CSE5401 Lec. 17

Primal-Dual Approximation Algorithms

1. Set cover without solving the LP: a "primal-dual" algorithm
2. Primal-dual algorithm for Feedback Vertex Set

Recall our linear program relaxation of the Set Cover problem: we have a universe $\{1,\dots,m\}$, and are given n subsets of the universe, $S_1,\dots,S_n \subseteq \{1,\dots,m\}$. We wish to find the minimum number of these S_i's so that all of $\{1,\dots,m\}$ is contained in at least one. More generally, we could say each set S_i has an associated cost $c_i > 0$ and we wish to minimize the total cost. Letting x_i be the relaxation of an indicator variable for whether or not we use S_i, we have the LP: minimize $\vec{c}^T\vec{x}$ s.t. $\sum_{i: j\in S_i} x_i \geq 1$ for $j \in \{1,\dots,m\}$, $\vec{x} \geq \vec{0}$.

The dual LP is then maximize $\sum_{j=1}^{m} w_j$ s.t. $\sum_{j\in S_i} w_j \leq c_i$ for $i \in \{1,\dots,n\}$, $\vec{w} \geq \vec{0}$.

We noted that if we solved the dual LP to find $\vec{w}^*$, and considered $\mathcal{A} = \{i : \sum_{j\in S_i} w_j^* = c_i\}$ (the indices of "tight" constraints) then $\bigcup_{i\in\mathcal{A}} S_i$ is a set cover, since for any j with $\min_{i: j\in S_i}\left(c_i - \sum_{j\in S_i} w_j\right) = \delta > 0$, we could increase w_j by δ while remaining feasible, thus increasing $\sum_{j=1}^{m} w_j$ by δ as well — hence, $\vec{w}$ would not be optimal for the dual.

(We also saw that if every j appears in at most d sets, $\mathcal{A}$ has cost at most d times the optimal value: it has cost

$$\sum_{i\in\mathcal{A}} c_i \overset{\text{def'n}}{=} \sum_{i\in\mathcal{A}} \sum_{j\in S_i} w_j^* \overset{\text{def'n}}{\leq} \sum_{j=1}^{m} d\cdot w_j^* \overset{\text{strong duality}}{=} d\cdot \sum_{i=1}^{n} x_i^* \overset{\text{relaxation}}{\leq} d\cdot \text{min. cost}.$$)

Solving LPs is polynomial time, but still relatively expensive. It turns out that we don't need to start with the optimal dual solution: the proof that the tight constraints give a set cover essentially describes an algorithm:

Primal-Dual Set Cover $(S_1, \dots, S_n \subseteq \{1, \dots, m\}, \vec{c})$

Initialize $\vec{w} \leftarrow \vec{0}$, $\mathcal{U} \leftarrow \emptyset$

While $\exists j : j \notin \bigcup_{i \in \mathcal{U}} S_i$

Increase w_j by $\min_{i : j \in S_i} \left(c_i - \sum_{j \in S_i} w_j\right)$

For each i s.t. $c_i = \sum_{j \in S_i} w_j$, add i to $\mathcal{U}$

Return $\mathcal{U}$.

If we're careful, this algorithm can be implemented to run in time $O(mn)$. It's immediate from our earlier proof that $\mathcal{U}$ is a set cover.

Theorem Primal-Dual Set Cover returns a set cover of cost at most d times optimal if every j is in at most d sets.

Proof: Notice that $\vec{0}$ is a feasible solution to the dual LP, and furthermore, our increase of w_j in each iteration gives another feasible $\vec{w}$. By weak duality, we know $\sum_{j=1}^{m} w_i \leq \sum_{i=1}^{n} c_i x_i^*$ where $\vec{x}^*$ is the primal optimum. In turn, since the primal LP is a relaxation, $\sum_{i=1}^{n} c_i x_i^*$ is at most the minimum cost of a set cover. Now, we again have our cover's cost $\sum_{i \in \mathcal{U}} c_i = \sum_{i \in \mathcal{U}} \sum_{j \in S_i} w_j \leq \sum_{j=1}^{m} d \cdot w_j \leq d \cdot \sum_{i=1}^{n} c_i x_i^*$ hence, at most d times the minimum cover's cost ▨

Feedback Vertex Set

Let's see another example, where it is crucial that we

don't need to solve the LP. Here, we'll be solving the "Feedback Vertex Set" problem: we are given an undirected graph $G=(V,E)$ and nonnegative costs $c_i \geq 0$ for the vertices. We wish to find $S \subseteq V$ of minimum cost s.t. for every cycle C in G, $C \cap S \neq \emptyset$.

Intuitively, this is a set cover problem, but where the universe is the set of cycles, and vertices corresp. to sets of cycles that contain them. Thus, we get an essentially similar LP relaxation:

minimize $\vec{c}^T \vec{x}$ s.t. $\sum_{i \in C} x_i \geq 1$ for all cycles C in G, $\vec{x} \geq \vec{0}$

... but now there is a (possibly) exponential # of constraints

What is the dual of this LP? We have a variable w_C for each cycle C in G, and maximize $\sum_{\text{cycle } C} w_C$ s.t. $\sum_{\substack{\text{cycle } C: \\ i \in C}} w_C \leq c_i$ for $i \in V$, $\vec{w} \geq 0$.

