Adv. Appl. Prob. **47,** 27–36 (2015)
Printed in Northern Ireland

NEW BOUNDS FOR THE TRAVELING SALESMAN CONSTANT

STEFAN STEINERBERGER,* *Universität Bonn*

Abstract

Let $X_1, X_2, \dots, X_n$ be independent and uniformly distributed random variables in the unit square $[0, 1]^2$, and let $L(X_1, \dots, X_n)$ be the length of the shortest traveling salesman path through these points. In 1959, Beardwood, Halton and Hammersley proved the existence of a universal constant β such that $\lim_{n\to\infty} n^{-1/2} L(X_1, \dots, X_n) = \beta$ almost surely. The best bounds for β are still those originally established by Beardwood, Halton and Hammersley, namely $0.625 \le \beta \le 0.922$. We slightly improve both upper and lower bounds.

Keywords: Traveling salesman constant; Beardwood–Halton–Hammersley theorem

2010 Mathematics Subject Classification: Primary 60D05; 60F17

1. Introduction and statement of results

For given points $x_1, x_2, \dots, x_n \subset [0, 1]^2$, let $L(x_1, \dots, x_n)$ denote the length of the shortest traveling salesman path through all these points. It was realized early (see, e.g. the 1940 work by Fejes [7], the 1951 work by Verblunsky [19], and the 1955 work by Few [8]) that there are uniform estimates

$$L(x_1, \dots, x_n) \le c_1\sqrt{n} + c_2$$

for some constants c_1 and c_2. If the points are chosen at random, we would expect an averaging effect from the self-similarity of the problem on multiple scales: finding the optimal path is in some sense 'equivalent' to finding the optimal path through the points in many small subsets of the unit square and then patching these together. That this is indeed the case constitutes one of the first limit theorems in combinatorial optimization [4].

Theorem 1. (Beardwood *et al.* [4].) *Let* $X_1, X_2, \dots, X_n, \dots$ *be independent and identically distributed (i.i.d.) uniformly distributed random variables in* $[0, 1]^2$. *There exists a universal constant* β *such that*

$$\lim_{n\to\infty} \frac{L(X_1, \dots, X_n)}{\sqrt{n}} = \beta$$

with probability 1.

The statement is by now classic and very well known (see, e.g. the textbooks of Applegate *et al.* [1], Finch [9], Gutin and Punnen [10], Steele [16], Venkatesh [18], Yukich [20], or even a popular-science book [6]). It is relatively easy to deduce that if the points X_i are randomly distributed with respect to some probability on measure on $[0, 1]^2$ and f denotes the absolutely

Received 1 April 2014; revision received 13 July 2014.
* Current address: Department of Mathematics, Yale University, 10 Hillhouse Avenue, New Haven, CT 06520, USA. Email address: stefan.steinerberger@yale.edu

continuous part of the measure, then

$$\lim_{n\to\infty} \frac{L(X_1, \dots, X_n)}{\sqrt{n}} = \beta \int_{[0,1]^2} f(x)^{1/2}\, dx.$$

The Beardwood–Halton–Hammersley (BHH) limit law holds for various other problems (e.g. the minimal spanning tree and Steiner trees) with a constant depending on the functional: a unified approach to the theory is given by Steele's limit theorem [15] (an extension was later given by Redmond and Yukich [14], who also provided bounds on the convergence rate). Interestingly, and despite considerable effort, the constant is not known in any of the aforementioned cases. In the case of the traveling salesman, Beardwood, Halton and Hammersley themselves proved that

$$0.625 = \tfrac{5}{8} \le \beta \le \beta_{\mathrm{BHH}} \sim 0.921\,16\dots,$$

where

$$\beta_{\mathrm{BHH}} = 2\int_0^\infty \int_0^{\sqrt{3}} \sqrt{z_1^2 + z_2^2}\, \exp(-\sqrt{3} z_1)\left(1 - \frac{z_2}{\sqrt{3}}\right) dz_2\, dz_1.$$

It should be noted that Beardwood, Halton and Hammersley actually claimed to prove the better result $\beta \le 0.920\,37\dots$ (a statement reiterated in many different books and papers); however, their computation relies on numerical integration and we believe this to be the origin of the error: for the convenience of the reader, we have quickly surveyed their argument (and the integral to be evaluated) below. Despite the relative fame of the BHH theorem, there has been no improvement in the constant over the years; a series of papers [11], [13], [17] carrying out numerical estimates with large data sets suggest that $\beta \sim 0.712$. The purpose of this paper is to draw some attention to the problem, describe the existing original arguments, and to improve them.

