SIAM J. Appl. Math.
Vol. 16, No. 1, 1968
Printed in U.S.A.

STEINER MINIMAL TREES*

E. N. GILBERT AND H. O. POLLAK†

Abstract. A Steiner minimal tree for given points $A_1, \cdots, A_n$ in the plane is a tree which interconnects these points using lines of shortest possible total length. In order to achieve minimum length the Steiner minimal tree may contain other vertices (Steiner points) beside $A_1, \cdots, A_n$. We find conditions which simplify the task of constructing a Steiner minimal tree. Some of these use relationships with the easily constructed (ordinary) minimal tree which achieves minimum length among all trees having only $A_1, \cdots, A_n$ as vertices. Other questions concern the relative lengths of these two trees in extreme or typical cases. A review of the existing literature is included.

1. Introduction. A *minimal tree* for points $A_1, A_2, \cdots, A_n$ in the plane is a tree having these points as its vertices and having minimal length, i.e., having the sum of the lengths of all its lines as small as possible. Simple algorithms are known for drawing minimal trees [7], [14]. However, one can sometimes construct still shorter trees connecting $A_1, \cdots, A_n$ by adding extra vertices beside the A_i. For example, if A_1, A_2, A_3, A_4 are four given vertices at the corners of a square, a shorter tree than the minimal tree for these vertices consists of four lines from A_1, A_2, A_3, A_4 to an extra vertex S at the center of the square.

An extra vertex S which is added to a tree to reduce its length is called a *Steiner point* after J. Steiner who considered the case $n = 3$ (see [1, pp. 354–361]).[1] It will always be assumed that at least three lines go to a Steiner point; for it is clear that no reduction in tree length is possible otherwise. Another condition will be added in §3. When any number of extra Steiner points may be added at will, the shortest possible tree is called the *Steiner minimal tree*. This paper will discuss the problem of finding a Steiner minimal tree. Since results about the Steiner minimal tree problem are scattered in the literature, we have included some sections (§3, §5, §6) which are largely reviews of known results. Some other kinds of minimal network problems from agriculture, industrial engineering, and telephony are found in [3], [5], [8], [9], [13], [17] and [19].

Z. A. Melzak [12] gave an algorithm for finding the Steiner minimal tree in a finite number of steps. The algorithm involves trying such a large number of possibilities that the computation is feasible only for very small values of n. In this paper we find a number of geometrical properties of Steiner trees, some of which could be used to rule out large numbers of cases immediately.

* Received by the editors November 23, 1966, and in revised form April 20, 1967.

† Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey 07971.

[1] However the problem with $n = 3$ dates back at least to Fermat [2, p. 21].

Section 12 is a glossary which summarizes the special terms and notations which this paper introduces.

2. Kinds of trees. Steiner minimal trees are difficult to find because a tree having locally minimum length need not have the absolutely minimum length. For instance, the tree of Fig. 1a cannot be shortened by small displacements of the Steiner points. Nevertheless Fig. 1b is shorter (it is the Steiner minimal tree for A_1, A_2, A_3, A_4).

By the *topology* of a tree we shall mean a connection matrix, or any equivalent description, specifying which pairs of points from the list $A_1, A_2, \cdots A_n, S_1, S_2, \cdots$ have a connecting line. Thus the topology specifies connections but not the positions of $S_1, S_2, \cdots$.

Since A_1 and A_2 have lines to a common Steiner point in Fig. 1a but not in Fig. 1b, these two figures have different topologies. Figure 1a is shorter than any other tree (e.g., Fig. 1d) with the same topology. Such a tree will be called a *relatively minimal tree* because the minimum is achieved relative to a given topology. Figures 1c and 1e are relatively minimal trees for two other topologies. If the topology prescribes no Steiner point at all, there is just one tree with that topology and it is relatively minimal by default.

It is convenient to restrict the search for a relatively minimal tree to trees having no lines of zero length. As a result certain given topologies may have no relatively minimal tree. Suppose, for example, that A_1 had been much closer to A_4, and A_2 had been much closer to A_3 in Fig. 1. In that case one would approach a minimum for the topology of Fig. 1a by letting the Steiner points come arbitrarily close. The limiting tree which one approaches will be called a *degenerate tree*; it is really a relatively minimal tree for a different topology (having one Steiner point connected to all four A_i in this example).

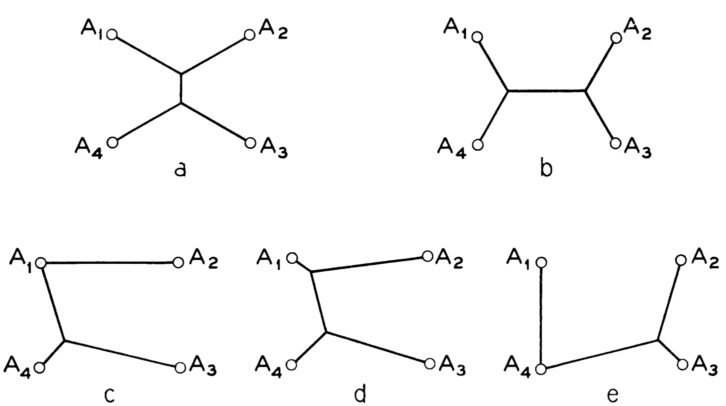


FIG. 1. *Trees for given points* A_1, A_2, A_3, A_4

One may regard Fig. 1d as a slightly perturbed copy of Fig. 1c although their topologies differ. Figure 1d is obtained from Fig. 1c by *splitting* vertex A_1 in two. In general to split a vertex V one disconnects two or more of the lines at V and connects them instead to a Steiner point V', located near V and connected to V by an extra line. After splitting it may happen, as in the case of Fig. 1d, that displacing V' away from V shortens the tree. If no small perturbation shortens a tree, even when splitting is allowed, then the tree is called a *Steiner tree*. Figures 1a and 1b are Steiner trees. A Steiner minimal tree is always a Steiner tree, and a Steiner tree is always a relatively minimal tree foı its topology.

A Steiner minimal tree is a minimal tree for all its vertices $A_1, \cdots, A_n$, $S_1, S_2, \cdots$, but the same need not be true for relatively minimal trees (Fig. 1c) nor even for Steiner trees. Indeed, one can achieve local minima of these two kinds by means of trees in which lines cross; a minimal tree cannot have crossing lines.

Relatively minimal trees have two useful properties. First of all, §4 will show that at most one relatively minimal tree exists for a given set A_1, $A_2, \cdots$ and a given topology. Secondly, §6 describes a construction (due to Melzak [12]) for the relatively minimal tree if it exists. Elementary considerations (§3.3 and §3.4) restıict the topology of a Steiner minimal tree to a finite set of possibilities which §7 enumerates. Thus the Steiner minimal tree may be obtained by constructing one relatively minimal tree for each topology. Unfortunately there are thousands of topologies to consider even when $n = 6$. Most of them are obviously poor possibilities; §8 gives criteria which rule out some topologies immediately.

3. Basic properties. This section collects together some elementary facts about the Steiner minimal tree problem. Many of these results are known but scattered in the literature. We obtain concise proofs of the known results as suggested by arguments from mechanics.

3.1. Mechanics. A tree will be interpreted as a mechanical system in which the potential energy is a sum of distances between adjacent vertices. Such a mechanical system is in stable equilibrium when the tree attains a relatively minimum length. For example, one may imagine lines of the tree to be elastic bands with the unusual property of having unit tension regardless of how much they are stretched. The points $A_1, A_2, \cdots, A_n$ are held fixed while the Steiner points are free to move. The equilibrium position is a relatively minimal tree.

Courant and Robbins [1, p. 392] give another interpretation involving a soap film between parallel plane plates separated by posts at $A_1, \cdots, A_n$. Since the film can change its topology by splitting (as defined in §2), the equilibrium tree is a Steiner tree. Steinhaus [16, p. 119] and Palermo [13] also give mechanical models.

3.2. Angles in a Steiner tree. In a Steiner tree no pair of lines meet at less than 120°. For suppose, on the contrary, that lines PR and RQ meet with angle $PRQ = \theta < 120°$. In the mechanical interpretation (§3.1) these two lines pull on point R with a resultant force of magnitude

$$F = 2 \cos \theta/2 > 1.$$

Now consider the effect of splitting vertex R. One adds a Steiner point S at R and replaces lines PR and RQ by PS, QS, and RS. The unit force of RS is inadequate to hold S at R against the combined force F exerted by QS and RS. Thus S is pulled away from R and one obtains a perturbed configuration with lower potential and shorter length, a contradiction.

3.3. Lines at a vertex. Because of §3.2 no vertex of a Steiner tree can have more than three lines. Consequently, we henceforth forbid vertices with four or more lines (such as the Steiner point mentioned at the beginning of §1). Thus every Steiner point of a Steiner tree has exactly three lines meeting at 120°.

At a Steiner point of a relatively minimal tree the three lines also meet at 120°. For in the mechanical interpretation (§3.2), the three unit force vectors acting at S can add to zero only if they are 120° apart.

Lines at $A_1, \cdots, A_n$ may meet at less than 120° in a relatively minimal tree. A relatively minimal tree is a Steiner tree if and only if all of its angles are 120° or more.

