Discussion Paper No. 655

# MINIMAL EUCLIDEAN NETWORKS WITH FLOW DEPENDENT COSTS--THE GENERALIZED STEINER CASE

A Monograph
by
Dan Trietsch*

May 1985**

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

**This monograph is a translation from Hebrew of some unpublished results from the author's Ph.D. dissertation, written under the supervision of Dr. Gabriel Y. Handler, Faculty of Management, Tel Aviv University, and submitted to the Senate of Tel Aviv University in March 1983. Slight changes have been made. The author is very grateful to Dr. Handler for his help over the years.

## Abstract

In this monograph we discuss in detail a problem introduced by Gilbert in 1967: on a Euclidean plane, a minimal total cost network connecting n given points is sought, where the cost of each arc depends on a flow cost function and on its length. In addition to the given points, generalized Steiner points (G-Steiner points) are allowed. We first discuss a simple case, where the network connects three given points. Some of the results can be generalized for $n > 3$ points; others are extendible only to a class of networks where the G-Steiner points are of degree 3 exactly--and we show some cases where this is assured; finally some results cannot be generalized to $n > 3$. We discuss all these $n > 3$ issues in the second part of the monograph.

## 1. Introduction

A network design problem often discussed in the literature [1] deals with choosing a subset of arcs from a given set (which includes a finite number of possible arcs connecting pairs of nodes), and assigning flows to the arcs according to a given flow demand in such a manner that the total cost incurred is minimal. This problem is NP-complete [12], and it seems to be one of the less tractable in that group [15]. It is therefore not surprising that not much work has been done to achieve possible savings by allowing extra nodes in the network (not belonging to the set of nodes which have to be connected). The Steiner tree problem, where the objective function is to minimize the sum of the Euclidean lengths of the arcs does allow such extra nodes—called Steiner points.

Since we have reason to believe that some of the basic literature we refer to is not widely known, we reiterate some of the major known results before proceeding to describe and derive our own.

The Steiner Minimal Tree Problem was introduced, it seems, by Courant and Robbins [6], under the name "The Street Network Problem," as a generalization of a classical problem which they called the Steiner Problem (after the famous German geometrician Jacob Steiner (1796-1864)). Actually, according to Kuhn [13], the classical version of the Steiner problem, namely, finding a point which minimizes the sum of distances from it to the vertices of a triangle, was first posed by Fermat in the beginning of the 17th century, and was solved geometrically for the non-degenerate case (where the point lies inside the triangle) by Torricelli before 1640. Other important contributors to the ultimate solution were Cavalieri (in 1647), Simpson (in 1750), and finally, F. Heinen (in 1834), who was the first to give the complete solution, including the degenerate case where the point coincides with one of the vertices of the triangle.

According to Courant and Robbins, the Minimal Steiner Tree Problem is: "Given n points, $A_1,\ldots,A_n$, [to] find a connected system of straight line segments of shortest total length such that any two of the given points can be joined by a polygon consisting of segments of the system" ([6], p. 360).

It can be shown that the "connected system" (the required network) is a tree with up to n - 2 Steiner points, each of degree three, with no intersecting arcs, and where no angle between adjacent arcs is less than 120°. Any such network is called a Steiner tree and the optimal one is called the Steiner minimal tree (SMT) [11].

Although Courant and Robbins solved some simple cases, the first finite algorithm capable of solving the problem is credited to Melzak [14]. Other important contributors are Gilbert and Pollak [11] and Cockayne [5]. Garey, Graham and Johnson [9] showed that the problem is NP-hard. Indeed, 10 point problems are generally considered to be intractable.

In their well-known comprehensive paper about Steiner trees, Gilbert and Pollak [11] conjectured that the ratio between the total length of the SMT and the minimal spanning tree is $\sqrt{3/4} \approx 0.866$ at least. This lower bound can actually be achieved, for instance, in the case of the SMT for the vertices of any equilateral triangle, and thus it is conjectured to be the greatest lower bound. Gilbert and Pollak also presented a general lower bound of 0.5, which is valid irrespective of the norm chosen. Pollak [16] proved the original conjecture for $n \leq 4$; Du, Hwang and Yao [8] proved it for $n \leq 5$; Chung and Hwang [4] proved a lower bound of 0.74309 (< 0.86603 of course), for any n; Chung and Graham [3] obtained a lower bound of 0.82046.

A further generalization of the Steiner tree problem was presented by

Gilbert [10]. We name it "The G-Steiner Minimal Network Problem" (G-Steiner may stand for generalized Steiner or Gilbert-Steiner). As our results hinge on this problem we redefine it as follows:

Definition 1: (The G-Steiner Minimal Problem): On a Euclidean plane, let a set N of n nodes be given, and let a set Q of bilateral nonnegative flow demands $q_{ij} = q_{ji}$ be given for all the possible pairs of nodes i,j. Also a function g(q) which assigns a cost per distance unit is given, such that:

(1) $g(0) = 0$

(2) $g(q) > 0; \forall\ q > 0$

(3) $g(q) \leqslant g(q + r); \forall\ r > 0$

(4) $g(q + r) \leqslant g(q) + g(r); \forall\ q,r > 0$

and such that if $d(\bar{\underline{X}}^{1a}, \bar{\underline{X}}^{2a})$ is the Euclidean distance between the endpoints of arc a in $E^2$, then the cost of assigning a total flow of $q_a$ to this arc is

(5) $g(q_a) \cdot d(\bar{\underline{X}}^{1a}, \bar{\underline{X}}^{2a})$.

The problem is to construct a network G(P,A) where $P \supseteq N$ and A is the set of arcs that span the set P of the original n nodes and possible extra nodes, and to assign all the flow demands $q_{ij}$ in Q to its arcs, so as to minimize the total cost incurred:

(6) $$Z = \sum_{a \in A} g(q_a) \cdot d(\bar{\underline{X}}^{1a}, \bar{\underline{X}}^{2a}).$$

Note that if a node i exists in N, so that all its flow demands are zeros, it can be "connected" by dummy arcs at zero cost (see equation 1). We shall assume, however, that such nodes do not exist, since their removal from the set N does not alter the real problem. Also note that practically, (1) to (4) are not restrictive at all, at least not for the Euclidean case.

We will name the solution of the problem the G-Steiner Minimal Network or the GSMN. Also, extra nodes will be referred to as G-Steiner points.

Clearly, if $g(q)$ is constant for any q, the Steiner tree problem is obtained as a special case. Thus, the G-Steiner problem is indeed a generalization of the Steiner problem as stated.

Gilbert [10] also generalized the construction known as the Steiner construction [14], whereby the Steiner point is found by a ruler and compass, to the case of any G-Steiner point of degree three. In the Steiner case an equilateral triangle is constructed outside the triangle using any of the three edges as a basis; the equilateral triangle is circumscribed by a circle. If all the angles in the triangle are less than 120°, then the Steiner point is located where the circle intersects the Simpson line [13], i.e., the line segment connecting the opposing apices of the equilateral triangle and the original triangle. However, if one angle is 120° or more, the vertex of that angle coincides with the Steiner point (this is a degenerate case). In Gilbert's generalized construction, the only difference is that an appropriate triangle of forces is constructed instead of the equilateral triangle. (Note than an equilateral triangle is a triangle of forces for equal vectors.) Figure 1 illustrates the G-Steiner construction for a case where $g(q_{12} + q_{13}) = a = 0.45$, $g(q_{12} + q_{23}) = b = 0.7$, and $q(_{23} + q_{13}) = c = 1$. The notation (1,2) for the third node of the triangle of forces is according to the notation developed by Cockayne [4] for the Steiner tree case. The construction is nondegenerate if node 3 is strictly out of the

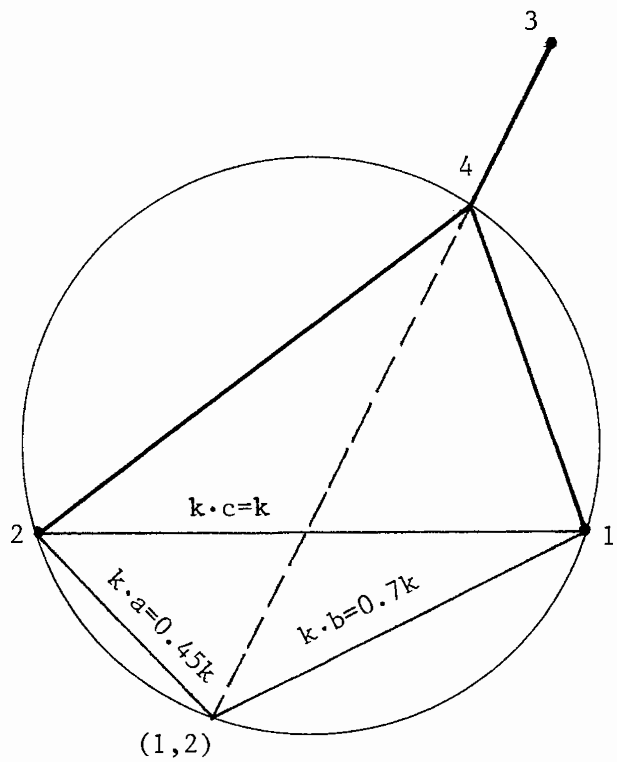



Figure 1

circle and strictly between the rays $\overline{(1,2),1}$ and $\overline{(1,2),2}$. Or else the solution is degenerate, and the G-Steiner point coincides with an original node.

Another generalization proved by Gilbert is that if the construction is nondegenerate, the weighted length $c \cdot d((1,2),3)$ is the value of the G-Steiner tree, i.e.,

$$a \cdot d(1,4) + b \cdot d(2,4) + c \cdot d(3,4) = c \cdot d((1,2),3) = |GST|. \quad (7)$$

Finally, the symmetric constructions give similar results. It follows that:

(8) $c \cdot d((1,2),3) = b \cdot d((3,1),2) = a \cdot d((2,3),1) = |GST|$

In [20] it was shown that for three nodes, the best possible improvement to be obtained by using one G-Steiner point is $1 - \sqrt{3/4} \approx 0.134$. This is similar to the result obtained for the Steiner tree case for $|N| = 3$, as shown by Gilbert and Pollak. This was the basis for a generalization of Gilbert and Pollak's conjecture to the G-Steiner case made there. Some interim results of [20] are repeated as part of this monograph. Recently, Du and Hwang [7] obtained a more elegant proof for the main theorem of [20]. They also showed by counterexample that the conjecture is not true for cases with more than one G-Steiner point with non-concave flow cost per distance unit function for $n \geq 4$. The conjecture remains open for concave functions and for $n = 3$ with two G-Steiner points.

In addition to the results of [20], the following results were obtained for $n = 3$, and are presented in this monograph:

(i) We prove that the best basic solution, i.e., the best solution such that no cycle exists which does not pass through at least one node of N, is the global optimum. This is contrary to an impression one might get from Gilbert's paper, where a nonbasic solution is presented as possible (and it is possible), without qualification.

(ii) Stable versus unstable solutions are defined, and we show that a local equilibrium in the G-Steiner points does not imply stability. Such stability is shown to be a necessary condition for the optimum. For the case of a stable network with a cycle we present a necessary condition which depends on the weights $g(q)$. In a series of theorems we prove that if a cycled stable topology exists at all it is unique, and if in addition g(q) is concave, then this cycled stable unique topology is optimal as well.

(iii) We show a sufficient condition for a cycled stable topology, which is the existence of a full G-Steiner tree together with the necessary condition of (ii), and discuss some other cases.

(iv) We develop a lower bound for the tree case--which is equal to its length if it is full.

For $n > 3$ we observe that potential rank four (or more) G-Steiner points may imply nonbasic solutions; further, these make the use of G-Steiner construction impossible, so a nonlinear search is required for the solution. Therefore, we want to discuss the class where the rank of G-Steiner points is three. For this class we show how the G-Steiner construction can be applied even if cycles are present by splitting nodes--if the solution is basic. But for this class we are also able to extend the basic-optimality result. We also characterize cases where the problem belongs to our class. The lower bound obtained for $n = 3$ is extended to $n > 3$. Finally we show two heuristics for this case, a greedy (or myopic) one and one based on aggregation. The myopic algorithm should start with a good regular network--which we can obtain by such methods as described in [1], etc.

In Section 2 we discuss the $n = 3$ case, and Section 3 is devoted to $n > 3$.

## 2. New Results for the Three Nodes Case

The case we discuss in this section is $N = \{1,2,3\}$, $Q = \{q_{12}, q_{23}, q_{31}\}$, $q_{ji} \equiv q_{ij} \geq 0$; $i,j \in N$, and the flow demands of two pairs at least are strictly positive (otherwise one node at least is separate).

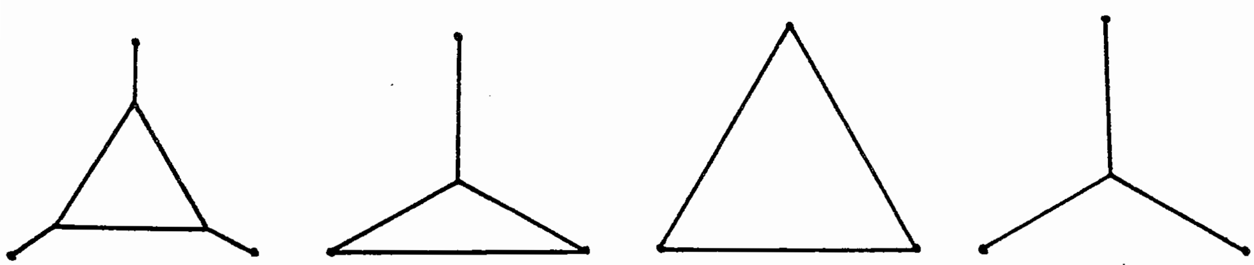
Definition 2: A G-Steiner cycle is a cycle which does not pass through any node of N.

