

## STEINER'S PROBLEM FOR SET-TERMINALS\*

BY

E. J. COCKAYNE (*University of Victoria*)

AND

Z. A. MELZAK<sup>1</sup> (*University of British Columbia*)

1. Let  $L_1, \dots, L_n$  be  $n$  separate lakes which are to be interconnected by a network of canals of minimum possible length. Or, let  $M_1, \dots, M_n$  be  $n$  metropolitan areas which are to be joined by the shortest possible network of roads. Again, let  $P_1, \dots, P_n$  be  $n$  metal plates in the plane which are to be soldered together by the least possible amount of wire of fixed diameter. Assuming spatial homogeneity, we formulate an abstract problem which underlies the given examples. By a net  $N$  we shall understand a finite set of plane rectifiable arcs; the sum  $L(N)$  of their lengths is the length of  $N$ . Let  $n \geq 3$  and let  $A = \{A_1, \dots, A_n\}$  be a plane set with exactly  $n$  components  $A_i$ ; it will be assumed throughout that each  $A_i$  is compact. Any such  $A$  will be called an  $n$ -terminal set and its  $n$  components will be called terminals. Our problem is:

(ST <sub>$n$</sub> ) given an  $n$ -terminal set  $A$ , to find the shortest net  $N$ , such that  $A \cup N$  is a connected set.

This is a direct generalization of Steiner's problem ( $S_n$ ) in which  $A$  consists of  $n$  points. In previous papers [1] and [2] we have given a solution of ( $S_n$ ); here we generalize our methods to solve (ST <sub>$n$</sub> ).

2. The existence of a minimal net  $N$  follows by a standard continuity and compactness argument, since each arc of  $N$  is clearly a straight closed segment and (as will be shown) there are at most  $2n - 3$  such arcs, and since each terminal is a compact set. Further, there are  $k$  points  $v_1, \dots, v_k$  (called vertices) and  $N$  consists of  $n + k - 1$  straight segments (called branches) each of which connects two terminals, or two vertices, or a vertex and a terminal. By the valency at a vertex  $v_i$  we understand the number  $w(v_i)$  of branches emanating from it. The following properties of the minimal net  $N$  are simply deducible by the same methods as their counterparts for ( $S_n$ ) in [1] and [2];

P1. Two branches can have at most one point in common, it is then their common endpoint and is either a vertex or a point in the boundary of a terminal.

P2. If two branches share an endpoint they subtend there an angle  $\geq 120^\circ$ .

P3. For each  $v_i$  ( $i = 1, \dots, k$ ),  $w(v_i) = 3$ .

P4. Any two of the three branches meeting at a vertex  $v_i$  subtend there the angle  $120^\circ$ .

P5. If a vertex  $v_i$  is joined by a branch to a terminal  $A_j$ , then it is joined to the point of  $A_j$  nearest to it (or to one of such points if there are several); the same holds for two terminals joined by a branch.

P6.  $0 \leq k \leq n - 2$ .

\*Received March 3, 1967; revised manuscript received May 29, 1967.

<sup>1</sup>Presently at Courant Institute for Mathematical Sciences, New York University, New York.

Suppose that the integer  $k$  and the vertices  $v_1, \dots, v_k$  are known. Our problem is then converted to the following one: given  $n + k$  disjoint compact connected sets  $U_1, \dots, U_{n+k}$  ( $= A_1, \dots, A_n, v_1, \dots, v_k$ ), to connect them together by some  $n + k - 1$  branches so that each branch joins two nearest points of a  $U_i$  and a  $U_j$ , and the sum of the lengths of all branches is minimum. The properties P1-P6 are then automatically satisfied.

