Lecture 6 (03.04.2013) Scribe: Michal Wlodarczyk
Lecturer: Marcin Mucha

Advanced Rounding Techniques & Asymmetric Travelling Salesman Problem

1 Dependent rounding: continuation

We begin with a problem stated in the end of lecture 5.

Theorem 1.1. *[Srinivasan et al. [4]] Let $G = (V, E)$ be a bipartite graph and $x : E \to (0, 1)$ be an arbitrary function. Then we can (algorithmically) construct binary random variables X_e such that*

1. *$\mathbb{E}[X_e] = x_e$ for each $e \in E$,*
2. *$\sum_{e\in\delta(v)} X_e \in \{\lfloor d_v \rfloor, \lceil d_v \rceil\}$ for each $v \in V$ where $d_v = \sum_{e\in\delta(v)} x_e$,*
3. *variables X_e for $e \in \delta(v)$ are negatively correlated for each $v \in V$.*

Proof. Let us consider the following algorithm

Algorithm 1: Dependent Rounding

```
while E ≠ ∅ do
    Let C ⊆ E by a cycle, or a non-extendable path;
    Let C = M_1 ∪ M_2, where M_i are matchings;
    α := min{x_e : e ∈ M_1} ∪ {1 − x_e : e ∈ M_2};
    β := min{1 − x_e : e ∈ M_1} ∪ {x_e : e ∈ M_2};
    with probability β/(α+β) do
        foreach e ∈ M_1 do x_e := x_e − α;
        foreach e ∈ M_2 do x_e := x_e + α;
    otherwise
        foreach e ∈ M_2 do x_e := x_e − β;
        foreach e ∈ M_1 do x_e := x_e + β;
    foreach e ∈ E do
        if x_e = 0 then set X_e = 0; E := E − e;
        if x_e = 1 then set X_e = 1; E := E − e;
```

Let $X_{e,k}$ be a random variable equal to the value of x_e after the $(k-1)$-th step. Similarly $M_{i,k}$, α_k and β_k denote matchings and coefficients chosen in the k-th step and $\mathbb{E}(\cdot|F_k)$ means estimated value conditioned on everything that happened before (i.e. it encompasses all previous random choices made by the algorithm). We will now prove each property by induction.

1. Obviously $X_{e,1} = x_e$ and in the last iteration k_0 we have $X_{e,k_0} = X_e$. Therefore it suffices to show $\mathbb{E}(X_{e,k+1}|F_k) = X_e$ for every F_k which implies $\mathbb{E}X_{e,k+1} = \mathbb{E}X_{e,k}$, by total expectation theorem.

$$\mathbb{E}(X_{e,k+1}|F_k) = \begin{cases} X_{e,k} - \alpha\frac{\beta}{\alpha+\beta} + \beta\frac{\alpha}{\alpha+\beta} = X_{e,k} & e \in M_{1,k} \\ X_{e,k} + \alpha\frac{\beta}{\alpha+\beta} - \beta\frac{\alpha}{\alpha+\beta} = X_{e,k} & e \in M_{2,k} \\ X_{e,k} & e \notin M_{1,k} \cup M_{2,k} \end{cases} \tag{1}$$

2. For $k = 1$ thesis is of course satisfied. If $M_{1,k} \cup M_{2,k}$ forms a cycle then every vertex v has zero or two adjacent edges that change value. The first case is trivial. In the second case, the edges belong to different matchings and $\sum_{e\in\delta(v)} X_{e,k+1} - \sum_{e\in\delta(v)} X_{e,k}$ equals $\alpha - \alpha$ (or $\beta - \beta$) $= 0$.

If $M_{1,k} \cup M_{2,k}$ is a path then the previous argument works for every vertex except for the beginning and the end of the path – v_1, v_2. However the path is maximal so $deg(v_1) = deg(v_2) = 1$. The last adjacent edge eventually gets rounded to 0 or 1 so $\sum_{e\in\delta(v_1)} X_e$ becomes one of the two integers closest to $\sum_{e\in\delta(v)} X_{e,k}$ what gives the thesis.

