COMBINATORICA
Akadémiai Kiadó – Springer-Verlag

COMBINATORICA 12 (3) (1992) 323–330

ON THE LONG EDGES IN THE SHORTEST TOUR THROUGH N RANDOM POINTS

WANSOO T. RHEE* and MICHEL TALAGRAND*

Received 15 June, 1989
Revised 19 April, 1990

Consider the shortest tour through n points $X_1,\dots,X_n$ independently uniformly distributed over $[0,1]^2$. Then we show that for some universal constant K, the number of edges of length at least $un^{-1/2}$ is at most $Kn\exp(-u^2/K)$ with overwhelming probability.

Introduction

The (in)famous Traveling Salesperson Problem (TSP) requires finding the shortest tour through n points $x_1,\dots,x_n$ of the plane, i.e.

$$\inf(\|x_{\sigma(n)} - x_{\sigma(1)}\| + \sum_{i=1}^{n-1} \|x_{\sigma(i+1)} - x_{\sigma(i)}\|) \tag{1}$$

where the inf is over all the permutations σ of $\{1,\dots,n\}$. In this paper we study the stochastic model where the points $X_1,\dots,X_n$ are independent uniformly distributed over $[0,1]^2$. (Our methods apply with routine modifications to much more general distributions). This model has received considerable attention.

The difficulty (and the fascination) if the TSP is the very non-constructive definition of the shortest tour. The structure of the shortest tour is likely to depend critically on the quantity we minimize, i.e. the sum of the lengths of the edges. It is very much unclear what is the relationship between the shortest tour, and a tour that would minimize, say, the sum of the squares of the lengths of the edges. There exist tours for which the sum of the squares of the edges is bounded by a universal constant. (See e.g. [5] for a simple proof using space-filling curves.) Thus it is natural to ask whether the sum of the squares of the edges of the shortest tour converges to some constant ([5]). While we do not answer this question, we will show that in a shortest tour the edges of length $\gg n^{-1/2}$ are exceptional. It follows in particular that the sum of the squares of the edges of the shortest tour is, with overwhelming probability, less than some universal constant.

AMS subject classification code (1991): 60 D 05
* This research is in part supported by an NSF grant.

Theorem 1. *There exists a constant K with the following property. Denote by $N(u)$ the number of edges that are $\gg un^{-1/2}$ in a shortest tour through $X_1,\ldots,X_n$. Then for all s*

$$P(N(u) \geq s) \leq \left(\frac{Kn\exp(-u^2/K)}{s}\right)^{s/K}.$$

It follows in particular that the longest edge is of order at most $(\log n/n)^{1/2}$.

With probability 1, any two different tours through $X_1,\ldots,X_n$ have different lengths, so there is a unique shortest tour. For simplicity, we will say "edge" instead of "edge in the unique shortest tour through $X_1,\ldots,X_n$".

For each point X_i, the length of the edges through X_i is at least the distance of X_i to its closest neighbor. It is a simple exercise to see that the number of points at distance $\geq un^{-1/2}$ of their closest neighbor is at least $n\exp(-u^2/K_1)$ with large probability where K_1 is a universal constant. Thus actually there are at least $n\exp(-u^2/K_1)$ edges of length $\geq un^{-1/2}$. On the other hand, Theorem 1 easily implies that with large probability there are at most $Kn\exp(-u^2/K)$ such edges. Thus at a superficial level we can argue that Theorem 1 is sharp. Nonetheless we must keep in mind that $K_1 \ll K$, and thus (for example) we know about the longest edge e, with large probability we have (say)

$$\frac{1}{10}\left(\frac{\log n}{n}\right)^{1/2} \leq e \leq 10^6\left(\frac{\log n}{n}\right)^{1/2}$$

while it does not seem unreasonable to hope that

$$(1-\varepsilon_n)c\left(\frac{\log n}{n}\right)^{1/2} \leq e \leq (1+\varepsilon_n)c\left(\frac{\log n}{n}\right)^{1/2}$$

for some constant c and $\varepsilon \to 0$.

The proof of Theorem 1 follows the natural approach: we show that the existence of long edges forces very special configurations. We then show that these are exceptional in the probabilistic sense. In Section 2 we prove a simple deterministic lemma, and we bring in the Poisson point process. In Section 3 we perform the main step, and in Section 4 we take care of the final computational details.