3.4. Number of Steiner points. Every tree has one more point than it has lines. A tree with given points $A_1, \cdots, A_n$ and s Steiner points must have $n + s - 1$ lines. Since each line has two ends, one obtains $2(n + s - 1)$ by summing, over all vertices, the number of incident lines. If n_k of the vertices A_i have k incident lines ($n_4 = n_5 = \cdots = 0$ for a Steiner tree), the sum in question is

$$3s + \sum_k kn_k = 2(n + s - 1) = 2s - 2 + 2\sum_k n_k .$$

Then

$$s = n_1 - 2 - n_3 - 2n_4 - \cdots .$$

In particular,

$$s \leqq n - 2,$$

with equality holding if and only if each A_i has one line.

3.5. Convex hull. In a relatively minimal tree all Steiner points lie in the convex hull of $A_1, \cdots, A_n$. Otherwise, some supporting line L of the convex hull would separate certain Steiner points, say $S_1, \cdots, S_r$, from $A_1, \cdots, A_n$. The total force on the set $S_1, \cdots, S_r$ is the vector sum of the tensions in the elastic bands crossing L. This sum has a nonzero component normal to L and so the tree cannot be in equilibrium. A much stronger result will be given in §8.2.

3.6. Maxwell's theorem. Draw unit vectors pointing outward from a relatively minimal tree in the directions of each of the lines incident at $A_1, \cdots, A_n$. Let F_i denote the sum of the unit vectors at A_i. In mechanical terms F_i is the external force needed at A_i to hold the tree in equilibrium. The length L of the tree has the simple formula

$$L = \sum_{i=1}^{n} A_i \cdot F_i . \tag{1}$$

Here any convenient origin may be used in measuring the position vectors A_i.

To prove (1) draw a unit force vector f_k pointing outward from each endpoint P_k of each line. Since there are $n + s - 1$ lines, there are $2(n + s - 1)$ vectors to draw. Now consider a sum of dot products defined by

$$\Sigma = \sum P_k \cdot f_k ,$$

where the sum extends over all $2(n + s - 1)$ endpoints. The terms of Σ may be grouped in two ways. First combine terms for which P_k is the same point. Each Steiner point appears in three of the terms of Σ. The contribution of a Steiner point S to Σ is zero because the three forces at S are in equilibrium. The contribution of a given point A_i to Σ is $A_i \cdot F_i$. Then Σ is the right-hand side of (1). The second grouping combines the two terms associated with the two ends of each line. If P_1P_2 is a line with length L_{12}, then $P_2 = P_1 + L_{12}u$, where u is a unit vector. The terms for this line contribute $P_1 \cdot (-u) + P_2 \cdot u = L_{12}$. Then Σ is also the left-hand side of (1).

The expression (1) for L is a special case of a theorem of J. C. Maxwell [10], [11]. J. D. Foulkes in an unpublished report seems to have been the first to recognize the connection between Maxwell's theorem and minimizing networks. Maxwell's application was to determine the minimum weight truss, made from pin-jointed rigid rods, holding a prescribed system of external forces $F_1, \cdots, F_n$ in equilibrium. In our application (1) is less useful because $F_1, \cdots, F_n$ are not known in advance.

In another form of Maxwell's theorem the A_k and F_k are regarded as complex numbers; one can then write

$$L = \sum_{k=1}^{n} A_k F_k^*. \tag{2}$$

The imaginary part of $A_k F_k^*$ is the moment of F_k about the origin. Then the condition for rotational stability ensures that the sum in (2) is real. However Re $A_k F_k^*$ is the dot product $A_k \cdot F_k$.

3.7. Full Steiner trees. A topology with $s = n - 2$ will be called a *full topology*. The corresponding relatively minimal tree will be called a *full Steiner tree*. According to §3.4 a topology is full if and only if each A_i has only one line. If $n \geqq 3$, the single line at A_i must lead to a Steiner point (otherwise the tree would be disconnected). A full Steiner tree is indeed a Steiner tree because there is no possibility of splitting $A_1, \cdots$, or A_n.

Moving along any path through a full Steiner tree, one changes direction at Steiner points, but only by $\pm 60°$. It follows that each of the F_k in (2) is one of the 6 complex numbers ± 1, $\pm\omega$, $\pm\omega^2$ ($\omega = \exp(i2\pi/3)$) if the tree is full and if the coordinate system is rotated suitably.

Every Steiner tree which is not full can be decomposed into a union of full Steiner trees as follows. Replace each vertex A_i having $k \geqq 2$ lines by k new vertices $A_{i1}, \cdots, A_{ik}$ all located at A_i (but regarded as disconnected from one another). Connect each of the k lines which were incident at A_i to a different one of $A_{i1}, \cdots, A_{ik}$. In this way the original tree is dissected into several smaller full trees which we call the *full components* of the original tree. Conversely, when the specified topology is not full, one can obtain the topologies of the full components. One may then try to construct the full components separately, afterward joining them together to produce the desired Steiner tree.

4. Uniqueness of relatively minimal trees. Let the vertices of a tree, including both $A_1, \cdots, A_n$ and the s Steiner points, be called $V_1, V_2, \cdots, V_{n+s}$. The topology of the tree determines an adjacency matrix,

$$a_{ij} = \begin{cases} 1 & \text{if } V_i V_j \text{ is a line of the tree,} \\ 0 & \text{otherwise.} \end{cases}$$

The length of the tree is

$$L = \sum_{i<j} a_{ij} | V_i - V_j |,$$

a function of the s Steiner points. Let the s Steiner points range over the entire plane (including configurations in which some lines have zero length). Each relative minimum of the function L corresponds to either a relatively minimal tree or a degenerate tree. We now show that L has a unique relative minimum.

The uniqueness will follow from a convexity property. Let T, T', and T'' be three trees with the given topology and having vertices V_i, V_i', and $V_i'' = pV_i + qV_i'$, where $p \geqq 0$, $q \geqq 0$, and $p + q = 1$. Let the lengths of these trees be called L, L', and L''. Then

$$\begin{aligned} L'' &= \sum a_{ij} | p(V_i - V_j) + q(V_i' - V_j') | \\ &\leqq \sum a_{ij} \{ p | V_i - V_j | + q | V_i' - V_j' | \}, \\ L'' &\leqq pL + qL'. \end{aligned}$$

All relative minima have the same length. For suppose, on the contrary, that trees T and T' are both relative minima although $L' < L$. Trees T'' with small p come arbitrarily close to T and have $L'' < L$. Then T is not a relative minimum, a contradiction.

The proof that only one relative minimum exists now proceeds by induction on s. The case $s = 0$ is trivial because the given topology specifies all connections and so there is only one tree. If there is always a unique relative minimum for problems with $s - 1$ Steiner points, consider a case with s Steiner points. If T and T' are two relative minima, they have the same length. For all (p, q) the intermediate trees T'' have this same length and so are relative minima. At least one pair of vertices (A_i, A_j) has a common Steiner point. Let this point be called S in T, S' in T' and $S'' = pS + qS'$ in T''. As p varies from 0 to 1, S'' moves on a straight line fron S to S'. But since T'' is a relative minimum, angle $A_iS''A_j$ must be 120°. Then S'' must move on a 120° circular arc $A_iSS'A_j$ from S to S'. These two loci agree only if $S = S'$. In that case T, T', and T'' are relative minima for a problem with $s - 1$ Steiner points in which S is prescribed as a new vertex A_{n+1}. Now the induction hypothesis shows that this problem has a unique solution, and hence $T = T'$.

5. The case $n = 3$**.** This section reviews the construction of the Steiner minimal tree for three points A, B, C. Let Δ denote the triangle ABC and let ϕ denote the largest angle of Δ, say $\angle ABC = \phi$.

If $\phi \geqq 120°$ no relatively minimal tree has a Steiner point S. For, according to §3.5, S must lie inside Δ; then

$$120° = \angle ASC > \angle ABS = \phi \geqq 120°.$$

The Steiner minimal tree is then just the minimal tree, consisting of AB and BC in this case.

If $\phi < 120°$, §3.2 shows that the Steiner minimal tree has a Steiner point

S. S lies inside Δ on a 120° arc with chord AB and also on another 120° arc through BC. These two arcs intersect at B and S; being arcs of circles they have no other intersections. Thus S can be found by constructing circles.

Figure 2 shows an even simpler construction. Let $AB'C$ denote the equilateral triangle exterior to Δ erected on the side AC. The 120° arc with chord AC is part of the circle circumscribing $AB'C$. Coxeter [2] shows that the Steiner point S, if it exists, lies on the line segment BB'. Thus S can be found as the intersection of BB' with the 120° arc AC. Moreover Coxeter shows, when S exists, that the length of the Steiner minimal tree is $| BB' |$. In his statement of these results he assumes that all angles of Δ are acute; however his proof applies as long as no angle exceeds 120°.

For the purposes of §6 we note here what happens to Fig. 2 if B may be so located that the Steiner minimal tree becomes degenerate. With B' in the position shown, B may lie anywhere to the left of the vertical line AC. A nondegenerate tree is obtained if B is in the region bounded by the 120° arc AC and the extended lines $B'A$, $B'C$. If B is outside this region, the Steiner point degenerates to one of A, B, or C as shown by the labels on the three remaining regions. In the degenerate cases the length of the Steiner minimal tree exceeds $| BB' |$.

In the nondegenerate case one might also construct the other exterior equilateral triangles $A'BC$ and ABC' and their circumscribing circles. The discovery that the three circles intersect at a common point (S) has been attributed to Napoleon (see Coxeter [2, p. 23]).

