SIAM J. COMPUT.
Vol. 24, No. 3, pp. 665–671, June 1995

015

A PRIORI BOUNDS ON THE EUCLIDEAN TRAVELING SALESMAN*

TIMOTHY LAW SNYDER† AND J. MICHAEL STEELE‡

Abstract. It is proved that there are constants c_1, c_2, and c_3 such that for any set S of n points in the unit square and for any minimum-length tour T of S (1) the sum of squares of the edge lengths of T is bounded by $c_1 \log n$, (2) the number of edges having length t or greater in T is at most c_2/t^2, and (3) the sum of edge lengths of any subset E of T is bounded by $c_3|E|^{1/2}$. The second and third bounds are independent of the number of points in S, as well as their locations. Extensions to dimensions $d > 2$ are also sketched. The presence of the logarithmic term in (1) is engaging because such a term is not needed in the case of the minimum spanning tree and several analogous problems, and, furthermore, we know that there always exists *some* tour of S (which perhaps does not have minimal length) for which the sum of squared edges is bounded independently of n.

Key words. Euclidean traveling salesman problem, inequalities, squared edge lengths, long edges

AMS subject classifications. 68R10, 05C45, 90C35, 68U05

1. Introduction.

The purpose of this note is to provide a priori bounds on quantities related to the edge lengths of an optimal traveling salesman (minimum-length) tour through n points in the unit square. By *a priori* we mean that the bounds are independent of the locations of the points.

Studies of a priori bounds were initiated by Verblunsky (1951) and Few (1955). Few showed that for any set S of n points in the unit square, the length of an optimal traveling salesman tour of S is at most $\sqrt{2n} + 1.75$. Few's result led to a series of improvements, culminating in Karloff (1989), where it was shown that Few's constant could be reduced to less than $\sqrt{2}$. Goddyn (1990) improved similar results in higher dimensions. Our results continue in this tradition by giving a priori inequalities for three other quantities related to the edge lengths of an optimal traveling salesman tour.

The interest in and subtlety of our inequalities comes from the fact that, in contrast to the minimum spanning tree (MST) problem, optimal solutions to the traveling salesman problem (TSP) are not invariant under monotone transformations of the edge weights. Before giving further details on this connection and other related work, we state our main results. We let $|e| = |x - y|$ denote the Euclidean length of the edge $e = xy$ with vertices x and y in $\mathbb{R}^2$ and, in settings where the order of the edges of an optimal tour is not important, we represent a traveling salesman tour by the edge set $\{e_1, e_2, \ldots, e_n\}$. In what follows, an "optimal" traveling salesman tour is a tour that is of minimum length when using Euclidean edge weights.

Our first theorem bounds the sum of squared edge lengths of any optimal traveling salesman tour.

THEOREM 1. *There exists a constant* $0 < c_1 < \infty$ *such that if* $T = \{e_1, e_2, \ldots, e_n\}$ *is an optimal traveling salesman tour of* $\{x_1, x_2, \ldots, x_n\} \subset [0, 1]^2$ *and if* $n \geq 2$, *then*

$$\sum_{i=1}^{n} |e_i|^2 \leq c_1 \log n. \tag{1.1}$$

Theorem 2 is a bound on the number of edges that are of length t or greater.

*Received by the editors February 24, 1992; accepted for publication (in revised form) January 21, 1994.

†Department of Computer Science, Georgetown University, Washington, DC 20057. The research of this author was supported in part by Georgetown University 1991 Summer Research Award, Georgetown University 1992 Junior Faculty Research Fellowship, and Georgetown College John R. Kennedy, Jr. Faculty Research Fund.

‡Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104. The research of this author was supported in part by the following grants: National Science Foundation grant DMS88-12868 and National Security Agency grants MDA904-89-H-2034, AFOSR-89-0301, and DAAL03-89-G-0092.

THEOREM 2. *There exists a constant* $0 < c_2 < \infty$ *such that, if* $\nu(n,t)$ *is the number of* $e_i \in T$ *such that* $|e_i| \geq t$, *then for all* $t > 0$ *and* $n \geq 1$,

$$\nu(n,t) \leq c_2/t^2. \tag{1.2}$$

Theorem 3 gives a bound on the total length of any k-edge subset of an optimal TSP tour.

