Chapter 50

Equidistribution of Point Sets for the Traveling Salesman and Related Problems

Timothy Law Snyder* J. Michael Steele**

Abstract

Given a set S of n points in the unit square $[0,1]^2$, an optimal traveling salesman tour of S is a tour of S that is of minimum length. A *worst-case point set* for the Traveling Salesman Problem in the unit square is a point set $S^{(n)}$ whose optimal traveling salesman tour achieves the maximum possible length among all point sets $S \subset [0,1]^2$, where $|S| = n$. An open problem is to determine the structure of $S^{(n)}$. We show that for any rectangle R contained in $[0,1]^2$, the number of points in $S^{(n)} \cap R$ is asymptotic to n times the area of R. One corollary of this result is an $O(n \log n)$ approximation algorithm for the worst-case Euclidean TSP. Analogous results are proved for the minimum spanning tree, minimum-weight matching, and rectilinear Steiner minimum tree. These equidistribution theorems are the first results concerning the structure of worst-case point sets like $S^{(n)}$.

1. Introduction

This paper deals with worst-case arrangements of points for problems in combinatorial optimization; it is shown that these point sets are equidistributed. For specificity, we concentrate on the Traveling Salesman Problem.

Given a set of points S in the unit square $[0,1]^2$, an *optimal traveling salesman tour* of S is a tour consisting of edges from the complete graph on S and having total length $\min_T \{ \sum_{e \in T} |e| : T \text{ is a tour of S} \}$. Here, $|e|$ denotes the Euclidean length of the edge e. We use TSP(S) to denote the set of edges of an optimal traveling salesman tour of S and $|\text{TSP}(S)|$ to denote the sum of the Euclidean lengths of the edges in TSP(S).

A *worst-case* optimal traveling salesman tour is a tour of total length

$$\rho_{\text{TSP}}(n) = \max_{\substack{S \subset [0,1]^2 \\ |S|=n}} |\text{TSP}(S)|. \tag{1.1}$$

In words, $\rho_{\text{TSP}}(n)$ is the maximum length, over all point sets in $[0,1]^2$ of size n, that an optimal traveling salesman tour can attain. Such a tour attaining this length is called a *worst-case TSP tour* and its associated point set a *worst-case TSP point set*.

The first works on the sequence $\rho_{\text{TSP}}(n)$ were the lower bounds of Fejes-Tóth (1940) and the upper bounds of Verblunsky (1951). Successive improvements to these bounds and their higher-dimensional analogues appeared in Few (1955), Supowit, Reingold, and Plaisted (1983), Moran (1984), Goldstein and Reingold (1988), Karloff (1989), and Goddyn (1990). One result we use later is that

$$\lim_{n \to \infty} \rho_{\text{TSP}}(n)/n^{1/2} = \beta_{\text{TSP}}, \tag{1.2}$$

where $\beta_{\text{TSP}} > 0$ is a constant (Steele and Snyder (1989)).

All these results deal with the worst-case *length* $\rho_{\text{TSP}}(n)$. There are no results, however, concerning the locations of the points that give rise to a worst-case tour. Let $S^{(n)}$ be a worst-case TSP point set. An engaging open problem is to determine the structure of

* Department of Computer Science, Georgetown University, Washington, DC 20057. Research supported in part by Georgetown University 1991 Summer Research Award and Georgetown University 1992 Junior Faculty Research Fellowship.

** Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104. Research supported in part by the following grants: NSF DMS92-11634, NSA MDA904-89-H-2034, AFOSR-89-0301, and ARO DAAL03-91-G-0110.

$S^{(n)}$. It is generally believed that $S^{(n)}$ is asymptotically a lattice; for example, Supowit, Reingold, and Plaisted (1983) conjectured a hexagonal grid for $S^{(n)}$.

