Recent Advances in Approximation Algorithms **Spring 2015**

Lecture 6 & 7: $O(\log(n)/\log\log(n))$ Approximation Algorithms for ATSP

Lecturer: Shayan Oveis Gharan *April 15th and 20th*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

In the next few lectures we will talk about three new variants of the classical randomized rounding technique, namely the rounding by sampling method, the pipage rounding method and the iterative rounding method. We will use these techniques to design approximation algorithms for Asymmetric TSP (ATSP) and the Steiner tree problems. The materials of the this lecture are based on the works of Asadpour, Goemans, Madry, Oveis Gharan and Saberi [Asa+10].

In an instance of ATSP we are given a set V of vertices and a nonnegative cost function $c : V \times V \to \mathbb{R}_+$ that satisfies the triangle inequality, i.e., for any triple of vertices u, v, w,

$$c(u,v) \leq c(u,w) + c(w,v).$$

The goal is to find the shortest tour that visits every vertex at least once. Note that if we find such a tour then by the triangle inequality we can shortcut the tour and obtain a Hamiltonian cycle of the same or smaller cost. The asymmetric terminology stands for the fact that the cost function is not necessarily symmetric, i.e., $c(u,v) \neq c(v,u)$. If the cost function is symmetric, then the problem is called Symmetric TSP or (STSP).

There is a natural linear programming relaxation for ATSP. This was first formulated by Dantzig, Fulkerson and Johnson [DFJ54] which is also known as subtour elimination polytope or Held-Karp LP relaxation (see also [HK70]).

$$\begin{aligned} \text{minimize} \quad & \sum_{u,v} c(u,v)x(u,v) & \\ \text{subject to} \quad & \sum_{u\in S, v\in \overline{S}} x(u,v) \geq 1 & \forall\, S \subsetneq V \\ & \sum_{v\in V} x(u,v) = \sum_{v\in V} x(v,u) = 1 & \forall\, u \in V \\ & x(u,v) \geq 0 & \forall\, u,v \in V. \end{aligned} \tag{6.1}$$

Observe that an *integral* tour that visits each vertex exactly once (i.e., a Hamiltonian cycle) is a feasible solution to the above LP. Therefore, the solution of above LP provides a *lower bound* on the cost of the optimum tour.

One can characterize the solution of ATSP as follows: Say a directed graph G is Eulerian if the indegree of every vertex is equal to its outdegree, and it is connected if the underlying undirected graph is connected. The goal in ATSP is to find an Eulerian connected graph of the smallest cost.

In other words, the set of feasible solutions of ATSP are the families of graphs that are connected and Eulerian. It is easy to optimize each of these two combinatorial properties on its own. That is a connected graph of minimum cost is just a minimum spanning tree and an Eulerian graph of minimum cost is a minimum cycle cover. In the past the main approaches taken by the researchers to tackle either of Asymmetric or Symmetric TSP were to start from an optimal solution of one of the two properties and then add as few edges as possible to get the second property [Chr76; FGM82; Blä02; GLS05; Kap+05; FS07].

i) Start with a minimum cost connected graph, and then add as few edges as possible to make it Eulerian.

ii) Start with an Eulerian graph, i.e., a minimum cost cycle cover, then add edges while preserving the Eulerianness until it becomes connected.

Until the last few years all works on ATSP were based on (ii) and obtained $\Theta(\log n)$ approximation algorithms and the Christofides' work on STSP was based on (i). Although it was expected that the Held-Karp relaxation can guide algorithms to find better quality solutions none of the classical works on TSP effectively used the LP relaxation.

In a nutshell, the recent works on these problems [Asa+10; OSS11; MS11; AKS12; Vis12; AO14; Sve15] effectively use the solution of LP. In particular, instead of simply selecting a minimum cost spanning tree, or a minimum cost cycle cover, they choose a (random) spanning tree (or a cycle cover) based on the solution of the LP. This implies that the cost of the tree or the cycle cover is no more than the optimum in expectation, but it has many more structures that the rounding algorithm can exploit to improve the approximation factor.

The work of [Asa+10] uses approach (i) for ATSP. Given a connected graph T, the Eulerian augmentation of T is a set of edges F such that $T + F$ is Eulerian. It is not hard to see that in undirected graphs one can find the smallest cost Eulerian augmentation in polynomial time; it is the minimum cost matching on the odd degree vertices of the given connected graph. In directed graph this is a variant of the maximum flow problem, so it is still solvable in polynomial time.

6.1 From Connected Subgraphs to Eulerian Connected Subgraphs

Let $G = (V, A, x)$ be the support graph of a (optimal) solution of LP (6.1) and let $T \subseteq A$ be a spanning tree of G. In this section we give an algorithm to find a minimum cost Eulerian augmentation of T, in addition, we show that if T has a certain combinatorial property called *thinness*, then the cost of the Eulerian augmentation is "small" with respect to the cost of the LP solution, $c(x)$.

First, we define the thinness property. Let $\delta^-(S) := \{(u, v) \in A : u \notin S, v \in S\}$ be the edges pointing to S and $\delta^+(S) := \{(u, v) \in A : u \in S, v \notin S\}$ be the edges pointing out of S. Also, let $\delta(S) = \delta^-(S) \cup \delta^+(S)$.

Definition 6.1. *We say that a tree T of $G = (V, A)$ is α-thin with respect to a function $x : A \to \mathbb{R}_+$ if for each set $S \subset V$,*

$$\big||T \cap \delta^-(S)| - |T \cap \delta^+(S)|\big| \leq \alpha \cdot x(\delta(S)).$$

Also we say that T is (α, s)-thin if it is α-thin and moreover,

$$c(T) \leq s \cdot c(x).$$

Note that thinness property can be defined for any subgraph of a graph G. We will refine the above definition later.

