Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 9464–9466, December 1990
Mathematics

The Steiner ratio conjecture of Gilbert and Pollak is true

(Steiner minimum trees/minimum spanning trees/minimax/convexity)

D.-Z. Du[†‡] and F. K. Hwang[§]

[†]Department of Computer Science, Princeton University, Princeton, NJ 08540; [‡]Institute of Applied Mathematics, Academia Sinica, Beijing, People's Republic of China 100010; and [§]AT&T Bell Laboratories, Murray Hill, NJ 07974

Communicated by R. L. Graham, September 12, 1990

ABSTRACT **Let P be a set of n points on the euclidean plane. Let $L_s(P)$ and $L_m(P)$ denote the lengths of the Steiner minimum tree and the minimum spanning tree on P, respectively. In 1968, Gilbert and Pollak conjectured that for any P, $L_s(P) \geq (\sqrt{3}/2)L_m(P)$. We provide an abridged proof for their conjecture in this paper.**

1. Introduction

Consider a set P of n points on the euclidean plane. A shortest network interconnecting P must be a tree, which is called a *Steiner minimum tree* and denoted by SMT(P). An SMT(P) may contain vertices not in P. Such vertices are called *Steiner points*, while vertices in P are called *regular points*. Computing SMT(P) has been shown to be an NP-hard problem (1). Therefore, it is of merit to study approximate solutions. A spanning tree on P is just a tree with vertex set P. The shortest spanning tree on P, called the *minimum spanning tree* on P, is denoted by MST(P). The Steiner ratio is defined to be

$$\rho = \inf_P \{L_s(P)/L_m(P)\},$$

where $L_s(P)$ and $L_m(P)$ are lengths of SMT(P) and MST(P), respectively. Since computing MST(P) is fast, MST(P) can be used as an approximate solution of SMT(P). In this case, the Steiner ratio is a measure for the performance of the approximation. Gilbert and Pollak (2) conjectured $\rho = \sqrt{3}/2$ and verified it for $n = 3$. The conjecture was then verified by Pollak (3) for $n = 4$, by Du *et al.* (4) for $n = 5$, and by Rubinstein and Thomas (5) for $n = 6$. Along another line, the lower bound of ρ for general n has been pushed up from 0.5 (by Moore as reported in ref. 2) to 0.57 by Graham and Hwang (6), to 0.74 by Chung and Hwang (7), to 0.8 by Du and Hwang (8), and to 0.824 by Chung and Graham (9). In either the small-n exact result or the general-n lower bound case, the lack of further progress was caused by a fast growth of computation load. In this paper, we will prove $\rho = \sqrt{3}/2$ without requiring much computation. For a general reference on SMTs, we cite a recent survey by Hwang and Richards (10).

It is well known (2) that an SMT(P) must satisfy the following conditions:

(*i*) All leaves are regular points.

(*ii*) Any two edges meet at an angle of at least 120°.

(*iii*) Every Steiner point has degree exactly three.

Conditions *ii* and *iii* together imply that every Steiner point is incident to exactly three edges; any two of them must meet at an angle of 120°.

A tree interconnecting P and satisfying conditions *i–iii* is called a *Steiner tree* (ST). Note that our definition of an ST allows two edges to cross. Its topology (the labeled graph structure of the network) is called a *Steiner topology*. An ST for n points can contain at most $n - 2$ Steiner points (2). If an ST has exactly $n - 2$ Steiner points, then it is called a full ST and its topology a full topology. Any ST can be decomposed into an edge-disjoint union of full STs. Therefore, it suffices to prove the Steiner ratio conjecture for full STs.

Our proof of the Steiner ratio conjecture shows that a convexity property of STs implies the maximality of the set of MST topologies at a point set where the Steiner ratio is achieved, which in turn implies that it suffices to consider only point sets that lie on a lattice of equilateral triangles.

2. A Maximin Problem

A full ST T can be determined by its full topology t and $2n - 3$ edge lengths of T. Without loss of generality we may normalize $l(T)$, the length of T, to be $\sqrt{3}/2$. Let x denote a vector of $2n - 3$ nonnegative numbers summing to 1 and let X denote the set of x. Usually, any edge length in T should be positive. But we will allow them to be zero so that any x in the compact set X is realizable by t. We denote T by $t(x)$. When x contains some zero elements, $t(x)$ can be seen as a limit of $t(y)$ as $y \to x$, where $t(y)$ has positive edge lengths.

Let $P(t; x)$ denote the set of regular points for $t(x)$, and let $L_t(x)$ denote the length of an MST for $P(t; x)$. Note that $P(t; x)$ is determined only up to rigid motions. For a spanning tree topology s, let $s(t; x)$ denote the spanning tree for $P(t; x)$ with topology s.