Our main result is that any sequence of worst-case TSP point sets is asymptotically equidistributed:

Theorem 1. *If $\{S^{(n)} : 1 \le n < \infty\}$ is a sequence of worst-case TSP point sets, then, for any rectangle $R \subset [0,1]^2$,*

$$\lim_{n\to\infty} \frac{1}{n} \left|S^{(n)} \cap R\right| = Area(R). \tag{1.3}$$

A corollary to Theorem 1 is that any worst-case Euclidean traveling salesman problem can be solved within a factor of $1+\epsilon$, for any $\epsilon > 0$, in $O(n \log n)$ time using the algorithm of Karp (1976). Even though Karp's algorithm is for points drawn uniformly from the unit square, we discuss in the concluding section how Theorem 1 and the limit (1.2) guarantee this performance for worst-case point sets.

The proof of Theorem 1 rests on the limit (1.2) and a characterization of $|S^{(n)} \cap R|$ in probabilistic terms. Even though $S^{(n)}$ is deterministic, a simple probabilistic identity plays a useful role. We show in a later section that the method provides analogous theorems for other problems, including the minimum spanning tree, minimum weight matching, and rectilinear Steiner tree.

In Section 2, we begin with two observations that simplify the proof of Theorem 1 in Section 3. Section 4 contains results and proofs for problems other than the Traveling Salesman Problem, and Section 5 concludes with speculations on extensions and applications of our results, including algorithmic ones.

2. Deviations from the Mean Number of Points per Cell

To prove Theorem 1, it suffices to show that if we divide the unit square $[0,1]^2$ into m^2 equally-sized parallel subsquares or *cells*, each having side length $1/m$, and label the cells Q_i, where $1 \le i \le m^2$, then, for all $1 \le i \le m^2$,

$$\lim_{n\to\infty} \frac{|S^{(n)} \cap Q_i|}{n} = \frac{1}{m^2}. \tag{2.1}$$

Let $s(n,i) = |S^{(n)} \cap Q_i|$; in words, $s(n,i)$ is the number of points in the ith cell of a worst-case TSP point set. Points that lie on the boundaries of the Q_i can be arbitrarily assigned. Furthermore, let

$$V_n = \frac{1}{m^2} \sum_{i=1}^{m^2} \Big(\{s(n,i)\}^{1/2} - \frac{1}{m^2} \sum_{j=1}^{m^2} \{s(n,j)\}^{1/2} \Big)^2. \tag{2.2}$$

One should note that if X is a random variable taking on the value $\{s(n,i)\}^{1/2}$ with probability $1/m^2$, then V_n is the variance of X. Our first lemma estimates the expected value of X.

Lemma 1. *As $n \to \infty$,*

$$\frac{1}{m^2} \sum_{j=1}^{m^2} \{s(n,j)\}^{1/2} = \frac{n^{1/2}}{m} + o(n^{1/2}). \tag{2.3}$$

Proof.

First write (1.2) as

$$\rho_{\text{TSP}}(n) = \beta n^{1/2} + r(n), \text{ where } r(n) = o(n^{1/2}); \tag{2.4}$$

for convenience, we have dropped the "TSP" from β_{TSP}.

Let W denote a closed walk on $S^{(n)} = \{x_1, x_2, \ldots, x_n\}$, i.e, W is a sequence of edges $(x_{i_1}, x_{i_2}), (x_{i_2}, x_{i_3}), \ldots, (x_{i_{k-1}}, x_{i_k}), (x_{i_k}, x_{i_1})$ that visits each point of $S^{(n)}$ at least once and begins and ends at the same point. Since W is feasible for the traveling salesman problem on $S^{(n)}$ even if W visits some points more than once and traverses some edges more than once, $\sum_{e\in W} |e| \le |\text{TSP}(S^{(n)})|$.

