UBC CPSC 536N: Sparse Approximations Winter 2013

Lecture 10 — February 4, 2013

Prof. Nick Harvey *Scribe: Alexandre Fréchette*

This lecture is about spanning trees and their polyhedral representation. Throughout the lecture, we "fix" our base graph $G = (V, E)$ to be undirected and connected, with $|V| = n$.

1 Spanning Trees

We start with the basic definitions.

Definition 1.1. A set $T \subseteq E$ is a *spanning tree* if T is connected and acyclic; or T is a maximal cyclic subgraph; or T is acyclic with $|T| = n-1$; or T is a minimal connected spanning subgraph.

It is easy to show that the above defining properties of spanning trees are equivalent. See Figure 1 for a example of a spanning tree.

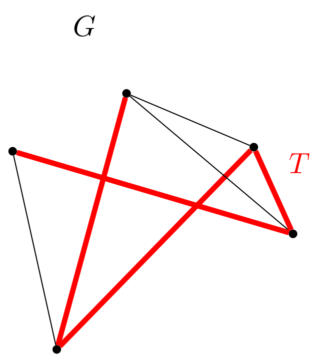

Figure 1: A spanning tree T of an undirected graph G.

Let's see if we can find defining inequalities for spanning trees. Since spanning trees are acyclic, we know they can't select too many edges in a set of nodes. Formally, for any $U \subseteq V$, let $E[U] = \{e = uv \in E : u, v \in U\}$.

Claim 1.2. Let T be a spanning tree of G. Then $|T \cap E[U]| \leq |U| - 1$ for all $U \subseteq V$.

Proof. Consider the subgraph $(U, T \cap E[U])$ of G. By the acyclic property of T, this subgraph is also acyclic. Any acyclic subgraph of $(U, E[U])$ can have at most $|U| - 1$ edges. So $|T \cap E[U]| \leq |U| - 1$. □

1.1 Polyhedral Representation

Let's now see how we can frame spanning trees and the inequality of Claim 1.2 in polyhedral terms.

Definition 1.3. For any $T \subseteq E$, the *characteristic vector* $\chi_T \in \{0,1\}^E$ is defined as

$$\chi_T(e) = \begin{cases} 1 & \text{if } e \in T \\ 0 & \text{otherwise} \end{cases}$$

Moreover, as with our previous notation, for any $C \subseteq E$, $\chi_T(C) = \sum_{e \in C} \chi_T(e) = |T \cap C|$.

Claim 1.2 shows that if T is a spanning tree and $U \subseteq V$ is arbitrary, then $\chi_T(E[U]) \leq |U| - 1$. This is a linear inequality constraint for the vector χ_T. We use these constraints (one for each $U \subseteq V$) to define a polyhedron. Let

$$Q = \{x \in \mathbb{R}^E_{\geq 0} : x(E) = n - 1,\ x(E[U]) \leq |U| - 1\, \forall U \subseteq V\}.$$

Note that $Q \subseteq [0,1]^E$. Indeed, we have that the single entries of the vectors in Q are bounded, as $0 \leq x(e = uv) \leq |\{u, v\}| - 1 = 1$ for every edge $e \in E$.

The following follows easily from our definition of spanning tree and Claim 1.2:

Corollary 1.4. Let T be any spanning tree, then $\chi_T \in Q$.

Hence, Q contains all characteristic vectors corresponding to spanning trees. The remaining work will be concerned with showing that (the extreme points of) Q are precisely the characteristic vectors of spanning trees.

First, we construct an alternate definition for Q.

Definition 1.5. For any $C \subseteq E$, let $\kappa(C)$ be the *number of connected components of* (V, C). Moreover, let $r(C) = n - \kappa(C)$.

An intuition behind $r(C)$ is that it is the largest acyclic set of edges that can be chose from C, that is $r(C) = \max\{|F| : F \subseteq C, F \text{ is acyclic}\}$.

Now define

$$P = \{x \in \mathbb{R}^E_{\geq 0} : x(E) = n - 1,\ x(C) \leq r(C)\, \forall C \subseteq E\}.$$

As we see next, the polytopes P and Q are equivalent. This will be a useful fact because, even though Q may be more "intuitive", P has easier to use structural constraints.