6. General case. Suppose that vertices $A_1, \cdots, A_n$ are given together with s Steiner points and a topology. The relatively minimal tree (if any) is found by an induction on s in which the construction of §5 is used several times to locate Steiner points. Since the construction of a tree with $s = 0$ is trivial, we shall consider a case $s \geqq 1$ and show that it can be reduced to drawing one or more trees with fewer Steiner points.

The first step is to find a Steiner point S which the topology connects to two vertices A_i, A_j. Such an S exists. For consider any full component (cf. §3.7) with $s' \geqq 1$ Steiner points and n' given points. Each of the given points has only one line and it leads to a Steiner point. Since $n' = s' + 2$, at least one Steiner point S must connect to two given points (there are two or more such Steiner points if $s' \geqq 2$).

Let A_i, A_j and B denote the three points connected to S. If B is a point A_k, the construction of §5 either locates S or decides that there is no solution. Suppose an S is located. If S and the lines A_iS, A_jS, A_kS were deleted from the relatively minimal tree, the result would be three separate relatively minimal trees, each containing one of A_i, A_j, A_k. Each tree contains fewer than s Steiner points and so can be constructed by the inductive procedure.

If B is a Steiner point, its location is not known. Nevertheless, the construction of Fig. 2 can proceed part way. B' can have only two locations (on opposite sides of the line A_iA_j). It may be necessary to try both possibilities although other properties, such as §3.5, can often rule out the wrong choice. If a relatively minimal tree T exists, it can be constructed as follows. First solve a simpler problem in which vertices A_i, A_j and S are removed and B' is added as one of the A_k. Connections to points other than S are the same as before; in addition, a line $B'B$ is required. The new minimal tree has the same length as the old one because $|B'B| = |A_iS| + |A_jS| + |BS|$. When the new tree is found, $B'B$ may be replaced by A_iS, A_jS and BS to get the desired tree.

One cannot always immediately rule out the wrong one of the two choices for B'. For a given B', the computation may then place B outside the "nondegenerate" region of Fig. 2, or it may conclude that the relatively minimal tree for the new topology (with B' replacing S, A_i, A_j) does not exist. One must then try the other choice of B'. If both choices fail, the given topology had no relatively minimal tree.

The inductive procedure replaces the problem with s Steiner points by one with $s - 1$. This in turn is replaced by an $s - 2$ point problem, etc.

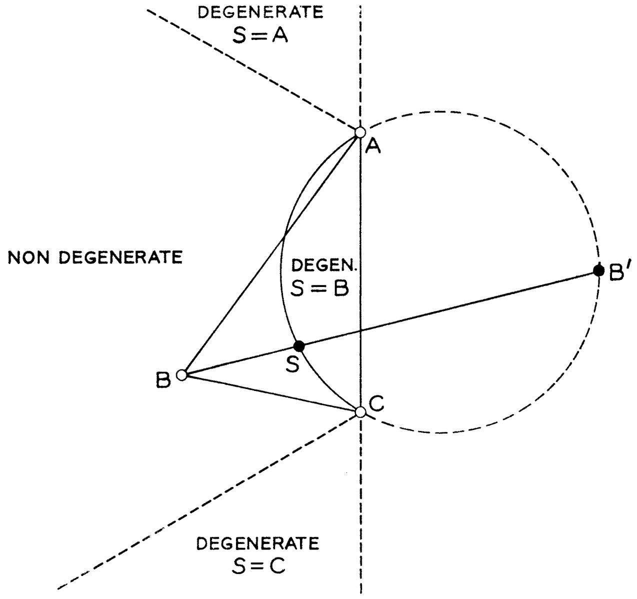


FIG. 2. *Construction of S; degenerate and nondegenerate locations for B*

The relatively minimal tree in each of these simpler problems has the same length as the desired relatively minimal tree. Thus, when the problem has been simplified so far that the solution is evident, the length is known immediately. For example, if the original tree is a full tree, so is each of the simpler trees; ultimately one reaches the situation shown in Fig. 2, and the length in that figure is $|B'B|$.

If there are s Steiner points, each of which might be in one of two places, it seems as though one must draw about 2^s trees to find the relatively minimal tree. In most hand computations the situation is not that bad because one can generally guess the correct choice for each B'. If the guesses are indeed correct and a relatively minimal tree is found, the uniqueness result of §4 shows that one need not try the other possibilities. When the computation is done by machine, it is harder to make good guesses. Some of the results of §8 may help.

7. Enumerations. To find a Steiner minimal tree by exhausting all possible relatively minimal trees one needs a way of listing the topologies. Steiner trees have at most three lines incident at each vertex A_k ; indeed even three lines can meet at A_k only by the rarest accident. For that reason this section will be concerned mainly with topologies in which each A_k has only one line or two. These topologies can be listed by an inductive procedure.

When counting or listing topologies the Steiner points are considered to be unlabeled; e.g., the tree of Fig. 1a is counted just once. However trees which would look alike with the labels $A_1, \cdots, A_n$ erased are counted as different trees if they are labeled differently. Figure 1a provides three topologies when the labels are permuted in all possible ways.

Full topologies (with $n = s + 2$) have an important role in the enumeration. Every topology has an underlying full topology, obtained as follows. According to §3.4, $s + 2$ of the A_k have one line and $n - s - 2$ have two lines. Remove these $n - s - 2$ vertices one at a time, merging the two lines at each removed vertex into a single line. The final result is a full tree for the remaining vertices.

If there are $f(s)$ full topologies for trees with s Steiner points and $F(n, s)$ topologies for trees which have n vertices A_k and s Steiner points, the $F(n,s)$ topologies can be obtained from the $f(s)$ topologies as follows. First pick a subset of $s + 2$ of $A_1, \cdots, A_n$ to serve as vertices of a full topology. Then the underlying full topology is drawn in one of $\binom{n}{s+2} f(s)$ ways Now add the remaining $n - s - 2$ vertices one at a time at interior points of the lines which have been drawn. The first vertex can go to one of $2s$

$+ 1$ lines. There are then $2s + 2$ lines for the second vertex, etc. The result is

$$F(n, s) = \binom{n}{s+2} f(s)(2s + 1)(2s + 2) \cdots (n + s - 2)$$

topologies.

The $f(s)$ full topologies can be obtained by induction on s. When $s = 0$, one has $f(0) = 1$; the single possibility consists of A_1 and A_2 with a line between them. Once all $f(s)$ full topologies for trees with s Steiner points are known, one can produce $(2s + 1)f(s)$ full topologies for trees with $(s + 1)$ Steiner points as follows. Pick any one of the $2s + 1$ lines in any one of the $f(s)$ trees and add an extra Steiner point S in the middle of the chosen line. Add a line from S to a new vertex A_{s+3}. Every one of the $f(s + 1)$ topologies is obtainable this way. For, as noted in §3.4, A_{s+3} is always connected to some Steiner point S; removing the line SA_{s+3} reduces the tree to one of the $f(s)$ trees. Thus $f(s + 1) = (2s + 1)f(s)$, which has the solution

$$f(s) = 2^{-s}(2s)!/s!. \tag{3}$$

Finally,

$$F(n, s) = 2^{-s}\binom{n}{s+2}(n + s - 2)!/s!. \tag{4}$$

The formula for $f(s)$ can also be obtained from a generating function of J. Riordan [15] which enumerates labeled trees by degrees.

Table 1 gives some of these numbers. As one might expect, the numbers

TABLE 1

Number of possible topologies for Steiner trees with n given vertices and s Steiner points

s \ n	3	4	5	6	7
0	3	12	60	360	2520
1	1	12	120	1200	12600
2		3	75	1350	22050
3			15	630	17640
4				105	6615
5					945
Total	4	27	270	3645	62370

grow large when n is not small. The results of the next section eliminate many of these possibilities at once. Figure 3 draws the full topologies for $s \leqq 7$. The labels $A_1, \cdots, A_{s+2}$ have been omitted from these drawings, so that each provides many topologies when the labels are added in all possible ways. The number of distinct labelings, which depends on symmetry properties, appears under each tree.