Definition 3: A simple cycle connects only nodes of N. (In the n = 3 case, the triangle $\Delta 1,2,3$ is the only simple cycle possible.)

Definition 4: A mixed cycle is a cycle which is neither G-Steiner nor simple (i.e., passes through some G-Steiner points and some nodes of N).

Definition 5: A G-Steiner tree (GST) is a G-Steiner network without any cycles. A GST which is minimal relative to all GSTs is a GSMT (G-Steiner minimal tree).

In Figure 2, parts (a), (b), (c) and (d) depict a G-Steiner cycle, a mixed cycle, a simple cycle and a tree case. Parts (a) and (b) of the figure are from [10].



a b c d

Figure 2

The following is Theorem 1 from [20], where the simple proof can be found.

Theorem 1: For any admissable instance of the G-Steiner network problem where $n = 3$, the three values $g(q_{31} + q_{12})$, $g(q_{12} + q_{23})$ and $g(q_{23} + q_{31})$ conform to the triangle inequality, at least weakly.

The implication of Theorem 1 is that the G-Steiner construction exists for any three nonnegative flows (with at least two strictly positive ones) and any function $g(q)$ which conforms to (1) through (4).

## On the Optimality of the Best Basic Solution

Definition 6: A G-Steiner network without G-Steiner cycles is called basic.

For $|N| = 3$, we now prove that the best basic solution is optimal. In order to do that we use a parallel shift procedure, depicted in Figure 3a. For a formal presentation of the parallel shift, the reader is referred to Definition 8 in [19]. The reader is also referred to Definition 9 there, of a maximal parallel shift. Armed with these, we can present the theorem.

Theorem 2: For the $|N| = 3$ case, if an optimal solution includes a G-Steiner cycle, two basic solutions exist which are also optimal.

Proof: (See Figure 3.) Assume that the network with the G-Steiner cycle $\overline{4,5,6,4}$ is optimal, and we have to show that the (basic) tree with the G-Steiner point 7 and the basic solution with the mixed cycle $\overline{3,8,9,3}$ are optimal. First we observe that since our solution is optimal, the three rays associated with $\overline{1,4}$, $\overline{2,5}$ and $\overline{3,6}$ must indeed intersect at a point such as 7, otherwise the system would not be in equilibrium, but would have an unresolved moment. Now, through 7 draw perpendiculars to $\overline{4,5}$, $\overline{5,6}$, and $\overline{6,4}$, respectively. We

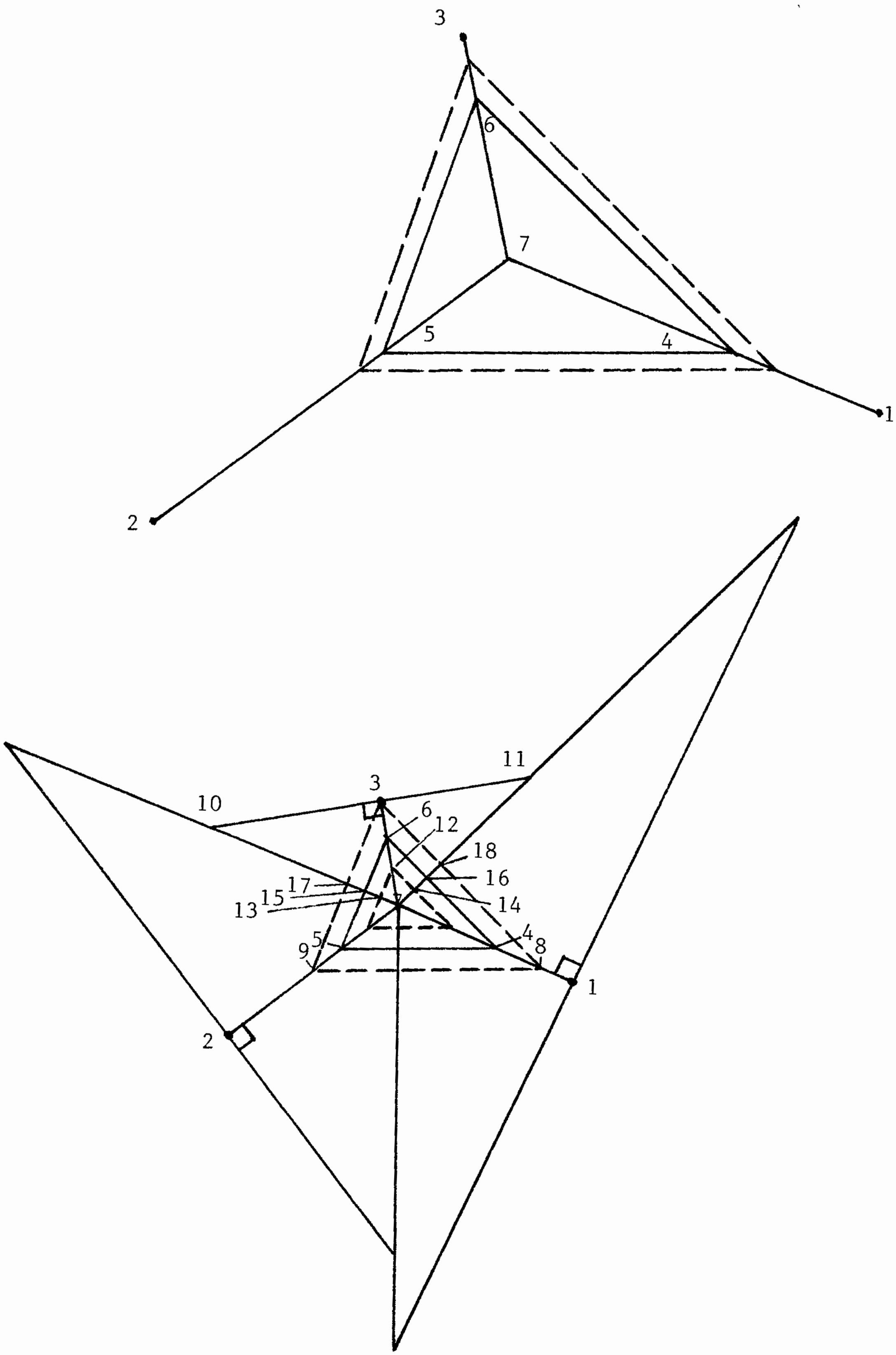



Figure 3

thus obtain three triangles. One of these is Δ7,10,11. Now, clearly this triangle is similar to the force triangle we would use to construct a G-Steiner point such as 6. I.e., (10,11) = k•c, d(7,11) = k•b and d(7,10) = k•a for some positive k, where the cost per distance of arcs $\overline{3,6}$, $\overline{4,6}$ and $\overline{5,6}$ are c, b and a, respectively. Clearly Δ3,8,9 is the result of a maximal parallel shift, and it is a proper triangle as a result of the three rays meeting at one point, 7, as discussed above (see Lemma 3, Section 3); or take any point such as 12, and a similar (nonmaximal) parallel shift can be defined for it, again resulting in a proper triangle with vertices on the three rays, intersecting $\overline{7,10}$ and $\overline{7,11}$ at 13 and 14. But, k•c•d(12,3) + k•b•d(12,14) + k•a•d(12,13) is the area of Δ7,10,11 and so is k•c•d(6,3) + k•b•d(6,16) + k•a•d(6,15) and k•b•d(3,18) + k•a•d(3,17), hence the value of the network within Δ7,10,11 is invariant under parallel shifts, and similarly for the other two triangles. Our result follows immediately. □

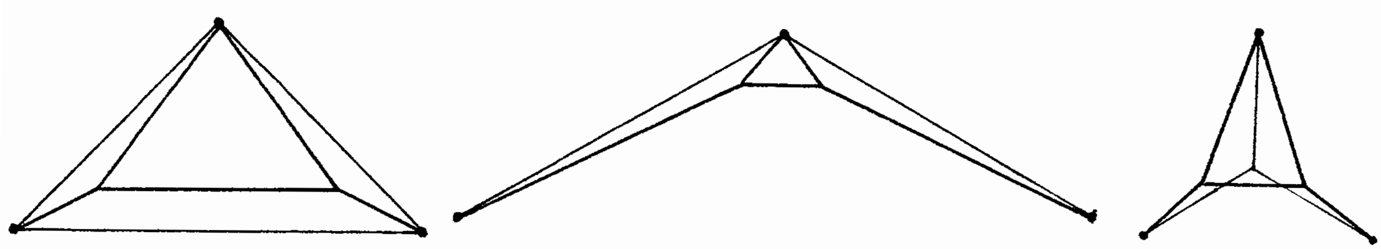
## Networks with Cycles for |N| = 3

Clearly, by Theorem 2 we do not have to bother with G-Steiner cycles. However, an optimal network may still have a simple or a mixed cycle. We now present two such examples, and a third example of a cycled network which is not optimal. This leads us to obtain a necessary condition for a "stable" cycle in the optimal network. The condition holds for a given situation where we apply a perturbation test for it, but it can also be checked independently of N, for a given Q and g(q). Later we show that the necessary condition implies that if a stable cycle exists it is the optimal solution for the flows assigned to its arcs; i.e., if we know that a flow $q_{ij}$ is not split to more than one path--and if g(q) is concave this is implied by a theorem due to Gilbert [10]--then a stable cyclic network is the global optimum, if one exists. If g(q) is not concave, it may happen that parts of the flow demands are dispatched through direct arcs, while others are sent through another part of the network (such an example is shown in [10]).

Let $q_{12} = q_{23} = q_{31} = 1$, and

$$g(q) = \begin{cases} c + q; & q > 0 , \\ 0 \quad ; & \text{otherwise}. \end{cases} \tag{9}$$

For cases (a) and (b) let c = 1/4, while for case (c), c = 2/3. This implies g(2)/g(1) = 1.8 for (a), (b) and 1.6 for (c). (These were chosen since $1.8 > \sqrt{3} > 1.6$; in Gilbert's G-Steiner cycle example, as in Figure 1a, $g(2)/g(1) = \sqrt{3}$ for the same flows.) Our three cases are depicted, respectively, in Figure 4.



a b c

Figure 4

For case (a) N is an isosceles triangle with an apex angle of 90°; for case (b) N is likewise, but with an angle of 120° (hence its Steiner tree degenerates); and (c) is an equilateral triangle case.

For (a) and (b) there is no other cycled topology with two Steiner points, but for (c) two symmetric ones exist. For (a) the simple cycle $\overline{1,2,3,1}$ is the optimal regular network (i.e., s.t. $P \equiv N$); for (b) the sides of the obtuse angle form a tree which is the regular optimum--this tree happens to be a degenerate Steiner tree; for (c) the regular optimum is any of the three possible trees. Now, for (c) the global optimum is the Steiner tree. In addition, (c) fails a perturbation test where we insert an additional Steiner point near the rank 2 node of N, and we (i) get improvement, and (ii) achieve equilibrium again only for the G-Steiner tree. Both (a) and (b) pass this perturbation test with flying colors. A $g(2)/g(1) = \sqrt{3}$ case would be indifferent to a "very small" perturbation here. We refer to cases such as (a) and (b) as (strictly) stable; a case such as (c) is unstable and a case such as $g(2)/g(1) = \sqrt{3}$ (for our Q) is stable, but not strictly.

Definition 7: A basic network is called strictly stable if inserting an additional G-Steiner point anywhere is detrimental. If such an insertion makes no difference, the network is stable, but not strictly.

Now we may obtain a necessary condition for stability, which is independent of N, for any given Q and g(q).

## A Necessary Condition for Stability

A function g(q) subject to (1) through (4), and a set $Q = \{q_{12}, q_{23}, q_{31}\}$ are given. Without loss of generality, we may assume $q_{ij} > 0$, $\forall\ i,j \in N$, since otherwise no cycles are possible. By Theorem 1 the triplet $g(q_{12})$; $g(q_{23})$; $g(q_{12} + q_{23})$ conforms to the triangle inequality, and likewise the triplets $g(q_{23})$; $g(q_{31})$; $g(q_{23} + q_{31})$ and $g(q_{31})$; $g(q_{12})$; $g(q_{31} + q_{12})$. Construct the respective triangles thus implied, say, $\Delta I$, $\Delta II$, and $\Delta$III as in Figure 5, and we are ready to state a theorem.