In the next theorem we show that if we have a (α, s)-thin tree with respect to a feasible solution x of LP (6.1), then we can obtain an ATSP tour of cost at most $(2\alpha + s)c(x)$. So, this reduces the problem of approximating ATSP to the problem of finding a (α, s) thin tree for small values of α and s.

Theorem 6.2. *Given an (α, s)-thin spanning tree $T \subseteq A$ with respect to x, we can find an Eulerian connected subgraph of G of cost $(2\alpha + s)c(x)$.*

Before proceeding to the proof of Theorem 6.2, we recall some basic network flow results related to circulations. A flow $y : A \to \mathbb{R}$ is called a *circulation* if $y(\delta^+(v)) = f(\delta^-(v))$ for each vertex $v \in V$. Hoffman's circulation theorem [Sch03, Theorem 11.2] gives a necessary and sufficient condition for the existence of a circulation subject to lower and upper capacities on arcs.

Theorem 6.3 (Hoffman's circulation theorem)**.** *For any directed graph $G = (V, A)$, and any two functions $f_l, f_u : A \to \mathbb{R}_+$, there exists a circulation y satisfying $f_l(a) \le y(a) \le f_u(a)$ for all $a \in A$* if and only if

1. *$f_l(a) \le f_u(a)$ for all $a \in A$ and*
2. *for all subsets $S \subset V$, we have $f_l(\delta^-(S)) \le f_u(\delta^+(S))$.*

Furthermore, if f_l and f_u are integer-valued, y can be chosen to be integer-valued.

We do not prove the above theorem. The proof of the existence of the flow y is by an argument similar to the proof of maxflow-mincut theorem. The integrality part follows from the fact that the extreme point solutions of the linear program of all feasible circulations is integral when f_l, f_u are integral. This is because the corresponding feasibility matrix of an extreme point is B_F for some $F \subset A$ of edges, such a matrix is *totally unimodular*, i.e., its determinant is 0 or 1 or -1. (see [Sch03, Corollary 12.2a] for more details).

Now, we are ready to prove Theorem 6.2.

Proof of Theorem 6.2. The minimum cost Eulerian augmentation problem for T can be formulated as a minimum cost circulation problem with lower capacity of 1 on the arcs of T (and no or infinite upper capacities). In particular, let

$$f_l(a) = \begin{cases} 1 & a \in \vec{T} \\ 0 & a \notin \vec{T}, \end{cases} \tag{6.2}$$

and let y be the minimum cost circulation subject to $y(a) \ge f_l(a)$ for all $a \in A$. By Theorem 6.3, y is an integral circulation; furthermore, y can be computed in polynomial time.

Circulation y corresponds to a directed (multi)graph H which contains T as a subgraph. Since it is a circulation, the indegree of every vertex is equal to its outdegree. Therefore H is an Eulerian connected graph, So, we just need to upper bound $c(y)$. Note that we only used the tree T to make sure that H is weakly connected, in other words any connected subgraph of $G = (V, A)$ would give us an Eulerian connected graph as well.

Next, we show that $c(y) \le (2\alpha + s)c(x)$ assuming T is (α, s)-thin. Define

$$f_u(a) = \begin{cases} 1 + 2\alpha \cdot x(a) & a \in T \\ 2\alpha \cdot x(a) & a \notin T. \end{cases}$$

In Claim 6.4 we show that there is a (fractional) circulation y' such that $f_l(a) \le y'(a) \le f_u(a)$ for every $a \in A$. Since y' is a feasible solution for the circulation problem satisfying (6.2) (with no upper bound), and y is the optimum of that problem, $c(y) \le c(y')$. Therefore,

$$c(y) \le c(y') \le c(f_u) = c(T) + 2\alpha \cdot c(x) \le (2\alpha + s) \cdot c(x).$$

This completes the proof of the theorem. □

Claim 6.4. *There is a circulation y' such that for all a, $f_l(a) \le y'(a) \le f_u(a)$.*

Proof. We use Theorem 6.3. First, by definition, for any a, $f_l(a) \le f_u(a)$. So, all we need to show is that for any subset $S \subset V$,

$$f_l(\delta^-(S)) \le f_u(\delta^+(S)).$$

To show the above we use the α-thinness of T and the feasibility of x in (6.1). Since $x(\delta^-(v)) = x(\delta^+(v))$ for all $v \in V$,

$$x(\delta^-(S)) = x(\delta^+(S)) \tag{6.3}$$

Irrespective of the orientation of the arcs of T, the number of arcs of T in $\delta^-(S)$ is at most $\alpha \cdot x(\delta(S))$ by definition of α-thinness. Thus,

$$f_l(\delta^-(S)) = |T \cap \delta^-(S)|.$$

On the other hand, we have

$$\begin{aligned} f_u(\delta^+(S)) &= |T \cap \delta^+(S)| + 2\alpha \cdot x(\delta^+(S)) \\ &= |T \cap \delta^+(S)| + \alpha \cdot x(\delta(S)), \end{aligned}$$

where in the second equality we have used (6.3). Therefore,

$$\begin{aligned} f_u(\delta^+(S)) - f_l(\delta^-(S)) &\geq -\big||T \cap \delta^+(S)| - |T \cap \delta^-(S)|\big| + \alpha \cdot x(\delta(S)) \\ &\geq 0, \end{aligned}$$

where in the second inequality we used the α-thinness of T. Therefore, $f_u(\delta^+(S)) \geq f_l(\delta^-(S))$. □

6.2 In Pursuit of Thin Trees

From now on we only need to think about the thin tree problem. The advantage of this new problem is that we do not need to worry about the direction of the arcs in the graph. From now on, we drop the direction of the edges and we think of G as an undirected graph and we say T is a α-thin spanning tree with respect to x, if for any set S,

$$|T \cap \delta(S)| \leq \alpha \cdot x(\delta(S)).$$

Therefore, we can treat our graph as an undirected graph. In addition, it is not hard to see that if we can find α-thin trees then we can find $(O(\alpha), O(\alpha))$ thin trees as well, so we can also drop the cost of the edges.