3. Let $S \subseteq \delta(v)$ for some vertex v. We will to show $\mathbb{E}(\prod_{e\in S})X_{e,k+1} \leq \mathbb{E}(\prod_{e\in S})X_{e,k}$ for all k, and the claim follows.
$|S \cap (M_{1,k} \cup M_{2,k})|$ could be 0,1 or 2. If it is 0 then of course nothing changes.

The second case: $S\cap(M_{1,k}\cup M_{2,k}) = \{e_1\}$. We can take advantage of the property $\mathbb{E}(X_{e,k+1}|F_k) = X_{e,k}$ proven in part 1.

In the last case $S \cap M_{1,k} = \{e_1\}, S \cap M_{2,k} = \{e_2\}$.

$$\mathbb{E}(X_{e_1,k+1}X_{e_2,k+1}|F_k) = \tag{2}$$

$$= \frac{\beta}{\alpha+\beta}(X_{e_1,k} - \alpha)(X_{e_2,k} + \alpha) + \frac{\alpha}{\alpha+\beta}(X_{e_1,k} + \beta)(X_{e_2,k} - \beta) = \tag{3}$$

$$= X_{e_1,k}X_{e_2,k} - \alpha\beta \leq X_{e_1,k}X_{e_2,k} \tag{4}$$

$$\mathbb{E}(\prod_{e\in S} X_{e,k+1}|F_k) = \prod_{e\in S-\{e_1,e_2\}} X_{e,k}\ \mathbb{E}(X_{e_1,k+1}X_{e_2,k+1}|F_k) \leq \prod_{e\in S} X_{e,k} \tag{5}$$

Knowing this and with part 1 proven, we can conclude

$$\mathbb{E}(\prod_{e\in S} X_e) \leq \cdots \leq \mathbb{E}(\prod_{e\in S} X_{e,k+1}) \leq \mathbb{E}(\prod_{e\in S} X_{e,k}) \leq \cdots \leq \mathbb{E}(\prod_{e\in S} x_e) = \prod_{e\in S} \mathbb{E}X_e \tag{6}$$

This proves $\mathbb{P}\left(\forall_{e\in S} X_e = 1\right) \leq \prod_{e\in S} \mathbb{P}(X_e = 1)$. The condition $\mathbb{P}\left(\forall_{e\in S} X_e = 0\right) \leq \prod_{e\in S} \mathbb{P}(X_e = 0)$ can be obtained in the same way. Each step of the algorithm can be done in polynomial time. The analysis is complete when we observe that in every step we erase at least one edge, so there are $O(n^2)$ steps. □

Remark 1.2. If the sums $\sum_{e\in\delta(v)} x_e$ happen to be integers then we can guarantee that $\sum_{e\in\delta(v)} X_e$ are respectively equal to them.

2 Swap rounding

In this section we will present another rounding technique, introduced in [3]. Let T_1, T_2 be some spanning trees in G and let $\alpha_1 + \alpha_2 = 1, \alpha_i \geq 0$. We will use characteristic function of a tree: $\chi_T(e) = 1 \Leftrightarrow e \in T$.

We want to sample spanning tree X (X_e would be binary variables saying if edge e belongs to the tree) with conditions

1. $\mathbb{E}X_e = \alpha_1 \chi_{T_1}(e) + \alpha_2 \chi_{T_2}(e)$

2. variables $\{X_e\}$ are negatively correlated

The easiest way to satisfy condition 1 is to sample tree T_1 with probability α_1 and T_2 with α_2. Unfortunately appearances of edges would be highly correlated. We will need a tool from the matroid theory.

Fact 1. *If spanning trees T_1, T_2 are different, we can choose edges $e_1 \in T_1 - T_2, e_2 \in T_2 - T_1$ so that $T_1 - e_1 + e_2$ and $T_2 - e_2 + e_1$ are also spanning trees.*

Remark 2.1. For simplicity, we work on spanning trees in this section, but the reasoning is the same for bases of any matroid.

