MINIMAX AND ITS APPLICATIONS: REVISIT THE PROOF OF GILBERT-POLLAK CONJECTURE

DING-ZHU DU*
Computer Science Department
University of Minnesota
Minneapolis, MN 55455 USA
and
Institute of Applied Mathematics
Chinese Academy of Sciences
Beijing 100080, PRC

Abstract. Consider the problem $\min_{x \in X} \max_{i \in I} f_i(x)$ where X is a convex set, I is a finite set of indices and $f_i(x)$'s are continuous concave functions of x. In this article, we study a characterization of $x \in X$ at which the minimax value is achieved. We also study some applications of the characterization.

Key words: Minimax, Steiner trees.

1. Introduction

Minimax is an important topic in optimization. There are two fundamental ideas to solve minimax problems.

The first is the search for a basis, tha is, for the problem

$$\min_{x \in X} \max_{y \in Y} f(x, y)$$

determine first a finite subset B of X such that

$$\min_{x \in X} \max_{y \in Y} f(x, y) = \min_{x \in B} \max_{y \in Y} f(x, y)$$

and then search an optimal x^* from B in finitely many steps.

The second is the determination of saddle point. A point (x^*, y^*) is called a *saddle point* of $f(x, y)$ on the set $X \times Y$ if for any $x \in X$ and $y \in Y$,

$$f(x^*, y) \leq f(x^*, y^*) \leq f(x, y^*).$$

It follows that for a saddle point (x^*, y^*),

$$\min_{x \in X} \max_{y \in Y} f(x, y) = f(x^*, y^*) = \max_{y \in Y} \min_{x \in X} f(x, y).$$

These two ideas have resulted two important mathematical branches. P. L. Chebyshev is probably the first person who made an important contribution to

* Support in part by the National Science Foundation under grant CCR-9208913.

the first idea. He discovered theory of best approximation. The second idea was extensively studied by Von Neumann. He initiated game theory. Since Von Neumann, many efforts have been made to find various sufficient conditions for a point being a saddle points. This involves a great deal of mathematics including fixed point theory.

While a hudge amount of materials about minimax in the literature exist, we select only a small part in this article. In fact, including all subjects about minimax should be the duty of a book instead of an article. In this article, we will commit ourselves only to recent developments on the first idea.

2. Chebyshev Theorem

The original problem considered by Chebyshev is as follows: Given a list of values of some real function:

$$y_k = f(x_k), \quad k = 0, \ldots, m,$$

find a polynomial p of degree at most n $(n < m)$ which provides the best approximation at these m points, that is, p minimizes

$$\max_{k=0,\ldots,m} |y_k - p(x_k)|.$$

Chebyshev gave a beautiful result about the solution of this problem.

First, consider $m = n + 1$. In this case, the polynomial of the best approximation exists and is unique. Chebyshev proved that a polynomial p is the best approximation if and only if for some h,

$$(-1)^k h + p(x_k) = y_k, \quad \text{for } k = 0, \ldots, n+1.$$

Furthermore, h and p can be constructed explicitly. This p is called a *Chebyshev interpolating polynomial.*

For general m, a subset of $n+2$ x_k's is called a *basis*. Each basis σ determines a Chebyshev interpolating polynomial p_σ and a value

$$h(\sigma) = \max_{x_k \in \sigma} |y_k - p_{\sigma^*}(x_k)|.$$

A basis σ^* is called an *extremal basis* if

$$h(\sigma^*) = \max_{\sigma} h(\sigma)$$

where σ is over all bases. Chebyshev showed the following.

Theorem 2.1 *There exists a unique polynomial of best approximation. A polynomial p is the polynomial of best approximation if and only if p is a Chebyshev interpolating polynomial for some extremal basis.*

There are other ways to characterize the extremal basis. In fact, Chebyshev also proved that σ^* is an extremal basis if and only if

$$h(\sigma^*) = \max_{k=0,\ldots,m} |y_k - p_{\sigma^*}(x_k)|.$$

(See [5].) For each polynomial p, define

$$I(p) = \{i \mid |y_i - p(x_i)| = \max_{k=0,\ldots,m} |y_k - p(x_k)|\}.$$

$I(p)$ is *maximal* if no polynomial q exists such that $I(p) \neq I(q)$ and $I(p) \subset I(q)$. From the second characterization of the extremal basis, it is not hard to prove the following.

