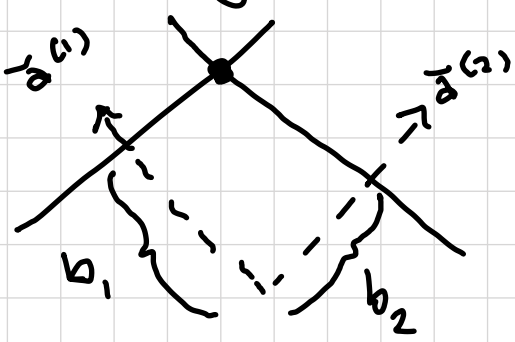
CSE 5401 Lec. 14

Linear Program Duality

1. Solutions to linear optimization problems
2. Weak and strong duality
3. Application to Set Cover

A standard optimization form of linear programming is to minimize a linear cost function subject to linear inequality constraints. The cost function is given by a vector $\vec{c} \in \mathbb{Q}^n$ and the constraints are given by $A \in \mathbb{Q}^{m \times n}$, $\vec{b} \in \mathbb{Q}^m$. Here we focus on rational inputs because this is usually how inputs are represented, but also it is crucial for exact solutions. Notice that we can clear the denominators and get integer costs and constraints $\vec{c} \in \mathbb{Z}^n$, $A \in \mathbb{Z}^{m \times n}$, $\vec{b} \in \mathbb{Z}^m$.

Not all linear programs have a finite optimum. For example, minimize $-x_1$ s.t. $x_1 \geq 0$: we can attain any negative value. But, when we do, the optimum is attained when some of the constraints hold with equality. If n (linearly independent) constraints hold with equality, this gives a system of equations with a unique solution. This solution is a <u>vertex</u> of the feasible region. It is always possible to return a vertex as an optimal solution to the linear program. In this case, the solution is given by



<u>Cramer's Rule</u>: Letting $S \subseteq m$, $A|_S$ be the $n \times n$ matrix given by the rows in S, and $A|_S^i$ be $A|_S$ with the <u>i^{th}</u>

column replaced by $\vec{b}$, the solution to $A|_S \vec{x} = \vec{b}$ is given by $x_i = \frac{\det(A|_S^i)}{\det(A|_S)}$.

Cramer's Rule is important because we can then bound the size of the integers in the rational solution $\vec{x}$:

$\det(A|_S) = \sum_{\text{perm. } \pi} \text{sgn}(\pi) \prod_{i=1}^{n} (A|_S)_{i,\pi(i)}$. There are $n! \leq 2^{n \log n}$ permutations, so if the input integers are all d digits, $\det(A|_S)$ has at most $d \cdot n^2 \log n$ digits, and similarly for $A|_S^i$. If each entry of $\tilde{x}$ is $\frac{1}{2\det(A|_S)}$-close to $\vec{x}^*$, rounding it to the nearest rational number with denominator $\frac{1}{\det(A|_S)}$ will recover $\vec{x}^*$.

We can without loss of generality write the linear program minimize $\vec{c}^T\vec{x}$ s.t. $A\vec{x} \geq \vec{b}$ as minimize $\vec{c}'^T\vec{x}$ s.t. $A'\vec{x} \geq \vec{b}'$, $\vec{x} \geq \vec{0}$

How? For each variable x_i we introduce the pair x_i^+, x_i^- and for each constraint $\vec{a}^{(j)T}\vec{x} \geq b_j$ we can put $a_i'^{(j)+} = a_i^{(j)}$, $a_i'^{(j)-} = -a_i^{(j)}$, and similarly, $c_i'^{+} = c_i$ and $c_i'^{-} = -c_i$. Now, for any feasible $\vec{x}$ for $A\vec{x} \geq \vec{b}$, if $x_i \geq 0$ put $x_i^+ = x_i$ and $x_i^- = 0$; otherwise, for $x_i < 0$, put $x_i^- = -x_i$ (≥ 0) and $x_i^+ = 0$. (Observe, all $x_i^+, x_i^- \geq 0$.)

We now have $a_i^{(j)} x_i = a_i'^{(j)+} x_i^+ + a_i'^{(j)-} x_i^-$ so $\vec{a}'^{(j)T}\vec{x}' = \vec{a}^{(j)T}\vec{x} \geq b_j$

And similarly, $c_i x_i = c_i'^{+} x_i^+ + c_i'^{-} x_i^-$ so $\vec{c}'^T\vec{x}' = \vec{c}^T\vec{x}$. Conversely, For any $\vec{x}'$, put $x_i = x_i^+ - x_i^-$. Then $a_i^{(j)} x_i = a_i^{(j)} x_i^+ + (-a_i^{(j)}) x_i^-$ so if $A'\vec{x}' \geq \vec{b}$, $A\vec{x} \geq \vec{b}$ also, and $c_i x_i = c_i x_i^+ + (-c_i) x_i^-$ so again $\vec{c}^T\vec{x} = \vec{c}'^T\vec{x}'$.

