Discussion Paper No. 635R

AN IMPROVED ALGORITHM FOR STEINER TREES

by

Dan Trietsch*

December 1984
Revised May 1985

---

*Visiting Assistant Professor, Department of Managerial Economics and Decision Sciences, J. L. Kellogg Graduate School of Management, Northwestern University, 2001 Sheridan Road, Evanston, Illinois 60201.

The author gratefully acknowledges very valuable comments by Gabriel Y. Handler (Faculty of Management, Tel Aviv University) and by Frank K. Hwang (AT&T Bell Laboratories, Murray Hill, New Jersey).

An Improved Algorithm for Steiner Trees

Abstract

We show two simple results for the Euclidean Steiner minimal tree problem: (i) all admissible partitions of the set of nodes (which the tree should span) correspond to degenerate full Steiner configurations, and can be systematically derived from these if degeneracy occurs; (ii) for any given full Steiner tree configuration there exists a lower bound on the value of the corresponding Steiner tree (this bound holds as an equality if the tree is full). Taken together these two results suggest a relatively efficient (though still exponential) branch and bound algorithm for the Steiner minimal tree construction. The algorithm uses the lower bound to discard some configurations after a brief check, and does away with the need to check all the possible partitions of the nodes set.

1. Introduction

The (Euclidean) Steiner tree problem [9,7] is to find the minimal spanning tree for a set of nodes $N = \{1,2,...n\}$, while using additional points (called Steiner points) if and where their inclusion would make the tree shorter.

A Steiner tree is a local minimal solution to the above problem. Steiner trees have up to $n - 2$ Steiner points, each of rank three; all angles between adjacent arcs are 120° at least; the arcs are straight and may not intersect each other. However, for $n > 3$ there may be (exponentially) many different Steiner trees, and it is difficult to find the Steiner minimal tree (SMT) for large sets unless the nodes are arranged very conveniently.

A Steiner tree with $n - 2$ Steiner points exactly is called full. In a full Steiner tree, all the nodes of N are leaves (i.e., of rank one). In a Steiner tree which is not full, some nodes of N are of rank two or (very rarely) three. A Steiner tree which is not full can always be partitioned to full subtrees (where a direct arc is considered a full Steiner tree for $n = 2$). In this text we shall refer to Steiner trees which are not full as degenerate. Also, we shall refer to any tree with $n - 2$ additional points of rank three, and where no arcs intersect as having a full Steiner configuration (note that we do not require the angles to be 120° at least). It turns out that some full Steiner configurations can be actually realized as full Steiner trees, while others would degenerate, i.e., some "would-be" Steiner points merge with nodes of N, or with each other. (If two Steiner points merge with each other, we obtain an additional point of rank four, which calls for the insertion of two Steiner points in its stead, but in such a manner that the configuration is changed. Therefore, in such cases the resulting Steiner tree can be associated with another full configuration.)

Cockayne [3] has introduced the "Steiner polygon" which is guaranteed to contain the SMT. If the Steiner polygon intersects itself, the problem can be partitioned to smaller independent subproblems (very advantageously). In this text we will assume that the Steiner polygon is not self-intersecting. This assumption is, obviously, not restrictive at all. To continue, if all the nodes of N are on the Steiner polygon, Cockayne showed that there are $(2(n-2))!/((n-2)!(n-1)!)$ candidate full configurations (see also [10]). It should be noted, however, that some of these configurations may not be realized as full Steiner trees due to degeneracy. The case is more complicated if some nodes of N are internal to the Steiner polygon. Those nodes of N <u>on</u> the polygon have a natural cyclic order, but internal points may be inserted between them in numerous ways. Cockayne concluded that there would be $(2(n-1))!/((n-2)!(n-i-1)!)$ full configurations if i points are within the Steiner polygon.

An algorithm capable of solving the minimal Steiner tree problem was proposed by Cockayne and Schiller [4], and later streamlined by Boyce and Seery [1]. The basic idea is to check all the possible partitions of N which conform to a long list of requirements (designed to discard some partitions faster), and for each part of these partitions to solve for the <u>minimal full</u> Steiner tree, finally settling on the partition for which the sum of the full partial trees is minimal. It is easy to show that such a scheme must converge to the optimum. Much time is expended in generating and checking all the partitions and solving for the parts. (On the other hand, only full trees or subtrees need be constructed in a direct manner, which saves some time in that part.)

In this paper we propose an algorithm which does away with the need to generate partitions of N, by identifying degenerate Steiner trees associated

with full configurations which cannot be realized as full Steiner trees. This should be an improvement by itself, but in addition we propose a lower bound on the length of any Steiner tree (full or degenerate) associated with a given full configuration, and this makes it possible to discard some full configurations after a brief check, without having to check whether they may be realized as full Steiner trees or not.

In the next section we discuss the Steiner construction and its extension to a contribution made by F. Heinen in 1834, which will be very useful for us below. After that we will proceed to establish the lower bound, show a procedure designed to identify degenerate Steiner trees associated with given full configurations (if they cannot be realized as full Steiner trees), and finally present the resulting algorithm.

The Generalized Heinen Dual Steiner Problem

Take the simplest case first, namely, $N = \{1,2,3\}$ (i.e., a triangle $\Delta 1,2,3$), and the Steiner construction is executed as follows: (i) Choose a pair of nodes in cyclic order, such as 1 and 2 (or 2 and 3, or 3 and 1) and represent them by point (1,2) which is located at the apex of the equilateral triangle $\Delta 1,2,(1,2)$ constructed outside of $\Delta 1,2,3$ (see Figure 1). (ii) Connect point (1,2) with node 3 (or (2,3) with 1, or (3,1) with 2). (A segment such as $\overline{(1,2),3}$ is called a Simpson line [8].) Circumscribe $\Delta 1,2,(1,2)$ by a circle, and denote the point where it intersects the segment $\overline{(1,2),3}$ as point 4. Now, if point 4 lies within $\Delta 1,2,3$ (or even coincides with one of the nodes), the Steiner tree is obtained by connecting each node of N with (the Steiner) point 4. On the other hand, if 4 is strictly out of $\Delta 1,2,3$, the Steiner tree is degenerate and consists of the sides of the obtuse angle of the triangle $\Delta 1,2,3$. (Note that this angle is greater than 120°.)