We now construct a closed walk W on $S^{(n)}$ as follows. Label the cells Q_i with the numbers 1 through m^2 by traversing the cells row by row, proceeding left to right in the first row, right to left in the second, and so on, alternating the direction of traversal in each row until all cells have been labeled. Next, construct in each cell Q_i an optimal traveling salesman tour of the points of $S^{(n)}$ that lie in Q_i. These disjoint, within-cell tours are then connected by placing an edge e_i between cell number i and cell number $i+1$, for $1 \le i \le m^2 - 1$. A closed walk W that traverses each of the within-cell tours once and each of the e_i twice can then be obtained.

To assess the length of W, let Q_i and Q_j be adjacent cells. Since any pair of points in $Q_i \cup Q_j$ can be connected by a line segment of length at most $5^{1/2}/m$,

$$\begin{aligned}\rho_{\mathrm{TSP}}(n) &= |\mathrm{TSP}(S^{(n)})| \\ &\le \sum_{e \in W} |e| \\ &= \sum_{i=1}^{m^2} |\mathrm{TSP}(S^{(n)} \cap Q_i)| + 2 \sum_{i=1}^{m^2-1} |e_i| \qquad (2.5) \\ &\le \sum_{i=1}^{m^2} |\mathrm{TSP}(S^{(n)} \cap Q_i)| + 2 \cdot 5^{1/2} m.\end{aligned}$$

We now use (2.4) in (2.5) along with the fact that $|\mathrm{TSP}(S^{(n)} \cap Q_i)|$ is at most $\rho_{\mathrm{TSP}}(s(n,i))$ scaled by the cell size $1/m$ to get

$$\begin{aligned}\rho_{\mathrm{TSP}}(n) &= \beta n^{1/2} + r(n) \\ &\le \frac{1}{m} \sum_{i=1}^{m^2} \rho_{\mathrm{TSP}}(s(n,i)) + 2 \cdot 5^{1/2} m \\ &\le \frac{1}{m} \sum_{i=1}^{m^2} \beta \{s(n,i)\}^{1/2} \qquad (2.6) \\ &\quad + \frac{1}{m} \sum_{i=1}^{m^2} r(s(n,i)) + 2 \cdot 5^{1/2} m,\end{aligned}$$

where, for all $1 \le i \le m^2$, the value $r(s(n,i)) = o(\{s(n,i)\}^{1/2}) = o(n^{1/2})$. Since m is fixed, we cancel β to find

$$\sum_{i=1}^{m^2} \{s(n,i)\}^{1/2} \ge m n^{1/2} + h(n), \qquad (2.7)$$

where $h(n) = o(n^{1/2})$. This proves half of the lemma. To obtain the other half, we note from the Schwarz inequality that $\sum_{i=1}^{m^2} \{s(n,i)\}^{1/2} \le m n^{1/2}$ since $\sum_{i=1}^{m^2} s(n,i) = n$.

□

We now use Lemma 1 and the characterization of V_n as a variance to show that V_n/n goes to zero.

Lemma 2. *For V_n as defined in Equation (2.2), we have*

$$V_n = o(n) \qquad (2.8)$$

as $n \to \infty$.

Proof.

First note that V_n can be written as

$$V_n = \frac{1}{m^2} \sum_{i=1}^{m^2} s(n,i) - \Big(\frac{1}{m^2} \sum_{j=1}^{m^2} \{s(n,j)\}^{1/2} \Big)^2, \qquad (2.9)$$

using the remark following definition (2.2) and the identity $\mathrm{Var}(X) = \mathrm{E}X^2 - (\mathrm{E}X)^2$. Since $\sum_{i=1}^{m^2} s(n,i) = n$, applying Lemma 1 to the second term of (2.9) gives

$$V_n = \frac{n}{m^2} - \Big(\frac{n^{1/2}}{m} + o(n^{1/2}) \Big)^2. \qquad (2.10)$$

On expanding, we see that $V_n = o(n)$.

□

3. Equidistribution in the Worst-Case TSP

To prove Theorem 1, recall that it suffices to show that $\lim_{n \to \infty} |S^{(n)} \cap Q_i|/n = 1/m^2$, for all $1 \le i \le m^2$.