Duality

On homework, you showed that if $A\vec{x} \geq \vec{b}$ (where $\vec{b} \geq 0$) is not feasible for $\vec{x}$ with $\|\vec{x}\|_1 = 1$, then there exists $\vec{w} \geq 0$ s.t. $\vec{w}^T A\vec{x} < \vec{w}^T\vec{b}$ for all $\vec{x}$ with $\|\vec{x}\|_1 = 1$ using the Min-Max Thm.

Essentially the same holds for a more standard form of Linear Programming:

Claim Either there exists $\vec{x}$ s.t. $A\vec{x} \geq \vec{b}$ for $0 \leq \vec{x}$, $\|\vec{x}\|_1 \leq M$ or there exists $\vec{w} \geq 0$ such that for all $\vec{x}$: $0 \leq \vec{x}$, $\|\vec{x}\|_1 \leq M$ $\vec{w}^T A \vec{x} < \vec{w}^T \vec{b}$.

Proof: First observe that we can extend to $0 \leq \vec{x}$, $\|\vec{x}\|_1 \leq M$ by introducing a new variable x_{n+1} with an all-0 column in A and writing $M \cdot A\vec{x} \geq \vec{b}$. Then for any $\vec{x}'$ with $\|\vec{x}'\|_1 \leq M$ we can put $x_{n+1} = 1 - \frac{\|\vec{x}'\|_1}{M}$ and $x_i = \frac{x_i'}{M}$ so $\|\vec{x}\|_1 = \frac{\|\vec{x}'\|_1}{M} + 1 - \frac{\|\vec{x}'\|_1}{M}$ and $M \cdot A\vec{x} = A\vec{x}'$ (and conversely, we can put $x_i' = M \cdot x_i$ for $i = 1, \ldots, n$)

Next, to remove the restriction $\vec{b} \geq 0$, we introduce x_0 and if $b_j < 0$, we put $a_0'^{(j)} = -2b_j$ and $b_j' = 0$ (otherwise, $a_0'^{(j)} = 0$ and $b_j' = b_j$).

Finally, we add the constraint $\sum_{i=1}^{n} x_i \geq \frac{1}{2}$ and put $A' = 2A$ for columns $1, \ldots, n$ (with $\|\vec{x}\|_1 = 1$, $\vec{x} \geq 0$). Now, if $\sum_{i=1}^{n} x_i \geq \frac{1}{2}$ and $x_0 + \sum_{i=1}^{n} = 1$, $x_0 \leq \frac{1}{2}$. so if $\vec{a}'^{(j)T}\vec{x}' \geq b_j'$ for $b_j \geq 0$ $2 \cdot \vec{a}^{(j)T}\vec{x}' \geq b_j$ and $-2b_j + 2\vec{a}^{(j)T}\vec{x}' \geq 0$ for $b_j < 0$ so $2\vec{a}^{(j)T}\vec{x}' \geq 2b_j x_0 \geq b_j$. Hence, $\vec{x} = 2\vec{x}'$ is feasible with $0 \leq \vec{x}$. (Conversely for $\vec{x}$ with $A\vec{x} \geq \vec{b}$ and $\|\vec{x}\|_1 = 1$, $x_0' = \frac{1}{2}$, $x_i' = \frac{x_i}{2}$ has $\|\vec{x}'\|_1 = 1$ and $A'\vec{x}' \geq \vec{b}'$.)

Thus, if there is no $\vec{x}'$ with $\|\vec{x}\| = 1$, $\vec{x}' \geq 0$ satisfying $A'\vec{x}' \geq \vec{b}'$, there is $\vec{w}' \geq 0$ with $\vec{w}^T A' \vec{x}' < \vec{w}^T \vec{b}'$ for all $\vec{x}'$ with $\|\vec{x}'\| = 1$, $\vec{x}' \geq 0$. In particular for $\vec{x}'$ s.t. $x_0' = \frac{1}{2}$, $x_i' = \frac{1}{2}x_i$ we have $\vec{w}_{1:n}'^T A\vec{x} + w_{n+1}' \underbrace{(\sum_{i=1}^{n} x_i)}_{\frac{1}{2}} < \vec{w}_{1:n}'^T \vec{b} + w_{n+1} \cdot \frac{1}{2}$ so: $\vec{w}_{1:n}^T A\vec{x} < \vec{w}_{1:n}'^T \vec{b}$ for all $\vec{x}$ with $\|\vec{x}\|_1 = 1$, $\vec{x} \geq 0$.