2. Preliminaries

We fix n, and we denote by Π a Poisson point on $\mathbb{R}^2$ of intensity n. For any Borel set $A \subset \mathbb{R}^2$, $\Pi(A)$ will consist of a set of N_A points uniformly distributed in A, where N_A itself is a Poisson random variable with mean $n|A|$, where $|A|$ is the area of A. We denote by $M(u)$ the number of edges $\geq un^{-1/2}$ in the shortest tour through the finite set $F = \Pi([0,1]^2)$.

Proposition 2. *It suffices to prove Theorem 1 with $M(u)$ instead of $N(u)$.*

Proof. We have

$$P\left(N_{[0,1]^2} = n\right) = \frac{n^n}{n!}e^{-n} \geq \frac{1}{C\sqrt{n}}$$

for some constant C. Thus

$$
\begin{aligned}
P(N(u) \geq s) &\leq C\sqrt{n}P(M(u) \geq s) \\
&\leq C\sqrt{n}\left(\frac{Kn}{s}\exp\left(-\frac{u^2}{K}\right)\right)^{s/K} =: f(s).
\end{aligned}
$$

Consider $L \geq K$, to be determined later. Let $g(s) = (\frac{Ln}{s}\exp(-u^2/L))^{s/L}$. It suffices to check that for $1 \leq s \leq n$, if $g(s) \leq 1$, then $f(s) \leq g(s)$. We turn to the rather tedious proof of that fact, that serves the purpose of shedding some light on the nature of a bound of type $g(s)$.

Supposing $g(s) \leq 1$, we have $a =: (Ln/s)\exp(-u^2/L) \leq 1$, and $\exp(-u^2/K) = (\frac{as}{Ln})^{L/K}$. Thus

$$
f(s) = C\sqrt{n}\left(\frac{Kn}{s}\left(\frac{s}{Ln}\right)^{L/K} a^{L/K}\right)^{s/K} = C\sqrt{n}\left(\frac{Kn}{s}\left(\frac{s}{Ln}\right)^{L/K}\right)^{s/K} a^{sL/K^2}.
$$

Since $1 \geq g(s) = a^{s/L} \geq a^{sL/K^2}$, it suffices to show that

$$
\begin{aligned}
(1) \qquad C\sqrt{n}\left(\frac{Kn}{s}\left(\frac{s}{Ln}\right)^{L/K}\right)^{s/K} &= C\sqrt{n}\left(\frac{K}{L^{L/K}}\left(\frac{s}{n}\right)^{L/K-1}\right)^{s/K} \\
&= C\sqrt{n}(As)^{s(L/K^2-1/K)} \leq 1
\end{aligned}
$$

where $A = (K^{K/(L-K)}/L^{L/(L-K)})(1/n)$. The function $s \to (As)^s$ decreases for $As \leq 1/e$, and in particular for $s \leq n$, if $nA \leq 1/e$. This is the case if L is large enough ($L \geq 2K$ suffices). Thus, it suffices to check (1) for $s = 1$, i.e. that

$$
C\sqrt{n}\left(\frac{Kn}{(Ln)^{L/K}}\right)^{1/K} = n^{1/2+1/K-L/K^2}\frac{CK^{1/K}}{L^{L/K^2}} \leq 1
$$

This holds as soon as $C\left(\frac{K^{1/K}}{L^{L/K^2}}\right) \leq 1$, $\frac{1}{2} + \frac{1}{K} - \frac{L}{K^2} \leq 0$. ∎

We now fix an arbitrary orientation of the shortest tour through F. Given $x \in F = \Pi([0,1]^2)$, we denote by x' the successor of x in this shortest tour through F.

Lemma 3. *For x, $y \in F$, we have*

$$
\|x - x'\| + \|y - y'\| \leq \|x - y\| + \|y' - x'\|.
$$

Proof. Otherwise we can build a shorter tour by deleting the edges xx', yy' and adding the edges xy, $x'y'$. ∎

We denote by $B(a,r)$ the disc of center a and radius r.

