

A NEW BOUND FOR EUCLIDEAN STEINER MINIMAL TREES

F. R. K. Chung and R. L. Graham

*Bell Laboratories
Murray Hill, New Jersey 07974*

INTRODUCTION

Let X denote a finite set of points in the (Euclidean) plane. A *minimum spanning tree* for X , denoted by $M(X)$, is a connected network of line segments joining various pairs of points of X so that the sum of the lengths of the line segments, called the *length* of $M(X)$, is as small as possible. We denote this minimum possible total length by $L_M(X)$.

It may happen for a set X that by embedding X in a suitable larger set Y , $L_M(Y)$ is actually less than $L_M(X)$. For example, if X consists of the three vertices of an equilateral triangle and Y consists of X together with the centroid of the triangle, then an easy calculation shows that

$$\frac{L_M(Y)}{L_M(X)} = \frac{\sqrt{3}}{2} = 0.866025\dots$$

For any set X we define a *Steiner minimal tree* $S(X)$ for X to be a minimum spanning tree $M(Y)$ having the least possible length over all sets $Y \supseteq X$. We denote this minimum possible length by $L_S(X)$. Although a set X may have many different Steiner minimal trees, the length $L_S(X)$ is uniquely determined. Points in $Y - X$ are called *Steiner points* of $S(X)$. Points in X are called *regular points*. Note that $L_S(X) \leq L_S(Y)$ for any $X \subseteq Y$.

The problem of determining $S(X)$ and $L_S(X)$ for a given set X has an old and venerable history, dating back to Steinhilber, Maxwell, Steiner, and, in a primitive form, to Fermat. As is the case with many combinatorial optimization problems, the (Euclidean) Steiner minimal tree problem is known to be *NP*-complete [4]. (For a complete discussion of this concept, see [5].) This has reinforced recent efforts to find good approximations for $S(X)$ that can be computed efficiently. One of the most natural of these is just the minimum spanning tree $M(X)$. A common measure of how well $M(X)$ approximates $S(X)$ in length is given by the worst-case ratio

$$\rho := \inf_X \frac{L_S(X)}{L_M(X)},$$

where X ranges over all finite sets in the plane.

A long-standing conjecture, due to Gilbert and Pollak [6], is the following:

Conjecture. $\rho = \sqrt{3}/2 = 0.866025\dots$

If true, this would be best possible, as the previously noted example of an equilateral triangle shows. This conjecture has been verified in the case where $|X| \leq 5$ (see [3, 10]) and also if $S(X)$ contains at most one Steiner point (see [9]).

Over the past few years, lower bounds for ρ have improved from the observation of E. F. Moore (see [6]) that $\rho \geq 1/2$ (in fact, for any metric space), to $\rho \geq 1/\sqrt{3} = 0.57735\dots$ by Graham and Hwang [7] (for any Euclidean space), and then to

$$\rho \geq (2 + 2\sqrt{3} - \sqrt{7 + 2\sqrt{3}})/3 = 0.74309\dots$$

by Chung and Hwang [1]. Very recently, Du and Hwang [11] proved that $\rho \geq 0.8$. The purpose of this paper is to tighten the bound on ρ even further with the following result.

THEOREM.

$$\rho \geq \rho_0 = 0.82416874\dots, \tag{1}$$

where ρ_0 is the unique real root of the polynomial

$$P(x) = x^{12} - 4x^{11} - 2x^{10} + 40x^9 - 31x^8 - 72x^7 + 116x^6 + 16x^5 - 151x^4 + 80x^3 + 56x^2 - 64x + 16$$

satisfying $0.8 < \rho_0 < 1$.

We point out that if X_n denotes a regular simplex in Euclidean n -space, then it is known [2] that the ratio $L_S(X_n)/L_M(X_n)$ can be arbitrarily close to $\sqrt{3}/(4 - \sqrt{2}) = 0.66984\dots$ for n sufficiently large so that the previously mentioned lower bound of $1/\sqrt{3}$ is not as bad as it looks. In the case of the *rectilinear* plane, with the (l_1) distance between (x, y) and (x', y') given by $|x - x'| + |y - y'|$, and where $L_{RS}(X)$ and $L_{RM}(X)$ denote the lengths of the rectilinear Steiner minimal tree for X and the rectilinear minimum spanning tree for X , respectively, Hwang [8] proved

$$\frac{L_{RS}(X)}{L_{MS}(X)} \geq \frac{2}{3}, \quad \text{for all } X,$$

and that further, this is best possible [as the four points $(\pm 1, \pm 1)$ show, for example]. Thus, the Euclidean plane seems to be a somewhat more stubborn case.

Before proceeding to the main results of the paper, we should comment on its general style. Even a casual perusal shows that a fair amount of algebraic computation is displayed. We felt that this was unavoidable for two reasons. First, in fairness to the reader who might really wish to verify the various claims, we tried to include enough detail so that it would actually be possible. In fact, we omitted most of the actual computations and, of course, all of the computer symbolic computations we needed. Second, on more than one occasion in the past, assertions made by various authors in proofs of bounds for this problem have proven to be incorrect (usually because of sketchy and incomplete arguments). We hope to avoid this pitfall by being more explicit than usual.

PROPERTIES OF STEINER MINIMAL TREES

In this preliminary section we will mention several properties of Steiner minimal trees that will be needed later. Some of these can be found in the excellent (but now somewhat out-of-date) survey of Gilbert and Pollak [6].

LEMMA 1. (i) In any Steiner minimal tree $S(X)$ for X we can assume without loss of generality that every Steiner point is incident to exactly three edges of $S(X)$, each of which meets the other two at 120° .

(ii) If $|X| = n$, then $S(X)$ has at most $n - 2$ Steiner points. If $S(X)$ has $n - 2$ Steiner points it is called a full Steiner tree.

(iii) Any Steiner minimal tree $S(X)$ for X can be decomposed into full Steiner trees in the following sense. There always exist unique subsets $X_i \subseteq X$ such that $|X_i \cap X_j| \leq 1$ for $i \neq j$, $S(X)$ restricted to X_i is a full Steiner tree on X_i , and these full trees partition the set of all edges of $S(X)$.

LEMMA 2 [10]. Let four points A, B, C, D be given so that both possible full Steiner trees exist. Then A and B are adjacent to a common Steiner point in the unique Steiner minimal tree if and only if the angle AEB is acute where E is the intersection of lines AD and BC (see FIGURE 1).

