Lower Bounds for Insertion Methods for TSP

Yossi Azar *

Abstract

We show that the random insertion method for the traveling salesman problem (TSP) may produce a tour $\Omega(\log\log n / \log\log\log n)$ times longer than the optimal tour. The lower bound holds even in the Euclidean Plane. This is in contrast to the fact that the random insertion method performs extremely well in practice. In passing we show that other insertion methods may produce tours $\Omega(\log n / \log\log n)$ times longer than the optimal one. No non-constant lower bounds were previously known.

*Department of Computer Science, Tel-Aviv University, Israel.

1 Introduction

The traveling salesman problem (TSP) is one of the most notorious NP-hard problems [GJ]. For the special case that distances satisfy the triangle inequality, many approximation algorithms have been developed and analyzed. The *approximation factor* of such an algorithm is the ratio between the length of the tour obtained by the algorithm and the optimal tour. The relative performance of different heuristics is measured by comparing their approximation factors and their running times. Rosenkrantz et al [RSL] defined and analyzed several heuristics. *Insertion methods* are a particularly important class of (tour-construction) heuristics. They work as follows: Vertices are inserted into the tour one at a time. A vertex is inserted between two consecutive vertices in the current tour where it fits best. More formally, after the ith insertion, the algorithm has a subtour T_i on a subset of i vertices S_i. Suppose that $v \notin S_i$ is the $(i+1)$st vertex inserted, and that (x, y) is an edge of T_i that minimizes $d(x, v) + d(v, y) - d(x, y)$. The new tour T_{i+1} (on vertices $S_i \cup v$) is obtained from T_i by deleting edge (x, y) and adding edges (x, v) and (v, y). (The initial tour is an edge of length zero between some vertex to itself.)

The algorithms in this family differ in the order in which vertices are inserted and thus may provide different tours. Clearly, there are $n!$ possible orders in which to insert the vertices. *Arbitrary insertion*, the generic algorithm in the family, inserts the vertices in an arbitrary order. Rosenkrantz et al [RSL] showed a $\lceil \log n \rceil + 1$ upper bound on the approximation factor of arbitrary insertion. They also showed that two specific schemes, *Nearest insertion* and *Cheapest insertion*, achieve an approximation ratio of 2. The question of whether the logarithmic growth permitted by their upper bound for the arbitrary insertion method can be achieved remained open. In fact, they knew of no example that achieved an approximation ratio of more than 4 and suggested that a constant upper bound may be possible. In contrast we prove:

Theorem 1.1 *There exist some insertion methods whose worst case approximation factor is* $\Omega(\log n / \log \log n)$. *The lower bound holds even in the Euclidean Plane.*

Another interesting insertion method is *random insertion*: the order in which the vertices are inserted is chosen uniformly at random. This method is of special interest since it performs better than nearest insertion and cheapest insertion in practice (see [Be],[GBDS],[LLRS]). Moreover, it is easier to implement and has lower running time. However, no better bounds on the performance of random insertion were known

([RSL], [LLRS]). It was tempting to think that random insertion may have a constant approximation factor. Surprisingly, we prove a non-constant lower bound for random insertion.

Theorem 1.2 *The worst case approximation factor of the random insertion method is* $\Omega(\log\log n/\log\log\log n)$ *even with probability* $1-o(1)$*. The lower bound holds even in the Euclidean Plane.*

It would be interesting to know if these techniques may also yield a non-constant lower bound for *Farthest insertion* method, when the farthest point is inserted at each step. This method performs better in practice than the other methods. The best known lower bound for this method is constant [Hu]: there is a metric space for which it is 6.5, and it is 2.43 for the plane.

Theorem 1.1 was proved independently and at the same time by Bafna, Kalyanasundaram and Pruhs [BKP]. The basic approach in this paper resembles the one of Bentley and Saxe in [BS] for the nearest neighbor algorithm and of Alon and Azar [AA] for on-line Steiner trees, but some different ideas are required.