Lemma 4. *Suppose that $x_1, \ldots, x_p \in F \cap B(a,r)$, and that $\|x_i' - a\| > 4r$ for $i \leq p$. Then $p \leq 5$.*

Proof. Fix i, $j \leq p$. Then, by lemma 3, we have

$$
\|x_i - x_i'\| + \|x_j - x_j'\| \leq \|x_i - x_j\| + \|x_i' - x_j'\|
$$

and thus

$$\|x_i' - a\| + \|x_j' - a\| - 4r \le \|x_i' - x_j'\|.$$

Since $\|x_i' - a\| > 4r$, $\|x_j' - a\| > 4r$, we have

$$\|x_i' - x_j'\| > \|x_i' - a\|, \quad \|x_j' - a\|$$

and so the angle between the lines ax_i', ax_j' is the largest in the triangle $ax_i'x_j'$. Thus this angle is $> \pi/3$. This implies the result. ∎

Corollary 5. *If* $\operatorname{card}(F \cap B(a,r)) \ge 11$, *there exists* $x \in F \cap B(a,r)$ *such that the two points adjacent to* x *in the shortest tour are in* $B(a,4r)$.

The following is rather obvious.

Proposition 6. *We can find* η, $\xi > 0$ *with the following property. If* $W(a)$ *denotes the event*

$$\operatorname{Card}(\Pi)\left(B(a,\eta n^{-1/2})\right) \ge 11);$$

$$\forall x,\ y,\ z \in \Pi(B(a, 4\eta n^{-1/2})),\ \|x - y\| + \|y - z\| \ge \|x - z\| + \xi n^{-1/2},$$

then $P(W(a)) \ge 1/2$.

Proof. Observe that $P(W(a))$ is independent of a and n; then take η such that $\pi\eta^2 \gg 11$ and then ξ small enough. ∎

The point of Proposition 6 is as follows.

Corollary 7. *If* $W(a)$ *occurs, and* $B(a,\eta n^{-1/2}) \subset [0,1]^2$, *there is a point* $x \in F \cap B(a,\eta n^{-1/2})$ *such that bypassing this point in the shortest tour through* F *shortens the tour by at least* $\xi n^{-1/2}$.

Proof. By bypassing x, we mean that if y, z are the neighbors of x in the shortest tour through π, we delete the edges xy, xz and replace them by the edge yz. If $\operatorname{card}\Pi \cap B(a,\eta n^{-1/2}) \ge 11$, Corollary 5 guarantees that we can find x in that set for which $\|y-a\|$, $\|z-a\| \le 4\eta n^{-1/2}$, and $W(a)$ guarantees that $\|x-y\|+\|y-z\|-\|y-z\| \ge \xi n^{-1/2}$. ∎

3. Main construction

We set $m = \lfloor\sqrt{n}\rfloor$. We divide $[0,1]^2$ into m^2 congruent squares C_i. Consider two such squares C_1, C_2, of centers z_1, z_2, and $k \ge 1$ such that $\|z_1 - z_2\| \ge 2^{k-2} n^{-1/2}$. (Thus $2^k \le 4\sqrt{2} n^{1/2}$).

Proposition 8. *There exists* $\alpha > 0$ *and* k_0, *such that if* $k \ge k_0$, *then there exists an event* $Z_k(C_1,C_2)$, *depending only on* $\Pi(B(z_1, \|z_2 - z_1\|))$ *with the following properties*

a) $P(Z_k(C_1,C_2)) \le \exp(-\alpha 2^{2k})$
b) *If the shortest tour through* F *contains an edge* xy, *with* $x \in C_1$, $y \in C_2$, *then* $Z_k(C_1,C_2)$ *occurs.*

Proof. Consider two parameters p, q, independent of n, to be determined later. We denote by L_0 the line $z_1 z_2$. We pick one of the sides of L_0 where there will be enough room to perform the following construction inside $[0,1]^2$. (That this is indeed possible should be obvious in a moment.) For $1 \le i \le 2^{k-p}$, we consider the lines L_i parallel to L_0, at distance $8i\eta n^{-1/2}$ of L_0, and on the side we have chosen. Consider now the lines H_1, H_2, perpendicular to L_0, that intersect the interval z_1z_2 respectively at $1/3$ and $2/3$ of its length. Denote by $a_{i,1}$ (resp. $a_{i,2^{k-q}}$) the intersection of L_i and H_1 (resp. L_i and H_2). We divide the interval $a_{i,1}a_{i,2^{k-q}}$ into $2^{k-q}-1$ equal subintervals to obtain points $(a_{i,\ell})_{\ell=1}^{2^{k-q}}$. We observe that, provided q has been taken large enough, the discs $B(a_{i,\ell}, 4\eta n^{-1/2})$ are all disjoint, and thus the events $W(a_{i,\ell})$ are independent. Since each has a probability $\ge 1/2$ given $i \le 2^{k-p}$, the event