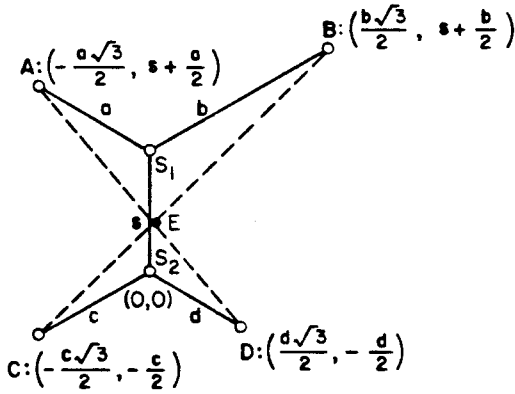


FIGURE 1

With coordinates assigned as shown in FIGURE 1 we have

$$\text{slope } AD = -\frac{s + (1/2)(a + d)}{(\sqrt{3}/2)(a + d)},$$

$$\text{slope } BC = \frac{s + (1/2)(b + c)}{(\sqrt{3}/2)(b + c)}.$$

Thus, by LEMMA 2, if A and B have a common Steiner point S_1 and the angle AEB is acute, then we have

$$\frac{s + (1/2)(a + d)}{(\sqrt{3}/2)(a + d)} \cdot \frac{s + (1/2)(b + c)}{(\sqrt{3}/2)(b + c)} \geq 1,$$

i.e.,

$$(2s + a + d)(2s + b + c) \geq 3(a + d)(b + c). \tag{2}$$

In fact, for $a' \leq a, b' \leq b, c' \leq c, d' \leq d$, we have

$$(2s + a' + d')(2s + b' + c') \geq 3(a' + d')(b' + c'). \tag{2'}$$

THE PROOF OF THE THEOREM

We now proceed to the proof of (1). It will be enough to show that for any finite set X in the Euclidean plane,

$$\frac{L_S(X)}{L_M(X)} \geq \rho_0, \tag{3}$$

where ρ_0 is defined as in (1). The proof is by induction on $|X|$. Inequality (3) certainly holds if $|X| \leq 4$ as we previously noted. Assume that (3) holds for all \bar{X} with $|\bar{X}| < n$ for a fixed integer $n \geq 5$ and suppose $|X| = n$ violates (3), i.e.,

$$\frac{L_S(X)}{L_M(X)} < \rho_0.$$

We will eventually derive a contradiction through a sequence of claims. The occasionally rather formidable algebraic manipulations needed in the proofs of the claims have been relegated to the Appendix.

CLAIM 1. Any Steiner minimal tree $S(X)$ on X must be a full Steiner tree. For if $S(X)$ were not full, then by LEMMA 1, we could partition the edges of $S(X)$ into (full) Steiner minimal trees on smaller sets $X_i \subseteq X$, and

$$\frac{L_S(X)}{L_M(X)} \geq \frac{\sum_i L_S(X_i)}{\sum_i L_M(X_i)} \geq \rho_0$$

since by induction, for each i

$$\frac{L_S(X_i)}{L_M(X_i)} \geq \rho_0.$$

Thus, we can also assume that some Steiner point S_1 of $S(X)$ is adjacent to two points A and B of X and another Steiner point S_2 .

Let $d(U, V)$ denote the (Euclidean) distance between U and V . Referring to FIGURE 2, we can assume (without loss of generality) that $d(A, S_1) = a \leq b = d(B, S_1)$ and $d(A, B) = 1$.

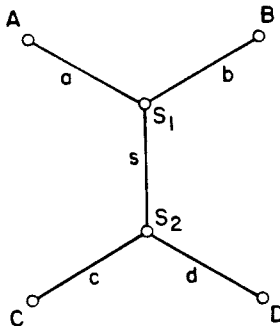


FIGURE 2

Note that $a^2 + ab + b^2 = 1$. Set

$$d(S_1, S_2) = s, \quad d(C, S_2) = c, \quad d(D, S_2) = d.$$

Note that at least one of C and D must be a Steiner point.

CLAIM 2. $s + c + d < ((\rho_0 - b)(2a + b - \rho_0))/(a + 2b - 2\rho_0)$.

Proof. Let s' denote $L_S(\{A, C, D\})$ and let X' denote $X - \{B\}$. Suppose

$$s + a + b + c + d - s' \geq \rho_0.$$

Now,

$$L_S(X') \leq L_S(X) - (s + a + b + c + d) + s'$$

since the right-hand side is the length of *some* tree connecting X' together. Since $|X'| < n$, then by induction

$$L_S(X') \geq \rho_0 L_M(X').$$

Combining these we obtain

$$\begin{aligned} \frac{L_S(X)}{L_M(X)} &\geq \frac{L_S(X) - (s + a + b + c + d) + s' + \rho_0}{L_M(X') + 1} \\ &\geq \frac{\rho_0 L_M(X') + \rho_0}{L_M(X') + 1} = \rho_0 \end{aligned}$$

which is a contradiction. Thus, we can assume

$$m + a + b - s' < \rho_0, \quad \text{where } m = s + c + d.$$

A straightforward computation gives

$$s' = (a^2 + am + m^2)^{1/2}.$$

Therefore,

$$(m + a + b - \rho_0)^2 < a^2 + am + m^2$$

which implies

$$m(a + 2b - 2\rho_0) < (\rho_0 - b)(2a + b - \rho_0).$$

Suppose we have $2\rho_0 > a + 2b$. Then by taking $a^2 + ab + b^2 = 1$, we obtain $b \geq 1.15$ or $b < 0.49$ which is impossible. Therefore $b < \rho_0$ and

$$m = s + c + d < \frac{(\rho_0 - b)(2a + b - \rho_0)}{(a + 2b - 2\rho_0)} := f(b).$$

Note that since $f(b)$ is a decreasing function in increasing b ,

$$m \leq f(1/\sqrt{3}) = 2.677\dots \blacksquare$$

CLAIM 3. *The angle AEB is acute (see FIGURE 1).*

CLAIM 3 is proved in the Appendix.

CLAIM 4. $b \leq 0.7574$ and $a \geq 0.3728$.

Proof. Choose S'_2 on the extension of S_1S_2 such that $d(S_1, S'_2) = f(b)$. It is easily checked that the angle AS'_2B is smaller than the angle AS_2B which is smaller than the acute angle AEB . Therefore by (2') we have

$$(2f(b) + a)(2f(b) + b) \geq 3ab.$$

Together with $a^2 + ab + b^2 = 1$, we obtain

$$b \leq 0.7574.$$

This also implies $a \geq 0.3728$. ■

CLAIM 5. (i) If C is a Steiner point, then there is a point $Y \in X$ with $d(A, Y) \leq d(A, C)$ or $d(A, Y) \leq d(A, S_2)$.