Proposition 2.2 *σ^* is an extremal basis if and only if $I(p_{\sigma^*})$ is maximal.*

3. Linear Programming

Chebyshev problem can be transformed to a linear programming as follows:

$$\begin{array}{rl} \min & z \\ \text{subject to} & -z \leq a_0 + a_1 x_k + \cdots + a_n x_k^n - y_k \leq z \\ & k = 0, \ldots, m. \end{array}$$

Note that this linear programming has $n+2$ variables and $2(m+1)$ constraints. For an extremal basis σ^*, p_{σ^*} would make $n+2$ constraints *active* (i.e., the equality sign holds for those constraints). This means that each extremal basis corresponds to a feasible basis of the above linear programming in the following standard form.

$$\begin{array}{rl} \min & z \\ \text{subject to} & u_k - z = a_0 + a_1 x_k + \cdots + a_n x_k^n - y_k = z - v_k \\ & u_k \geq 0, \quad v_k \geq 0 \\ & k = 0, \ldots, m. \end{array}$$

Linear programming are closely related to minimax problems. In fact, there are several ways to transform linear programming to a minimax problem. For example, consider a linear programming

$$\begin{array}{rl} \min & cx \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}$$

and its dual

$$\begin{array}{rl} \max & b^T y \\ \text{subject to} & A^T y \leq c. \end{array}$$

For any feasible solution x of the original linear programming and any feasible solution y of the dual linear programming, $cx \geq b^T y$. The equality sign holds only if the two feasible solutions are actually optimal solutions for the two linear programming, respectively. This is equivalent to the following minimax problem achieves the minimax value 0.

$$\min_{(x,y)} \max(cx - b^T y, -x, Ax - b.b - Ax, A^T y - c).$$

4. Du-Hwang Theorem

In the previous two sections, we see already two problem in the following form:

$$\min_{x \in X} \max_{i=1,\cdots,m} f_i(x).$$

Now, we consider it with a little general conditions. We assume that X is a polytope in $\mathbf{R}^n$ and $f_i(x)$'s are continuous concave functions of x. We will extend Chebyshev's idea to this problem.

The simplest case is $m = n = 1$. As shown in Figure 1, the minimum value of a concave function $f_1(x)$ on the interval $[a, b]$ is achieved at a or b. For $m = 1$ and

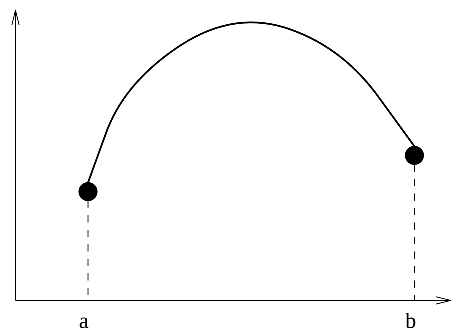

Fig. 1. The minimum point of a concave function.

general n, it is well-known that the minimum value of $f_1(x)$ is achieved at a vertex of the polytope X. What we are interested in this article is the case $m > 1$. If $m > 1$ and $n = 1$, then as shown in Figure 2, $g(x) = \max_{i=1,\cdots,m} f_i(x)$ is a piecewise concave function. Thus, the minimum value of $g(x)$ on the interval $[a, b]$ is achieved at an endpoint of a concave piece.

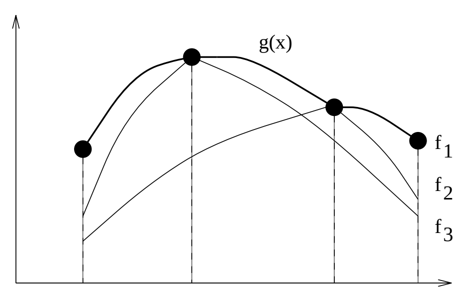

Fig. 2. A piecewise concave function $g(x)$.

Similarly, for $m > 1$, the polytope P can be divided into small regions in each of which $g(x)$ is concave. These small regions can be defined by

$$X_i = \{x \in X \mid f_i(x) = g(x)\}.$$

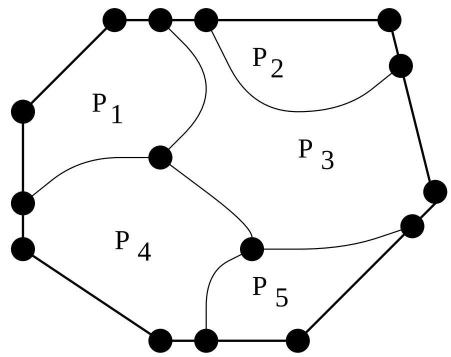

Fig. 3. g-vertices.

However, they may not be convex. Thus, where the minimum value of $g(x)$ can be achieved is not so easy to see. Du and Hwang [10] found that the minimum value of $g(x)$ can still be achieved at a *vertex* of small regions where the vertex is defined in the following way.