THEOREM 3. *There exists a constant* $0 < c_3 < \infty$ *such that for each* $E = \{e_{i_1}, e_{i_2}, \ldots, e_{i_k}\} \subseteq T$, *we have*

$$\sum_{i \in E} |e_i| \leq c_3\sqrt{k}. \tag{1.3}$$

It is interesting to compare these results to their minimum spanning tree analogues. Steele and Snyder (1989) proved MST analogues to (1.2) and (1.3), but these proofs were predicated on a solution to the MST problem via a greedy algorithm and thus were not applicable to the TSP. The best TSP analogue to (1.2) was therefore $\nu_{\text{TSP}}(n,t) \leq c_{\text{TSP}}\sqrt{n}/t$, for some constant c_{TSP}. The bounds (1.2) and (1.3), however, are independent of n, the number of points, as well as the locations of the points. For this reason, we say that (1.2) and (1.3) are *fully a priori* inequalities.

Inequalities like (1.1) are important in simulations and investigations in which square root computations required for Euclidean lengths are deemed to be too expensive (cf. the discussion in Steele (1990)). It was observed in Steele (1990) in an application of the space-filling curve heuristic that one could obtain a result like (1.1) for the MST, but without the logarithmic factor. Although this result might make the logarithmic term of (1.1) seem disappointing, it is actually best possible since Bern and Eppstein (1993) recently showed that there exist a positive constant c and point sets S with $|S| \to \infty$ such that $\sum_{e \in T} |e|^2 \geq c \log |S|$.

Part of the interest in these results comes from the fact that there are closely connected inequalities that exhibit strikingly different behavior from the optimal TSP tour. In particular, there is a constant c' and for all n there is a nonminimal length tour T' of S with $|S| = n$ such that, for all $n \geq 2$, $\sum_{e \in T'} |e|^2 \leq c'$. These tours can be obtained via the space-filling heuristic as noted in the discussion of the MST. Neumann (1982) showed that such tours can be obtained by appropriately generalizing the Pythagorean theorem, a construction that, upon reflection, almost parallels that of some space-filling curves.

The curious issue for the TSP is that although there is *some* tour T' that makes $\sum_{e \in T'} |e|^2$ particularly small, the Bern and Eppstein (1993) result tells us that a traveling salesman tour T minimizing $\sum_{e \in T} |e|$ need not do nearly so well. Because of the matroidal properties of the MST, these issues do not arise in its analysis; analyzing the optimal TSP is more difficult.

In the final section, we will comment further on this as well as problems concerning points in $[0,1]^d$ for dimension $d > 2$. In §2, we prove a technical result that is applied in §3 to prove our main results.

2. Edge lemmas. The second lemma of this section explicates a property of edges in a TSP tour that will be useful in the next section, where we prove our main results. Our first lemma gives a simple geometric bound concerning diagonals of quadrilaterals. In the statement of Lemma 1, the term "diagonal" is used to denote a segment connecting nonadjacent vertices of a quadrilateral, regardless of whether the quadrilateral is convex.

LEMMA 1. *Let* L_1 *and* L_2 *be two nonintersecting line segments satisfying* $r \leq |L_i| \leq \beta r$, *where* $\beta > 1$ *and* $r > 0$. *Suppose the midpoints of* L_1 *and* L_2 *are separated by distance* λ, *where* $\lambda \leq \min\{\frac{1}{2}|L_1|, \frac{1}{2}|L_2|\}$. *If the endpoints of* L_1 *and* L_2 *are joined to form a quadrilateral with sides* L_1, L_2, S_1, *and* S_2, *then* $|S_i| \leq \frac{1}{2}(\beta - 1)r + 3\lambda$ *for* $i = 1, 2$.

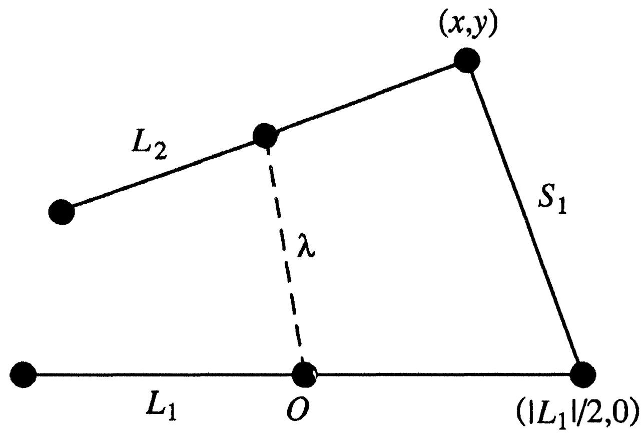

FIG. 1. *The lines L_1 and L_2 in the proof of Lemma* 1.