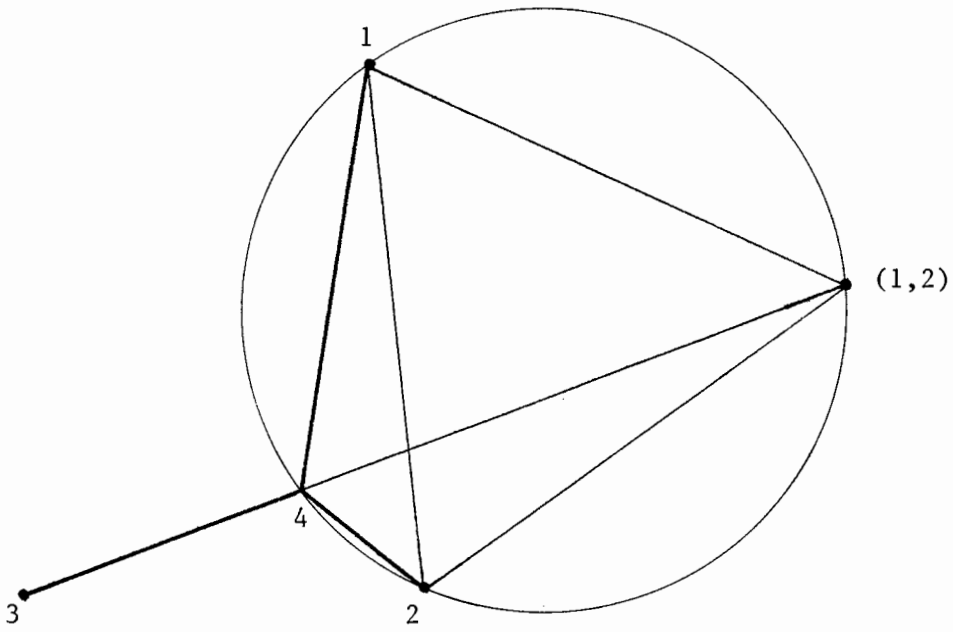
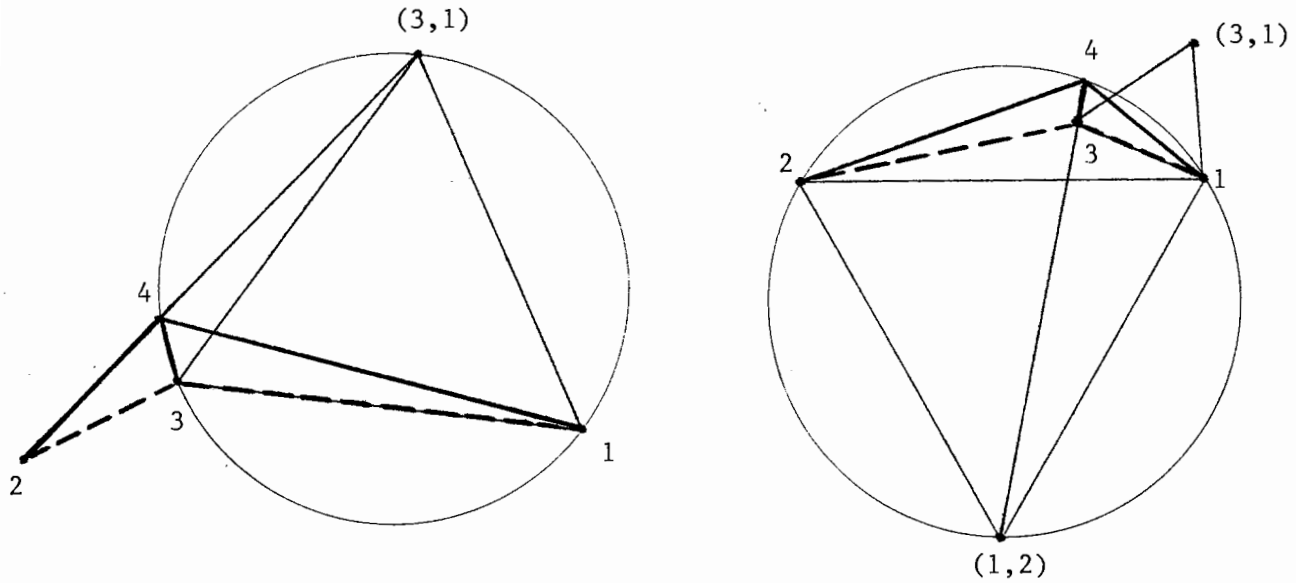


Figure 1

According to Kuhn [8], the first person to solve the orignal Fermat problem (of minimizing the total distance to the three vertices of a triangle) completely was F. Heinen. (The Fermat problem is, obviously, "Steiner's" problem for n = 3.) Heinen's contribution was the solution of the degenerate case, which was neglected by earlier contributors (such as Torricelli [before 1640], Cavalieri [1647], and Simpson [1750]). In connection with the degenerate case, Heinen showed that if the triangle has an angle greater than 120° the Steiner construction would yield a point out of the triangle which (locally) minimizes the sum of its distances to the two acute angled vertices minus the distance to the third (obtuse) vertex. Figure 2 illustrates this point, where we minimize d(4,1) + d(4,2) - d(4,3). The figure has two parts

reflecting the two possible ways in which a degeneracy may manifest itself, as a result of our choice of which two vertices to represent. The respective minimal Steiner trees are shown in broken lines in the figure.



a b

Figure 2

A more careful analysis, however, shows that the minimum obtained for the difference above is not the global minimum. To obtain the global minimum one should proceed with the Steiner construction in the "wrong" way, i.e., construct the equilateral triangle on the same hemisphere as the original triangle $\Delta 1,2,3$. For our purpose we do not want this real minimum, but rather the local one. However, we shall have to devise a check to make sure that we have a "legitimate" version of the problem. This can be done by checking if

the Heinen point and point 3 are or are not on the same side of the line connecting 1 and 2. For our purpose they should indeed be on the same side (i.e., yield two weak linear inequalities of the same sign). This can be checked for very easily. The example at the end of the paper illustrates this point, and shows why the inequalities should be weak.