Let us first give a new definition for the vertex of a polytope. Consider the polytope $X = \{x \mid a_j^T x \geq b_j, j = 1, \cdots, k\}$. Denote $J(x) = \{j \mid a_j^T x = b_j\}$. A point x in X is a vertex if $J(x)$ is maximal, i.e., there does not exist $y \in X$ such that $J(x)$ is a proper subset of $J(y)$. This definition is different from the traditional one: x is a vertex if $x = \frac{1}{2}y + \frac{1}{2}z$ for y, $z \in X$ implies $x = y = z$. However, they are equivalent for polytopes.

Now, a point x in X is called a *g-vertex* if $J(x) \cup M(x)$ is maximal where

$$M(x) = \{i' \mid f_i(x) = g(x)\}.$$

Theorem 4.1 (Du and Hwang [11]) *The minimum value of $g(x)$ is achieved at a g-vertex.*

Proof. Let x^* be a minimum point for $g(x)$. Since all $f_i(x)$ are continuous, there is a neighborhood V of x^* such that for any $x \in V$, $M(x) \subseteq M(x^*)$. Let $Y = \{x \in X \mid a_j^T x = b_j \text{ for } j \in J(x^*)\}$. Then x^* is a *relative interior point* of Y, that is, for any $x \in Y$ and for sufficiently small number λ, $x^* + \lambda(x^* - x) \in Y$. Consider a g-vertex $\hat{x}$ such that $M(x^*) \cup J(x^*) \subseteq M(\hat{x}) \cup J(\hat{x})$, i.e., $M(x^*) \subseteq M(\hat{x})$ and $J(x^*) \subseteq J(\hat{x})$. The latter inclusion implies that $\hat{x} \in Y$. We will show that $\hat{x}$ is also a minimum point. Therefore, the theorem is proved.

For contradiction, suppose that $\hat{x}$ is not a minimum point. Choose a positive λ sufficiently small such that

$$x(\lambda) = x^* + \lambda(x^* - \hat{x}) \in V \cap Y.$$

Thus, $M(x(\lambda)) \subseteq M(x^*) \subseteq M(\hat{x})$. Consider an index $i \in M(x(\lambda))$. Since x^* is a minimum point of $g(x)$, we have

$$f_i(x^*) < f_i(\hat{x}), \quad \text{and} \quad f_i(x^*) \leq f_i(x(\lambda)).$$

Note that

$$x^* = \frac{\lambda}{1+\lambda}\hat{x} + \frac{1}{1+\lambda}x(\lambda).$$

By the concavity of $f_i(x)$,

$$f_i(x^*) \geq \frac{\lambda}{1+\lambda}f_i(\hat{x}) + \frac{1}{1+\lambda}f_i(x(\lambda)) > f_i(x^*),$$

a contradiction. □

Let us make some remarks on this minimax theorem.

Remark 1. A function f is *pseudo-concave* in a region if for any x and y in the region and for any $\lambda \in [0, 1]$,

$$f(\lambda x + (1-\lambda)y) \geq \min(f(x), f(y)).$$

The pseudo-concavity is clearly weaker than the concavity. In the theorem, the concavity of f_i can be replaced by the pseudo-concavity. For this replacement, the proof needs to be modified as follows: Choose a minimum point x^* with maximal $J(x)$ and a point $\hat{x}$ in Y with $M(x^*) \subseteq M(\hat{x})$. Find the existence of V as above. By the pseudo-concavity of $f_i(x)$,

$$f_i(x^*) \geq \min(f_i(\hat{x}), f_i(x(\lambda)) \geq f_i(x^*),$$

for $i \in M(x(\lambda))$, $x(\lambda) = x^* + \lambda(\hat{x} - x^*) \in Y \cap V$ and $\lambda > 0$. It follows that for $x(\lambda) \in Y \cap V$, $x(\lambda)$ is a minimum point. Note that all minimum points form a closed set. There exists the maximum value λ^* such that $x(\lambda^*)$ is a minimum point. Clearly, $x(\lambda^*)$ cannot be a relative interior point of Y. (Otherwise, by the above argument, we can obtain a larger λ such that $x(\lambda)$ is a minimum point.) Thus, $J(x^*)$ is a proper subset of $J(x(\lambda^*))$, contradicting the choice of x^*. We state the result of this remark in the following.