This fact leads to the following algorithm. As long as $T_1 \neq T_2$ pick e_1, e_2 as in Fact 1, set $T_1 := T_1 - e_1 + e_2$ with probability α_2 and $T_2 := T_2 - e_2 + e_1$ with probability α_1. When $T_1 = T_2$ return T_1.

Please note that edges from $T_1 \cap T_2$ are taken for sure, the ones from $T_1 - T_2$ are rejected with probability α_2 and otherwise they will not be picked any more. Similarly for $T_2 - T_1$. The algorithm stops because $|(T_1 - T_2) \cup (T_2 - T_1)|$ decreases in every step.

Theorem 2.2 (Chekuri et al. [3])**.** *The above schema guarantees negative correlation.*

Remark 2.3. This method is easily generalized for larger sets of trees $T_1 \dots T_n$ and coefficients $\alpha_1 + \dots + \alpha_n = 1$ by iteratively launching the algorithm for some pair T_i, T_j and setting weight $\alpha_i + \alpha_j$ for the resulting tree.

3 Maximum entropy sampling

Let us recall one of the LP formulations for the spanning tree problem.

$$\forall_{\emptyset \neq S \subseteq V} \quad x(E(S)) \leq |S| - 1 \tag{7}$$

$$x(E) = |V| - 1 \tag{8}$$

$$\forall_e \quad x_e \geq 0 \tag{9}$$

Let x lie in the interior of the solutions' polytope. We want to sample integer solution X satisfying:

1. X is a spanning tree (i.e. a vertex of the polytope)

2. $\mathbb{E}X_e = x_e$

3. variables $\{X_e\}$ are negatively correlated

Let π_T be the probability of drawing tree T. The idea is to choose the distribution in such a way it would maximize the entropy function $-\sum_T \pi_T \log \pi_T$ (we sum over the set of all spanning trees in the graph).

Theorem 3.1 (Asadpour, Saberi [2]). *Distribution satisfying 1-3 and maximizing entropy can be expressed as*

$$\pi_T = \prod_{e \in T} \lambda_e \tag{10}$$

for some weights (λ_e) which may be found in polynomial time.

The question remains how to effectively draw trees knowing the numbers (λ_e). We will need a generalized version of the Kirchhoff's theorem.

Theorem 3.2. *Consider a matrix M defined as follows: $M_{ii} = \sum_{e \in \delta(v_i)} \lambda_e$ and for $i \neq j$: $M_{ij} = -\lambda_{ij}$ if $ij \in E$ and 0 otherwise. The algebraic complement of the field $(1,1)$ (which is the determinant of M without first row and first column) equals $\sum_T \prod_{e \in T} \lambda_e$.*

What is the probability that some particular edge e_1 does not belong to T?

$$\mathbb{P}(e_1 \notin T) = \sum_T \prod_{e \in T} \lambda_e^1 \tag{11}$$

where $\lambda_{e_1}^1 = 0$ and $\lambda_e^1 = \lambda_e$ otherwise. Knowing this we can draw a random number $y \in [0,1]$ and if $y \leq \mathbb{P}(e_1 \notin T)$ then we reject e_1 (and remove it from the graph, setting $\lambda_e^1 = 0$), otherwise we include it in the tree (and contract it). We iterate this procedure until all edges are fixed – all we need to know are probabilities

$$\mathbb{P}(e_k \notin T | e_1 \dots e_{k-1} \, \text{fixed}) = \frac{\sum_T \prod_{e \in T} \lambda_e^k}{\sum_T \prod_{e \in T} \lambda_e^{k-1}} \tag{12}$$

where λ_e^k is defined in a way similar to λ_e^1 before. All of them could be obtained using Theorem 3.2 with modified weights. Please note we only use the sums in division so weights do not need to be normalized.

Theorem 3.3 (Asadpour, Saberi [2]). *The above sampling algorithm satisfy conditions 1-3.*

4 Techniques from the flow theory

In this section we present a few theorems concerning flows in graphs which will turn out helpful in the next section. Firstly we recall some definitions.