$$H_i: \text{ "less than } 2^{k-q-2} \text{ events } W(a_{i,\ell}),\ \ell = 1, \dots, 2^{k-q} \text{ occur"}$$

has probability $\le \exp(-2^{k-q-3})$, by standard estimates on the tails of the binomial law. Consider the event $Z_k(C_1, C_2) = \bigcap_{i \le 2^{k-p}} H_i$. Then

$$P(Z_k(C_1, C_2)) \le \exp(-2^{2k-p-q-3})$$

so that a) holds. Also, provided that p has been taken large enough, all the discs $B(a_{i,\ell}, 4\eta n^{-1/2})$ are contained in $B(z_1, \|z_2 - z_1\|)$. Thus $Z_k(C_1, C_2)$ depends only on $\Pi(B(z_1, \|z_2 - z_1\|))$.

It remains to show the following. Suppose that for $1 \le j \le 2^{k-p}$, we can find $1 \le \ell(1) < \ell(j) < \ldots < \ell(2^{k-q-2}) \le 2^{k-q}$ such that all the events $W(a_{i,\ell(j)})$ occur for $1 \le j \le 2^{k-q-2}$. Then the shortest tour through F cannot contain an edge xy for $x \in C_1$, $y \in C_2$. Suppose for contradiction, that this were the case. Corollary 7 shows that each set $W(a_{i,\ell(j)})$ contains a point x_j such that bypassing that point in the shortest tour shortens this tour by at least $\xi n^{-1/2}$. Moreover, the proof of Corollary 7 shows that no two points x_j are adjacent in the shortest tour. Thus bypassing them all results in a saving $\ge 2^{k-q-2}\xi n^{-1/2}$. We are going to show that if p and q are suitable, when we replace the edge xy by the edges $xx_1, x_1x_2, \ldots, x_{2^{k-q-2}}y$, we increase the length by $< 2^{k-q-2}\xi n^{-1/2}$. We have thus constructed a shorter tour, a contradiction.

Denote by y_j the orthogonal projection of x_j on the line L_0. Using the formula $(a^2+b^2)^{1/2} \le a + b^2/(2a)$, we see that

$$\|x_{j+1} - x_j\| \le \|y_{j+1} - y_j\| + \frac{(2\eta n^{-1/2})^2}{2\|y_{j+1} - y_j\|}.$$

Now

$$\begin{aligned} \|y_{j+1} - y_j\| &\ge \|a_{0,\ell(j+1)} - a_{0,\ell(j)}\| - 2\eta n^{-1/2} \\ &= \frac{\|z_1 - z_2\|}{3 \cdot 2^{k-q}} - 2\eta n^{-1/2}. \end{aligned}$$

Since $\|z_1 - z_2\| \ge 2^{k-2} n^{-1/2}$ we see that if q has been taken large enough for some universal constant C_1 we have

$$\|x_{j+1} - x_j\| \le \|y_{j+1} - y_j\| + C_1 2^{-q} n^{-1/2}$$

and thus

$$(1)\qquad \sum_{j=1}^{2^{k-q}-1} \|x_{j+1}-x_j\| \le C_1 2^{k-2q} n^{-1/2} + \sum_{j=1}^{2^{k-q}-1} \|y_{j+1}-y_j\|.$$

(The reader will observe the essential fact that the q in the exponent of 2 has a factor 2.) We have

$$\|x-x_1\| \le 1/m + \|z_1-x_1\|$$

and

$$\|z_1-x_1\| \le \|z_1-y_1\| + \frac{\|x_1-y_1\|^2}{2\|z_1-y_1\|}.$$

Now

$$\|x_1-y_1\| \le (8i+1)\eta n^{-1/2} \le 2^{k-p+4}\eta n^{1/2}$$

and

$$\|z_1-y_1\| \ge \|z_1-a_{0,1}\| - \eta n^{-1/2} \ge \frac{\|z_1-z_2\|}{3} - \eta n^{-1/2}.$$