(ii) If D is a Steiner point, then there is a point $Z \in X$ with $d(B, Z) \leq d(B, D)$ or $d(B, Z) \leq d(B, S_2)$.

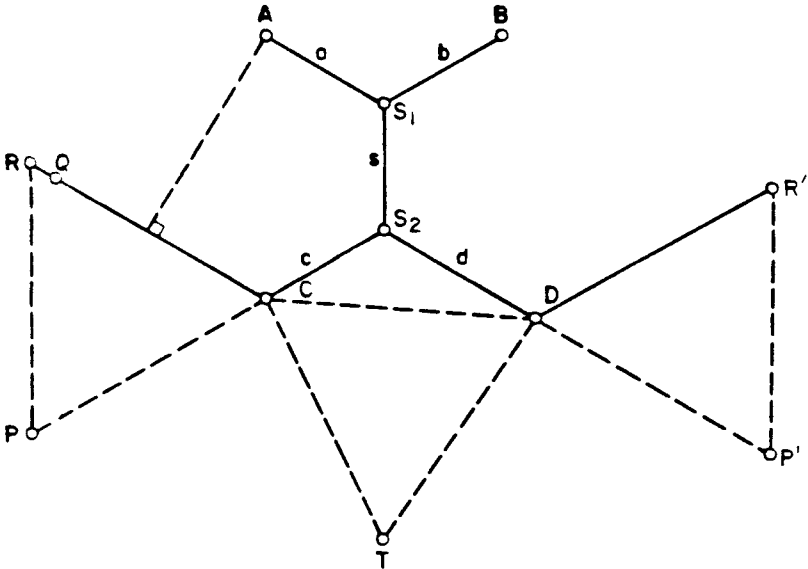


FIGURE 3

Proof of (i). Suppose C is a Steiner point but no such point Y exists. Referring to FIGURE 3, locate the point R so that RC is parallel to AS_1 and $d(R, C) = s + 2a - c$ so that $d(A, R) = d(A, C)$. If some point of X lies on the line segment RC , then we are done.

Suppose some Steiner point Q lies on RC . Let us move along the edges of $S(X)$ on $AS_1S_2CQQ_1Q_2 \dots Q_i \dots Y$ by always turning to the right at the Steiner points Q_i . If the angle ACQ is no more than 60° , we have $d(Q, C) \leq d(A, Q) \leq d(A, C)$ because of the minimality of $S(X)$. Also, we note that the angle Q_1QA is no more than 60° . Therefore we have $d(A, Q_i) \leq d(A, C)$ and $d(A, Y) \leq d(A, C)$ if Q_i is not in

the region AS_1S_2C . If the angle ACQ is greater than 60° , it can easily be seen that the angle AQC is less than 60° . By a similar argument we have

$$d(A, Y) \leq \max \{d(A, C), d(A, S_2)\}.$$

We may assume some Steiner point Q' of $S(X)$ connected to C lies on the extension of the line segment CR .

We next need the following inequality which is proved in the Appendix:

$$a + b + c + d + s + d(R, C) \geq L_S(A, C, D, R) + \rho_0. \tag{4}$$

By deleting the point B and the five edges incident to S_1 or S_2 and reconnecting the points (excluding B) by $S(A, C, D, Q')$ we have

$$L_S(A, C, D, Q') \leq L_S(A, C, D, R) + d(R, Q')$$

and so, by induction and (5),

$$\begin{aligned} \frac{L_S(X)}{L_M(X)} &\geq \frac{(L_S(X) - (a + b + c + d + s + d(C, R)) + L_S(A, C, D, R)) + \rho_0}{L_M(X - \{B\}) + 1} \\ &\geq \frac{L_S(X - \{B\}) + \rho_0}{L_M(X - \{B\}) + 1} \geq \rho_0 \end{aligned}$$

which is a contradiction. ■

Proof of (ii). The argument used to prove (ii) is analogous to that used to prove (i) except that in place of (4) we need the inequality (4') (also proved in the Appendix):

$$a + b + c + d + s + d(D, R') \geq L_S(B, C, D, R') + \rho_0. \tag{4'}$$

To complete the proof of (1) we need a final inequality (proved in the Appendix):

$$a + b + c + d + s \geq d(C, D) + \rho_0 \omega, \tag{5}$$

where

$$\omega = \min \{1 + \omega_1, 1 + \omega_2, \omega_1 + \omega_2\}$$

and

$$\omega_1 = \max \{d(A, C), d(A, S_2)\},$$

$$\omega_2 = \max \{d(B, D), d(B, S_2)\}.$$

Granting these, we then have

$$\begin{aligned} \frac{L_S(X)}{L_M(X)} &\geq \frac{(L_S(X) - (a + b + c + d + s)) + d(C, D) + \rho_0 \omega}{L_M(X - \{A, B\}) + \omega} \\ &\geq \frac{L_S(X - \{A, B\}) + \rho_0 \omega}{L_M(X - \{A, B\}) + \omega} \\ &\geq \rho_0 \end{aligned}$$

which again is a contradiction.

In the last sequence we used the induction hypothesis on the set $X - \{A, B\}$, using the fact (by CLAIM 5) that there are points of $X - \{A, B\}$ relatively close to A and B .

This completes the proof of the theorem. ■

APPENDIX

Proof of CLAIM 3. Suppose the angle AEB is not acute. LEMMA 2 indicates that the full Steiner tree on $ABCD$ in the “other” direction (i.e., the full Steiner tree in which A and C are joined to the same Steiner point) does not exist. This implies (as shown in [10]) that

$$\angle ACS_2 + \angle BDS_2 < 60^\circ \quad \text{or} \quad \angle CAS_1 + \angle DBS_1 < 60^\circ.$$

Case 1. $\angle ACS_2 + \angle BDS_2 < 60^\circ$.

We then have

$$\frac{a + s}{c + s} = \frac{\sin \theta_1}{\sin (60^\circ + \theta_1)},$$

$$\frac{b + s}{d + s} = \frac{\sin \theta_2}{\sin (60^\circ + \theta_2)},$$

where $\theta_1 = \angle ACS_2$ and $\theta_2 = \angle BDS_2$.