Definition 1. A circulation in a directed graph $G = (V, A)$ is a function $c : A \longrightarrow \mathbf{R}^+$ satisfying $\sum_{a \in in(v)} c(a) = \sum_{a \in out(v)} c(a)$ for every vertex v. For given cost function $\omega : A \longrightarrow \mathbf{R}^+$ circulation cost equals $\sum_{a \in A} c(a)\omega(a)$.

Definition 2. A cut in an undirected graph $G = (V, E)$ is a division $V = V_1 \cup V_2, V_1 \cap V_2 = \emptyset$. A size of a cut is defined as $|E(V_1, V_2)| = |\{v_1 \in V_1, v_2 \in V_2, v_1 v_2 \in E\}|$ (or as a sum of weights between V_1 and V_2 in the general case). A minimal cut is a cut with the smallest size.

Theorem 4.1 (Hoffman [6]). *Let $G = (V, A)$ be a directed graph and functions $l, u : A \longrightarrow \mathbf{N}$ satisfy $l \leq u$. There exists a circulation c so $l(a) \leq c(a) \leq u(a)$ for every edge a iff*

$$\forall_{S \subset V} \sum_{a \in A(S, V-S)} l(a) \leq \sum_{a \in A(S, V-S)} u(a) \tag{13}$$

Moreover for every cost function ω and integer bounds there exists an integer circulation minimizing the cost and it may be found in polynomial time. This remains true if some of $u(a)$ are infinite.

We now focus on the **Karger's algorithm**. It searches for a minimal cut in a connected undirected graph $G = (V, E)$. The algorithm is surprisingly simple. While $|V| > 2$ take a random edge and contract it (note that multiple edges may appear, we keep them). The two vertices in the end represent a cut.

Theorem 4.2. *For any minimal cut the probability of having it drawn is $\geq \frac{1}{n^2}$.*

Proof. Let c equal the size of the minimal cut. This implies all vertices have degree $\geq c$, so $|E| \geq \frac{cn}{2}$. We get minimal cut V_1, V_2 if no edges from $E(V_1, V_2)$ get erased. Probability of drawing one of them in k-step is at most $\frac{c}{c(n-k+1)/2}$, so

$$\mathbb{P}(success) \geq (1 - \tfrac{c}{cn/2})(1 - \tfrac{c}{c(n-1)/2}) \dots (1 - \tfrac{c}{c3/2}) = \tag{14}$$

$$= (1 - \tfrac{2}{n})(1 - \tfrac{2}{n-1}) \dots (1 - \tfrac{2}{3}) = \tfrac{n-2}{n} \tfrac{n-3}{n-1} \dots \tfrac{1}{3} = \tfrac{2}{n(n-1)} \geq \tfrac{1}{n^2} \tag{15}$$

□

Remark 4.3. The number of minimal cuts in a graph is $\leq n^2$.

Similar bound can be obtained for slightly bigger cuts.

Theorem 4.4 (Karger [5]). *The number of connected cuts of size $\leq \alpha c$ is $\leq n^{2\alpha}$ for any half-integer α. This remains true in weighted case.*

Please note that Karger's algorithm returns only connected cuts (i.e. cuts with both sides inducing connected subgraphs). If a cut (V_1, V_2) is minimal then of course V_i are connected, but otherwise we have to keep that in mind.

5 ATSP approximation

ASYMMETRIC TRAVELING SALESMAN PROBLEM
Input: Set V, cost function $w : V \times V \longrightarrow \mathbf{R}^+$ satisfying triangle inequality
Output: Hamiltonian cycle with the smallest cost

Theorem 5.1 (Asadpour et al. [1]). *There is an $O\left(\frac{\log n}{\log \log n}\right)$-approximation for ATSP*

Although the difference between $\log n$ and $\frac{\log n}{\log \log n}$ seems negligible, it took long time to show that approximation better than $O(\log n)$ is reachable.