For our modified system with x_{n+1} and $a_{n+1}'^{(i)} = 0$, $a_i'^{(j)} = M \cdot a_i^{(j)}$, we also have for all $\vec{x}$ with $\|\vec{x}\|_1 \leq M$, $\vec{x} \geq 0$. ▨

Just like how the players' strategies can be formulated as linear programs, the "hardest" choice of $\vec{w}$ can be found by another linear

program; this is known as the dual program, where we call the original program the primal program:

Primal minimize $\vec{c}^T\vec{x}$ s.t. $A\vec{x} \geq \vec{b}$ for $\vec{x} \geq 0$

Dual maximize $\vec{b}^T\vec{w}$ s.t. $\vec{w}^T A \leq \vec{c}^T$ for $\vec{w} \geq 0$.

The following relationship essentially always holds for such pairs

Theorem ("Weak Duality") If the primal and dual have feasible solutions, dual opt. $\leq$ primal opt.

Proof: For every feasible pair $\vec{x}$ and $\vec{w}$, $\vec{w}^T A\vec{x} \geq \vec{w}^T\vec{b}$ and $\vec{w}^T A\vec{x} \leq \vec{c}^T\vec{x}$ so $\vec{c}^T\vec{x} \geq \vec{w}^T\vec{b}$. This holds for the optimal solutions in particular. ▨

For linear programs, more is true: similar to the min-max Theorem, the two problems have the same value:

Theorem ("Strong Duality") If the primal and dual have feasible solutions, their optimal values are equal.

Proof: By Weak Duality, it suffices to prove $\max \vec{w}^T\vec{b} \geq \min \vec{c}^T\vec{x}$.

Note that Weak Duality implies that these optimal solutions may be assumed to be vertex solutions (since the values are bounded)

Suppose $\vec{x}^*$ is such a solution attaining $\min \vec{c}^T\vec{x} = k$ with $\|\vec{x}^*\|_1 = M$

Then if we add a constraint $-\vec{c}^T\vec{x} \geq -k+\varepsilon$ for any $\varepsilon > 0$, the system is infeasible (even for $\|\vec{x}\|_1 \leq M + \frac{1}{\delta}$ for $\delta > 0$) so our earlier claim gives that there exist $\vec{w}$ and w' such that $\vec{w}^T A\vec{x} - w'\vec{c}^T\vec{x} < \vec{w}^T\vec{b} - w'(k-\varepsilon)$ — in particular, for $\vec{x} = \vec{x}^* - \frac{1}{\delta}\vec{e}_i$

giving $\vec{w}^T(A\vec{x}^* - \vec{b}) + w'(k - \varepsilon - \underbrace{\vec{c}^T\vec{x}^*}_{=k} + \frac{1}{\delta}c_i) < \frac{1}{\delta}(\vec{w}^T A)_i$ (ith std. basis vec.)

i.e., $c_i - \delta\cdot\varepsilon < (\frac{\vec{w}^T}{w'}A)_i$. For sufficiently small δ, rounding $\frac{\vec{w}^T}{w'}$ (≤ 0) gives a feasible dual solution. And, for $\vec{x} = \frac{1}{2}\vec{x}^*$, it satisfies

$\frac{1}{2}\vec{w}^T A\vec{x}^* - w'\cdot\frac{1}{2}\vec{c}^T\vec{x}^* < \vec{w}^T\vec{b} - w'(k-\varepsilon)$ (subtract $\frac{1}{2}\vec{w}^T\vec{b}$ ...)

$\frac{1}{2}\vec{w}^T(A\vec{x}^* - \vec{b}) + w'(k-\varepsilon-\frac{1}{2}k) < \frac{1}{2}\vec{w}^T\vec{1}$ so $k - 2\varepsilon < \frac{\vec{w}^T}{w'}\vec{b}$. Letting $\varepsilon \to 0$, we find indeed that the dual optimum is at least k. ▨ ($\vec{w}^T(A\vec{x}^* - \vec{b}) \le 0$)

Indeed, the primal and dual solutions tell us a lot about the problem, such as which vertices can be optimal solutions

Corollary ("Complementary Slackness") In a pair of optimal primal and dual solutions $\vec{x}^*$ and $\vec{w}^*$ for all j either $\vec{a}^{(j)T}\vec{x}^* = b_j$ or $w_j = 0$ and either $\vec{w}^{*T}A_i = c_i$ or $x_i = 0$.