It turns out that it is very easy to generalize Heinen's Dual Steiner Problem for any given full Steiner topology which proves to be degenerate. One simply carries out the Steiner construction in the usual manner, but without stopping upon degeneracy. Then, if any angles of 60° are found at some Steiner points, this indicates that the solution minimizes the sum of all the edges which form at most one angle of 60° minus those edges which form two angles of 60° (at the Steiner points). The generalized Heinen problem is to locate the Steiner points (whether they are "proper" or not) according to the given configuration. It is possible to ascertain that for a full Steiner configuration which proves degenerate, any of the legitimate Steiner constructions would yield the same "Heinen tree."

We now show that the Heinen objective function is a lower bound on the value of the Steiner tree with the same "parent" full topology.

A Lower Bound for Steiner Trees Associated with a Given Full Topology

It is well documented [7], that if the Steiner construction does not degenerate, the "length" of the Steiner tree is equivalent to that of the Simpson line, e.g., the segment $\overline{(1,2),3}$ (which we may denote as d((1,2),3) or, even more explicitly for the nondegenerate $N = \{1,2,3\}$ case: $|MST| = d((1,2),3)$. Due to symmetry, it follows that $d((1,2),3) = d(1,(2,3)) = d((3,1),2)$. Lemma 1 below states a similar result for the degenerate case as well. Also, the reader can easily verify that in the degenerate case this value is equal to the "length" of the Heinen tree.

Before we proceed, however, let us state briefly that for $|N| > 3$ the Steiner construction of any given full configuration is carried out by iterative representation of consecutive nodes or representative points, similar to the representation described above, until a Simpson line is obtained, and then "folding backwards" to obtain the tree. The various full configurations correspond to the various possible pairings of the nodes. It is customary to keep node n as an end point of the final Simpson line (e.g., $\overline{(1,2),3}$; but other representations and pairs exist for the same topologies, such as $\overline{(2,3),1}$ or $\overline{(3,1),2}$). We also refer to segments such as $\overline{(1,2),3}$--where a pair of parenthesis signifies that the two points within it are represented by the apex of an equilateral triangle constructed in the manner discussed above--as Cockayne notations. Another example of a Cockayne notation might be $\overline{(1,((2,3),(4,5))),6}$, where $N = \{1,2,3,4,5,6\}$ (see Figure 3); here nodes 2 and 3 are represented by (2,3), nodes 4 and 5 and represented by (4,5), (2,3) and (4,5) are represented by ((2,3), (4,5)) and, finally, 1 and ((2,3), (4,5)) are represented by (1,((2,3), (4,5))) to obtain the indicated segment. When performing the construction for this notation, we draw the segment from 6 to the point intersecting the circle circumscribing the equilateral triangle $\Delta 1$, ((2,3),(4,5)), (1,((2,3),(4,5))) associated with the last representation, where a Steiner point, say point 9, is located; we now have two segments, $\overline{1,9}$ and $\overline{((2,3),(4,5)),9}$ which we treat in a similar manner, and so on.

We are now ready to state and prove some results.

Lemma 1. All possible representations of the same full Steiner tree configuration yield Simpsons lines of equal length--whether the construction degenerates or not.

Note. The lemma is well-known for the nondegenerate case. The proof for our more general case is essentially the same, and we present it for completeness.