Theorem 4.2 *Let $g(x) = \max_{i \in I} f_i(x)$ where f_i's are continuous pseudo-concave functions and I is a finite set of indices. Then the minimum value of $g(x)$ over a polytope is achieved at a g-vertex.*

Remark 2. An interior point x of X is a g-vertex iff $M(x)$ is maximal. In general, for any g-vertex, there exists an extreme subset Y of X such that $M(x)$ is maximal over Y. A point x in X is called a *critical point* if there exists an extreme set Y such that $M(x)$ is maximal over Y. Thus, every g-vertex is a critical point. However, the inverse is false. For example, in Figure 3, the interior boundary of X_2 consists of critical points which are not g-vertices.

Remark 3. A similar result holds for the following minimax problem:

$$\min_x \max_{x \in I(x)} f_i(x)$$

where $I(x)$ is a finite index set varying as x varies. The following is a useful form. The proof is similar to the proof of Theorem 4.1 (Figure 4).

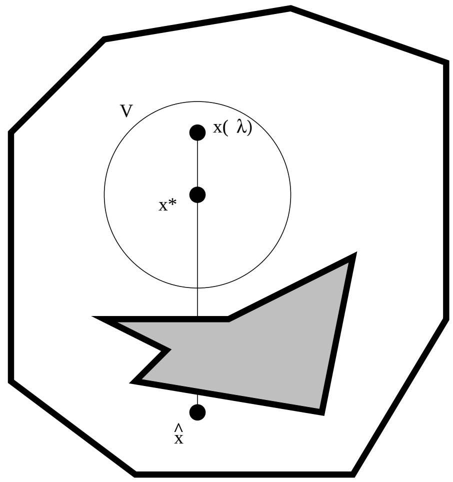

Fig. 4. $I(x)$ is defined on a subset of X

Theorem 4.3 *Let $g(x) = \max_{i \in I} f_i(x)$ where f_i's are continuous and pseudo-concave functions in the interior of a convex region X and $I(x)$ is a finite index set defined on a compact subset X' of X. Denote $M(x) = \{i \in I(x) \mid f_i(x) = g(x)\}$. Suppose that for any $x \in X$, there exists a neighborhood of x such that for any point y in the neighborhood, $M(y) \subseteq M(x)$. If the minimum value of $g(x)$ over X is achieved at an interior point of X', then this minimum value is achieved at a critical point, i.e., a point with maximal $M(x)$ over X'. Moreover, if x is an interior minimum point in X' and $M(x) \subseteq M(y)$ for some $y \in X'$, then y is a minimum point.*

Remark 4. Du and Pardalos [13] proved that the finite index set I in Theorem 1 can be replaced by a compact set. Their theorem can be stated as follows.

Theorem 4.4 *Let $f(x, y)$ be a continuous function on $X \times I$ where X is a polytope in $\mathbf{R}^m$ and I is a compact set in $\mathbf{R}^n$. Let $g(x) = \max_{y \in Y} f(x, y)$. If $f(x, y)$ is concave with respect to x, then the minimum value of $g(x)$ over X is achieved at some* critical point.

The proof of this theorem is also the same as the proof of Theorem 4.1 except that the existence of the neighborhood V needs to be derived from the compactness of I and the existence of $\hat{x}$ needs to be derived by Zorn's lemma.

5. Geometric Inequalities

Theorem 4.1 was first used in a proof of the following geometric inequality.

Theorem 5.1 *Let D, E and F be three points on three edges BC, CA and AB of a triangle ABC, respectively. Let $per(\triangle ABC)$ denote the perimeter of the triangle ABC. Then*

$$per(\triangle DEF) \geq \min(per(\triangle AEF), per(\triangle BFD), per(\triangle CDE)).$$

This inequality was proposed by Debrummer in 1956 and by Oppenhein in 1960. It was appeared in *American Mathematics Monthly* as the 4964^{th} problem in 1961. During 1961-1967, it obtained several proofs given by Dresel [?], Breusch [?], Croft [?], Zalgaller [?], and Szekers [?]. Using Theorem 2.1, Du [?] gave a new proof. This proof is not the simplest one. However, it is more general. In fact, it is suitable for similar inequalities. We introduce this proof as follows.

Proof of Theorem 5.1. Let us fix DEF and vary ABC. Consider the following function

$$f(A,B,C) = \min(per(\triangle AEF), per(\triangle BFD), per(\triangle CDE)).$$

As shown in Figure 5, A varies in the area W_A bounded by EF and extensions of DE and DF. Similarly, B and C varies in areas W_B and W_C, respectively. Define