Apart from the above simplifications, the problem of finding a thin tree has several inherent hardness. For example, when we drop the direction of the edges, then there is no efficient algorithm to test the thinness of a given tree. We will talk about this later.

In this section we will use the rounding by sampling method to round the solution of Held-Karp relaxation to an $(O(\log(n)/\log\log(n), O(\log(n)/\log\log(n))$ thin tree. Let us first describe the result of applying the classical randomized rounding method of Raghavan and Thompson [RT87]. In this method we independently round each element (edge) of the LP solution. More precisely, for each edge e, we include e in T with probability $\alpha \cdot x_e$ independent of other edges for some suitable parameter $\alpha > 0$. Let X_e be a random variable indicating that $e \in T$. The two main properties of the independent randomized rounding are the following.

i) Any linear function of variables x_e's will remain invariant in expectation, i.e., for any $f: E \to \mathbb{R}$,

$$\mathbb{E}\left[\sum_{e \in E} f(e) X_e\right] = \sum_{e \in E} f(e) x_e.$$

For example, $\mathbb{E}[c(T)] = c(x)$.

ii) Since the variables X_e for $e \in E$ are independent, we can use strong concentration bounds such as Chernoff bounds to argue that any Lipschitz function of these indicator variables is concentrated around its expected value.

Unfortunately, this method does not preserve the combinatorial property of the LP solution; although G is fractionally 2-edge connected, T may be disconnected with high probability unless $\alpha = \Omega(\log(n))$. In particular, if G is a complete graph where $x_{u,v} = 2/(n-1)$, for each vertex $v \in V$, the degree of v in T is zero with probability

$$(1 - 2\alpha/(n-1))^{n-1} \approx \exp(-2\alpha).$$

To make sure that T is connected, we need $\alpha \geq \Omega(\log(n))$. But, in that case $\mathbb{E}\left[c(T)\right] \geq \log(n)c(x)$, so this method can only give us an $O(\log(n))$-thinness. But, it is not hard to see that indeed it does give us an $O(\log(n))$-thinness, so we get another $O(\log(n))$ approximation algorithm for ATSP (see [Goe+09]).

6.2.1 Rounding by Sampling Method

Now, let us describe the rounding by sampling method. Our goal is to design a rounding method that preserves the connectivity with high probability (ideally with probability 1) while satisfying the two main properties of the independent randomized rounding method.

To make sure that we get connectivity for free, we sample a spanning tree based on $x(.)$, so the edges will be correlated, but as we will see later we sample our tree from a weighted uniform distribution to make sure that property (ii) is satisfied.

To round x into a spanning tree we need to show that x or an adjustment of x is in the spanning tree polytope. Recall that the spanning tree polytope is just the convex hull of the indicator vectors of all spanning trees of G.

$$\begin{aligned} &\sum_e z(e) = n-1 & & \\ &\sum_{e \in E(S)} z(e) \leq |S| - 1 & &\forall S \subseteq V \\ &z(e) \geq 0 & &\forall e \in E. \end{aligned} \tag{6.4}$$

Edmonds [Edm70] proved that above linear program is exactly the convex-hull of all spanning trees of graph G, i.e., extreme point solutions of above linear program are exactly the spanning trees of G. Indeed, for any matroid (or any intersection of two matroids) the corresponding linear programming relaxation is integral, we will say more about it in the next lecture.

Now, let us see how we can adjust x to get a point in the above polytope. For an edge $e = (u, v)$ let

$$z(e) = (1 - 1/n)(x(u, v) + x(v, u)); \tag{6.5}$$

it is not hard to see that z is in the spanning tree polytope. For any set $S \subset V$,

$$\begin{aligned} \sum_{e \in E(S)} z(e) &= \frac{1}{2}\left(\sum_{v \in S} z(\delta(v)) - z(E(S, \overline{S}))\right) \\ &= \frac{1 - 1/n}{2}\left(\sum_{v \in S} x(\delta(v)) - x(E(S, \overline{S}))\right) \\ &\leq \frac{1 - 1/n}{2}(2|S| - 2) \leq (|S| - 1) - 1/n. \end{aligned} \tag{6.6}$$

Any feasible point of a polytope can be written as a convex combination of vertices of that polytope (see Carathéodorys theorem). So, we can write z as a convex combination of vertices of the spanning tree polytope, i.e., as a convex combination of spanning trees of G. So, we can write,

$$z = \beta_1 T_1 + \cdots + \beta_k T_k.$$

where $T_1, \ldots, T_k$ are spanning trees of G. Any convex-combination defines a distribution. Therefore, we can define a distribution μ, where $\mathbb{P}_{T\sim\mu}[T = T_i] = \beta_i$. Now, we can round the solution x simply by choosing a random spanning tree from μ. Observe that by definition μ preserves the marginal probabilities imposed by z, $\mathbb{P}_{T\sim\mu}[e \in T] = z(e)$. Therefore, linear functions are preserved in expectation, for any function $f : E \to \mathbb{R}$,

$$\mathbb{E}_\mu\left[\sum_{e\in E} f(e)X_e\right] = \sum_{e\in E} f(e)z(e). \tag{6.7}$$

For example,

$$\mathbb{E}[c(T)] = c(z) = (1 - 1/n)c(x).$$

Furthermore, unlike the independent randomized rounding method, the rounded solution is always *connected*.