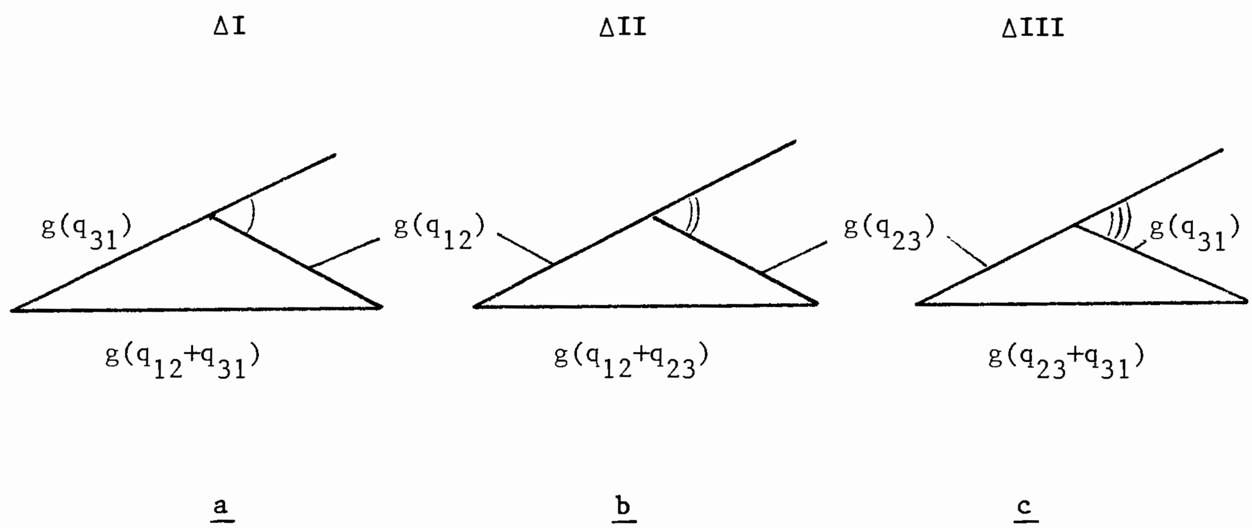
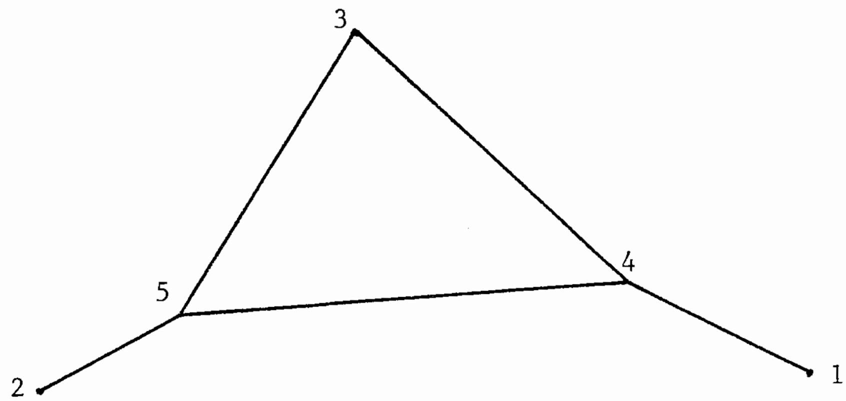



Figure 5

Theorem 3: A necessary condition for (strict) stability in a cycled network for Q, g(q) is that the sum of the apex angles of $\Delta$I, $\Delta$II and $\Delta$III is (greater than, or at least) not less than 360°.

Proof: By negation, assume otherwise and we have to show that the network fails the perturbation test. Now, in a stable network the G-Steiner points are vector equilibrium points, so the angles must be as implied by the three triangles. Figure 6 depicts such a case, and the acute angles at 4 and 5 should be equal to the complementary angles of the apices of $\Delta$I and $\Delta$II, respectively. But under out assumption the angle at 3 is smaller than the respective apex complementary angle for $\Delta$III, and it follows almost directly that it cannot stand the test. □




<u>Figure 6</u>

<u>Theorem 4</u>: If the necessary condition of Theorem 3 holds, then sets N always exist with such a configuration that a cycle with two G-Steiner points exists for them, other configurations always exist implying cycles with one G-Steiner point, and still others exist implying a simple cycle.

<u>Proof</u>: By construction, trivial. □

In Figure 4c we showed a case of an unstable and nonunique cycled network. We now show that if the condition of Theorem 3 holds, there is at most one such cycle at local equilibrium (and obviously, it is also stable then).

Theorem 5: The existence of a stable solution implies it is also unique.

Proof: Such a solution can pass the perturbation test, hence it is the unique solution of a convex problem, namely, where to locate the three G-Steiner points (and recall that at least one of them must merge with a node of N).

□

We now proceed to prove that if such a stable cycled network exists it must be optimal. Two cases are considered: (a) the best cycle-less network is a nondegenerate G-Steiner tree (as in Figure 4a); (b) the best tree has no G-Steiner point (as in Figure 4b). Below we will also show that in case (a) the cycled network is implied by the necessary condition, i.e., if the necessary condition of Theorem 3 is satisfied and the tree optimum has a G-Steiner point, the optimum is cycled. (However, that proof will require some extensive preparation.)

Theorem 6: For $N = \{1,2,3\}$; $Q = \{q_{12}, q_{23}, q_{31}\}$, and an admissable function $g(q)$, if a strictly stable solution and a nondegenerate G-Steiner tree exist, then the cycled network is better.

Proof: (Figure 7): Strictly stable implies that $g(q_{23} + q_{31})$ is greater than the vector sum of a vector from 3 to 4 with a magnitude of $g(q_{31})$ and a vector from 3 to 5 with a magnitude of $g(q_{23})$. Denote this sum vector's magnitude as $\ell$, hence, as we said

(10) $$g(q_{23} + q_{31}) > \ell.$$

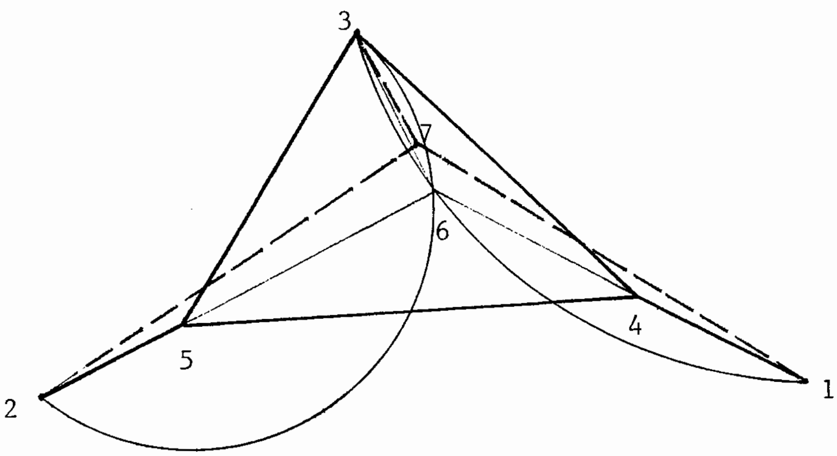



Figure 7

Let 7 be the G-Steiner point, and 6 be the point where $\overline{1,4}$ (extended) meets $\overline{2,5}$ (extended), then

(11) $\sphericalangle\ 1,7,2 < \sphericalangle\ 1,6,2$

(12) $\sphericalangle\ 2,7,3 > \sphericalangle\ 2,6,3 > \sphericalangle\ 3,4,5$

(13) $\sphericalangle\ 1,7,3 > \sphericalangle\ 1,6,3 > \sphericalangle\ 4,5,3$,

since the G-Steiner construction is not degenerate for $\Delta 1,2,3$. By (11), (12) and (13), clearly we can execute the same G-Steiner construction for $\Delta 3,4,5$; hence the opposite is also true, and 7 is the G-Steiner point for a triangle, say $\Delta 3,8,9$ congruent to $\Delta 3,4,5$ but (it is easy to verify) smaller. Figure 8

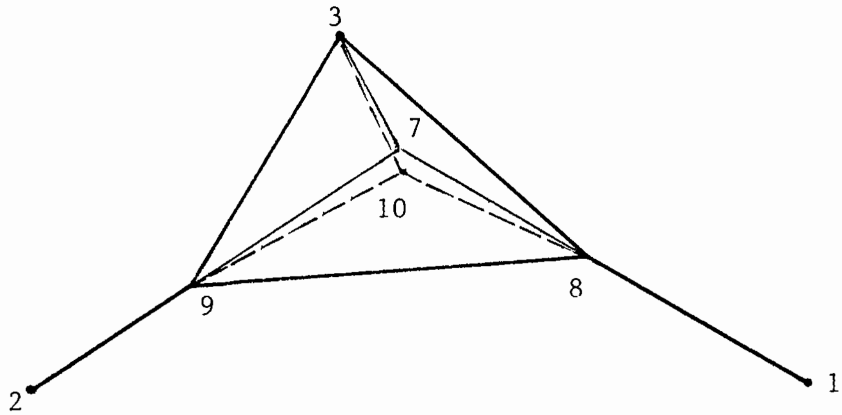



Figure 8

depicts this triangle, and point 10 there is obtained by performing the G-Steiner construction for $\Delta 3,8,9$ with $\ell$, $g(q_{31})$ and $g(q_{23})$ for the connections to 3, 8, and 9, respectively. I.e., 10 is "similar" to 6. Now, if we take the sub-network $\Delta 3,8,9$, clearly using it as depicted costs us the same as assigning $q_{12} + q_{31}$ to $\overline{8,10}$; $q_{12} + q_{23}$ to $\overline{9,10}$ and paying $\ell$ per distance unit on $\overline{10,3}$, since $\Delta 3,8,9$ is effectively a Steiner cycle for these costs. Denote this cost by P, and we can write again

$$(14) \qquad P = d(3,8) \cdot g(q_{13}) + d(3,9) \cdot g(q_{23}) + d(8,9) \cdot g(q_{12}) =$$

$$d(3,10) \cdot \ell + d(8,10) \cdot g(q_{12} + q_{31}) + d(9,10) \cdot g(q_{12} + q_{23}).$$

But 10 is optimal for this sub-network and these same costs, so if we move, say, to 7 we will have to pay more, say, Q. I.e.,

$$(15) \qquad Q = d(3,7)\cdot\ell + d(7,8)\cdot g(q_{12} + q_{31}) + d(7,9)\cdot g(q_{12} + q_{23}) > P.$$

Now, if for "some reason" we decide that $\ell$ is infeasible, and change it for $g(q_{23} + q_{31})$ we pay even more than Q, and certainly more than P. But this was our situation before inserting $\Delta 3,8,9$ instead of this part of the tree! Shifting back to $\Delta 3,4,5$ will imply even further gains, since it is stable, and therefore a local minimum. □

Theorem 7: For $N = \{1,2,3\}$, $Q = \{q_{12}, q_{23}, q_{31}\}$, and an admissible function $g(q)$, a strictly stable cycled network, if one exists, is better than any regular (two arc) tree.

Proof: Clearly it is enough to prove for a regular tree if it is optimal within the tree set. Assume this optimal tree is $\overline{2,3} + \overline{3,1}$, and the G-Steiner construction would degenerate to it, of course. Several mutually exclusive and exhaustive possibilities exist, and we discuss them one by one:

(a) The cycle includes node 3 and two G-Steiner points. In this case, depicted in Figure 9, a similar analysis to that of Theorem 6 yields for the value V of our network

$$(16) \qquad V = d(3,6)\cdot\ell + d(2,5)\cdot g(q_{12} + q_{23}) + d(1,4)\cdot g(q_{12} + q_{31})$$

which is clearly better than the tree value

$$(17) \qquad T = d(2,3)\cdot g(q_{12} + q_{23}) + d(1,3)\cdot g(q_{12} + q_{31}),$$

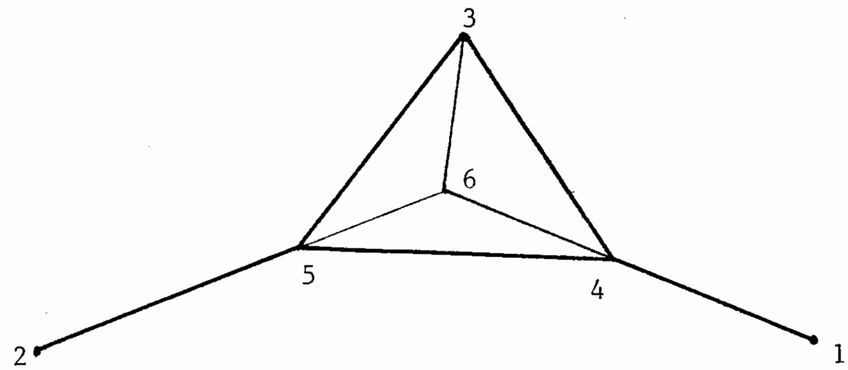



Figure 9

since otherwise the G-Steiner construction for locating 6 would degenerate.

(b) The cycle includes one node of N, but not 3. This one is simply impossible, since the Steiner construction degenerates to 3, obviously a sum vector directed from, say, 2 and with a magnitude of not more than b cannot be enough to "pull" the cycle to it. A similar result holds for 1.

(c) Nodes 1,2 of N are on the cycle but 3 is not. Like (b), this is impossible.

(d) Nodes 3 and, without loss of generality, say 2 are on the cycle, and 1 is not. Figure 10 depicts this case. Point 5 is located where two sum vectors associated with 3 (this is the one discussed above with magnitude $\ell$) and 2 (with magnitude m, similarly) meet. Since we have a stable case, both $\ell$ and m must not be greater than $g(q_{23} + q_{31})$ and $g(q_{12} + q_{23})$, respectively.

From here on the case is similar to case (a).