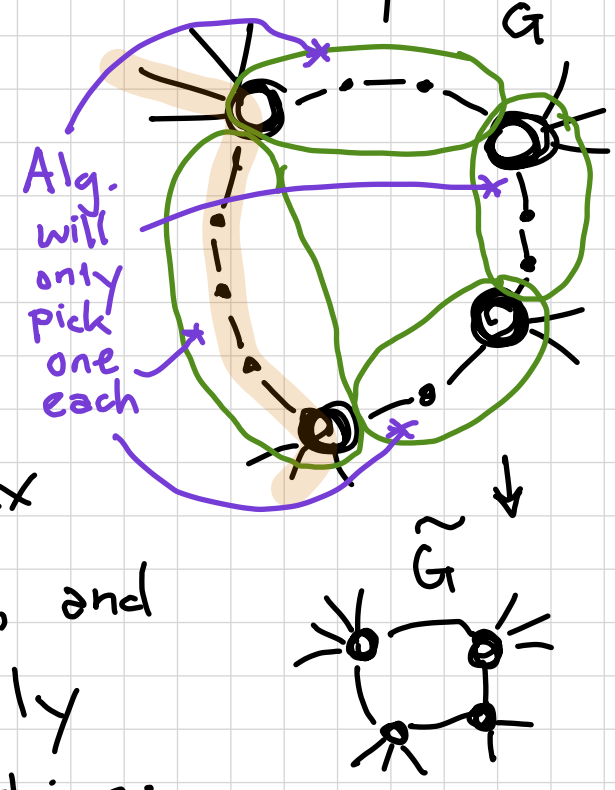
This seems bad! But, remember: we only increase the value of one dual variable at a time until the constraint for i becomes tight. Since there are only $|V|$ vertices, we only set $|V|$ dual variables to nonzero values — so we can use the nonzero values as a poly. size representation of $\vec{w}$.

How do we identify a dual variable (= cycle) to use? Recall that we only obtained a reasonable approximation guarantee when an element appeared in few sets. But, here, the same cycle may contain many

vertices, so we wouldn't get a reasonable approx

Once we pick a vertex, all cycles that contain it are covered, so it will suffice to consider G with that vertex deleted — any remaining cycles of G are cycles in this graph. We can also delete all vertices of degree-1 in this (new) graph, since these cannot be in any cycles. Then, in turn, for any path P of degree-2 vertices in G, whenever some $i \in P$ is chosen, its neighbors have deg 1 after it is deleted, as do the next vertices on the path, and so on. Thus: <u>at most one vertex from each P may be chosen</u>

We therefore focus on the graph $\tilde{G}$ induced by collapsing paths of degree-2 vertices in G to edges; $\tilde{G}$ only has vertices of degree 3 and higher. Since the algorithm can only include one vertex per degree-2 segment on a cycle in G, and the number of such segments is generally equal to the number of degree >2 vertices (unless the entire cycle has degree 2), the number of vertices it will use on a cycle in G is at most twice the number of vertices on the corresponding cycle in $\tilde{G}$



<u>Lemma</u> If G has no degree-1 vertices, then we can find a cycle of length at most $2\log_2|V|$ in the path-collapsed graph $\tilde{G}$ in time $O(|V|+|E|)$.

<u>Proof</u>: We use a breadth-first search in $\tilde{G}$ started

from an arbitrary degree >2 vertex. As soon as we find a vertex reachable via a path of length ℓ that we'd already marked as reachable (via a path of length $\leq \ell$) we return the cycle obtained by the union of these two paths, of length 2ℓ. Notice that until we find such a vertex, every vertex reachable up to distance ℓ must be distinct, and since every vertex in $\tilde{G}$ has degree ≥ 3, there are $\geq 2^\ell$ vertices up to distance ℓ. Since there are only $|V|$ vertices total, $2^\ell \leq |V|$, so $\ell \leq \log_2 |V|$ ▨

Thus, we will generally be able to use cycles that only include $2 \cdot 2\log_2|V|$ vertices in our final set, which will give a $4\log_2|V|$-approximation. Here's our final algorithm:

Primal-Dual Feedback Vertex Set ($G = (V, E)$)

Initialize $\vec{w} \leftarrow \vec{0}$, $S \leftarrow \emptyset$
While G has a cycle
  While G has a degree-1 vertex v remove v from G.
  Find a cycle C with at most $2\log_2|V|$ vertices of degree >2 in G using breadth-first search
  Increase w_C by $\min_{i \in C}\left(c_i - \sum_{\substack{\text{cycle } C' : \\ i \in C'}} w_{C'}\right)$, where i^* attains min
  Add i^* to S
  Remove i^* from G
Return S

Theorem The above algorithm returns a feedback vertex set of cost $\leq 4\log_2|V|$ times the minimum.

Proof: The cost of the returned S is $\sum_{i \in S} c_i = \sum_{i \in S} \sum_{C : i \in C} w_C$

for the dual solution $\vec{w}$ we construct. This is equal to $\sum_{\text{cycle } C} |C \cap S| w_C$. Notice, we only set $w_C > 0$ if w_C had at most $2\log_2|V|$ degree >2 vertices in G at the iteration we found it, where $S \cap C = \emptyset$ on that iteration. Thus, we know that the algorithm can choose at most $4\log_2|V|$ vertices from C across all later iterations, giving $|S \cap C| \leq 4\log_2|V|$ for the final set S. Thus, $\sum_{i \in S} c_i \leq 4\log|V| \sum_{\text{cycles } C} w_C$ $\leq 4\log|V| \sum_{i=1}^{n} c_i x_i^*$ for the primal opt. $\vec{x}^*$ by weak duality since $\vec{w}$ is feasible. Now, since the primal is a relaxation of the feedback vertex set problem, $\sum_{i=1}^{n} c_i x_i^*$ is at most the minimum cost of a feedback vertex set. Thus, S has cost at most $4 \log |V|$ times greater. ▨