Euclidean TSP

Lecture 3, Jan 27 2015
Lecturer: Arnab Bhattacharyya

Wednesday, January 28, 2015
10:59 AM

Recall: TSP NP-hard. In fact, no approx possible for arbitrary distances. For (symmetric) metric TSP, we saw 3/2 - approx (Christofides '76). Also, for this problem, NP-hard to have a 220/219 - approx (Papadimitriou - Vempala '00)

Polynomial time approx scheme (PTAS):

For a minimization problem P, a PTAS for P is an algorithm A that, given input instance I and a parameter $\varepsilon > 0$, returns a solution of cost $\leq (1+\varepsilon) \cdot OPT_I$ in time $O(n^{f(\varepsilon)})$.

(No PTAS for metric TSP)

Euclidean TSP

Input: n pts in $\mathbb{R}^2$ with $d(x, y) = \|x-y\|_2$

Output: shortest tour visiting all n points.

Also, NP-hard (complete?)

But PTAS for this problem!! (Arora '98, Mitchell '99)

↳ and generalization to any const. dimension d (points in $\mathbb{R}^d$)

· No PTAS for $d = O(\log n)$ (Trevisan '97)

Best so far: PTAS in time $2^{poly(1/\varepsilon)} \cdot n$ (Bartal - Gottlieb '13)

We will see Arora's $n^{O(1/\varepsilon)}$ time PTAS.

How to show PTAS

- A "Structure Theorem" showing that there's a solution of cost $\leq (1+\varepsilon) \cdot OPT$ that has special local structure
- Find structured solution by divide / conquer or DP.

Arora's TSP PTAS

Three steps:

Perturbation (round instance so pts are on grid corners)

... space into squares)

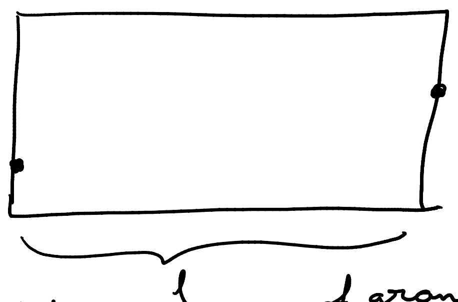
- Randomized dissection (partition of
- Portal-respecting tours (apply DP to find tours that enter/exit squares at specific positions)

(assume $1/\varepsilon$ integral for simplicity)

Perturbation

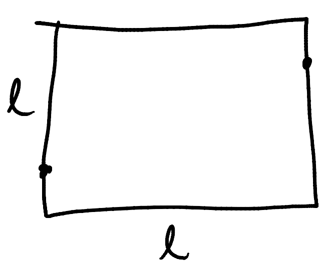
Want: - each pt to have integer coordinates in $[0, \frac{9n}{\varepsilon}]^2$
- smallest nonzero distance between pts to be ≥ 4.

- Translate pts so that they are in a bounding box with longer side length 1.



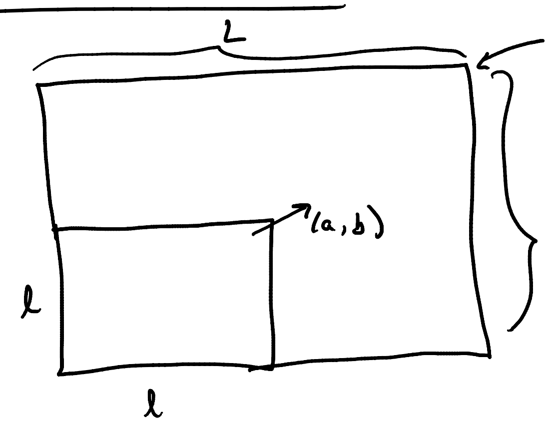
Note: OPT ≥ 2

- Place a grid inside this box of granularity $\frac{\varepsilon}{8n}$
- Move each pt to its nearest grid node
- Cost of TSP changes by $\leq n \cdot 2 \cdot \frac{\varepsilon}{8n} \leq \frac{\varepsilon}{4} \leq \frac{\varepsilon}{8} \cdot \text{OPT}$
- Now divide all coordinates by $\frac{\varepsilon}{32n}$.
- All coordinates become integral and min. distance is ≥ 4.
- Make bounding box square of side length $\frac{32n}{\varepsilon} = l$



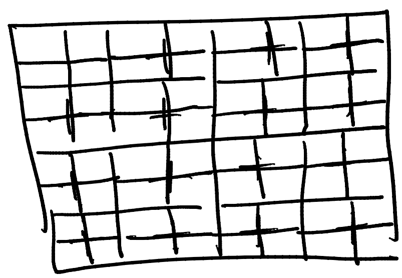
Randomized dissection

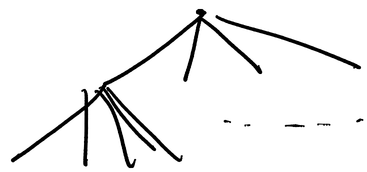
Shift:



$2l \le L \le 4l$, power of 2.