Since $\theta_1 + \theta_2 < 60^\circ$, we have

$$\begin{aligned} (c + s) + (d + s) &\geq (a + s) \frac{\sin (60^\circ + \theta_1)}{\sin \theta_1} + (b + s) \frac{\sin (60^\circ + \theta_2)}{\sin \theta_2} \\ &\geq (a + s) \frac{\sin (60^\circ + \theta_1)}{\sin \theta_1} + (b + s) \frac{\sin (60^\circ + \theta_1)}{\sin (60^\circ - \theta_1)} \\ &:= \omega(\theta_1). \end{aligned}$$

Now suppose

$$\frac{d\omega}{d\theta_1} = - \frac{(a + s) \sin 60^\circ}{(\sin \theta_1)^2} + \frac{(b + s) \sin 60^\circ}{(\sin (60^\circ - \theta_1))^2}$$

has a zero at $\bar{\theta}$. It can be shown that

$$\frac{d^2\omega}{d\theta^2}(\bar{\theta}) \geq 0.$$

Therefore ω has a minimum at $\bar{\theta}$ and

$$\omega(\bar{\theta}) = (\sqrt{a + s} + \sqrt{b + s})^2.$$

Thus

$$c + d > a + b + 2\sqrt{(a + s)(b + s)}$$

and $f(b) \geq a + b + s + 2\sqrt{(a + s)(b + s)}$. Since $\theta_1, \theta_2 < 60^\circ$, we have $c \geq a$ and $d \geq b$.

Choose B_0, C_0, D_0 on BS_1, CS_2, DS_2 , respectively, such that $d(B_0, S_1) = d(C_0, S_2) = d(D_0, S_2) = a$. Then the subtree of $S(x)$ connecting A, B_0, C_0, D_0 is minimal. Therefore, $s \geq (\sqrt{3} - 1)a$.

We have

$$f(b) \geq \sqrt{3}a + b + 2\sqrt{\sqrt{3}a(b + (\sqrt{3} - 1)a)}.$$

Combining this with the fact that $a^2 + ab + b^2 = 1$ and $b \geq a$, we get

$$b \leq 0.561 \quad \text{or} \quad b \geq 0.853$$

which is impossible. Therefore Case 1 cannot occur.

Case 2. $\angle CAS_1 + \angle DBS_1 < 60^\circ$.

Then by a method similar to that in Case 1, we have

$$a + b > c + d + 2\sqrt{(c + s)(d + s)}.$$

We consider the following subcases:

Subcase (i). $c \leq a - s$ and $d \leq b - s$.

Since $\angle ACR$ and $\angle BDR'$ (see FIGURE 3) are not more than 60° , by an argument similar to that used in CLAIM 5, it can be shown that there are regular points $d(A, Y) \leq d(A, S_2)$ and $d(B, Z) \leq d(B, S_2)$. Since $s \leq a \leq b$, the minimal spanning tree on A, B, S_2 consists of AS_2 and BS_2 . Therefore we have

$$a + b + s \geq \frac{\sqrt{3}}{2} (d(A, S_2) + d(B, S_2))$$

and

$$\begin{aligned} \frac{L_S(X)}{L_M(X)} &\geq \frac{L_S(X) - a - b - s + (\sqrt{3}/2)(d(A, S_2) + d(B, S_2))}{L_M(X - \{A, B\}) + d(A, Y) + d(B, Z)} \\ &\geq \frac{L_S(X - \{A, B\}) + (\sqrt{3}/2)(d(A, S_2) + d(B, S_2))}{L_M(X - \{A, B\}) + d(A, S_2) + d(B, S_2)} \\ &\geq \frac{\rho_0 L_M(X - \{A, B\}) + \rho_0(d(A, S_2) + d(B, S_2))}{L_M(X - \{A, B\}) + d(A, S_2) + d(B, S_2)} \end{aligned}$$

which is a contradiction.

Subcase (ii). $c > a - s$ and $b \leq b - s$.

We know that

$$2s + c + d \leq a + b.$$

Choose C' on the extension of CS_2 such that $d(C', S_2) = a + b - 2s$. Let E' denote the intersection of AS_2 and BC' . Then the angle $AE'B$ is greater than the angle AEB . Therefore from (2) we know that

$$(2s + a)(2s + b + c') \leq 3a(b + c'),$$

where $c' = a + b - 2s$, i.e.,

$$(2s + a)(a + 2b) \leq 3a(a + 2b - 2s)$$

which implies

$$s \leq \frac{(2b + a)a}{2(b + 2a)} = s_0 < a.$$

We want to show that

$$(2 - \sqrt{3})a + b > \rho_0 \sqrt{b^2 + s_0^2 + bs_0}.$$

Using $a^2 + ab + b^2 = 1$, this can be derived by straightforward manipulation (using the VAXIMA symbolic system, for example). In fact, this holds for $1 > b \geq 0.2$.

Now choose C'' on CS_2 so that $d(C'', S_2) = a - s$. Then $a + b + s + d(C'', S_2) - L_S(\{A, S_2, C\}) - \rho_0 d(B, S_2) \geq 0$ since $L_S(\{A, S_2, C''\}) = \sqrt{3}a$. We consider

$$\begin{aligned} \frac{L_S(X)}{L_M(X)} &\geq \frac{L_S(X) - a - b - s - d(C'', S_2) + L_S(\{A, S_2, C\}) + \rho_0 d(B, S_2)}{L_M(X - \{B\}) + d(B, Z)} \\ &\geq \frac{L_S(X - \{B\}) + \rho_0 d(B, S_2)}{L_M(X - \{B\}) + d(B, S_2)} \\ &\geq \rho_0. \end{aligned}$$

Again this is a contradiction.

Subcase (iii). $c \leq a - s$ and $d > b - s$.

As in (ii) we can get

$$s \leq \frac{(2a + b)b}{2(a + 2b)} := s_1 < b,$$

and

$$a + (2 - \sqrt{3})b > \rho_0 \sqrt{a^2 + as_1 + s_1^2}$$

(which holds for $1 > b > 0.43$). We can also derive a contradiction by the same method as in Subcase (ii). ■

Proof of (4). Referring to FIGURE 3, we first note that $c \geq a$ in this case. It is easily seen that

$$a + b + c + d + s + d(C, R) \leq L_S(A, B, C, D, R).$$

Choose P so that PCR forms an equilateral triangle (see FIGURE 3). Thus, PCS_2 forms a straight line and

$$L_S(A, B, D, P) = a + b + c + d + s + d(C, R)$$

(by applying LEMMA 2 and using the fact that PCR is equilateral).

Thus, by (2) we have

$$(2s + a + d)(2s + b + c') \geq 3(a + d)(b + c'), \tag{6}$$

where

$$c' = c + d(C, R) = s + 2a.$$

In fact, for $d' < d$ we have

$$(2s + a + d')(2s + b + c') \geq 3(a + d')(b + d'). \tag{6'}$$

If $f(b) \leq s + c' + d$, then by a calculation similar to that used in CLAIM 2 we can prove

$$a + b + c' + d + s - L_S(A, P, D) \geq \rho_0$$

which is equivalent to (4).

