (g_1+g_2)^T x \geq \delta_1 + \delta_2\}$. For any $\tilde{x} \in S_1 \cap S_2$ we have that $g_1^T \tilde{x} \geq \delta_1$ & $g_2^T \tilde{x} \geq \delta_2$. Therefore, $S_1 \cap S_2 \subseteq \{x \mid (g_1+g_2)^T x \geq \delta_1 + \delta_2\}$. Based on this approach for constructing halfspaces, we consider the following problem:

$$d' = \sup_{\substack{\gamma \in \mathbb{R} \\ \delta, \delta_1, \delta_2 \in \mathbb{R} \\ g, g_1, g_2 \in \mathbb{R}^n}} \gamma \quad \text{s.t.} \quad \begin{array}{l} \operatorname{epi}(c^T x) \subseteq \{(x,t) \mid g^T x - t \leq \delta\} \\ S_1 \subseteq \{x \mid g_1^T x \geq \delta_1\} \\ S_2 \subseteq \{x \mid g_2^T x \geq \delta_2\} \\ g = g_1 + g_2 \;;\; \delta + \gamma = \delta_1 + \delta_2 \end{array}$$

For any $g, g_1, g_2 \in \mathbb{R}^n$ and $\gamma, \delta, \delta_1, \delta_2 \in \mathbb{R}$ such that $S_1 \subseteq \{x \mid g_1^T x \geq \delta_1\}$, $S_2 \subseteq \{x \mid g_2^T x \geq \delta_2\}$, $g = g_1 + g_2$, $\delta + \gamma = \delta_1 + \delta_2$, we have that $(S_1 \cap S_2) \times (-\infty, \gamma) \subseteq \{(x,t) \mid g^T x - t > \delta\}$. Therefore, $d' \leq d^*$, and in turn $d' \leq p^*$. This approach of constructing separating hyperplanes of intersections of sets is called <u>Lagrange Duality</u> & the associated dual problem is called the <u>Lagrange Dual Problem</u>.

Q: When is $d' = p^*$? That is, when do we have strong duality holding with the Lagrange Dual problem?

We have that $d' \leq d^* \leq p^*$, so we need that both of these inequalities hold with equality.

- For $d^* = p^*$, we need that $S_1 \cap S_2$ is convex.
- For $d' = d^*$, we have the following proposition:

<u>Proposition:</u> Suppose $S_1, S_2 \subseteq \mathbb{R}^n$ are closed, convex sets such that $ri(S_1) \cap ri(S_2) \neq \emptyset$. Then for any $g \in \mathbb{R}^n$, $\delta \in \mathbb{R}$ such that $S_1 \cap S_2 \subseteq \{x \mid g^T x \geq \delta\}$, there exists $g_1, g_2 \in \mathbb{R}^n$, $\delta_1, \delta_2 \in \mathbb{R}$ such that:

- $S_1 \subseteq \{x \mid g_1^T x \geq \delta_1\}$ • $g = g_1 + g_2$
- $S_2 \subseteq \{x \mid g_2^T x \geq \delta_2\}$ • $\delta = \delta_1 + \delta_2$

Let's further simplify the Lagrange dual problem:

As $\operatorname{epi}(c^T x) \subseteq \{(x,t) \mid g^T x - t \leq \delta\}$, we have that: $\operatorname{epi}(f) \subseteq \{(x,t) \mid g^T x - c^T x \leq \delta\}$. As there is no restriction on x in the definition of $\operatorname{epi}(f)$, we have that $g = c$ and $\delta \geq 0$. Thus, the Lagrange dual problem for our original problem simplifies as:

$$d' = \sup_{\substack{\gamma \in \mathbb{R} \\ \delta, \delta_1, \delta_2 \in \mathbb{R} \\ g_1, g_2 \in \mathbb{R}^n}} \gamma \quad \text{s.t.} \quad \begin{array}{l} S_1 \subseteq \{x \mid g_1^T x \geq \delta_1\} \\ S_2 \subseteq \{x \mid g_2^T x \geq \delta_2\} \\ c = g_1 + g_2 \;;\; \delta + \gamma = \delta_1 + \delta_2 \\ \delta \geq 0 \end{array}$$

This can be rewritten as:

$$\boxed{d' = \sup_{\substack{g_1, g_2 \in \mathbb{R}^n \\ \delta_1, \delta_2 \in \mathbb{R}}} \delta_1 + \delta_2 \quad \text{s.t.} \quad \begin{array}{l} S_1 \subseteq \{x \mid g_1^T x \geq \delta_1\} \\ S_2 \subseteq \{x \mid g_2^T x \geq \delta_2\} \\ g_1 + g_2 = c \end{array}}$$

From the preceding proposition, we have that:

$d' = p^*$ if S_1 & S_2 are closed, convex, $ri(S_1) \cap ri(S_2) \neq \emptyset$.

One can develop Lagrange duality more generally in the following ways:

- The constraint set is $S = \bigcap_{i=1}^{k} S_i$ for $k > 2$. This case follows from what we've done by induction.
- The objective function is not linear. There are several ways to deal with this, one being absorbing the objective into the constraints and obtaining a linear objective.

11/6 **L12:**

Last Time: Lagrange Duality: Geometrically, this is an approach to obtain separating hyperplanes of intersections of sets by 'adding' separating hyperplanes for the individual sets.

- $p^* = \inf_{x \in \mathbb{R}^n} c^T x$ s.t. $x \in S_1$, $x \in S_2$ (*)

The Lagrange Dual of this problem is:

$$d' = \sup_{\substack{g_1, g_2 \in \mathbb{R}^n \\ \delta_1, \delta_2 \in \mathbb{R}}} \delta_1 + \delta_2 \quad \text{s.t.} \quad \begin{array}{l} S_1 \subseteq \{x \mid g_1^T x \geq \delta_1\} \\ S_2 \subseteq \{x \mid g_2^T x \geq \delta_2\} \\ g_1 + g_2 = c \end{array}$$

Proposition: Suppose $S_1, S_2 \subseteq \mathbb{R}^n$ are sets such that:

- S_1, S_2 are convex
- $ri(S_1) \cap ri(S_2) \neq \emptyset$

Then for any $g \in \mathbb{R}^n$ & $\delta \in \mathbb{R}$ such that $S_1 \cap S_2 \subseteq \{x \mid g^T x \geq \delta\}$ there exists $g_1, g_2 \in \mathbb{R}^n$ & $\delta_1, \delta_2 \in \mathbb{R}$ such that

- $S_1 \subseteq \{x \mid g_1^T x \geq \delta_1\}$ • $g = g_1 + g_2$
- $S_2 \subseteq \{x \mid g_2^T x \geq \delta_2\}$ • $\delta = \delta_1 + \delta_2$

Remarks:

- For the optimization problem (*), we have that weak duality holds in general for the Lagrange dual problem, i.e. $p^* \geq d'$. Under the conditions of the preceding proposition, we have that $p^* = d'$, i.e. strong duality holds.
- The condition that $ri(S_1) \cap ri(S_2) \neq \emptyset$ is called Slater's condition.
- All settings which mention dual problems, strong duality, etc... (unless otherwise stated) will be referring to Lagrange Duality.
- In this proposition, the convexity condition is typically the more important one in practice.

11/6 L12 cont'd:

Optimization problems in nonlinear programming form

$$p^* = \inf_{x\in\mathbb{R}^n} f_0(x) \quad \text{s.t. } f_i(x)\le 0,\ i=1,\dots,k$$
$$g_j(x)=0,\ j=1,\dots,m$$

Here $f_0, f_1, \dots, f_k, g_1, \dots, g_m : \mathbb{R}^n \to \mathbb{R}$

In this context, a convex optimization problem is one in which the f_i's are convex functions and the g_j's are affine functions. This is the perspective taken by Boyd & Vandenberghe and by Rockafellar.

Q: How do we think about Lagrange duality in this context? We'll consider the special case with only inequality constraints:

$$p^* = \inf_{x\in\mathbb{R}^n} f_0(x) \quad \text{s.t. } f_i(x)\le 0 \quad i=1,\dots,k$$

This problem can be reformulated as:

$$p^* = \inf_{x\in\mathbb{R}^n,\ t\in\mathbb{R}^k} f_0(x) \quad \text{s.t. } (x_1,t_1)\in epi(f_1)\ t_1\in(-\infty,0]$$
$$\vdots$$
$$(x_k,t_k)\in epi(f_k)\ t_k\in(-\infty,0]$$

This dual problem in terms of separating hyperplanes is:

$$d^* = \sup_{\gamma\in\mathbb{R},\ g\in\mathbb{R}^n,\ \delta\in\mathbb{R},\ \lambda\in\mathbb{R}^k} \gamma$$

s.t. $\{(x,t_0,t_1,\dots,t_k) \mid (x,t_0)\in epi(f_0)\} \subseteq \{(x,t_0,t_1,\dots,t_k)\mid g^Tx+\lambda^T t_{1:k}-t_0\le\delta\}$

$\{(x,t_0,t_1,\dots,t_k)\mid (x,t_i)\in epi(f_i)\ \&\ t_i\in(-\infty,0] \text{ for } i=1,\dots,k;\ t_0\in(-\infty,\gamma)\}$

$\subseteq \{(x,t_0,t_1,\dots,t_k)\mid g^Tx+\lambda^T t_{1:k}-t_0>\delta\}$

Remark: • $x, t_1, \dots, t_k$ are variables in the original problem. By the Lagrange duality principle, we'll search for separating hyperplanes involving the constraints individually & add them up.

• Before considering the Lagrange dual problem, we have strong duality, i.e. $d^*=p^*$, if f_0 is convex and the constraint set in $(x,t_{1:k})$ is convex (this would be true, if, for example, $f_1,\dots,f_k$ are also convex functions).

