RANDOM MINIMAL TREES*

E. N. GILBERT†

1. Introduction. Let $P_1, P_2, \ldots, P_n$ be $n$ points in the plane which must be joined by means of lines to form a connected linear graph (or network). For example the $P_i$ might be $n$ terminals which must be connected together electrically (see [5]). Alternatively the $P_i$ may be stations in a communication network and the lines two-way channels (see [6]). Only $n - 1$ lines need be drawn in order to obtain a tree network which contains paths between each pair $P_i, P_j$ of points. According to Cayley's formula (see [7, p. 28]) there are $n^{n-2}$ trees which can be drawn to interconnect $n$ points.

Suppose $c(P_i, P_j)$ is the cost of providing a line $P_i P_j$ in the tree. The cost of the tree is then defined to be the sum of the costs $c(P_i, P_j)$ of its lines. Usually the Euclidean distance from $P$ to $Q$ is used for $c(P, Q)$. All cost functions to be considered here will be distances, although not necessarily Euclidean. For example, in a communication network within a city one might let $c(P, Q)$ be the distance from $P$ to $Q$ measured along city streets via the shortest route.

A minimal tree for $P_1, \ldots, P_n$ is a tree which joins these points for the lowest possible cost. Loberman and Weinberger [5], Prim [6], and Kruskal [4] gave simple algorithms for constructing minimal trees. Note that $P_1, P_2, \ldots, P_n$ are the only vertices allowed for the trees in this paper. If one were free to use additional points as vertices one could in general connect $P_1, \ldots, P_n$ together with even lower cost (take $P_1, P_2, P_3$ as the vertices of an equilateral triangle, the cost as Euclidean distance, and consider allowing the centroid of the triangle as an extra vertex).

Suppose that $P_1, \ldots, P_n$ are chosen independently and at random with constant probability density $A^{-1}$ from a plane bounded region of area $A$. Then the cost of the minimal tree is a random variable with a mean $L(n)$ which will be estimated here. Verblunsky [8] and Beardwood, Halton, and Hammersley [1] studied a similar problem for minimal traveling salesman paths. For large $n$ the expected (Euclidean) length of the shortest traveling salesman path through $n$ points is asymptotic to const. $\times (An)^{1/2}$. The constant lies between .625 and .922; a Monte Carlo experiment estimated the value at .75. In [1] the authors remark that a similar asymptotic result holds for the minimal tree but with some other constant. One problem will be to estimate this constant. A theoretical argument will show

* Received by the editors April 3, 1964, and in revised form June 19, 1964.
† Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey.
that the constant is less than $2^{-1/2}$. In §8 experimental evidence suggests .68 as a good approximation.

The bound $2^{-1/2}$ is obtained by means of an exact analysis of a (not quite minimal) tree which is called "exodic" because certain of its paths radiate outward from a central point. An exact formula in §4 gives the mean length of the exodic tree connecting $n$ random points in the unit circle $\| P \| \leq 1$. The cases of Euclidean metric and Manhattan metric are considered in detail in §§6, 7. In a wide class of metrics $\| P \|$, the mean length of a random exodic tree connecting points produced by a Poisson process of mean $N$ in the unit circle $\| P \| \leq 1$ becomes asymptotic to $(\frac{1}{2} \pi N)^{1/2}$ for large $N$ (§5).

2. The unit circle. Let $x_i$, $y_i$ be Cartesian coordinates of $P_i$. Let vector addition and multiplication by scalars be defined in the usual way

$$P_i + P_j = (x_i + x_j, y_i + y_j), \quad \mu P_i = (\mu x_i, \mu y_i).$$

The cost function $c(P, Q)$ will always be assumed to have the form $\| P - Q \|$ where $\| \cdot \|$ is a norm satisfying $\| \mu P \| = |\mu| \| P \|$. Some norms of particular interest are

- **Euclidean norm** $\| P \| = (x^2 + y^2)^{1/2}$,
- **Manhattan norm** $\| P \| = |x| + |y|$,
- **Max norm** $\| P \| = \max\{|x|, |y|\}$.