Theorem 2. *We have*

$$\tfrac{5}{8} + \tfrac{19}{5184} \le \beta \le \beta_{\mathrm{BHH}} - \varepsilon_0$$

for some explicit

$$\varepsilon_0 > \tfrac{9}{16} 10^{-6}.$$

We have an explicit representation of ε_0 as an integral in $\mathbb{R}^7$: a concentration of measure effect turns Monte Carlo estimates into a highly stable method and suggests that actually

$$\varepsilon_0 \sim 0.0148\dots;$$

however, we consider the underlying idea to be of greater interest than the actual numerical improvement—in addition, certain natural generalizations of our method should be able to give at least $\beta \le 0.891$ if we assume that certain integrals in high dimensions can be evaluated (details are given below). While additional improvements of the upper bound may lead to integrals whose evaluations become nontrivial, the approach is conceptually clear: further improving the lower bound, however, seems more challenging and in need of new ideas. As of this moment, we know of no methodical approach by which this could be accomplished.

2. Proof of the upper bound

2.1. Reduction to Poisson processes

The core of the proof is in the stochastic treatment of n random points in $[0, 1]^2$ locally on the scale $n^{-1/2}$. At this scale the law of small numbers (see, e.g. [3]) implies that the process behaves essentially like a Poisson process of intensity n. This property was exploited by Beardwood, Halton and Hammersley; using their result, we can replace the n random points with a Poisson process of intensity n, which simplifies further computations (this argument was pointed out to me by J. Michael Steele).

Lemma 1. *Let $\mathcal{P}_n$ denote a Poisson process with intensity n on $[0, 1]^2$. Then*

$$\lim_{n\to\infty} \frac{\mathbb{E}L(\mathcal{P}_n)}{\sqrt{n}} = \beta.$$

The idea is rather simple: the number of points in a Poisson process (i.e. the Poisson distribution) has mean n and variance n. This means that we usually expect $|\#\mathcal{P}_n - n| \sim \sqrt{n}$, which is rather small compared to n. The expected length of a traveling salesman path lies somewhere between $\sim\beta\sqrt{n - \sqrt{n}}$ and $\sim\beta\sqrt{n + \sqrt{n}}$, the difference of which is approximately 1 and, thus, of smaller order—if we now assume that the BHH result holds, this implies convergence for all cases concentrated here. The remaining case, $|\#\mathcal{P}_n - n| \gtrsim \sqrt{n}$, has exponentially decaying probability: it suffices to use the uniform bound of Few [8] to bound the arising error. We leave the details to the interested reader. The inverse statement (i.e. that the result for the Poisson process implies the result for n random points) is actually due to Beardwood, Halton and Hammersley.

2.2. The original argument

In this section we describe the original argument due to Beardwood *et al.* [4]. A very similar argument was used some years earlier by Few [8]. Let X be a Poisson process with intensity n in the unit square $[0, 1]^2$. We look at the set

$$X^* = \left\{x \in X : \pi_2(x) \le \frac{\sqrt{3}}{\sqrt{n}}\right\},$$

where π_2 is the projection onto the second component. Instead of asking for a traveling salesman tour through all n points, we merely ask for one through this particular strip. The entire unit square is cut into strips and within each strip a local path gets constructed: in the end they all get connected to yield a traveling salesman path through the entire set. The simplest solution locally within a strip is to order the points in X^* with respect to the first coordinate, i.e. order them in such a way that

$$\pi_1(x_1) < \pi_1(x_2) < \pi_1(x_3) < \cdots,$$

and then simply connect the points in that order. This is illustrated in Figure 1.

