Directed Steiner Tree and the Lasserre Hierarchy

Thomas Rothvoß

Presented by Martin Koutecký

Combinatorics and Graph Theory PhD Seminar, 2014, MFF UK

Directed Steiner Tree

D(Directed Steiner Tree):

Input: directed weighted graph $G = (V, E, c)$, **root** $r \in V$, **terminals** $X \subseteq V$

Goal: Find a tree T connecting r and X, minimizing $c(T)$.

W.l.o.g. G is acyclic.

T(Nodes in ℓ layers (Zelikovsky '97)): For every $l \geq 1$, there is a tree T (potentially using edges in the metric closure) of cost $c(T) \leq \ell \cdot |X|^{1/l} \cdot OPT$ such that every r-s path (with $s \in X$) in T contains at most l edges.

That is: Modulo $O(\log |X|)$ factor, we may assume $\ell = \log |X|$ layers.

Known results:

- Generalizes Set Cover, (Non-Metric / Multi-level) Facility Location, Group Steiner Tree
- $\Omega(\log^{2-\epsilon} n)$-hard (Halperin, Krauthgamer '03)
- $|X|^\epsilon$-apx in poly by sophisticated greedy algo (Zelikovsky '97)
- $O(\log^3 |X|)$-apx in $n^{O(\log |X|)}$ time by more sophisticated greedy algo (Charikar, Chekuri, Cheung, Goel, Guha and Li '99)
- LP's have integrality gap $\Omega(\sqrt{k})$ already for 5 layers; existing attempts fail (Alon, Moitra, Sudakov '12).

Lasserre Hierarchy

Intuition: Take a polytope $K = \{x \in \mathbb{R}^n | Ax \geq b\}$ such that $K_I := K \cap \{0,1\}^n$ precisely corresponds to **solutions** of a given problem (e.g. Independent Set). Then every point $x \in \text{conv}(K_I)$ corresponds to a **probability distribution** of valid solutions. The problem is that in general $K \neq K_I$, so $x \in K$ may not correspond to a distribution of valid solutions.

What is wrong with $x \in K \setminus \text{conv}(K_I)$? It only gives us valid *marginal* probabilities, but it might not satisfy e.g. $\Pr[X_i = X_j = 1] \in [\max\{x_i + x_j - 1, 0\}, \min\{x_i, x_j\}]$ for $i \neq j$. What about introducing variables for every $I \subseteq [n], |I| = 2$? That would ensure the solution to be sensible w.r.t. second moments. Now generalize to higher moments...

Notation: Let $\mathcal{P}_t([n]) := \{I \subseteq [n] \mid |I| \leq t\}$ be the set of all index sets of cardinality at most t and let $y \in \mathbb{R}^{\mathcal{P}_{2t+2}([n])}$ be a vector with entries y_I for all $I \subseteq [n]$ with $|I| \leq 2t + 2$. Intuitively $y_{\{i\}}$ represents the original variable x_i and the new variables y_I represent $\prod_{i \in I} x_i$.

D(Moment matrix): $M_{t+1}(y) \in \mathbb{R}^{\mathcal{P}_{t+1}([n]) \times \mathcal{P}_{t+1}([n])}$:

$$M_{t+1}(y))_{I,J} := y_{I \cup J} \quad \forall |I|, |J| \leq t + 1.$$

D(Moment matrix of slacks): For the ℓ-th ($\ell \in [m]$) constraint of the LP $A^T x \geq b$, we create $M_t^\ell(y) \in \mathbb{R}^{\mathcal{P}_t([n]) \times \mathcal{P}_t([n])}$:

$$M_t^\ell(y)_{I,J} := (\sum_{i=1}^{n} A_{li} y_{I \cup J \cup \{i\}}) - b_l y_{I \cup J}$$

D(t-th level of the Lasserre hierarchy): Let $K = \{x \in \mathbb{R}^n \mid Ax \geq b\}$. Then $\text{Las}_t(K)$ is the set of vectors $y \in \mathbb{R}^{\mathcal{P}_{2t+2}([n])}$ that satisfy

$$M_{t+1}(y) \succeq 0; \qquad M_t^\ell(y) \succeq 0 \quad \forall \ell \in [m]; \qquad y_\emptyset = 1.$$

Furthermore, let $\text{Las}_t^{\text{proj}} := \{(y_{\{1\}}, \ldots, y_{\{n\}}) \mid y \in \text{Las}_t(K)\}$ be the projection on the original variables.

Intuition: $M_{t+1}(y) \succeq 0$ ensures *consistency* (y behaves *locally* as a distribution) while $M_t^\ell(y) \succeq 0$ guarantees that y satisfies the l-th linear constraint.