We may assume

$$f(b) \geq s + c' + d = 2(s + a) + d. \tag{7}$$

Suppose $d \geq b$. By (6') we have

$$(2s + a + b)(2s + b + c') \geq 3(a + b)(b + c'),$$

i.e.,

$$s \geq \frac{\sqrt{13b^2 + 40ab + 28a^2} - b - 2a}{6} \geq a.$$

From (7) we have

$$4a + b \leq f(b).$$

This implies

$$g(a, b) := \rho_0^2 + (-4b - 10a)\rho_0 + 3b^2 + 11ab + 4a^2 \leq 0.$$

On the other hand, we note that $g(a, b)$ can be viewed as a function of a single variable b since $a^2 + ab + b^2 = 1$ and the derivative of $g_1(b) := g(a, b)$ with respect to b is

$$\begin{aligned} \frac{d}{db} g_1(b) &= -4\rho_0 + 10\left(\frac{2b + a}{2a + b}\right)\rho_0 + \frac{14a^2 - 16b^2 - 4ab}{2a + b} \\ &\geq \frac{14a^2 - 4b^2 - 2ab}{2a + b} \geq 0 \end{aligned}$$

since $da/db = -(2b + a)/(2a + b)$, and $\rho_0 \geq 0.76 \geq b \geq a \geq 0.37$. By a straightforward calculation we have $g(a, b) \geq g_1(1/\sqrt{3}) > 0$, which is a contradiction. Therefore we may assume $d < b$. It follows from earlier arguments that there is a regular point Z such that

$$\mathbf{d}(B, Z) \leq \mathbf{d}(B, S_2) \quad \text{or} \quad \mathbf{d}(B, Z) \leq \mathbf{d}(B, D).$$

We consider the following two possibilities.

Case 1. $\mathbf{d}(B, Z) \leq \mathbf{d}(B, S_2)$.

Suppose the following is true:

$$a + b + c' + s - \mathbf{d}(A, P) - \rho_0 \mathbf{d}(B, S_2) \geq 0. \tag{8}$$

Then we have

$$\begin{aligned} \frac{L_S(X)}{L_M(X)} &\geq \frac{(L_S(X) - (a + b + c' + s + \mathbf{d}(C, R))) + L_S(A, R, C) + \rho_0 \mathbf{d}(B, Z)}{L_M(X - \{B\}) + \mathbf{d}(B, Z)} \\ &\geq \frac{L_S(X - \{B\}) + \rho_0 \mathbf{d}(B, Z)}{L_M(X - \{B\}) + \mathbf{d}(B, Z)} \\ &\geq \rho_0 \end{aligned}$$

which is a contradiction. Therefore we must have

$$a + b + c' + s - \sqrt{3}(a + s) - \rho_0(b + s) < 0,$$

i.e.,

$$s \geq \frac{(3 - \sqrt{3})a + (1 - \rho_0)b}{\rho_0 - 2 + \sqrt{3}} \geq 1.5b,$$

since $a \geq 0.3728$ and $b \leq 0.7574$.

From (7) we have

$$2a + 3b \leq f(b).$$

However, this implies $f(b) \geq 4a + b$, which has been shown to be impossible.

Case 2. $\mathbf{d}(B, S_2) \leq \mathbf{d}(B, Z) \leq \mathbf{d}(B, D)$.

In this case we have $d \geq b - s$. Suppose the following is true:

$$a + b + c + d + s - \mathbf{d}(C, D) - \rho_0(1 + \mathbf{d}(B, D)) \geq 0.$$

Then we have

$$\begin{aligned} \frac{L_S(X)}{L_M(X)} &\geq \frac{L_S(X) - a - b - c - d - s + \mathbf{d}(C, D) + \rho_0(1 + \mathbf{d}(B, Z))}{L_M(X - \{A, B\}) + 1 + \mathbf{d}(B, Z)} \\ &\geq \frac{L_S(X - \{A, B\}) + \rho_0(1 + \mathbf{d}(B, Z))}{L_M(X - \{A, B\}) + 1 + \mathbf{d}(B, Z)} \\ &\geq \rho_0, \end{aligned}$$

which is a contradiction.

Thus we may assume

$$g_2(a, b, c, d, s) := a + b + c + d + s - \mathbf{d}(C, D) - \rho_0(1 + \mathbf{d}(B, D)) \leq 0.$$

Since g_2 is increasing in c and decreasing in d we then have

$$\begin{aligned} g_3(a, b, s) &:= g_2(a, b, 0, b, s) \\ &= 2a + 2b + s - 1 - \rho_0(1 + b + s) \leq 0. \end{aligned}$$

By choosing $d' = b - s$ in (6') we have

$$s \geq \frac{\sqrt{4b^2 + 13ab + 10a^2} - b - 2a}{3} \geq \frac{b + a}{3}.$$

Therefore

$$g_3(a, b, s) \geq 2a + b - 1 - \rho_0 + 7(1 - \rho_0)(a + b)/3 \geq 0.02 > 0$$

since $a > 0.37$ and $a + b \geq 1.12$. This again yields a contradiction. This completes the proof of (4). ■

The proof of (4') is basically the same as that of (4) so the details will be omitted.

Proof of (5). The one remaining inequality to prove is (5). We consider several cases (refer to FIGURE 3).

Case 1. $\mathbf{d}(A, C) \leq \mathbf{d}(A, S_2), \mathbf{d}(B, D) \leq \mathbf{d}(B, S_2)$.

In this case (5) holds if it holds for $c = d = 0$. However, since it does hold for $c = d = 0$, Case 1 is completed.

Case 2. $\mathbf{d}(A, C) \leq \mathbf{d}(A, S_2), \mathbf{d}(B, D) \geq \mathbf{d}(B, S_2)$.

It is easily checked that $d \geq b - s = d'$ and $s \leq a$.

Suppose we have

$$a + (2 - \sqrt{3})b \geq \rho_0 \mathbf{d}(A, S_2).$$

Then we can choose D' on S_2D with $\mathbf{d}(S_2, D') = b - s$. Suppose $a + b + d' + s - L_S(B, S_2, D') \geq \rho_0 \mathbf{d}(A, S_2)$. Then

$$\begin{aligned} \frac{L_S(X)}{L_M(X)} &\geq \frac{(L_S(X) - (a + b + d' + s) + L_S(B, S_2, D')) + \rho_0 \mathbf{d}(A, S_2)}{L_M(X - \{A\}) + \mathbf{d}(A, S_2)} \\ &\geq \frac{L_S(X - \{A\}) + \rho_0 \mathbf{d}(A, S_2)}{L_M(X - \{A\}) + \mathbf{d}(A, S_2)} \\ &\geq \rho_0, \end{aligned}$$

which is a contradiction.