We start with a LP relaxation.

$$\begin{aligned}
&\min && \sum_a w_a x_a && (16)\\
&\forall_v && \sum_{a \in in(v)} x_a = \sum_{a \in out(v)} x_a = 1 && (17)\\
&\forall_{\emptyset \neq S \subsetneq V} && \sum_{a \in out(S)} x_a \geq 1 && (18)\\
&\forall_a && x_a \geq 0 && (19)
\end{aligned}$$

Let x^* be an optimal solution of the above LP. Let us define $z(uv) = \frac{n-1}{n}(x^*(u,v) + x^*(v,u))$. Please note z belongs to the interior of the LP for spanning trees (7).

$$\begin{aligned}
&\forall_{\emptyset \neq S \subseteq V} && x(E(S)) \leq |S| - 1 && (20)\\
& && x(E) = |V| - 1 && (21)\\
&\forall_e && x_e \geq 0 && (22)
\end{aligned}$$

We also assign costs to undirected edges so, that the cost of uv is equal to $\min(w(u,v), w(v,u))$. We will use the same symbol w to denote this new cost function, as it is always clear from the context which cost is being used.

Definition 3. We call a spanning tree R (α, β)-thin in relation to z if its cost is not bigger than cost of z times β and

$$\forall_{\emptyset \neq S \subset V} |E(S, V-S) \cap R| \leq \alpha z(E(S, V-S)) \tag{23}$$

Theorem 5.2. *We can find an* $(O\left(\frac{\log n}{\log \log n}\right)$*, 2)-thin tree in relation to* z*.*

Proof. We draw tree R using maximum entropy sampling. $\mathbb{E}(w(R)) = w(z)$ therefore from Markov inequality we get $w(R) \leq 2w(z)$ with probability $\geq \frac{1}{2}$.

Consider a cut S with size s in relation to z. Then $\mathbb{E}(|E(S, V-S) \cap R|) = s$. Let us take $\alpha = \frac{10 \log n}{\log \log n}$. As MES gives us the negative correlation we can use the Chernoff bound (see Lecture 5).

$$\mathbb{P}(|E(S, V-S) \cap R| \geq \alpha s) \leq \left(\frac{e^{\alpha - 1}}{\alpha^\alpha}\right)^s \leq n^{-2.5s} \tag{24}$$

(we omit the technical proof of the last inequality).

We want to estimate the probability of having at least one cut too large (larger more than α times its weight in z) – we call this event B. Please note LP 16 guarantees that all (directed) cuts in x^* have sizes in $[1, \frac{n}{2}]$. Therefore size of the smallest cut in z must be $\geq 2(1 - \frac{1}{n})$. We can group all cuts according to their sizes $[2(1-\frac{1}{n}), 3(1-\frac{1}{n})), [3(1-\frac{1}{n}), 4(1-\frac{1}{n})) \ldots [(n-1)(1-\frac{1}{n}), n(1-\frac{1}{n})]$ and use Theorem 4.4.

$$\mathbb{P}(B) \leq \sum_{k=3}^{n} n^k n^{-2.5(k-1)(1-\frac{1}{n})} = O(\sum_{k=3}^{n} n^{-1.5}) = O(\frac{1}{\sqrt{n}}) \tag{25}$$

Therefore we sample desired tree with high probability. □

Let T_0 be a (undirected) tree from theorem 5.2. We transform T_0 into a directed tree T replacing each edge uv with (u,v) or (v,u) choosing the one with smaller cost.

Theorem 5.3. *We can find an integer circulation containing T with cost $\leq (2\alpha+2)OPT$.*

Proof. We want to use Theorem 4.1 with $l = \chi_T, u = \chi_T + 2\alpha x^*$ (α as defined in proof of Theorem 5.2). We need to ensure that

$$\forall_{S\subset V} \sum_{a\in A(S,V-S)} l(a) \leq \sum_{a\in A(V-S,S)} u(a) \tag{26}$$

Indeed, as T_0 was $(\alpha, 2)$-thin

$$\sum_{a\in A(S,V-S)} l(a) = \sum_{a\in A(S,V-S)} \chi_T(a) \leq \sum_{e\in E(S,V-S)} \alpha z(E(S,V-S)) \leq \sum_{a\in A(S,V-S)} 2\alpha x^*(a) \tag{27}$$