LEMMA 1. $L_t(x)$ *is continuous in* x.

Proof: $L_t(x) = \min_s l(s(t; x))$, where s is a spanning tree topology. Since each $l(s(t; x))$ is continuous in x, so is the minimum. ∎

Define

$$L(t) = \max_x L_t(x).$$

Since there are only finitely many t, there exists a t^* such that

$$L(t^*) = \max_t L(t).$$

Then proving the Steiner ratio conjecture is equivalent to proving $L(t^*) \leq 1$. We prove $L(t^*) \leq 1$ by induction on n. The proof is trivial for $n = 1$ and 2. Thus we assume $n \geq 3$. Call a point $x \in X$ maximum if $L_{t^*}(x) = L(t^*)$. It can be proved as follows.

LEMMA 2. *If* $L(t^*) > 1$, *then a maximum point is an interior point of* X.

Note that

$$L(t^*) = \max_{P(t^*;x)} \min_s l(s(t^*; x)).$$

In the next section we will prove that $l(s(t^*; x))$ is convex with respect to x.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MST, minimum spanning tree; SMT, Steiner minimum tree; ST, Steiner tree.

3. Convexity

In this section we prove convexity for the problem and its important consequences.

LEMMA 3. *Let* t *be a full topology and* s *a spanning tree topology. Then* l(s(t; x)) *is a convex function with respect to* x.

Proof: Let A and B be two regular points of the full topology t. We first show that the distance between A and B, $d(A, B)$, is a convex function of x.

Since the Steiner ratio is invariant to a rotation of the ST, we may assume a fixed orientation for the edges of the ST. Find the path in $t(x)$ that connects the points A and B. Suppose the path has k edges with lengths $x_{1'}, \ldots, x_{k'}$ and with directions $e_1, \ldots, e_k$, respectively, where $e_1, \ldots, e_k$ are unit vectors (fixed by the given orientation) in the order from A to B. It is easy to see that $d(A, B) = \|x_{1'}e_1 + \ldots + x_{k'}e_k\|$, where $\|\cdot\|$ is the euclidean norm and is convex. Note that the part inside the norm is linear with respect to x. Thus, it is straightforward to verify that $d(A, B)$ is a convex function with respect to x.

Since the sum of convex functions is also a convex function, it follows immediately from the above that $l(s(t; x))$ is a convex function of x. ∎

Let $M(t; x)$ denote the set of topologies m such that $m(t; x)$ is an MST for $P(t; x)$.

LEMMA 4. *For every* x *there is a neighborhood of* x *such that for any* y *in the neighborhood* M(t; y) ⊆ M(t; x).

Proof: For contradiction, suppose that there exists a sequence of points $\{y_k\}$ converging to x such that for each y_k an MST topology $m_k \in M(t; y_k)\backslash M(t; x)$ exists. Since the number of spanning tree topologies is finite, there exists a subsequence $\{y_{k'}\}$ of $\{y_k\}$ such that $y_{k'}$ has an MST topology m'. Thus $l(m'(t; y_{k'})) \leq l(m(t; y_{k'}))$ for each $y_{k'}$ and each $m \in M(t; x)$. Letting $k' \to \infty$, we obtain $l(m'(t; x)) \leq l(m(t, x))$. Hence $m' \in M(t; x)$. ∎

Lemma 2 tells us that if $L(t^*) > 1$, then every maximum point is an interior point of the simplex X. The next lemma uses this property to give another important property of a maximum point.

LEMMA 5. *Suppose that* x *is an interior maximum point and that* y *is a point in X satisfying* M(t*; x) ⊆ M(t*; y). *Then* y *is also a maximum point.*

Proof: For any m in $M(t^*, x)$, define $A(m) = \{z \in X_{t^*} \,|l(m(t^*; z)) \leq L_{t^*}(x)\}$. By *Lemma 4*, $A(m)$ is a convex region. We first claim that the union of all $A(m)$ for m in $M(t^*; x)$ covers a neighborhood of x. In fact, if such a union does not cover any neighborhood of x, then in every neighborhood of x, we can find a point z such that $\min\{l(m(t^*; z)) \,|m \in M(t^*; x)\} > L_{t^*}(x)$. However, by *Lemma 4*, we know that for z sufficiently close to x, $L_{t^*}(z) = \min\{l(m(t^*; z)) \,|m \in M(t^*; x)\}$. Thus, there exists z in X such that $L_{t^*}(z) > L_{t^*}(x)$, contradicting the assumption that $L_{t^*}(x) = L(t^*)$.