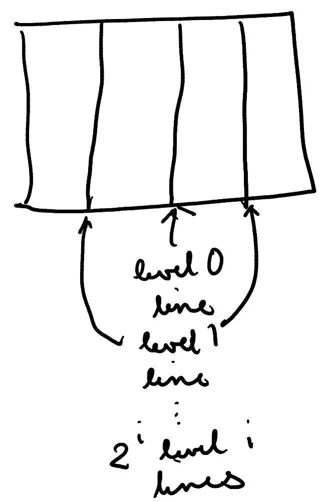
Dissection:

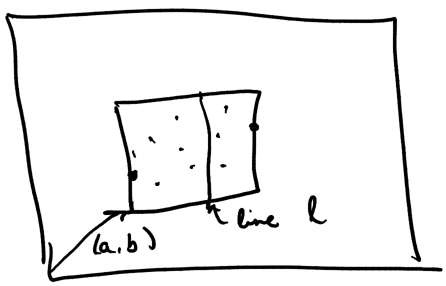




4-ary tree

Enclosing box: level 0 square
4 level 1 squares
16 level 2 squares
⋮
2^{2i} level i squares





<u>Obs</u>: $\Pr_{a,b \le L/2}\left[\text{line } l \text{ is at level } i\right] \le \frac{2^i}{L/2} = \frac{2^{i+1}}{L}$

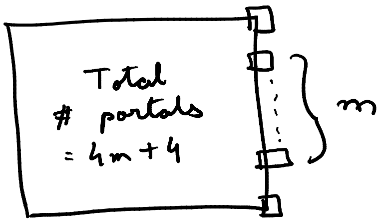
at most 2^i level i lines that could be reached by shifts of size $< L/2$.

Portal tours

- Each grid line has special points on it called "portals"
- A level i line has $2^{i+1} \cdot m$ equally spaced portals and also each corner of a level i square on the line is a portal.

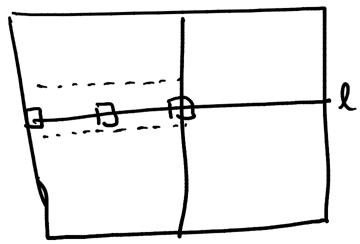
Portal tour is a tour

a portal.



Portal tour is a tour that always crosses grid lines at portals.

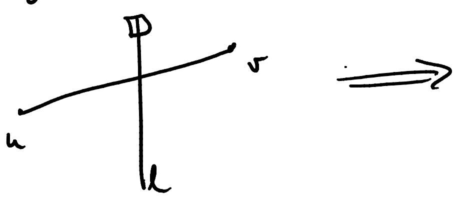
Obs: Without shifting, portal tour may be much longer than OPT.

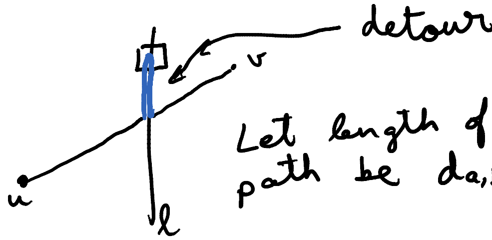


Optimal tour crosses l lots of times but a portal tour has to have many long detours

Thm: For a, b random integers in $[0, L/2)$, with prob $\geq \frac{1}{2}$, there is a portal tour of the (a, b)-dissection of cost $\leq \text{OPT} + \frac{8 \log L}{m} \cdot \text{OPT}$.

Pf: We bound $\mathbb{E}[\text{OPT}_{a,b,m} - \text{OPT}] \leq \frac{4 \log L}{m} \cdot \text{OPT}$. Thm follows from Markov.