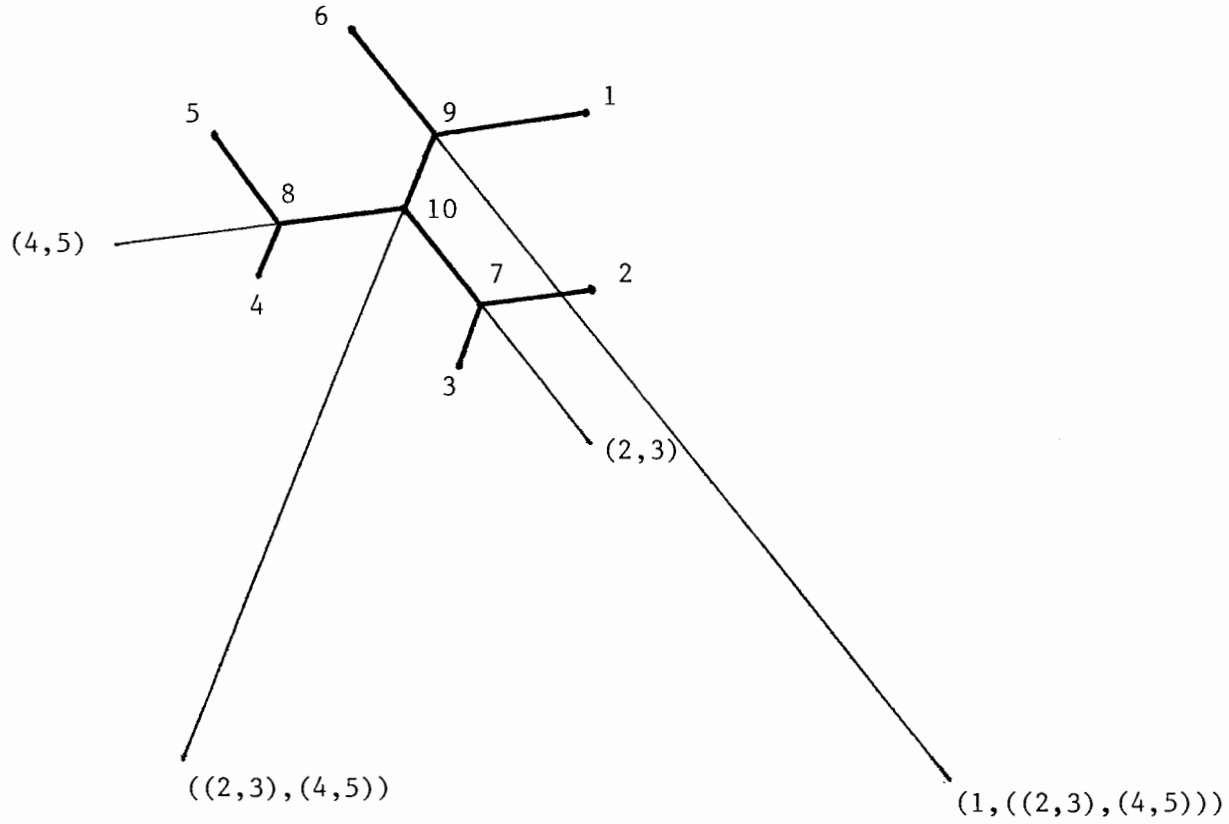


Figure 3

Proof. We first prove for $|N| = 3$, and proceed by induction. For $N = \{1,2.3\}$

we have to show that $d((1,2),3) = d(1,(2,3)) = d((3,1),2)$. By construction we have (see Figure 2-b):

$$d((1,2),1) = d(2,1),$$

$$d(1,3) = d(1,(3,1)),$$

$$\sphericalangle(1,2),1,3 = \sphericalangle 2,1,(3,1) = \sphericalangle 3,1,2 + 60°.$$

$$\Rightarrow \Delta(1,2),1,3 \simeq \Delta 2,1,(3,1);$$

$$\Rightarrow d((1,2),3) = d((3,1),2).$$

And, similarly, $d((1,2),3) = d(1,(2,3))$.

Now assume the lemma holds for $|N| = k - 1$; $k = 4,5,\ldots$, and we have to prove for $|N| = k$, for a given full Steiner configuration. But, for any such configuration there is at least one pair of consecutive nodes, say $u,v \in N$, which are connected by direct links to a common Steiner point (see Lemma in [10]). Take this pair, and represent it by point (u,v). Clearly we now have $k - 1$ points, and for all the alternative full configurations defined in this reduced set, the assumption implies that the resulting segments are equal. Now take one of these alternatives where point (u,v) is an end point, and another point, say x, represents the other $k - 2$ nodes of N; there is exactly one such segment $\overline{(u,v),x}$. We now have a triangle $\Delta u,v,x$ and two more alternative segments, which are the only ones we did not consider before, namely, $\overline{u,(v,x)}$ and $\overline{(x,u),v}$ but this is the case of three points so these two must be of the same length as $\overline{(u,v),x}$, and all the other representations which include (u,v) as a pair. □

Note. The proof implies that there are $2n - 3$ possible Simpson lines or respresentations for each full configuration. This is true since there are n nodes, $n - 2$ Steiner points, and hence $2n - 3$ edges in a full Steiner tree.

For each such edge, a unique representation exists.

Lemma 2. For $N = \{1,2,3\}$, $|MST| \geq d((1,2),3)$; or, in words: the length of the minimal Steiner tree for three nodes is not less than the Simpson line obtained by the Steiner construction.

Proof. Without degeneracy $|MST| = d((1,2),3)$, so it suffices to prove the degenerate case. By Lemma 1, $d((1,2),3) = d(1,(2,3)) = d((3,1),2)$, so without loss of generality we may assume that the degeneracy is as depicted in Figure 2-a. In that case:

$$|MST| = d(1,3) + d(2,3).$$

Also, by construction

$$d((3,1),3) = d(1,3)$$
$$\Rightarrow |MST| = d((3,1),3) + d(1,2).$$

But, points (3,1), 2 and 3 form a triangle, and by the triangle inequality we obtain:

$$d((3,1),2) \leq d((3,1),3) + d(2,3) = |MST|. \quad \square$$

Using the two lemmas, we can now state and prove our lower bound theorem:

Theorem 1. Let ST be any Steiner tree with a given full configuration or a degeneragte version thereof, and let d denote the length of each of the Simpson lines associated with it, then $|ST| \geq d$.

Proof. If there is no degeneracy, the theorem holds as an equality, so we may assume degeneracy, and proceed by induction. By Lemma 2 the theorem holds for $n = 3$. Assume it holds for $n = k - 1$; $k = 4,5...$, and it remains to prove for $n = k$. Since there is degeneracy, there exists a node $v \in N$ such that rank $(v) \geq 2$. If rank $(v) = 2$ then $|ST|$ is the sum of two partial Steiner trees--each of which may be full or degenerate in itself, and for each of which the theorem holds by assumption (since none of them has more than $k - 1$ nodes of N), and it is easy to see that our required result follows for the sum, too. Now, suppose that the only vertex $v \in N$ of rank more than one is of rank three (the only case not yet proved), then angles of 120° exactly must be formed there, and three full Steiner subtrees are connected to it. Represent each of these by a segment including v as an endpoint, and denote the other endpoints as x_1, x_2 and x_3. Clearly, v is (also) a Steiner point for $\Delta x_1,x_2,x_3$ (see Figure 4) and our situation is exactly the same as it would be if x_1, x_2, x_3, and v were nodes of N (i.e., $N = \{x_1,x_2,x_3,v\}$, which we did not stipulate. Due to symmetry, we may assume without loss of generality that the configuration is indicated by the segment $\overline{((x_1,x_2),x_3),v}$. But v is a Steiner point of $\Delta x_1,x_2,x_3$, hence the length of our Steiner tree is $d((x_1,x_2),x_3)$, or any other side of the equilateral triangle $\Delta((x_1,x_2), x_3, ((x_1,x_2),x_3)$. On the other hand, $d((x_1,x_2),x_3), v)$ is the distance between an apex of this triangle to its opposite side, and our result follows immediately. □

Having established a lower bound on the value of Steiner trees associated with given full configurations, we proceed to investigate the admissible partitions of the set N, and their relationship to degenerate trees associated with full configurations.