We now show that $L_{t^*}(y) = L(t^*)$. Suppose to the contrary that $L_{t^*}(y) > L(t^*)$. Note that $M(t^*; x) \subseteq M(t^*; y)$. Thus, for every $m \in M(t^*; x)$, $l(m(t^*; y)) < L_{t^*}(x)$. We claim that for all positive numbers c, the point $x + c(x - y)$ is not in $A(m)$ for every $m \in M(t^*; x)$. In fact, if the point $x + c(x - y)$ for some positive c is in $A(m)$, then the point x as an interior point of the segment $[y, z]$, where $z = x + c(x - y)$ can be written as $x = \lambda y + (1 - \lambda)z$, with $0 < \lambda = c/(1 + c) < 1$. By *Lemma 3*, we have

$$l(m(t^*; x) \leq \lambda l(m(t^*; y)) + (1 - \lambda)l(m(t^*; z))$$

$$< L_{t^*}(x),$$

contradicting that $m \in M(t^*; x)$. Finally, the fact that $x + c(x - y)$ for all $c > 0$ is not in $A(m)$ for every $m \in M(t^*; x)$ contradicts that the union of all $A(m)$ covers a neighborhood of x. ∎

Let x be a maximum point. Suppose that x can be moved to a point y, which is a vector of $2n - 3$ nonnegative numbers but not necessarily summing to 1, such that $M(t^*; x) \subseteq M(t^*; y)$ and $m(t^*; x) = m(t^*; y)$ for $m \in M(t^*; x)$. We remark that y is also a maximum point. To see this, note that there always exists a positive number h such that $hy \in X$. Since $M(t^*; hy) = M(t^*; y) \supseteq M(t^*; x)$, by *Lemma 5*, hy is also a maximum point. Hence $L_{t^*}(x) = L_{t^*}(hy) = h \cdot L_{t^*}(y)$. It follows that $h = 1$ and $y \in X$. Therefore, y is a maximum point, by use of *Lemma 5* again. In the next section, we will use this remark in some proofs.

4. Critical Structure

We have transformed the Steiner ratio conjecture into a maximum problem and showed some important properties of the maximum points. In this section, we translate these properties back to the original geometrical problem, which allows us to reduce the range of the problem to one special class of point sets.

Let t be a full topology and x a parameter vector. Denote by $\Gamma(t; x)$ the union of MSTs for $P(t; x)$. The following two lemmas were given by Rubinstein and Thomas (11). They are helpful for determining the structure of $\Gamma(t; x)$.

LEMMA 6. *Two MSTs can never cross; i.e., edges meet only at vertices.*

Let $C(t(x))$ denote the convex hull of $P(t; x)$. From *Lemma 6* we can see that $\Gamma(t; x)$ divides $C(t(x))$ into smaller areas, each of which is bounded by a polygon with vertices all being regular points. Such a polygon is called a polygon of $\Gamma(t; x)$ if it is a subgraph of $\Gamma(t; x)$.

LEMMA 7. *Every polygon of* Γ(t; x) *has at least two equal longest edges.*

A $\Gamma(t; x)$ is said to be *critical* if $M(t; x)$ is maximal—i.e., there does not exist $y \in X_t$ such that $M(t; x) \subset M(t; y)$. Since the number of MST topologies is finite, the set of critical Γ cannot be empty for any t. Furthermore, by *Lemma 5*, there must exist a maximum point x such that $\Gamma(t^*; x)$ is critical. When t is a full topology, the vertex set $P(t; x)$ is determined by $2n - 3$ edge lengths. An edge length is called *independent* if it can be changed while all other edge lengths are held fixed.

LEMMA 8. *Suppose that* L(t*) > 1 *and that all edges in any triangulation of* C(t*(x)) *are independent. Then* Γ(t*; x) *is critical only if it partitions* C(t*(x)) *into exactly* $n - 2$ *equilateral triangles.*

Proof: We prove that $\Gamma(t^*; x)$ cannot be critical if one of the following occurs:

(*i*) $\Gamma(t^*; x)$ has a free edge, an edge not on any polygon of $\Gamma(t^*; x)$.

(*ii*) $\Gamma(t^*; x)$ has a polygon of more than three edges.

(*iii*) Conditions *i* and *ii* do not occur, but $\Gamma(t^*; x)$ has a triangle not being equilateral.