2 The lower bound proofs

We first prove Theorem 1.1. The metric space considered is the Euclidean plane. All the points are in the unit square. Let x be an integer, $x \geq 5$. We construct a set of n points, $x^{3x} < n \leq 2x^{3x}$, such that the length of the optimal TSP tour on these points is $\Theta(1)$ whereas the length of some insertion method tour is $\Omega(x) = \Omega(\log n/\log\log n)$. This yields a lower bound of $\Omega(\log n/\log\log n)$, as needed.

The points consist of $x+1$ major layers and x minor layers, where each layer is a set of equally spaced points on a horizontal line of length 1. Let $a_i = x^{-3i}$ and $l_i = 1/a_i$ for $0 \leq i \leq x$. Thus $a_0 = 1$, $a_1 = x^{-3}$ and $a_x = x^{-3x}$. The coordinates of the j'th points in major layer number i, denoted by $v_{i,j}$, is (ja_i, b_i) for $0 \leq i \leq x$ and $0 \leq j \leq l_i$. Hence in major layer 0 there are only two points, in major layer 1 there are x^3+1 and so on up to major layer number x which contains $x^{3x}+1$ points. Let $b_0 = 0$. The vertical distance between major layer number i and major layer number $i+1$ is $c_i = b_{i+1} - b_i$, where for all $0 \leq i \leq x-1$, $c_i = a_i/x$. For $0 \leq i \leq x-1$ the minor layer i is precisely in the middle between major layer i to major layer $i+1$ and it is a copy of major layer i without the left most points. Thus, the coordinates of the j'th points in minor layer i, denoted by $y_{i,j}$, is $(ja_i, b_i + c_i/2)$ for $0 \le i \le x-1$ and $1 \le j \le l_i$.

The order of inserting the points of the major layers is layer by layer $0 \le i \le x+1$. In each major layer from left to right $0 \le j \le l_i$. The points of minor layer i are inserted after the points of major layer $i+1$ (and before major layer $i+2$). In each minor layer the inserting order is by decreasing indices i.e., from right to left.

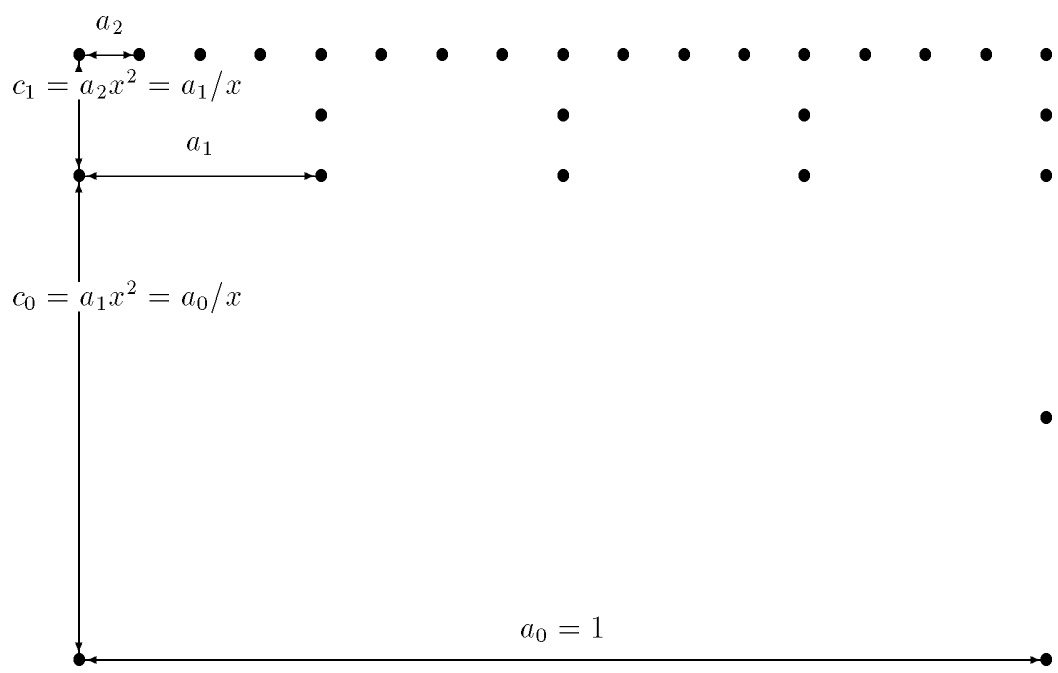

Figure 1: An example of the construction

First, observe that

$$x^{3x} < n \le x^{3x} + 1 + 2 \cdot \sum_{i=0}^{x-1}(1/a_i + 1) = (x^{3x} + 1) + 2 \cdot \sum_{i=0}^{x-1}(x^{3i} + 1) \le 2x^{3x}.$$