This is a discrete problem for which an efficient algorithm has been given by Prim [3]:

- (1) Join  $U_1$  to its nearest neighbor, say  $U_2$ ,
- (2) replace  $U_1$  and  $U_2$  by the union of  $U_1, U_2$  and the shortest straight segment joining them,
- (3) repeat the same procedure for the new  $(n + k - 1)$ -terminal set, and keep on repeating until only one set remains. Actually, Prim's algorithm was originally intended for the case of point-terminals; however, it works equally well for any compact connected terminals. Moreover, if the nearest neighbor of some  $U_i$  can be connected to it by several segments of the same minimal length, or if  $U_i$  has several nearest neighbors, we perform the connection in all possible ways and get then the set of all connecting networks of the same minimal length.

Therefore our problem  $(ST_n)$  will be solved if we can show how to find for each  $k$  ( $0 \leq k \leq n - 2$ ) all admissible vertex sets  $V = \{v_1, \dots, v_k\}$  satisfying P1-P5. For once this is done, we can perform Prim's algorithm for each admissible vertex set, augmented by the fixed  $n$ -terminal set, and we find the vertex set(s) leading to the shortest connecting net.

This shows the principle of the proof; actual work will be arranged somewhat differently so as to cut down on the number of geometrical constructions by taking advantage of certain structural properties.

3. The solution of Steiner's problem  $(S_n)$ , given in [1] and [2] depends on the following elementary geometrical lemma which will be referred to as the equilateral construction. Let  $A = \{a_1, a_2, a_3\}$  be the vertices of a triangle  $T$  no angle of which is  $\geq 120^\circ$ . Let  $a_{12}, a_{23}, a_{13}$  be the three third vertices of the equilateral triangles built outward on the sides of  $T$  (with the obvious notation), then the straight segments  $a_{12}a_3, a_{23}a_1, a_{13}a_2$  intersect in a point  $v$  inside  $T$ . Moreover, the straight segments  $a_1v, a_2v, a_3v$  subtend pairwise the angle  $120^\circ$  at  $v$ , so that the minimal net  $N$  for the 3-terminal set  $A$  consists just of those three segments. Finally, we have

$$L(N) = |a_{12}a_3| = |a_{23}a_1| = |a_{13}a_2|. \tag{1}$$

The equilateral construction is of fundamental importance; it enables us to solve  $(S_n)$  and a suitable generalization of it will lead to the corresponding solution of  $(ST_n)$ . The crux of the matter is that the points  $a_{12}, a_{23}, a_{13}$  can be constructed by reference to the given 3-terminal set  $A$  alone.

Suppose now that  $A$  is a given  $n$ -terminal set of  $(S_n)$ ; let the minimal connecting net  $N$  have the vertices  $v_1, \dots, v_k$  and suppose that of the  $n$  terminal points exactly  $n_i$  have valency  $i$  ( $i = 1, 2, 3$ , no terminal can have valency  $\geq 4$  by P2). Counting the total number  $B$  of branches by vertex- and terminal-multiplicities, we get

$$B = (3k + n_1 + 2n_2 + n_3)/2, \quad n = n_1 + n_2 + n_3,$$

so that

$$B = (3k + n + n_2 + 2n_3)/2.$$

On the other hand, a tree with  $n + k$  vertices has exactly  $n + k - 1$  branches and so  $B = n + k - 1$ , which gives us

$$k = n - 2 - n_2 - 2n_3 .$$

This proves P6; further, it shows that if  $k$  has its largest possible value  $n - 2$  then each terminal has valency 1.

We outline now the solution of  $(S_n)$  beginning with the construction at all admissible vertex sets for the case when the  $n$ -terminal set is complete; here  $k = n - 2$ . An example illustrating the construction for  $n = 6$  is given in Fig. 1.