Proof. Without loss of generality, we can assume that L_1 is the longer of the two lines and that it is oriented along the x axis with its midpoint at the origin. We can also assume that L_2 lies entirely in the upper half plane. Let (x, y) denote the rightmost endpoint of L_2 and let S_1 be the line segment determined by (x, y) and $(|L_1|/2, 0)$ (see Fig. 1, which illustrates a convex quadrilateral). By the triangle inequality, the segment from the origin to (x, y) is at most $\frac{1}{2}|L_2| + \lambda$, so $x \leq \frac{1}{2}|L_2| + \lambda$.

We also claim that $\frac{1}{2}|L_2| - \lambda \leq x$. To prove the claim, consider the disk D of radius λ centered at the midpoint m of L_2. Since $\lambda \leq \frac{1}{2}|L_2|$, the point (x, y) must lie outside the interior of D. Since L_2 lies entirely in the upper half plane, the endpoints of L_2 must lie in the shaded regions in Fig. 2, with (x, y) constrained to lie in the first quadrant. Letting $(x,' 0)$ be the point where the x axis intersects the circle with center m passing through (x, y), it is clear from the figure that $x \geq x'$. However, the origin–x' segment is greater than or equal to $\frac{1}{2}|L_2| - \lambda$ since $\frac{1}{2}|L_2| - \lambda$ is the minimum distance from $(x,' 0)$ to D. This proves the claim and yields

$$\frac{1}{2}|L_2| - \lambda \leq x \leq \frac{1}{2}|L_2| + \lambda. \tag{2.1}$$

Since L_1 and L_2 do not intersect, $0 < y < 2\lambda$. Combining this with (2.1) gives us $|S_1| \leq \left|x - \frac{1}{2}|L_1|\right| + y \leq \frac{1}{2}(\beta - 1)r + 3\lambda$, as claimed. □

LEMMA 2. *Let $\{e_1, e_2, \ldots, e_n\}$ denote the edges of an optimal traveling salesman tour of $\{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}^2$. For each e_i satisfying $r \leq |e_i| \leq \beta r$, where $r > 0$, let D_i denote the disk of radius $\alpha|e_i|$ centered at the midpoint of e_i, where $\alpha = 1/13$ and $\beta = 3/2$. Then, for any three disks D_{i_1}, D_{i_2}, and D_{i_3}, the intersection $D_{i_1} \cap D_{i_2} \cap D_{i_3}$ is empty.*

Proof. Suppose at first that $D_{i_1} \cap D_{i_2} \cap D_{i_3} \neq \emptyset$ and that the edges e_{i_1}, e_{i_2}, and e_{i_3} share no common vertex. Without loss of generality, let $i_j = j$ for $j = 1, 2, 3$. We show that if D_1, D_2, and D_3 have a point in common, then it is possible to construct a shorter tour through $\{x_1, x_2, \ldots, x_n\}$. It is well known that edges of an optimal Euclidean traveling salesman tour cannot intersect. We can therefore assume that $e_1 = a_1b_1$, with midpoint m_1 and endpoint a_1 to the left and b_1 to the right, is oriented along the x axis. Similarly we can assume the midpoint m_2 of $e_2 = a_2b_2$ lies above e_1 and the midpoint m_3 of $e_3 = a_3b_3$ lies above e_2, as illustrated in Fig. 3.

Since the endpoints of the e_i, $\{a_1, a_2, a_3, b_1, b_2, b_3\}$, are distinct and are on the tour, there is a pair a_i, b_j with $i \neq j$ such that a_i and b_j are joined by a path that contains none of the