Claim 1.6.

$$P = Q.$$

Proof. We proceed in two steps:

- $P \subseteq Q$;
  Say $x \in P$. Given $U \subseteq V$, let $C = E[U]$. We know that each node in $V \setminus U$ is a singleton connected component in (V, C), hence $\kappa(C) \geq n - |U|$. Moreover, U itself will form at least one big connected components (possibly many more), so in fact $\kappa(C) \geq n - |U| + 1$. Hence,

  $$x(E[U]) = x(C) \leq r(C) = n - \kappa(C) \leq |U| - 1$$

  and thus $x \in Q$.

- $Q \subseteq P$;
  Say $x \in Q$. Given $C \subseteq E$, let the connected components of (V, C) be $\{(V_i, C_i)\}_{i=1}^{\kappa}$, where

naturally $\kappa = \kappa(C)$. Then $x(C) = \sum_{i=1}^{\kappa} x(C_i)$. Since $x \geq 0$ and $C_i \subseteq E[V_i]$, we have that

$$x(C) = \sum_{i=1}^{\kappa} x(C_i) \leq \sum_{i=1}^{\kappa} x(E[V_i]) \leq \sum_{i=1}^{\kappa} |V_i| - 1 = |V| - \kappa = n - \kappa(C) = r(C),$$

so $x \in P$.

□

Both P and Q represent the *spanning tree polytope*. Now we state our main results.

Theorem 1.7. Let x be an extreme point of Q. Then $x = \chi_T$ for some spanning tree T.

It is perhaps not clear at this point whether this theorem is useful — suppose we want to find a spanning tree that optimizes a linear cost function. Can we simply solve the linear program $\max \{ w^\mathsf{T} x : x \in P \}$? One issue is that P is defined by exponentially many linear constraints, so it is not immediately clear that this linear program can be solved in polynomial time[1].

Our theorem is implied from the following integrality result:

Lemma 1.8. Let x be an extreme point of Q. Then $x \in \{0,1\}^E$.

Let's use the lemma to show our theorem.

Proof of Theorem 1.7. Let x be an extreme point of Q. Since $x \in \{0,1\}^E$ from Lemma 1.8, $x = \chi_T$ for *some* $T \subseteq E$. We know that $|T| = x(E) = n - 1$. For T to be an extreme point, we only need it to be acyclic (according to Definition 1.1). Suppose T is not acylic, and let $C \subseteq T$ be a cycle. Note that $\kappa(C) = n - |C| + 1$ as C is connected. Hence, $r(C) = |C| - 1$. However, $x(C) = |T \cap C| = C$, so $x(C) \not\leq r(C)$, and thus $x \notin P$, so $x \notin Q$ by Claim 1.6; contradiction. So T is acyclic, and is a spanning tree. □

Proving the lemma requires additional machinery.

2 Submodularity, Lattices and Chains

So the bulk of the work remaining is in the proof of Lemma 1.8. Recall that we already proved an integrality result for the st-flow polyhedron, where we used the concept of totally unimodular matrices. However, in our situation, we will need new tools.

We first start with a technical result.

Claim 2.1. Let $A, B, C \subseteq E$ be disjoint set of edges. Then,

$$\kappa(C) - \kappa(A \cup C) \geq \kappa(B \cup C) - \kappa(A \cup B \cup C).$$

Proof. First, note that we can focus on the case where $|B| = 1$. Indeed, induction gives us the remaining cases. As a quick argument, suppose $B = \{b_i\}_{i=1}^t$ and that the result holds for any

[1] The linear program $\max \{ w^\mathsf{T} x : x \in P \}$ *can* be solved in polynomial time by a greedy combinatorial algorithm. The duality theory of linear programming is convenient for proving correctness of that algorithm. Unfortunately we don't have time to discuss this further.

B such that $|B| < t$, then

$$\begin{aligned}\kappa(B \cup C) - \kappa(A \cup B \cup C) &= \kappa(\{b_i\}_{i=1}^t \cup C) - \kappa(A \cup \{b_i\}_{i=1}^t \cup C) \\ &= \kappa(C' \cup \{b^t\}) - \kappa(A \cup C' \cup \{b^t\})\end{aligned}$$

where $C' = C \cup \{b_i\}_{i=1}^{t-1}$, using the result for $|B| = 1$ gives us

$$\begin{aligned}&\leq \kappa(C') - \kappa(A \cup C') \\ &= \kappa(C \cup \{b_i\}_{i=1}^{t-1}) - \kappa(A \cup C \cup \{b_i\}_{i=1}^{t-1})\end{aligned}$$

and using the induction hypothesis yields

$$\kappa(B \cup C) - \kappa(A \cup B \cup C) \leq \kappa(C) \cup \kappa(A \cup C).$$

So assume $B = \{b\}$. The left-hand-side of the inequality we want to show corresponds to the number of components of (V, C) that get connected by A. Similarly, the right-hand-side is the number of components of $(V, B \cup C)$ that get tied together by A. Intuitively, there are less components to tie together in $(V, B \cup C)$ that there are in (V, C), which explains the inequality.