Here  $(ab)$  denotes the third vertex of the equilateral triangle based on  $ab$ , it is supposed that as we move along  $ab$  from  $a$  to  $b$ ,  $(ab)$  is on the right. The equilateral construction enables us to find successively  $(a_6a_5)$ ,  $(a_4a_3)$  and  $(a_2a_1)$ , then  $((a_2a_1)(a_6a_5))$ , and finally the straight segment  $[(a_2a_1)(a_6a_5)]$ ,  $(a_4a_3)$  which we call an axis for the net  $N$  in question. Observe that the net  $N$  of Fig. 1 has also other axes, for instance,  $[(a_2a_1), ((a_6a_5)(a_4a_3))]$ . Repeated use of (1) shows that both these axes have length  $L(N)$ . In the general case we start with the given  $n$ -terminal set  $A = \{a_1, \dots, a_n\}$  and no axis is a priori known; however, the number of possible orderings and bracketings of  $a_1, a_2, \dots, a_n$ , which are admissible as axes, is clearly finite. Thus, when  $A$  is given, we construct all possible axes and after finding the shortest one(s) we reconstruct therefrom the minimal net(s).

If  $k < n - 2$  in the minimal net, we can split  $A$  into certain subsets called components, and we find the minimal subnets over each component separately; these are joined then as in Prim's algorithm. Figure 2 shows an example of a 9-terminal set which is not complete. Here there are two components: a 3-terminal component  $\{a_2, a_3, a_4\}$  and a 4-component  $\{a_6, a_7, a_8, a_9\}$ . Each component is complete, in the sense of having the largest possible number of vertices (= number of its terminals - 2). In addition,

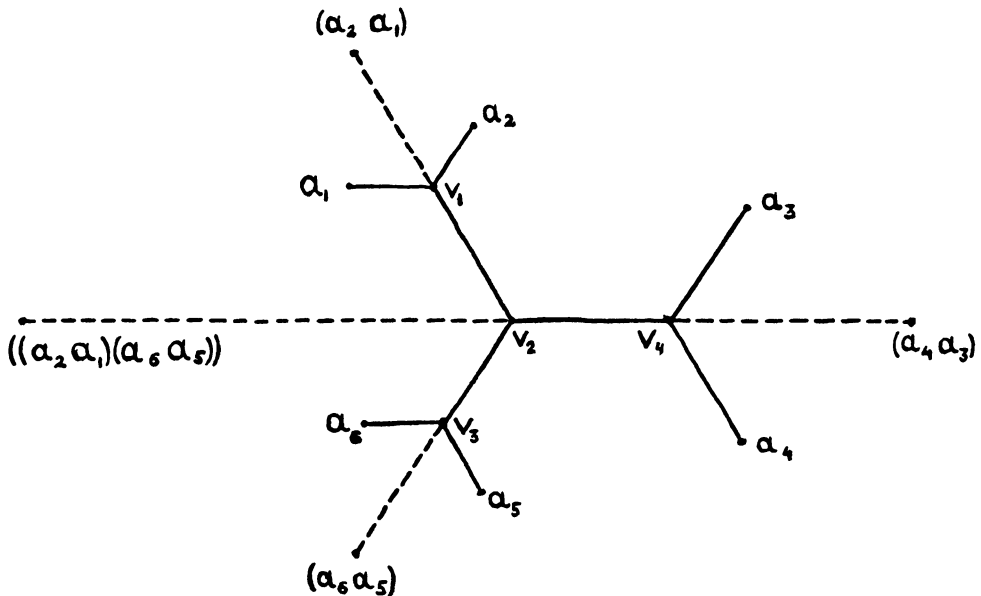


Fig. 1

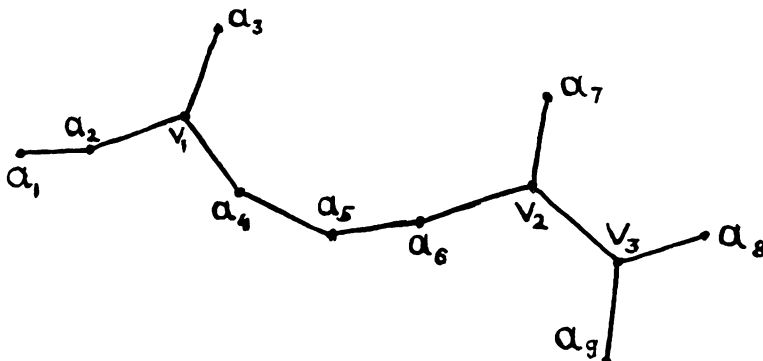


Fig. 2

there are two residual terminals,  $a_1$  and  $a_5$ . The minimal subnet (or the set of all minimal subnets) is constructed over each component separately, in the same way as before.