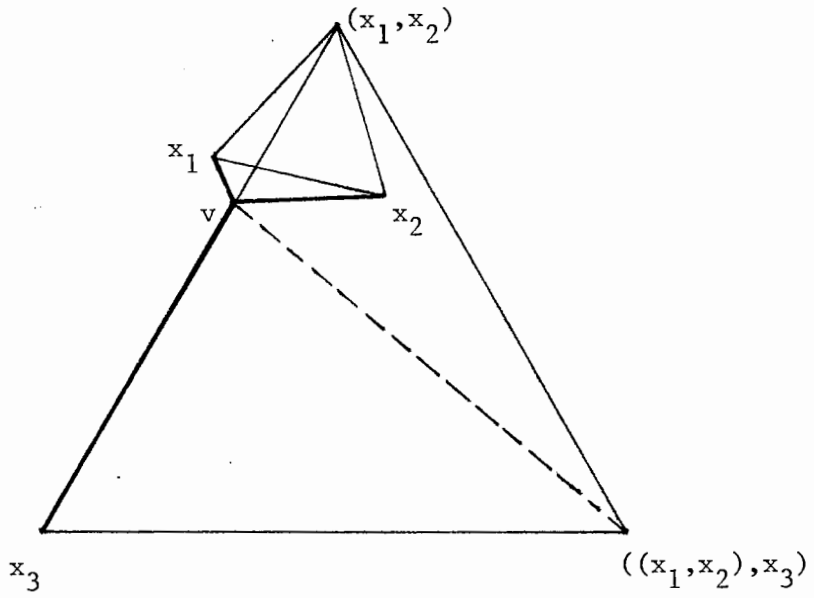


Figure 4

Full Configurations and Admissible Partitions

Obviously, when a full configuration degenerates, a partition (or division) of N is implied. However, it is important to show that any partition which might be implied by the SMT, is associated with a full configuration which degenerates to it, or at least one such full configuration. (If the SMT contains a node $v \in N$ s.t. rank (v) = 3, then there would be more than one full configuration which degenerates to it. Otherwise there is just one.) This result will enable us to consider only such partitions which are implied by degeneracy of full configurations, and

achieve considerable savings in computational time. To that end we first define admissible partitions as follows:

Definition. A partition of N is called admissible, if upon constructing SMTs for its parts we obtain a Steiner tree for N (i.e., all angles are 120° at least, and no arcs intersect), and each such SMT is full (relative to its subset of N).

As an example, take a rectangle of 1 × 5 (i.e., N = {1,2,3,4}, as in Figure 5-a.

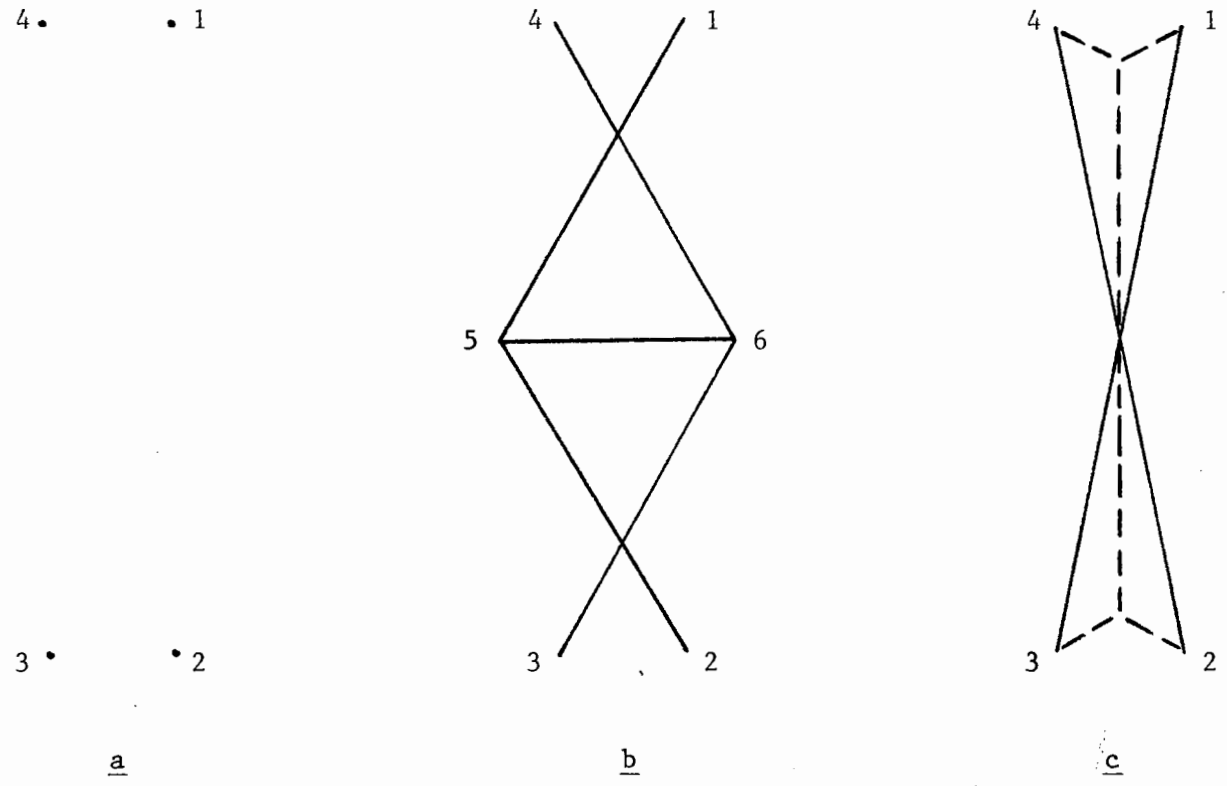


Figure 5

For this case there is no admissible partition, except N itself. However, Cockayne and Schiller's algorithm might generate 24 more partitions for it, such as $\{1,2\}$, $\{2,3,4\}$; $\{1,4\}$, $\{2,3,4\}$ (and six similar partitions to a triangle and a segment); or $\{1,2\}$, $\{2,3\}$, $\{3,4\}$ (and 15 similar partitions to three segments).

We proceed to show that every admissible partition implies at least one degenerate full configuration, and every degenerate full configuration either implies an admissible partition or is not optimal (i.e., the SMT can be found even if we ignore it).