Manhattan distance is an appropriate norm for networks which follow city streets.

The **unit circle** is the set of all points $P$ with $\| P \| \leq 1$. The unit circle is a convex figure (see [3, p. 74]). For example if $\| P \|$ denotes Manhattan distance the unit circle is a square with corners on the $x$ and $y$ axes and with sides of (Euclidean) length $2^{1/2}$.

This paper will be concerned mostly with trees in the unit circle. Results about the unit circle can supply asymptotic information about trees in regions of other shape. In the case of Euclidean metric Beardwood, Halton, and Hammersley [1] state that their analysis of the traveling salesman problem applies, with minor modifications, to the minimal tree problem. Then the expected length of the minimal tree joining $n$ random points in a region of area $A$ is asymptotically const. $\times (An)^{1/2}$ for large $n$. The constant of proportionality is the same for all regions and so can be estimated by studying the unit circle. Although there has been no proof of a similar asymptotic formula for other metrics, I believe such formulas hold very generally. In connection with traveling salesman paths on the earth's sur-
Fig. 1. Exodic tree for 31 points (Euclidean metric)

face the authors of [1] state that their results could be derived for distances measured along the surface of a spheroid.

3. Exodic trees. Let $n$ points in the plane be given. The following construction connects these points by a tree (not necessarily minimal). Call the points $P_1, P_2, \ldots, P_n$, choosing the subscripts to order the points according to their distances from the origin:

$$ ||P_1|| \leq ||P_2|| \leq \cdots \leq ||P_n||. $$

For $i = 2, 3, \ldots, n$ connect $P_i$ to a point $P_j$ chosen from $P_1, \ldots, P_{i-1}$ to minimize $||P_i - P_j||$. That these $n - 1$ connections form a tree follows easily by induction on $i$. If the connections made at $P_2, \ldots, P_{i-1}$ form a tree, the additional connection at $P_i$ creates paths from $P_i$ to $P_1, \ldots, P_{i-1}$ but completes no closed cycles.

Along a path from the first point $P_1$ to any other point $P_i$ the intermediate points $P_j$ occur with monotone increasing "distance" $||P_j||$ from the origin. For this reason the tree just constructed will be called an exodic tree.

The cost of an exodic tree is

$$ \sum_{i=2}^{n} \min_{j<i} ||P_i - P_j||, $$
a formula which is simpler, for analytical purposes, than the cost of the minimal tree. Of all trees in which paths from $P_1$ to $P_4$ proceed through points $P_j$ of increasing $\| P_j \|$ the exodic tree has least cost. Thus there is reason to hope that the cost of the exodic tree is only slightly more than the cost of the minimal tree.

Fig. 1 shows the exodic tree for 31 points placed at random in the unit Euclidean circle. This exodic tree has cost 7.36 which is slightly more than the minimal cost 6.95. The exodic tree and the minimal tree agree except for 6 connections.

Prim [6] constructs a minimal tree in great circle metric which connects Washington, D. C., with the 48 (in 1957) state capitals. When the same cities are connected by an exodic tree centered on Topeka, Kansas, 39 of the 48 lines are the same as in the minimal tree.

4. Random trees. Let $L(n)$ denote the expected cost of the exodic tree connecting $n$ points placed at random in the unit circle $\| P \| \leq 1$. To derive an analytical expression for $L(n)$ consider first an exodic tree connecting a (random) number of points with mean $N$ placed in the unit circle by a Poisson process. Let $S(N)$ be the expected length of this second random exodic tree. $S(N)$ and $L(n)$ are related by

$$S(N) = \sum_{n=2}^{\infty} e^{-N} \frac{N^n}{n!} L(n).$$

$S(N)$ will be found first; then $L(n)$ may be obtained as the coefficient of $N^n$ in the power series for $n! e^N S(N)$.