¿From the definition (2.2) of V_n,we have that for all i satisfying $1 \leq i \leq m^2$,

$$\left|\{s(n,i)\}^{1/2} - \frac{1}{m^2}\sum_{j=1}^{m^2}\{s(n,j)\}^{1/2}\right| \leq m V_n^{1/2}. \quad (3.1)$$

Applying first Lemma 2 then Lemma 1 to (3.1) gives

$$\begin{aligned}\{s(n,i)\}^{1/2} &= \frac{1}{m^2}\sum_{j=1}^{m^2}\{s(n,j)\}^{1/2} + o(n^{1/2}) \\ &= \frac{n^{1/2}}{m} + o(n^{1/2}).\end{aligned} \quad (3.2)$$

Squaring (3.2) yields Theorem 1.

□

4. Equidistribution in the MST, Matching, and Steiner Problems

The method just used for the TSP can be applied to the minimum spanning tree, the minimum-length matching, and the rectilinear minimum Steiner tree. If $\mathrm{L} = \mathrm{L}(S)$ denotes the length associated with any of these, then we can define $\rho_{\mathrm{L}}(n) = \sup_{S:|S|=n} \mathrm{L}(S)$ and let $S_{\mathrm{L}}^{(n)}$ be such that $\mathrm{L}(S_{\mathrm{L}}^{(n)}) = \rho_{\mathrm{TSP}}(n)$. To show that $S_{\mathrm{L}}^{(n)}$ is asymptotically equidistributed boils down to checking that L satisfies two conditions:

1. $\rho_{\mathrm{L}}(n) = \beta_{\mathrm{L}} n^{1/2} + o(n^{1/2})$, where $\beta_{\mathrm{L}} > 0$ is constant;
2. $\rho_{\mathrm{L}}(n) \leq m^{-1}\sum_{i=1}^{m^2} \rho_{\mathrm{L}}(s_{\mathrm{L}}(n,i)) + o(n^{1/2})$, where $s_{\mathrm{L}}(n,i)$ is the number of points of $S_{\mathrm{L}}^{(n)}$ contained in the cell Q_i.

Condition 1 has been proved for the minimum spanning tree, minimum matching, and rectilinear Steiner tree problems (cf., Steele and Snyder (1989) and Snyder (1992)), and Condition 2 can be verified for these problems by the method used in the proof of Lemma 1.

For example, if $\mathrm{L}(S) = \mathrm{MST}(S)$ denotes the total length of a minimum spanning tree of S, we first form minimum spanning trees $\mathrm{MST}(S_{\mathrm{MST}}^{(n)} \cap Q_i)$ on the points of $S_{\mathrm{MST}}^{(n)}$ in the cells Q_i, where $1 \leq i \leq m^2$. The trees within cells can then be interconnected at total cost $O(m) = o(n^{1/2})$ by adding $m^2 - 1$ edges, each costing no more than $5^{1/2}/m$. This forms a heuristic tree on $S_{\mathrm{MST}}^{(n)}$. Since the lengths $|\mathrm{MST}(S_{\mathrm{MST}}^{(n)} \cap Q_i)|$ are no greater than the worst-case (within-cell) lengths $m^{-1}\rho_{\mathrm{MST}}(s_{\mathrm{MST}}(n,i))$, Condition 2 follows.

Checking these conditions for each of the problems yields the following.

Theorem 2. *If* $\{S_{\mathrm{L}}^{(n)} : 1 \leq n < \infty\}$ *is a sequence of worst-case point sets for the function* L, *where* L *is the minimum spanning tree, the minimum matching, or the rectilinear minimum Steiner tree, then, for any rectangle* $R \subset [0,1]^2$,

$$\lim_{n\to\infty} \frac{1}{n}\left|S_{\mathrm{L}}^{(n)} \cap R\right| = Area(R). \quad (4.1)$$

5. Concluding Remarks

The asymptotic equidistribution of worst-case point sets for the problems we have considered offers some support to the conjectures of Supowit, Reingold, and Plaisted (1983) that worst-case point sets are approximated by lattices as $n \to \infty$. It is still a major open problem to resolve these conjectures.