Take any admissible partition, and construct the SMTs for its parts. By the definition of an admissible partition we obtain a Steiner tree. For a while assume that all the nodes of N are of rank one or two (we will discuss the rank three case later). For each $v \in N$ s.t. rank $(v) = 2$, iteratively, draw the bisector of its angle, and mark a point s on it at a distance of ε ($\varepsilon > 0$, but small enough); now connect s by two new links to the two points adjacent to v in ST and erase the two old edges which were incident to v (before we drew the bisector). Now s (and all similar points thus obtained) is, topologically speaking, a Steiner point and we have obtained a full Steiner configuration tree. Furthermore, since for any given topology (and hence for any given configuration) the problem of locating the Steiner points so as to minimize the total length of the tree is strictly convex, it has a unique solution; and it follows that if we start with the full configuration obtained by our procedure, and look for a Steiner tree (full or not) to conform with it, we will obtain our original Steiner tree. (The uniqueness of the Steiner tree for a given configuration has an important role in the algorithm.)

It remains to discuss the case of rank three nodes. Such nodes represent

the rare occurrence where two Steiner points are merged into the same node of N. It follows that in order to obtain a full Steiner tree, we have to insert a pair of interconnected Steiner points for each node of rank three; the four "free" links associated with the pair would be connected to the rank three node itself and to the three points adjacent to it. This can be done in more than one way. (Towards the end of the paper we give an example of how our algorithm would identify an SMT with such a rank three node--depicted in Figure 9.)

We have shown that any Steiner tree, and particularly the SMT, can be obtained from one of the full configurations defined for N. This is enough in order to proceed and describe a suggested technique of doing it, and present our algorithm. However, we should note that <u>not</u> every full configuration implies a Steiner tree. We illustrate this point by the example (mentioned above) where $N = \{1,2,3,4\}$ is the set of vertices of a rectangle, depicted in Figure 4. Take the full configuration implied by $\overline{(1,2),\ (3,4)}$ (we could use the notation $\overline{((1,2),3),\ 4}$ instead), as Figure 4-b shows, this full configuration gives rise to a Heinen tree with the negative link connecting the two Steiner points. Now, if we look for the solution of the convex problem implied, namely, to locate points 5 and 6 so that the sum $d(1,5) + d(2,5) + d(3,6) + d(4,6) + d(5,6)$ will be minimized, we obtain the tree shown in part c of the figure, where points 5 and 6 merge with each other. This is <u>not</u> a Steiner tree, since the additional point is of rank four. Indeed, if we insert two Steiner points instead (as depicted by dotted lines in Figure 5-c), we obtain the SMT (for this case). However, the SMT is associated with another full configuration, namely $\overline{(4,1),\ (2,3)}$ (or $\overline{(1,(2,3)),\ 4}$).

We now describe a procedure designed to find Steiner trees associated with degenerate full Steiner configurations, if they exist. Our procedure

will not find a Steiner tree for cases where the solution of the convex problem is not a Steiner tree, such as the case illustrated above and the "wrong" direction case, however it will identify these cases, and we can simply dismiss them then and there. The procedure is as follows:

Obtaining Steiner Trees from Full Configurations

(a) For a given full configuration (defined on N or any subset of N), execute the Steiner construction to obtain a Steiner tree or a Heinen tree.

(b) Look for degeneracies. (These can be identified by angle tests, since "negative" links form angles of 60° on both sides.) If none exist, go to (f); else, go to (c).

(c) For a given Heinen tree, look for the longest "negative" link connecting a Heinen point to a node of N. If no such link exists (i.e., the "negative" links connect only Heinen points), stop (the full topology does not yield a Steiner tree). Else, go to (d).

(d) The partition implied by the longest "negative" link, is that a Steiner point merges with the node of N incident to it. Obtain the notations of these two parts of the tree, and then go to (e).

(e) Check if the longest negative link is associated with a legitimate Steiner construction (i.e., if the Heinen point and the node it is connected with form weak linear inequalities of the same sign with the line connecting the two other points as obtained in (d)). If the construction is legitimate, return to (a) with each of the two parts separately; else, stop (the full topology does not yield a Steiner tree).

(f) Check all angles at the degenerate points. If all are 120° at least, stop (sucess). If any of them is less than 120°, pick the

most acute one, identify the full subtree associated with inserting a Steiner point for it (as in the greedy hueristic of [2]), and perform steps (a) to (e) on the subtree. (This will imply a later return to (f), but the procedure is finite.)

Before proceeding any further, it may be instrumental to show, by counterexamples, that the simple procedure of just eliminating all the "negative" links in one stroke, may fail. Our first counterexample, depicted in Figure 6, shows that one of the "negative" links, namely $\overline{2,5}$ may "turn positive" when a longer "negative" link, namely $\overline{3,6}$ is eliminated.

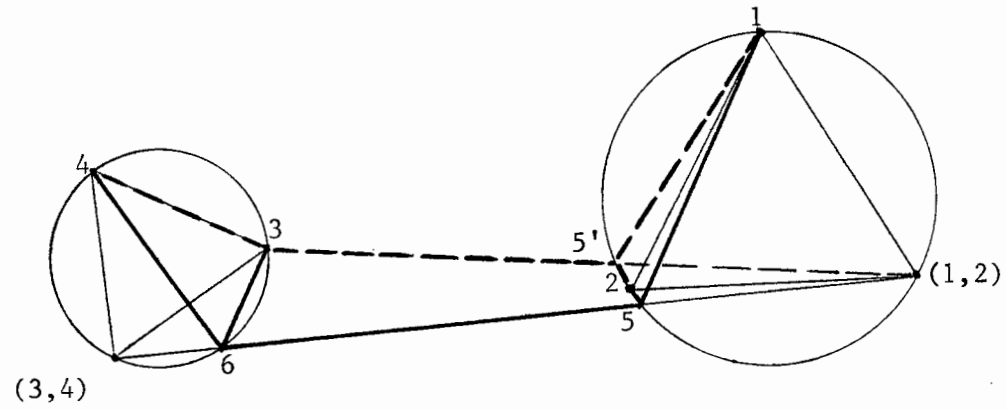


Figure 6

The second counterexample, depicted in Figure 7, shows that a new "negative" link, namely $\overline{1,5}$, may appear after another, again longer, "negative" link is eliminated.