It is conceivable, although unlikely, that the Steiner minimal tree will have $n_3 > 0$ vertices A_k with three lines. The topologies for such trees are obtainable by drawing topologies for trees with $n - n_3$ given vertices and

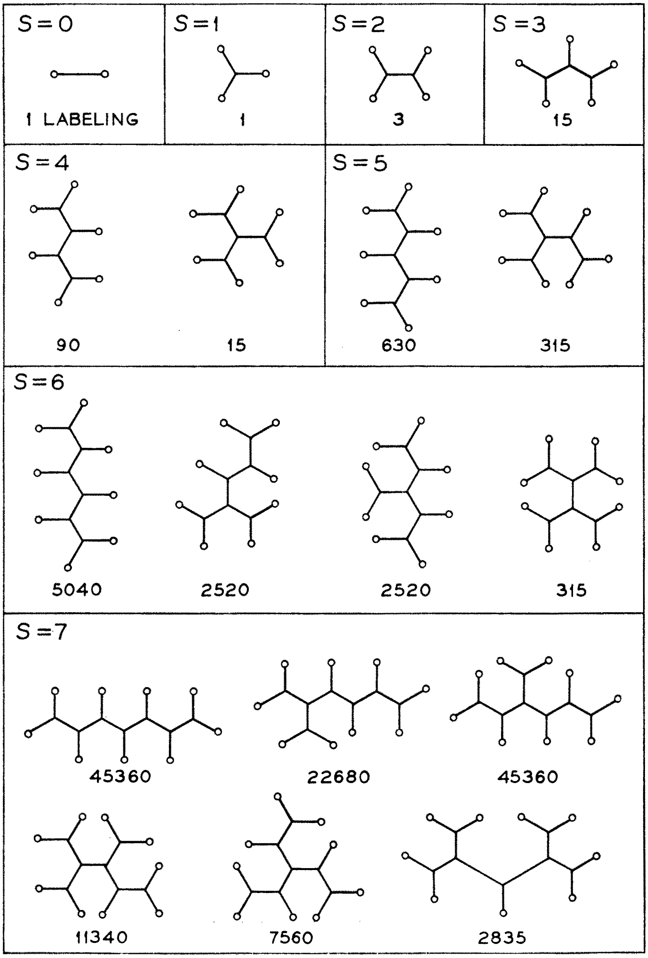


FIG. 3. *Full Steiner trees with s Steiner points*

$s + n_3$ Steiner points and then labeling n_3 of the Steiner points as vertices A_k. Thus there are

$$\binom{n}{n_3} F(n - n_3, s + n_3)(s + n_3)!/s!$$

topologies of this kind.

8. Restrictions on Steiner minimal trees. In this section we derive a number of necessary conditions on Steiner minimal trees. Some of them can help determine the correct choice of B' in the construction algorithm of §6. Others can rule out certain relatively minimal trees as candidates for the Steiner minimal tree.

8.1. Lune property. Let AB be any line of a minimal tree. (The result to follow will hold even when one is not free to add Steiner points.) Let $L(A, B)$ be the region consisting of all points X satisfying

$$|XA| < |AB| \quad \text{and} \quad |XB| < |AB|.$$

$L(A, B)$ is the lune-shaped intersection of circles of radius $|AB|$ centered on A and B. No other vertex of the tree can lie in $L(A, B)$. For, if X_0 were such a vertex, the tree would contain either a path from X_0 to A not containing B or a path from X_0 to B not containing A. In the latter case, for example, one could shorten the tree by deleting AB and adding X_0A.

In a Steiner minimal tree, $L(A, B)$ must also contain no interior point X_0 of another line of the tree. For the same argument again applies with X_0 added as a Steiner point.

8.2. Wedge property. Let W be any open wedge-shaped region having angle 120° or more and containing none of $A_1, \cdots, A_n$. Then W contains no Steiner point of a relatively minimal tree.

The proof is by contradiction. Suppose W contains a Steiner point S but no A_i. If there are several Steiner points in W, we may introduce a Cartesian coordinate system (x, y) with positive x-axis along the angle bisector of the wedge and pick S to be a Steiner point with largest x coordinate. Of the three lines at S, one leaves S in a direction within $\pm 60°$ of the positive x-axis. This line cannot leave W and so cannot end at one of $A_1, \cdots, A_n$. Moreover, its endpoint has a larger x coordinate than S, a contradiction.

The union of all such wedges W contains no Steiner point. Thus all Steiner points lie in the complement K of this union. K is a very small set in some cases. For example, if A_1, A_2, A_3 are three vertices of an equilateral triangle, the corresponding K contains a single point at the center of the triangle. More generally K consists of one or more closed regions bounded by arcs of circles. K is always a subset of the convex hull of $A_1, \cdots, A_n$ because each supporting line of the convex hull bounds a 180° wedge free of vertices $A_1, \cdots, A_n$.

8.3. Double wedge properties. Two lines which cross at 120° cut the plane into two 60° wedges and two 120° wedges. Let R_1 and R_2 denote the two closed 60° wedges. If $R_1 \cup R_2$ contains all of $A_1, \cdots, A_n$, then §8.2 shows it contains all Steiner points too. Let X denote the point at which R_1 and R_2 meet. First, if X is not one of $A_1, \cdots, A_n$, then the Steiner minimal tree contains at most one line having one endpoint in R_1 and one endpoint in R_2. Secondly, if X is one of $A_1, \cdots, A_n$, there are no such lines, aside from lines having X as one endpoint.

When they apply, these properties severely restrict the topology of the Steiner minimal tree. In particular, if X is one of $A_1, \cdots, A_n$, the Steiner minimal tree may be found simply by drawing two Steiner minimal trees, one for vertices in R_1 and another for vertices in R_2. These two trees connect at X to form the desired tree.

We prove the first result by contradiction. If X is not one of $A_1, \cdots, A_n$, X is not a Steiner point either; otherwise one of the three lines at X leads to a vertex outside $R_1 \cup R_2$. Suppose the Steiner minimal tree has more than one line joining R_1 to R_2. Let P_1P_2 be one such line with P_1 in R_1 and P_2 in R_2. Either there is another path from P_1 to R_2 or another path from P_2 to R_1; suppose the former case. Let the other path from P_1 to R_2 end with a line Q_1Q_2 with Q_1 in R_1 and Q_2 in R_2 (Q_1 may equal P_1 but $Q_2 \neq P_2$ and none of P_1, P_2, Q_1, Q_2 is X). One can derive two new trees from the Steiner minimal tree by adding a line P_2Q_2 and removing one of P_1P_2 or Q_1Q_2. To obtain a contradiction we shall show that one of these new trees is shorter than the Steiner minimal tree.

Since angle P_2XQ_2 is less than 60°, the law of sines shows that P_2Q_2 cannot be the largest side of the triangle P_2XQ_2. If, for instance, $|P_2Q_2| \leqq |XP_2|$, then the new tree which provides the contradiction will be the one with P_1P_2 removed. For X is closer to P_2 than any other point of R_1; in particular,

$$|P_1P_2| > |XP_2| \geqq |P_2Q_2|.$$

The second result has a similar proof by contradiction. Suppose X is A_k and suppose the Steiner minimal tree contains a line P_1P_2 with P_1 in R_1, P_2 in R_2, and neither P_1 nor P_2 equal to X. Remove P_1P_2. The tree breaks into two components, one of which contains X. This component must contain either P_1 or P_2. If it contains P_1, add a line XP_2 to get a new tree. Again $|XP_2| < |P_1P_2|$ and the new tree is shorter than the Steiner minimal tree, a contradiction.

8.4. Distance between Steiner points. Suppose S_1 and S_2 are two Steiner points which are connected by a line in the Steiner minimal tree. S_1 and S_2 each have two other lines. Let the shortest of these four lines have length L_0. Then the line S_1S_2 must have length at least $(\sqrt{3} - 1)L_0$.

To prove this result let P_1 and P_2 be the two points which lie on the two lines to S_1 at distance L_0 away from S_1. Likewise, let P_3 and P_4 be two points lying at distance L_0 from S_2 on its two lines. The points P_1, P_2, P_3, P_4 lie at the corners of a rectangle, and the part of the tree which connects P_1, P_2, P_3, P_4 together has the appearance of Fig. 1a in which the A_i are renamed P_i. The sides of the rectangle are $| P_1P_2 | = | P_3P_4 | = \sqrt{3}\, L_0$ and $| P_2P_3 | = | P_4P_1 | = L_0 + |S_1S_2|$. If $L_0 + | S_1S_2 | < \sqrt{3}\, L_0$, as in Fig. 1a, then a shorter network connecting P_1, P_2, P_3, P_4 has the appearance of Fig. 1b.

8.5. Lines A_iA_j. Since a minimal tree for $A_1, \cdots, A_n$ is easy to construct, it is natural to ask if drawing a minimal tree M first can help in finding the Steiner minimal tree. We show now that, if the Steiner minimal tree is to have any lines of the form A_iA_j (i.e., if the tree is not full), then each of these lines must also be lines of the minimal tree[2] M.

Suppose A_1A_2 is a line in the Steiner minimal tree. The vertices $A_1, \cdots, A_n$, together with all the Steiner points, fall into two classes C_1, C_2. C_1 contains all vertices which can be reached from A_1 via a path which does not contain A_2. C_2 contains the remaining vertices; they may be reached from A_2 via a path not containing A_1. No pair of points P_1, P_2 with P_i in C_i, $i = 1, 2$, are closer together than A_1 and A_2. For if a closer pair P_1, P_2 existed, the Steiner minimal tree could be shortened by replacing the line A_1A_2 by P_1P_2.

Now consider the steps given by Prim [14] for drawing the minimal tree M. The algorithm starts at any point, say A_1, and adds lines one at a time to form a tree which ultimately grows to become the minimal tree. At each step the added line is the (or one of the) shoitest possible. At some step the added line must go to a point of C_2. Since A_1A_2 is a shortest line between C_1 and C_2, A_1A_2 may be taken as the added line.

This result rules out all topologies in which an incorrect line A_iA_j appears. An even stronger result, which unfortunately is not true, would state that the minimal tree consists of A_1A_2 and the two minimal trees for the vertices in C_1 and the vertices in C_2. Figure 4 shows a counterexample.

8.6. Diamond property. By the diamond $D(A, B)$ of a line AB we shall mean the open set containing the points P such that both angles APB and BPA are less than 30°. $D(A, B)$ is a rhombus having angles 60°, 120°, 60°,

[2] Section 10 will discuss another connection with the minimal tree.