The Lagrange Dual problem is:

$$d' = \sup_{\gamma\in\mathbb{R};\ g,g_1,\dots,g_k,\tilde g_1,\dots,\tilde g_k\in\mathbb{R}^n;\ \lambda,\lambda_1,\dots,\lambda_k,\tilde\lambda_1,\dots,\tilde\lambda_k\in\mathbb{R}^k;\ \delta,\delta_1,\dots,\delta_k,\tilde\delta_1,\dots,\tilde\delta_k\in\mathbb{R}} \gamma$$

s.t. $\{(x,t_0,t_1,\dots,t_k)\mid (x,t_0)\in epi(f_0)\}\subseteq\{(x,t_0,t_1,\dots,t_k)\mid g^Tx+\lambda^T t_{1:k}-t_0\le\delta\}$

$\{(x,t_0,t_1,\dots,t_k)\mid (x,t_i)\in epi(f_i)\}\subseteq\{(x,t_0,t_1,\dots,t_k)\mid g_i^Tx+\lambda_i^T t_{1:k}\ge\delta_i\}\ i=1,\dots,k$

$\{(x,t_0,t_1,\dots,t_k)\mid t_i\le 0\}\subseteq\{(x,t_0,t_1,\dots,t_k)\mid \tilde g_i^Tx+\tilde\lambda_i^T t_{1:k}\ge\tilde\delta_i\}\ i=1,\dots,k$

$$g=g_1+g_2+\dots+g_k+\tilde g_1+\dots+\tilde g_k$$
$$\lambda=\lambda_1+\lambda_2+\dots+\lambda_k+\tilde\lambda_1+\dots+\tilde\lambda_k$$
$$\delta+\gamma=\delta_1+\delta_2+\dots+\delta_k+\tilde\delta_1+\dots+\tilde\delta_k$$

• From the condition involving $epi(f_0)$, we have that $\lambda=0$

• From the condition involving $t_i\le 0$ we have that $\tilde g_i=0\ \forall i$ & that $(\tilde\lambda_i)_j=0$ if $i\ne j$

• Similarly $(\lambda_i)_j=0$ if $i\ne j$

With this simplification, we have

$$d' = \sup_{\gamma\in\mathbb{R};\ g,g_1,\dots,g_k\in\mathbb{R}^n;\ \delta,\delta_1,\dots,\delta_k,\tilde\delta_1,\dots,\tilde\delta_k\in\mathbb{R};\ \bar\lambda,\bar{\bar\lambda}\in\mathbb{R}^k} \gamma$$

s.t. $epi(f_0)\subseteq\{(x,t_0)\mid g^Tx-t_0\le\delta\}$ $\qquad g=g_1+\dots+g_k$

$epi(f_i)\subseteq\{(x,t_i)\mid g_i^Tx+\bar\lambda_i t_i\ge\delta_i\}$ $\qquad 0=\bar\lambda+\bar{\bar\lambda}$

$\{t_i\mid t_i\le 0\}\subseteq\{t_i\mid \bar{\bar\lambda}_i t_i\ge\tilde\delta_i\}\ i=1,\dots,k$ $\qquad \delta+\gamma=\delta_1+\dots+\delta_k+\tilde\delta_1+\dots+\tilde\delta_k$

• From the condition on $t_i\le 0$ we have that $\bar{\bar\lambda}_i\le 0,\ i=1,\dots,k$ & that $\tilde\delta_i\le 0,\ i=1,\dots,k$. This leads to the following problem:

$$d' = \sup_{\gamma\in\mathbb{R};\ g,g_1,\dots,g_k\in\mathbb{R}^n;\ \bar\lambda\in\mathbb{R}^k;\ \delta,\delta_1,\dots,\delta_k\in\mathbb{R}} \gamma$$

s.t. $epi(f_0)\subseteq\{(x,t_0)\mid g^Tx-t_0\le\delta\}$ $\qquad g=g_1+\dots+g_k$

$epi(f_i)\subseteq\{(x,t_i)\mid g_i^Tx+\bar\lambda_i t_i\ge\delta_i\}$ $\qquad \bar\lambda_i\ge 0$

$i=1,\dots,k$ $\qquad \delta+\gamma\le\delta_1+\dots+\delta_k$

This can be expressed as:

$$d' = \sup_{g,g_1,\dots,g_k\in\mathbb{R}^n;\ \bar\lambda\in\mathbb{R}^k;\ \delta,\delta_1,\dots,\delta_k\in\mathbb{R}} \delta_1+\dots+\delta_k-\delta$$

s.t. $g^Tx-f_0(x)\le\delta\ \forall x\in\mathbb{R}^n$ $\qquad g=g_1+\dots+g_k$

$g_i^Tx+\bar\lambda_i f_i(x)\ge\delta_i\ \forall x\in\mathbb{R}^n,\ i=1,\dots,k$ $\qquad \bar\lambda_i\ge 0\ i=1,\dots,k$

• For each $x\in\mathbb{R}^n$ the constraints can be simplified as:

$$\delta_1+\dots+\delta_k-\delta\le(g_1+\dots+g_k)^Tx+\left(\sum_{i=1}^k\bar\lambda_i f_i(x)\right)+f_0$$
$$g=g_1+\dots+g_k;\ \bar\lambda_i\ge 0\quad i=1,\dots,k$$

Thus, we end up with

$$d' = \sup_{\bar\lambda\in\mathbb{R}^k} h(\bar\lambda) \quad \text{s.t. } \bar\lambda_i\ge 0\quad i=1,\dots,k$$

where $h(\bar\lambda)=\inf_{x\in\mathbb{R}^n} f_0(x)+\sum_{i=1}^k\bar\lambda_i f_i(x)$

This problem is called the Lagrange Dual problem in most textbooks and references. The function $h(\bar\lambda)$ is called the <u>dual function</u>.

11/8 L13:

Consider an optimization problem in the following form

$$(*)\quad p^*=\inf_{x\in\mathbb{R}^n} f_0(x)\quad \text{s.t. } f_i(x)\le 0\quad i=1,\dots,k$$
$$g_j(x)=0\quad j=1,\dots,m$$

Here $f_0,f_1,\dots,f_k,g_1,\dots,g_m:\mathbb{R}^n\to\mathbb{R}$

The Lagrange dual problem is:

$$d^*=\sup_{\lambda\in\mathbb{R}^k,\ \mu\in\mathbb{R}^m} h(\lambda,\mu)\quad \text{s.t. } \lambda_i\ge 0\quad i=1,\dots,k$$
$$h(\lambda,\mu)=\inf_{x\in\mathbb{R}^n} f_0(x)+\sum_{i=1}^k\lambda_i f_i(x)+\sum_{j=1}^m\mu_j g_j(x)$$

Here the function $h:\mathbb{R}^k\times\mathbb{R}^m\to\mathbb{R}$ is called the dual function.

In general, $d^*\le p^*$ and this is called weak (Lagrangian) duality.

Q: When does strong duality hold?

A: Under the following conditions, strong duality is garanteed to hold:

• $f_0,f_1,\dots,f_k:\mathbb{R}^n\to\mathbb{R}$ are all convex functions.

• $g_1,\dots,g_m:\mathbb{R}^n\to\mathbb{R}$ are all affine functions

• $\exists\tilde x\in\mathbb{R}^n$ s.t. $g_j(\tilde x)=0,\ j=1,\dots,m$ & $f_i(\tilde x)<0,\ i=1,\dots,k$

↳ This last condition is Slater's condition

Note: Strong duality can hold with the Lagrange dual problem even if the preceding conditions are not satisfied. These conditions are simply sufficient conditions for strong duality to hold with the Lagrange dual problem.

One further fact, which we note without proof is the following: If $f_0,f_1,\dots,f_k,g_1,\dots,g_m$ are all affine functions, then strong duality holds with the Lagrange dual problem under the following weaker requirement:

• $\exists\tilde x\in\mathbb{R}^n$ s.t. $g_j(\tilde x)=0,\ j=1,\dots,m$ $f_i(\tilde x)\le 0,\ i=1,\dots,k$

In an optimization problem of the form (*), the <u>Lagrangian</u> is the function defined as follows:

$$L(x,\lambda,\mu)=f_0(x)+\sum_{i=1}^k\lambda_i f_i(x)+\sum_{j=1}^k\mu_j g_j(x)$$

From this perspective, the dual function may be viewed as:

$h(\lambda,\mu)=\inf_{x\in\mathbb{R}^n} L(x,\lambda,\mu)$ & the dual problem may be viewed as:

$$d^*=\sup_{\lambda\in\mathbb{R}^k,\mu\in\mathbb{R}^m}\left(\inf_{x\in\mathbb{R}^n} L(x,\lambda,\mu)\right)\quad \text{s.t. } \lambda_i\ge 0\quad i=1,\dots,k$$

One can also check that: $p^*=\inf_{x\in\mathbb{R}^n}\left[\sup_{\lambda\in\mathbb{R}^k,\mu\in\mathbb{R}^m} L(x,\lambda,\mu)\right.$ s.t. $\lambda_i\ge 0$ $i=1,\dots$

To see this, note that

$$\sup_{\lambda\in\mathbb{R}^k,\ \mu\in\mathbb{R}^m} L(x,\lambda,\mu) \text{ s.t. } \lambda_i\ge 0\ i=1,\dots,k = \begin{cases} f_0(x) & \text{if } f_i(x)\le 0\ i=1,\dots,k;\ g_j(x)=0\ j=1,\dots,m \\ +\infty & \text{otherwise}\end{cases}$$

11/8 L13 cont'd:

Thus, weak duality with the Lagrange dual problem may also be viewed as follows:

$$p^* = \inf_{x \in \mathbb{R}^n} \sup_{\lambda \in \mathbb{R}^k, \mu \in \mathbb{R}^m, \lambda_i \geq 0\ i=1,\dots,k} L(x,\lambda,\mu)$$
$$\geq \sup_{\lambda \in \mathbb{R}^k, \mu \in \mathbb{R}^m, \lambda_i \geq 0\ i=1,\dots,k} \inf_{x \in \mathbb{R}^n} L(x,\lambda,\mu) = d^*$$

In general, we have the following result, which provides an alternative justification for weak duality with the Lagrange dual problem:

Proposition [Min-Max Inequality]:

Consider a function $f: \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ and any sets $\mathcal{U} \subseteq \mathbb{R}^p$, $\mathcal{V} \subseteq \mathbb{R}^q$. Then we have that:

$$\inf_{u \in \mathcal{U}} \sup_{v \in \mathcal{V}} f(u,v) \geq \sup_{v \in \mathcal{V}} \inf_{u \in \mathcal{U}} f(u,v).$$

Proof: Fix $\tilde{v} \in \mathcal{V}$. Then we have that $f(u,\tilde{v}) \leq \sup_{v \in \mathcal{V}} f(u,v)$ for each $u \in \mathcal{U}$. Thus we can conclude that $\inf_{u \in \mathcal{U}} f(u,\tilde{v}) \leq \inf_{u \in \mathcal{U}} \sup_{v \in \mathcal{V}} f(u,v)$. As $\tilde{v} \in \mathcal{V}$ is arbitrary, we have the desired result. ∎