Let $C$ denote the area of the unit circle. The circle $\| P \| \leq R$ of radius $R$ has area $\pi R^2$ and contains an expected number $NR^2$ of points of the Poisson pattern.

If $Q$ is one of the Poisson pattern points, the rule for constructing the exodic tree connects $Q$ to its nearest neighbor (if any) inside a circle of radius $\| Q \|$. This connection will exist and have length less than $r$ if and only if another pattern point $P$ lies in the intersection $K(Q, r)$ of the circles $\| P \| \leq \| Q \|$ and $\| P - Q \| \leq r$.

Let $CA(Q, r)$ denote the area of $K(Q, r)$ so that $NA(Q, r)$ is the expected number of Poisson pattern points in $K(Q, r)$. Consider two events $E_1, E_2$ defined by:

$$E_1 : \\text{no Poisson pattern points } P \text{ have } \| P \| \leq \| Q \|;$$

$$E_2 : \\text{no Poisson pattern points } P \text{ lie in } K(Q, r).$$

The probabilities of these events are

$$P(E_1) = \exp\{-N\| Q \|^2\}, \quad P(E_2) = \exp\{-NA(Q, r)\}.$$  

If $E_1' E_2$ occurs, the connection at $Q$ exists and has length longer than $r$.  

Then the cumulative distribution function for the length of the connection at \( Q \) (counting no connection as length zero) is

\[
p(r) = 1 - P(E_2E_1') = 1 - P(E_2) + P(E_1).
\]

The last step follows because \( E_1 \) implies \( E_2 \).

The expected length of the connection at \( Q \) is

\[
f(Q) = \int_0^{2!Q_1} r \, dp(r) = 2 \| Q \| - \int_0^{2!Q_1} p(r) \, dr
\]

\[
(1) = \int_0^{2!Q_1} (P(E_2) - P(E_1)) \, dr
\]

\[
= \int_0^{2!Q_1} \exp \{-NA(Q, r)\} \, dr - 2 \| Q \| \exp \{-N \| Q \|^2\}.
\]

Since an element of area \( dx_1 \, dx_2 \) has probability \( (N/C)dx_1 \, dx_2 \) of containing a pattern point \( Q \), this element contributes an expected length \( (N/C)f(Q)dx_1 \, dx_2 \) to the exodic tree. The expected total length is

\[
S(N) = \frac{N}{C} \int_{1Q_1 \leq 1} f(Q) \, dx_1 \, dx_2.
\]

After expanding \( S(N) \exp(N) \) into a power series and using (1) and (2) one finds

\[
L(n) = \frac{n}{C} \int_{1Q_1 \leq 1} \left\{ \int_0^{2!Q_1} [1 - A(Q, r)]^{n-1} \, dr
\]

\[
- 2 \| Q \| (1 - \| Q \|^2)^{n-1} \right\} \, dx_1 \, dx_2.
\]

In §§6, 7 the integrals in (3) will be evaluated in detail for the Euclidean and Manhattan norms.

Formulas (1), (2), (3) generalize easily to exodic trees in a \( D \)-dimensional unit sphere. \( C \) becomes the volume of the unit sphere. \( K(Q, r) \) is defined by the same inequalities as before and has \( D \)-dimensional volume \( CA(Q, r) \). A \( D \)-fold integral over the unit sphere replaces the double integral in (2).

Formulas (1), (2), (3) also apply to certain other cost functions beside norms. The triangle inequality \( \| P + Q \| \leq \| P \| + \| Q \| \) was never used in deriving (1), (2), (3). Thus these formulas hold for cost functions which satisfy the other axioms of a norm but not the triangle inequality. In that case the unit circle \( \| P \| \leq 1 \) is not convex. In the next section the triangle inequality will be used and these more general cost functions are prohibited.
5. Bounds for exodic trees. The results in [1] suggest that $S(N)$ should grow like a constant multiple of $N^{1/2}$ for large $N$. This section will derive upper and lower bounds on $S(N)$ to show that the constant multiplier is $(\pi/2)^{1/2}$ if the unit circle has continuously turning tangent. It is curious that this multiplier is the same for all such norms $\| Q \|$.