Thus, since $\|z_1-z_2\| \ge 2^{k-2}n^{-1/2}$ for $k \ge k_0$, where k_0 is universal, we see that for a universal constant C_2 we have

$$\|x-x_1\| \le \|z_1-y_1\| + C_2 2^{k-2p} n^{-1/2}.$$

A similar inequality holds for $\|x_{2^{k-q-2}}-y\|$.

Since

$$\|z_1-y_1\| + \left(\sum_{i=1}^{2^{k-q-2}-1} \|y_{i+1}-y_i\|\right) + \|y_{2^{k-q-2}}-z_2\| = \|z_1-z_2\| \le \|x-y\| + 2/m,$$

we see that the sum of lengths of the edges $xx_1, x_1x_2, \ldots, x_{2^{k-q-2}}y$ is less than

$$\|x-y\| + 2/m + C_1 2^{k-2q} n^{-1/2} + 2C_2 2^{k-2p} n^{-1/2}.$$

If we take q such that $C_1 2^{-q} < 2^{-4}\xi$ and then p such that $2C_2 2^{-2p} < 2^{-q-4}\xi$, we see that this is less than

$$\|x-y\| + 2^{k-q-3}\xi n^{-1/2} + 2/m.$$

For $k \ge k_0$, (k_0 universal) this is less than $\|x-y\| + 2^{k-q-2}\xi n^{-1/2}$. This completes the proof. ∎

4. Details

For the simplicity of computations, we denote by K a universal constant, that may vary at each occurrence.

We denote by $N(k)$ the number of edges in a shortest tour through F that are of length $\ge 2^{k-1}n^{-1/2}$ and $\le 2^k n^{-1/2}$. Consider a paving of the plane with squares of side $3{\cdot}2^k/m$. Consider such a square G, and the square G' of side $2^k/m$ located in the center of G. Consider the event $H(G)$: "There is an edge of length $\geq 2^{k-1}n^{-1/2}$ and $\leq 2^k n^{-1/2}$ that has an endpoint in G'".

We assume that the coordinates of the vertices of G are of the type p/m $(p\in\mathbb{Z})$, so that $G\cap[0,1]^2$ and $G'\cap[0,1]$ are union of small squares C_i. A segment of length $\leq 2^k n^{-1/2} \leq 2^k/m$ having an endpoint in G' has the other endpoint in G. Thus we have, with the notations of Proposition 8, that $H(G)\subset H'(G):=\cup Z(C_1,C_2)$ where the union is over all choices $C_1\subset G\cap[0,1]^2$, $C_2\subset G'\cap[0,1]^2$. It follows that

$$P(H'(G)) \leq 9\cdot 2^{2k}\exp(-\alpha 2^{2k}) := p$$

and that $P(H(G'))=0$ if $G'\cap[0,1]^2=\emptyset$. The number q of squares G such that $G'\cap[0,1]^2\neq\emptyset$ is easily seen to be at most

$$\left(\frac{m}{3\cdot 2^k}+1\right)^2 \leq \frac{Km^2}{2^{2k}}$$

since $2^k \leq 4\sqrt{2}n^{1/2}$. Since $H'(G)$ depends only on $\Pi(G)$, the events $H'(G)$ are independent. If M is the number of events $H(G)$ that occur, we have

$$P(M\geq s) \leq \binom{q}{s}p^s \leq \left(\frac{eq}{s}\right)^s p^s = \left(\frac{eqp}{s}\right)^2$$

so that
$$P(M\geq s)\leq\left(\frac{Km^2}{s}\exp(-\alpha 2^{2k})\right)^s.$$

The square G' can be covered by 2^{10} discs of radius $<2^{k-1}n^{-1/2}$. Corollary 5 shows that there can be at most $10\cdot 2^{10}$ edges of length $\geq 2^{k-1}n^{-1/2}$ in a shortest tour through F that meet G', and there is no such edge unless $H(G)$ occurs. Thus, if we denote by $N'(k)$ the number of edges in a shortest tour through F that are of length $\geq 2^{k-1}n^{-1/2}$ and $\leq 2^k n^{-1/2}$, and for which one endpoint is in a square G', we have

$$P\left(N'(k)\geq 5\cdot 2^{11}s\right)\leq\left(\frac{Km^2}{s}\exp(-\alpha 2^{2k})\right)^s.$$