Suppose we have an optimization problem of the form:

$$p^* = \inf_{x \in \mathbb{R}^n} f_0(x). \quad \text{s.t.} \quad \begin{array}{ll} f_i(x) \leq 0 & i=1,\dots,k \\ g_j(x) = 0 & j=1,\dots,m \end{array}$$

Suppose further that: $f_0, f_1, \dots, f_k: \mathbb{R}^n \to \mathbb{R}$ are convex & $g_1, \dots, g_m: \mathbb{R}^n \to \mathbb{R}$ are affine. Then strong duality holds (follows from Slater's condition). In addition, the primal and dual optimal values are attained, i.e. there exists optimal solutions of the primal and dual problems. What can we conclude? Suppose $x^* \in \mathbb{R}^n$ is a primal optimal solution & $(\lambda^*, \mu^*) \in \mathbb{R}^k \times \mathbb{R}^m$ are dual optimal solutions. Then we have that:

$$p^* = f_0(x^*) = \inf_{x \in \mathbb{R}^n} \sup_{\lambda \in \mathbb{R}^k, \lambda_i \geq 0\ i=1,\dots,k, \mu \in \mathbb{R}^m} L(x,\lambda,\mu)$$
$$\geq \sup_{\lambda \in \mathbb{R}^k, \lambda_i \geq 0\ i=1,\dots,k, \mu \in \mathbb{R}^m} \inf_{x \in \mathbb{R}^n} L(x,\lambda,\mu)$$
$$= \sup_{\lambda \in \mathbb{R}^k, \lambda_i \geq 0,\ i=1,\dots,k, \mu \in \mathbb{R}^m} h(\lambda,\mu) = h(\lambda^*,\mu^*) = d^*$$

When strong duality holds, $d^* = p^*$, thus we can conclude that $f_0(x^*) = \inf_{x \in \mathbb{R}^n} L(x,\lambda^*,\mu^*)$. Thus x^* is a minimizer of the function $f_0(x) + \sum_{i=1}^k \lambda_i^* f_i(x) + \sum_{j=1}^m \mu_j^* g_j(x)$.

This is a convex function in x. Hence,

$$0 \in \partial f_0(x^*) + \sum_{i=1}^k \lambda_i^* \partial f_i(x^*) + \sum_{j=1}^m \mu_j^* \partial g_j(x^*).$$

We also observe that $f_0(x^*) = f_0(x^*) + \sum_{i=1}^k \lambda_i^* f_i(x^*) + \sum_{j=1}^m \mu_j^* g_j(x^*)$

As x^* is a feasible point in the primal problem & (λ^*, μ^*) is feasible for the dual problem, we have that $f_i(x^*) \leq 0$ $i=1,\dots,k$; $g_j(x^*) = 0$ $j=1,\dots,m$; $\lambda_i^* \geq 0$ $i=1,\dots,k$

$\therefore \sum_{i=1}^k \lambda_i^* f_i(x^*) = 0$. But each $\lambda_i^* f_i(x^*) \leq 0$ based on feasibility of x^* & λ^*. Consequently, we conclude that $\lambda_i^* f_i(x^*) = 0$ $i=1,\dots,k$

To summarize, we have that

- $f_i(x^*) \leq 0$ $i=1,\dots,k$
- $g_j(x^*) = 0$ $j=1,\dots,m$
- $\lambda_i^* \geq 0$ $i=1,\dots,k$
- $\lambda_i^* f_i(x^*) = 0$ $i=1,\dots,k$
- $0 \in \partial f_0(x^*) + \sum_{i=1}^k \lambda_i^* \partial f_i(x^*) + \sum_{j=1}^m \mu_j^* \partial g_j(x^*)$

These five conditions are collectively called the KKT conditions. The condition that $\lambda_i^* f_i(x^*) = 0$ $i=1,\dots,k$ is called complementary slackness.

11/11 L13:

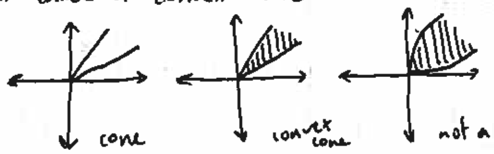
Conic optimization problems & their Lagrange duals:

$$p^* = \inf_{x\in\mathbb{R}^n} c^Tx \quad \text{s.t. } Ax=b;\ x\in K$$

Here $A\in\mathbb{R}^{m\times n}$, $b\in\mathbb{R}^m$, $c\in\mathbb{R}^n$, $K\subseteq\mathbb{R}^n$ is a convex cone

Definition: A set $K\subseteq\mathbb{R}^n$ is a cone if $x\in K\Rightarrow \alpha x\in K$ for all $\alpha\geq 0$. If K is further a convex set, then K is called a convex cone.

Ex:



Remark: A set $K\subseteq\mathbb{R}^n$ is a convex cone iff for any $x^{(1)}, x^{(2)}\in K$ and any $\alpha_1,\alpha_2\geq 0$, we have that $\alpha_1 x^{(1)}+\alpha_2 x^{(2)}\in K$.

We'll see next how specific choices of K lead to different families of optimization problems.

Definition: The nonnegative orthant in $\mathbb{R}^n$ is the set of all vectors with nonnegative entries:

$\mathbb{R}^n_+ = \{x\in\mathbb{R}^n \mid x_i\geq 0 \quad i=1,\dots,n\}$. Conic optimization problems with the nonnegative orthant as the choice of cone are called linear programs.

Q: What kinds of convex programs can we solve via linear programming? In other words, what convex sets can be specified as the intersection of an affine space and a nonnegative orthant?

Proposition: Let $S\subseteq\mathbb{R}^n$ be a set specified as follows: $S=\left\{x\in\mathbb{R}^n \;\middle|\; \begin{array}{ll} a^{(i)T}x\leq b_i & i=1,\dots,k \\ c^{(j)T}x=d_j & j=1,\dots,m\end{array}\right\}$

That is, S is the intersection of a finite collection of halfspaces and an affine space. Then one can optimize a linear function over S via linear programming.

Proof: Suppose we have the following optimization problem:

$$(\dagger)\quad p^* = \inf_{x\in\mathbb{R}^n} c^Tx \quad \text{s.t. } x\in S$$

This problem can be expressed equivalently as:

$$p^* = \inf_{x\in\mathbb{R}^n;\, y\in\mathbb{R}^k} c^Tx \quad \text{s.t. } \begin{array}{ll} a^{(i)T}x+y_i=b_i & i=1,\dots,k \\ c^{(j)T}x=d_j & j=1,\dots,m \\ y_i\geq 0 & i=1,\dots,k\end{array}$$

Any vector in $\mathbb{R}^n$ can be expressed as the difference of two entrywise nonnegative vectors in $\mathbb{R}^n$. Thus, we have that

$$p^* = \inf_{x^{(+)},x^{(-)}\in\mathbb{R}^n,\, y\in\mathbb{R}^k} c^Tx^{(+)}+(-c)^Tx^{(-)}+0^Ty$$

$$\text{s.t. } \begin{array}{ll} a^{(i)T}x^{(+)}+(-a)^{(i)T}x^{(-)}+y_i=b_i & i=1,\dots,k \\ c^{(j)T}x^{(+)}+(-c^{(j)})^Tx^{(-)}+0^Ty=d_j & j=1,\dots,m\end{array} \quad \begin{pmatrix}x^{(+)}\\x^{(-)}\\y\end{pmatrix}\in\mathbb{R}^{2n+k}_+$$

■

Remark: In many contexts, the optimization problem (†) for a set specified as the intersection of a finite number of halfspaces are called linear programs.

- Sets that are specified as the intersection of a finite number of halfspaces are called polyhedra. Bounded polyhedra are called polytopes.

Q: What sets can be described via linear programming in an efficient manner? (Efficient means a polynomial function in which the set lies.)

Example:

- S = unit ball of ℓ_∞ norm $=\{x\in\mathbb{R}^n \mid |x_i|\leq 1,\ i=1,\dots,n\}$

This set can be described using the previous proposition as a standard form LP of size $4n$ based on the following operation:

$$S=\left\{x\in\mathbb{R}^n \;\middle|\; \begin{array}{ll} x_i\leq 1 & i=1,\dots,n \\ -x_i\leq 1 & i=1,\dots,n\end{array}\right\}$$

- S = unit ball of ℓ_1 norm $=\{x\in\mathbb{R}^n \mid \sum_{i=1}^n |x_i|\leq 1\}$

This set is equivalently expressed as:

$S=\{x\in\mathbb{R}^n \mid \sum_{i=1}^n \pm x_i\leq 1\}$ — all 2^n possible sign patterns

This description has 2^n linear inequalities. One can instead specify S as follows:

$$S=\left\{x\in\mathbb{R}^n \;\middle|\; \exists y\in\mathbb{R}^n \text{ s.t. } \begin{array}{l} -x_i\leq y_i \\ x_i\leq y_i\end{array}\ i=1,\dots,n;\ \sum_{i=1}^n y_i\leq 1\right\}$$

In this description, which has $2n+1$ linear inequalities, we have added n additional variables, attained an efficient representation in $\mathbb{R}^{2n}$, & then projected out the additional variables. Thus, if we have a description of a set $S\subseteq\mathbb{R}^n$ as follows:

$S=\{x\in\mathbb{R}^n \mid \exists y\in\mathbb{R}^k \text{ s.t. } Ax+By=b,\ \binom{x}{y}\in\mathbb{R}^n_+\}$, then we ~~have~~ can solve the following optimization problem (†) as:

$$p^* = \inf_{x\in\mathbb{R}^n;\, y\in\mathbb{R}^k} c^Tx+0^Ty \quad \text{s.t. } Ax+By=b;\ \binom{x}{y}\in\mathbb{R}^n_+.$$

Such descriptions of convex sets are called lift-and-project descriptions, & they are very useful in an optimization context. In particular, if the number of additional variables required is a polynomial function of n, then lift-and-project descriptions lead to an efficient approach for optimization.