The lower bound is the simpler one.

**Theorem 1.** For all choices of norm $\| Q \|$, 

$$S(N) \geq N^3 \{1 - \beta N^{-1}\} \int_0^{2\sqrt{N}} e^{-1/2} \, dt,$$

where $\beta = \frac{1}{2} + \sqrt{2} = 1.664$.

**Proof.** The theorem will follow from a bound

(4) \[ A(Q, r) \leq \frac{1}{2} r^2. \]

To prove (4) recall that $CA(Q, r)$ is the area of that part $K(Q, r)$ of the circle $\| P - Q \| \leq r$ (of area $Cr^2$) contained within a second circle $\| P \| \leq \| Q \|$ through $Q$. Since circles are convex and have central symmetry, the intersection $K(Q, r)$ contains at most half the area of $\| P - Q \| \leq r$; then (4) follows.

On the circle $\| Q \| = R$, (4) and (1) combine to give

(5) \[ f(Q) \geq \int_0^{2R} \exp \left\{ -\frac{1}{2} N r^2 \right\} \, dr - 2R \exp \left\{ -NR^2 \right\}. \]

Let $\varphi(R)$ denote the function on the right-hand side of (5). A bound on $S(N)$ will follow by substituting $\varphi(\| Q \|)$ for $f(Q)$ in (2). Since the integrand then depends only on $R = \| Q \|$, a convenient form for the integral is

$$S(N) \geq \frac{N}{C} \int_0^{1} \varphi(R) \, d(CR^2) = 2N \int_0^{1} \varphi(R)R \, dR.$$

After an interchange of order of integration,

$$S(N) \geq N \int_0^{2} \exp \left\{ -\frac{1}{2} N t^2 \right\} \int_0^{1} 2R \, dR \, dt - 4N \int_0^{1} R^2 \exp \{ -NR^2 \} \, dR$$

$$\geq N^{1/2} \int_0^{2N^{1/2}} \exp \left\{ -\frac{1}{2} t^2 \right\} \, dt - \beta N^{-1/2} \int_0^{2N^{1/2}} t^2 \exp \left\{ -\frac{1}{2} t^2 \right\} \, dt$$

$$= \{N^{1/2} - \beta N^{-1/2}\} \int_0^{2N^{1/2}} \exp \left( -\frac{1}{2} t^2 \right) \, dt + 2\beta \exp (-2N),$$

which completes the proof.

The asymptotic series for the error function shows that the lower bound in Theorem 1 is

$$\left(\frac{\pi N}{2}\right)^{1/2} \{1 - \beta N^{-1}\} + O(\exp (-2N)) \text{ as } N \to \infty.$$
An exactly similar lower bound for the mean length of a random exodic tree in a $D$-dimensional unit sphere has the asymptotic form

$$S(N) \geq N^{(D-1)/D}2^{1/D}1(1 + D^{-1})[1 + o(1)].$$

The upper bound is less precise and requires that the unit circle satisfy a smoothness condition.

**THEOREM 2.** If the unit circle has a continuously turning tangent at every point, then as $N \to \infty$;

$$\lim \sup N^{-1/2}S(N) \leq (\frac{1}{2}\pi)^{1/2}.$$

**Proof.** The proof depends on finding a lower bound on $A(Q, r)$ which approximates $\frac{1}{2}r^2$ for small $r$. First note that

$$A(\mu Q, \mu r) = \mu^2 A(Q, r)$$

for any positive $\mu$; this follows from the property $\|\mu P\| = \mu\|P\|$ of the norm. In particular take $\mu = 1/\|Q\|$ to get

$$A(Q, r) = \|Q\|^2 A(q, u),$$

where $q = Q/\|Q\|$ (a point on the unit circle) and $u = r/\|Q\|$.