First, assume that *i* occurs. Embedding $\Gamma(t^*; x)$ into a triangulation of $C(t^*(x))$, we can find a triangle containing the free edge e. Let e' be an edge of the triangle not in $\Gamma(t^*; x)$ such that in an MST containing e, removing e and adding e' will result in another spanning tree. [Such an edge e' must exist, for if the triangle has only one edge not in $\Gamma(t^*; x)$, then this edge must have the desired property; if the triangle has two edges not in $\Gamma(t^*; x)$, then the one that lies between the two connected components of the MST after removing e meets the requirement.] Clearly, $l(e) < l(e')$. Now, we decrease the length of e' and fix all other edge lengths in the triangulation. Let l be the length of the shrinking e'. At the beginning, $l = l(e')$. At the end, $l = l(e) < l(e')$. For each l, denote by $\overline{P}(l)$ the corresponding set of regular points. Then $\overline{P}(l(e')) = P(t^*; x)$. Consider the set L of all $l \in [l(e), l(e')]$ satisfying the condition that $\overline{P}(l) = P(t^*; y)$ for a maximum point $y = y(l)$. Since the set of maximum points is a closed set and contains the point x, the set L is nonempty and closed. Therefore, there exists a minimal element l^* in L. Suppose that $P(t^*; y) = \overline{P}(l^*)$. Since both x and y are maximum points, the length of an MST for $P(t^*; x)$ equals that for $P(t^*; y)$. Furthermore, since decreasing e' does not affect the length of any edge in $\Gamma(t^*; x)$, $M(t^*; x) \subseteq M(t^*; y)$. Suppose that $M(t^*; y) = M(t^*; x)$. Clearly, $l^* \neq l(e)$ since, when $l^* = l(e)$, dropping e and adding e' will give one more MST. By *Lemma 2*, y has no component being zero. This means that there exists a neighborhood of l^* such that for l in it, the full ST of topology t exists for the point set $\overline{P}(l)$. Thus, there exists an $l < l^*$ such that $\overline{P}(l) = P(t^*; z)$ for some length vector z. From *Lemma 4*, we know that there exists a neighborhood y such that for y' in it, $M(t; y') \subseteq M(t; y)$. Thus, z can be chosen also to satisfy that $M(t; z) \subseteq M(t; y)$. Note that for every $m, m' \in M(t; x)$, $l(m(t; z)) = l(m'(t; z))$, and $M(t; y) = M(t; x)$. It follows that $M(t; z) = M(t; x)$. By the remark we made at the end of Section *3*, z is also a maximum point, a contradiction to the assumption that l^* is minimal. Therefore $M(t^*; x) \subset M(t^*; y)$ and $\Gamma(t^*; x)$ is not critical.

For the other two cases (*ii* and *iii*), we can give similar proofs by decreasing the length of an edge not in $\Gamma(t^*; x)$ in case *ii* and by increasing the length of all shortest edges (uniformly) in $\Gamma(t^*; x)$ in case *iii*. ■

5. Inner Spanning Trees

Let T be a full ST. If we inflate the edges of T to have an ε width, then T is a polygonal region with a boundary. Two regular points are called adjacent in T if they are consecutive on the boundary. Connecting all adjacent pairs of regular points by straight lines, we obtain a polygon that may have lines crossing each other in the plane. When such a polygon is viewed as drawn on a spiral surface, it remains simple with no crossing lines. Note that all regular points are on the boundary. We call the polygon the *characteristic area* of T. For a nonfull ST T we define the characteristic area as the union of the characteristic areas of its component full Steiner subtrees. Clearly, T lies inside of its characteristic area. A spanning tree is called an *inner spanning tree* for T if it lies inside of the characteristic area of T.

It turns out that *Lemmas 2–8* still hold for inner spanning trees and *Lemma 1* follows as a corollary of *Lemma 4* (a rigorous treatment will be published elsewhere). $C(t(x))$ is now the characteristic area of $t(x)$ instead of the convex hull, and $\Gamma(t; x)$ the union of minimum inner spanning trees for $P(t; x)$. Again, $\Gamma(t; x)$ divides $C(t(x))$ into polygons. Consider a triangulation of these polygons. Since all regular points are on the boundary of $C(t(x))$, these triangles form a sequence that does not loop back into a cycle. We can construct these triangles one by one in the order of the sequence. Then, when a new triangle is added, only one side of it is already fixed by the preceding triangle. The remaining two sides can take any lengths as long as the triangular inequality is satisfied. Hence all edges of the triangulation are independent. Since $\Gamma(t^*; x)$ can be embedded into a triangulation, all edges of $\Gamma(t^*; x)$ are independent. By *Lemma 8*, we have the following.

LEMMA 9. *If* $L(t^*) > 1$, *then a critical* $\Gamma(t^*; x)$ *must partition* $C(t(x))$ *into* $n - 2$ *equilateral triangles*.