An, Kleinberg and Shmoys [AKS12] used the above simple idea to design an improved approximation algorithm for the TSP path problem. We will not talk about their ideas in this course. Also, see [MOS11; LOS12] for other applications of this idea.

Unfortunately, the above simple idea does not necessarily satisfy property (ii) of the independent randomized rounding method, and because of that the rounded solution may not be a thin tree with probability 1.

Example 6.5. *Let G be a complete graph and $z(e) = 2/n$ for all $e \in E$. It is easy to see that z is in the spanning tree polytope. We can write z as a distribution of trees where for each $v \in V$, with probability $1/n$ we choose a star with root v. It is easy to see that for any edge $e = \{u, v\}$,*

$$\mathbb{P}[e \in T] = \mathbb{P}[u \text{ is the root}] + \mathbb{P}[v \text{ is the root}] = 2/n.$$

But, any tree in this distribution is $\Theta(n)$ thin because the root has degree n in the tree while $z(\delta(v)) \leq 2$ for all v.

In the rest of this section we show that if we carefully choose the distribution μ, we are guaranteed to preserve (ii). In particular, we will show that there is a weighted uniform spanning tree distribution such that the marginal probability of each edge e is equal to $z(e)$. This shows that sums of of the indicator random variables of edges are concentrated around the expected value because of negative correlation property between the edges (see notes of Lecture 3).

6.2.2 Introduction to Convex Programming

In this part we give a short overview of the convex programming and Lagrangian dual, and we refer to [BV06] for more details. The standard form of a convex program is defined as follows:

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad \forall 1 \leq i \leq m, \\ & h_i(x) = 0 \quad \forall 1 \leq i \leq m', \end{aligned} \tag{6.8}$$

where $f_0, f_1, \ldots, f_m$ are convex functions and $h_1, \ldots, h_{m'}$ are affine functions, i.e., $h_i(x) = a_i x$ for some vector $a_i \in \mathbb{R}^n$. It is easy to see that the set of feasible solutions of the above program is convex, i.e., if

x_1, x_2 are feasible solutions then so is $(x_1 + x_2)/2$. In general, one can use ellipsoid algorithm or interior point method to approximate the value of the optimum solution of the above program within ϵ accuracy in time polynomial in the size of the program and $\log(1/\epsilon)$.

Given the above program, next, we describe how to write the Lagrangian dual. Let λ_i be the Lagrangian dual variable for the constraint $f_i(x) \leq 0$ and ν_i be the dual variable of the constraint $h_i(x) = 0$. The Lagrangian, $L(x, \lambda, \nu)$ is defined as follows:

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{m'} \nu_i h_i(x).$$

The Lagrangian dual function is defined as follows:

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu).$$

It is easy to see that g is a concave function of λ, ν as it is the minimum of many affine functions.

The weak duality says that for any $\lambda \geq 0$ and ν, and for any feasible x, $g(\lambda, \nu) \leq f_0(x)$, i.e., if x^* is the optimum solution of (6.8),

$$\sup_{\lambda \geq 0, \nu} \inf_x L(x, \lambda, \nu) \leq f_0(x^*).$$

Note that weak duality always hold for convex or nonconvex problems. We say that the strong duality holds if the above equation holds with equality.

Strong duality does not necessarily hold for all convex programs. We say a point x is in the interior of (6.8) if for all $1 \leq i \leq m$, $f_i(x) < 0$. Slater condition says that the strong duality holds for the convex program (6.8) if there is a point x in the interior of the program. We will not say more above convex programs here, and we refer interested readers to [BV06].

6.2.3 Maximum Entropy Rounding by Sampling Method

There are many ways to write a feasible point of a polytope as a convex combination of its vertices. The idea is to use a distribution that maximizes the randomness while preserving the marginal probability of the edges. Roughly speaking, we don't want to enforce any additional artificial structure when writing z as a convex combination of spanning trees. More formally, we write z as a distribution of spanning trees that has the maximum possible *entropy* among all distributions that preserve marginal probability of edges. Asadpour and Saberi first studied and used the *maximum entropy rounding* scheme for sampling a random matching in a bipartite graph with given marginal probabilities [AS07; AS09].

Let $\mathcal{T}$ be the collection of all the spanning trees of $G = (V, E)$. Recall that the entropy of a probability distribution $p : \mathcal{T} \to \mathbb{R}_+$, is simply $\sum_{T \in \mathcal{T}} -p(T) \log(p(T))$. The maximum entropy distribution $p^*(\cdot)$ with respect to given marginal probabilities z is the optimum solution of the following convex program (CP):

$$\begin{aligned} \inf \quad & \sum_{T \in \mathcal{T}} p(T) \log p(T) & \\ \text{subject to} \quad & \sum_{T \ni e} p(T) = z(e) & \forall e \in E, \\ & p(T) \geq 0 & \forall T \in \mathcal{T}. \end{aligned} \tag{6.9}$$

Note that it is implicit in the constraints of this convex program that, for any feasible solution $p(.)$, we have $\sum_T p(T) = 1$ since

$$n - 1 = \sum_{e \in E} z(e) = \sum_{e \in E} \sum_{T \ni e} p(T) = (n-1) \sum_T p(T).$$

Let H^* denote the optimum value of the convex program. Observe that if we remove the equality constraint the optimum distribution of the above convex program will just be the uniform spanning tree distribution. Since any graph has at most n^{n-2} spanning trees [Cay89], H^* always satisfies

$$H^* \geq \log(1/|\mathcal{T}|) \geq -\log(n^{n-2}) \geq -n\log n. \tag{6.10}$$

The above convex program is feasible whenever z belongs to the spanning tree polytope. In addition, the objective function is bounded and the feasible region is compact (closed and bounded), so the infimum is attained and there exists an optimum solution $p^*(\cdot)$. Furthermore, since the objective function is strictly convex, this maximum entropy distribution $p^*(\cdot)$ is unique.