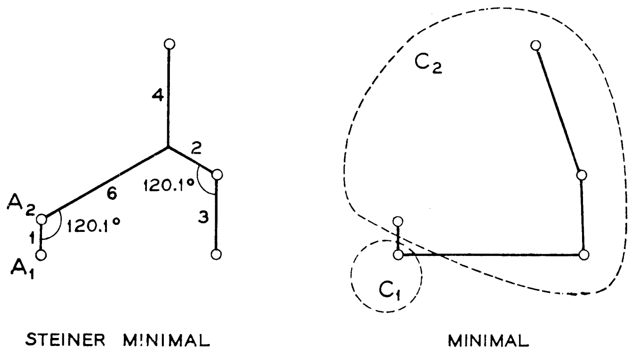


FIG. 4. *A minimal tree which does not contain the minimal tree for the vertices in* C_2

120° and with AB as longest diagonal. If one draws diamonds about each line of a minimal (e.g., Steiner minimal) tree, one will find that no two diamonds intersect.

Before proving this result we mention one of its consequences. Suppose the n points of a minimal tree are known only to lie in a certain region R. Let U be the union of all diamonds $D(P, Q)$ with P and Q both points of R. For example, if R is a unit circle, then U is a circle of radius $2/3^{1/2}$. Since the $n - 1$ diamonds about the lines of the minimal tree are disjoint subsets of U, one obtains

$$\sum_{i=1}^{N-1} 3^{-1/2} L_i^2/2 \leqq \text{area } (U),$$

with L_i the length of the ith line in the tree. The length of the tree is now

$$\begin{aligned}\sum_{i=1}^{n-1} L_i &\leqq \left\{\sum_{i=1}^{n-1} L_i^2 \sum_{j=1}^{n-1} 1\right\}^{1/2} \\ &\leqq 2^{1/2}3^{1/4}\{(n-1) \text{ area } (U)\}^{1/2}\end{aligned}$$

because of Schwarz's inequality.

For example, if $A_1, \cdots, A_n$ lie in a unit circle, the minimal tree (and hence the Steiner minimal tree) has length less than $3.81(n-1)^{1/2}$. This result may be compared with the expected length of the minimal tree obtained when $A_1, \cdots, A_n$ are placed at random in the unit circle. In [4] one of us showed that the expected length is less than about 1.25 $\cdot(n-1)^{1/2}$; experimental values lie near $1.20(n-1)^{1/2}$. The bound for a unit square is $\{5.464(n-1)\}^{1/2}$. For $n \leqq 12$ this is a slight improvement on a bound $2 + (2.8n)^{1/2}$ obtained by Verblunsky [18] (his bound applies even to the minimum length path connecting $A_1, \cdots, A_n$).

For still another comparison consider the regular point lattice consisting of all points of the form $a(iu + jv)$, where u, v are unit vectors 60° apart, a is some positive number, and i, j range over all integer values. Take A_1, A_2, $\cdots$ to be all the lattice points lying in the unit circle. When a is small, the number of lattice points is $n = 2\pi(3)^{-1/2}a^{-2}$ approximately and the length of the minimal tree is

$$(n - 1)a \sim \{2\pi n/3^{1/2}\}^{1/2} = 1.90\, n^{1/2}.$$

We conjecture, but cannot prove, that this length is asymptotic for large n to the length of the longest minimal tree having n vertices, all in the unit circle.

The diamond property is a best possible result in the sense that the 60° angles cannot be increased. To see this note that two lines can meet at 60° (as in the case that A_1, A_2, A_3 are located at corners of an equilateral triangle). Then two diamonds can have a side in common.

In the proof of the diamond property which follows we use several major simplifications suggested by J. H. van Lint. Let AB and XY be two lines of a minimal tree. Let C_A and C_B be open circles of radius $| AB |$ centered on A and B. Then the lune $L(A, B)$ is the intersection $C_A \cap C_B$. The proof will consider four cases; these depend mainly on how X and Y are situated with respect to C_A and C_B. Because of the lune property (§8.1) we need not consider cases in which X or Y belong to $C_A \cap C_B$.

In case 1 both C_A and C_B contain a point X or Y. Since X and Y are not in $C_A \cap C_B$, we may suppose XY labeled so that X is in C_A and Y is in C_B. Now a contradiction will be obtained by constructing a tree shorter than the supposed minimal tree. As a first step remove the line AB. The minimal tree now breaks into two components, one of which contains the line XY and a vertex A or B. If A is disconnected from X and Y, add the line XA to the tree. If B is disconnected from X and Y, add the line YB. In either event one obtains a shorter tree.

Since at least one of C_A and C_B contains neither X nor Y, we shall suppose AB labeled so that X and Y are in C_B'. Also let AB be the longer of the two lines, say $| XY | \leqq | AB | = 1$. It suffices to show in the remaining cases that, under these restrictions on X and Y, diamonds $D(A, B)$ and $D(X, Y)$ can overlap only if A is in $L(X, Y)$.

In case 2, X and Y are both in $C_A' \cap C_B'$. If P is any point of $D(A, B)$, the distances $| PX |$ and $| PY |$ must exceed $3^{-1/2}$. For the closest points of $C_A' \cap C_B'$ to $D(A, B)$ are the two corners of the lune $L(A, B)$. If P is a point of $D(A, B) \cap D(X, Y)$, the angle XPY is more than 120° and so $| XY |^2 > | XP |^2 + | YP |^2 + | XP |\cdot| YP | \geqq 1$, contradicting the assumption that AB is the longer line.

In the remaining cases at least one of X and Y belongs to C_A. Without loss of generality, suppose the labels X, Y have been so chosen that X

belongs to C_A (and so to $C_A \cap C_B'$). Y belongs to C_B'. The two final cases depend on the size of the angle YAB.

In case 3, $\angle YAB \leqq 60°$. Then, as shown in Fig. 5a, Y lies in a region bounded by a unit circle about B and the extended line segments AZ_1 and AZ_2. If $D(A, B)$ and $D(X, Y)$ overlap, a corner P of $D(X, Y)$ must lie inside $C_A \cap C_B$. We now find some restrictions on the location of P. First of all, P is as close to B as possible if XY is a chord of the circle C_B. Since $|XY| \leqq 1$, $|PB| \geqq 3^{-1/2}$; then P lies outside a circle through corners Q_1, Q_2 of $D(A, B)$. Also, since $|PY| \leqq 3^{-1/2}$, P must lie in one of the circles of radius $3^{-1/2}$ centered on the corners Z_1, Z_2 of $L(A, B)$. Thus P is in one of the shaded regions of Fig. 5a.

Suppose P is in the upper region. Then Y lies within distance $3^{-1/2}$ of Z_1 and X lies within distance 1 of Z_1. Neither line PX nor PY crosses AQ_1 or BQ_1 (although PX and AQ_1 may be the same line). Then the (open) diamonds $D(A, B)$ and $D(X, Y)$ do not overlap.

Finally, in case 4, X is in $C_A \cap C_B'$ and $\angle YAB > 60°$. If X and Y lie on opposite sides of the extended line AB, the diamonds can only intersect if A is in $L(X, Y)$, a contradiction. Suppose X and Y both lie above AB in Fig. 5b. Again a corner P of $D(X, Y)$ lies in $C_A \cap C_B$ outside the circle through Q_1 and Q_2; then the diamonds intersect only if P is in $D(A, B)$. We now show that P cannot lie in $D(A, B)$ even if the constraints on X and Y are relaxed slightly. Let X and Y range over the 120° sector lying above the extended lines AB and AZ_1. Look for positions of X and Y which minimize the angle PAB. Clearly one can take X to be on the extended line AB and Y on AZ_1 as in Fig. 5b. Since both angles XAY and XPY are then 120°, X, A, P and Y all lie on a 120° circular arc. Then $\angle PAY = \angle PXY = 30°$ so that $\angle PAB = 60° - 30° = 30°$ in the minimizing case. Again neither PX nor PY crosses AQ_1, and the diamonds do not overlap.

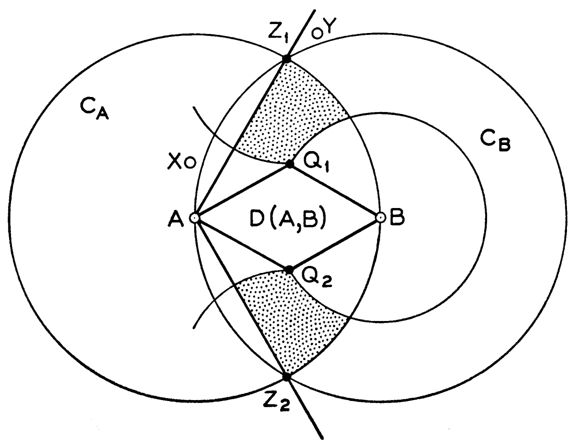
8.7. The deciding region. At each step in the construction algorithm of §6 one encounters the situation shown in Fig. 2. A and C are fixed points which are to be connected, if possible, to a common Steiner point S. The location of the third point B is not yet fixed, and so S may be anywhere on the 120° arc AC. One can sometimes conclude immediately, from the location of nearby vertices, that A and C cannot have a Steiner point S anywhere on this arc.