T(Lasserre properties): Let $K = \{x \in \mathbb{R}^n \mid Ax \geq b\}$ and $y \in \text{Las}_t(K)$. Then the following holds:

(a) $\text{conv}(K \cap \{0,1\}^n) = \text{Las}_n^{\text{proj}}(K) \subseteq \ldots \subseteq \text{Las}_0^{\text{proj}}(K) \subseteq K$.

(b) We have $0 \leq y_I \leq y_J \leq 1$ for all $I \supseteq J$ with $0 \leq |J| \leq |I| \leq t$.

(c) Let $I \subseteq [n]$ with $|I| \leq t$. Then

$$K \cap \{x \in \mathbb{R}^n \mid x_i = 1 \; \forall i \in I\} = \emptyset \implies y_I = 0.$$

(d) Let $I \subseteq [n]$ with $|I| \leq t$. Then

$$y \in \text{conv}(\{z \in \text{Las}_{t-|I|}(K) \mid z_{\{i\}} \in \{0,1\} \; \forall i \in I\}).$$

(e) Let $S \subseteq [n]$ be a subset of variables such that not many can be equal to 1 at the same time:

$$\max\{|I| : I \subseteq S; x \in K; x_i = 1 \; \forall i \in I\} \leq k < t.$$

Then we have

$$y \in \text{conv}(\{z \in \text{Las}_{t-k}(K) \mid z_{\{i\}} \in \{0,1\} \; \forall i \in S\}).$$

(f) For any $|I| \leq t$ we have $y_I = 1 \Leftrightarrow \bigwedge_{i \in I}(y_{\{i\}} = 1)$.

(g) For $|I| \leq t$: $(\forall i \in I : y_{\{i\}} \in \{0,1\}) \implies y_I = \prod_{i \in I} y_{\{i\}}$.

(h) Let $|I|, |J| \leq t$ and $y_I = 1$. Then $y_{I \cup J} = y_J$.

Remark: Property (e) is very strong and does not hold for the Sherali-Adams or Lovász-Schrijver hierarchy. For example, it implies that after $t = O(\frac{1}{\varepsilon})$ rounds, the integrality gap for the Knapsack polytope is bounded by $1 + \varepsilon$ (taking S as all items that have profit at least $\varepsilon \cdot OPT$). Another example is that the Independent Set polytope $\{x \in \mathbb{R}_+^V \mid x_u + x_v \leq 1 \; \forall \{u, v\} \in E\}$ describes the integral hull after $\alpha(G)$ rounds of Lasserre.

Vector representation: For each event $\bigcap_{i \in I}(x_i = 1)$ with $|I| \leq t$ there is a vector v_I representing it in a consistent way:

L(Vector Representation Lemma):Let $y \in \text{Las}_t(K)$. Then there is a family of vectors $(\mathbf{v}_I)_{|I| \leq t}$ such that $\langle \mathbf{v}_I, \mathbf{v}_J \rangle = y_{I \cup J}$ for all $|I|, |J| \leq t$. In particular $\|\mathbf{v}_I\|_2^2 = y_I$ and $\|\mathbf{v}_\emptyset\|_2^2 = 1$.

The linear program

Idea: Send a unit flow from the root to each terminal $s \in X$ (represented by variables $f_{s,e}$). On each edge we have to pay $y_e = \max\{f_{s,e} | s \in X\}$. We abbreviate $\delta^+(v)$ and $\delta^-(v)$ the edges outgoing and incoming to v, respectively. Also, $y(E') := \sum_{e \in E'} y_e$.

$$\min \sum_{e \in E} c_e y_e$$

$$\begin{aligned}
\sum_{e \in \delta^+(v)} f_{s,e} - \sum_{e \in \delta^-(v)} f_{s,e} &= \begin{cases} 1 & v = r \\ -1 & v = s \\ 0 & \text{otherwise} \end{cases} && \forall s \in X \; \forall v \in V \\
f_{s,e} &\leq y_e && \forall s \in X \; \forall e \in E \\
y(\delta^-(v)) &\leq 1 && \forall v \in V \\
0 \leq y_e &\leq 1 && \forall e \in E \\
0 \leq f_{s,e} &\leq 1 && \forall s \in X \; \forall e \in E
\end{aligned}$$

Lasserre strenghtening: Now we make the choice $t := 2\ell$. Our variable indices are $\mathcal{V}_t = \{(P, H) \mid P \subseteq E; H \subseteq X \times E; |P| + |H| \leq 2t + 2\}$ – that is $\text{Las}_t(K) \subseteq [0,1]^{\mathcal{V}_t}$. Let $Y = (Y_{P,H})_{(P,H) \in \mathcal{V}_t} \in \text{Las}_t(K)$ be an optimum solution for the Lasserre relaxation, which can be computed in time $n^{O(t)}$. We abbreviate $OPT_f := \sum_{e \in E} c_e y_{\{e\}}$ as the objective function value.