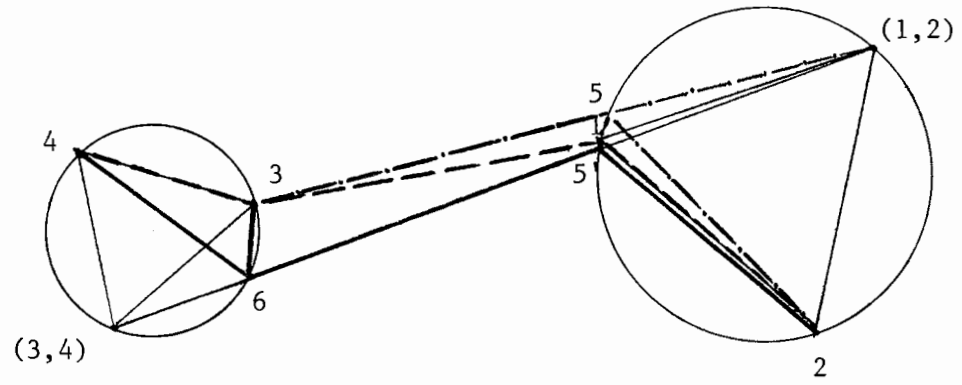


Figure 7

Furthermore, Figure 8 illustrates a case where the first degeneracy indicated (to point 3) later proves to be nondegenerate in the final solution, thus necessitating the reinsertion of Steiner points at step (f).

Armed with our lower bound (Theorem 1) and with a procedure for obtaining degenerate Steiner trees from full configurations (i.e., the corresponding Heinen trees), we are now ready to state our algorithm.

The Proposed Algorithm

We assume that N cannot be partitioned by the Steiner polygon. Otherwise, the problem would be decomposed to smaller subproblems, each of which would conform to our assumption separately.

Theorem 1 gives us a lower bound on the value of a Steiner tree associated with a full configuration. We also have a procedure to identify

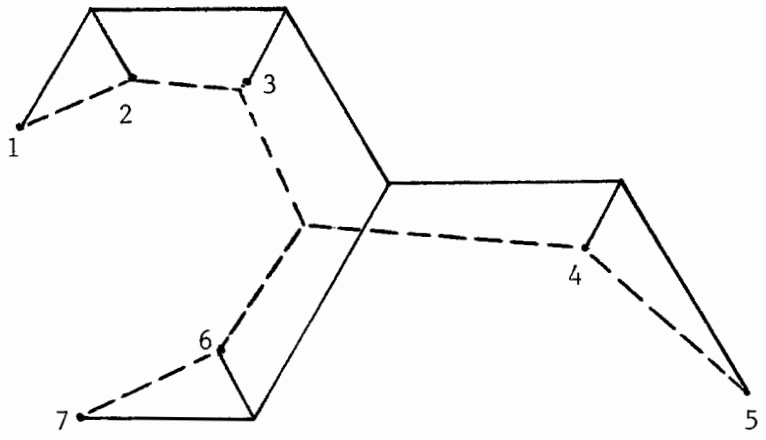


Figure 8

the Steiner tree associated with any given full configuration, if one exists. It is well within the power of Cockayne and Shiller's algorithm to scan all the full configurations associated with any set of nodes N, and using all these, we proceed as follows:

(a) By a fast heuristic (e.g., see [2]) find a satisficing feasible initial solution which will supply us with a tentative Steiner tree and an upper bound; go to (b) with the first full configuration.

(b) Check the lower bound associated with the full configuration. If it is below the upper bound call the procedure for obtaining Steiner trees from full configurations. Return from the procedure upon one of the following:

(i) A Steiner tree is found which is better than the current tentative solution, in which case go to (c); or

(ii) there is no Steiner tree associated with the full topology. In which case go to (d); or

(iii) the Steiner tree obtained is worse than the upper bound, in which case go to (d).

(c) Update the tentative solution and the upper bound (since a better Steiner tree has been found). Continue to (d).

(d) If the full configurations are not exhausted, go back to (b) with the next full configuration. Else, stop. (The current tentative solution is the global optimum.)

An Example

As an example we show an $N = \{1,2,3,4,5,6\}$ case depicted in Figure 9. There are five possible cyclic orders for this instance, since node 6 can be inserted between any pair--the other nodes are on the Steiner polygon--and this gives rise to 70 full topologies which our algorithm would have to scan. (For a convex hexagon there are 14 full topologies.) Now, a greedy heuristic would find the SMT here immediately, but since this SMT is not full, we will still have to check out some full configurations with promising lower bounds. Some of these would converge to the SMT, and we show one. Begin with the cyclic order 1-2-3-4-5-6-1 and the full configuration $\overline{(1,((2,3),(4,5))),6}$. After constructing the Heinen tree, we would observe a large negative link connected to node 6. This degeneracy yields two segments, $\overline{1,6}$ and $\overline{((2,3),(4,5)),6}$. $\overline{1,6}$ is a full Steiner tree for $\{1,6\}$, and the other segment yields a partial Steiner tree where a Steiner point merges with node 6 (again). Finally, in step (f), we find that we have a legitimate Steiner tree, which happens to be the SMT for this case. Note that this is a very special case since node 6 turns out to be of rank three. More regular cases are simpler in the sense that only one full "parent" topology exists for an SMT. In this example, if we carry out the check as required in step (e), in the degeneracies identification part of the algorithm for obtaining Steiner trees from full topologies, we find that 6 is on the line, so the inequalities are of the "same" sign only if we take them as weak inequalities, which is why

we only stipulated "weak" to begin with. Indeed in this case the opposite representation would be fine as well, and the same final result would be obtained.