Next, we show that if z is in the interior of the spanning tree polytope of G then $p^*(T)$ is a weighted uniform distribution of spanning trees of G. Note that the vector z obtained from the LP relaxation of the ATSP indeed satisfies this assumption. See (6.6) for the proof.

For this purpose, we write the Lagrangian dual of CP (6.9). For every $e \in E$, we associate a Lagrangian multiplier δ_e to the constraint corresponding to the marginal probability $z(e)$, and we define the Lagrangian function by

$$L(p,\delta) = \sum_{T\in\mathcal{T}} p(T)\log p(T) - \sum_{e\in E}\delta_e\left(\sum_{T\ni e} p(T) - z(e)\right).$$

This can also be written as:

$$L(p,\delta) = \sum_{e\in E}\delta_e z(e) + \sum_{T\in\mathcal{T}}\left(p(T)\log p(T) - p(T)\sum_{e\in T}\delta_e\right).$$

The Lagrange dual to CP (6.9) is now

$$\sup_{\delta}\inf_{p\geq 0} L(p,\delta). \tag{6.11}$$

The inner infimum is easy to solve. As the contributions of the $p(T)$'s are separable, we have that, for every $T \in \mathcal{T}$, $p(T)$ must minimize the convex function

$$p(T)\log p(T) - p(T)\delta(T),$$

where $\delta(T) = \sum_{e\in T}\delta_e$. Taking derivatives, we derive that

$$0 = 1 + \log p(T) - \delta(T),$$

or

$$p(T) = e^{\delta(T)-1}. \tag{6.12}$$

The above equality must be satisfied for any optimal solution of the convex program and any tree T assuming the strong duality holds. These constraints are also known as the KKT conditions. The above equality already shows that the maximum entropy distribution is a weighted uniform distribution after a suitable change of variables (see below). Next, we will write the dual program.

$$\inf_{p\geq 0} L(p,\delta) = \sum_{e\in E}\delta_e z(e) - \sum_{T\in\mathcal{T}} e^{\delta(T)-1}.$$

Using the change of variables $\gamma(e) = \delta_e - \frac{1}{n-1}$ for $e \in E$, the Lagrange dual (6.11) can therefore be rewritten as

$$\sup_{\gamma}\left[1 + \sum_{e\in E} z(e)\gamma(e) - \sum_{T\in\mathcal{T}}\exp(\gamma(T))\right]. \tag{6.13}$$

Our assumption that z is in the interior of the spanning tree polytope implies that the strong duality holds. In particular, any point in the interior of a bounded polytope can be written as a convex combination where all of the vertices of the polytope have positive coefficients (we leave the proof of this as an exercise). So, z can be written as a convex combination of all spanning trees in $\mathcal{T}$ such that the coefficient corresponding to each spanning tree is positive. But this means that there is a point $p(.)$ in the interior of program (6.9). So, the convex program (6.9) satisfies the Slater's condition and the strong duality holds. Therefore, if γ^* is the optimum of (6.13),

$$p^*(T) = \exp(\gamma^*(T)). \tag{6.14}$$

The above distribution is indeed a weighted uniform spanning tree distribution for $w(e) = \exp(\gamma^*(e))$. Summarizing, the following theorem holds.

Theorem 6.6. *Given a vector z in the interior of the spanning tree polytope on $G = (V, E)$, there exist $w : E \to \mathbb{R}_+$ such that the corresponding weighted uniform spanning tree distribution μ satisfies,*

$$\mathbb{P}\left[e \in T\right] = z(e)$$

for all $e \in E$.

It is worth noting that the requirement that z is in the interior of the spanning tree polytope (as opposed to being just in this polytope) is necessary. Let G be a triangle and z be the vector $(\frac{1}{2}, \frac{1}{2}, 1)$. In this case, z is in the polytope (but not in its relative interior) and there are no λ_e^*'s that would satisfy the statement of the theorem (however, one can get arbitrarily close to $z(e)$ for all $e \in E$).

Computing Maximum Entropy Distribution. Although the convex program (6.8) has exponentially many variables, we can use the ellipsoid algorithm to find a near optimal solution of the dual and use that the get a near optimal solution of the problem. Note that by the matrix tree theorem, for any $\gamma : E \to \mathbb{R}$, we can calculate the value of the dual. There is only one technical problem: To use the ellipsoid algorithm we also need to show that there exists a solution to (6.13) such that $\gamma(e) = \text{poly}(n)$ for all $e \in E$. The proof of this is nontrivial we refer interested readers to [SV14] for the proof. One can also use multiplicative update algorithm to design a more efficient algorithm to approximate the maximum entropy distribution, we refer the interested readers to [Asa+10].