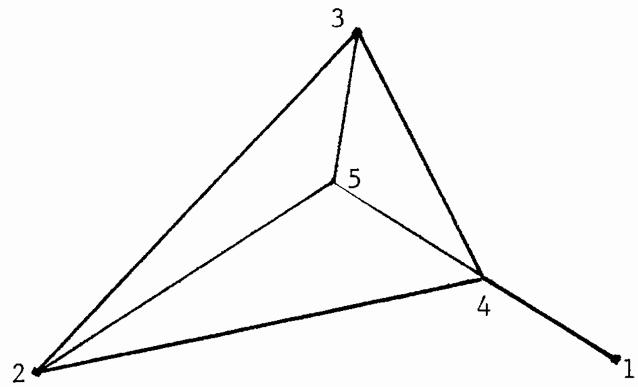



Figure 10

(e) The cycle is simple. Now we have three sum vectors instead of two (as in (d)) or one (as in (a)), but the same proof still holds. □

So far we have shown that the existence of a cycled network implies its optimality for the flows assigned to it. We also have a necessary condition for cycles (Theorem 3). We wish to obtain a sufficient condition as well. (We do not have a necessary <u>and</u> sufficient one yet, however.) To that end, and as an end in itself, we first characterize all the cases which are possible, if no flow demand $q_{ij}$ is split. A sufficient condition for such "unity" has been proved by Gilbert [10], and that is if g(q) is concave.

Characterization of the Cyclic Cases Without Splitting Flow Demands

Whether g(q) is concave or we just decide not to consider splitting flows, we characterize the cyclic cases for this case only.

The first part is simple: Theorem 3 is a necessary condition for cycles. If it is not satisfied, cycles are out of the question. In this case the GSMT (degenerate or not) is our optimum. The same applies if the necessary condition applies as an equality, where we may settle for the best tree without losing optimality. We can also, in this case, and if and only if the G-Steiner tree is full, find a cycled solution, with a mixed or (rarely) simple cycle. In this case every convex combination of the two basic solutions is a nonbasic solution.

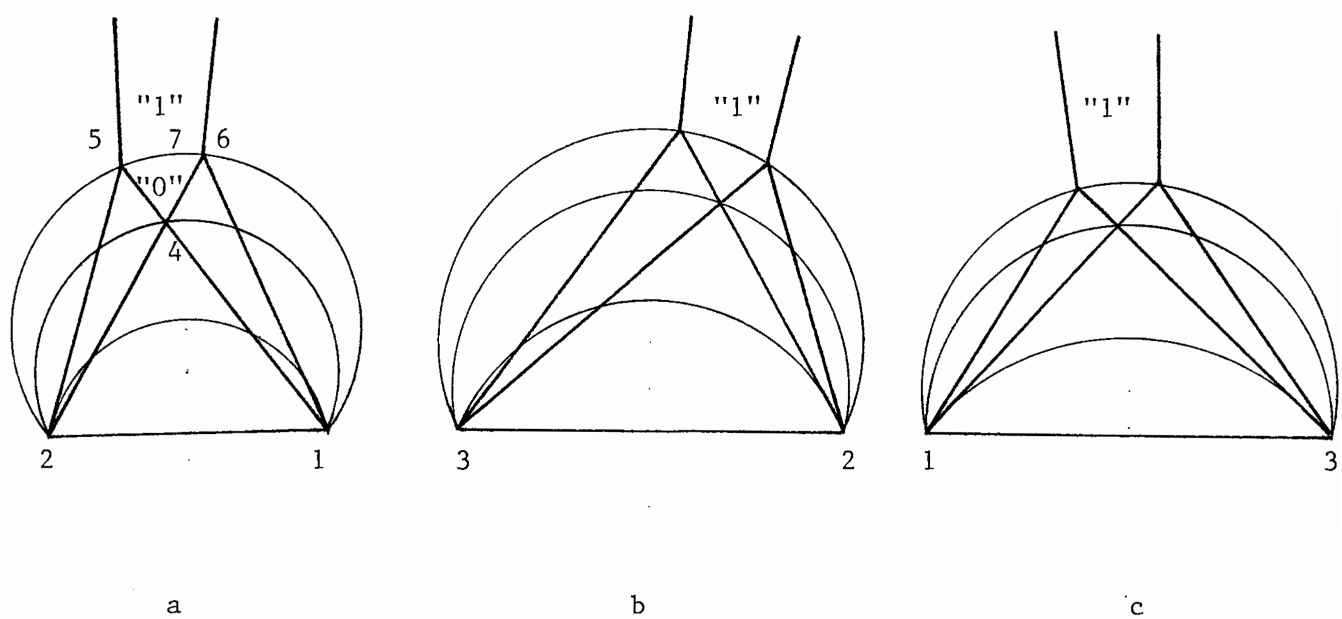



Figure 11

If the necessary condition is obtained strictly, the analysis is a little more complicated. We refer to Figure 11, with its three parts, a, b, and c. We discuss mainly part a, but similar results are implied for parts b and c.

Pick a side of $\Delta 1,2,3$, say $\overline{1,2}$ as a basis. The length d(1,2) is taken as some proportion of $g(q_{23} + q_{31})$. Point 4 is located in such a manner that $\angle 1,2,4$ is the complementary to the apex angle of Figure 5a, while $\angle 2,1,4$ is likewise with Figure 5b. The arc 1,6,5,2 is the locus of points subtending $\overline{1,2}$ at the respective complementary angle of Figure 5c, and points 5,6 are on the extended rays $\overline{1,4}$ and $\overline{2,4}$, respectively. (The arc 1,6,5,2 must be above 1,4,2 due to Theorem 3.) From 5 and 6 we extend rays away from $\overline{1,2}$ in the direction implied by the vector sum of vectors directed from 1 and 2 to the point (5 or 6) with magnitudes of $g(q_{13})$ and $g(q_{23})$, respectively (Figure 5c is the appropriate force triangle that we can use for this purpose). To the two arcs we have we add one, below (as can be shown), subtending $\overline{1,2}$ at an angle such as $\angle 1,6,2$ in Figure 8. Parts b and c are obtained similarly for $\overline{2,3}$ and $\overline{3,1}$, respectively. The triangles $\Delta 1,2,4$, $\Delta 1,2,5$, and $\Delta 1,2,6$ are similar to triangles forming mixed cycles of one N-node. E.g., $\Delta 1,2,4$ is similar to $\Delta 4,5,3$ in Figure 8.

Our characterization is based on the location of 3. Referring to part a again, as follows:

(a) If 3 is between the rays extended from 5 and 6 (or on one of them), above arc 5,6 (but not on it), then a point 7 exists on arc 5,6 which, if connected to 1 and 2 yields vector equilibrium as required, while the angles $\angle 1,2,7$ and $\angle 2,1,7$ will be enough to pass the perturbation test. In this case, and only in one of the three such cases, a cycle with one G-Steiner point is obtained, and 7 it is, with 1 and 2 on the cycle in our case.

(b) If and only if 3 is in the lined area above $\angle 5,4,6$ and up to arc $\overline{5,6}$, including the boundary, a simple cycle is indicated.

(c) If 3 is below the low arc, the G-Stiner construction degenerates.

(d) If none of the above happens for any of the three parts, two G-

Steiner points are indicated. If 3 is to be the N-node for this case, then in part (a) it must be below 1,4,2. However, even if 3 is below 1,4,2, it may happen that 2 or 1 is also below its respective arc, so more than one candidate exists, and we may have to check both (however, only one of them will be realizable).

The next theorem concludes our characterization.

Theorem 8: IF the G-tree is full and Theorem 3 satisfied, a stable cycled network exists.

Proof: In this case, respective to Figure 11 a, b, and c, the G-Steiner point is above the lower arc, so case (c) in the characterization is not indicated. All other cases yield cycles of one kind or another. □

Note that the opposite is not true, e.g., Figure 4b, where this sufficient condition does not hold but a stable cycle exists.

## A Lower Bound for G-Steiner Trees

Figure 12 depicts the areas where node 3 should be so that the G-Steiner construction does not degenerate, or so that it degenerates to 1, to 2, or to 3. (Without loss of generality, assume 3 is above $\overline{1,2}$.) If there is a degeneracy to 3 we call it type I degeneracy, while if it is to a side (1 or 2) we call it type II (although actually there is no real difference). Denote the value of the G-Steiner tree by T, then

Theorem 9: $T \geqslant c \cdot d((1,2),3) = g(q_{23} + q_{31}) \cdot d((1,2),3)$.

Proof: Without degeneracy, this is an equality [10]. Therefore, assume degeneracy. But even if there is degeneracy, then it is easy to show that

(18) $c \cdot d((1,2),3) = b \cdot d(2,(3,1)) = a \cdot d(1,(2,3))$.

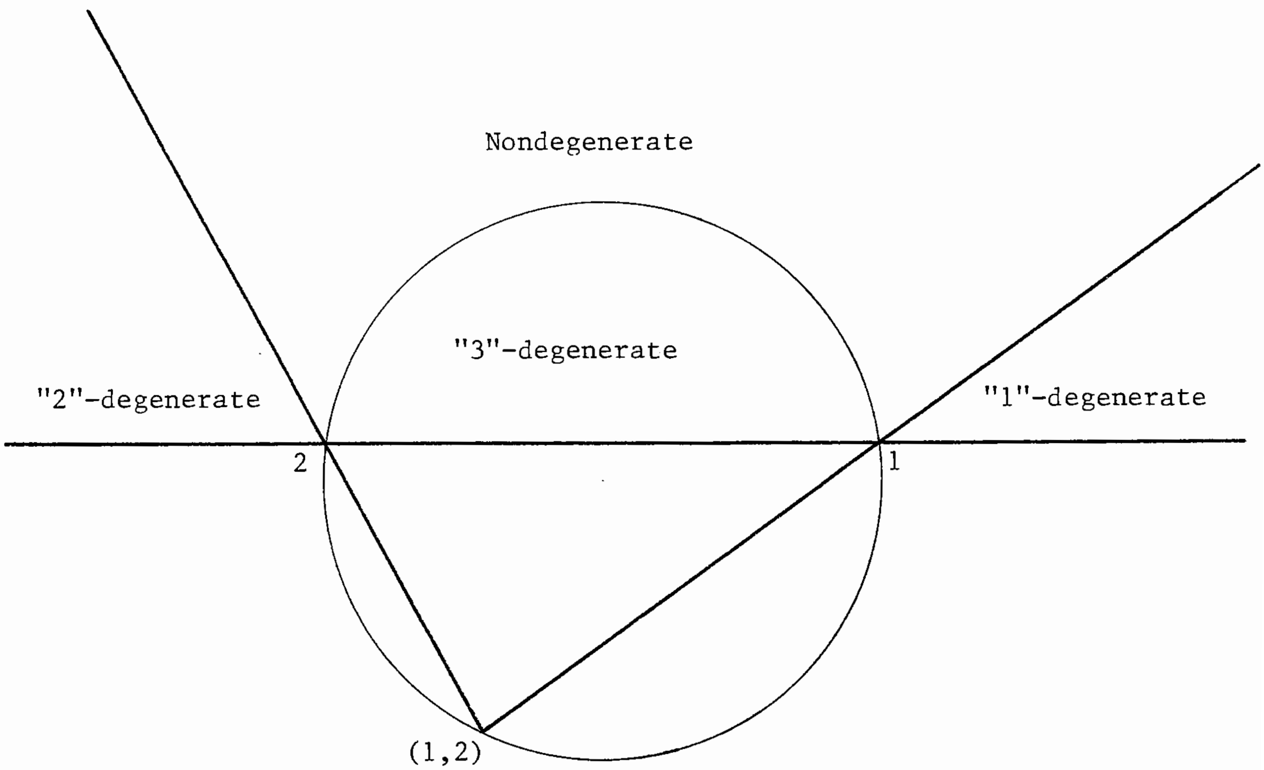



Figure 12

Hence, we may assume without loss of generality any type of degeneracy we wish, and use the appropriate construction. Assume Type I degeneracy, then, and Figure 13 depicts our case. For $\Delta 1,2,4$ the GSMT is $\overline{1,4} + \overline{4,2}$, and its value is

(19) $c \cdot d((1,2),4) = c \cdot d((1,2),3) + c \cdot d(3,4)$.

Node 3 is not the optimal G-Steiner point for $\Delta 1,2,4$, hence

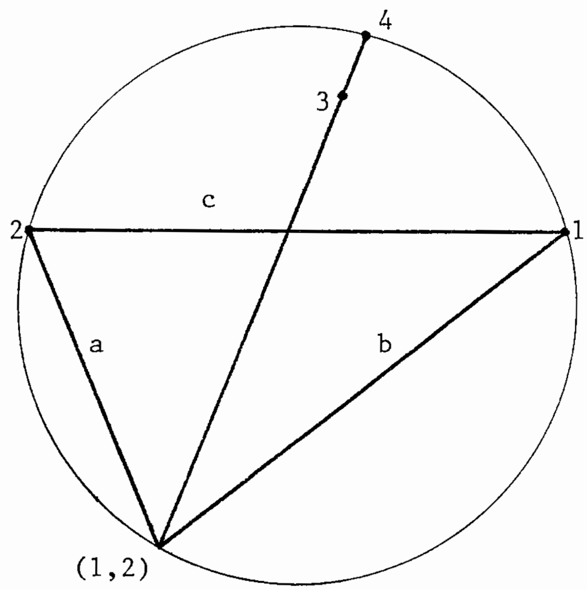



Figure 13

(20) $c \cdot d(3,4) + a \cdot d(1,3) + b \cdot d(2,3) \geqslant c \cdot d((1,2),4).$

But,

(21) $T = a \cdot d(1,3) + b \cdot d(2,3).$

So by subtracting $c \cdot d(3,4)$ from both sides of (20), and using (19) we obtain

(22) $T = a \cdot d(1,3) + b \cdot d(2,3) \geqslant d((1,2),3).$ $\square$

The Gilbert and Pollak Conjecture

In this section we review a theorem implying that the Gilbert and Pollak conjecture holds for $|N| = 3$ if we only allow one G-Steiner point.