To formalize the argument, consider the following three cases:

- The endpoints of b are in the same component of C;
  In this case, $\kappa(B \cup C) = \kappa(C)$ and $\kappa(A \cup B \cup C) = \kappa(A \cup C)$ as adding b does not connect two distinct components of (V, C). Hence, $\kappa(C) - \kappa(A \cup C) = \kappa(B \cup C) - \kappa(A \cup B \cup C)$.

- The endpoints of b are in the same component of $A \cup C$, but not in the same component of C alone;
  So adding b does not connect two distinct components of $(V, A \cup C)$, so $\kappa(A \cup B \cup C) = \kappa(A \cup C)$, but does connect two components of (V, C), so $\kappa(B \cup C) = \kappa(C) - 1$. Hence, $\kappa(B \cup C) - \kappa(A \cup B \cup C) = \kappa(C) - 1 - \kappa(A \cup C) \leq \kappa(C) - \kappa(A \cup C)$.

- The endpoints of b are not in the same component of $A \cup C$;
  This means b connects two distinct components both in $(V, A \cup C)$ and (V, C), so $\kappa(B \cup C) = \kappa(C) - 1$ and $\kappa(A \cup B \cup C) = \kappa(A \cup C) - 1$, and thus $\kappa(B \cup C) - \kappa(A \cup B \cup C) = \kappa(C) - 1 - \kappa(A \cup C) + 1 = \kappa(C) - \kappa(A \cup C)$.

□

The key new tool we use for our integrality result is in the following concept.

Definition 2.2. A set function $f : \wp(X) \to \mathbb{R}$ is *submodular* if

$$f(S) + f(T) \geq f(S \cup T) + f(S \cap T)$$

for all $S, T \in \wp(X)$. Here $\wp(X)$ denotes the **power set** of X, i.e., the collection of all subsets of X.

Claim 2.3. The function $r : \wp(E) \to \mathbb{R}$ from Definition 1.5 is submodular.

Proof. We show that for any $S, T \in \wp(E)$,

$$r(S) + r(T) \geq r(S \cup T) + r(S \cap T).$$

This is quite straightforward from the result of Claim 2.1:

$$\begin{aligned} \kappa(C) - \kappa(A \cup C) &\geq \kappa(B \cup C) - \kappa(A \cup B \cup C) \\ \kappa(C) + \kappa(A \cup B \cup C) &\geq \kappa(A \cup C) + \kappa(B \cup C). \end{aligned}$$

Let $A = S \setminus T$, $B = T \setminus S$ and $C = S \cap T$, then

$$\begin{aligned} \kappa(S \cap T) + \kappa(S \cup T) &\geq \kappa(S) + \kappa(T) \\ -\kappa(S \cap T) - \kappa(S \cup T) &\leq -\kappa(S) - \kappa(T) \\ (n - \kappa(S \cap T)) + (n - \kappa(S \cup T)) &\leq (n - \kappa(S)) + (n - \kappa(T)) \\ r(S \cap T) + r(S \cup T) &\leq r(S) + r(T). \end{aligned}$$

□

The submodularity of r will yield interesting structure in the tight constraints of P.

Definition 2.4. Let $x \in P$. A set $C \subseteq E$ is *tight* for x if $x(C) = r(C)$. Let $\mathcal{T}_x = \{C \subseteq E : x(C) = r(C)\}$ be the collection of tight sets at x.

Note that E is always in $\mathcal{T}_x$, since $x(E) = n - 1 = r(E)$. One of the main trick to unravel such tight sets structure is *uncrossing*.

Claim 2.5. Let S and T be tight for $x \in P$. Then $S \cup T$ and $S \cap T$ are also tight.