Discussion

The Steiner problem is known to be NP-Hard [5], and indeed our algorithm requires exponential time, since it has to check an exponential number of full configurations; note, though, that if we start with a good initial solution, it is reasonable to expect that most full configurations would be eliminated

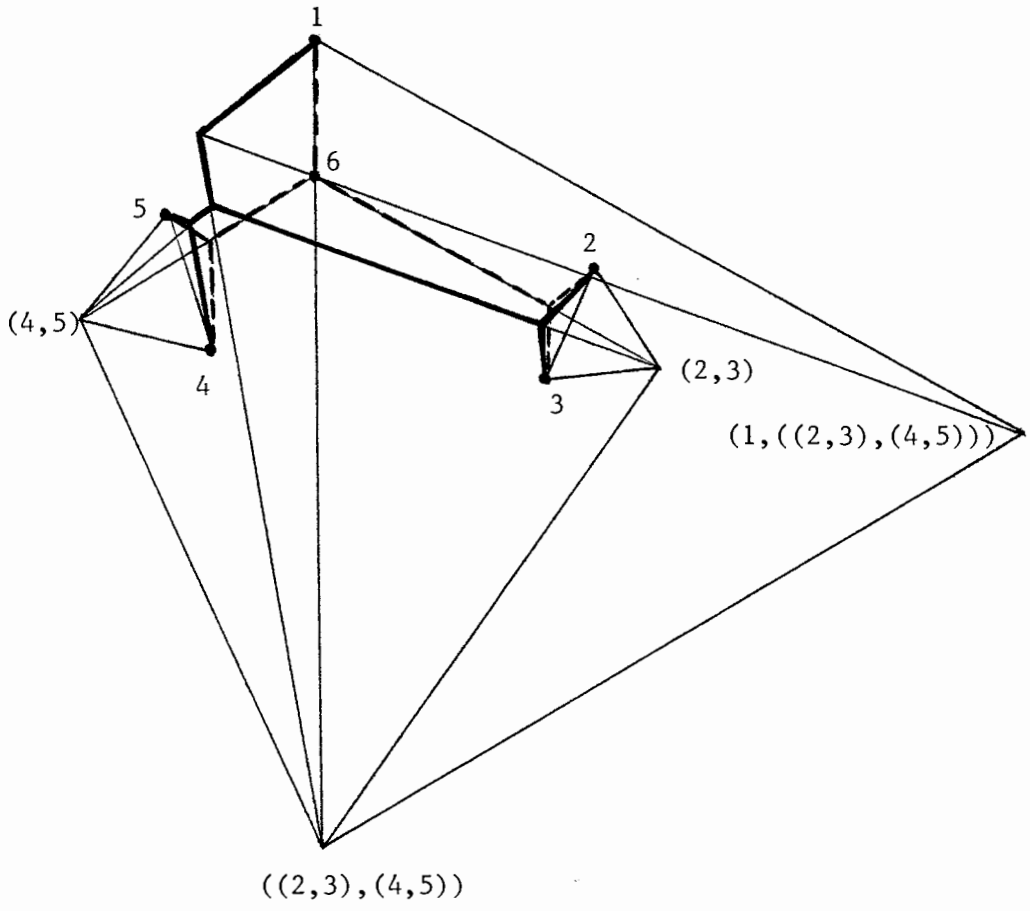


Figure 9

in step (b) without calling the tree obtaining procedure at all, by using the lower bound. Also note that all these full configurations are just a few of the configurations and partitions generated and checked by the previous algorithm. It is true that if a full configuration has a promising (low) lower bound it may require some time to check it out completely; however, the convexity of the degeneracy identification problem implies that on average this part of the procedure should be very fast. Indeed, this problem can be solved by a ropes and weights analog model which suggests that a polynomial procedure may be found for its solution. (Certainly, it is not NP-complete.)

An important generalization of the Steiner tree problem, first discussed by Gilbert [6] (or see [11] for an overview) introduces weights to the arcs. It might be worthwhile to adapt the algorithm suggested in this paper to this generalized weighted case, or at least some restricted versions thereof. (In the general weighted case, additional points may be of rank four or more, rendering the (generalized) Steiner construction inappropriate.)

The proposed algorithm was not implemented on a computer, so far, and there is cetainly room for contributions in this endeavor. The important question of the number of points which can be solved for will have to wait for such an implementation.

References

1. W. M. Boyce and J. B. Seery, STEINER 72, An Improved Version of Cockayne and Schiller's Program STEINER for the Miniman Network Problem, Bell Telephone Computing Science Technical Report #35 (1975).

2. S. K. Chang, The Generation of Minimal Trees with a Steiner Topology, J. ACM, 19, 4 (1972), 699-711.

3. E. J. Cockayne, On the Efficiency of the Algorithm for Steiner Minimal Trees, SIAM J. Appl. Math., 18 (1970), 150-159.

4. E. J. Cockayne and D. G. Schiller, Computation of Steiner Minimal Trees, in D. G. A. Welsh and D. R. Woodal (eds.), Combinatorics, Institute of Applied Mathematics (1972), 53-72.

5. M. R. Garey, R. L. Graham and D. S. Johnson, The Complexity of Computing Steiner Minimal Trees, SIAM J. Appl. Math., 32 (1977), 835-859.

6. E. N. Gilbert, Minimum Cost Communication Networks, Bell System Technical Journal 46 (1967), 2209-2227.

7. E. N. Gilbert and H. O. Pollak, Steiner Minimal Trees, SIAM J. Appl. Math., 16 (1968), 1-29.

8. H. W. Kuhn, "Steiner's" Problem Revisited, in G. B. Dantzig and B. C. Eaves (eds.), "Studies in Optimization," Studies in Mathematics, 10, 52-70, The Mathematical Association of America.

9. Z. A. Melzak, On the Problem of Steiner, Canad. Math. Bull., 4 (1961), 143-148.

10. D. Trietsch, Euler's Problem of Polygon Division and Full Steiner Topologies--A Duality, Discussion Paper No. 625, CEMSMS, Northwestern University (1984).

11. D. Trietsch and G.Y. Handler, The Gilbert and Pollak Conjecture--A Generalization, to appear in Networks.