Thus we must have

$$a + (2 - \sqrt{3})b \leq \rho_0 \sqrt{a^2 + as + s^2}.$$

This implies

$$s \geq \frac{\sqrt{a^2(4 - 3\rho_0^2) + 28b^2} + \sqrt{3(-16b^2 - 8ab) + 16ab - a\rho_0}}{2\rho_0} := s_0.$$

We want to prove

$$g_4(a, b, s) := a + b + s - \rho_0(\sqrt{a^2 + as + s^2} + 1) \geq 0.$$

Since g_4 is increasing in s , by straightforward calculation we have

$$g_4(a, b, s_0) \geq (\sqrt{3} - 1)b + s_0 - \rho_0 \geq 0.16 > 0.$$

Recalling that $b \leq 0.7574$ and $a \geq 0.3728$, we therefore have

$$\begin{aligned} a + b + c + d + s &\geq g_4(a, b, s) + \mathbf{d}(C, D) + \rho_0(1 + \omega_1) \geq \mathbf{d}(C, D) + \rho_0(1 + \omega_1) \\ &\geq \mathbf{d}(C, D) + \rho_0 \omega \end{aligned}$$

and (5) is proved.

Case 3. $\mathbf{d}(A, C) \geq \mathbf{d}(A, S_2), \mathbf{d}(B, D) \leq \mathbf{d}(B, S_2)$.

Since any point on $L_S(X) - L_S(A, B, S_2)$ cannot be within distance b of B , there is a regular point Z' in the shaded area of FIGURE 4.

We know that

$$\begin{aligned} \frac{\sqrt{3}}{2} L_M(A, B, Z') &\leq L_S(A, B, Z') \\ &\leq \sqrt{(a + b + s)^2 + d^2} + (a + b + s)d. \end{aligned} \tag{9}$$

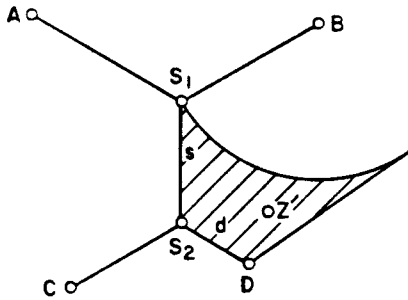


FIGURE 4

We now consider

$$g_5(a, b, d, s) := a + b + s - \rho_0 \frac{2}{\sqrt{3}} \sqrt{(a + b + s)^2 + d^2 + (a + b + s)d}.$$

Note that

$$\frac{a + b + s}{\sqrt{(a + b + s)^2 + d^2 + (a + b + s)d}} = \frac{\sin \theta}{\sin 60^\circ},$$

where θ is the angle $S_2 DW$ (see FIGURE 5).

If $\sin \theta \geq \rho_0$, we have $g_5(a, b, d, s) \geq 0$ and

$$\begin{aligned} \frac{L_S(X)}{L_M(X)} &= \frac{L_S(X) - a - b - s + \rho_0 L_M(A, B, Z')}{L_M(X - \{A, B\}) + L_M(A, B, Z')} \\ &= \frac{L_S(X - \{A, B\}) + \rho_0 L_M(A, B, Z')}{L_M(X - \{A, B\}) + L_M(A, B, Z')} \\ &\geq \rho_0 \end{aligned}$$

which is impossible.

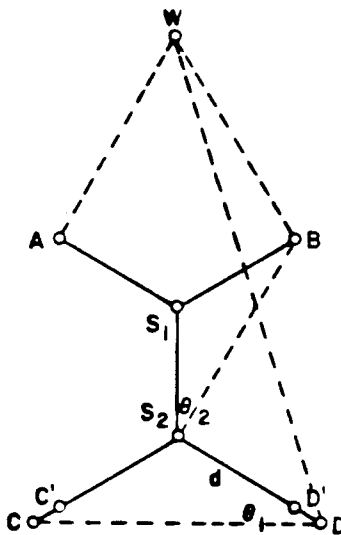


FIGURE 5

We may assume $\sin \theta < \rho_0 := \sin \theta_0$. Hence

$$d \geq (a + b + s) \frac{\sin(60^\circ - \theta_0)}{\sin \theta_0} := d_0. \quad (10)$$

Suppose we have

$$b + (2 - \sqrt{3})a \geq \rho_0 \mathbf{d}(B, S_2),$$

where $\mathbf{d}(C', S_2) = a - s$. Then

$$\frac{L_S(X)}{L_M(X)} \geq \frac{L_S(X - \{B\}) + \rho_0 \mathbf{d}(B, S_2)}{L_M(X - \{B\}) + \mathbf{d}(B, S_2)} \geq \rho_0$$

which is impossible. We may therefore assume

$$b + a\sqrt{3} \leq \rho_0 \sqrt{b^2 + bs + s^2}. \quad (11)$$

Consider the function

$$g_6(a, b, s) := 2a + b + d_0 - \mathbf{d}(C', D') - \rho_0(1 + \sqrt{b^2 + bs + s^2}),$$

where $\mathbf{d}(D', S_2) := d_0$.

Note that g_6 is decreasing in s since

$$g_6(a, b, s + \varepsilon) - g_6(a, b, s) = \varepsilon(t - t \cos \theta_1 - \rho_0 \cos \theta_2) < 0$$

when $t = \sin(60^\circ - \theta_0)/\sin \theta_0 \leq 0.1$ and $\theta_1 = \angle S_2 D' C'$, $\theta_2 = \angle S_1 S_2 B \leq 60^\circ$. Thus

$$g_6(a, b, s) \geq g_6(a, b, a) = 2a + b - 2\rho_0.$$

If $b \leq 0.6539$, then $g_6(a, b, s) \geq 0$ and

$$\begin{aligned} \frac{L_S(X)}{L_M(X)} &\geq \frac{L_S(X) - (a + b + s + c' + d') + L_S(A, C', D') + \rho_0 \mathbf{d}(B, Z')}{L_M(X - \{B\}) + \mathbf{d}(B, Z')} \\ &\geq \frac{L_S(X - \{B\}) + \rho_0 \mathbf{d}(B, Z')}{L_M(X - \{B\}) + \mathbf{d}(B, Z')} \\ &\geq \rho_0, \quad \text{a contradiction.} \end{aligned}$$

Thus, we may assume $b \geq 0.6539$. Also, if $a + b + s - \rho_0(1 + \mathbf{d}(A, C)) \geq 0$, then we have

$$\frac{L_S(X)}{L_M(X)} \geq \frac{L_S(X) - a - b - s_0 + \rho_0(1 + \mathbf{d}(A, C))}{L_M(X - \{A, B\}) + 1 + \mathbf{d}(A, C)} \geq \rho_0$$

which is impossible. Consequently, we may assume

$$\mathbf{d}(A, C) \geq \frac{a + b + s - \rho_0}{\rho_0}. \quad (12)$$

We now consider

$$g_7(a, b, c, d, s) := c + d + s - f(b)$$

for a, b, c, d and s satisfying

$$a^2 + ab + b^2 = 1 \tag{13}$$

together with (10), (11), (12) and $0.6539 \leq b \leq 0.7574$.