Proof. Since $x \in P$, we know that $r(S \cup T) \geq x(S \cup T)$ and $r(S \cap T) \geq x(S \cap T)$. Furthermore, note that $x(S \cup T) + x(S \cap T) = x(S) + x(T)$. Indeed, $x(S) = x(S \setminus T) + x(S \cap T)$, $x(T) = x(S \setminus T) + x(S \cap T)$ and $x(S \cup T) = x(S \setminus T) + x(T \setminus S) + x(S \cap T)$, so

$$x(S \cup T) + x(S \cap T) = x(S \setminus T) + x(T \setminus S) + x(S \cap T) + x(S \cap T) = x(S) + x(T).$$

Piecing those two observations together yields

$$r(S \cup T) + r(S \cap T) \geq x(S \cup T) + x(S \cap T) = x(S) + x(T) = r(S) + r(T) \geq r(S \cup T) + r(S \cap T),$$

and thus equality must hold throughout, so

$$r(S \cup T) + r(S \cap T) = x(S \cup T) + x(S \cap T).$$

Finally, since $x(S \cup T) \leq r(S \cup T)$ and $x(S \cap T) \leq r(S \cap T)$, we must in fact have individual equality $x(S \cup T) = r(S \cup T)$ and $x(S \cap T) = r(S \cap T)$, and thus $S \cup T$ and $S \cap T$ are also tight. □

Claim 2.5 say that $\mathcal{T}_x$ forms a *lattice* — $S, T \in \mathcal{T}_x$ implies $S \cup T, S \cap T \in \mathcal{T}_x$.

Definition 2.6. A sequence of sets $\{C_i\}_{i=1}^{k}$ is a *chain* if $C_i \subseteq C_{i+1}$ for all $i \in [k-1]$.

The properties of maximal chains in lattices will be the key to get our desired result.

Again fix x to be an extreme point of P. Let $\emptyset = C_0 \subset C_1 \subset C_2 \subset \ldots \subset C_{k-1} \subset C_k = E$ be an inclusion-wise maximal chain in $\mathcal{T}_x$. Let $C'_i = C_i \setminus C_{i-1}$ — so $C_i = \bigcup_{j \le i} C'_j$ and the C'_i's are disjoint.

Lemma 2.7. For all $S \in \mathcal{T}_x$,

$$\chi_S \in \mathsf{span}(\{\chi_{C_i} : i \in [k]\}).$$

Proof. Fix a tight set $S \in \mathcal{T}_x$. First, suppose there exists a $j \in [k]$ such that $S \cap C'_j$ is a proper subset of C'_j (i.e. $S \cap C'_j \notin \{\emptyset, C'_j\}$ — see Figure 2). Let $C^* = (S \cap C'_j) \cup C_{j-1} = (S \cap C_j) \cup C_{j-1}$. Since $\mathcal{T}_x$ is a lattice, $C^* \in \mathcal{T}_x$. However, $C_{j-1} \subset C^* \subset C_j$; contradiction to the maximality of our chain.

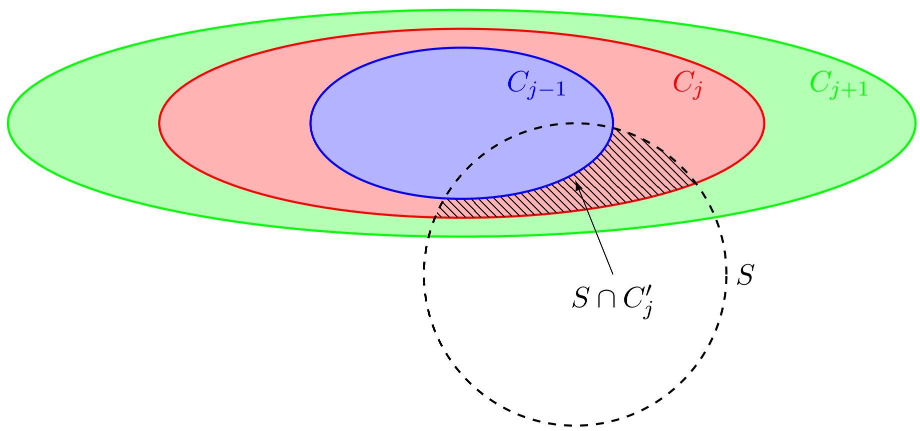

Figure 2: A set $S \in \mathcal{T}_x$ partially intersecting a C'_j.