11/13 L14:

Definition: The <u>semidefinite</u> cone in $\mathbb{S}^n$ - the space of $n\times n$ symmetric matrices - is the collection of positive semi-definite matrices:

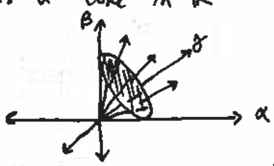
$$\mathbb{S}^n_+ = \{X \mid X\in \mathbb{S}^n,\ X\succeq 0\}$$

Remarks: $\mathbb{S}^n_+$ is a convex cone. For $X_1, X_2\in \mathbb{S}^n_+$, $\alpha_1,\alpha_2\geq 0$, $y^T(\alpha_1X_1+\alpha_2X_2)y = \alpha_1 y^TX_1y + \alpha_2 y^TX_2y$

$\Rightarrow \alpha_1X_1+\alpha_2X_2\in \mathbb{S}^n_+$

- $\mathbb{S}^2_+$ can be viewed as a cone in $\mathbb{R}^3$:

$\begin{bmatrix}\alpha & \gamma\\ \gamma & \beta\end{bmatrix}\succeq 0$



det ≥ 0; trace ≥ 0

$\alpha,\beta\geq 0$; $\alpha\beta\geq\gamma^2$

Conic optimization problems in which the cone is the semidefinite cone are called <u>semidefinite programs</u>: $p^* = \inf_{X\in\mathbb{S}^n} Tr(CX)$ s.t. $Tr(A_jX)=b_j$ $j=1,\ldots,m$, $X\in\mathbb{S}^n_+$

Here $C\in\mathbb{S}^n$, each $A_j\in\mathbb{S}^n$, and each $b_j\in\mathbb{R}$

<u>Remark:</u>
- By restricting the matrix variable in a semidefinite program to be diagonal (for example, by adding linear equality constraints that require the off diagonal entries to be zero), we obtain a linear program. Thus, linear programs are a subclass of semidefinite programs.
- In linear programming, the constraint set is a polyhedron, i.e. specified by the intersection of a finite collection of halfspaces. However, the semidefinite programs can have constraint sets that are not polyhedra (see the example of $\mathbb{S}^2_+$). Here, semidefinite programs are a strictly more general family of optimization problems than linear programs.

Q: What kinds of sets can be described via semidefinite programming? What sets can be described efficiently?

Both of these are actively studied research questions. However, semidefinite programs have been used widely in practice in many different application areas.

<u>Example:</u>
- $\mathcal{S} = \{X\in\mathbb{S}^n \mid \lambda_{max}(X)\leq 1\} = \{X\in\mathbb{S}^n \mid X\preceq I\}$
$= \{X\in\mathbb{S}^n \mid I - X\succeq 0\} \to$ can be described via SDP.
- $\mathcal{S} = \{X\in\mathbb{S}^n \mid \sigma_{max}(X)\leq 1\}$; $\sigma(M) = \sqrt{\lambda(MM^T)}$ $M\in\mathbb{R}^{m\times n}$

Hence, $\mathcal{S} = \{X\in\mathbb{S}^n \mid -1\leq\lambda_i(X)\leq 1\ \ i=1,\ldots,n\}$
$= \{X\in\mathbb{S}^n \mid -I\preceq X,\ X\preceq I\} \to$ can be described via SDP.

<u>Conic Duality:</u>

<u>Definition:</u> Let $K\subseteq\mathbb{R}^n$ be a cone. The <u>dual</u> of K is:
$K^* = \{y\in\mathbb{R}^n \mid y^Tx\geq 0\ \ \forall x\in K\}$

Remarks: K^* is a convex cone, even if K is not.

Examples: $(\mathbb{R}^n_+)^* = \mathbb{R}^n_+$; $(\mathbb{S}^n_+)^* = \mathbb{S}^n_+$

For a subspace W, the dual $W^* = \{y\in\mathbb{R}^n \mid y^Tx\geq 0\ \forall x\in W\}$
$= \{y\in\mathbb{R}^n \mid y^Tx = 0\ \ \forall x\in W\}$ as W is a subspace $= W^\perp$

To derive a dual of a conic optimization problem, we will appeal to the Lagrange duality principle by considering separating hyperplanes for each of the constraints & then combining them. From a few lectures ago, the Lagrange dual of $p^* = \inf_{x\in\mathbb{R}^n} c^Tx$ s.t. $Ax=b$, $x\in K$

is: $d^* = \sup_{\delta_1,\delta_2\in\mathbb{R};\, g_1,g_2\in\mathbb{R}^n} \delta_1+\delta_2$ s.t. $\{x\mid Ax=b\}\subseteq\{x\mid g_1^Tx\geq\delta_1\}$
$K\subseteq\{x\mid g_2^Tx\geq\delta_2\}$, $c = g_1+g_2$

What constraints do these conditions imply on $g_1, g_2, \delta_1, \delta_2$?

- $\{x\mid Ax=b\}\subseteq\{x\mid g_1^Tx\geq\delta_1\}$

$\Rightarrow g_1$ must be of the form $A^T\mu$ for $\mu\in\mathbb{R}^m$

Therefore, $g_1^Tx = \mu^T(Ax) = \mu^Tb$ for all x satisfying $Ax=b$

$\Rightarrow \delta_1\leq\mu^Tb$

- $K\subseteq\{x\mid g_2^Tx\geq\delta_2\}$; $\tilde{x}\in K\Rightarrow g_2^T\tilde{x}\geq\delta_2$
$\Rightarrow g_2^T(\alpha\tilde{x})\geq\delta_2\ \ \forall\alpha\geq 0 \Rightarrow \delta_2\leq 0$
$\Rightarrow g_2\in K^*$

Hence, the Lagrange dual problem simplifies as:

$\sup_{\delta_1,\delta_2\in\mathbb{R};\, g_1,g_2\in\mathbb{R}^n;\, \mu\in\mathbb{R}^m} \delta_1+\delta_2$ s.t. $g_1 = A^T\mu$, $\delta_1\leq\mu^Tb$, $c = g_1+g_2$
$g_2\in K^*$, $\delta_2\leq 0$

<u>This can be simplified further</u> as:

$$\boxed{d^* = \sup_{\mu\in\mathbb{R}^m} \mu^Tb \text{ s.t. } c - A^T\mu\in K^*}$$

Strong duality holds if $\exists\,\tilde{x}\in ri(K)$ s.t. $A\tilde{x}=b$

The Lagrange dual of an LP is:
$d^* = \sup_{\mu\in\mathbb{R}^m} b^T\mu$ s.t. $c - A^T\mu\in\mathbb{R}^n_+$

For an LP, strong duality holds if $\exists\,\tilde{x}$ s.t. $\tilde{x}\in\mathbb{R}^n_+$ & $A\tilde{x}=b$

The Lagrange dual of the SDP:
$p^* = \inf_{X\in\mathbb{S}^n} Tr(CX)$ s.t. $Tr(A_jX)=b_j$; $j=1,\ldots,m$; $X\in\mathbb{S}^n_+$

is:
$d^* = \sup_{\mu\in\mathbb{R}^m} b^T\mu$ s.t. $C - \sum_{j=1}^m A_j\mu_j\in\mathbb{S}^n_+$

Strong duality holds if $\exists\,\tilde{X}$ s.t. $\tilde{X}\succ 0$ & $Tr(A_j\tilde{X}) = b_j$ for each $j=1,\ldots,m$.

- Expressions of the type $C - \sum_{j=1}^m A_j\mu_j\in\mathbb{S}^n_+$ where we require an affine function of a variable to be positive semidefinite are called Linear Matrix Inequalities (LMIs)

11/18 L16: <u>Integer Programming:</u>

$p^* = \inf_{x\in\mathbb{Z}^n} f(x)$ s.t. $x \in S$

Here $f:\mathbb{Z}^n\to\mathbb{R}$ & $S\subseteq \mathbb{Z}^n$. We'll investigate a subclass of Integer Programming problems called <u>(mixed) integer linear programming:</u>

(†) $\inf_{x\in\mathbb{R}^n} c^T x$ s.t. $Ax=b$, $x\in \underline{\mathbb{Z}^n_+} \to \mathbb{R}^n_+\cap\mathbb{Z}^n$

This problem is an integer linear program (ILP), & if some of the decision variables are allowed to be reals (rather than being restricted to being integers), then the resulting problem is called a mixed integer linear program (MILP).

If we wish to use convex optimization to solve this problem, we need a way to compute convex hulls of constraint sets in integer linear programs efficiently.

Q: How do we describe the convex hull of the set $\{x \mid Ax=b,\ x\in\mathbb{Z}^n_+\}$? When can this convex hull be described efficiently?

<u>Fundamental Theorem of Integer Programming:</u>

Let $A\in\mathbb{R}^{m\times n}$ & let $b\in\mathbb{Z}^m$. Then the set $S=\{x \mid Ax=b,\ x\in\mathbb{Z}^n_+\}$ can be described as $\mathrm{conv}(S)=\{x \mid \tilde{A}x\le\tilde{b},\ x\in\mathbb{R}^n\}$ for some $\tilde{A}\in\mathbb{Z}^{q\times n}$, $\tilde{b}\in\mathbb{Z}^q$. In other words, $\mathrm{conv}(S)$ is a polyhedron.

This theorem says nothing about the efficiency of describing the convex hull of S. In principle q could be an exponential function of n.

A natural relaxation of (†) is the following linear program:

$$p' = \inf_{x\in\mathbb{R}^n} c^T x \quad \text{s.t. } Ax=b,\ x\in\mathbb{R}^n_+$$

In general $p'\le p^*$ as we're minimizing over a larger set in the relaxation. A different way to see this is that

$\mathrm{conv}\{x \mid Ax=b,\ x\in\mathbb{Z}^n_+\}\subseteq\{x \mid Ax=b,\ x\in\mathbb{R}^n_+\}$ with the inclusion being strict in general.

Q: When is the above inclusion an equality?

<u>Definition:</u> Let $S\subseteq\mathbb{R}^n$ be a convex set. The set S is <u>integral</u> if $S=\mathrm{conv}(S\cap\mathbb{Z}^n)$.

With this definition, our question becomes:

Under what conditions on $A\in\mathbb{Z}^{m\times n}$ & $b\in\mathbb{Z}^m$ is the polyhedron $\{x \mid Ax=b,\ x\in\mathbb{R}^n_+\}$ integral?

<u>Definition:</u> Let $A\in\mathbb{Z}^{m\times n}$ be a matrix of full row rank. Then A is a <u>unimodular matrix</u> if every $m\times m$ submatrix has determinant equal to $+1$ or -1. (or 0)

<u>Proposition:</u> Let $A\in\mathbb{Z}^{m\times m}$ be non-singular. Then A is unimodular iff $A^{-1}b\in\mathbb{Z}^m$ for every $b\in\mathbb{Z}^m$.