Other Applications. Given a set $\mathcal{M} \subseteq \{0,1\}^m$, and a point z in the interior of the convex hull of all vectors of $\mathcal{M}$. We say a distribution $\mu : \mathcal{M} \to \mathbb{R}_+$ is a product distribution if there are weights $w : [m] \to \mathbb{R}_+$ such that for any $S \in \mathcal{M}$,

$$\mathbb{P}\left[S\right] \propto \prod_{i \in S} w(i).$$

We can use the maximum entropy convex program to show that there is a product distribution that preserve $z(i)$ as the marginal probability of element i. This product distribution is not necessarily negatively correlated. Even in the case where $\mathcal{M}$ is the set of perfect matchings of a bipartite graph, some of the edges may be positively correlated in a product distribution of matchings. In the special case where $\mathcal{M}$ is the set of bases of a given vector set $\{v_1, \ldots, v_m\} \in \mathbb{R}^d$, the corresponding product distribution is negatively correlated and it has almost all of the properties of the random spanning tree distributions.

Singh and Vishnoi [SV12] show that the maximum entropy distribution can be (approximately) computed if and only if one can (approximately) calculate $\sum_{M \in \mathcal{M}} \prod_{i \in M} w(i)$ for any given set of weights $w : [m] \to \mathbb{R}_+$. For example, although it is NP-hard to (exactly) count the number of perfect matchings of a bipartite graph, one can use the result of Jerrum, Sinclair and Vigoda [JSV04] to approximately compute the above weighted sum of matchings for a given weight function w, and use that to (approximately) calculate the maximum entropy distribution.

6.2.4 From Random Spanning Trees to Thin Trees

In this section we show that a weighted uniform random spanning tree T sampled from the maximum entropy distribution, μ, that preserve $z(e)$ as the marginal probability of each edge $e \in E$ is $\alpha = O(\log n/\log\log n)$-thin with high probability. Combining this with Theorem 6.2 gives an $O(\log(n)/\log\log(n))$ approximation algorithm for ATSP.

The proof is basically, a Chernoff bound, union bound argument. Recall that the edges of a (weighted) uniform spanning tree distribution are negatively correlated (see lecture 2). So, by Panconesi and Srinivasan theorem (lecture 3) we can use Chernoff types of bound to say the the number of edges of a random spanning is concentrated around its expected value with high probability.

First, observe that since μ preserves $z(e)$ as the marginal of edge e, the expected number of edges of a random spanning tree in any cut $(S, \overline{S})$ is exactly $z(\delta(S)) \approx x(\delta(S))$. The main difficulty in the proof is because of the fact that there are exponentially many cuts in G, and a random spanning tree T is thin, if in each of these exponentially many cuts, $(S, \overline{S})$, T has at most $\alpha \cdot x(\delta(S))$ edges. For a vertex v, since $z(\delta(v)) \approx 2$, by Chernoff bound

$$\mathbb{P}\left[|T \cap \delta(v)| \le 2\alpha\right] \le \left(\frac{e}{1+\alpha}\right)^{2(1+\alpha)}.$$

The above is just an inverse polynomial function of n for $\alpha = \Theta(\log(n)/\log\log(n))$. So, to get a high probability result over exponentially many cuts we need the RHS to be at most $1/2^n$ or $\alpha = \Theta(n)$.

The idea is to use the Benczúr and Karger cut counting argument [BK96]. Karger [Kar93] observed that although a graph has exponentially many cuts, the number of cuts of size close to the minimum value is only polynomially large in n (see the next theorem). So, Becnzúr and Karger came up with the idea of applying Chernoff bound to cuts of size r times the minimum for each $r \in \{1, 2, \dots, n\}$, and then using a union bound for that takes into account the number of cuts of that size.

First, we prove the following result due to Karger [Kar93].

Theorem 6.7 (Karger [Kar93])**.** *For any k-edge connected graph $G = (V, E)$ and any $r \ge 1$ the number of cuts of value at most $r \cdot k$ is at most*

$$O(r \cdot n^{2r}).$$

Proof. This proof is based on [Kar00, Lemma 3.2]. First, we show a simple claim. Fix a spanning tree T of G. We show that for any $F \subseteq T$, there is only one cut of G where its intersection with T is exactly F,

$$|\{(S, \overline{S}) : E(S, \overline{S}) \cap T = F\}| = 1.$$

To see that, fix a vertex u on one side of the cut. Then, for every other vertex v we count the number of edges of F in the unique path from u to v. If this number is even, then v must be the same side of the cut as u, otherwise v is on the other side. Therefore, the number of cuts whose intersection with T has size at most ℓ for some integer $\ell \ge 1$ is exactly

$$\sum_{i=1}^{\ell} \binom{n-1}{i} = O(n^\ell).$$

To prove the lemma, we show that there is a small number of spanning trees of G such that for any cut of size at most $r \cdot k$ at least one of them has a small number of edges in that cut. Nash-Williams [Nas61] show that any k-edge connected graph contains $k/2$ edge disjoint spanning trees. Let $T_1, \dots, T_{k/2}$ be $k/2$ edge disjoint spanning trees of G. For any cut $(S, \overline{S})$ of size $|E(S, \overline{S})| \le rk$, the expected number of edges of a uniformly random tree among $T_1, \ldots, T_{k/2}$ in $(S, \overline{S})$ is at most $2r$,

$$\mathbb{E}\left[|T_i \cap E(S, \overline{S})|\right] \leq \frac{rk}{k/2} = 2r.$$

Therefore, at least one of these trees has at most $2r$ edges in $(S, \overline{S})$. This already gives a bound of $\frac{k}{2} \cdot n^{2r}$ on the number of cuts of size $r \cdot k$.