Let length of new path be $d_{a,b,m}(u,v)$

$$\mathbb{E}[d_{a,b,m}(u,v) - d(u,v)]$$

$$\leq \sum_{l \text{ crossed}} \sum_{i=0}^{\log L - 1} \Pr[l \text{ is in level } i] \cdot \frac{L}{m 2^{i+1}}$$

$$= \sum_{l \text{ crossed}} \sum_{i=0}^{\log L - 1} \frac{2^{i+1}}{L} \cdot \frac{L}{m \cdot 2^i} = \frac{2 \log L}{m} \cdot \#\text{of lines crossed}$$

$$\#\text{of lines crossed} \leq |x_u - x_v| + |y_u - y_v| + 2$$

$$\leq \sqrt{2(|x_u - x_v|^2 + |y_u - y_v|^2)} + 2$$

Use that $|x_u - x_v|$, $|y_u - y_v| \geq 4$.

$$\leq 2 \cdot d(u,v)$$

So, $\mathbb{E}[d_{a,b,m}(u,v) - d(u,v)] \leq \frac{4 \log L}{m} \cdot d(u,v)$

Summing over all edges (u,v) of opt tour gives result. $\square$

<u>Remark</u>: You might worry that the detour crosses horizontal lines at non-portal locations. But note that intersection of l with any horizontal line is always a portal!

Set $m = (4 \log L)/\varepsilon = O((\log(n/\varepsilon))/\varepsilon)$

$OPT_{a,b,m} - OPT \leq \varepsilon \cdot OPT$.

How to find $OPT_{a,b,m}$?

<u>Use dynamic programming!</u>

Use one state for any square and any set of possible ways of entering and exiting this square.

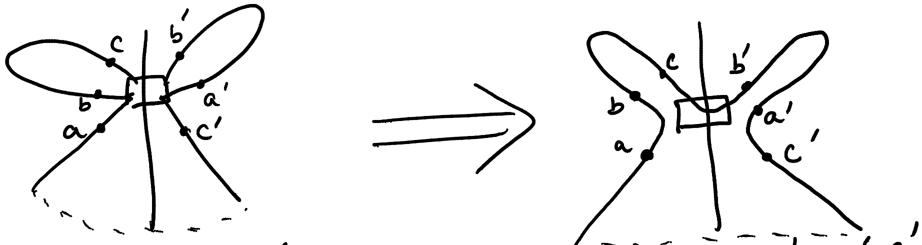
To bound # of states, need to restrict how many times each portal is visited.

<u>Def</u>: A portal tour is <u>2-light</u> if it goes through the portal at most twice.

<u>Lemma</u>: Without loss of generality, a portal tour is 2-light.

<u>Proof by picture</u>:

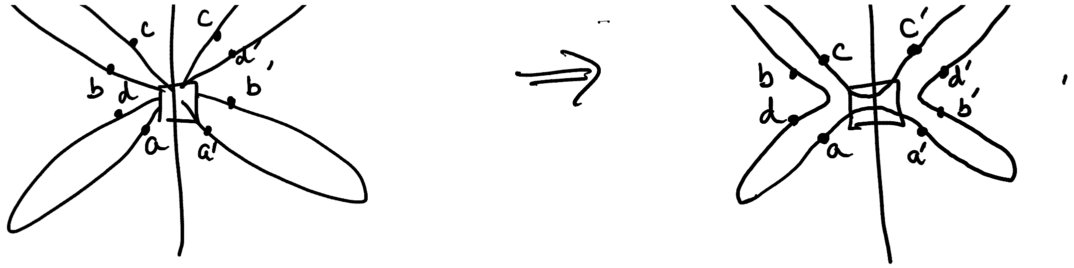
<u>Odd</u>



Visiting order: a, a', b', b, c, c' $\qquad$ a, b, c, b', a', c'

<u>Even</u>

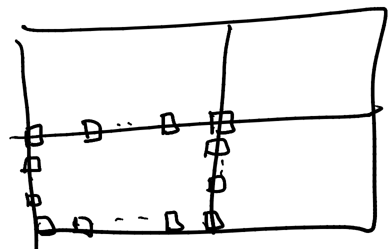
Even



Visiting order : a, a', b', b, c, c', d', d

a, a', b', d', c', c, b, d

Keep info about the portals used to enter each square and the order in which the tour uses these portals



For a square s and a set of pairs of portals $(s_1, t_1), (s_2, t_2), \ldots, (s_k, t_k)$, let $A[s, \{(s_1,t_1), \ldots, (s_k,t_k)\}]$ be cost of cheapest set of paths $p_1, \ldots, p_k$ that

(1) each p_i goes from s_i to t_i

(2) together $p_1, \ldots, p_k$ contain all points in s.