For fixed $q$, $A(q, u)$ is a monotone increasing function of $u$ with the behavior $A(q, u) = \{\frac{1}{2} + o(1)\}u^2$ near $u = 0$. Then for every number $b < \frac{1}{2}$ there is a lower bound of the form

$$A(q, u) \geq \begin{cases} bu^2 & \text{if } 0 \leq u \leq U, \\ bU^2 & \text{if } U \leq u, \end{cases}$$

with $0 < U \leq 2$. Since the unit circle has continuously turning tangent at every point, $U$ can be chosen small enough so that (6) holds at every point $q$ on the unit circle. Then

$$A(Q, r) \geq \begin{cases} br^2 & \text{if } 0 \leq r \leq U\|Q\|, \\ bU^2\|Q\|^2 & \text{if } U\|Q\| \leq r. \end{cases}$$

Now substitute (7) for $A(Q, r)$ in (2) to get an upper bound $\psi(\|Q\|)$ on $f(Q)$. As in Theorem 1, a bound

$$S(N) \leq 2N \int_0^1 \psi(R)R \, dR$$

follows. The integration proceeds as in Theorem 1 with the final result

$$S(N) \leq \frac{1}{2} \left( \frac{N\pi}{b} \right)^{1/2} + O(1).$$

Then $\lim \sup N^{-1/2}S(N) \leq \frac{1}{2}(\pi/b)^{1/2}$ for all $b \leq \frac{1}{2}$ and the theorem follows.

The smoothness assumption in Theorem 2 rules out norms, such as the
Manhattan norm, for which the unit circle has corners. I believe that Theorem 2 still applies when there are only finitely many corners. However, since §7 solves the Manhattan norm case exactly, this generalization will not be pursued.

Beardwood, Hammersley, and Halton give const. \( \times (An)^{1/2} \) as an asymptotic formula for the expected (Euclidean) length of the minimal tree joining \( n \) points in a region of area \( A \). In particular the length in the unit circle would be asymptotic to const. \( \times (\pi n)^{1/2} \). Then Theorems 1 and 2 show that the constant is less than \( 2^{-1/2} = .707 \).

6. Euclidean norm. In this section the norm \( ||P|| \) will be Euclidean length. In this special case the integrations indicated in §4 can be simplified.

To find an expression for \( A(Q, r) \) note that \( K(Q, r) \) is bounded by two circular arcs. One arc has radius \( R = ||Q|| \); let its angle be called \( 2\varphi \). The other arc has radius \( r \) and another angle, say \( 2\vartheta \). The angles are related to each other by \( 2\vartheta + \varphi = \pi \), and to \( r \) and \( R \) by

\[
r = 2R \cos \vartheta = 2R \sin \frac{1}{2} \varphi.
\]

The area of \( K(Q, r) \) is \( r^2(\vartheta - \frac{1}{2} \sin 2\vartheta) + R^2(\varphi - \frac{1}{2} \sin 2\varphi) \). Eliminate \( r \) and \( \varphi \) to get \( A(Q, r) = R^2 a(\vartheta) \), where

\[
a(\vartheta) = 1 + \frac{2\vartheta \cos 2\vartheta - \sin 2\vartheta}{\pi}.
\]

Then (1) becomes

\[
f(Q) = 2R \left\{ \int_0^{\pi/2} \exp (-Na(\vartheta)R^2) \sin \vartheta d\vartheta - \exp (-NR^2) \right\}
\]

after a change of variable of integration from \( r \) to \( \vartheta \).