We do not know how to decompose  $A$  into components a priori. However, we observe that there is a finite number of such possible decompositions. Specifically, let  $A = \{a_1, \dots, a_n\}$ , and let

$$A = A_1 \cup A_2 \cup \dots \cup A_p \cup R; \quad (2)$$

(2) is called a decomposition if: (a) the set  $A_i$  has  $n(i) \geq 3$  points, (b) for  $i \neq j$ ,  $A_i$  and  $A_j$  have at most one point in common, (c) each  $a_k$  belongs to at most three sets  $A_i$ , (d)  $R$  is disjoint from  $\bigcup_i A_i$ . If, in addition, each set  $A_i$  in (2) considered as an  $n(i)$ -terminal set is complete, i.e., if a minimal net with  $n(i) - 2$  vertices exists for it, we call the decomposition (2) admissible. If, further, there are exactly  $k$  vertices in an admissible decomposition, we call it an admissible  $k$ -decomposition.

The procedure is now clear. For a given  $n$ -terminal set  $A$  and a given integer  $k$  we construct first all admissible  $k$ -decompositions; their number is certainly finite. For each such admissible  $k$ -decomposition we construct the minimal subsets over each component separately; finally, these minimal subnets over components and the residual set  $R$  are joined together by Prim's algorithm. When this is done for  $k = 0, 1, \dots, n - 2$ , we have a finite number of connecting networks which must contain the minimal net of  $A$  (or all the minimal nets of  $A$ , if there are several). We observe a stability property of components: if a terminal of a component is perturbed by a sufficiently small amount then, in general, only the minimal subnet over that component changes. There are, however, some exceptional cases when two components share a terminal of valency 2, and the two branches meeting there subtend an angle  $120^\circ$ , or when similar circumstance occurs with respect to the residual set.

4. In this section we prove a lemma which generalizes the equilateral construction to the case of a 3-terminal set, with sets as terminals. When  $a$  and  $b$  are points, let the point  $(ab)$  be as defined before; suppose now that  $A$  and  $B$  are compact connected disjoint sets and put

$$(AB) = \{(ab) : a \in A, b \in B\}.$$

Then  $(AB)$  is likewise a compact connected set, it may be called the equilateral sum of  $A$  and  $B$ . However, the sets  $(AB)$  and  $(BA)$  do not coincide. We list some elementary

properties of the equilateral sum: if  $A$  is a point then  $(AB)$  and  $B$  are congruent under a rotation of  $60^\circ$  about  $A$ ; if  $c$  is an extreme point of  $(AB)$  then  $c = (ab)$  where  $a$  and  $b$  are extreme points of  $A$  and  $B$  respectively; the distributive laws hold: if  $A = C \cup D$  then  $(AB) = (CB) \cup (DB)$ , the same for  $B = C \cup D$ . The following properties hold for  $(AB)$  whenever  $A$  and  $B$  have them; convexity, arcwise-connectedness, being the smooth boundary of a region, being a simple polygon. Let  $d(X, Y)$  denote the distance between two compact disjoint sets  $X$  and  $Y$ .