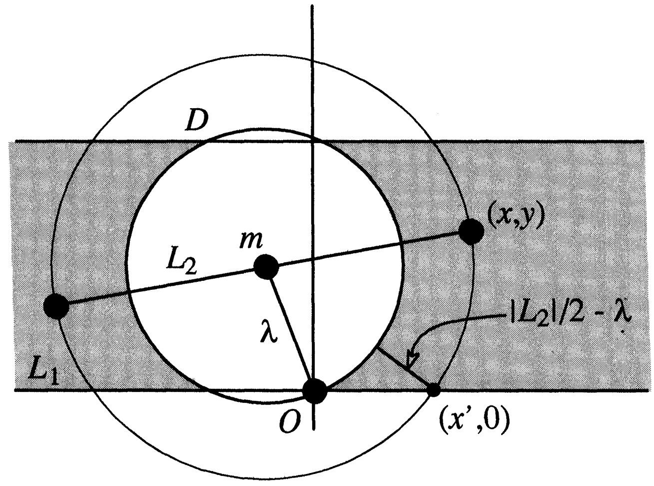

FIG. 2. *The disk D and its relation to x, L_2, and λ.*

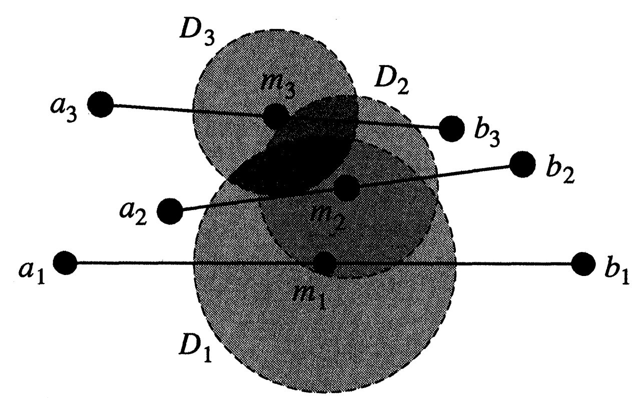

FIG. 3. *Three nonintersecting lines of a TSP tour and their D_i. Here, $\alpha = 1/2$ for visual clarity.*

edges e_1, e_2, and e_3. We now claim that we can construct a shorter tour by replacing edges e_i and e_j with edges a_ia_j and b_ib_j. This contradiction will establish the lemma.

For specificity, assume that $i = 2$ and $j = 3$, as shown in Fig. 4. We form a new path from a_2 to b_3 by deleting e_2 and e_3 and adding the edges a_2a_3 and b_2b_3. Since D_1, D_2, and D_3 have a point in common, the midpoints of e_2 and e_3 can be separated by at most the summed radii of D_2 and D_3, which is $\alpha|e_2| + \alpha|e_3|$. Setting $\lambda = \alpha(|e_2| + |e_3|)$ and recalling that $r \le |e_i| \le \beta r$, we note that $\lambda \le \alpha(|e_2| + \beta|e_2|) \le \frac{1}{2}|e_2|$; similarly, $\lambda \le \frac{1}{2}|e_3|$. In addition, we have $\lambda \le 2\alpha\beta r$. These facts allow us to apply Lemma 1 to estimate the net change Δ in the path length as

$$\begin{aligned}\Delta &= |a_2 - a_3| + |b_2 - b_3| - |e_2| - |e_3| \\ &\le 2\left[\frac{1}{2}(\beta - 1)r + 3\lambda\right] - 2r \\ &\le 2\left[\frac{1}{2}(\beta - 1)r + 6\alpha\beta r\right] - 2r \\ &= (\beta - 3 + 12\alpha\beta)r.\end{aligned}$$

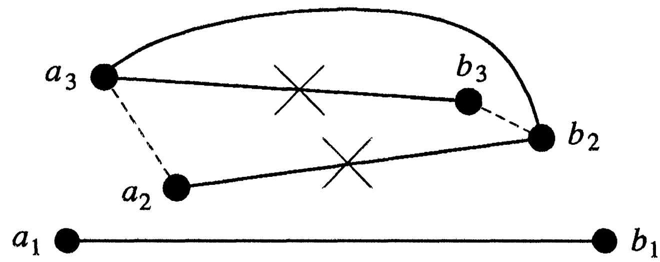

FIG. 4. *Rebuilding the a_2 to b_3 path when $i = 2$ and $j = 3$ in Lemma 2. The curved arc is a path, the $\times$'ed edges have been removed, and the dashed edges have been added.*

The choices $\beta = 3/2$ and $\alpha = 1/13$ guarantee that $\Delta < 0$.

For the case of $i = 1$ and $j = 3$, one obtains identical bounds on the change in the tour length when replacing e_i and e_j with $a_i a_j$ and $b_i b_j$. Without loss of generality, the $i = 2$, $j = 3$ and $i = 1$, $j = 3$ cases are the only cases that need to be considered.