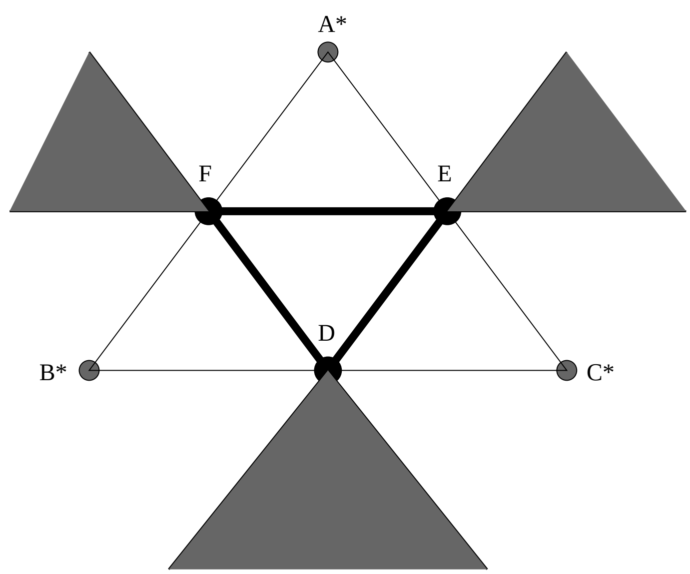

Fig. 5. W_A, W_B and W_C.

$$X = \left\{ (A,B,C) \in W_A \times W_B \times W_C \;\middle|\; \begin{array}{l} B, D, C \text{ are colinear,} \\ C, E, A \text{ are colinear,} \\ A, F, B \text{ are colinear} \end{array} \right\}.$$

We want to prove that for $(A,B,C) \in X$,

$$f(A,B,C) \leq per(\triangle DEF). \tag{1}$$

Note that three points (x_1, y_1), (x_2, y_2) and (x_3, y_3) are collinear if and only if

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = 0.$$

Thus, X is a polyhedran of dimension three, which is an unbounded region. To obtain a polytope, consider $\triangle A^*B^*C^*$ with D, E, F as its middle points of three edges (see Figure 5). Let $\bar{W}_A$ be the bounded part obtained from cutting W_A by a line ℓ_A parallel to EF. If ℓ_A is sufficiently far away from EF, then A^* is an interior point of $\bar{W}_A$. Similarly, we can define polygons $\bar{W}_B$ and $\bar{W}_C$. Suppose that the polytope $\bar{X}$ is obtained from the definition of X by using $\bar{W}_A \times \bar{W}_B \times \bar{W}_C$ to replace $W_A \times W_B \times W_C$. Clearly, to prove (1), it suffices to prove that for every $\bar{X}$,

$$\max_{(A,B,C)\in\bar{X}} f(A,B,C) = per(\triangle DEF). \tag{2}$$

Note that $per(\triangle AEF)$, $per(\triangle BFD)$ and $per(\triangle CDE)$ are convex functions with respect to (A,B,C). By Theorem 4.1, the maximum value of $f(A,B,C)$ over $\bar{X}$ is achieved at a g-vertex. If this g-vertex is an interior point of $\bar{X}$, then it must be (A^*,B^*,C^*). In this case, (2) holds. If this g-vertex (A,B,C) is on the boundary of $\bar{X}$, then at least one of A, B and C is on the boundary of $\bar{W}_A$ or $\bar{W}_B$ or $\bar{W}_C$. Without loss of generality, assume that A is on the boundary of $\bar{W}_A$. If A is on EF or the extensions of DE and DF, then one of $per(\triangle AEF)$, $per(\triangle BFD)$ and $per(\triangle CDE)$ equals twice the length of an edge of $\triangle DEF$ which is smaller than $per(\triangle DEF)$. Thus, this A must be on ℓ_A. When ℓ_A is sufficiently far from EF, BF and CE are almost parallel. In the limiting case that BF and CE are parallel, either B lies in $\triangle B^*DF$ and is different from B^* or C lies in $\triangle C^*ED$ and is different from C^* (Figure 6). Thus, either $per(\triangle BDF) < per(\triangle B^*DF) = per(\triangle DEF)$ or $per(\triangle CED) < per(\triangle DEF)$. Therefore, in this case (2) also holds when ℓ_A moves sufficiently far from $\triangle A^*B^*C^*$. $\square$

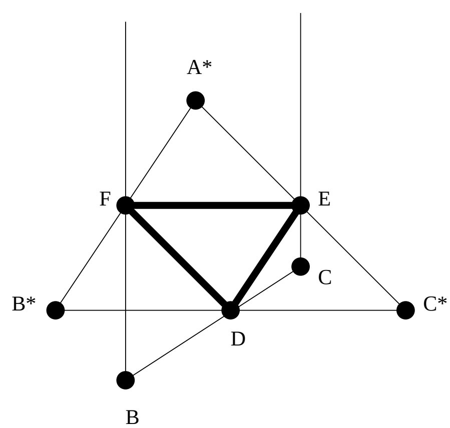

Fig. 6. BF and CE are almost parallel.