Since \( f(Q) \) depends only on \( R \), polar coordinates \( (R, \alpha) \) simplify the integral (2) to

\[
S(N) = (N/\pi) \int_0^{2\pi} d\alpha \int_0^1 f(Q)R \, dR;
\]

hence,

\[
S(N) = 4N \int_0^1 R^2
\]

\[
\left\{ \int_0^{\pi/2} \exp (-Na(\vartheta)R^2) \sin \vartheta d\vartheta - \exp (-NR^2) \right\} \, dR.
\]

The \( R \) integration in (9) can be performed first to leave a single integral on \( \vartheta \) which contains an error function of \( (2Na(\vartheta))^{1/2} \) in the integrand.

By transforming the integral (3) to polar coordinates, \( L(n) \) becomes an
integral of a finite sum of products of powers of $R$, $\vartheta$, $\cos \vartheta$, and $\sin \vartheta$. Thus $L(n)$ is an elementary integral for each $n$. To achieve a slightly shorter formula for $L(n)$, integrate (9) on $\vartheta$ by parts to get

$$S(N) = \frac{(32N^2/\pi)}{\int_0^1 R^4 \int_0^{\pi/2} \exp(-Na(\vartheta)R^2) \cos^2 \vartheta \sin \vartheta \, d\vartheta \, dR.}$$ (10)

Now expand $S(N) \exp(N)$ into a power series to get

$$L(n) = \frac{32n(n - 1)}{\pi} \sum_{k=0}^{n-2} \left( \begin{array}{c} n - 2 \\ k \end{array} \right) (-1)^k J(k),$$ (11)

where

$$J(k) = \frac{\int_0^{\pi/2} \{a(\vartheta)\}^k \vartheta \cos^2 \vartheta \sin \vartheta \, d\vartheta}{2k + 5}.$$

**Table 1**

Mean lengths of exodic trees connecting $n$ points in the unit circle

<table>
<thead>
<tr>
<th>$n$</th>
<th>$(\frac{n^2}{2})^{1/3}$</th>
<th>Euclidean</th>
<th>Manhattan and Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.772</td>
<td>0.9054</td>
<td>0.9333</td>
</tr>
<tr>
<td>3</td>
<td>2.171</td>
<td>1.517</td>
<td>1.563</td>
</tr>
<tr>
<td>4</td>
<td>2.507</td>
<td>1.989</td>
<td>2.052</td>
</tr>
<tr>
<td>5</td>
<td>2.803</td>
<td>2.383</td>
<td>2.458</td>
</tr>
<tr>
<td>6</td>
<td>3.070</td>
<td>2.725</td>
<td>2.811</td>
</tr>
<tr>
<td>7</td>
<td>3.316</td>
<td>3.029</td>
<td>3.126</td>
</tr>
<tr>
<td>8</td>
<td>3.545</td>
<td>3.306</td>
<td>3.411</td>
</tr>
<tr>
<td>9</td>
<td>3.760</td>
<td>3.561</td>
<td>3.673</td>
</tr>
<tr>
<td>10</td>
<td>3.963</td>
<td>3.798</td>
<td>3.917</td>
</tr>
<tr>
<td>11</td>
<td>4.157</td>
<td>4.020</td>
<td>4.146</td>
</tr>
<tr>
<td>12</td>
<td>4.342</td>
<td>4.231</td>
<td>4.362</td>
</tr>
<tr>
<td>13</td>
<td>4.519</td>
<td>4.431</td>
<td>4.567</td>
</tr>
<tr>
<td>14</td>
<td>4.689</td>
<td>4.621</td>
<td>4.763</td>
</tr>
<tr>
<td>15</td>
<td>4.854</td>
<td>4.804</td>
<td>4.950</td>
</tr>
<tr>
<td>16</td>
<td>5.013</td>
<td>4.979</td>
<td>5.130</td>
</tr>
<tr>
<td>17</td>
<td>5.168</td>
<td>5.149</td>
<td>5.302</td>
</tr>
<tr>
<td>18</td>
<td>5.317</td>
<td>5.311</td>
<td>5.470</td>
</tr>
<tr>
<td>19</td>
<td>5.463</td>
<td>5.488</td>
<td>5.633</td>
</tr>
<tr>
<td>20</td>
<td>5.605</td>
<td>5.618</td>
<td>5.787</td>
</tr>
<tr>
<td>21</td>
<td>5.743</td>
<td>5.797</td>
<td>5.937</td>
</tr>
<tr>
<td>22</td>
<td>5.879</td>
<td>6.087</td>
<td>6.087</td>
</tr>
<tr>
<td>23</td>
<td>6.011</td>
<td>6.188</td>
<td>6.188</td>
</tr>
</tbody>
</table>
The binomial coefficient in (11) arises from expanding \((1 - a\phi R^2)^{n-2}\) by the binomial theorem. As a check (11) gives the exact value for two points,