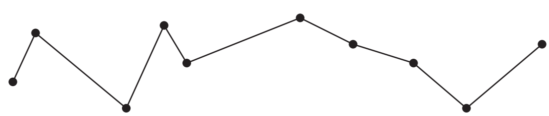

FIGURE 1: A strip containing some points.

https://doi.org/10.1239/aap/1427814579 Published online by Cambridge University Press

Theorem 3. ([4].) *Let X be a Poisson process with intensity n in $[0, 1]^2$, and let F be the length of the path constructed in the way described above. Then*

$$\lim_{n\to\infty} \frac{\mathbb{E}F(X)}{\sqrt{n}} = 0.921\,16\ldots.$$

Sketch of the proof. We restrict the Poisson process with intensity n to the strip $\pi_2(x) \leq \sqrt{3}/\sqrt{n}$. Then the real random variables

$$\left\{\pi_1(x) \colon \pi_2(x) \leq \frac{\sqrt{3}}{\sqrt{n}}\right\}$$

are distributed following a Poisson process with intensity $\sqrt{3n}$ on $[0, 1]$. Ordering the points with respect to the increasing first coordinate will give x-coordinates whose consecutive differences are exponentially distributed, i.e.

$$\mathbb{P}(|\pi_1(x_{i+1}) - \pi_1(x_i)| = z) \sim \sqrt{3n}\exp(-\sqrt{3n}z) \quad \text{for large } n,$$

while the y-coordinates are i.i.d. distributed following the uniform distribution on $[0, \sqrt{3}/\sqrt{n}]$. Therefore, the expected distance in joining one point to the next is given by

$$\mathbb{E}\|x_i - x_{i+1}\| = 2n\int_0^\infty \int_0^{\sqrt{3/n}} \sqrt{z_1^2 + z_2^2}\exp(-\sqrt{3n}z_1)\left(1 - \frac{\sqrt{n}z_2}{\sqrt{3}}\right) \mathrm{d}z_2\,\mathrm{d}z_1.$$

Substitution allows us to rewrite the integral as

$$\mathbb{E}\|x_i - x_{i+1}\| = \left(2\int_0^\infty \int_0^{\sqrt{3}} \sqrt{z_1^2 + z_2^2}\exp(-\sqrt{3}z_1)\left(1 - \frac{z_2}{\sqrt{3}}\right) \mathrm{d}z_2\,\mathrm{d}z_1\right)\frac{1}{\sqrt{n}}.$$

Since we are actually joining all n points, we have to jump from one strip to another approximately $\sqrt{n/3}$ times and each time the jump is of order approximately $1/\sqrt{n}$; this implies that the contribution arising from these jumps is of order $\mathcal{O}(1)$ and the total expected length is simply given by n times the expected length of a single jump, which gives

$$\left(2\int_0^\infty \int_0^{\sqrt{3}} \sqrt{z_1^2 + z_2^2}\exp(-\sqrt{3}z_1)\left(1 - \frac{z_2}{\sqrt{3}}\right) \mathrm{d}z_2\,\mathrm{d}z_1\right)\sqrt{n} \sim 0.921\,16\sqrt{n}.$$

The underlying 'layer' method is easily extended to other settings: an example is given in [5].

2.3. Changing variables

The argument contains all the necessary ingredients for our improved local construction: following the steps outlined above, we will study Poisson processes with intensity n in the strip

$$\left\{(x, y) \in \mathbb{R}^2 \colon (0 \leq x \leq 1) \wedge \left(0 \leq y \leq \frac{\sqrt{3}}{\sqrt{n}}\right)\right\},$$

which, following the same variable transformation as above, turns into studying local properties of the Poisson process with intensity 1 in the infinite strip $\{(x, y) \in \mathbb{R}^2 \colon 0 \leq y \leq \sqrt{3}\}$. We construct the Poisson distribution indirectly in the following way: since we are interested in the lengths of paths through a local number of points and the strip has a translation symmetry, we may assume that the first point is given by $p_1 = (0, y_1)$, where y_1 is uniformly distributed on $[0, \sqrt{3}]$. Adding now iteratively exponentially distributed random variables with parameter $\sqrt{3}$ to the first variables and replacing the second component by independent uniformly distributed random variables in $[0, \sqrt{3}]$ yields the Poisson process with intensity 1 in the strip.

FIGURE 2: Changing a zigzag path into something more effective.