To complete the proof, note that any vertex shared by any of e_{i_1}, e_{i_2}, and e_{i_3} can be replaced with two vertices that are viewed as being joined by an edge of length 0. The above analysis can then be applied as before without change to obtain a contradiction. $\square$

3. A priori edge-length bounds. We are now in position to prove our main results. Label the edges of an optimal tour T of $\{x_1, x_2, \dots, x_n\} \subset [0, 1]^2$ in order as $e_1, e_2, \dots, e_n$. We first construct disks D_i of radius $\alpha|e_i|$ and center at the midpoint of e_i for each $1 \le i \le n$, where $\alpha = 1/13$. Let $\psi_i(\cdot)$ denote the indicator function of D_i, i.e., for all $x \in \mathbb{R}^2$, $\psi_i(x) = 1$ if $x \in D_i$; otherwise $\psi_i(x) = 0$. Let A be the set of all i such that $r \le |e_i| \le \beta r$, where $\beta = 3/2$. We then claim that

$$\sum_{i \in A} \psi_i(x) \le 2\psi(x), \tag{3.1}$$

where $\psi(\cdot)$ is the indicator function of the square $[-1, 2]^2$.

To prove the claim, note that for $\beta = 3/2$ and $\alpha = 1/13$, Lemma 2 tells us that no three disks of A intersect. Hence, the point $x \in \mathbb{R}^2$ can belong to at most two disks associated with A. Furthermore, since any disk with center in $[0, 1]^2$ and radius bounded by $\alpha\beta r$ is contained in $[-\alpha\beta r, 1 + \alpha\beta r]^2 \subset [-1, 2]^2$, we need only concern ourselves with the square $x \in [-1, 2]^2$. This proves the claim.

If we now integrate (3.1) over x, we obtain a basic bound on a subset of the squared edge lengths of an optimal TSP tour:

$$\sum_{r \le |e_i| \le \beta r} |e_i|^2 \le c, \tag{3.2}$$

where $c = 18\alpha^{-2}\pi^{-1}$. The bound is then used as follows.

$$\begin{aligned} \sum_{i=1}^{n} |e_i|^2 &\le 1 + \sum_{n^{-1/2} \le |e_i| \le \sqrt{2}} |e_i|^2 \\ &\le 1 + \sum_{k=1}^{m} \sum_{\beta^{k-1} n^{-1/2} \le |e_i| \le \beta^k n^{-1/2}} |e_i|^2. \end{aligned} \tag{3.3}$$

where m is the least integer k such that $\beta^k n^{-1/2} \geq \sqrt{2}$. It suffices to take $m = \lceil \log_{3/2}(\sqrt{2n}) \rceil$; applying (3.2) to (3.3) yields the bound

$$\sum_{i=1}^{n} |e_i|^2 \leq c_1 \log n, \tag{3.4}$$

where c_1 is constant as required by Theorem 1. We remark that explicit constants have been given only to facilitate checking; there is little hope of obtaining the best possible bounds on c_1 and related values.

Returning to (3.1) and integrating, we see that since $|e_i| \geq r$ for all $i \in A$,

$$|A|\pi\alpha^2 r^2 \leq 18. \tag{3.5}$$

However, $|A| = \big|\{ i : r \leq |e_i| \leq \beta r \}\big|$, so for $c = 18/(\pi\alpha^2)$, we have

$$\big|\{ i : r \leq |e_i| \leq \beta r \}\big| \leq c r^{-2}. \tag{3.6}$$

We can now bound

$$\begin{aligned} \nu(n,t) &= \big|\{ i : |e_i| \geq t \}\big| \\ &\leq \sum_{k=0}^{m_t - 1} \big|\{ i : \beta^k t \leq |e_i| \leq \beta^{k+1} t \}\big|, \end{aligned} \tag{3.7}$$

where $m_t = \min_j\{\beta^j t \geq \sqrt{2}\}$. We then use (3.6) to obtain

$$\begin{aligned} \nu(n,t) &\leq c \sum_{k=0}^{m_t - 1} (\beta^k t)^{-2} \\ &\leq c t^{-2} \sum_{k=0}^{\infty} \beta^{-2k} \\ &= \frac{c}{1-\beta^{-2}} t^{-2}, \end{aligned} \tag{3.8}$$

which is Theorem 2, with $c_2 = c\beta^2/(\beta^2 - 1)$.