We can prove a better bound noting that by Markov inequality a uniformly random tree of $T_1, \ldots, T_{k/2}$ has at most $2r$ edges in $(S, \overline{S})$ with probability $1/4r$. In the worst case $1/4r$ fraction of trees have $\lfloor 2r \rfloor$ edges and the rest of them have $\lfloor 2r \rfloor + 1$ edges.

Therefore, if we first choose a uniformly random tree, say T_i, and we count all cuts of T_i with at most $2r$ edges, we have counted at least $1/4r$ fraction of all cuts of size at most $r \cdot k$. So, the number of cuts of size at most $r \cdot k$ is at most $O(r \cdot n^{2r})$. □

Let $\alpha = \Theta(\log n / \log\log n)$ such that $\alpha^\alpha \geq \text{poly}(n)$. First, for any set $S \subset V$, by Chernoff bound

$$\begin{aligned}\mathbb{P}\left[|T \cap \delta(S)| > \alpha \cdot z(\delta(S))\right] &\leq \left(\frac{e}{\alpha}\right)^{\alpha \cdot z(\delta(S))} \\ &\leq n^{-3z(\delta(S))},\end{aligned}$$

using the definition of α.

Now, by Theorem 6.7 there are at most $4n^{2r}$ cuts of size at most r times the minimum cut of $G = (V, E, z)$ for any $r \geq 1$ (note that although we proved Theorem 6.7 for unweighted graphs the same proof holds for G). Now, by the definition of z, G is $2(1 - 1/n)$ edge connected (we think of it as 2 for simplicity), so there are at most $4n^{2r}$ cuts of size at most $2r$ for any $r \geq 1$. Therefore, by union bound, the probability that there is a cut $(S, \overline{S})$ where $|T \cap \delta(S)| > \alpha \cdot z(\delta(S))$ is at most

$$\sum_{r=2}^{\infty} 4n^{2r} n^{-3(r-1)} \leq O(1/n)$$

where each term is an upper bound on the probability that there exists a violated cut of size within $[2r-1, 2r]$. Thus, $T \sim \mu$ is a α-thin spanning tree with probability $(1 - O(1/n)$.

The Upshot The rounding by sampling method does slightly better than the independent randomized rounding method (asymptotically it is better by a factor of $1/\log\log(n)$). The main reason is that instead of sampling edges independently, we use a correlated distribution, i.e., we sample a random spanning tree, this gives the connectivity of the sample for free; so we just need to use Chernoff bound to bound the probability that the tree has more than $\alpha \cdot z(\delta(S))$ edges in a cut $(S, \overline{S})$. The $1/\log\log(n)$ gain is established because the right tail of the Chernoff bound is slightly stronger than its left tail, i.e., for a random variable X,

$$\mathbb{P}\left[X > \alpha \cdot \mathbb{E}\left[X\right]\right] \leq \left(\frac{e}{\alpha}\right)^{\alpha \cdot \mathbb{E}[X]}.$$

Note that for $\alpha = \Theta(\log(n)/\log\log(n))$ the RHS of the above is $1/\text{poly}(n)$. On the other hand,

$$\mathbb{P}\left[X < \mathbb{E}\left[X\right]/2\right] \leq e^{-\mathbb{E}[X]}.$$

So, to get a connected subgraph in the independent randomized rounding method one needs to sample each edge with probability $\Theta(\log(n) z(e))$, or equivalently let $\alpha = \Theta(\log(n))$.

6.2.5 A Tight Example

We conclude this lecture by providing a tight example of the above analysis. We construct a graph where a random spanning tree from the maximum entropy distribution is $\Omega(\log n/\log\log n)$-thin with high probability.

Suppose that we have a complete graph $G = (V, E)$, and $z(e) = 2/(n-1)$ for all $e \in E$. Since all of the edges of G have the same probability, by symmetry, the maximum entropy distribution assigns the same weight to all edges. So, we just sample a uniform spanning tree of a complete graph. But, in assignment 2 we showed that a random spanning tree of a complete graph has a vertex with degree $\Theta(\log(n)/\log\log(n))$ with high probability. This shows that the thinness of a random spanning tree from maximum entropy distribution can be $\Omega(\log(n)/\log\log(n))$.

References

[AKS12] H.-C. An, R. Kleinberg, and D. B. Shmoys. "Improving christofides' algorithm for the s-t path TSP". In: *STOC*. 2012, pp. 875–886 (cit. on pp. 6-2, 6-6).

[AO14] N. Anari and S. Oveis Gharan. "Effective-Resistance-Reducing Flows and Asymmetric TSP". 2014. URL: http://arxiv.org/abs/1411.4613 (cit. on p. 6-2).

[AS07] A. Asadpour and A. Saberi. "An approximation algorithm for max-min fair allocation of indivisible goods". In: *STOC*. 2007, pp. 114–121 (cit. on p. 6-7).

[AS09] A. Asadpour and A. Saberi. *Maximum Entropy Selection: a Randomized Rounding Method*. 2009 (cit. on p. 6-7).

[Asa+10] A. Asadpour, M. X. Goemans, A. Madry, S. Oveis Gharan, and A. Saberi. "An $O(\log n/\log\log n)$-approximation Algorithm for the Asymmetric Traveling Salesman Problem". In: *SODA*. 2010, pp. 379–389 (cit. on pp. 6-1, 6-2, 6-9).

[BK96] A. A. Benczúr and D. R. Karger. "Approximating s-t Minimum Cuts in $\tilde{O}(n^2)$ Time". In: *STOC*. 1996, pp. 47–55 (cit. on p. 6-10).