2.4. Counting zigzags

The key observation in our improvement is the following: the BHH method is locally quite bad if we encounter what we will informally call a zigzag structure in the points: four consecutive points with a small difference in the x-coordinate but a large difference in the y-coordinate. More precisely, we will say that four points p_1, p_2, p_3, p_4 (ordered such that their x-coordinates increase) form a zigzag if

$$\|p_1 - p_3\| + \|p_3 - p_2\| + \|p_2 - p_4\| \le \|p_1 - p_2\| + \|p_2 - p_3\| + \|p_3 - p_4\|.$$

Given a zigzag, it is advantageous to locally change the structure of the path. This is illustrated in Figure 2.

We introduce some notation. Let x_2, x_3, and x_4 be i.i.d. variables distributed according to the exponential law $\sqrt{3}\exp(-\sqrt{3}z)$, and let y_1, y_2, y_3, and y_4 be i.i.d. random variables uniformly distributed in $[0, \sqrt{3}]$. We define four random points via

$$p_1 = (0, y_1), \qquad p_2 = (x_2, y_2), \qquad p_3 = (x_2 + x_3, y_3), \qquad p_4 = (x_2 + x_3 + x_4, y_4).$$

We know from the previous section that, for all $1 \le i \le 3$,

$$\mathbb{E}\|p_i - p_{i+1}\| = \left(2\int_0^\infty \int_0^{\sqrt{3}} \sqrt{z_1^2 + z_2^2}\exp\left(-\sqrt{3}z_1\right)\left(1 - \frac{z_2}{\sqrt{3}}\right)\mathrm{d}z_2\,\mathrm{d}z_1\right) \sim 0.92\ldots.$$

Given these four points, we introduce a stochastic event (A):

(A) $\|p_1 - p_3\| + \|p_3 - p_2\| + \|p_2 - p_4\| \le \|p_1 - p_2\| + \|p_2 - p_3\| + \|p_3 - p_4\|.$

Furthermore, we will introduce the respective (random) difference

$$X = (\|p_1 - p_2\| + \|p_2 - p_3\| + \|p_3 - p_4\|) - (\|p_1 - p_3\| + \|p_3 - p_2\| + \|p_2 - p_4\|).$$

Lemma 2. *We have*

$$\mathbb{E}(X \mid \mathrm{A})\mathbb{P}(\mathrm{A}) \ge \tfrac{9}{4}10^{-6}.$$

Proof. Since we are only trying to show a positive lower bound, rough estimates suffice. We study the event B, illustrated in Figure 3, and defined as

$$\left(x_2 \le \frac{\sqrt{3}}{9}\right) \wedge \left(x_3 \le \frac{\sqrt{3}}{9}\right) \wedge \left(x_4 \le \frac{\sqrt{3}}{9}\right) \wedge \left(\min(y_1, y_3) \ge \frac{8\sqrt{3}}{9}\right) \wedge \left(\max(y_2, y_4) \le \frac{\sqrt{3}}{9}\right).$$

These variables are independent and all distributions are explicitly given; thus, for $2 \le i \le 4$, we have

$$\mathbb{P}\left(x_i \le \frac{\sqrt{3}}{9}\right) = \int_0^{\sqrt{3}/9} \sqrt{3}\mathrm{e}^{-\sqrt{3}z}\,\mathrm{d}z = 1 - \frac{1}{\mathrm{e}^{1/3}},$$

https://doi.org/10.1239/aap/1427814579 Published online by Cambridge University Press

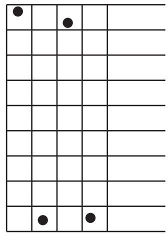

FIGURE 3: An instance of the event B.

while $\mathbb{P}(\min(y_1, y_3) \geq 8\sqrt{3}/9) = \mathbb{P}(\max(y_2, y_4) \leq \sqrt{3}/9) = \frac{1}{81}$. Then

$$\mathbb{P}(\mathrm{B}) = \left(1 - \frac{1}{\mathrm{e}^{1/3}}\right)^3 \left(\frac{1}{81}\right)^2 \geq 3 \times 10^{-6}.$$