So for every tight set $S \in \mathcal{T}_x$, we must have $S \cap C'_j \in \{\emptyset, C'_j\}$ for all $j \in [k]$. Since $S \subseteq E = C_k$, there is a $J_S \subseteq [k]$ so that $S = \bigcup_{j \in J_S} C'_j$, so

$$\chi_S = \sum_{j \in J_S} \chi_{C'_j} = \sum_{j \in J_S} (\chi_{C_j} - \chi_{C_{j-1}})$$

and thus $\chi_S \in \mathsf{span}(\{\chi_{C_j} : j \in [k]\})$. □

We now have the tools required to finalize our target result.

Proof of Lemma 1.8. Let x^* be our extreme point. Note that it is a basic feasible solution, so its tight constraints span the whole space $\mathbb{R}^E$. Formally, this means $\mathsf{span}(\{\chi_S : S \in \mathcal{T}_{x^*}\} \cup \{e_i : x^*_i = 0\}) = \mathbb{R}^E$. One thing to note here is that the tight constraints at x^* are not only from $\mathcal{T}_{x^*}$. They can also be tight non-negativity constraints coming from the requirement that $x \ge 0$ in P. That is why we also add the possibly tight $e_i^T x^* \ge 0$.

By Lemma 2.7, we have that $\mathsf{span}(\{\chi_{C_i} : i \in [k]\} \cup \{e_i : x^*_i = 0\}) = \mathbb{R}^E$, where $\{C_i\}_{i=1}^k$ is our inclusion-wise maximal chain in $\mathcal{T}_{x^*}$.

Our first step is to reorganize the coordinates so that all the zero entries of x^* are at the end (i.e. there is a l so that $x^*_i = 0$ implies $i > l$). Hence, since x^* is a basic feasible solution,

$x^* = \begin{bmatrix} y^* \\ z^* \end{bmatrix}$ is the (unique) solution to

$$\begin{bmatrix} & & & \chi_{C_1}^T & & & & \\ & & & \chi_{C_2}^T & & & & \\ & & & \vdots & & & & \\ & & & \chi_{C_k}^T & & & & \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} r(C_1) \\ r(C_2) \\ \vdots \\ r(C_k) \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where $y \in \mathbb{R}^l$ and $z \in \mathbb{R}^{m-l}$.

Because of the lower portion of the constraint matrix, any solution $\begin{bmatrix} y \\ z \end{bmatrix}$ to the above must have $z = 0$. Hence, y^* is the unique solution to

$$\begin{bmatrix} \chi_{C_1 \cap [l]}^T \\ \chi_{C_2 \cap [l]}^T \\ \vdots \\ \chi_{C_k \cap [l]}^T \end{bmatrix} y = \begin{bmatrix} r(C_1) \\ r(C_2) \\ \vdots \\ r(C_k) \end{bmatrix},$$

where $\chi_{C_i \cap [l]} \in \mathbb{R}^l$ is χ_{C_i} restricted to the first l coordinates. Notice that this restriction doesn't affect the inclusion property of our chain, that is $C_i \cap [l] \subseteq C_{i+1} \cap [l]$ for all $i \in [k-1]$. Hence, we can reorder the columns (again) so that the matrix is a form where the 1s are all left-aligned (i.e., no 1 is to the right of a 0) and lower rows have more ones (i.e., no 0 is beneath a 1). For example,

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & 1 & \dots & 1 \end{bmatrix} y = \begin{bmatrix} r(C_1) \\ r(C_2) \\ \vdots \\ r(C_k) \end{bmatrix}.$$

Finally, the above system has a unique solution y^*, so it has full column rank. Hence, we can delete rows to get a square $l \times l$ non-singular matrix. This last updated matrix must be lower triangular. Indeed, there are no full zero rows or identical rows by non-singularity. Hence, y^* is the unique solution to

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix} y = \begin{bmatrix} r(C_{\alpha_1}) \\ r(C_{\alpha_2}) \\ \vdots \\ r(C_{\alpha_l}) \end{bmatrix}.$$

Hence, $y_1^* = r(C_{\alpha_1})$ and $y_1^* + y_2^* = r(C_{\alpha_2})$, and more generally,

$$y_i^* = r(C_{\alpha_i}) - r(C_{\alpha_{i-1}}).$$

Thus, y^* is integral since the $r(C_j)$'s are integral, and thus x^* is integral. □