Let $(a_0, b_0, c_0, d_0, s_0)$ denote the 5-tuple for which g_6 achieves its minimum value m_0 . It is easily seen that a_0, b_0, c_0, d_0 , and s_0 must satisfy (10), (11), and (12) with equality. Thus, g_7 can be considered a function of the single variable b . We now outline the (brute-force) method used to prove that $g_7(b) > 0$ for $0.6539 \leq b \leq 0.7574$. The actual mechanism by which this was done was through the symbolic computation system VAXIMA, a modified form of the Macsyma system at the Massachusetts Institute of Technology.

Step 1. Determine the derivatives $da/db, dc/db, dd/db, ds/db$ in terms of a, b, c, d, s .

Step 2. Show that in the ranges

$$0.37 \leq a \leq 0.5, \quad 0.65 \leq b \leq 0.76, \quad 0 \leq s, c, d \leq 1$$

we have $|da/db| \leq 2, |ds/db| \leq 2, |dc/db| \leq 10, |dd/db| \leq 1$, and $df(b)/db \leq 0$. It then follows that $dg_7/db \geq -15$.

Step 3. Compute the values of $g_7(b_i)$ for $b_i = 0.65 + i/200, 0 \leq i \leq t$, where $b_{t-1} < 0.76 < b_t$. It turns out that $g_7(f_i) > 0.1$ for $0 \leq i \leq t$.

Step 4. Estimate $g_7(b)$ for $0.65 \leq b \leq 0.76$ using the results from Steps 2 and 3. In particular, since g_7 is continuous, then for any $b \in [0.65, 0.76]$ we can choose the largest b_i satisfying $b > b_i$ and conclude

$$\begin{aligned} g_7(b) &\geq g_7(b_i) - 15(b - b_i) \\ &\geq 0.1 - 0.075 > 0. \end{aligned}$$

By this method we can prove that $g_7(b) > 0$. However, this contradicts CLAIM 2 so we conclude that Case 3 cannot occur.

Case 4. $d(A, C) \geq d(A, S_2), d(B, D) \geq d(B, S_2)$.

We want to prove that

$$g_8(a, b, c, d, s) := a + b + c + d + s - d(C, D) - \rho_0(1 + \omega_3) \geq 0,$$

where $\omega_3 = \min \{d(A, C), d(B, D)\}$ for a, b, c, d and s satisfying (13) and

$$s + c + d \leq \frac{(\rho_0 - b)(2a + b - \rho_0)}{(a + 2b - 2\rho_0)}, \tag{14}$$

$$(2s + a + d)(2s + b + c) \geq 3(a + d)(b + c), \tag{15}$$

$$0.6539 \leq b \leq 0.7574. \tag{16}$$

Note that (14) and (15) follow from CLAIMS 1 and 2.

Let $(a_0, b_0, c_0, d_0, s_0)$ denote the 5-tuple for which g_7 achieves its minimum value m . Suppose $d(A, C) < d(B, D)$. For a small positive value ε we note that

$$g_8(a_0, b_0, c_0, d_0 - \varepsilon, s_0) < g_8(a_0, b_0, c_0, d_0, s_0)$$

and $(a_0, b_0, c_0, d_0 - \varepsilon, s)$ satisfies (13)–(15), and $d_0 - \varepsilon > 0$ since $\mathbf{d}(B, D) > \mathbf{d}(B, S_2)$. This contradicts the minimality of $(a_0, b_0, c_0, d_0, s_0)$. Similarly, if $\mathbf{d}(A, C) > \mathbf{d}(B, D)$, we derive a contradiction. Thus we may assume $\mathbf{d}(A, C) = \mathbf{d}(B, D)$, i.e.,

$$(a_0 + c_0)s_0 + a_0^2 + c_0^2 - a_0c_0 = (b_0 + d_0)s_0 + b_0^2 + d_0^2 - b_0d_0. \tag{17}$$

Suppose $(a_0, b_0, c_0, d_0, s_0)$ does not satisfy equality in (15). We consider the 5-tuple $(a_0 - \varepsilon_a, b_0 + \varepsilon_b, c_0 + \varepsilon_c, d_0 - \varepsilon_d, s_0)$, where $\varepsilon_a, \varepsilon_b, \varepsilon_c, \varepsilon_d$ are small positive values chosen such that equality holds in (13) and (14), and $\mathbf{d}(A', C') = \mathbf{d}(A, C) = \mathbf{d}(B', D')$ (see FIGURE 6). Looking at first-order approximations, we conclude:

- (i) $(2a_0 + b_0)\varepsilon_a = (2b_0 + a_0)\varepsilon_b$,
- (ii) $\varepsilon_d(s_0 + 2a_0 - c) = \varepsilon_c(s_0 + 2c - a)$,
- (iii) $\varepsilon_b(s_0 + 2b_0 - d) = \varepsilon_d(s_0 + 2d_0 - b)$.

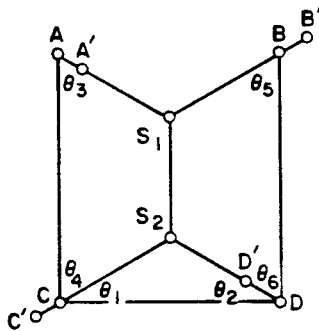


FIGURE 6

Thus

$$\begin{aligned} \varepsilon_a + \varepsilon_d - \varepsilon_b - \varepsilon_c &\geq 3\varepsilon_b \left(\frac{b_0 - d_0}{s_0 + 2d_0 - b_0} + \left(\frac{2b_0 + a_0}{b_0 + 2a_0} \right) \left(\frac{c_0 - a_0}{s_0 + 2c_0 - a_0} \right) \right) \\ &:= 3\varepsilon_b g_9(a_0, b_0, c_0, d_0, s_0). \end{aligned}$$

Since $\mathbf{d}(A, C) = \mathbf{d}(B, D)$, we have $b_0 \geq a_0, c_0 \geq d_0$. It suffices to consider the following three cases:

Subcase 1. $b_0 \geq a_0 \geq c_0 \geq d_0$.