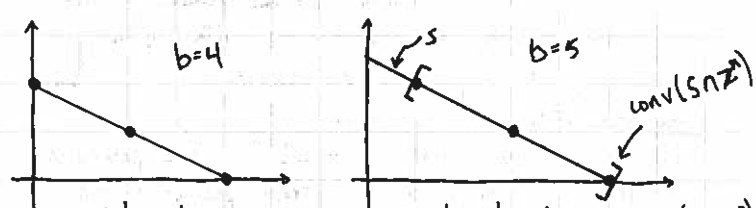
<u>Proof:</u> Suppose A is unimodular. We have that $A^{-1}=\frac{1}{\det(A)}\cdot\mathrm{adj}(A)$ by Cramer's Rule. As $A\in\mathbb{Z}^{m\times n}$, we have that $\mathrm{adj}(A)\in\mathbb{Z}^{m\times n}$. Further, as $\det(A)=\pm1$, we have that $A^{-1}\in\mathbb{Z}^{m\times m}$. Hence, $A^{-1}b\in\mathbb{Z}^m$ for any $b\in\mathbb{Z}^m$.

In the other direction, $A^{-1}b\in\mathbb{Z}^m$ for every $b\in\mathbb{Z}^m$ implies $A^{-1}\in\mathbb{Z}^{m\times m}$. Therefore $\det(A^{-1})\in\mathbb{Z}$. Similarly, $\det(A)\in\mathbb{Z}$. But we also know that $\det(A)\cdot\det(A^{-1})=1$. Hence, $\det(A)=\pm1$. ■

<u>Theorem:</u> (Dantzig, 1960's) Let $A\in\mathbb{Z}^{m\times n}$. The set $S=\{x \mid Ax=b,\ x\in\mathbb{R}^n_+\}$ is integral for every $b\in\mathbb{Z}^m$ if and only if A is unimodular.

11/20 <u>L17:</u>

A polyhedron: $\{x\in\mathbb{R}^2 \mid x+2x_2=b,\ x\in\mathbb{R}^2_+\}$. We'll consider this polyhedron for different choices of $b\in\mathbb{Z}$. Note here $A=[1\ 2]$ which is not unimodular.



integral   not integral: $S\ne\mathrm{conv}(S\cap\mathbb{Z}^2)$

Remark: Unimodular matrices <u>need not</u> have all entries being $0, +1, -1$. Ex. $\begin{pmatrix}2&3\\1&2\end{pmatrix}$.

<u>Definition:</u> An integer matrix $A\in\mathbb{Z}^{m\times n}$ is <u>totally unimodular</u> if every square submatrix of A has determinant equal to 0 or ±1. [Here the square submatrices can be of any size.]

Remark: There is no restriction on the rank of a totally unimodular matrix. So n could be less than m, unlike with a unimodular matrix.

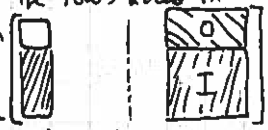
- The entries of a totally unimodular matrix must be $0, \pm1$.
- Totally unimodular matrices of full row-rank are unimodular.

11/20 L17 cont'd:

Proposition: Fix an integer matrix $A \in \mathbb{Z}^{m\times n}$. Then A is totally unimodular if and only if the matrix $[A\ I] \in \mathbb{Z}^{m\times(n+m)}$ is unimodular.

Proof: [A totally unimodular $\Rightarrow$ $[A\ I]$ unimodular] $\Rightarrow$
Select any $m\times m$ submatrix of $[A\ \ I]$. If this submatrix is a submatrix of A, then we have that the determinant of the submatrix equals $0, \pm 1$ as A is totally unimodular. If the submatrix is not a submatrix purely of A and contains columns of both A and of I, then we can conclude that the determinant of our $m\times m$ submatrix is equal to the determinant of a smaller square submatrix of A. As A is totally unimodular, we are done. Finally, if our $m\times n$ submatrix equals I, we are again done as the determinant equals 1.

[$[A\ I]$ unimodular $\Rightarrow$ A totally unimodular] $\Leftarrow$
For any square submatrix of A (of any size), consider an $m\times m$ submatrix of $[A\ I]$ formed by
(i) appending the remaining rows (for the same columns)
(ii) appending columns of the identity matrix that correspond to the indices of the rows added in part (i). This larger $m\times m$ submatrix of $[A\ \ I]$ has determinant equal to 0 or ± 1 (as $[A\ I]$ is unimodular) & further this determinant is equal to the determinant of the smaller square submatrix of A that we started with (up to sign). Therefore, A is totally unimodular. ■



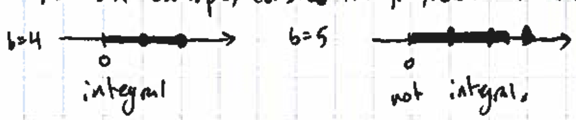
Theorem: [Hoffman & Kruskal, 1950's]
Fix an integer matrix $A \in \mathbb{Z}^{m\times n}$. The polyhedron $\{x \mid Ax \le b,\ x \in \mathbb{R}^n_+\}$ is an integral polyhedron for every $b \in \mathbb{Z}^m$ if & only if A is totally unimodular.
[Here $Ax \le b$ is a componentwise inequality]

Proof: For each $b \in \mathbb{Z}^m$, the polyhedron $\{x \mid Ax \le b, x \in \mathbb{R}^n_+\}$ is integral if and only if

$$\left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^{n+m} \,\middle|\, Ax+y=b,\ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^{n+m}_+ \right\}$$

is an integral polyhedron. Based on the earlier result by Dantzig, this latter polyhedron is integral for b if and only if the matrix $[A\ \ I]$ is unimodular. This in turn is equivalent to A being totally unimodular. ■

Remark: As with unimodularity, we could have an integral polyhedron $\{x \mid Ax \le b, x \in \mathbb{R}^n_+\}$ for a particular b even if A is not totally unimodular.
As an example, consider the polyhedron $\{x \in \mathbb{R} \mid 2x \le b, x \in \mathbb{R}_+\}$



Note that $[2] \in \mathbb{Z}^{1\times 1}$ is not totally unimodular.

ACM 113

12/02 L18: Review

$p^* = \inf_{x\in\mathbb{R}^n} f(x) : x \in S$] General optimization problem.

Here $f:\mathbb{R}^n \to \mathbb{R}$ and $S \subseteq \mathbb{R}^n$.

Dual perspective:

$p^* = \sup_{\gamma\in\mathbb{R}} \gamma : \mathrm{epi}(f) \cap \{S \times (-\infty, \gamma)\} = \emptyset$.

Q: How do we certify that two sets have an empty intersection?

→ show that they are on two sides of a hyperplane

Q: ~~How~~ When does this work?

→ If the two sets are convex.

This leads to convex optimization problems:

$$p^* = \inf_{x\in\mathbb{R}^n} \underbrace{f(x)}_{\text{convex function}} : x \in \underbrace{S}_{\text{convex set}}$$

- Observation: All optimization problems can be reformulated as convex optimization problems (in fact with a linear objective).
- The difficulty is that convex sets may not be easy or tractable to describe.

Q: How do we describe convex sets efficiently?

→ Consider intersections of known convex sets that are easy to describe.

Q: How do we think about duals of convex optimization problems in which the constraint set is described as an intersection of convex sets?

→ Lagrange Duality.

- Conic Programming as a way to describe convex sets
  → LP, SDP, ... → dependent on how you define K.

$$\inf_{x\in\mathbb{R}^n} c^Tx \quad \text{s.t.} \quad Ax = b, \quad x \in K \to \text{convex cone}$$

- Integer Programming:

$$\inf_{x\in\mathbb{R}^n} c^Tx \quad \text{s.t.} \quad Ax = b, \quad x \in \mathbb{Z}^n_+$$

- If the set $\{x \mid Ax = b, x \in \mathbb{R}^n_+\}$ is an integral set, then we can solve an integer program via linear programming in an efficient manner

Q: When is a polyhedron integral?

→ Unimodularity / Total Unimodularity.

↳ $Ax = b$ is integral ↳ $Ax \leq b$ is integral $\forall b$.

- Solving integer programs via LP

Let $P = \{x \mid Ax = b, x \in \mathbb{R}^n_+\}$

$$p^* = \inf_{x\in\mathbb{R}^n} c^Tx : x \in P \cap \mathbb{Z}^n.$$
$$= \inf_{x\in\mathbb{R}^n} c^Tx : x \in \mathrm{conv}(P \cap \mathbb{Z}^n) \quad \text{because objective is linear.}$$
$$= \inf_{x\in\mathbb{R}^n} c^Tx : x \in P \quad (\text{if } P \text{ is integral})$$

- For exam: reformulate problem to LP, SDP, use
  How to say it can be reformulated in standard form.
- Problem 5 from Midterm:

$$p^* = \inf_{x\in\mathbb{R}^n} \tfrac{1}{2}\|x-a\|^2 : x \in B_f$$

The dual is reformulated as:

$$\inf_{g\in\mathbb{R}^n} \tfrac{1}{2}\|g-a\|^2 + f^{\text{dual}}(g_a).$$

$$p^* = \sup_{\gamma\in\mathbb{R},\ g\in\mathbb{R}^n,\ \delta\in\mathbb{R}} \gamma \quad \text{s.t.} \quad \mathrm{epi}\left(\tfrac{1}{2}\|x-a\|^2\right) \subseteq \{(x,t) \mid g^Tx - t \leq \delta\}$$
$$S \times (-\infty, \gamma) \subseteq \{(x,t) \mid g^Tx - t > \delta\}$$

Think of this as constraints on γ, g, δ; ~~we only really care about x~~ we want to remove x.

- From the first constaint: $g^Tx - t \leq \delta$ for all (x,t) s.t. $t \geq \frac{1}{2}\|x-a\|^2$. This is a condition on g, δ.

The extremal value is when $t = \frac{1}{2}\|x-a\|^2$

In other words,

$$g^Tx - \tfrac{1}{2}\|x-a\|^2 \leq \delta \quad \forall x \in \mathbb{R}^n.$$

$$= g^Tx - \tfrac{1}{2}x^Tx + a^Tx + \tfrac{1}{2}a^Ta = -\tfrac{1}{2}x^Tx + (g+a)^Tx - \tfrac{1}{2}(g+a)^T(g+a)$$
$$+ \tfrac{1}{2}(g+a)^T(g+a) - \tfrac{1}{2}a^Ta$$

$$= -\tfrac{1}{2}\|x-(g+a)\|^2 + \tfrac{1}{2}(g+a)^T(g+a) - \tfrac{1}{2}a^Ta$$

Worst case x is when $x = (g+a) \rightsquigarrow -\frac{1}{2}\|x-(g+a)\|^2 = 0$.