We call such a critical structure an *equilateral configuration*. Note that in unconstrained MSTs all edge lengths need not be independent. So the concept of minimum inner spanning trees is crucial to our proof.

6. The Main Result

We are now ready to prove our theorem.

THEOREM 1. *For every ST* T, *there exists an inner spanning tree* N *such that* $l(T)/l(N) \geq \sqrt{3}/2$.

Clearly, the Steiner ratio conjecture is a corollary of *Theorem 1*.

Now, in order to derive a contradiction to $L(t^*) > 1$, it suffices to show that for any ST $t(x)$ with a critical equilateral configuration, *Theorem 1* holds for $t(x)$.

Note that an equilateral configuration contains $n - 2$ equilateral triangles that form a framework fixing all regular points. Let a be the length of edge of the equilateral triangles in the configuration. If we divide the plane into a lattice of equilateral triangles with edge length a, then all regular points in an equilateral configuration can be placed on the lattice points. The following lemma is easy to prove.

LEMMA 10. *An MST for* n *lattice points has length at least* $(n - 1)a$. *For the point set* $P(t; x)$ *with a critical equilateral configuration, the minimum inner spanning tree has length exactly* $(n - 1)a$.

Proof: The first part is obvious. The second part follows immediately from the fact that any minimum inner spanning tree is an MST of an equilateral configuration. ■

By *Lemma 9*, we can see that for the point set $P(t; x)$ with a critical equilateral configuration, every minimum inner spanning tree is an MST in the plane. Thus, to show that *Theorem 1* holds for $t(x)$, it suffices to verify the Gilbert–Pollak conjecture for the point set $P(t; x)$ with a critical equilateral configuration.

We now study a different kind of tree. Given three directions each two of which meet at an angle of 120°, a shortest network interconnecting a given set P of points and having edges all parallel to the three directions is called a *minimum hexagonal tree* on P. Let $L_h(P)$ denote the length of a minimum hexagonal tree for P. By noting that each euclidean edge can be replaced by two hexagonal edges with the length increased by at most a factor of $\sqrt{3}/2$, J. Weng gave the following lemma in an unpublished work.

LEMMA 11. $L_s(P) \geq (\sqrt{3}/2)L_h(P)$.

For a given set P of points, let $H(P)$ be the hexagonal grid constructed by running three lines parallel to the edges of the equilateral configuration through each point of P and each intersection point of these lines. It can be shown:

LEMMA 12. *There exists a minimum hexagonal tree that uses only segments of* $H(P)$.

Clearly, when we consider a minimum hexagonal tree on $P(t; x)$ with the property in *Lemma 12* and with segments all parallel to edges of the equilateral configuration, it must be an MST. Thus we have

$$L_s(P) \geq (\sqrt{3}/2)L_h(P) = (\sqrt{3}/2)(n - 1)a = (\sqrt{3}/2)L_m(P).$$

Theorem 1 is proved.

We thank a referee for many helpful comments. D.-Z.D. is supported by the Center for Discrete Mathematics and Theoretical Computer Science, a National Science Foundation Science and Technology Center, under Grant STC88-09648, and by the National Science Foundation of China.

1. Garey, M. R., Graham, R. L. & Johnson, D. S. (1977) *SIAM J. Appl. Math.* **32**, 835–859.
2. Gilbert, E. N. & Pollak, H. O. (1968) *SIAM J. Appl. Math.* **16**, 1–29.
3. Pollak, H. O. (1978) *J. Combinatorial Theory Ser. A* **24**, 278–295.
4. Du, D.-Z., Hwang, F. K. & Yao, E. N. (1985) *J. Combinatorial Theory Ser. A* **35**, 396–400.
5. Rubinstein, J. H. & Thomas, D. A. (1991) *J. Combinatorial Theory Ser. A.*, in press.
6. Graham, R. L. & Hwang, F. K. (1976) *Bull. Inst. Math. Acad. Sinica* **4**, 177–182.
7. Chung, F. R. K. & Hwang, F. K. (1978) *SIAM J. Appl. Math.* **34**, 27–36.
8. Du, D.-Z. & Hwang, F. K. (1983) *Trans. Am. Math. Soc.* **278**, 137–148
9. Chung, F. R. K. & Graham, R. L. (1985) *Ann. N.Y. Acad. Sci.* . **440**, 328–346.
10. Hwang, F. K. & Richards, D. S. (1991) *Networks*, in press.
11. Rubinstein, J. H. & Thomas, D. A. (1991) *Ann. Oper. Res.*, in press.