Consider all possible Steiner minimal trees with three or more given vertices $A_1 = A$, $A_2 = C$, A_3, A_4, $\cdots$, where A_3, A_4, $\cdots$ range over the entire plane and in which A and C have a common Steiner point S on the particular 120° arc shown in Fig. 2. A point P in the plane will be called a *deciding point* if P is not a vertex of any of the Steiner minimal trees. The set of all deciding points is the *deciding region* which we now determine, approximately. In any problem for which one of the A_k belongs to the deciding region, one can decide, without further calculation, that the Steiner minimal tree does not have a Steiner point S as shown in Fig. 2.

Wherever S may be, §8.1 shows that the union of the two lunes $L(S, A)$ and $L(S, C)$ must contain no vertex A_k. A point P in the intersection

$$I = \bigcap_S \{L(S, A) \cup L(S, C)\}$$

(S ranging over the 120° arc AC) cannot be a vertex A_k, no matter where S lies. Thus I is a subset of the deciding region. Figure 6 shows I. It is



(a) *Case 3*

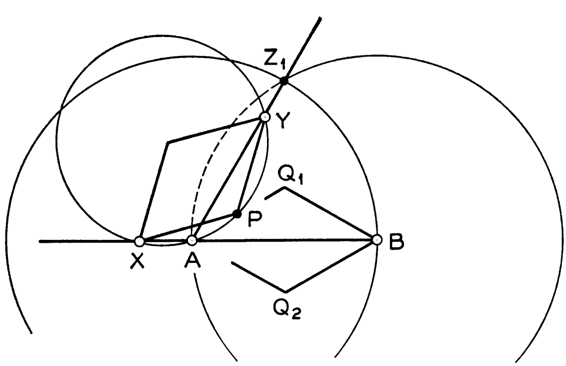


FIG. 5. *Proof of the diamond property*
(b) *Case 4*

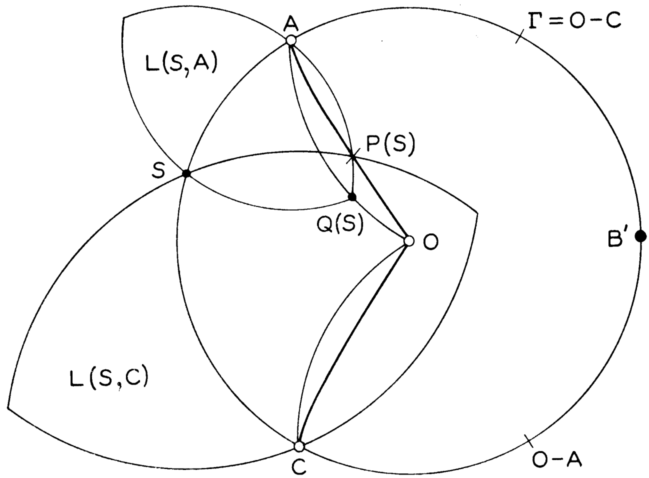


FIG. 6. *The region I*

bounded by the 120° arc AC and two other arcs. These arcs leave A and C and meet at O, the center of the circle through A, B', and C. The arc from O to A is the locus of points $P(S)$ satisfying the simultaneous equations

$$|P(S) - S| = |A - S|, \qquad |P(S) - B| = |B - S|$$

for points S lying on the 120° arc AC and closer to A than to C.

In Fig. 6 the arc $P = P(S)$ is described in Cartesian coordinates by an equation of fourth degree. However this boundary can be replaced by a simpler one without sacrificing much of the area of I. The new arc in question is $P = Q(S)$, where the point $Q(S)$ is the corner of $L(S, A)$ lying in I. We now show that this arc does indeed lie in I and that it is a 60° arc of a circle centered on $\Gamma = O - C$.

The line segment $B'S$ extended beyond S, meets SA and SC at 120° (cf. §5). Thus angle ASB' is 60°. Since angle $ASQ(S)$ is also 60°, $Q(S)$ lies on the line SB'. Since Γ is the midpoint of the arc AB', angles $AS\Gamma$ and $\Gamma SB'$ are both 30°. Then ΓS is a perpendicular bisector of the side $AQ(S)$ in the equilateral triangle $ASQ(S)$. But then $|\Gamma A| = |\Gamma Q(S)|$, i.e., $Q(S)$ moves on a circle of center Γ.

The line segment SB' contains both $Q(S)$ and, for the same reason, a corner $R(S)$ of $L(S, C)$. $Q(S)$ and $R(S)$ lie on SB' at distances $|SA|$ and $|SC|$ away from S. Since it is assumed that S lies closer to A than to C, $Q(S)$ lies between S and $R(S)$ on a chord of $L(S, C)$. Thus, for every S, a point of the 60° arc AO lies either in $L(S, A)$, $L(S, C)$ or both; the arc AO then belongs to I.

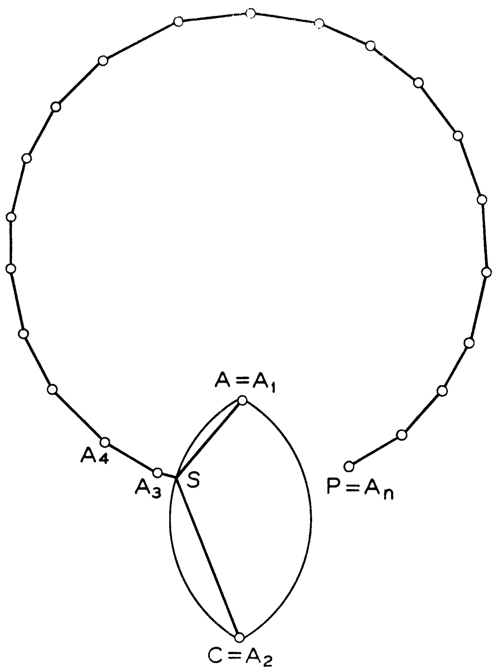


FIG. 7. *P is not a deciding point if angle APC* $< 120°$

The region I of Fig. 6 may not be the entire deciding region but it cannot be much different. Figure 7 shows a construction to prove that any point P for which angle APC is less than 120° cannot be a deciding point. Such points P lie outside the region bounded by the two 120° arcs through A and C. For any such P one can find a possible Steiner point S which lies closer to both A and to C than does P. Now suppose that points A_3, A_4, $\cdots$ are spaced close together as shown along a large circular arc from P to S. The tree shown is a Steiner tree. It seems intuitively obvious that it is the Steiner minimal tree, but the proof is difficult. The proof will appear as part of a forthcoming paper by R. L. Graham.

9. Generalizations. Many of the results about Steiner minimal trees do not depend on properties of the Euclidean plane.[3] In this section $A_1, \cdots, A_n$ and Steiner points will lie in some Riemannian space (possibly of dimension more than 2). Relatively minimal trees and Steiner minimal

[3] In the case of minimal trees, the algorithms of Prim and Kruskal require only that the $n(n-1)/2$ "distances" between pairs of the n given points must satisfy the triangle inequality. Hanan [6] discusses the Steiner minimal tree problem using $|x - X| + |y - Y|$ for the distance between (x, y) and (X, Y).

trees may be defined as in the Euclidean plane problem; now the lines of the trees are geodesic arcs.

The mechanical model (§3.1) generalizes directly and again one concludes (as in §3.2) that lines cannot meet at less than 120° in a Steiner tree. To generalize §3.3 and §3.4 it suffices to show that at most three vectors can meet at angles of 120° or more. Suppose $v_1, \cdots, v_k$ are k such unit vectors. Observe that

$$0 \leqq |v_1 + \cdots + v_k|^2 = \sum_i |v_i|^2 + \sum_{i \neq j} v_i \cdot v_j .$$

Since v_i and v_j meet at 120° or more, $v_i \cdot v_j \leqq -\frac{1}{2}$. Then

$$0 \leqq k - k(k-1)/2 = k(3-k)/2,$$

an inequality which is satisfied only by $k = 0, 1, 2, 3$.

Sections 3.5 and 3.6 do not generalize to arbitrary Riemannian spaces, but §3.5 and (1) of §3.6 hold in Euclidean spaces of any dimension. Likewise the uniqueness theorem of §4 requires the Euclidean metric but holds in any dimension. Of the properties in §8 only the lune property (§8.1) and the connection between the minimal tree and Steiner minimal tree (§8.5) hold in arbitrary Riemannian spaces. The wedge (§8.2) and double wedge (§8.3) properties hold in any Euclidean space, but the arguments used in §8.4, §8.6, and §8.7 apply only in the plane. R. L. Graham and J. H. van Lint supplied the following argument which shows that the diamond property also holds in Euclidean spaces of dimension 3 or more.

The diamond $D(A, B)$ is defined again as in the beginning of §8.6. In dimension 3 or higher the "diamond" is really a pair of truncated cones placed base to base. The proof begins as in §8.6. The argument for case 1 is valid in any dimension, and so we again suppose the labels A, B are so chosen that X and Y are both in C_B'; likewise we now suppose A and B are in C_Y'. Now let P be a point at which $D(A, B)$ and $D(X, Y)$ intersect and look for a contradiction.