This reduces to: $\delta \geq \frac{1}{2}\|g+a\|^2 - \frac{1}{2}a^Ta \ ; g \in \mathbb{R}^n$

- From the second constraint, we require that

$S \subseteq \{x \mid g^Tx - \gamma \geq \delta\}$

This condition says that $g^Tx \geq \delta + \gamma \ \forall x \in B_f$.

$\iff (-g)^Tx \leq -(\delta+\gamma) \ \forall x \in B_f$.

This reduces to $f^{\text{dual}}(-g) \leq -(\delta+\gamma)$

The dual problem becomes:

$$\sup_{\gamma\in\mathbb{R},\ \delta\in\mathbb{R},\ g\in\mathbb{R}^n} \gamma \quad \text{s.t.} \quad f^{\text{dual}}(g) \leq -(\delta+\gamma)$$
$$\delta \geq \tfrac{1}{2}\|g+a\|^2 - \tfrac{1}{2}a^Ta$$

$$\iff \sup_{\delta\in\mathbb{R},\ g\in\mathbb{R}^n} -\delta - f^{\text{dual}}(-g) \quad \text{s.t.} \quad \delta \geq \tfrac{1}{2}\|g+a\|^2 - \tfrac{1}{2}a^Ta.$$

- HW 5 Question 1.6

Ex 1: Knapsack:

intractable $\inf_{x \in \mathbb{R}^n} v^T x$ s.t. $c^T x \le B$, $x \in \{0,1\}^n$ Worst case: 2^n combinations

Ex 2: Assignment:

tractable $\inf_{M \in \mathbb{R}^{n \times n}} \operatorname{tr}(PM)$ s.t. $P \in \{n \times n$ permutation matrices$\}$: $n!$

↳ efficient description of convex hull

- HW 7 Question 3.1

$A \in \{-1, 0, 1\}^{m \times n}$

A has at most one +1 & one -1 in each column $\Rightarrow$ A is totally unimodular.

Consider any square submatrix, choose case by case examples and show it holds.

12/4 L19: Solving Optimization Problems

Aside: Convex functions that are difficult to compute:

$$f(w) = \sup_{x \in \mathbb{R}^n} w^T x : x \in S \,;\, f: \mathbb{R}^n \to \mathbb{R},\ S \subset \mathbb{R}^n$$

Minimizing quadratics:

$$\inf_{x \in \mathbb{R}^n} \tfrac{1}{2} x^T A x - b^T x$$

Here $A \in S^n, b \in \mathbb{R}^n$. If $A \succ 0$ then $\frac{1}{2}x^T A x - b^T x$ is a convex function. Setting the gradient equal to zero, the optimal solution of this problem is given by the $\hat{x}$ that solves the following system:

$$A\hat{x} = b.$$

Linear systems can be solved via Gaussian elimination (for example). Based on this, we can solve

$$\inf_{x \in \mathbb{R}^n} \tfrac{1}{2} x^T A x - b^T x \quad \text{s.t. } Cx = d$$

Here $A \in S^n, A \succ 0, b \in \mathbb{R}^n, C \in \mathbb{R}^{m \times n}, d \in \mathbb{R}^m$. This problem can again be solved by reducing to a system of linear equations.

The next natural problem to aim to solve is:

$$\inf_{x \in \mathbb{R}^n} f(x) \text{ where } f: \mathbb{R}^n \to \mathbb{R} \text{ is a convex function.}$$

We could try to solve this by Newton's method, which reduces to solving a sequence of quadratic minimization problems. To do this, we need to be able to evaluate the second-order Taylor expansion of f about any point in a tractable fashion (i.e. compute gradients/Hessians efficiently).

Based on the preceding ideas, we could try to minimize convex functions subject to linear equality constraints.

Q: How do we handle inequality constraints?

To make things concrete, we'll focus on Linear Programs: $\inf_{x \in \mathbb{R}^n} c^T x$ s.t. $a_i^T x \le b_i$ $i = 1, \dots, k$

Here $c, a_1, \dots, a_k \in \mathbb{R}^n$, $b_1, \dots, b_k \in \mathbb{R}$

One approach to computing p^* is to consider the following problem: $p^*(\tau) = \inf_{x \in \mathbb{R}^n} c^T x + \tau \sum_{i=1}^{k} \underbrace{-\log(b_i - a_i^T x)}_{\text{"Log barrier functions"}}$

If we initialize Newton's method for this problem with $\tilde{x}$ that satisfies $a_i^T \tilde{x} < b_i$ for all $i = 1, \dots, k$, then we'll stay within the feasible region. For $\tau > 0$, computing $p^*(\tau)$ entails minimization of a convex function.

We can try to compute p^* by taking the limit $\tau \to 0$ and considering the sequence of corresponding optimal values $p^*(\tau)$. In practice, τ must not approach zero too fast. To implement this method, we need $\tilde{x}$ s.t. $a_i^T \tilde{x} < b_i$ $\forall i = 1, \dots, k$. To obtain such a point, we could solve

$$\inf_{x \in \mathbb{R}^n, s \in \mathbb{R}} s \quad \text{s.t. } a_i^T x \le b_i + s,\ i = 1, \dots, k$$

To make these ideas go through, we need to appeal to a notion called self-concordance. This notion was developed by Nesterov & Nemirovski. An excellent exposition on this idea is in the book "A Mathematical View of Interior-Point Methods" by Renegar.

Q: How about solving Integer Linear Programs?

$$p^* = \inf_{x \in \mathbb{R}^n} c^T x \quad \text{s.t. } x \in P \cap \mathbb{Z}^n$$

Here P is a polyhedra. If P is integral, you can remove the integer constraint and we are done. But what about more general polyhedra?

Suppose we solve the LP relaxation:

$$\inf_{x \in \mathbb{R}^n} c^T x \quad \text{s.t. } x \in P$$

and obtain an optimal solution $\hat{x} \in \mathbb{Z}^n$. This means that we've solved our ILP & $c^T \hat{x} = p^*$. Suppose instead that $\hat{x} \notin \mathbb{Z}^n$; in particular, suppose $\hat{x}_1 \notin \mathbb{Z}$. We can now consider the following two problems:

$$p_1^* = \inf_{x \in \mathbb{R}^n} c^T x \quad \text{s.t. } x \in P \cap \{x \mid x_1 \le \lfloor \hat{x}_1 \rfloor\} \cap \mathbb{Z}^n$$

$$p_2^* = \inf_{x \in \mathbb{R}^n} c^T x \quad \text{s.t. } x \in P \cap \{x \mid x_1 \ge \lceil \hat{x}_1 \rceil\} \cap \mathbb{Z}^n$$

This step is called branching. Computing each of p_1^* & p_2^* is itself an ILP, and $p^* = \min\{p_1^*, p_2^*\}$.

Q: When do we not need to branch?

- If a subproblem gives an integral solution for its LP relaxation, we don't need to branch further.
- If the LP relaxation of a subproblem has an optimal value greater than or equal to a known upper bound on p^* then we don't need to branch further.

We initialize this procedure with an upper bound on p^* of $+\infty$.

This method is called Branch-and-Bound.

Fin

ACM 113 Exam 1 Prep

Matrix Decompositions:

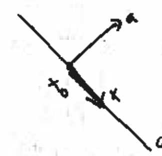
- LU decomposition: square matrix $A = LU$ where L is lower triangular and U is upper triangular
  - ↳ LDU: L —"— w/ ones along diagonal, same for U and D is diagonal
- Block LU decomposition: $\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & 0 \\ 0 & D-CA^{-1}B \end{bmatrix}\begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix} = \begin{bmatrix} I_p & BD^{-1} \\ 0 & I_q \end{bmatrix}\begin{bmatrix} A-BD^{-1}C & 0 \\ 0 & D \end{bmatrix}\begin{bmatrix} I_p & 0 \\ D^{-1}C & I_q \end{bmatrix}$
- Schur complement: $M/D := A - BD^{-1}C$; $M/A = D - CA^{-1}B$
- Rank factorization: $m \times n$ matrix A of rank r → $A = CF$ where C = $m\times r$ full column rank, $F = r \times n$ full row rank
- Cholesky decomposition: square, hermitian, positive definite matrix $A = U^* U$ where U is upp. trian. w/ real + diag. entries
- QR decomposition: $m \times n$ matrix A w/ linearly independent columns. $A = QR$ where Q is $m \times m$ unitary and R is upp. triang. $m \times n$.

Eigen decomposition: square matrix A w/ linearly independent eigenvec. (not necessarily distinct eigenvalues).
$A = VDV^{-1}$ where D is diagonal of eig. val. cols of V are eigen. vectors.
- For any real symmetric matrix, $A = VDV^T = Q\Lambda Q^T$ → Q is orthogonal
- Complex normal matrix $(A^* A = AA^*)$, $A = U \Lambda U^*$ where U is unitary

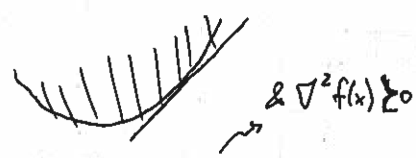
- Schur Decomposition: square matrix $A = UTU^*$ where U is unitary, T is upper triangular w/ λ on diagonal
- SVD: $m \times n$ matrix $A = UDV^*$ D is nonnegative diagonal. U & V satisfy $U^* U = I = V^* V$; D has singular values

Boyd:

- Least Squares: $\min \|Ax - b\|_2^2 = \sum_{i=1}^k (a_i^T x - b_i)^2$ → sol: $(A^T A)x = A^T b \Rightarrow x = (A^T A)^{-1} A^T b$
- Linear programming: $\min\ c^T x$ s.t. $a_i^T x \leq b_i$
- Hyperplane is a set of the form $\{x \mid a^T x = b\}$
  - set of points with a constant inner product with a
  - hyperplane with normal vector a. b determines offset from origin



- Halfspace: $\{x \mid a^T x \leq b\}$ $a \neq 0$ are convex but not affine
- Norm cone: $C = \{(x,t) \mid \|x\| \leq t\} \subseteq \mathbb{R}^{n+1}$ → convex cone
- Convexity: preserved under: intersection
  - affine functions
  - projection onto some of it's coordinates
  - Minkowski sum