The results also prove that the worst-case traveling salesman problem on $S^{(n)}$ is amenable to Karp's probabilistic algorithm for the TSP, even though $S^{(n)}$ is entirely deterministic. For n points selected uniformly from $[0,1]^2$, Karp's algorithm runs in $O(n \log n)$ time with probability one, and, for all $\epsilon > 0$, the tour formed by the algorithm is within a factor of $1 + \epsilon$ of optimum with probability one, as $n \to \infty$.

Using Karp's algorithm, given $\epsilon > 0$, we can guarantee the construction of a tour T of $S^{(n)}$ in $O(n \log n)$ time such that the total length of T is at most $1 + \epsilon$

times the optimal length $\rho_{\text{TSP}}(n)$, as $n \to \infty$. These results are entirely deterministic, and they can be proved by observing that Karp's analysis uses a probabilistic counterpart to the limit theorem expressed by (1.2) along with asymptotic equidistribution of point sets uniformly selected from the unit square. Though the limit (1.2) for the worst-case length of the TSP has been known for several years, an equidistribution result for $S^{(n)}$ was required in order to guarantee both the time and performance bounds for Karp's algorithm applied to $S^{(n)}$. Until there are definitive tests for worst-case point sets, however, this result is only of theoretical interest; whether it can be used for point sets other than $S^{(n)}$ is an open problem.

Several other open problems are motivated by our results. For any finite set of points S, the *Steiner problem* is to find a minimum-length tree $T = (V, E)$ such that V contains S. The added points $Q = V - S$ are the *Steiner points* of T. Though any metric can be used to assign costs to edges, two metrics of interest are the Euclidean and rectilinear (L_1).

Let $S^{(n)}$ be a worst-case point set for the rectilinear Steiner problem, and let T be a rectilinear minimum Steiner tree of $S^{(n)}$. The asymptotic equidistribution in Theorem 2 applies only to $S^{(n)}$, and not to the set Q of Steiner points of T; we conjecture that Q is asymptotically equidistributed, as well.

For the Euclidean Steiner problem, the limit result for Condition 1 in Section 4 has yet to be established. We believe such a result holds, and it would imply that a worst-case point set for the Euclidean Steiner problem is asymptotically equidistributed. It is also likely that the Steiner points in the Euclidean case are asymptotically equidistributed.

A final open problem concerns the greedy matching. Though Condition 1 in Section 4 holds for this problem, the methods used here to verify Condition 2 fail since a greedy matching is not a matching of minimum length. Hence, this problem also remains open.

References

Fejes-Tóth, L.T. (1940). "Über ein geometrische Satz," *Math. Zeit.* **46**, 299–327.

Few, L. (1955). "The shortest path and the shortest road through n points in a region," *Mathematika* **2**, 141–144.

Goddyn, L. (1990). "Quantizers and the worst-case Euclidean traveling salesman problem," *J. Comb. Theory, Series B* **50**, 65–81.

Karloff, H.J. (1989). "How long can a Euclidean traveling salesman tour be?" *SIAM J. Disc. Math.* **2**, 91–99.

Karp, R.M. (1976). "The probabilistic analysis of some combinatorial search algorithms," in *Algorithms and Complexity: New Directions and Recent Results*, J.F. Traub, ed., Academic Press, New York, 1–19.

Moran, S. (1984). "On the length of optimal TSP circuits in sets of bounded diameter," *J. Comb. Theory*, Series B **37**, 113–141.

Snyder, T.L. (1992). "On minimal rectilinear Steiner trees in all dimensions," to appear in *Discrete & Comp. Geom.*; also in *Proceedings of the Sixth Annual Symposium on Computational Geometry (1990)*, Association for Computing Machinery.

Steele, J.M. and Snyder, T.L. (1989). "Worst-case growth rates of some classical problems of com-