**LEMMA 1.** *Let  $A_1, A_2, A_3$  be three compact connected pairwise disjoint sets. Suppose that a minimal connecting net  $N$  for the 3-terminal set  $A = \{A_1, A_2, A_3\}$  consists of three straight segments  $a_1v, a_2v, a_3v$  ( $a_i \in A_i$ ) meeting at a vertex  $v$ , and let the rotation  $a_1 \rightarrow a_2 \rightarrow a_3 \rightarrow a_1$  be counter-clockwise about  $v$ . Then  $(A_1A_2), (A_2A_3), (A_3A_1)$  contain points  $(a_1a_2), (a_2a_3), (a_3a_1)$  closest respectively to  $A_3, A_1, A_2$ ; the three segments  $a_3(a_1a_2), a_1(a_2a_3), a_2(a_3a_1)$  intersect at  $v$ ; and finally*

$$L(N) = d(A_1, (A_2A_3)) = d(A_2, (A_3A_1)) = d(A_3, (A_1A_2)) \quad (3)$$

To prove it we note that the function  $f(x) = \sum_1^3 d(x, A_i)$  attains its minimum for  $x = v$ ; we have then  $d(v, A_i) = |va_i|$ , with  $a_i \in A_i$ . Applying the equilateral construction to the triangle  $a_1a_2a_3$  leads at once to the proof of Lemma 1.

5. Given an  $n$ -terminal set  $A = \{A_1, \dots, A_n\}$ , we solve the problem  $(ST_n)$  by the same sequence of steps as in  $(S_n)$ . The higher equilateral sums, e.g.,  $((A_2A_1)(A_6A_5))$ , are unambiguously defined, and by an axis, e.g.,  $[((A_2A_1)(A_6A_5)), (A_4A_3)]$  we understand the straight segment joining the points in  $X = ((A_2A_1)(A_6A_5))$  and in  $Y = (A_4A_3)$  for which the minimum  $d(X, Y)$  is attained. We note the following

**THEOREM 1.** *Let  $A = \{A_1, \dots, A_n\}$  be an  $n$ -terminal set and suppose that each  $A_i$  is a simple polygon. Then the minimal connecting net  $N$  can be found by a finite sequence of Euclidean constructions (employing ruler and compass in the traditional sense).*

To prove this it suffices to observe that:

(a) the equilateral sum of two polygons is a polygon and hence constructible by Euclidean means, and

(b) the (closest) distance between two polygons can be found by Euclidean means (for it is the distance between a pair of vertices).

Theorem 1 leads immediately to

**THEOREM 2.** *Let  $A = \{A_1, \dots, A_n\}$  be an  $n$ -terminal set and suppose that each  $A_i$  is arbitrarily well approximable by simple polygons. Then the minimal connecting set  $N$  can be found by a finite sequence of Euclidean constructions to within arbitrary accuracy.*

For if  $A_1$  and  $A_2$  are approximated by the polygons  $P_1$  and  $P_2$ , then  $(A_1A_2)$  is approximated by  $(P_1P_2)$ .

6. The effective algorithm for our problem, described in the previous sections, is in its present form not very practicable for the cases when  $n$  exceeds 10–15. Of course, one can take advantage of various minimizing conditions to exclude many of the possibilities for admissible vertex sets and admissible decompositions. The reader interested in such matters is referred to the forthcoming paper "Steiner Minimal Trees" by F. N. Gilbert and H. O. Pollak, to appear shortly in the S.I.A.M. Journal.

7. The second author (Z.A.M) acknowledges gratefully the help of the National Research Council (Ottawa) in the form of a Senior Fellowship for 1966–67. Both authors wish to thank the referee and E. N. Gilbert of the Bell Telephone Laboratories, for constructive criticism and improving suggestions.

#### REFERENCES

1. Z. A. Melzak, *On the problem of Steiner*, Canad. Math. Bull. 4, (2), 143–148 (1961)
2. E. J. Cockayne, *On the Steiner problem* (to appear in Canad. Math. Bull.)
3. R. C. Prim, *Shortest connecting networks*, Bell System Tech. Jr. 31, 1398–1401 (1957)