- Convexity of functions. f is convex iff dom f is convex & $f(y) \geq f(x) + \nabla f(x)^T (y-x)$
- Sublevel sets: The α-sublevel set of $f: \mathbb{R}^n \to \mathbb{R}$ is $C_\alpha = \{x \in \text{dom} f \mid f(x) \leq \alpha\}$
  - ↳ sublevel sets of a convex function are convex → converse not true
  - ↳ if f is concave, its α-superlevel set $\{x \in \text{dom} f \mid f(x) \geq \alpha\}$ is convex
- Epigraph: $\{(x, f(x)) \mid x \in \text{dom} f\}$ → f is convex iff its epigraph is convex
- Hypograph: hypo $f = \{(x,t) \mid t \leq f(x)\}$ → f is concave iff hypo f is convex
- Jensen's inequality: if f is convex, $x_1, \ldots, x_k \in \text{dom} f$ & $\theta_1, \ldots, \theta_k \geq 0$ w/ $\sum \theta_i = 1$
  Then $f(\theta_1 x_1 + \cdots + \theta_k x_k) \leq \theta_1 f(x_1) + \cdots + \theta_k f(x_k)$ → extends to sets & $\int$
- Function composition $f(x) = h(g(x))$

f is convex if	h is convex &	nondecreasing (in each arg.)	& g is convex (in each arg. For multi dim)
convex	convex	nonincreasing	concave
concave	concave &	nondecreasing	concave
concave	concave &	nonincreasing	convex

Appendix:

~~Norm~~ Inner product: $\langle x,y\rangle = x^T y = \sum_{i=1}^n x_i y_i$ $\angle(x,y) = \cos^{-1}\left(\frac{x^T y}{\|x\|_2\|y\|_2}\right)$ → orthogonal if $x^T y = 0$

Norm $\|x\|_2 = (x^T x)^{1/2} = (x_1^2 + \dots + x_n^2)^{1/2}$ $\langle X,Y\rangle$ (m×n matrices) $= \mathrm{tr}(X^T Y) = \sum_{i=1}^m \sum_{j=1}^n X_{ij} Y_{ij}$

Cauchy Schwartz: $|x^T y| \le \|x\|_2 \|y\|_2$

$\|X\|_F = \mathrm{tr}(X^T X)^{1/2} = \left(\sum_{i=1}^m \sum_{j=1}^n X_{ij}^2\right)^{1/2}$

Operator norms: $\|\cdot\|_a$ and $\|\cdot\|_b$ are norms on $\mathbb{R}^m$ and $\mathbb{R}^n$ resp. Operator norm of $X \in \mathbb{R}^{m\times n}$ induced by them is
$\|X\|_{a,b} = \sup\{\|Xu\|_a \mid \|u\|_b \le 1\}$

→ if they are both Euclidean norms, then $\|X\|_2 = \sigma_{max}(X) = (\lambda_{max}(X^T X))^{1/2}$

Analysis:

Interior point: $x \in C \subseteq \mathbb{R}^n$ is an interior point if $\exists\, \varepsilon > 0$ for which $\{y \mid \|y-x\|_2 \le \varepsilon\} \subseteq C$

Closure: $\mathrm{cl}\, C = \mathbb{R}^n \setminus \mathrm{int}(\mathbb{R}^n \setminus C)$

↳ closed iff it contains the limit point of every convergent sequence in it

Boundary: $\mathrm{bd}\, C = \mathrm{cl}\, C \setminus \mathrm{int}\, C$

Sup: if $\sup C \in C$, the sup is attained

Continuity: $f: \mathbb{R}^n \to \mathbb{R}^m$ is continuous at $x \in \mathrm{dom} f$ if $\forall\, \varepsilon > 0\ \exists\, \delta$ s.t.
$y \in \mathrm{dom} f$ $\|y-x\|_2 \le \delta \Rightarrow \|f(y) - f(x)\|_2 \le \varepsilon$ ⇝ $\lim_{i\to\infty} f(x_i) = f(\lim_{i\to\infty} x_i)$

Closed: A function is closed if for each $\alpha \in \mathbb{R}$ the sublevel set C_α is closed. ≡ epi f is closed

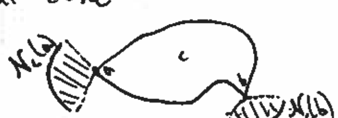
~ if f is continuous ($\mathbb{R}^n \to \mathbb{R}$) & dom f is closed, f is closed. if dom f is open, f is closed iff $f \to \infty$ along every sequence converging to the boundary point of dom f

---

Linear Algebra:

- Range: $A \in \mathbb{R}^{m\times n}$ $R(A) = \{Ax \mid x \in \mathbb{R}^n\}$ → subspace of $\mathbb{R}^m$
- Null space $N(A) = \{x \mid Ax = 0\}$
- Orthogonal complement: V is subspace of $\mathbb{R}^n$ $V^\perp = \{x \mid z^T x = 0\ \forall z \in V\}$
  $N(A) = R(A^T)^\perp$, $R(A) = N(A^T)^\perp$
- $\det A = \prod_{i=1}^n \lambda_i$, $\mathrm{tr}\, A = \sum_{i=1}^n \lambda_i$; $\|A\|_2 = \max|\lambda_i| = \max\{\lambda_1, -\lambda_n\}$; $\|A\|_F = (\sum_{i=1}^n \lambda_i^2)^{1/2}$
- Pseudo-inverse: $A = U\Sigma V^T$ → $A^\dagger = V\Sigma^{-1}U^T \in \mathbb{R}^{n\times m}$ → $A^\dagger b$ is sol. to least squares
- Schur complement $X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$ $A \in S^k$ if $\det A \ne 0$ then $S = C - B^T A^{-1} B$ is schur complement
  $X \succ 0$ iff $A \succ 0$ & $S \succ 0$ $\det X = \det A \det S$
  if $A \succ 0$ $X \succeq 0$ iff $S \succeq 0$ $X\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix}$ ⇝ $v = B^T A^{-1} u + Sy$ so $y = S^{-1}(v - B^T A^{-1} u)$ so $x = (A^{-1} + A^{-1} B S^{-1} B^T A^{-1})u - A^{-1} B S^{-1} v$
- Suppose $A \succ 0$, consider min $u^T A u + 2v^T B^T u + v^T C v$ w/ variable u $u = -A^{-1}Bv$ and optimal val is
  $\inf_u \begin{bmatrix} u \\ v \end{bmatrix}^T \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = v^T S v$

Normal Cone:



Dual Problem: min $x^T x$ s.t. $Ax = b$
↳ $-(1/4)v^T A A^T v - b^T v$ ⇒ unconstrained concave quadratic max.

Lagrange form:

min $f_0(x)$
s.t. $f_i(x) \le 0$
$h(x) = 0$

[convex if $f_0(x)$ is convex, $f_i(x) \le 0$ is affine & $a^T x = b$]

⇒ $L(x,\lambda,\nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)$
dom $L = D \times \mathbb{R}^m \times \mathbb{R}^p$

lagrange multiplier

inequality equality

domain $D = \bigcap_{i=0}^m \mathrm{dom} f_i \cap \bigcap_{i=1}^p \mathrm{dom} h_i$

Taylors Theorem:

$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \dots + \frac{f^{(k)}(a)}{k!}(x-a)^k + \text{Rem.}$ ← is exact at some ξ b/w x & a

ACM 113 Exam 2 Prep

- "Feasibility problem" - find out whether constraint set is empty
- Box constraints: $l \le x \le u \equiv l - x \le 0$ & $x - u \le 0$.
- Equivalent problems:
  - change of variables
  - slack variables: introducing slack variables for linear inequalities preserves convexity (makes it an affine func)

    ↳ ex: if $f_i(x) \le 0$ iff $\exists s_i \ge 0 : f_i(x) + s_i = 0$ then we can reformulate

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t} & f_i(x) \le 0 \\ & h_i(x) = 0 \end{array} \equiv \begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & s_i \ge 0 \\ & f_i(x) + s_i = 0 \\ & h_i(x) = 0 \end{array}.$$

[Saying that $s_i = -f_i(x) \ge 0$]

  - Eliminating equality constraints: retains convexity.

    ↳ if we can explicitly parameterize all solutions of the equality constraints $h_i(x) = 0$ using some parameter $z \in \mathbb{R}^k$ then we can eliminate the equality constraints.

    ↳ Ex: $\phi : \mathbb{R}^k \to \mathbb{R}^n$: x satisfies $h_i(x)$ iff $\exists z \in \mathbb{R}^k : x = \phi(z)$. Then

$$\equiv \quad \begin{array}{ll} \min & \tilde{f}_i(z) = f_0(\phi(z)) \\ \text{s.t.} & \tilde{f}_i = \phi(z) \le 0. \end{array}$$

[equivalent to original problem].

  - Eliminating linear inequality constraints: $Ax = b$ can be absorbed.

    ↳ solution to $Ax = b$ is in general given by $Fz + x_0$ where $z \in \mathbb{R}^k$ & $F \in \mathbb{R}^{n \times k}$.

$$\equiv \quad \begin{array}{ll} \min & f_0(Fz + x_0) \\ \text{s.t.} & f_i(Fz + x_0) \le 0. \end{array}$$

  - Introducing equality constraints: Assuming new constraints are linear, this preserves convexity

    ↳ ex: if $f = f(Ax+b)$, then you can write it as $f(y)$ w/ $y = Ax + b$

- Optimizing over some variables:
  - Preserves convexity: $\inf\limits_{x,y} f(x,y) = \inf\limits_{x} \inf\limits_{y} f(x,y)$

LP:
$$\begin{array}{ll} \min & c^T x + d \\ \text{s.t.} & Gx \preceq h \\ & Ax = b \end{array}$$

Convex!

Standard form:
$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & x \succeq 0 \\ & Ax = b \end{array}$$

Inequality form:
$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax \preceq b. \end{array}$$

- Converting LPs into standard form:

$$\begin{array}{ll} \min & c^T x + d \\ \text{s.t.} & Gx \preceq h \\ & Ax = b \end{array} \xrightarrow[s_i]{\text{slack var.}} \begin{array}{ll} \min & c^T x + d. \\ \text{s.t} & Gx + s = h \\ & Ax = b \\ & s \succeq 0 \end{array} \xrightarrow{\substack{x = x^+ - x^- \\ x^+, x^- \ge 0}} \begin{array}{ll} \min & c^T x^+ - c^T x^- + d \\ \text{s.t.} & Gx^+ - Gx^- + s = h \\ & Ax^+ - Ax^- = b \\ & s \succeq 0 \\ & x^+ \succeq 0, \; x^- \succeq 0 \end{array}$$

→ stack x^+, x^-, s so you have a standard LP.