Definition 8: For given Q and g(q), a spanning tree network without G-Steiner points is called a regular tree, and the best of these is called the regular minimal tree or the RMT. The best network possible without G-Steiner points is called (similarly) the RMN.

Definition 9: For given Q and g(q), a triangle $\Delta 1,2,3$ is called most improvable (or MIT) if it minimizes the ratio $|GSMT|/|RMT|$ relative to all triangles. If, in addition, d(1,2) = 1, then we call it the normalized most improvable triangle (or NMIT).

While MITs can be shown to exist and are uncountable, there is just one NMIT for any Q and g(q) pair. We proceed to show the existence of MITs.

Lemma 1: If (a;b;c) strictly conforms to the triangle inequality, then so does $(a(-a + b + c);\ b(a - b + c);\ c(a + b - c))$.

Note: If we would allow weak triangle inequality for the triplet (a;b;c), we would get a similar weak inequality in the lemma. However, the weak inequality case does not give rise to any improvement potential by G-Steiner points (direct connections are called for there), hence it is of no importance to us.

Lemma 2: For triplet (a;b;c) which conforms strictly to the triangle inequality, triplet (x;y;z) solves the equation sytem:

$$(23) \qquad bx + ay = cx + az = cy + bz = \text{constant} > 0,$$

if and only if

(24) $\{x:y:z\} = \{ka(-a + b + c) : kb(a - b + c) : kc(a + b - c)\}$

for some $k > 0$.

Lemmas 1 and 2 serve to prove the next theorem, which identifies MITs as triangles for which any of the three regular trees is the regular minimal tree. The proof is by negation [20].

Theorem 10: For any triplet (a,b,c) which conforms strictly to the triangle inequality, there exists a unique set of similar MITs, one of which is the NMIT, such that for $x = d(2,3)$, $y = d(1,3)$ and $z = d(1,2) = 1$, equations (23) and (24) hold.

Finally, the next theorem shows that the best possible improvement for $|N| = 3$ by adding just one G-Steiner point is achieved for the regular (nonweighted) Steiner case when the triangle N is equilateral. Theorem 10, by the way, is our first indication that the more symmetric our case, the more improvement potential we have (we only say that for $|N| = 3$).

Theorem 11: For $|N| = 3$, $|GSMT|/|RMT| \geqslant \sqrt{3/4}$.

For an elegant new proof of this theorem, the reader is referred to Du and Hwang [7].

## 3. G-Steiner Networks for $|N| > 3$

In this section we lift the restriction $|N| \leqslant 3$. Our aim is to generalize as many results as possible for this case. Indeed we have some success, but we are far short of full success. To be specific, the result about the best basic solution being basic does not generalize (a counter-example is shown), unless we restrict ourselves to a class of G-Steiner networks with rank three G-Steiner points only. For this class, however, the G-Steiner construction is extended (by defining pseudo-trees as trees with split nodes). Since the G-Steiner construction hinges on rank three G-Steiner points, we either have basic optimality assurance <u>and</u> the construction, or <u>neither</u> of these.

Since the rank three G-Steiner class is so important, we characterize cases where it is sure to be optimal (the regular Steiner case is a trivial example). This may be a trait of g(q) or a result of the flows themselves.

When the G-Steiner construction exists, the lower bound of Theorem 9 can be extended for any basic network with a given tree or pseudo-tree configuration.

Theorem 5, about the uniqueness of a stable cycle is clearly extendible at least for the strictly stable case for a <u>given</u> configuration, and we do not repeat it.

To continue, in addition to Definitions 2 through 4 (G-Steiner cycle, simple cycle, and mixed cycle) we now define two more concepts:

<u>Definition 10</u>: A stochastic junction is an intersection of two or more arcs where all flows continue in the same direction they enter (i.e., 180°).

Note that stochastic junctions may look like G-Steiner points of even rank, yet the difference is substantial—most importantly since they do not present a location problem.

<u>Definition 11</u>: A stochastic cycle is a cycle connecting stochastic points only.

As for the cycles already defined, if they include stochastic junctions but without changes in direction there, they are maintained. However, we do not discuss mixtures of stochastic cycles with the other kinds, where some angles may be formed at stochastic junctions, others at nodes of N, and still others at G-Steiner points. Definition 6 (basic network) also holds if we ignore stochastic junctions when crossed "directly." The next definition clarifies this a little more.

Definition 12: A basic cycle is a cycle with no angles at stochastic junctions, and with one node of N at least incorporated in it. If no node of N is incorporated we call the cycle nonbasic.

The difference between a stochastic cycle and a nonbasic one can be very small but they are coneptually very different. To illustrate, Figure 14 shows

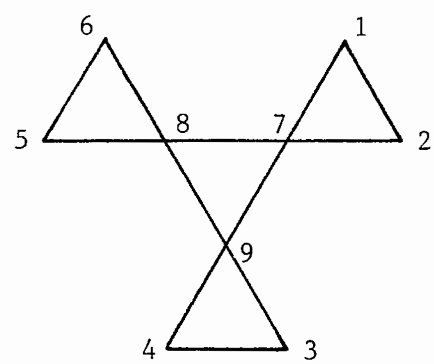



Figure 14

a case of $N = \{1,2,\ldots,6\}$ where, all flow demands, Q (with $q_{ij} = q_{ji}$, as required), and the function g(q) are, respectively:

(25) $$Q = \begin{bmatrix} 0, & 1, & h, & 1, & h, & h \\ 1, & 0, & h & h, & 1, & h \\ h, & h, & 0, & 1, & h, & 1 \\ 1, & h, & 1, & 0, & h, & h \\ h, & 1, & h, & h, & 0, & 1 \\ h, & h, & 1, & h, & 1, & 0 \end{bmatrix},$$

(26) $$g(q) = \begin{cases} \varepsilon + q; & \varepsilon, q > 0 \\ 0 & ; q = 0. \end{cases}$$

First assume h = 0; then a small enough $\varepsilon$ exists to justify a direct arc for each flow of 1, as depicted in the figure. E.g., $\varepsilon = 0.2$ is small enough. Now cycle 7,8,9,7 is a stochastic cycle. However, if h is set to some very small positive value (such that $h/\varepsilon$ is "small enough"), then we should not yet change the design, but let the h flows go through their respective shortest paths, such as $\overline{1,7,8,6}$ for $q_{16}$, etc. Now the formerly stochastic cycle becomes a nonbasic cycle, since 7,8 and 9 are now legitimate G-Steiner points (even if, due to symmetry, we would not move them).

Note that in this case the G-Steiner construction cannot serve us to locate the junctions optimally, since the G-Steiner points are of rank > 3 (and besides, we do not have an "anchor" where we can begin).

If we want to maintain the G-Steiner construction, then, we have to restrict ourselves to the class of rank three G-Steiner points networks.

<u>Definition 13</u>: The class of all G-Steiner networks where all G-Steiner points (if any) are of rank three is called the rank three G-Steiner class (R3GS class or R3GSC in short).

Results for the R3GS Class

Theorem 12: For the R3GS class, if an optimal solution includes a G-Steiner cycle, then the best basic solution is optimal too.

Proof: As in Theorem 2, we proceed by maximal parallel shifts of the arcs on the G-Steiner cycle. Lemma 3 (below) stipulates that it will result in a closed cycle; and a direct inductive extension of the proof to Theorem 2 shows that optimality is preserved during the shift. It remains to discuss how the shift may terminate. (i) It may terminate with a basic cycle (success); (ii) it may terminate with the cycle eliminated, i.e., if it was the only cycle, with a tree (success); and, finally (iii) it may get stuck in a point where two G-Steiner points merge to form a rank four G-Steiner point in such a manner that it cannot be resplit to two (other) G-Steiner points. Case (iii), depicted in Figure 15 (where it applies to one direction of the shift only, but other examples exist where both directions would so "fail"), is the reason why we have to restrict our theorem to the R3GS class. □

Lemma 3: For R3GSC, given a G-Steiner cycle at equilibrium, if we perform a parallel shift along its arcs (starting anywhere), the result is still a closed cycle, or (in case of some maximal shifts) the cycle degenerates to single Steiner point(s).

Proof: Initially, we show by counterexample that equilibrium is necessary here. Figure 16 takes care of that for us. The rest of the proof is by induction. We first show that in a triangle the lemma holds (see Figure 3), since if we have equilibrium then $\Delta 7,6,4$ is similar to $\Delta 7,3,8$; $\Delta 7,4,5$ is similar to $\Delta 7,8,9$; and $\Delta 7,5,6$ is similar to $\Delta 7,9,3$ and our result follows immediately. Assume now the lemma holds for $k - 1$ sided cycles, for

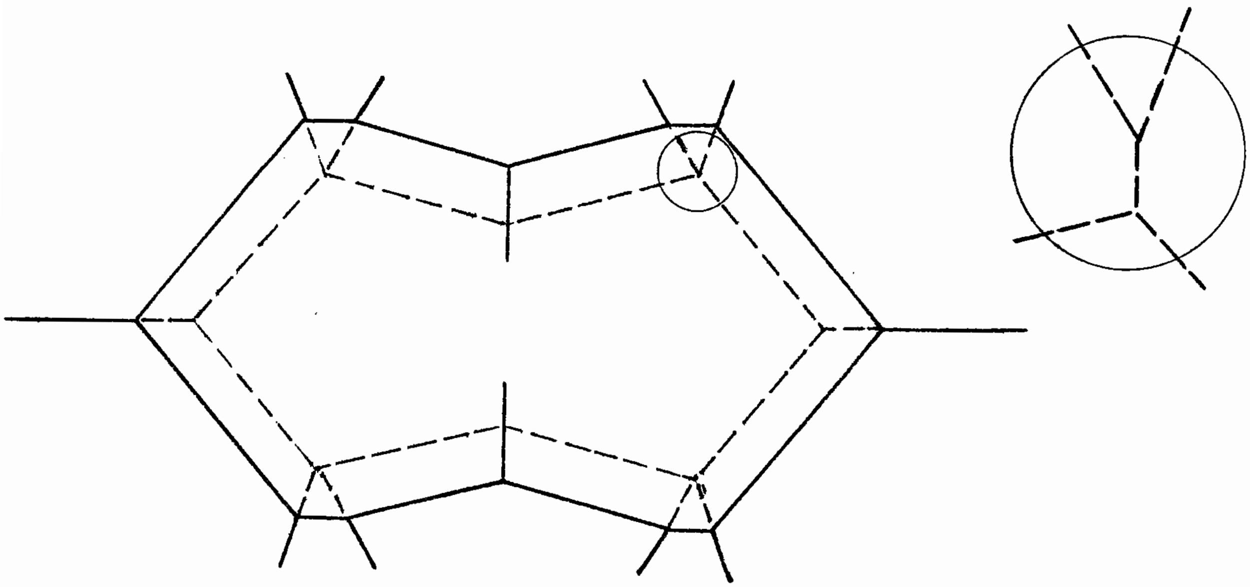


Figure 15

k = 4,5,..., and we have to prove for k. Three mutually exclusive and exhaustive possibilities exist:

1. At least one pair of sides have just one other side between them and if we extend them they meet outside the cycle in such a manner that if we ignore the intermediate side we have k - 1 sides (Figure 17a).

2. No pair as in (1) exists, but two such arcs exist where after extension they meet within or on the cycle (Figure 18).

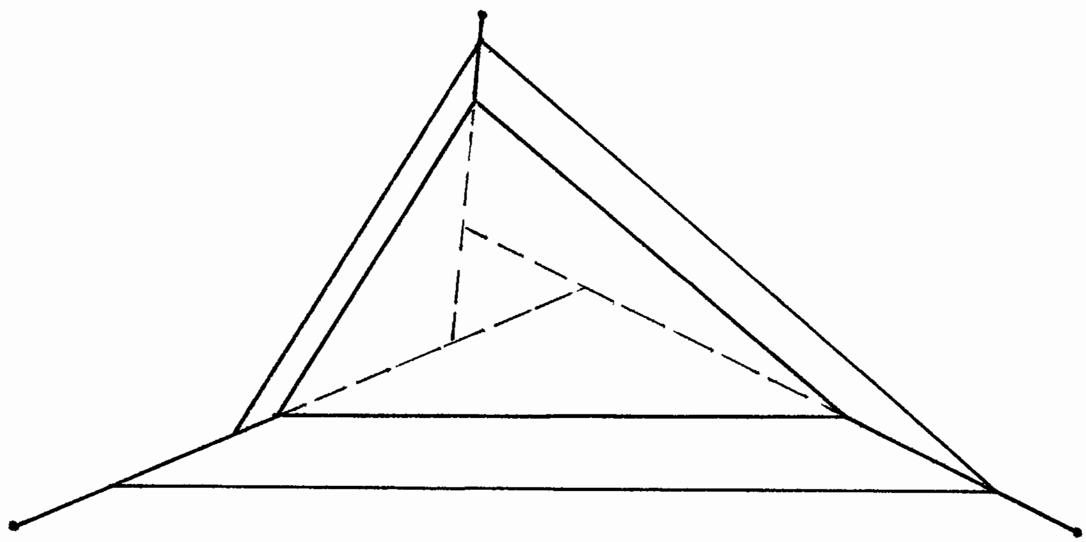


Figure 16

3. Neither 1. nor 2., so our cycle must be a parallelogram (hence k = 4; Figure 19).

Before discussing each of the three caes, we need two more definitions. Of the three arcs at each G-Steiner point, two are sides of the cycle, and the third is directed out of the cycle, or into it (see Figure 15). We define three vectors for each G-Steiner point, as follows.