We will only address either groups of y_e variables (then we write $y_H := Y_{H,\emptyset}$ for $H \subseteq E$), or we address groups of $f_{s,e}$ variables for the same terminal $s \in X$. Then we write $f_{s,H} := Y_{\emptyset, \{(s,e) | e \in H\}}$.

The rounding algorithm

Idea: Sample a set T of paths from a distribution that depends on Y. Start at layer 0 and go through all layers and for each path P (ending in node u) that is sampled so far, extend it to $P \cup \{(u, v)\}$ with probability $\frac{y_{P \cup \{(u,v)\}}}{y_P}$.

(1) $T := \emptyset$
(2) FOR ALL $e \in \delta^+(r)$ DO
  (3) independently, with prob. $y_{\{e\}}$, add path $\{e\}$ to T
(4) FOR $j = 1, \ldots, \ell - 1$ DO
  (5) FOR ALL $u \in V_j$ and all r-u paths $P \in T$ DO
    (6) FOR ALL $e \in \delta^+(u)$ DO
      (7) independently with prob. $\frac{y_{P \cup \{e\}}}{y_P}$ add $P \cup \{e\}$ to T
(8) return $E(T)$.

Remark: We *do not* remove partial paths, because they will be useful in the analysis.

Notation: $V(P)$ is the set of vertices on path P, $E(T)$ the set of all edges on any path of T, $V(T)$ all vertices of T.

Analysis

We will:

(i) show that for each edge e the probability to be included is $\Pr[e \in E(T)] \leq y_{\{e\}}$.

(ii) prove that for each terminal $s \in X$, the probability to be connected by a path satisfies $\Pr[s \in V(T)] \geq \Omega(\frac{1}{\ell})$.

Part (i) provides that the expected cost for the sampled paths is at most OPT_f, while part (ii) implies that after repeating the sampling procedure $O(\ell \log |X|)$ times, each terminal will be connected to the root with high probability.

Upper bounding the expected cost

Notation: Let $Q(v) := \{P \mid P \text{ is } r\text{-}v \text{ path}\}$ be the set of paths from the root to v. For an edge e let $Q(e)$ be the set of r-v paths that have e as last edge.

L(1):Let P be an r-v path with $v \in V$. Then $\Pr[P \in T] = y_P$.

Each edge e is sampled with probability at most $y_{\{e\}}$:

L(2):For any edge $e \in E$, one has $\sum_{P \in Q(e)} y_P \leq y_{\{e\}}$.

P(2):By induction over the layers. Use Lasserre property (d) and (h).

L(3):$\Pr[e \in E(T)] \leq y_{\{e\}}$ and $E[c(E(T))] \leq \sum_{e \in E} c_e y_{\{e\}}$

P(3):Using Lemmas 1 and 2 and linearity of expectation.

Lower bounding the success probability

L(4):Fix a terminal $s \in X$ and an r-v path P' for some $v \in V$. Then

a) $\sum_{P \in Q(s)} y_P = 1$
b) $\sum_{P \in Q(s): P' \subseteq P} y_P \leq y_{P'}$.

P(4):Consecutively using Lasserre properties (e), (b), (f).

Now fix a terminal $s \in X$ and let $Z := |T \cap Q(s)|$ be the r. v. that yields the number of sampled paths that end in s.

C(5):$E[Z] = 1$ **P**(5):By lemmas 1 and 4(a).

Curiously, we have to prove an *upper* bound on Z in order to *lower* bound $\Pr[Z \geq 1]$.

L(6):$E[Z | Z \geq 1] \leq l + 1$ **P**(6):By lemmas 1 and 4

L(7):$\Pr[Z \geq 1] \geq \frac{1}{l+1}$ **P**(7):By the law of total probability.

Our integrality gap is $O(\ell \log |X|)$:

T(8):Let $Y \in \text{L}{\small AS}_t(K)$ be a given $t = 2\ell$ round Lasserre solution. Then one can compute a feasible solution $H \subseteq E$ with $E[c(H)] \leq O(\ell \log |X|) \cdot \sum_{e \in E} y_{\{e\}}$. The expected number of Lasserre queries and the expected overhead running time are both polynomial in n.
P(8):Repeat the sampling algorithm for $2\ell \log |X|$ many times and let H be the union of the sampled paths.

C(9):$|X|^\epsilon$-apx algo in poly time, or take $\ell = \log |X| \Rightarrow O(\log^3 |X|)$-apx in quasipoly ($n^{O(\log |X|)}$) time.

Q(Open):Is there a convex relaxation with polylog($|X|$) integrality gap that can be solved in poly time? It would suffice to have a polynomial time oracle that takes a path $P \subseteq E$ and outputs the Lasserre entry y_P.