- Linear Fractional Programming

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & Gx \preceq h \\ & Ax = b \end{array} \xrightarrow{\text{if the feasible set is nonempty}} \begin{array}{l} \min \; c^T y + dz \\ Gy - hz \preceq 0 \\ Ay - bz = 0 \\ e^T y + fz = 1 \\ z \ge 0. \end{array}$$

$f_0(x) = \dfrac{c^T x + d}{e^T x + f}$ dom $f_0 = \{x \mid e^T x + f > 0\}$

To show equiv, first note that if x is feasible then $y = \dfrac{x}{e^T x + f}$ $z = \dfrac{1}{e^T x + f}$ is feasible w/ same objective function: $c^T y + dz = f_0(x)$.

QP:

General form: $\min \ \frac{1}{2}x^T P x + q^T x + r$
s.t. $Gx \preceq h$
$Ax = b$

$P \in S^n_+,\ G \in \mathbb{R}^{m\times n},\ A \in \mathbb{R}^{p\times n}$

Least Squares:
$\min \|Ax+b\|_2^2 = x^T A^T A x - 2b^T A x + b^T b$ is an unconstrained QP

• sol: $x = A^+ b$. If constrained to $l_i \le x_i \le u_i$, no analytical sol exists.

• Distance b/w polyhedra:

$P_1 = \{x \mid A_1 x \preceq b_1\}$ & $P_2 = \{x \mid A_2 x \preceq b_2\}$ $\quad \text{dist}(P_1, P_2) = \inf \{\|x_1 - x_2\|_2 \mid x_1 \in P_1, x_2 \in P_2\}$

$\equiv \min \ \|x_1 - x_2\|_2^2$
s.t. $A_1 x_1 \preceq b_1$
$A_2 x_2 \preceq b_2$

• Second-order cone programming:

$\min \ f^T x$
s.t. $\|A_i x + b_i\|_2 \le c_i^T x + d_i$
$Fx = g$

$A_i \in \mathbb{R}^{n_i \times n},\ F \in \mathbb{R}^{p\times n},\ A \in \mathbb{R}^{k\times n}$
↳ cone constraint $\equiv$ to requiring $(Ax+b,\ c^T x + d)$ to lie in second order cone $\mathbb{R}^{k+1}$.

SDP: If k is S^n_+, the cone of PSD $k\times k$ matrices, the conic form problem is called an SDP:

$\min \ c^T x$
s.t. $x_1 F_1 + \dots + x_n F_n + G \preceq 0$ → all matrices but A are in S^k, $A \in \mathbb{R}^{p\times n}$
$Ax = b$

$\equiv$ $\min \ \text{tr}(CX)$
s.t. $\text{tr}(A_i X) = b_i$ where everything is PSD.
$X \succeq 0$ $\quad \text{tr}(CX) = \sum_{ij} C_{ij} X_{ij} \equiv$ inner prod of matrices.

Inequality form: $\min \ \text{tr}\ c^T x$
s.t. $x_1 A_1 + \dots + x_n A_n \preceq B$

• Duality: Lagrange dual function is concave even when original problem is not convex
Weak duality: holds for all problems.
Strong duality: primal is convex + Slater's condition:

Slater's: for $\min \ f_0(x)$
s.t. $f_i(x) \le 0$,
$Ax = b$

Slater: $\exists\, x \in \text{relint}\, D$: $f_i(x) < 0$
$Ax = b$
↳ strictly feasible point.

→ if all f_1, f_k are affine then strong duality holds if
$\exists\, x \in \text{relint}\ D$: $f_i(x) \le 0 \quad i = 1, \dots, k$
$f_i(x) < 0 \quad i = k+1, \dots, m$
$Ax = b$.

• Strong Duality holds for any LP provided the primal problem is feasible

• KKT: • necessary & sufficient for strong duality.
• If primal problem is convex, KKT are sufficient for points to be primal & dual optimal.
i.e. if f_i are convex & h_i are affine & $\tilde{x}, \hat{\lambda}, \hat{v}$ are any points that satisfy KKT,
$\tilde{x}, \hat{\lambda}, \hat{v}$ are primal & dual optimal w/ strong duality.

1st conditions $\tilde{x}$ is primal feasible.
Last condition: Gradient vanishes for dual optimality.
Equality constraints enforce strong duality.

ACM 113 Exam 2 Prep cont'd

- Unimodularity: An $m \times n$ matrix A is unimodular if it has rank m, it is integral, and $\det(B) = 0, \pm 1$ for every $m \times m$ submatrix B of A. A square matrix is unimodular if it is integral and has $\det = \pm 1$.

- Unimodular operations:
  - Interchange two columns
  - Add an integer multiple of a column to another column
  - Multiply a column by -1.

- Let U be an $n \times n$ nonsingular matrix. The following are equivalent:
  (i) U is unimodular
  (ii) U and U^{-1} are both integral
  (iii) U^{-1} is unimodular
  (iv) For all $x \in \mathbb{R}^n$, Ux is integral iff x is integral
  (v) U is obtained from the identity matrix by a sequence of unimodular operations.

- Integer Farkas lemma: Let A be a rational matrix & b a rational vector. The system $Ax = b$ admits no integral solution iff $\exists\, u \in \mathbb{R}^m : uA \in \mathbb{Z}^n,\ ub \notin \mathbb{Z}$.

- Minimax theorem: Let $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ be compact convex sets. If $f : X \times Y \to \mathbb{R}$ is a continuous function that is concave-convex, i.e.
  $f(\cdot, y) : X \to \mathbb{R}$ is concave for fixed y, and
  $f(x, \cdot) : Y \to \mathbb{R}$ is convex for fixed x

  Then: $\max\limits_{x \in X} \min\limits_{y \in Y} f(x,y) = \min\limits_{y \in Y} \max\limits_{x \in X} f(x,y)$.

Examples:

- Minimum Euclidean Distance: Min. distance to an affine set is $\min \frac{1}{2}\|x\|_2^2 : Ax = b$.
  - ↳ convex & satisfies Slater's conditions (strong duality always holds for convex quadratic problems)
    - $\therefore p^* = d^* \rightarrow$ Optimal value can be computed analytically as $p^* = d^* = \frac{1}{2} b^T (AA^T)^{-1} b$.
  - ↳ optimal point: for every ν, the point $x(\nu) = -A^T\nu$ achieves the minimum in the definition of the dual function $g(\nu)$. $x^* := x(\nu^*)$, $\nu^* = -(AA^T)^{-1}b$ denotes the optimal dual variable. $x^* = A^T(AA^T)^{-1}b$ is optim

- Linear Optimization:
  - LP in inequality form: $p^* = \min_x c^T x : Ax \leq b$. $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Assume feasibility so strong duality holds.
    - Dual form: $p^* = d^* = \max_\lambda -b^T\lambda : \lambda \geq 0,\ A^T\lambda + c = 0$. → standard form

- SVM: Training set: $(x_i, y_i)_{i=1}^m$: $L(w,b) = \sum_{i=1}^m (1 - y_i(w^T x_i + b))_+$.
  - ↳ Control robustness & guarantee unicity: $\min_{w,b} C \cdot L(w,b) + \frac{1}{2}\|w\|_2^2$. $C > 0$ controls tradeoff b/w robustness & performance
  - Reformulation as QP: $\min_{w,b,v} \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^m v_i : v \geq 0,\ y_i(w^T x_i + b) \geq 1 - v_i\ \ i = 1, \dots, m$.

$$\equiv \min_{w,b,v} \frac{1}{2}\|w\|_2^2 + C v^T \mathbf{1} : v \geq 0 \quad v + Z^T w + by \geq \mathbf{1}.$$

Lagrangian: $\mathcal{L}(w,b,\lambda,\mu) = \frac{1}{2}\|w\|_2^2 + C v^T\mathbf{1} + \lambda^T(\mathbf{1} - v - Z^T w - by) - \mu^T v$. [$\mu$ is sign constraint on v]

$$g(\lambda,\mu) = \min_{w,b} \mathcal{L}(w,b,\lambda,\mu)$$

- ↳ solve for w by taking derivatives $\rightarrow w(\lambda,\mu) = Z\lambda$.
- ↳ Taking derivatives w.r.t. v : $C\mathbf{1} = \lambda + \mu$
- ↳ — " — w.r.t. b : $\lambda^T y = 0$.

$$\therefore g(\lambda,\mu) = \begin{cases} \lambda^T\mathbf{1} - \frac{1}{2}\|Z\lambda\|_2^2 & \text{if } \lambda^T y = 0,\ \lambda + \mu = C\mathbf{1} \\ +\infty & \text{o.w.} \end{cases}$$

Dual problem: $d^* = \max_{\lambda \geq 0, \mu \geq 0} g(\lambda,\mu) = \max_\lambda \lambda^T\mathbf{1} - \frac{1}{2}\lambda^T Z^T Z \lambda : 0 \leq \lambda \leq C\mathbf{1},\ \lambda^T y = 0$.

↳ strong duality holds b/c original problem is a QP.

→ better computational cost b/c $k = Z^T Z \in S_+^m$, m var. + m constraints instead of n & m.

ACM 113 Exam 2 Prep cont'd:

PCA: usual formulation of PCA is not convex

$$\max \; x'Ax \quad \text{s.t.} \; x'x = 1.$$

A is usually symmetric PSD (*)

consider instead

$$\max \; x^T A x \quad \text{s.t.} \; x'x \le 1. \quad (**)$$

$\rightsquigarrow A = P'\Sigma P$

A is PSD $\Rightarrow$ Σ has non-negative diagonals.

$\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_n \ge 0.$

$x = P'y$

$\rightsquigarrow f(y) = y'Ay = x'P'APx = x'\Sigma x = \sigma_1 x_1^2 + \cdots + \sigma_n x_n^2.$

↳ not convex!

Max of (* *) will occur at $x'x = 1$. $\therefore$ subtract σ_1 from f, or for points on boundary of $\mathcal{X}$. Will not change locations of points on boundary because it lowers all values on boundary by same value.

let $g(y) = f(y) - \sigma_1 y'y$. $\rightarrow$ ensures no new global minima on interior of $\mathcal{X}$.

~ Because P is orthogonal, $y'y = x'x$. $\therefore$

$$g(x) = \sigma_1 x_1^2 + \cdots + \sigma_n x_n^2 - \sigma_1 (x_1^2 + \cdots x_n^2) = (\sigma_2 - \sigma_1) x_2^2 + \cdots + (\sigma_n - \sigma_1) x_n^2.$$

B/c $\sigma_1 \ge \sigma_i \; \forall i$, each coefficient is zero or negative.

$\therefore$ (i) g is convex

(ii) g is optimized when $x_2 = x_3 = \cdots = x_n = 0$. ($x'x = 1$ implies $x_1 = \pm 1$ & $y = P(\pm 1, 0, \ldots, 0)'$.