Here, it is immediate that $g_9(a_0, b_0, c_0, d_0, s_0) \geq 0$.

Subcase 2. $c_0 \geq d_0 \geq b_0 \geq a_0$.

Then

$$\begin{aligned} g_9(a_0, b_0, c_0, d_0, s_0) &\geq \frac{b_0 - d_0}{s_0 + 2d_0 - b_0} + \frac{c_0 - a_0}{s_0 + 2c_0 - a_0} \\ &\geq \frac{(b_0 + c_0 - a_0 - d_0)s_0 + b_0c_0 - a_0d_0}{(s_0 + 2d_0 - b_0)(s_0 + 2c_0 - a_0)} \\ &\geq 0. \end{aligned}$$

Subcase 3. $b_0 \geq d_0, c_0 \geq a_0$.

Then

$$g_9(a_0, b_0, c_0, d_0, s_0) \geq \frac{(b_0 + c_0 - a_0 - d_0)s_0 + b_0c_0 - a_0d_0}{(s_0 + 2d_0 - b_0)(s_0 + 2c_0 - a_0)} \geq 0.$$

Thus we have $\varepsilon_a + \varepsilon_d \geq \varepsilon_b + \varepsilon_c$. It follows that $(a_0 - \varepsilon_a, b_0 + \varepsilon_b, c_0 + \varepsilon_c, d_0 - \varepsilon_d, s)$ satisfies (15) since the angle determined by $A'C'$ and $B'D'$ containing S_1 is no more than 90° . Therefore,

$$\begin{aligned} g_8(a_0, b_0, c_0, d_0, s_0) - g_8(a_0 - \varepsilon_a, b_0 + \varepsilon_b, c_0 + \varepsilon_c, d_0 - \varepsilon_d, s_0) \\ = \varepsilon_a - \varepsilon_b - \varepsilon_c(1 - \cos \theta_1) + \varepsilon_d(1 - \cos \theta_2) \\ > (1 - \cos \theta_1)(\varepsilon_a - \varepsilon_b - \varepsilon_c + \varepsilon_d) \geq 0, \end{aligned}$$

where $\theta_1 = \angle S_2CD$ and $\theta_2 = \angle S_2DC$. However, this contradicts the minimality of $(a_0, b_0, c_0, d_0, s_0)$. Thus $(a_0, b_0, c_0, d_0, s_0)$ satisfies the inequalities (13), (14), (15), and (17), with equality. Since we have five variables and four equations, a_0, c_0, d_0, s_0 , and g_8 can all be viewed as functions of the single variable b_0 . Again we use a brute-force technique to prove $g_8(b) \geq 0$ for $1/\sqrt{3} \leq b \leq 0.76$ as follows:

Step 1. Determine $da/db, dc/db, dd/db, ds/db$ in terms of a, b, c, d, s using (13)–(17).

Step 2. Determine dg_8/db in terms of a, b, c, d, s . It turns out that

$$|dg_8/db| \leq 1500.$$

Step 3. Determine d^2g_8/db^2 in terms of a, b, c, d, s . It turns out for this computation that

$$|d^2g_8/db^2| \leq 6 \times 10^8.$$

Step 4. Sample $(dg_8/db)(b)$ at the points $b_i \in (1/\sqrt{3}, 0.58]$ where

$$b_i = 1/\sqrt{3} + i(2.2 \times 10^{-8}), \quad i = 1, 2, \dots, 1.3 \times 10^5.$$

All these values of $(dg_8/db)(b)$ happen to be greater than 60.

Step 5. Sample $g_8(b)$ at points b_j in $[0.58, 0.76]$ where

$$b_j = 0.58 + j(6 \times 10^{-5}).$$

All these values of g_8 are greater than 0.1.

Now from Steps 3 and 5 we know that

$$\frac{df_8}{db}(b) \geq \frac{df_8}{db}(b_i) - 6 \times 10^8 \cdot 2.2 \times 10^{-8} > 0$$

for $b \in [1/\sqrt{3}, 0.58]$. Since $g_8(1/\sqrt{3}) = 0$, we have $g_8(b) \geq 0$ for $b \in [1/\sqrt{3}, 0.58]$. From Steps 2 and 5, we have, for b in $[0.58, 0.76]$,

$$g_8(b) \geq g_8(b_j) - 1500 \cdot 6 \times 10^{-5} > 0.$$

This completes the proof of (5). ■

We point out that the actual value of ρ_0 is obtained by solving the system of equations formed by taking equality in (13), (14), (15), and (17) together with the

equation $g_8(1/\sqrt{3}) = 0$. We remark in conclusion that it would appear that new ideas will be needed to prove that $\rho \geq \sqrt{3}/2$, although this does not diminish our belief in its truth.

REFERENCES

1. CHUNG, F. R. K. & F. K. HWANG. 1978. A lower bound for the Steiner tree problem. *SIAM J. Appl. Math.* **34**: 27–36.
2. CHUNG, F. R. K. & E. N. GILBERT. 1976. Steiner trees for the regular simplex. *Bull. Inst. Math. Acad. Sinica* **4**: 313–325.
3. DU, D. Z., F. K. HWANG & E. Y. YAO. The Steiner ratio conjecture is true for five points. *J. Combin. Theory Ser. A*. In press.
4. GAREY, M. R., R. L. GRAHAM & D. S. JOHNSON. 1977. The complexity of computing Steiner minimal trees. *SIAM J. Appl. Math.* **32**: 835–859.
5. GAREY, M. R. & D. S. JOHNSON. 1979. *Computers and Intractability, a Guide to the Theory of NP-Completeness*. W. H. Freeman and Co. San Francisco, Calif.
6. GILBERT, E. N. & H. O. POLLAK. 1968. Steiner minimal trees. *SIAM J. Appl. Math.* **16**: 1–29.
7. GRAHAM, R. L. & F. K. HWANG. 1976. Remarks on Steiner minimal trees, I. *Bull. Inst. Math. Acad. Sinica* **4**: 177–182.
8. HWANG, F. K. 1976. On Steiner minimal trees with rectilinear distance. *SIAM J. Appl. Math.* **30**: 104–114.
9. KALLMAN, R. 1973. On a conjecture of Gilbert and Pollak on minimal trees. *Stud. Appl. Math.* **52**: 141–151.
10. POLLAK, H. O. 1978. Some remarks on the Steiner problem. *J. Combin. Theory Ser. A* **24**(3): 278–295.
11. DU, D. Z. & F. K. HWANG. 1983. A new bound for the Steiner ratio. *Trans. Amer. Math. Soc.* **278**: 137–148.