We now note that this argument applies also when we shift the paving by a vector of components $(\varepsilon_1 2^k/m, \varepsilon_2 2^k/m)$ for ε_1, $\varepsilon_2=0,1,-1$. We thus get, since we can assume $s\geq 1$:

$$P\left(N(k)\geq 9\cdot 5\cdot 2^9 s\right)\leq 9\left(\frac{Km^2}{s}\exp(-\alpha 2^{2k})\right)^s \leq \left(\frac{4Km^2}{s}\exp(-\alpha 2^{2k})\right)^s$$

and hence
$$P(N(k)\geq s)\leq\left(\frac{Kn}{s}\exp\left(-\frac{2^{2k}}{K}\right)\right)^{s/K}.$$

We now prove Theorem 1. Given n, denote by k the largest integer such that $2^k \leq u$. Then we have

$$(2)\quad P(M(u)\geq s)\leq\sum_{\ell\geq 0}P(N(k+\ell)\geq 2^{-\ell-1}s)\leq\sum_{\ell\geq 0}\left(\frac{Kn}{s2^{-\ell}}\right)^{s2^{-\ell}/K}\exp\left(-\frac{s2^{2k+\ell}}{K^2}\right).$$

The function $\left(\frac{A}{t}\right)^2$ increases for $t \le \frac{A}{e}$. Thus, for $s \le \frac{Kn}{e}$, we have $\left(\frac{Kn}{s2^{-\ell}}\right)^{s2^{-\ell}/K} \le \left(\frac{Kn}{s}\right)^{s/K}$. Also, since we can suppose $s \ge 1$, we have $s2^{2k+\ell} \ge s2^{2k} + 2^\ell$ so that the left hand side of (2) is bounded by

$$\left(\frac{Kn}{s}\exp\left(-\frac{2^{2k}}{K}\right)\right)^{s/K} \sum_{\ell\ge 0} \exp\left(-\frac{2^\ell}{K^2}\right) \le K' \left(\frac{Kn}{s}\exp\left(-\frac{2^{2k}}{K}\right)\right)^{s/K}$$
$$\le \left(\frac{K'^K Kn}{s}\exp\left(-\frac{2^{2k}}{K}\right)\right)^{s/K}.$$

This completes the proof. ∎

References

[1] J. J. BARTHOLDI, and L. K. PLATZMAN: Heuristic based on space filling curves for combinatorial problems in Euclidean space, *Management Science* **34** (1988), 281–305.

[2] J. BEARDWOOD, J. H. HALTON, and J. V. HAMMERSLEY: The shortest path through many points, *Proc. Cambridge Philosophial Society* **55** 299–327.

[3] W. RHEE, and M. TALAGRAND: A sharp deviation inequality for the stochastic traveling salesman problem, *Annals of Probability* **17** (1989), 1–8.

[4] J. M. STEELE: Complete convergence of short paths and Karp's algorithm for the TSP, *Math. Oper. Res.* **6** (1981), 374–378.

[5] J. M. STEELE: Seedlings in the Theory of Shortest Paths, *Disorder in Physical Systems: A Volume in Honor of J. M. Hammerslet* (G. Grimmett and D. Welsh eds.), Cambridge University Press, 277–306, London, 1990.

WanSoo T. Rhee

Faculty of Management Science
1775 College Road
The Ohio State University
Columbus, OH 43210
U.S.A.
rhee.1@osu.edu

Michel Talagrand

Equipe d'Analyse-Tour 46
U.R.A. au C.N.R.S. n° 754
4 Place Jussieu
75230 Paris Cedex 05 France
mit@frunip62.bitnet
and
Department of Mathematics
The Ohio State University
231 West 18th Avenue
Columbus, Ohio 43210-1174
U.S.A.
talagrand.1@osu.edu