Note that the area of $\triangle AEF$ can be computed by the following formula:

$$S_{\triangle AEF} = \frac{1}{2}\begin{vmatrix} x_A & y_A & 1 \\ x_E & y_E & 1 \\ x_F & y_F & 1 \end{vmatrix}$$

which a linear function with respect to A where x_A and y_B are coordinates of A. Thus, a similar argument yields the following.

Theorem 5.2 *Let D, E and F be three points on three edges BC, CA and AB of a triangle ABC, respectively. Then*

$$\max(S_{\triangle AEF}, S_{\triangle BFD}, S_{\triangle CDE}) \geq S_{\triangle DEF} \geq \min(S_{\triangle AEF}, S_{\triangle BFD}, S_{\triangle CDE}).$$

Since $S_{\triangle AEF}$ is linear and $per(\triangle AEF)$ is convex with respect to A, the ratio $per(\triangle AEF)/S_{\triangle AEF}$ is psuedo-convex in A. Note that

$$per(\triangle AEF)/S_{\triangle AEF} = 2/r_{\triangle AEF}$$

where $r_{\triangle AEF}$ is the radius of the circle inscribed $\triangle AEF$. Therefore, the above argument also yields the following.

Theorem 5.3 *Let D, E and F be three points on three edges BC, CA and AB of a triangle ABC, respectively. Then*

$$\max(r_{\triangle AEF}, r_{\triangle BFD}, r_{\triangle CDE}) \geq r_{\triangle DEF}.$$

6. Approximation Performance

Many optimization problems are NP-hard. So, their optimal solutions are unlikely computed in polynomial time. For these problems, polynomial-time approximations are useful. One way to design a polynomial-time approximation is as follows: put some restriction on feasible solutions so that the optimal solution under this restriction can be computed in polynomial time and use this optimal solution for the restricted problem to approximate the optimal solution for the original problem.

To be explicit, consider the problem

$$\min_{k \in K} \phi_k(x)$$

which is NP-hard. Let $I \subseteq K$ such that

$$\min_{i \in I} \phi_i(x)$$

can be computed in polynomial time. Now, we use the second one to do an approximation of the first one. Usually, the performance of approximation is measured by the following ratio:

$$\rho = \min_x \frac{\min_{k \in K} \phi_k(x)}{\min_{i \in I} \phi_i(x)}.$$

Clearly, the larger is this ratio, the better is the approximation. Proving the lower bound of this ratio can be transformed to a minimax problem. In fact, suppose that we want to prove $\rho \geq \rho_0$. Then it suffices to prove that for any x,

$$\min_{k \in K} \phi_k(x) \geq \rho_0 \min_{i \in I} \phi_i(x).$$

This is equivalent to that for any x and $k \in K$,

$$\phi_k(x) - \rho_0 \min_{i \in I} \phi_i(x) \geq 0,$$

that is,

$$\max_{i \in I}(\phi_k(x) - \rho_0 \phi_i(x)) \geq 0.$$

Thus, it suffices to prove that for any $k \in K$,

$$\min_x \max_{i \in I}(\phi_k(x) - \rho_0 \phi_i(x)) \geq 0.$$

For example, let us consider the Steiner tree problem.

The Steiner tree problem is a classic intractable problem with many applications in the design of computer circuits, long-distance telephone lines, or mail routing, etc. Given a set P of points in a metric space, the problem is to find a shortest network interconnecting the points in the set. The optimal solution of this problem is called the *Steiner minimum tree* on the point set P. The Steiner minimum tree may have some vertices not in P. Such vertices are called *Steiner points* while the vertices in P are called *regular points*.

A *spanning tree* on P is a tree interconnecting all points in P under restriction that all edges are between the points in P. In the other words, no Steiner point is allowed to exist. The minimum spanning tree is the shortest spanning tree. While the Steiner minimum tree problem is intractable, the minimum spanning tree can be computed pretty fast. Thus, we can use the minimum spanning tree to approximate the Steiner minimum tree. In this case, the approximation performance ratio is called the *Steiner ratio*.