At the same time, it follows from a simple computation that in the event B, we always have

$$X \geq \tfrac{3}{4}.$$

This implies that the event B is a subset of the event A and, thus,

$$\mathbb{E}(X \mid \mathrm{A})\mathbb{P}(\mathrm{A}) \geq \mathbb{E}(X \mid \mathrm{B})\mathbb{P}(\mathrm{B}) \geq \tfrac{9}{4}10^{-6}.$$

Proof of the upper bound. We follow the original idea of Beardwood *et al.* [4] and partition the unit square into strips: since we are dealing with a Poisson process, the behavior within each strip is independent of that in all other strips; focusing on one strip, we deal with a Poisson process of intensity n. For any set of random points arising from the Poisson process, we order them with increasing x-coordinate, i.e.

$$\pi_1(x_1) \leq \pi_1(x_2) \leq \cdots \leq \pi_1(x_k),$$

and consider the 4-tuples (x_1, x_2, x_3, x_4), (x_5, x_6, x_7, x_8), and so on (with possibly up to three points left at the end of each strip). Whether or not any of these 4-tuples contains a zigzag structure is an independent event: the computations in the previous section then imply that, with probability at least 3×10^{-6}, a zigzag yielding a reduction of length at least $3/(4\sqrt{n})$ is present. There are $n/4 - O(\sqrt{n})$ 4-tuples to consider, implying the reduction in length is of order $(n/4 - O(\sqrt{n}))(3/(4\sqrt{n}))(3 \times 10^{-6})$ and, thus,

$$\beta \leq \beta_{\mathrm{BHH}} - \tfrac{9}{16}10^{-6}.$$

Remark. The reduction in length was achieved by looking at $n/4 - O(\sqrt{n})$ independent events: usual arguments would allow us to conclude that the predicted reduction in length is actually tightly concentrated around its mean. This, however, is not necessary for our type of argument: we already know that β describes the limiting behavior almost surely and so the expected length of any construction of deterministic paths is necessarily an upper bound on β.

Remark. These problems exhibit a concentration of measure phenomenon implying the stability of Monte Carlo estimates, which will then usually imply much stronger results. For comparison, we computed ten samples of a million random points each, which suggests that

$$\mathbb{P}(\mathrm{A} \sim 0.1418, \qquad \mathbb{E}(X \mid \mathrm{A}) \sim 0.4187, \qquad \text{and, thus,} \quad \mathbb{E}(X \mid \mathrm{A})\mathbb{P}(\mathrm{A}) \geq 0.059,$$

with a standard deviation of 0.0003 and 0.001, respectively. This would imply that indeed

$$\beta \leq 0.906\,32.$$

2.5. Numerical estimates

Our result was aimed towards the clearest presentation of the idea. Improvements of the idea are rather obvious; however, they require somewhat accurate bounds for certain finite-dimensional integrals. One particular generalization is as follows: we could study not merely zigzags, but all 24 possible paths through six points leaving the first and the last points invariant; let us consider all 24 permutations over the symbols {2, 3, 4, 5} and denote the existence of an improved path as the stochastic event (C):

$$\text{(C)}\quad \inf_{\pi \in S_4(\{2,3,4,5\})} \|p_1 - p_{\pi(2)}\| + \sum_{i=2}^{4} \|p_{\pi(i+1)} - p_{\pi(i)}\| + \|p_6 - p_{\pi(5)}\| < \sum_{i=1}^{5} \|p_{i+1} - p_i\|.$$

The respective improvement is given by

$$Z = \sum_{i=1}^{5} \|p_{i+1} - p_i\| - \inf_{\pi \in S_4(\{2,3,4,5\})} \|p_1 - p_{\pi(2)}\| + \sum_{i=2}^{4} \|p_{\pi(i+1)} - p_{\pi(i)}\| - \|p_6 - p_{\pi(5)}\|.$$

Monte Carlo methods (10 samples of 50 000 sets of points each) suggest that

$$\mathbb{P}(\text{C}) \sim 0.3721 \quad \text{and} \quad \mathbb{E}(Z \mid \text{C}) \sim 0.4990,$$

with a standard deviation of 0.02 and 0.004, respectively. These values would suggest $\beta \leq 0.8902$. As already hinted at in the introduction, there is a natural limit to these improvements: we study paths through random points with an additional restriction on their movement in one of the two dimensions, which corresponds to a different functional and this difference will be a great hindrance to further major improvements.