\[
L(2) = \frac{128}{45\pi} = .905415,
\]
a result which had been known earlier (see [2, pp. 41, 66]). Although the integrals \(J(k)\) are elementary they become increasingly laborious as \(k\) increases. To prepare Table 1 it was found easiest to compute \(J(k)\) numerically.

7. Manhattan and max norms. Consider the mapping \((x, y) \rightarrow (x', y')\) defined by \(x' = x - y, y' = x + y\) and with inverse mapping \(x = \frac{1}{2}(x' + y'), y = \frac{1}{2}(y' - x').\) Since the mapping is linear, it carries Poisson patterns in the \((x, y)\)-plane into Poisson patterns in the \((x', y')\)-plane. Moreover the Manhattan norm of \((x, y)\) equals the max norm of \((x', y').\) Then the max norm has the same \(S(N)\) and the same \(L(n)\) as the Manhattan norm.

Throughout this section \(\|P\|\) will denote Manhattan norm. The function \(A(Q, r)\) depends not only on \(r\) and on \(R = \|Q\|\) but also on a coordinate \(z\) which measures the distance from \(Q\) to the nearest corner of the (square) circle of radius \(R.\)

Consider a point \(Q = (R - \frac{1}{2}z, \frac{1}{2}z),\) \(0 \leq z \leq R,\) at Manhattan distance \(z\) away from the nearest corner \((R, 0).\) The region \(K(Q, r)\) can have three different shapes, depending on the size of \(r.\)

(i) \(r \leq z:\) The circle \(\|P - Q\| = r\) has two corners

\[
V_1 = (R - \frac{1}{2}z - r, \frac{1}{2}z), \quad V_2 = (R - \frac{1}{2}z, \frac{1}{2}z - r),
\]
satisfying \(\|V_i\| < R.\) Then \(K(Q, r)\) is half the circle \(\|P - Q\| \leq r\) and has area \(Cr^2/2.\)

(ii) \(z \leq r < 2R - z:\) In this case \(\|V_1\| < R \leq \|V_2\|\) and \(K(Q, r)\) is the rectangle with corners \((R, 0), (R - \frac{1}{2}r, -\frac{1}{2}r), V_1, (R - \frac{1}{2}z - \frac{1}{2}r, \frac{1}{2}z + \frac{1}{2}r).\)

(iii) \(2R - z \leq r < 2R:\) \(K(Q, r)\) is the rectangle with corners \((R, 0), (R - \frac{1}{2}r, -\frac{1}{2}r), (-\frac{1}{2}r, R - \frac{1}{2}r), (0, R).\)

\(A(Q, r)\) has the following formula:

\[
A(Q, r) = \begin{cases} 
\frac{1}{2} r^2 & \text{if } 0 \leq r < z \\
\frac{r(z + r)}{4} & \text{if } z \leq r < 2R - z \\
\frac{1}{2} Rr & \text{if } 2R - z \leq r \leq 2R 
\end{cases}
\]

(case i),

(case ii),

(case iii).
This function may be substituted into (2) and (3) to get $S(N)$ and $L(n)$. It is convenient to change variables of integration from $x_1$, $x_2$ to $R$, $z$ and use

$$\int_{\|x\| \leq 1} \cdots dx_1 dx_2 = 8 \int_0^1 \int_0^R \cdots \frac{1}{2} dz dR.$$ 

The factor $\frac{1}{2}$ is the Jacobian of $x_1$, $x_2$ with respect to $z$, $R$. The factor $8$ is needed because the range $0 \leq z \leq R < 1$ describes only $\frac{1}{8}$ of the unit circle.