Let π be the plane containing B, Y and P. Let A^0 be a point in π obtained by rotating A about the axis BP. There are two such points in π; take A^0 to be the one farther away from Y. Since A is in C_Y' and $|AY| \leqq |A^0Y|$, A^0 is also in C_Y'. Likewise, by a rotation about axis YP, X moves to a point X^0 in $\pi \cap C_B'$. We now shall show that the two lines A^0B and X^0Y are situated in the plane π in a way which contradicts what was proved in cases 2, 3, and 4 of §8.6. We may assume that A^0B is the longer of the two lines. Both X^0 and Y are in C_B'. Also the diamonds $D(A^0, B)$ and $D(X^0, Y)$ intersect at P because $\angle A^0PB = \angle APB > 120°$ and $\angle X^0PY = \angle XPY > 120°$. Then §8.6 shows that A^0 is in $L(X^0, Y)$. However, A^0 is in C_Y' and so is not in $L(X^0, Y)$, a contradiction.

Suppose the n vertices of a minimal tree are known to lie in a D-dimensional set R. Again let U be the union of all diamonds $D(P, Q)$ as P and Q range over R. The $n - 1$ disjoint diamonds about lines of the minimal tree have total volume less than vol (U). As in §8.6, Hölder's inequality now shows

$$\text{length of min. tree} \leqq v^{-1/D}\{\text{vol } (U)\}^{1/D}(n - 1)^{(D-1)/D},$$

where v is the volume of a diamond about a line of unit length:

$$v = D^{-1}(2 \cdot 3^{1/2})^{-(D-1)}s(D - 1),$$

where $s(D - 1)$ denotes the volume of a unit sphere in $(D - 1)$-dimensional space.

Steiner trees on the surface of a sphere have some interest in connection with networks which join together a widespread set of cities. Even the case of three cities located at corners of an equilateral spherical triangle offers some complication. If the triangle is small, the Steiner minimal tree will have a Steiner point, as in the plane. However note that, unlike the situation of §4, there is now a second relatively minimal tree with the same topology. Its Steiner point is at the diametrically opposite point on the sphere. If the triangle is large enough (cities 109° 28′ or more apart), then the angles of the triangle are all larger than 120°, so a tree with just two lines and no Steiner point is another Steiner tree to consider. This tree is the Steiner minimal tree if the cities are more than 114° 41′ apart.

The angle 109° 28′ is just the angle of the great circle arcs between pairs of vertices of a regular tetrahedron. Since these arcs meet at 120°, a Steiner tree for the four vertices of a tetrahedron can be drawn using three of these arcs. However, such a Steiner tree will contain two arcs belonging to the same (triangular) face of the tetrahedron. A shorter Steiner tree is obtained by placing a Steiner point in the center of this face. This tree is probably a Steiner minimal tree, but that has not been proved.

The construction of §6 does not generalize conveniently to the sphere. The locus of points S such that $\angle\ ASB = 120°$ is no longer a circle.

10. The .86603 $\cdots$ conjecture. The length of a minimal tree is an easily computed upper bound on the length of the Steiner minimal tree. To get a lower bound consider, for each set of given points $A_1, A_2, \cdots$, the ratio L_S/L_M of the length of the Steiner minimal tree to the length of the minimal tree. Let ρ be the least upper bound of this ratio as the given points range over all possibilities. If we knew ρ, then ρL_M would be a convenient lower bound for L_S. We shall find that $\rho \geqq \frac{1}{2}$. Actual experience with drawing trees in the plane strongly suggests the result $\rho = 3^{1/2}/2 = .86603 \cdots$, which we conjecture but do not prove.

E. F. Moore gave us the following proof that $\rho \geqq \frac{1}{2}$. If a Steiner minimal tree has length L_S, one can find a traveling salesman tour of $A_1, A_2, \cdots$ of length $2L_S$. To obtain the tour consider the graph G obtained by replacing each line of the Steiner minimal tree by two lines in parallel. Since an even number of lines are incident at every vertex of G, G has an Euler line. The Euler line has length $2L_S$ and is a traveling salesman tour of $A_1, \cdots, A_n$. Then $2L_S \geqq L_T$, the length of the minimal traveling salesman tour. Any line of the minimal traveling salesman tour may be deleted to obtain a tree (without Steiner points). Then the minimal tree has length $L_M < L_T \leqq 2L_S$.

This argument does not depend on properties of the Euclidean plane and, in fact, applies even to arbitrary metric spaces. One can achieve savings approaching 50% if the space is peculiar enough. For example, let points be k-tuples of real numbers $x = (x_1, x_2, \cdots, x_k)$, and let the metric be "Manhattan distance",

$$d(x, X) = |x_1 - X_1| + |x_2 - X_2| + \cdots + |x_k - X_k|.$$

Take $A_1, A_2, \cdots$ to be the k points

$$\begin{aligned} A_1 &= (1, 0, 0, \cdots, 0), \\ A_2 &= (0, 1, 0, \cdots, 0), \\ &\vdots \\ A_k &= (0, 0, 0, \cdots, 1). \end{aligned}$$

Then $d(A_i, A_j) = 2$ $(i \neq j)$ and $m = 2(k - 1)$. However, the tree which connects $A_1, \cdots, A_k$ directly to the point $S = (0, 0, \cdots, 0)$ has length k. Then $\rho \leqq k/(2(k - 1))$, which is close to $\frac{1}{2}$ if k is large.

Better bounds are possible if one uses special properties of the space or of n. For example, if $n = 3$, one obtains $L_S/L_M \geqq \frac{3}{4}$ for any metric. To prove this result let A_1, A_2, A_3 be given, say with $d(A_2, A_3)$ greater than both $d(A_1, A_2)$ and $d(A_1, A_3)$. Then $m = d(A_1, A_2) + d(A_1, A_3)$. If the Steiner minimal tree has length $L_S < L_M$, it must consist of three paths from A_1, A_2, A_3 to a common Steiner point S. Then,

$$\begin{aligned} 4L_S &= 4\{d(A_1, S) + d(A_2, S) + d(A_3, S)\} \\ &= 2\{d(A_1, S) + d(S, A_2)\} + 2\{d(A_2, S) + d(S, A_3)\} \\ &\qquad\qquad + 2\{d(A_3, S) + d(S, A_1)\} \\ &\geqq 2d(A_1, A_2) + 2d(A_2, A_3) + 2d(A_1, A_3) \\ &\geqq 2L_M + d(A_2, A_3) + d(A_2, A_3) \\ &\geqq 2L_M + d(A_1, A_3) + d(A_1, A_2) = 3L_M. \end{aligned}$$

The bound $\frac{3}{4}$ is attained by the case $k = 3$ of the Manhattan metric example cited above.

In the Euclidean plane one obtains the stronger bound $L_S/L_M \geqq 3^{1/2}/2$ when $n = 3$. Note first that $L_S/L_M = 3^{1/2}/2$ if A_1, A_2, A_3 are at vertices of an equilateral triangle. If A_1, A_2, A_3 do not form an equilateral triangle, let S be their Steiner point (if there is no Steiner point, then $L_S/L_M = 1$ and the desired bound is satisfied). Suppose A_1 is the vertex (or perhaps one of the two vertices) closest to S. Then a minimal tree consists of A_2A_1 and A_1A_3 and

$$L_M = |A_2A_1| + |A_1A_3|,$$

while

$$L_S = |A_1S| + |A_2S| + |A_3S|.$$

Let A_2' be the point on the line segment A_2S such that $|A_2'S| = |A_1S|$. Similarly locate A_3 on A_3S such that $|A_3'S| = |A_1S|$. Then,

$$\begin{aligned} L_S &= |A_1S| + |A_2'S| + |A_2A_2'| + |A_3'S| + |A_3A_3'| \\ &= |A_2A_2'| + |A_3A_3'| + 3^{1/2}\{|A_2'A_1| + |A_3'A_1|\}/2 \\ &\geqq 3^{1/2}\{(|A_2A_2'| + |A_2'A_1|) + (|A_3A_3'| + |A_3'A_1|)\}/2 \\ &\geqq 3^{1/2}\{|A_2A_1| + |A_3A_1|\}/2 = 3^{1/2}L_M/2. \end{aligned}$$

In the Euclidean plane no set of points $A_1, \cdots, A_n$ is known for which $L_S/L_M < 3^{1/2}/2$ even when $n > 3$. Thus we conjecture $\rho = 3^{1/2}/2$. To prove this conjecture it would suffice to show that $L_S/L_M \geqq 3^{1/2}/2$ for all full Steiner trees. For, if a Steiner tree is not full, it may be decomposed into full components as in §3.7. Let σ_i be the length of the ith full component and let μ_i be the length of the minimal tree for the same set of given vertices. The union of the separate minimal trees is a tree (perhaps not minimal) for all of $A_1, A_2, \cdots$ and having length $\mu_1 + \mu_2 + \cdots \geqq L_M$. Also $L_S = \sigma_1 + \sigma_2 + \cdots$. Then the conjecture would be proved if one knew $\sigma_i/\mu_i \geqq 3^{1/2}/2$.

In higher dimensional Euclidean spaces L_S/L_M can be smaller. Figure 8 shows Steiner trees which connect the vertices of a tetrahedron, an octahedron, and a cube in 3-dimensional Euclidean space. These trees have L_S/L_M equal to .813052, .819615, and .874231, respectively; the first two are less than $3^{1/2}/2 = .866026$.