3. Proof of the lower bound

3.1. The original argument

Proving an upper bound can be done (and has been done) by constructing an explicit path. To prove a lower bound we have to pursue an entirely different strategy since we have very little idea what an optimal path could look like: we already know, however, that it is sufficient to prove lower bounds on the expected length of the traveling salesman path through points of a Poisson process with intensity n in $[0, 1]^2$. The only real basic information about paths at our disposal is that, for every point, there are two points to which that particular point is connected: suppose now that, for every point p, these two points are also at the same time the two nearest neighbors of p. Beardwood *et al.* [4] showed that the average sum of the distances to the two nearest neighbors is explicitly computable and certainly any traveling salesman has to be longer than that. We may assume that the Poisson process is actually distributed with the intensity n on all of $\mathbb{R}^2$: adding more points can only decrease the expected distance and allows us to disregard the behavior of the process close to the boundary of $[0, 1]^2$. The following lemma is usually a basic exercise for students (see, e.g. [12]).

Lemma 3. *Let P_n be a Poisson process on $\mathbb{R}^2$ with intensity n. Then, for any fixed point $p \in \mathbb{R}^2$, the probability distribution of the distance between p and the nearest point in $\mathcal{P}_n$ is given by*

$$f(r) = 2\pi n r \mathrm{e}^{-\pi n r^2},$$

https://doi.org/10.1239/aap/1427814579 Published online by Cambridge University Press

and the distance to the second-nearest neighbour is given by

$$g(r) = 2\pi^2 n^2 r^3 \mathrm{e}^{-\pi n r^2}.$$

It follows from standard calculations that the distance r to the nearest point has expectation

$$\int_0^\infty r f(r)\,\mathrm{d}r = \frac{1}{2\sqrt{n}},$$

while the distance to the next-to-nearest point has expectation

$$\int_0^\infty r g(r)\,\mathrm{d}r = \frac{3}{4\sqrt{n}}.$$

Given a traveling salesman path, every point is connected to two other points—in the worst case, these are the nearest and the next-to-nearest points in all cases, yielding a lower bound of

$$\beta \geq \left(\tfrac{1}{2} + \tfrac{3}{4}\right)\tfrac{1}{2} = 0.625,$$

which is the original result of Beardwood *et al.* [4].

3.2. An improvement

The previous argument assumed that it is always the worst case that occurs: every point is connected to its two closest neighbors. This, however, is not possible if we have the following constellation of points: a point a with closest point b at distance r_1 and its second-closest point c at distance $r_2 > r_1$ and third-closest point d at distance $r_3 > r_1 + 2r_2$. Our proof rests on an analysis of this situation.

Lemma 4. *Let P_n be a Poisson process on $\mathbb{R}^2$ with intensity n. Then, for any fixed point $p \in \mathbb{R}^2$, the probability distribution of the distance between p and the closest, second-closest, and third-closest points is given by*

$$h(r_1, r_2, r_3) = \begin{cases} \mathrm{e}^{-n\pi r_3^2}(2n\pi)^3 r_1 r_2 r_3 & \text{if } r_1 < r_2 < r_3, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. We study the probability of precisely one point at distance $(r_1, r_1 + \varepsilon)$ (event A), precisely one point in $(r_2, r_2 + \varepsilon)$ (event B), and precisely one point in $(r_3, r_3 + \varepsilon)$ (event C) and no points in between (event D). The probabilities for these events including their expansion up to first order in ε are

$$\begin{aligned}
\mathbb{P}(\mathrm{A}) &= ((r_1 + \varepsilon)^2 - r_1^2) n\pi \mathrm{e}^{(-(r_1+\varepsilon)^2 + r_1^2)n\pi} = 2n\pi\varepsilon r_1 + O(\varepsilon^2), \\
\mathbb{P}(\mathrm{B}) &= ((r_2 + \varepsilon)^2 - r_2^2) n\pi \mathrm{e}^{(-(r_2+\varepsilon)^2 + r_2^2)n\pi} = 2n\pi\varepsilon r_2 + O(\varepsilon^2), \\
\mathbb{P}(\mathrm{C}) &= ((r_3 + \varepsilon)^2 - r_3^2) n\pi \mathrm{e}^{(-(r_3+\varepsilon)^2 + r_3^2)n\pi} = 2n\pi\varepsilon r_3 + O(\varepsilon^2), \\
\mathbb{P}(\mathrm{D}) &= \mathrm{e}^{-r_1^2\pi n} \mathrm{e}^{(-r_2^2 + (r_1+\varepsilon)^2)\pi n} \mathrm{e}^{(-r_3^2 + (r_2+\varepsilon)^2)\pi n} = \mathrm{e}^{-n\pi r_3^2} + O(\varepsilon).
\end{aligned}$$