Proof: $\vec{w}^{*T}\vec{b} = \vec{c}^T\vec{x}^*$ so $\vec{c}^T\vec{x}^* = \vec{w}^{*T}A\vec{x}^* = \vec{w}^{*T}\vec{b}$. Therefore $\vec{w}^{*T}(A\vec{x}^* - \vec{b}) = \vec{0}$ giving the first conclusion and $(\vec{w}^{*T}A - \vec{c}^T)\vec{x}^* = \vec{0}$ giving the second. ▨

It's easiest to see why the dual is useful via examples

Using duality for Set Cover

Recall the set cover problem: we have a universe of m elements that we'll represent by $\{1,\dots,m\}$, and a collection of n subsets $S_i \subseteq \{1,\dots,m\}$ for $i = 1,\dots,n$. We went to find the smallest $\mathcal{S} \subseteq \{1,\dots,n\}$ such that $\bigcup_{i\in\mathcal{S}} S_i = \{1,\dots,m\}$. We say $\mathcal{S}$ is a "set cover".

How can we represent set cover as an integer linear program?

We'll use variables $x_i = \begin{cases} 1 & \text{if } S_i \in \mathcal{S} \\ 0 & \text{o.w.} \end{cases}$ with a constraint for each element $j = 1,\dots,m$ ensuring some $S_i \in \mathcal{S}$ contains j: $\sum_{i:j\in S_i} x_i \ge 1$. To ensure x_i takes only values 0 or 1, we'd include $0 \le x_i$ and $x_i \le 1$. But actually, $x_i \le 1$ is unnecessary: the objective is minimize $\sum_{i=1}^{n} x_i$. In any feasible solution with some $x_i > 0$, we can put $x_i = 1$ and this can only improve the objective while still satisfying all constraints containing x_i.

So, relaxing the requirement that the x_i's are integer, we get the program minimize $\sum_{i=1}^{n} x_i$ s.t. $\forall j \sum_{i:j\in S_i} x_i \geq 1$ and $\vec{x} \geq 0$.
The dual of this program is maximize $\sum_{j=1}^{m} w_j$ s.t. $\forall i \; 1 \geq \sum_{j\in S_i} w_j$, $\vec{w} \geq 0$.

Claim In any optimal solution $\vec{w}^*$ to the dual, the set $\mathcal{d} = \{i : 1 = \sum_{j\in S_i} w_j^*\}$ is a set cover.

Proof: Suppose some j^* is not in $\mathcal{d} = \{i : 1 = \sum_{j\in S_i} w_j\}$. Then $\min_{i:j^*\in S_i} 1 - \sum_{j\in S_i} w_j = \varepsilon > 0$
Increasing w_{j^*} by ε gives another feasible solution with objective greater by $\varepsilon > 0$, so $\vec{w}$ was not optimal. ▨

How large is this compared to optimal? Sometimes it's good:

Theorem Let $d = \max_i |S_i|$. The set $\mathcal{d} = \{i : 1 = \sum_{j\in S_i} w_j^*\}$ for any optimal dual solution $\vec{w}^*$ is at most d times the smallest set cover.

Proof: Since $1 = \sum_{j\in S_i} w_j^*$ for every $i \in \mathcal{d}$, summing over $\mathcal{d}$ gives $|\mathcal{d}| = \sum_{i\in\mathcal{d}} \sum_{j\in S_i} w_i^* \leq \sum_{j=1}^{m} d \cdot w_j^* = d \cdot \sum_{i=1}^{n} x_i^*$ for an optimal primal solution $\vec{x}^*$ by strong duality. Since the primal is a relaxation of the integer program for the set cover problem, $\sum_{i=1}^{n} x_i^*$ is at most the size of the smallest set cover ▨

In particular, vertex cover is a special case of set cover with $d=2$: recall that the problem is, for an undirected graph $G=(V,E)$ to find a minimum size $S \subseteq V$ s.t. for every $(u,v) \in E$ $u \in S$ or $v \in S$. Here, E is the universe and each $v \in V$ corresponds to the set $S_v = \{e \in E : v \text{ is an endpoint of } e\}$. $(u,v) \in E$ is in S_u and S_v and no other sets, so $d=2$, and we get a vertex cover that is at most twice the smallest size. This is conjectured to be the best possible approximation for vertex cover; if d is large, though

It is possible to obtain a tighter approximation by other methods. We'll find many other applications for duality later in the course.