Definition 14: An external vector is directed from the G-Steiner point out of (or into) the cycle in the direction of the third connection to the G-Steiner point, and its magnitude is $g(q_a)$ where $q_a$ is the flow there.

Definition 15: An inner vector is any one of the two vectors extending from a G-Steiner point in the direction of a side of the cycle incident to it, with a magnitude of $g(q_a)$, where $q_a$ is the flow there.

Definition 16: $q^*_{ij}$ is the total flow on arc i,j.

We proceed to discuss our three cases, one by one.

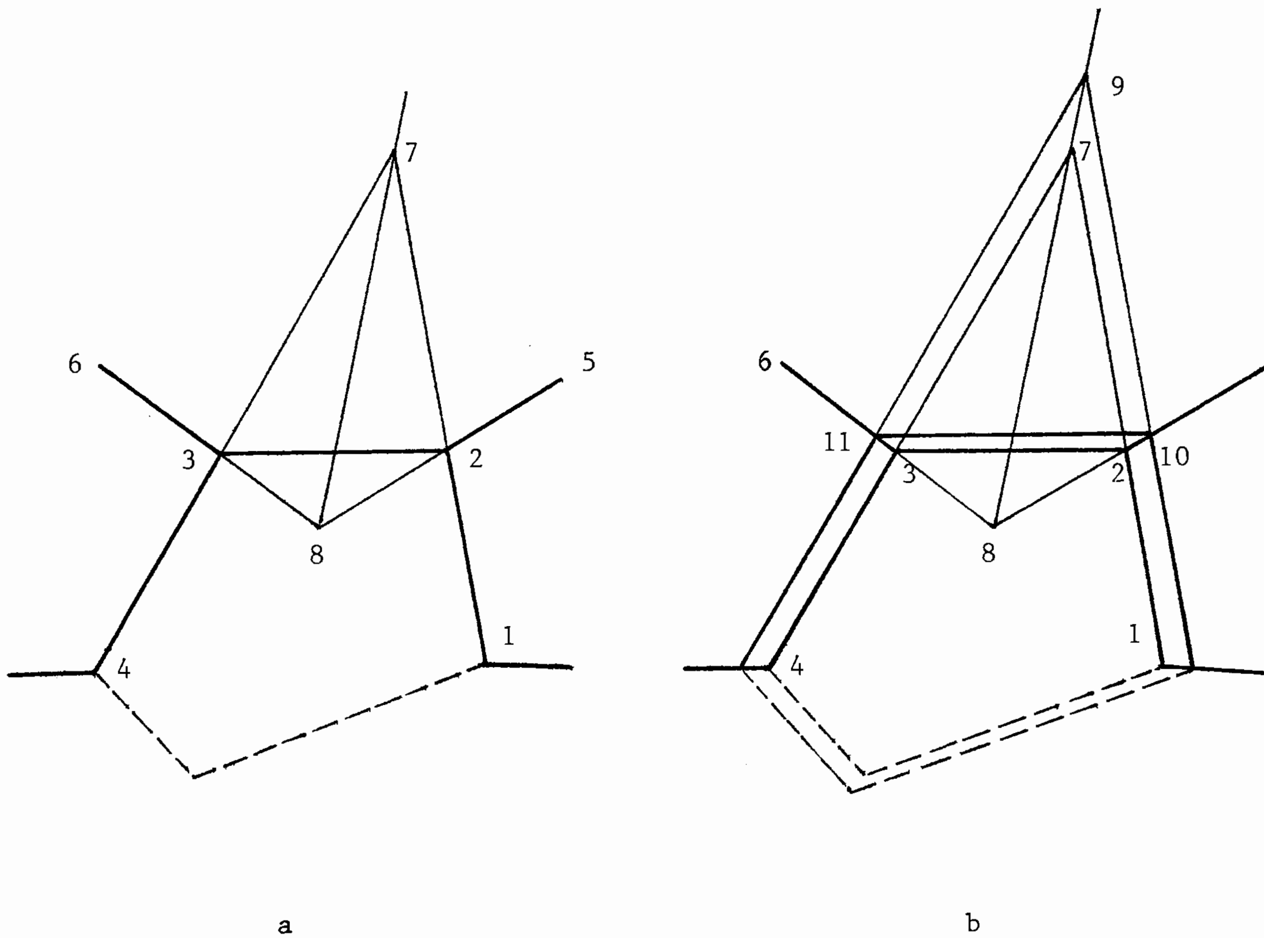



Figure 17

Case 1 (Figure 17): Point 7 is where the extended lines of $\overline{1,2}$ and $\overline{3,4}$ meet, and 8 is the meeting point of $\overline{2,5}$ and $\overline{3,6}$ (the direction of the external vectors), i.e., 8 is the action point of the two external vectors of 2 and 3. From 7 extend an external vector to nullify the two internal ones associated with it (with magnitudes $g(q^*_{12})$, $g(q^*_{34})$). Now, this external vector must be in the direction $\overline{8,7}$, since it is actually the vector sum of the two external vectors meeting there as mentioned before. To see this more clearly note that the external vector in 2 is the sum of an internal vector in the direction $\overline{7,2,1}$ with magnitude $g(q^*_{12})$ plus a vector in the direction $\overline{2,3}$ at $g(q^*_{23})$; the external vector at 3 is a similar vector sum of a vector in the direction of $\overline{7,3,4}$ at $g(q^*_{34})$ minus a vector in the direction of $\overline{2,3}$ at $g(q^*_{23})$). Adding these two vectors, the $\overline{2,3}$ directed ones cancel out, and those internal at 7 remain. But, before and after we constructed 7 we had a k and k - 1 sided cycles at equilibrium, hence the resultant external vector in 7 must operate through point 8, otherwise we would not have an equilibrium. Now, for the k - 1 sided cycle, as per Figure 17b, we perform a parallel shift, and we just have to show that $\overline{10,11}$ is parllel to $\overline{2,3}$ to complete our proof for this case. Note that $\Delta 8,9,10$ and $\Delta 8,7,2$ are similar (parallel shift), and so are $\Delta 8,9,11$ and $\Delta 8,7,3$. Therefore,

(27) $$d(2,8)/d(8,10) = d(7,8)/d(8,9) = d(3,8)/d(8,11).$$

Hence $\Delta 8,2,3$ is similar to $\Delta 8,10,11$, and $\overline{10,11}$ is indeed parallel to $\overline{2,3}$.

Case 2 (Figure 18): By careful observation of Figure 18, we can see that even though the external resultant vector at 8 is now directed into the cycle, we still have the same case as above.

Case 3 (Figure 19): This is actually a limiting case of 1, and if we "cut" the network as indicated in the figure, the resultant vector depicted must be parallel to the sides or it would not be able to nullify them exactly, and no equilibrium would be possible. Even if this resultant is compressed to a point the same proof applies. Incidentally, in this case a parallel shift

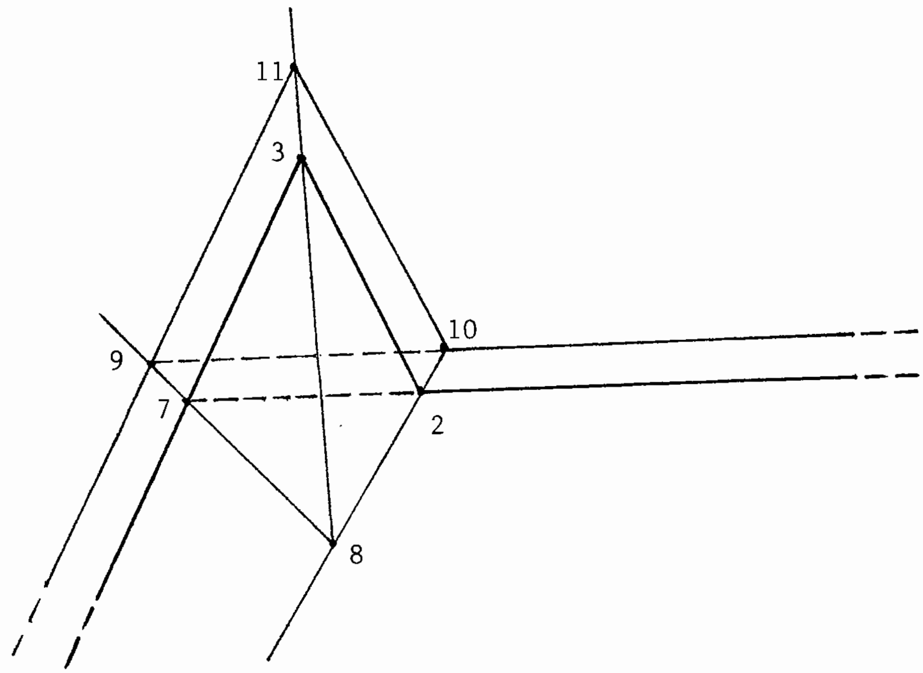



Figure 18

inwards would result in a GST depicted in broken lines. This completes our proof of the lemma. □

We mention again, however, that our result applies to R3GSC only. E.g., Figure 15 depicts a cycle where a parallel shift outwards results in a basic case, but an inwards one gets stuck with rank 4 G-Steiner points. If these cannot be "resplit" as indicated in the circle in the figure, we have a nonbasic solution there. This could conceivably happen in both directions, and not merely one as here.

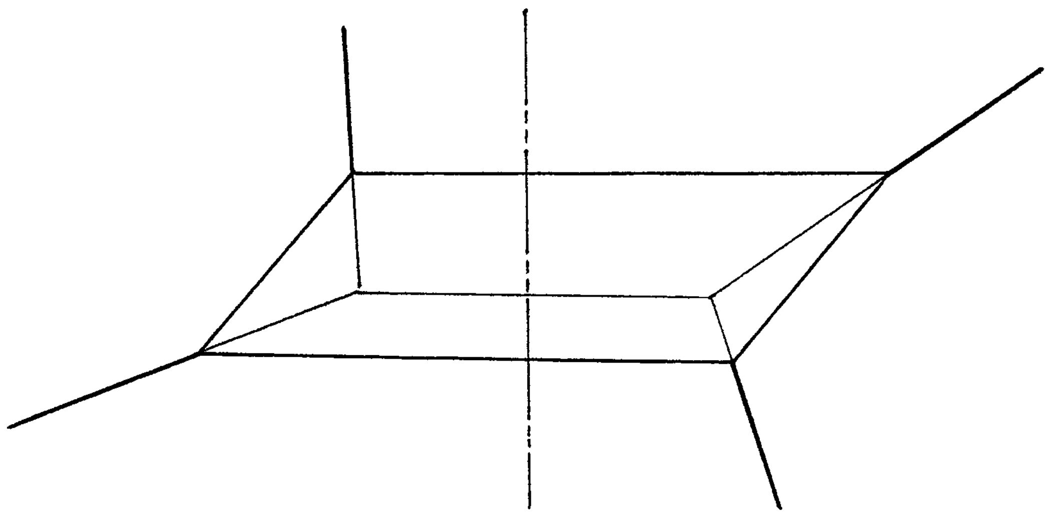


Figure 19

## Trees and Pseudo-Trees

Definition 17: A pseudo-tree is a cycled network without G-Steiner cycles (i.e., basic), where for each cycle, one of the nodes of N (which must be) on it, is counted twice, or "split" conceptually, thus opening the cycle.

Examples: In Figure 1 case a is not basic, case d is a tree, but cases b and c are cycled basic networks and respective possible pseudo-trees are shown for them in Figure 20. In the figure the split points are shown slightly apart for illustration. In both cases other nodes of N could be chosen for splitting. Our second example is the case depicted in Figures 4a and 21. Here we split node 3 to $3'$ and $3''$, and use the G-Steiner construction as per the Cockayne notation $(1,(2,3')),3''$ or (as in the figure) $(3'',1),(2,3')$.