This immediately implies the statement.

Proof of the lower bound. We start by showing that both the nearest as well as the next-to-nearest points of any element in $\{a, b, c\}$ also lie in the set. Let x be some other point with $x \notin \{a, b, c\}$. Then

$$\|b - x\| \geq \|a - x\| - r_1 \geq r_3 - r_1 > 2r_2 \geq r_1 + r_2 \geq \|b - c\|,$$

https://doi.org/10.1239/aap/1427814579 Published online by Cambridge University Press

FIGURE 4: Point a and the three closest points of a.

and, therefore, the second-closest point from b is a or c. By the same token,

$$\|c - x\| \geq \|a - x\| - r_2 \geq r_3 - r_2 \geq r_1 + r_2 \geq \|b - c\|,$$

and, therefore, the second-closest point from c is a or b. Cases of equality have probability 0 and can be ignored. See Figure 44.

If we simply connect every point to its two closest neighbors, we end up with a triangle where every point is connected to the two other points but no other point except those. This is clearly not possible for a traveling salesman path. Let us first compute the frequency of such an event. Using Lemma 4, the probability of all of these distance relations being true for a fixed point a is

$$\int_0^\infty \int_{r_1}^\infty \int_{r_1+2r_2}^\infty \mathrm{e}^{-n\pi r_3^2}(2n\pi)^3 r_1 r_2 r_3 \,\mathrm{d}r_3 \,\mathrm{d}r_2 \,\mathrm{d}r_1 = \frac{7}{324}.$$

There is a lack of independence: if it is true for a, it is likely to be true for b and c as well; thus, we have only the trivial bound

$$\tfrac{1}{3}\tfrac{7}{324}n = \tfrac{7n}{972}$$

on the number of triples of points with this property. However, if the case occurs then the algorithm connecting every point to its two nearest neighbors has an expected length which can be bounded from above by

$$r_1 + r_2 + 2||a - c|| \leq 3(r_1 + r_2),$$

where the distance $\|a - c\|$ has to be counted twice because the algorithm cannot 'see' that it has created a triangle and counts the distance twice.

In the case of three points isolated from the rest, there is one special case which is the easiest to connect to the remaining points: this is when b can be connected to a point d having distance r_3 from a, c can be connected to a different point e also at distance r_3 from a, and, additionally, b lies on the line $\overline{ad}$ and c lies on $\overline{ae}$. This is illustrated in Figure 5. In this case, the required length is

$$\|d - b\| + \|b - a\| + \|a - c\| + \|c - e\| \geq (r_3 - r_1) + r_1 + r_2 + (r_3 - r_2) = 2r_3.$$

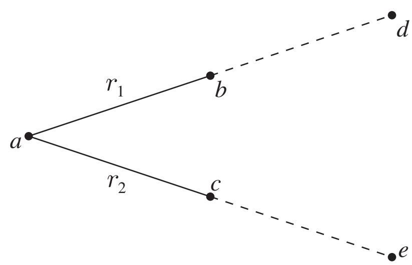

FIGURE 5: The best of the worst case.

https://doi.org/10.1239/aap/1427814579 Published online by Cambridge University Press

This implies that whenever we are in this particular configuration, the actual path has to be at least a length $2r_3 - 3(r_1 + r_2)$ longer than what the greedy algorithm suggests. Note that

$$2r_3 - 3(r_1 + r_2) \geq r_2 - r_1 \geq 0,$$

and that we always gain something at this point. In expectation, this is an average length of

$$\frac{324}{7}\int_0^\infty \int_{r_1}^\infty \int_{r_1+2r_2}^\infty (2r_3 - 3r_1 - 3r_2)\mathrm{e}^{-n\pi r_3^2}(2n\pi)^3 r_1 r_2 r_3 \,\mathrm{d}r_3\,\mathrm{d}r_2\,\mathrm{d}r_1 = \frac{57}{112}\frac{1}{\sqrt{n}}.$$