The *topology* of a tree is the adjacent relation or the adjacent matrix of the tree. Let $t(P)$ denote the minimum tree with topology t on the point set P. Let $\ell(t(P))$ denote the length of the tree $t(P)$. Suppose that all topologies of trees interconnecting P form a set K and all topologies of spanning trees on P form a set I. Then the Steiner minimum tree and the minimum spanning tree problems can be represented respectively as follows:

$$\min_{t \in K} \ell(t(P)) \quad \text{and} \quad \min_{s \in I} \ell(s(P)).$$

The lengths of the Steiner minimum tree and the minimum spanning tree on the point set P are denoted respectively by $L_S(P)$ and $L_M(P)$. From the above analysis, to prove a lower bound ρ_0 for the Steiner ratio, it suffices to prove that for any $t \in K$

$$\min_P \max_{s \in I}[\ell(t(P)) - \rho_0 \ell(s(P))] \geq 0.$$

A topology t in K is *full* if every regular point is a leaf. If a regular point is not a leaf, then this topology can be decomposed at this point into two or more subtree topologies. In this way, every topology t in K can be decomposed into edge-disjoint full topologies $t_1, \cdots, t_h$ respectively interconnecting subsets $P_1, \cdots, P_h$ of P. Note that the union of minimum spanning trees for $P_1, \cdots, P_h$ is a spanning tree for P. Thus,

$$\forall k \quad \ell(t_k(P_k)) \geq \rho_0 L_M(P_i) \Longrightarrow \ell(t(P)) \geq \frac{\sqrt{3}}{2} L_M(P).$$

It follows that to prove the lower bound ρ_0 for the Steiner ratio, it suffices to prove that for every full topology t in K,

$$\min_P \max_{s \in I} [\ell(t(P)) - \rho_0 \ell(s(P))] \geq 0. \tag{3}$$

7. Gilbert-Pollak Conjecture

In 1968, Gilbert and Pollak [19] conjectured that the Steiner ratio in the Euclidean plane is $\sqrt{3}/2$. Through many efforts [2, 3, 6, 7, 8, 21, 23], this conjecture was finally proved by Du and Hwang [11, 9, 10]. Their proof is motivated from the transformation in Section 6.

Note that the Steiner minimum tree in the Euclidean plane has the following properties.

(S1) All leaves are regular points.

(S2) Any two edges meet at an angle of at least 120^o.

(S3) Every Steiner point has degree at least three.

A tree satisfying the above three conditions and interconnecting all regular points (i.e., all points in P) is called a *Steiner tree*. Clearly, in a full Steiner tree, every angle equals 120^o. Thus, the full Steiner tree can be determined by all its edge-lengths provided the topology of the tree is fixed. Let us give the following notations.

$t(x)$: the full Steiner tree with topology t and edge-lengths $(x_1, \cdots, x_{2n-3})$ $(= x)$.
$P(t; x)$: the set of all leaves of the tree $t(x)$.
$s(t; x)$: the spanning tree with topology s for the point set $P(t; x)$.

Now, (3) can be written as

$$\min_x \max_{s \in I} (x_1 + \cdots + x_{2n-3} - \frac{\sqrt{3}}{2} \ell(s(t; x)) \geq 0 \tag{4}$$

where I is the set of spanning tree topologies for the set of n points. Note that for any $\alpha > 0$, $P(t; \alpha x)$ is similar to $P(t; x)$. Thus, $\ell(s(t; \alpha x)) = \alpha \ell(s(t; x))$. This means that among all similar point sets, we need to consider only one. So, it suffices to consider x with $x_1 + \cdots + x_{2n-3} = 1$. Define

$$f_{t,s}(x) = 1 - \frac{\sqrt{3}}{2} \ell(s(t; x))$$

and

$$X = \{x = (x_1, \cdots, x_{2n-3}) \mid x_1 \geq 0, \cdots, x_{2n-3} \geq 0, x_1 + \cdots + x_{2n-1} = 1\}.$$

To show (4), it suffices to prove that for every full Steiner tree topology t,

$$\min_{x \in X} \max_{s \in I} f_{t,s}(x) \geq 0.$$

The next lemma shows that $f_{t,s}(x)$ is a concave function in x.

Lemma 7.1 *$f_{t,s}(x)$ is a concave function in x.*

Proof. It suffices to prove that $\ell(s(t;x))$ is a convex function in x. Let A and B be two regular points. We show that the distance between A and B, $d(A,B)$ is a convex function of x. Find a path in T which connects points A and B. Suppose the path has k edges with lengths $x_{1'},...,x_{k'}$ and with directions $e_1,...,e_k$, respectively, where $e_1,...,e_k$ are unit vectors. Then $d(A,B) = \|x_{1'}e_1 + \cdots + x_{k'}e_k\|_D$. Note that a norm is a convex function and the part inside the norm is linear with respect to x. Thus, $d(A,B)$ is a convex function with respect to x. Finally, we notice that the sum of convex functions is also a convex function. □

By Theorem 4.1, the problem is reduced to the problem of finding the minimax value at critical points. Note that the transformation between the Steiner ratio problem and the minimax problem is based on a mapping between sets of n points in the Euclidean plane and points in the $(2n-3)$-dimensional space. Thus, each critical point corresponds to a set of n points with a nice geometric structure, called a critical structure. Finally, verify the conjecture on the point set with critical structure.