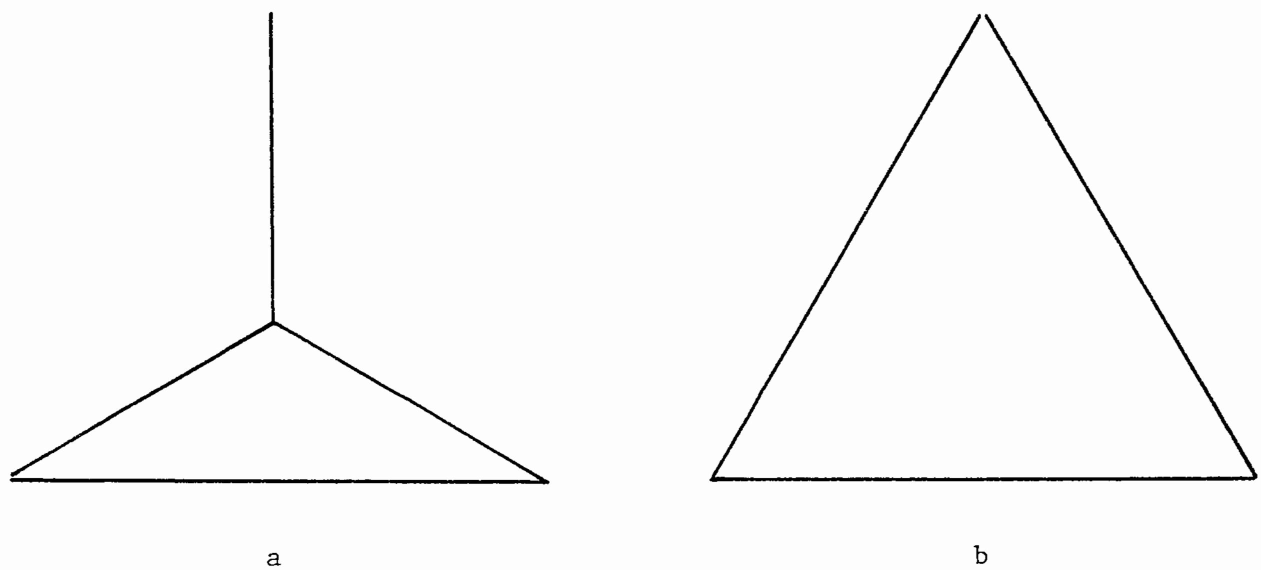



Figure 20

Now, in the R3GS class, we have Theorem 12 which assures us that nonbasic solutions may be discarded without danger of losing optimality, and we can also obtain a G-Steiner pseudo-tree for any basic solution. Hence, for this class we have a countable set of solutions, each executable by the G-Steiner construction. However, if two G-Steiner points degnerate "into each other," and if this happens for the topology representing "resplitting" as well (as in the circle in Figure 15), then we cannot carry out the G-Steiner construction. However, we are no longer in the R3GS class in this case, so if we restrict ourselves to this class we can dismiss it. If we do not discard such cases, their solution usually requires the use of nonlinear programming search methods.

Since we are discussing the R3GS class now, it is interesting to note when we can tell in advance that this class contains the optimum.

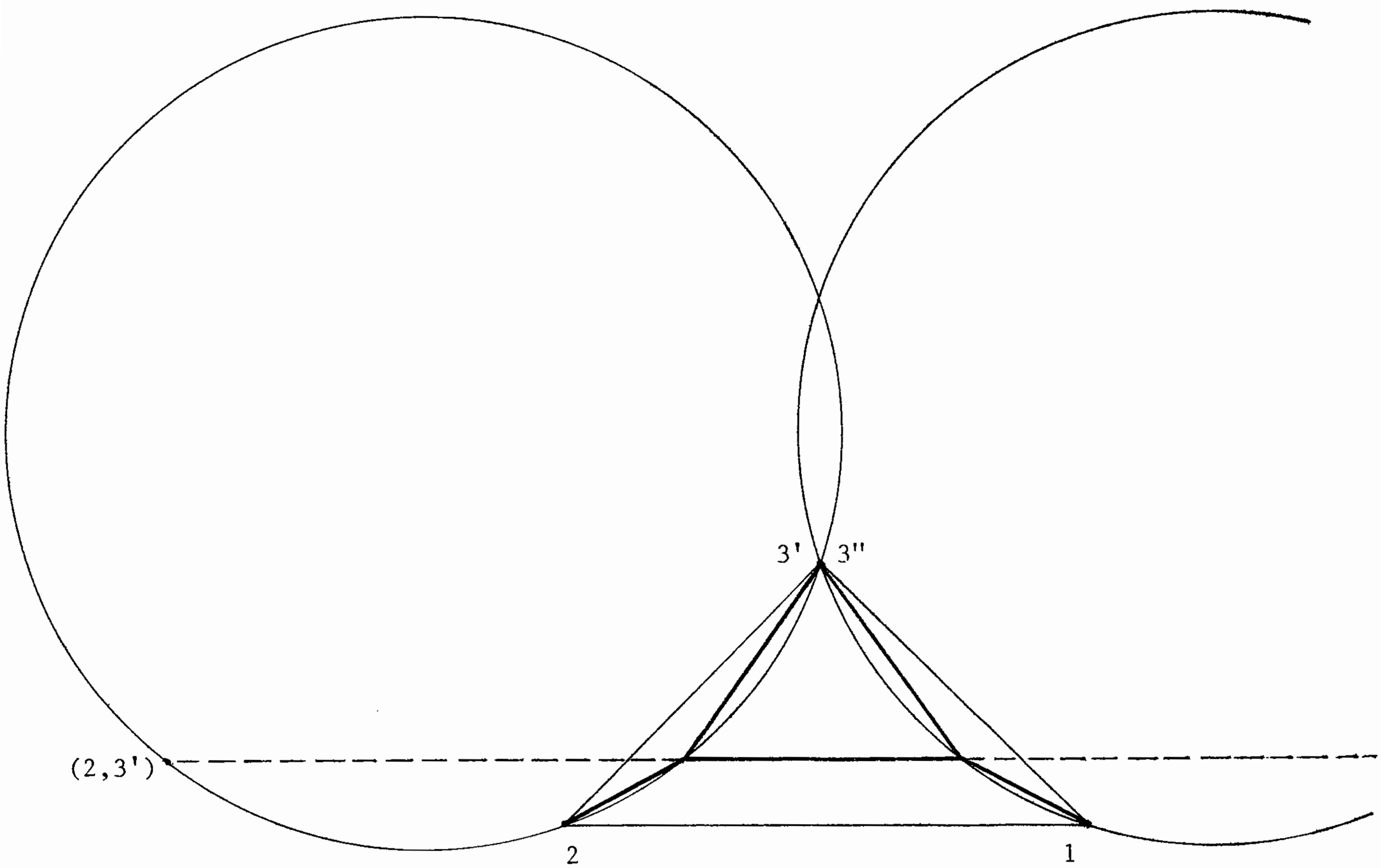



Figure 21

Cases Where G-Steiner Points Must be of Rank Three

The next theorem is a sufficient condition for the optimality of rank three G-Steiner points.

Theorem 13: If, for any $q_1, q_2 > 0$

(28) $$g^2(q_1) + g^2(q_2) > g^2(q_1 + q_2)$$

holds, no rank four or more G-Steiner points or stochastic points need be considered.

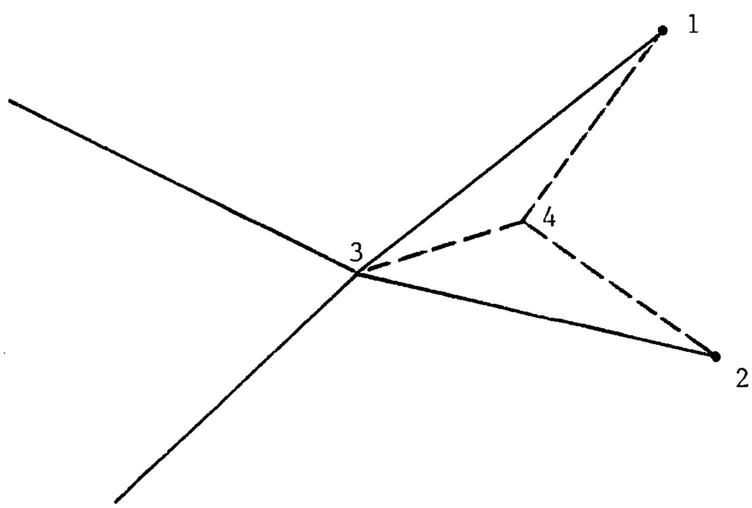



Figure 22

Proof (Figure 22): By negation. Assume a rank > 4 junction point exists, then at least one angle formed there is not obtuse, such as ∢ 1,3,2 in the figure. For this angle, by inserting a G-Steiner point such as 4, we perform a perturbation test, and the flows assigned to the arcs are $q_{13}^*$ on $\overline{1,4}$; $q_{23}^*$ on $\overline{2,4}$ and $q_{13}^* + q_{23}^* - q_{12}$ on $\overline{3,4}$. Note that $q_{13}^*$ and $q_{23}^*$ are $q_{13}^* + q_{12}$ and $q_{23} + q_{12}$, respectively, if we take $q_{13}, q_{23}$ as all the flows from 1 and 2 to all nodes of N except 2 and 1, respectively. Denote $q_{13}$ (as discussed above) by $q_1$, $q_{23}$ by $q_2$ and $q_{12}$ by $\varepsilon$, and clearly $\varepsilon \geqslant 0$, then the flows on $\overline{1,4}$, $\overline{2,4}$ and $\overline{3,6}$ are $q_1 + \varepsilon$, $q_2 + \varepsilon$ and $q_1 + q_2$, respectively, as implied by the figure. Now, at point 4, if it does not degenerate (which we assume for a while), a vector equilibrium exists for the three vectors directed to 1, 2, and 3 with magnitudes of $g(q_1 + \varepsilon)$, $g(q_2 + \varepsilon)$, and $g(q_1 + q_2)$, respectively. But, by (3) and (28)

$$(29) \qquad g^2(q_1 + \varepsilon) + g^2(q_2 + \varepsilon) \geq g^2(q_1) + g^2(q_2) > g^2(q_1 + q_2).$$

Hence, by Pythagoras' Theorem, $\sphericalangle$ 1,4,2 is obtuse! It follows that 4 does not degenerate, at least not to 3, and point 3 does not pass the perturbation test. 4 may degenerate to 1 or 2, though, but the rank of 3 would be reduced by one on that event as well. □

One of the cases where this sufficient condition holds is, of course, the regular Steiner case, where $g(q) = \text{const.}; \forall\, q > 0$. Actually, let

$$(30) \qquad g_{ab\alpha}(q) = \begin{cases} a + b \cdot q^{\alpha}; \ q > 0; \ a,b,\alpha \geq 0; \ a + b > 0 \\ 0 \quad ; \ q = 0 \end{cases}$$

be a family of functions. Then for $b = 0$ the regular Steiner case is obtained, while for $a = 0$, $\alpha = 0.5$ a boundary case is obtained where (28) becomes an equality. For any $\alpha \in [0, 0.5)$, however, (28) holds.

On the other hand, for $a = 0$, $\alpha = 1$, a direct arc is indicated for each pair i,j where $q_{ij} > 0$. (Note that for this case g is convex.)

If $\alpha \in [0.5, 1)$ it may happen that rank four or more points are indicated, but it need not necessarily happen, since condition (28) is not necessary. For $a = 0$, $\alpha = 0.5$ we can certainly assume that rank four points do not exist except as "marginally degenerate" rank three points, which can still be found by the G-Steiner construction! Incidentally, $a = 0$, $\alpha = 0.5$ fits nicely the case of water pipes; hence, water networks can be considered to belong to the R3GS class.

Now let $\alpha \equiv 1$, $a > 0$, namely,

$$g_{ab}(q) = \begin{cases} a + b \cdot q; & a,b,q > 0 \\ 0 & ; q = 0. \end{cases} \tag{31}$$

I.e., for positive flows we have a linear function plus a positive intercept for g. Clearly $g_{ab}$ can serve at least as an approximation for many applications. Unfortunately, for such flow cost functions, cases where (28) is violated can be constructed easily; e.g., let $a \to 0^+$, and we approach $g_{ab\alpha}(q)$ for $a = 0$, $\alpha = 1$, where we use direct connections (Figure 15). A similar situation occurs when the flows are very large. Hence, for $g_{ab}(q)$, Q should be taken into account when we try to determine if (28) holds (while in Theorem 13 (28) was supposed to hold for any Q).

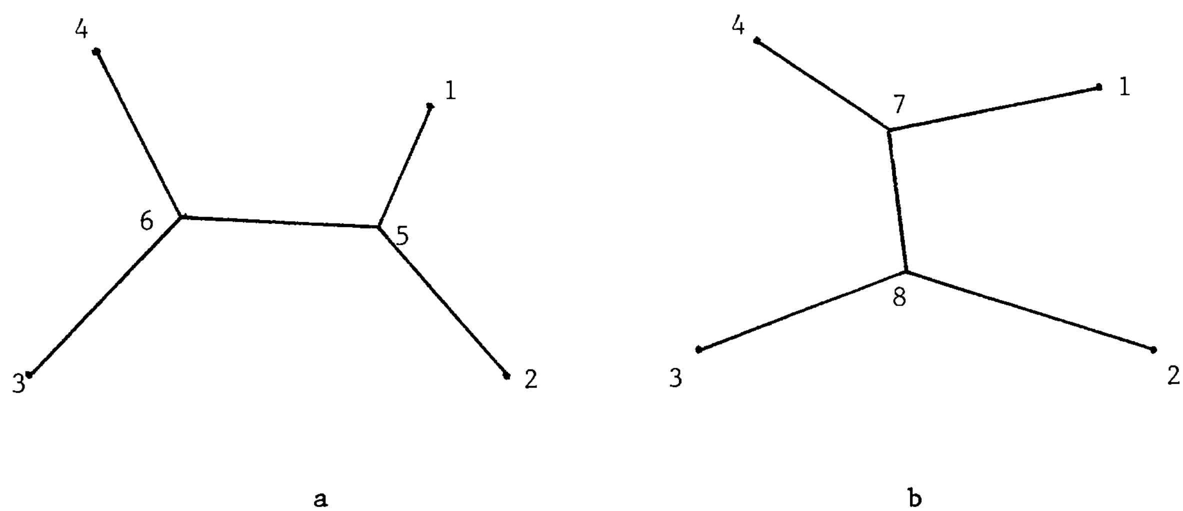



Figure 23

Figure 23 shows the two full G-Steiner tree topologies for $|N| = 4$. Let $g_i$ be the cost $g_{ab}(\sum_{j \neq i} q_{ij})$, i.e., the cost of an arc carrying all the flows of node i to the other nodes. Similarly, let $g_{56}$ and $g_{78}$ be the costs $g_{ab}(q^*_{56})$ and $g_{ab}(q^*_{78})$, respectively. Then, for Figure 23a, at point 5, we may derive the following expressions:

(32) $$g^2_{56} = g^2_{ab}(q^*_{56}) = a^2 + 2abq^*_{56} + (bq^*_{56})^2,$$

(33) $$g^2_1 = a^2 + 2abq^*_{15} + (bq^*_{15})^2,$$

(34) $$g^2_2 = a^2 + 2abq^*_{25} + (bq^*_{25})^2.$$

In addition, recall that

(35) $$q^*_{15} + q^*_{25} \geqslant q^*_{56}.$$

Then (28) holds if

(36) $$a^2 + 2ab(q^*_{15} + q^*_{25} - q^*_{56}) + b^2((q^*_{15})^2 + (q^*_{25})^2 + (q^*_{56})^2) > 0.$$

A sufficient condition for this, in turn, is

(37) $$a^2 + 2ab(q^*_{15} + q^*_{25} - q^*_{56}) - 2b^2 q^*_{25}\, q^*_{15} > 0.$$

Other sufficient conditions would be

(38) $$(q^*_{15})^2 + (q^*_{25})^2 > (q^*_{56})^2,$$

(39) $a^2 > 2b^2 q^*_{15} q^*_{25} = 2b^2 g_1 g_2 \Leftrightarrow a > b\sqrt{2g_1 g_2}$.