One might further conjecture, by analogy with the situation in the plane, that the corners of the regular simplex in D-dimensional Euclidean space provide a minimum of L_S/L_M. No counterexample to this conjecture is known. Some short Steiner trees for D-dimensional simplexes have been found. Their L_S/L_M values appear in Table 2. Further details are omitted

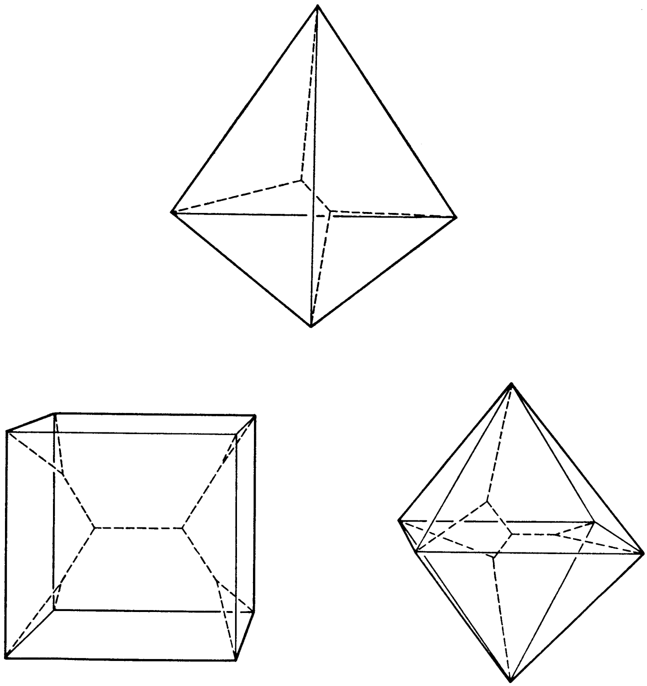


FIG. 8. *Steiner trees for the tetrahedron, cube, and octahedron*

because these Steiner trees are only conjectured to be Steiner minimal trees.

An upper bound on L_S/L_M for D-dimensional simplexes with D odd may be obtained as follows. Let the center of the simplex be called C, i.e.,

$$C = \frac{A_1 + \cdots + A_{D+1}}{D + 1}.$$

For $i = 1, 2, \cdots, (D + 1)/2$ connect the three points S_{2i-1}, A_{2i}, C by a Steiner minimal tree. These $(D + 1)/2$ trees join at C to form a single tree connecting $A_1, \cdots, A_{D+1}$ together via C and the $(D + 1)/2$ Steiner points. A routine calculation of the length of this tree shows that

$$L_S/L_M \leqq (4D)^{-1}\{(D - 1)^{1/2}(D + 2)^{1/2} + 3^{1/2}(D + 1)\}.$$

When $D = 3$ or 5, one obtains the tree of Table 2. When D is large, one can have L_S/L_M near $4^{-1}(1 + 3^{1/2}) = .683010$.

11. Random trees. It is unusual to find a real plane network, such as one which interconnects a given set of cities, having L_S/L_M near .86603. To get a more typical value one may distribute $A_1, \cdots, A_n$ by a suitable

TABLE 2

Length of a Steiner tree joining the vertices of a D-dimensional simplex

D	L_S/L_M
2	$\frac{1}{2}\cdot 3^{1/2} = .866026$
3	$(1 + 6^{1/2})/(2^{1/2}\cdot 3) = .813052$
4	$(\frac{1}{2}\cdot 5^{1/2} + \frac{1}{2}\cdot 3^{1/2} + 6^{1/2})/(2^{1/2}\cdot 4) = .783748$
5	$(6^{1/2} + 3^{3/2})/10 = .764568$
11	$(3^{1/2}\cdot 14 + 6^{1/2}\cdot 3)/44 = .7181178$

random process and then ask for the expected value of L_S/L_M. Some results for the case $n = 3$ follow. Most of them were given to us by J. H. van Lint.

Let θ be the largest angle in the triangle $A_1A_2A_3$. One can show that L_S/L_M for such a triangle is at least as large as L_S/L_M for an isosceles triangle with largest angle θ. Then $L_S/L_M \geqq \cos(\pi/3 - \theta/2)$ if $\pi/3 \leqq \theta \leqq 2\pi/3$, and so

$$E(L_S/L_M) \geqq 1 - P(2\pi/3) + \int_{\pi/3}^{2/\pi 3} \cos(\pi/3 - \theta/2)\, dP(\theta),$$

where $P(\theta)$ is the cumulative probability distribution function for θ.

A simple random way to choose A_1, A_2, A_3 is to place these vertices independently on the circumference of a circle using a constant probability density for each vertex. If α_i is the angle at vertex A_i in the triangle $A_1A_2A_3$, the triple $(\alpha_1, \alpha_2, \alpha_3)$ is uniformly distributed over the set $0 \leqq \alpha_1$, $0 \leqq \alpha_2$, $0 \leqq \alpha_3$, a triangle in the plane $\alpha_1 + \alpha_2 + \alpha_3 = \pi$. Then,

$$P(\theta) = \begin{cases} \{(3\theta/\pi) - 1\}^2, & \pi/3 \leqq \theta \leqq \pi/2, \\ 1 - 3\{1 - \theta/\pi\}^2, & \pi/2 \leqq \theta \leqq \pi, \end{cases}$$

and one obtains $E(L_S/L_M) \geqq .98$.

Another random process picks six independent random numbers from a common Gaussian distribution to serve as Cartesian coordinates of A_1, A_2, A_3. After a long derivation, which we omit, one obtains the probability density function for $\pi/2 \leqq \theta \leqq \pi$. It is

$$P'(\theta) = \{9/\pi D^3\}\{2 \cos\theta \tan^{-1}[(2 + \cos\theta)/D] + D\},$$

where $D = (4 - \cos^2\theta)^{1/2}$. A numerical integration then shows that the probabilities of finding θ in the three ranges 60°–90°, 90°–120°, and 120°–180° are .25, .31, and .44. Then a crude lower bound on $E(L_S/L_M)$ is $.25 \cos 30° + .31 \cos 15° + .44 = .96$.

The vertex at the largest angle of the triangle $A_1A_2A_3$ must lie in the lune about the opposite side. This observation suggests a third random process in which one places a single vertex A_3 at random with constant probability density in the lune about a given line A_1A_2. The probabilities of finding θ in 60°–90°, 90°–120°, and 120°–180° are now .36, .31, and .33. Then $E(L_S/L_M) > .36 \cos 30° + .31 \cos 15° + .33 = .94$.

12. Glossary.

Term	Where defined
A_i (given point)	§1
deciding point, deciding region	§8
degenerate tree	§2
$D(A, B)$ (diamond)	§8
double wedge	§8
$F(n, s), f(s)$ (numbers of topologies)	§7
full ($s = n - 2$)	§3.7
L_M (length of minimal tree)	§10
L_S (length of Steiner minimal tree)	§10
$L(A, B)$ (lune)	§8
minimal tree	§1
n (number of A_i)	§1
relatively minimal tree	§2
s (number of Steiner points)	§3.4
splitting	§2
Steiner minimal tree	§1
Steiner point	§1
Steiner tree	§2
topology of a tree	§2
wedge	§8

REFERENCES

[1] RICHARD COURANT AND HERBERT ROBBINS, *What Is Mathematics?*, Oxford University Press, New York, 1941.

[2] H. S. M. COXETER, *Introduction to Geometry*, John Wiley, New York, 1961.

[3] H. S. EDWARDS AND H. Z. HARDAWAY, *New concepts in exchange outside plant engineering*, Bell System Tech. J., 44 (1965), pp. 373–379.

[4] E. N. GILBERT, *Random minimal trees*, this Journal, 13 (1965), pp. 376–387.

[5] IRWIN GREENBERG AND RAYMOND A. ROBERTELLO, *The three factory problem*, Math. Mag., 38 (1965), pp. 67–72.

[6] M. HANAN, *On Steiner's problem with rectilinear distance*, this Journal, 14 (1966), pp. 255–265.

[7] J. B. KRUSKAL, JR., *On the shortest spanning subtree of a graph and the traveling salesman problem*, Proc. Amer. Math. Soc., 7 (1956), pp. 48–50.

[8] HUA LOO-KENG AND OTHERS, *Applications of mathematical methods to wheat harvesting*, Chinese Math., 2 (1962), pp. 77–91.

[9] HAROLD W. KUHN AND R. E. KUENNE, *An efficient algorithm for the numerical*

solution of the generalized Weber problem in spatial economics, J. Regional Science, 4 (1966), pp. 21–33.

[10] J. C. MAXWELL, Cambridge Philos. Mag., (1864).

[11] ———, Trans. Roy. Soc. Edinburgh, 26 (1869).

[12] Z. A. MELZAK, *On the problem of Steiner*, Canad. Math. Bull., 4 (1961), pp. 143–148.

[13] F. P. PALERMO, *A network minimization problem*, IBM J. Res. Develop., 5 (1961), pp. 335–337.

[14] R. C. PRIM, *Shortest connection networks and some generalizations*, Bell System Tech. J., 36 (1957), pp. 1389–1401.

[15] JOHN RIORDAN, *The enumeration of labeled trees by degrees*, Bull. Amer. Math. Soc., 72 (1966), pp. 110–112.

[16] H. STEINHAUS, *Mathematical Snapshots*, Oxford University Press, New York, 1960.

[17] W. J. VAN DE LINDT, *A geometrical solution of the three factory problem*, Math. Mag., 39 (1966), pp. 162–165.

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

[19] ALFRED WEBER, *Über den Standort der Industrien*, Tübingen, 1909, Chap. III. Translated by C. J. Friedrich as *Alfred Weber's Theory of the Location of Industries*, Chicago, 1929.