Altogether, this gives the lower bound

$$\beta \geq \left(\tfrac{1}{2} + \tfrac{3}{4}\right)\tfrac{1}{2} + \tfrac{7}{972}\tfrac{57}{112} = \tfrac{5}{8} + \tfrac{19}{5184}.$$

Acknowledgements

I am grateful to Steven Finch for comments on the history of the problem, and indebted to J. Michael Steele for encouraging me to write this paper, many helpful remarks on how to improve the exposition, and a series of very valuable discussions. The author was supported by grant SFB 1060 from the German Research Foundation (DFG).

References

[1] Applegate, D. L., Bixby, R. E., Chvátal, V. and Cook, W. J. (2006). *The Traveling Salesman Problem: A Computational Study*. Princeton University Press.
[2] Avram, F. and Bertsimas, D. (1992). The minimum spanning tree constant in geometrical probability and under the independent model: a unified approach. *Ann. Appl. Prob.* **2,** 113–130.
[3] Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximation*. Oxford University Press.
[4] Beardwood, J., Halton, J. H. and Hammersley, J. (1959). The shortest path through many points. *Proc. Camb. Philos. Soc.* **55,** 299–327.
[5] Borovkov, A. A. (1962). A probabilistic formulation of two economic problems. *Dokl. Akad. Nauk SSSR* **146,** 983–986 (in Russian).
[6] Cook, W. J. (2012). *In Pursuit of the Traveling Salesman: Mathematics at the Limits of Computation*. Princeton University Press.
[7] Fejes, L. (1940). Über einen geometrischen Satz. *Math. Z.* **46,** 83–85.
[8] Few, L. (1955). The shortest path and the shortest road through n points. *Mathematika* **2,** 141–144.
[9] Finch, S. R. (2003). *Mathematical Constants* (Encyclopedia Math. Appl. **94**). Cambridge University Press.
[10] Gutin, G. and Punnen, A. P. (eds) (2002). *The Traveling Salesman Problem and Its Variations*. Kluwer, Dordrecht.
[11] Johnson, D. S., McGeoch, L. A. and Rothberg, E. E. (1996). Asymptotic experimental analysis for the Held–Karp traveling salesman bound. In *Proceedings of the Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*. ACM, New York, pp. 341–350.
[12] Mathai, A. M. (1999). *An Introduction to Geometrical Probability: Distributional Aspects with Applications* (Statist. Distributions Models Appl. **1**). Gordon and Breach, Amsterdam.
[13] Percus, A. G. and Martin, O. C. (1996). Finite size and dimensional dependence in the Euclidean traveling salesman problem. *Phys. Rev. Lett.* **76,** 1188–1191.
[14] Redmond, C. and Yukich, J. E. (1994). Limit theorems and rates of convergence for Euclidean functionals. *Ann. Appl. Prob.* **4,** 1057–1073.
[15] Steele, J. M. (1981). Subadditive Euclidean functionals and nonlinear growth in geometric probability. *Ann. Prob.* **9,** 365–376.
[16] Steele, J. M. (1997). *Probability Theory and Combinatorial Optimization* (CBMS-NSF Regional Conf. Ser. Appl. Math. **69**). SIAM, Philadelphia, PA.
[17] Valenzuela, C. L. and Jones, A. J. (1997). Estimating the Held–Karp lower bound for the geometric TSP. *Europ. J. Operat. Res.* **102,** 157–175.
[18] Venkatesh, S. S. (2012). *The Theory of Probability: Explorations and Applications*. Cambridge University Press.
[19] Verblunsky, S. (1951). On the shortest path through a number of points. *Proc. Amer. Math. Soc.* **2,** 904–913.
[20] Yukich, J. E. (1998). *Probability Theory of Classical Euclidean Optimization Problems* (Lecture Notes Math. **1675**). Springer, Berlin.