Note also that

$$b_x = \sum_{i=0}^{x-1} c_i = \sum_{i=0}^{x-1} a_i/x = (1/x)\sum_{i=0}^{x-1} x^{-3i} \le 2/x \le 1$$

and therefore all the points lie, indeed, in the unit square.

Next observe that the length of the optimal spanning tree is $O(1)$ and therefore the length of the optimal TSP tour is also $O(1)$. Indeed, one can take the horizontal line in the last layer (major layer number x) together with vertical lines from it to any other point.

The total length of this tree is

$$1+\sum_{i=0}^{x-1} c_i(\frac{1}{a_i}+1) \leq (1+\sum_{i=0}^{x-1} 2c_i/a_i) = (1+x\frac{2}{x}) = 3.$$

On the other hand, we should analyze the tour generated by the insertion method. It is straightforward to see that after completing the first two major layers ($i=0,1$) and the first minor layer ($i=0$) the tour is $v_{0,0}, v_{1,0}, v_{1,1}, \ldots, v_{1,l_1}, y_{0,1}, v_{0,1}, v_{0,0}$.

We will prove by induction (where the previous case is the base case) that after adding the vertex $v_{i,j}$ where $i>1$ the tour looks like that (see fig. 2)

$$\begin{aligned}
v_{0,0}, v_{1,0}, \ldots, v_{i-1,0}, v_{i-1,0}, v_{i,0},\\
v_{i,1}, v_{i,2}, \ldots, v_{i,j},\\
v_{i-1,1}, v_{i-1,2}, \ldots, v_{i-1,l_{i-1}}\\
y_{i-2,l_{i-2}}, y_{i-2,l_{i-2}-1}, \ldots, y_{i-2,1},\\
v_{i-2,1}, v_{i-2,2}, \ldots, v_{i-2,l_{i-2}},\\
y_{i-3,l_{i-3}}, \ldots, y_{i-3,1},\\
.\\
.\\
v_{2,1}, \ldots, v_{2,l_2},\\
y_{1,l_2}, y_{1,l_2-1}, \ldots, y_{1,1},\\
v_{1,1}, v_{1,2}, \ldots, v_{1,l_1},\\
y_{0,1},\\
v_{0,1}, v_{0,0}.
\end{aligned}$$

After adding $y_{i-1,j}$ where $i>1$ the tour looks like that (see fig. 3)

$$\begin{aligned}
v_{0,0}, v_{1,0}, \ldots, v_{i-1,0}, v_{i,0},\\
v_{i,1}, v_{i,2}, \ldots, v_{i,l_i},\\
y_{i-1,l_{i-1}}, y_{i-1,l_{i-1}-1}, \ldots, y_{i-1,j},\\
v_{i-1,1}, v_{i-1,2}, \ldots, v_{i-1,l_{i-1}}
\end{aligned}$$