Since x^* is a circulation we have $\sum_{a\in A(S,V-S)} 2\alpha x^*(a) = \sum_{a\in A(V-S,S)} 2\alpha x^*(a)$ and so we get

$$\sum_{a\in A(S,V-S)} l(a) \leq \sum_{a\in A(V-S,S)} 2\alpha x^*(a) \leq \sum_{a\in A(V-S,S)} u(a) \tag{28}$$

The cost of T_0 is the same as that of T by definition, and so by Theorem 5.2 is at most $2OPT$. Circulation x^* is a relaxation of ATSP problem so its cost is $\leq OPT$ and so the cost of the circulation whose existence follows from Hoffman's Theorem is at most that of $\chi_T + 2\alpha x^*$, i.e. $\leq (2\alpha+2)OPT$. Therefore cheapest circulation with lower bound $l = \chi_T$ and no upper bound also has cost $\leq (2\alpha+2)OPT$. One can also ensure that it is integral by Hoffman's Theorem. □

Proof of Theorem 5.1. Let c be the circulation from Theorem 5.3. It represents a connected multigraph over V. Moreover it is Eulerian and let $v_1 \dots v_m$ be some corresponding Eulerian cycle. We can change it greedily into Hamiltonian cycle by shortcutting vertices which have already occurred. As the weight function w satisfies the triangle inequality, shortcutting will not increase the cost of the path. □

6 In the next lecture

Lecture 7 will concern metric embeddings. Here is the general idea. Image we try to design an approximation algorithm for a problem that is hard for general metric spaces (or we just do not know how to approximate it well), but easy for specific spaces like ℓ_1, ℓ_2, ℓ_∞, etc. One could try to embed the given metric space into one of these metrics with as little distortion as possible, solve the problem there, and pull back the solution.

In the next lecture we consider the problem of approximate (or low-distortion) embeddings of general metrics into tree metrics, i.e. shortest path metrics induced by (weighted) trees.

More specifically, for a given metric (V, d) we would like to find a mapping ϕ of V into a vertex set of some tree T (a *tree-embedding*), so that

$$d(x, y) \leq d_T(\phi(x), \phi(y)) \leq \alpha d(x, y)$$

and α should be relatively small, but may depend on n.

Unfortunately that goal is impossible to achieve even for a cycle. We will show in the next lecture that in this case we need to have $\alpha = \Omega(n)$. Therefore we relax our goal slightly. We construct a distribution on tree-embeddings, such that

$$d(x, y) \leq d_T(\phi(x), \phi(y))$$

and

$$\mathbb{E} d_T(\phi(x), \phi(y)) \leq \alpha d(x, y),$$

where ϕ is a random mapping into a tree T (also random).

One can easily see that by removing a random edge from a cycle, we obtain such a distribution with $\alpha = 2$. What is a bit surprising, and very useful, is that we can actually construct such a distribution for any metric, with $\alpha = O(\log n)$.

References

[1] A. Asadpour, M. X. Goemans, A. Madry, S. O. Gharan, and A. Saberi. An O(log n/ log log n)-approximation algorithm for the asymmetric traveling salesman problem. *SODA*, pages 379–389, 2010.

[2] A. Asadpour and A. Saberi. Maximum entropy selection: a randomized rounding method. *submitted*.

[3] C. Chekuri, J. Vondrák, and R. Zenklusen. Dependent randomized rounding via exchange properties of combinatorial structures. *FOCS*, pages 575–584, 2010.

[4] R. Gandhi, S. Khuller, S. Parthasarathy, and A. Srinivasan. Dependent rounding and its applications to approximation algorithms. *J. ACM*, 53(3):324–360, 2006.

[5] D. R. Karger. Global min-cuts in rnc, and other ramifications of a simple min-cut algorithm. *ACM-SIAM Symposium on Discrete algorithms*, pages 21–30, 1993.

[6] A. Schrijver. Combinatorial optimization. *Algorithms and Combinatorics, Springer*, 2003.