$A(Q, r)$ is composed only of pieces of polynomials in $r$, $R$, $z$ and the integral in (3) is elementary. After some routine but tedious analysis one finds

$$L(n) = \sum_{k=1}^{n-1} \binom{n-1}{k} (-1)^k \left\{ A(k) + nB(k) \right\} - \frac{2}{3} (n - 1),$$

where

$$A(k) = \frac{4(2^{-k} - 1)}{k(2k + 1)} + \frac{4 - 2^{1-k}}{(k + 1)(2k + 3)},$$

$$(2k + 3)B(k) = \frac{2^{1-k}}{2k + 1} + \frac{k(4 - 2^{-k})}{(k + 1)(k + 2)} - 4.$$ 

For $n = 2, 3$, (12) gives the exact values $L(2) = 14/15$, $L(3) = 1313/840$. Numerical values for larger $n$ appear in Table 1. As expected, $L(n)$ stays close to $(\pi n/2)^{1/2}$ for both Euclidean and Manhattan norms. As $n$ increases both formulas (11) and (12) become sums of large numbers with alternating sign.

**Table 2**

Minimal trees connecting $n$ random points in the unit square

<table>
<thead>
<tr>
<th>$n$</th>
<th>Number of trees</th>
<th>Mean length</th>
<th>Standard deviation</th>
<th>$\cdot 8 n^{1/2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>100</td>
<td>.502</td>
<td>.26</td>
<td>.96</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>.880</td>
<td>.27</td>
<td>1.18</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>1.114</td>
<td>.30</td>
<td>1.36</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>1.352</td>
<td>.31</td>
<td>1.52</td>
</tr>
<tr>
<td>20</td>
<td>8</td>
<td>3.02</td>
<td>.15</td>
<td>3.04</td>
</tr>
<tr>
<td>30</td>
<td>8</td>
<td>3.85</td>
<td>.23</td>
<td>3.72</td>
</tr>
<tr>
<td>50</td>
<td>8</td>
<td>4.85</td>
<td>.18</td>
<td>4.80</td>
</tr>
<tr>
<td>80</td>
<td>8</td>
<td>6.08</td>
<td>.13</td>
<td>6.07</td>
</tr>
<tr>
<td>100</td>
<td>7</td>
<td>6.78</td>
<td>.10</td>
<td>6.80</td>
</tr>
</tbody>
</table>
8. **Minimal trees.** Ruth Weiss wrote a computer program using Prim's algorithm to find the minimal tree connecting a set of given points. The cost function is Euclidean distance. Her program constructed minimal trees for 439 random sets of points in the unit square. The results appear in Table 2.

As a check, the mean Euclidean distance between two points in the unit square may be found analytically to be

\[ \frac{2 + 2^{1/2}}{15} + \frac{1}{3} \log (1 + 2^{1/2}) = .522, \]

as compared with the experimental value .502. The exact standard deviation for this distance is \( \sigma = .245 \) instead of .26.

In Table 2, \(.68n^{1/2}\) seems to be a good numerical approximation to the mean length for large \( n \). Then for a large number \( n \) of points in a region of area \( A \) the expected length of the minimal tree should be near \(.68 (An)^{1/2}\). This is only slightly less than the asymptotic mean length \( \left(\frac{1}{2}An\right)^{1/2} \) derived for the exodic tree in §4. It is also surprisingly close to Beardwood, Halton, and Hammersley’s experimental estimate \(.75 (An)^{1/2}\) for traveling salesman paths.

**REFERENCES**