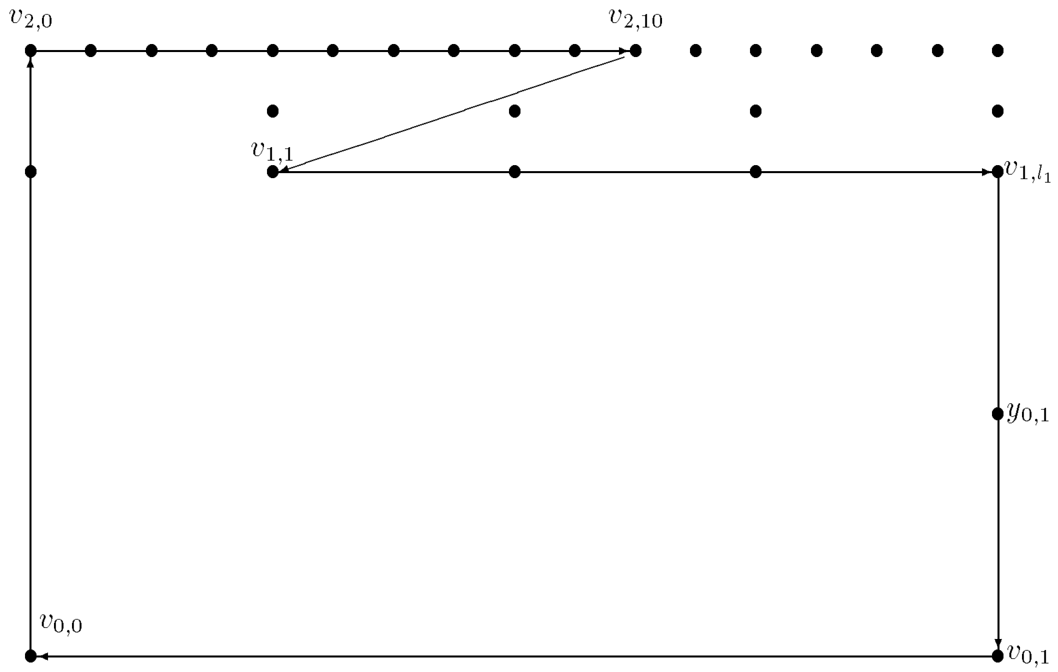

Figure 2: After adding $v_{2,10}$

$$\begin{aligned} y_{i-2,l_{i-2}}, y_{i-2,l_{i-2}-1}, \ldots, y_{i-2,1}, \\ v_{i-2,1}, v_{i-2,2}, \ldots, v_{i-2,l_{i-2}}, \\ y_{i-3,l_{i-3}}, \ldots, y_{i-3,1}, \\ . \\ . \\ v_{2,1}, \ldots, v_{2,l_2}, \\ y_{1,l_2}, y_{1,l_2-1}, \ldots, y_{1,1}, \\ v_{1,1}, v_{1,2}, \ldots, v_{1,l_1}, \\ y_{0,1}, \\ v_{0,1}, v_{0,0} \end{aligned}$$

It is not difficult to verify that after the last vertex has been added, the length of the tour described above lies between $2x$ and $2x+2$ as needed.

In order to prove that the tour is as described we prove the following statements. The first is that $v_{i,0}$ for $i > 1$ ($i = 0, 1$ are the base cases) is inserted between $v_{i-1,0}$