Theorem 3 now results from (3.8) by first noting that $n - \nu(n,x)$ is the number of edges in T of length less than x, then writing

$$\begin{aligned} \sum_{e_i \in E} |e_i| &= \sum_{\substack{e_i \in E \\ |e_i| < t}} |e_i| + \sum_{\substack{e_i \in E \\ |e_i| \geq t}} |e_i| \\ &\leq \sum_{\substack{e_i \in E \\ |e_i| < t}} |e_i| + \sum_{\substack{e_i \in T \\ |e_i| \geq t}} |e_i| \\ &\leq t|E| + \int_t^{\sqrt{2}} x \, d(n - \nu(n,x)) \\ &\leq t|E| - \int_t^{\sqrt{2}} x \, d\nu(n,x). \end{aligned} \tag{3.9}$$

Integrating the rightmost term of (3.9) by parts and then applying (3.8), we obtain

$$
\begin{aligned}
-\int_t^{\sqrt{2}} x\, d\nu(n,x) &= t\nu(n,t) + \int_t^{\sqrt{2}} \nu(n,x)\, dx \\
&\le \frac{c_2}{t} + \int_t^{\infty} \frac{c_2}{x^2}\, dx \\
&\le \frac{2c_2}{t}.
\end{aligned} \tag{3.10}
$$

Inserting (3.10) into (3.9) and setting $t = |E|^{-1/2}$ yields Theorem 3, with $c_3 = 1 + 2c_2$.

4. Concluding remarks. The preceding arguments can be generalized without difficulty to higher dimensions. The key idea is that in Lemma 2 we showed that if three of the D_i associated with edges of a TSP tour had a point in common, then we could find three edges e_1, e_2, and e_3 that were close together and nearly parallel.

We can obtain a proper analogue in dimensions $d > 2$ if we consider the possibility that a large number $N(d)$ of d-spheres $D_i = D(m_i, \alpha|e_i|) \subset \mathbb{R}^d$ intersect and exploit the fact that the surface of any sphere in $\mathbb{R}^d$ can be covered with a finite number $M(\epsilon)$ of spherical caps with polar angle ϵ. In summary, one can prove the following theorem.

THEOREM. *There exist positive constants c_d and c'_d such that for any traveling salesman tour T of $\{x_1, x_2, \ldots, x_n\} \subset [0, 1]^d$ and for all $n \geq 2$,*

$$\sum_{e \in T} |e|^d \leq c_d \log n, \tag{4.1}$$

and

$$\nu_d(t) = \left|\{ e \in T : |e| \geq t \}\right| \leq c'_d / t^d. \tag{4.2}$$

Furthermore, there exists a positive constant c''_d such that for any $E = \{e_{i_1}, e_{i_2}, \ldots, e_{i_k}\} \subseteq T$, we have

$$\sum_{i \in E} |e_i| \leq c''_d k^{(d-1)/d}. \tag{4.3}$$

Acknowledgment. We thank a kind and meticulous referee who corrected some errors and who provided useful suggestions for simplifying our proof of Lemma 2.

REFERENCES

[1] M. BERN AND D. EPPSTEIN (1993), *Worst-case bounds for subadditive geometric graphs*, Proceedings of the Ninth Annual Symposium on Computational Geometry, Association for Computing Machinery, pp. 183–188.

[2] L. FEW (1955), *The shortest path and the longest road through n points in a region*, Mathematika, 2, pp. 141–144.

[3] L. GODDYN (1990), *Quantizers and the worst-case Euclidean traveling salesman problem*, J. Combin. Theory, Ser. B, 50, pp. 65–81.

[4] H. J. KARLOFF (1989), *How long can a Euclidean traveling salesman tour be?*, SIAM J. Discrete Math., 2, pp. 91–99.

[5] D. J. NEUMANN (1982), *A Problem Seminar*, Springer-Verlag, New York, NY.

[6] J. M. STEELE (1990), *Probabilistic and worst-case analyses of classical problems of combinatorial optimization in Euclidean space*, Math. Oper. Res., 15, pp. 749–770.

[7] J. M. STEELE AND T. L. SNYDER (1989), *Worst-case growth rates of some problems from combinatorial optimization*, SIAM J. Comput., 18, pp. 278–287.

[8] S. VERBLUNSKY (1951), *On the shortest path through a number of points*, Proc. Amer. Math. Soc., 2, pp. 904–913.