Remark: Our final answer will correspond to $A[s, \emptyset]$, where s is a unit square,

Remark: At a leaf node, There will be only ≤ 1 point p. The pt p will be contained in only one path p_i (a straight line segment from s_i to p and a straight line segment from p to t_i) and the other paths p_j will be straight line segments from s_j to t_j. We can search all $i \in [k]$ to solve this case.

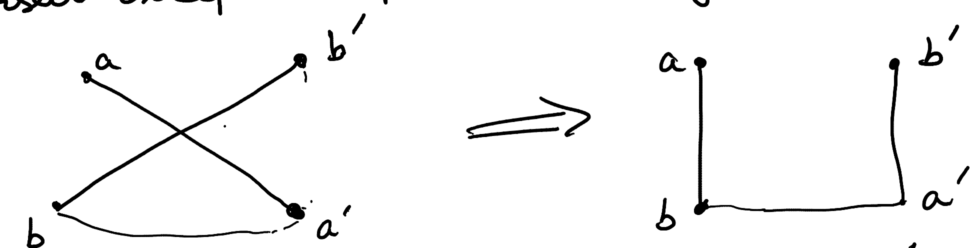
Remark: The DP itself is straightforward. $A[s, \cdots]$ can be computed from $A[s_1, \cdots]$, $A[s_2, \cdots]$, $A[s_3, \cdots]$, $A[s_4, \cdots]$ by trying all consistent combinations ... $s_1 \ldots s_4$ are

of entry and exit portals. $(s_1, s_2, \ldots, s_?$ …
the children of s in the quadtree)

Total # of states $\leq$ (# of squares) $\cdot \underbrace{3^{4m+4}}_{\substack{\text{Each portal} \\ \text{used 0, 1, or} \\ \text{2 times by} \\ \text{2-lightness}}} \cdot \underbrace{(8m+8)!}_{\substack{\text{\# of possible} \\ \text{pairings of} \\ \text{the} \leq 8m+8 \\ \text{portal visits}}}$

As noticed in class, for $m = \log\left(\frac{1}{\varepsilon} \log \frac{n}{\varepsilon}\right)$, this is only quasi-poly $(2^{\log^2 n})$, not poly (n).

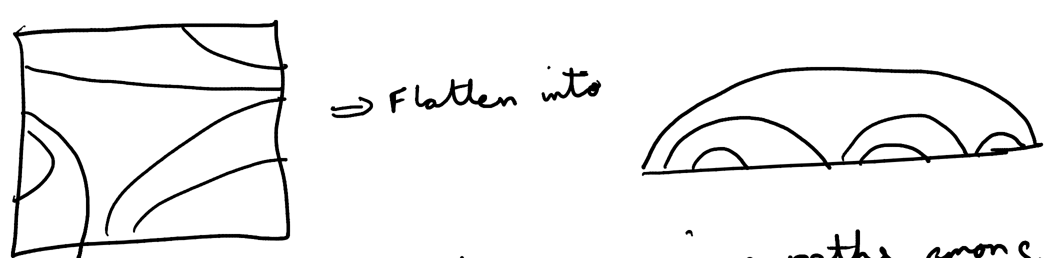
To get poly (n) bound, observe that we may assume $p_1, \ldots, p_k$ are vertex-disjoint inside s (meaning they don't intersect except at portals). Why? Again shortcutting:



Visiting order: a, a', b, b'

a, b, a', b'
cost not more by
Triangle inequality

So, now, need to bound number of ways there can be k non-crossing paths among $\leq 8m+8$ portal occurrences.



The number of such ways of non-crossing paths among $\leq 8m+8$ portal occurrences bounded by number of well-matched parentheses pairs when there are $\leq 4m+4$ of each opening & closing kind. Latter bounded by Catalan number $O(2^{8m})$.

So, total # of states $\leq$ (# of squares) $\cdot$ $O(3^{4m} \cdot 2^{om})$

$$= n^{O(1/\varepsilon)}$$

if $m = O\left(\frac{1}{\varepsilon} \lg \frac{n}{\varepsilon}\right)$.