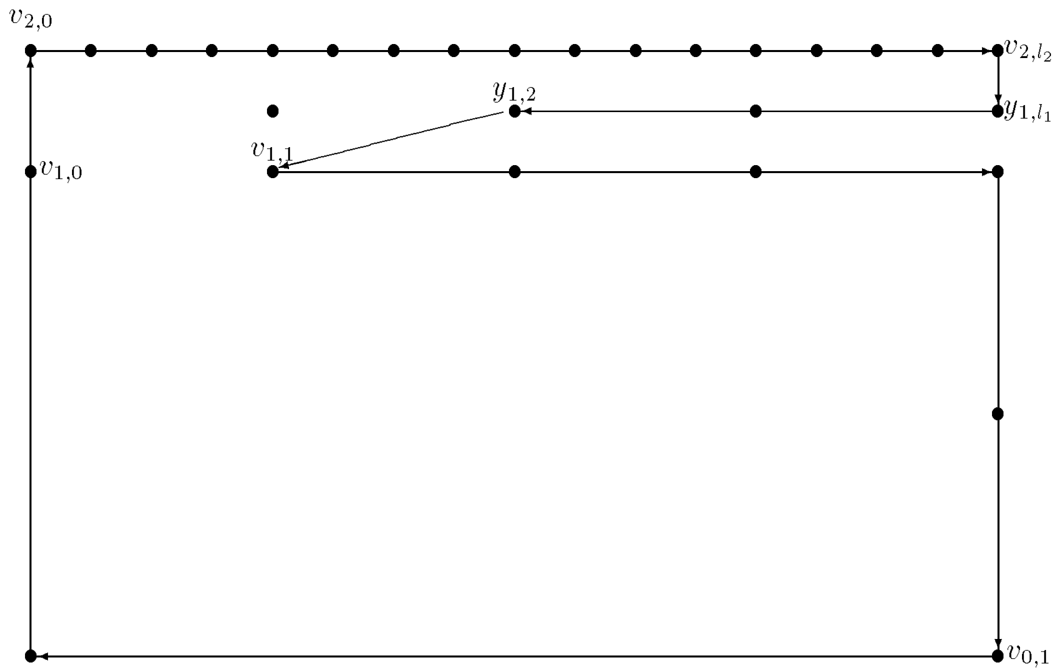

Figure 3: After adding $y_{1,2}$

and $v_{i-1,1}$. Second, $v_{i,j}$ ($i > 1$ and $j > 0$) is inserted between $v_{i,j-1}$ and $v_{i-1,1}$. Third, $y_{i-1,l_{i-1}}$ is inserted between v_{i,l_i} and $v_{i-1,1}$. At last, $y_{i-1,j}$ is inserted between $y_{i-1,j+1}$ and $v_{i-1,1}$. That would complete the proof by induction.

The first case is straightforward, since by simple geometry the cost of inserting $v_{i,0}$ is seen to be less than $2c_{i-1}$ whereas other ways of insertion lead to costs at least $2c_{i-1}$. The third case is as easy since by simple geometry, the cost of inserting $y_{i-1,l_{i-1}}$ between v_{i,l_i} and $v_{i-1,1}$ is smaller than $c_{i-1}/2$ where it is at least $c_{i-1}/2$ for any other choice. Next we prove the second case. First observe that inserting $v_{i,j}$ between $v_{i,j-1}$ and $v_{i-1,1}$ costs less than $2a_i$. The other reasonable candidates are between consecutive vertices of major/minor layers k such that $k \leq i-1$ (horizontal edges) apart from minor $i-1$ which is not on the tour yet. This requires a cost of at least $2(\sqrt{l^2+z^2}-l)$ where $l = a_k/2$ and $z = c_k/2$. But

$$2(\sqrt{l^2+z^2}-l) = a_k(\sqrt{1+(c_k/a_k)^2}-1) = a_k(\sqrt{1+1/x^2}-1).$$

Thus, by using the inequality (for $t \geq 0$)

$$\sqrt{1+t}-1 = \frac{t}{\sqrt{1+t}+1} \geq \frac{t}{1+t/2+1} = \frac{2t}{4+t}$$

we get that the insertion cost is bounded below by

$$a_k \frac{2/x^2}{4+1/x^2} = a_k \frac{2}{4x^2+1} = a_{k+1} \frac{2x^3}{4x^2+1} > 2a_{k+1} \geq 2a_i$$

for $x \geq 5$, which completes the proof of case 2. For case 4 note that inserting $y_{i-1,j}$ between consecutive vertices in major layer $i-1$ or i costs at least $c_{i-1}/2$. Moreover, as in the previous case inserting it between consecutive vertices of major/minor layers k such that $k \leq i-2$ costs at least $2a_{k+1} \geq 2a_{i-1} > c_{i-1}/2$. On the other hand we will show that inserting it between $y_{i-1,j+1}$ and $v_{i-1,1}$ costs at most $c_{i-1}/2$. The last statement is obvious for $j=1$. For $j>1$ draw a vertical line from $A = y_{i-1,j}$ until it hits (at point D) the line connecting $B = y_{i-1,j+1}$ and $C = v_{i-1,1}$. Clearly $|AD| \leq c_{i-1}/4$. But,

$$|AB|+|AC|-|BC| \leq |AD|+|DB|+|AD|+|DC|-(|BD|+|DC|) = 2|AD| \leq c_{i-1}/2$$

which completes the proof of case 4 and therefore the proof of Theorem 1.1.