Of these, (38) has the advantage that it does not depend on a and b. E.g., take the case $q_{ij} = 1$; $\forall$ i,j, and (38) holds. Analog bounds can be obtained for points such as 6, 7 and 8, of course.

On occasions where close nodes of N generate a high flow demand, as implied for instance by the gravity model in transportation, (38) tends to hold.

Bounds for G-Steiner Network

As a conjecture, for the concave g(q) case we already have the generalized Gilbert and Pollak conjecture bound of

(40) $|GSMN|/|RMN| \geq \sqrt{3/4}$.

However, to calculate this bound, even under the conjecture, we still have to find the regular minimal network for the $|RMN|$ value! This problem is NP-complete [12]. Similarly the bound of [10] for the $g_{ab}$ case is NP-hard.

For a given full configuration of a G-Steiner tree or pseudo-tree in the R3GS class, however, we can extend Theorem 9, to obtain a lower bound by any of the segments associated with the G-Steiner construction. Before showing this extended theorem, we need some more preparations.

We have already defined a full G-Steiner tree (GFST). We now extend this to the pseudo-tree case.

Definition 18: A full G-Steiner pseudo-tree (GFSPT) is a G-Steiner pseudo-tree with m split nodes and n + m - 2 G-Steiner points.

Note that we need split nodes just for cycles associated with internal faces of the planar graph in which the network is embedded; e.g., in a figure such as "θ," two nodes have to be split, even though three cycles can be identified there; however, by two splits we can obtain an "H," which is a tree.

Theorem 14: For any G-Steiner tree or pseudo-tree which is not full, a full configuration exists, which degenerates to it, and if the G-Steiner tree is stable, then the full configuration is unviable, and would lead to degeneracy during the G-Steiner construction.

Proof: By a tentative insertion of G-Steiner points near nodes of N with ranks of two or more, between pairs of arcs there, we simultaneously obtain a full configuration and a perturbation test. If our tree or pseudo-tree is stable, then clearly the test must fail. □

Now, we are ready to extend Theorem 9.

Theorem 15: For any full configuration of a G-Steiner tree or pseudo-tree, if we carry out the representations as per its Cockayne notation, until we obtain a weighted segment, the value associated with it is a lower bound for the resulting tree or pseudo-tree, whether or not degeneracy occurs.

Proof: For $n = 3$, $m = 0$ we proved this in Theorem 9. For other cases, a trivial induction suffices. □

See [18] for an application of this theorem in the regular Steiner case. The same ideas apply to the G-Steiner case as well.

Heuristics

Two heuristics suggest themselves for the G-Steiner network design problem. We discuss them for the R3GS class, but they can be extended to the more general case.

The first heuristic is by aggregation, where we solve for two kinds of networks essentially: (i) an outer major backbone network connecting the aggregated subsets; and (ii) internal networks within the subsets. At a more advanced stage, local improvements can be made at the interface, as per the myopic heuristic, which we describe in some detail below as our second heuristic. This hierarchy aggregation idea can be extended to super-backbone, intermediate and internal networks, etc. The heuristic is exponential, but by specifying enough levels and using fast heuristics for the aggregation itself, it can be performed in polynomially bounded time.

Our second heuristic, the greedy or myopic heuristic is, as the names imply, based on local improvements, performed by order of immediate benefit. We start with a "good" initial network with P = N ("regular"), or even with some G-Steiner points. Now, by the perturbation test procedure we look for the best (if any) G-Steiner point of rank three which can be incorporated, insert it and reiterate until a local minimum is achieved. (So far we have an almost direct generalization of an algorithm described in [2] for Steiner trees.) We now check for "double insertions" as defined below. These have the capability of creating stable mixed cycles even if we had none before, and insert them in a greedy manner as well. Of course, we can choose to be a little less myopic, and look two or more steps ahead before choosing which candidate G-Steiner point to insert at each iteration. It now remains to define the double insertion, and afterwards we show an example.

Definition 19: For any rank three G-Steiner point, with the three arcs associated with it, terminating at points $x_1$, $x_2$ and $x_3 \in P$, a double insertion is achieved by adding another G-Steiner point, directly connected to

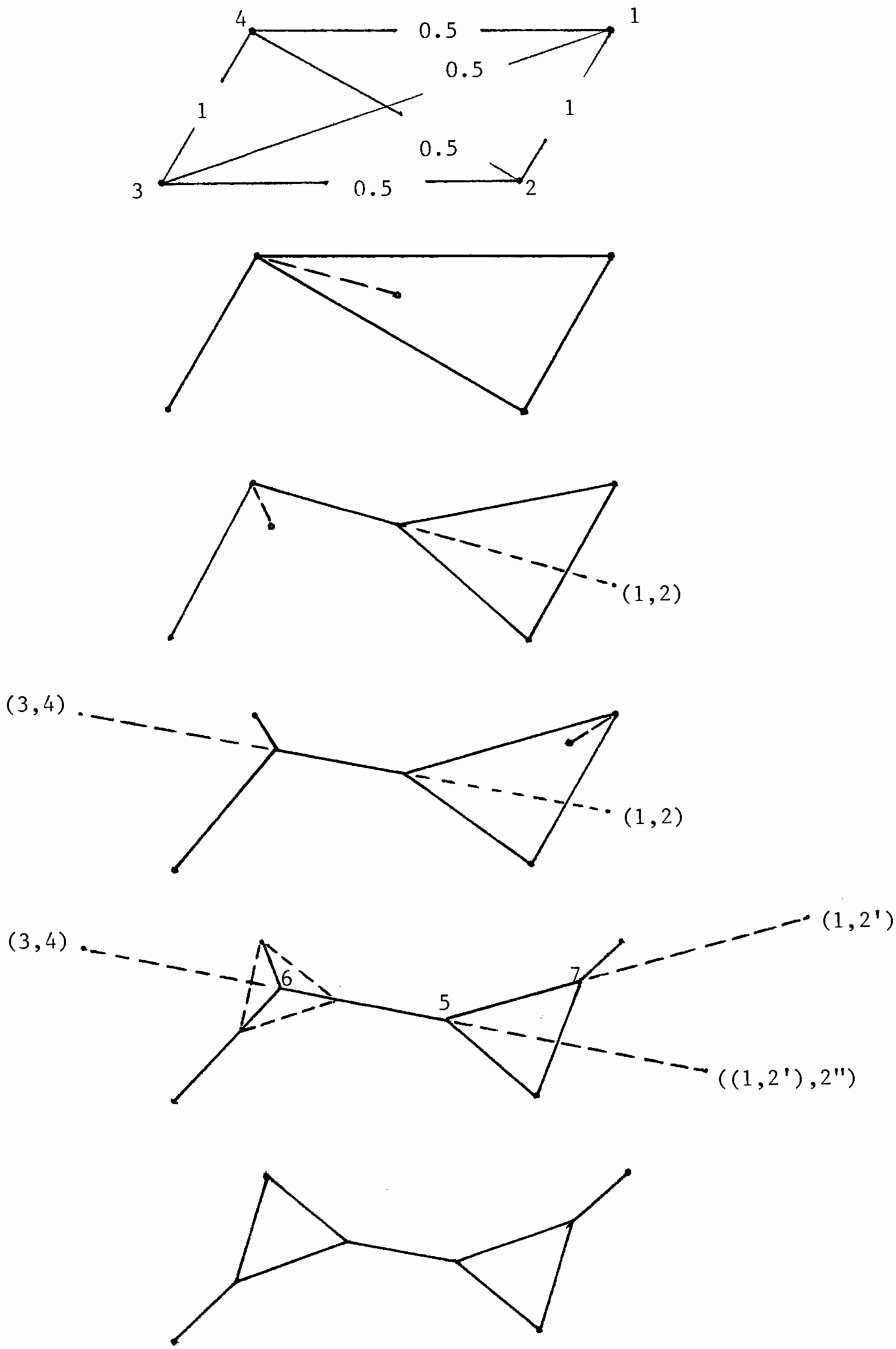



Figure 24

the former G-Steiner point and to two of the set $\{x_1,x_2,x_3\}$.

Note that in the greedy heuristic, with or without the double insertion procedure, after inserting a tentative G-Steiner point, all the other G-Steiner points already incorporated in the curernt tentative solution may have to be relocated to maintain equilibrium. The following example illustrates this point.

The Greedy Heuristic--An Example (Figure 24):

$N = \{1,2,3,4\}$, and it forms a parallelogram with a basis of 2 units and sides of 1 unit, with angles of 60° and 120°. $q_{12} = q_{34} = 1$, $q_{ij} = 0.5$ for any other pair, as illustrated in the top of the figure. For g(q) we take $g_{0.25,1}(q)$ of the $g_{ab}(q)$ family as in (31). Note that the problem is symmetric.

A best regular network for this case is depicted in the second part of the figure, costing 8.1651. (If we exchange $\overline{1,4}$ with $\overline{2,3}$ we get another best regular network, symmetric to the first, but note that neither of them is symmetric in itself.)

With this network we start inserting G-Steiner points as illustrated, while the objective function value drops from 8.1651 to 8.021 in the first four "single" insertions. However, more can be achieved by a double insertion landing us in the optimal solution (here), which also happens to be attractively symmetric, with a value of 7.94, or 2.75% improvement (almost 1% at the last "double" step). Furthermore, had we begun with the symmetric tree $\overline{1,2} + \overline{2,4} + \overline{4,3}$ we would obtain the same result in the end (with two double insertions), with an improvement of 5.4% this time. True, the tree is not optimal, but it is difficult to find the optimal regular network in the general case, so we may have landed on an inferior one, and thus the potential value of this heuristic is more than in the nonweighted Steiner case, where the RMT is easy to find.

Finally we note that the aggregation heuristic can serve to obtain a valid but rough lower bound on our objective function value, but it would be difficult to apply and not very tight. Besides that use, it seems that the greedy heuristic is more attractive, but both can be used, and the better result chosen.

## References

[1] D. E. Boyce (ed.), Transportatation Network Design. Trans. Res. 13B, 1 (1979) (a special issue).

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

[3] F. R. K. Chung and R. L. Graham, A New Bound for Steiner Minimal Trees (to appear).

[4] F. R. K. Chung and F. K. Hwang, A Lower Bound for the Steiner Tree Problem. SIAM J. Appl. Math. 34 (1978), 27-36.

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

[6] R. Courant and H. Robbins, What is Mathematics? Oxford University Press (1941).

[7] D. Z. Du and F. K. Hwang, On a Conjecture by Trietsch and Handler on the Flow-Dependent Steiner Ratio.

[8] D. Z. Du, F. K. Hwang and E. Y. Yao, The Steiner Ratio Conjecture is True for Five Points. J. Comb. Theory Series A (to appear).

[9] 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.

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

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

[12] D. S. Johnson, J. K. Lenstra and A. H. G. Rinooy Kan, The Complexity of the Network Design Problem. Networks 8 (1978), 279-285.

[13] H. W. Kuhn, "Steiner's" Problem Revisited. In G. B. Dantzig and B. C. Eaves (eds.), Studies in Optimization, Studies in Mathematics 10 (1974), 52-70, the Mathematical Association of America.

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

[15] A. D. Pearman, The Structure of the Solution Set to Network Optimization Problems. Transp. Res. 13B (1979), 81-90.

[16] H. O. Pollak, Some Remarks on the Steiner Problem. J. Comb. Theory Series A 24 (1978), 278-295.

[17] R. C. Prim, Shortest Connection Networks and Some Generalizations. Bell Sytem Technical Journal 36 (1957), 1389-1401.

[18] D. Trietsch, An Improved Algorithm for Steiner Trees. Discussion Paper No. 635R, CEMSMS, Northwestern University (1985).

[19] D. Trietsch, Connecting Euclidean Networks--The Steiner Case. Discussion Paper No. 640, CEMSMS, Northwestern University (1985).

[20] D. Trietsch and G. Y. Handler, The Gilbert and Pollak Conjectures--A Generalization. To appear in Networks.