[Blä02] M. Bläser. "A new approximation algorithm for the asymmetric TSP with triangle inequality". In: *SODA*. 2002, pp. 638–645 (cit. on p. 6-1).

[BV06] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2006 (cit. on pp. 6-6, 6-7).

[Cay89] A. Cayley. "A theorem on trees". In: *Quart. J. Math.* 23 (1889), pp. 376–378 (cit. on p. 6-8).

[Chr76] N. Christofides. *Worst Case Analysis of a New Heuristic for the Traveling Salesman Problem*. Report 388. Pittsburgh, PA: Graduate School of Industrial Administration, Carnegie-Mellon University, 1976 (cit. on p. 6-1).

[DFJ54] G. Dantzig, R. Fulkerson, and S. Johnson. "Solution of a large-scale traveling-salesman problem". In: *Operations Research* 2 (1954), pp. 393–410 (cit. on p. 6-1).

[Edm70] J. Edmonds. "Submodular functions, matroids and certain polyhedra". In: *Combinatorial Structures and Their Applications*. New York, NY, USA: Gordon and Breach, 1970, pp. 69–87 (cit. on p. 6-5).

[FGM82] A. M. Frieze, G. Galbiati, and F. Maffioli. "On the worst-case performance of some algorithms for the asymmetric traveling salesman problem". In: *Networks* 12 (1982), pp. 23–39 (cit. on p. 6-1).

[FS07] U. Feige and M. Singh. "Improved approximation ratios for traveling salesman tours and path sin directed graphs". In: *APPROX*. 2007, pp. 104–118 (cit. on p. 6-1).

[GLS05] D. Gamarnik, M. Lewenstein, and M. Sviridenko. "An improved upper bound for the TSP in cubic 3-edge-connected graphs". In: *Oper. Res. Lett.* 33.5 (Sept. 2005), pp. 467–474 (cit. on p. 6-1).

[Goe+09] M. X. Goemans, N. J. A. Harvey, K. Jain, and M. Singh. "A Randomized Rounding Algorithm for the Asymmetric Traveling Salesman Problem". 2009. URL: http://arxiv.org/abs/0909.0941 (cit. on p. 6-5).

[HK70] M. Held and R. Karp. "The traveling salesman problem and minimum spanning trees". In: *Operations Research* 18 (1970), pp. 1138–1162 (cit. on p. 6-1).

[JSV04] M. Jerrum, A. Sinclair, and E. Vigoda. "A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries". In: *J. ACM* 51.4 (July 2004), pp. 671–697 (cit. on p. 6-9).

[Kap+05] H. Kaplan, M. Lewenstein, N. Shafrir, and M. Sviridenko. "Approximation algorithms for asymmetric TSP by decomposing directed regular multigraphs". In: *J. ACM* (2005), pp. 602–626 (cit. on p. 6-1).

[Kar00] D. R. Karger. "Minimum cuts in near-linear time". In: *J. ACM* 47.1 (Jan. 2000), pp. 46–76 (cit. on p. 6-10).

[Kar93] D. R. Karger. "Global min-cuts in RNC, and other ramifications of a simple min-out algorithm". In: *SODA*. Austin, Texas, United States: Society for Industrial and Applied Mathematics, 1993, pp. 21–30 (cit. on p. 6-10).

[LOS12] B. Laekhanukit, S. Oveis Gharan, and M. Singh. "A Rounding by Sampling Approach to the Minimum Size k-Arc Connected Subgraph Problem". In: *ICALP (1)*. 2012, pp. 606–616 (cit. on p. 6-6).

[MOS11] V. H. Manshadi, S. Oveis Gharan, and A. Saberi. "Online Stochastic Matching: Online Actions Based on Offline Statistics". In: *SODA*. 2011, pp. 1285–1294 (cit. on p. 6-6).

[MS11] T. Moemke and O. Svensson. "Approximating Graphic TSP by Matchings". In: *FOCS*. 2011, pp. 560–569 (cit. on p. 6-2).

[Nas61] C. S. J. A. Nash-Williams. "Edge disjoint spanning trees of finite graphs". In: *J. London Math. Soc.* 36 (1961), pp. 445–45 (cit. on p. 6-10).

[OSS11] S. Oveis Gharan, A. Saberi, and M. Singh. "A Randomized Rounding Approach to the Traveling Salesman Problem". In: *FOCS*. 2011, pp. 550–559 (cit. on p. 6-2).

[RT87] P. Raghavan and C. D. Tompson. "Randomized rounding: a technique for provably good algorithms and algorithmic proofs". In: *Combinatorica* 7.4 (Dec. 1987), pp. 365–374 (cit. on p. 6-4).

[Sch03] A. Schrijver. *Combinatorial Optimization*. English. Vol. 2. Algorithms and Combinatorics. Springer, 2003 (cit. on p. 6-3).

[SV12] A. Sebö and J. Vygen. "Shorter Tours by Nicer Ears:" CoRR abs/1201.1870. 2012 (cit. on p. 6-9).

[SV14] M. Singh and N. K. Vishnoi. "Entropy, optimization and counting". In: *STOC*. 2014, pp. 50–59 (cit. on p. 6-9).

[Sve15] O. Svensson. "Approximating ATSP by Relaxing Connectivity". 2015. URL: http://arxiv.org/abs/1502.02051 (cit. on p. 6-2).

[Vis12] N. K. Vishnoi. "A Permanent Approach to the Traveling Salesman Problem". In: *FOCS*. 2012, pp. 76–80 (cit. on p. 6-2).