Next we prove Theorem 1.2. Recall that in the proof of Theorem 1.1, we constructed a set T of m vertices and an order π, for which the length of the tour constructed by the insertion method for the set T using order π is larger by a factor of $\Omega(\log m/\log\log m)$ from the optimal tour. Denote the vertices in order π by $u_1, \ldots, u_m$. For $1 \leq i < m$ replace u_i by a set S_i of $n_i = m^{2(m-i)} - m^{2(m-i-1)}$ vertices in the same location. Let S_m be the set of the one vertex u_m, and thus $n_m = 1$. Let $S = \cup_i S_i$. Clearly $|S| = \sum_i n_i = m^{2(m-1)}$. It is immediate that the optimal tour for the set S has the same length as the optimal tour for the set T (essentially all the non-zero length edges are the same). For $1 \leq i < m$ denote by A_i the event that by chosing at random an order on the set S the first occurrence of a vertex in S_i is after the first occurrence of a vertex in $\cup_{j=i+1}^m S_j$. Clearly

$$\Pr[A_i] = \frac{\sum_{j=i+1}^m n_i}{\sum_{j=i}^m n_i} = \frac{m^{2(m-i-1)}}{m^{2(m-i)}} = \frac{1}{m^2}\ .$$

Let A be the event that neither of the A_i has happened. Clearly

$$\Pr[A] \geq 1 - \frac{m-1}{m^2} > 1 - \frac{1}{m}\ .$$

It is straightforward to check that for all orders in the event A, the tours constructed by the random insertion for the set S are the same, (up to the order of vertices in each S_i). Moreover, they have the same length as the tour constructed by the insertion method for the set T using the order π, since all positive edges are the same. Thus we conclude that with probability $1 - o(1)$ the tour constructed by the random insertion method on the set S of $n = m^{2(m-i)}$ is longer by a factor of $\Omega(\log m / \log\log m) = \Omega(\log\log n / \log\log\log n)$ from the optimal tour. This completes the proof of Theorem 1.2.

3 Acknowledgement

I would like to thank Noga Alon and John Hershberger for helpful discussions. I would like to thank Cor Hurkens and the referee for pointing out and fixing errors in an earlier draft.

References

[AA] N. Alon and Y. Azar, *On-line Steiner trees in the Euclidean plane, Discrete and Computational Geometry* 10 (1993) pp. 113-121.

[Be] J.L. Bentley, *Experiments on Traveling Salesman Heuristics, in Proc. 1st Annual ACM-SIAM SODA* San-Francisco, California, 1990 pp. 91-99.

[BKP] V. Bafna, B. Kalyanasundaram and K. Pruhs, *Not all insertion methods yields constant approximate tours in the Euclidean plane,* Manuscript.

[BM] B. Bollobás and A. Meir, *A traveling salesman problem in the k-dimensional unit cube, Operations Research Letters*, 11 (1992) 19-21.

[BS] J.L. Bentley and J.B. Saxe, *An analysis of two heuristics for the Euclidean Traveling salesman, 18th Annual Allerton Conference on Communication, Control and Computing,* Monticello, 1980, pp. 41-49.

[GBDS] B.L. Golden, L.D. Bodin, T. Doyle and W. Stewart, *Approximate traveling salesman algorithms, Oper. Res. 28,*(1980), pp. 694-711.

[GJ] M. R. Garey and D. S. Johnson, *Computers and Intractability: a guide to the theory of NP-completeness*, Freeman and Company, New York, 1979.

[Hu] C. Hurkens, *Nasty TSP instances for Farthest Insertion, 2nd IPCO,* Pittsburgh 1992.

[IW] M. Imase and B.M. Waxman, *Dynamic Steiner tree problem, SIAM J. Disc Math. 4,* (1991), pp. 369-384.

[LLRS] E. Lawler, J. Lenstra, A. Rinnooy Kan, and D. Shmoys, *The Traveling Salesman Problem*, Wiley, New York, 1985.

[Me] A. Meir, *A geometric problem involving the nearest neighbor algorithm, Operations Research Letters 6* (1987), pp. 289-291.

[Ne] D.J. Newman, *A problem seminar, Springer,* Berlin 1982, 9, Problem 57.

[RSL] D.J. Rosenkrantz, R.E. Stearns and P.M. Lewis II, *An analysis of several heuristics for the traveling salesman problem, SIAM J. Computing 6*, (1977), pp. 563-581.