For a technical reason, we also need to modify Gilbert-Pollak conjecture at the beginning. This modification is necessary because the critical structure obtained above is not nice enough to be able to handle. This modification will make the critical struction much nicer. In the next section, we give the proof in details.

8. Refine the Proof of Du and Hwang

In the following, we will refine the proof for Gilbert-Pollak conjecture by using Theorem 4.3. We will show how to modify Gilbert-Pollak conjecture, how to determine the critical structure and how to verify the conjecture for the point set with critical structure.

8.1. Characteristic Area and Inner Spanning Trees

Consider a full Steiner tree $t(x)$. Two regular points are called *adjacent* if one can be reached from the other by always moving in a *clockwise* direction or always moving in a *counterclockwise* direction. Clearly, each regular point has two other adjacent regular points.

Now, consider two adjacent regular points A and B with the path $AS_1 \cdots S_k B$ connecting them. Note that there is a point S_i such that A lies inside of every angle on the path $AS_1 \cdots S_i$ and B lies inside of every angle on the path $S_i \cdots S_k B$. Thus, connecting A to $S_1,...,S_i$ and B to $S_i,...,S_k$, we obtain $\triangle AS_1S_2$, $\triangle AS_2S_3$, $\cdots$, $\triangle AS_{i-1}S_i$, $\triangle AS_iB$, $\triangle BS_iS_{i+1}$, ..., $\triangle B_{k-1}S_k$. Pasting these triangle along their edges such that every point between them has a neighborhood isometric to a neighborhood in the Euclidean plane, we obtain a simply connected region either in the plane or in a multilayer Rimann surface (Figure 7). Call this region a *cell*. Pasting all cells along all edges in $t(x)$ results in an area such that every point on $t(x)$ has a neighbothood isometric to a neighborhood in the Euclidean plane. Clearly, the area is a simply connected region in a multilayer Rimann surface. It is not unique (Figure 7(a)(b)). However, they all satisfy the following properties:

(R1) Every point has a neighborhood isometric to a neighborhood in the Eu-

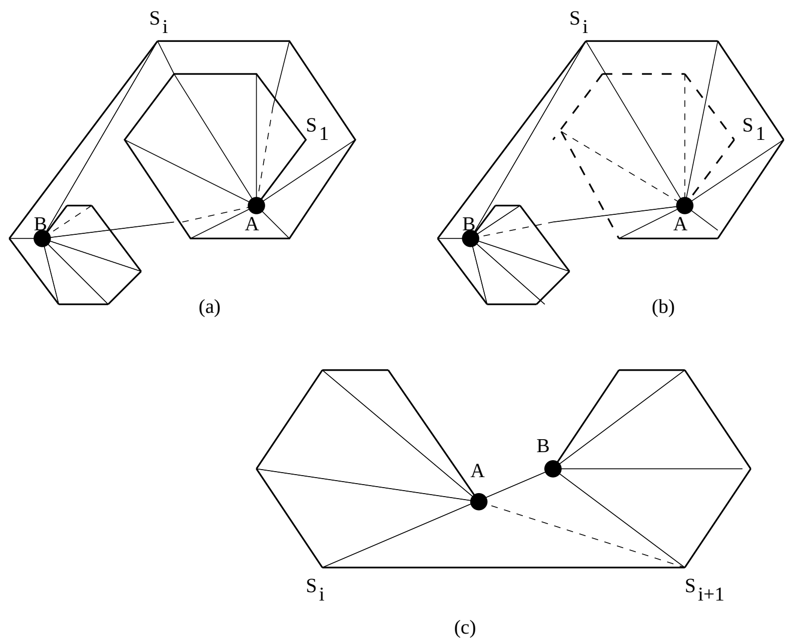

Fig. 7. Simple connected region in multilayer Rimann surface.

clidean plane.

(R2) All regular points lie on the boundary.

(R3) $t(x)$ lies inside of the area.

Those areas are called *characteristic area* of $t(x)$. We will use $C(t;x)$ to represent any one of them.

As x varies the characteristic area $C(t;x)$ varies. For some x, $t(x)$ may have self-intersection in the Euclidean plane but has no self-intersection in $C(t;x)$ (see Figure 8). Let us allow such x together with $C(t;x)$ in our consideration. Let $X(t;x)$

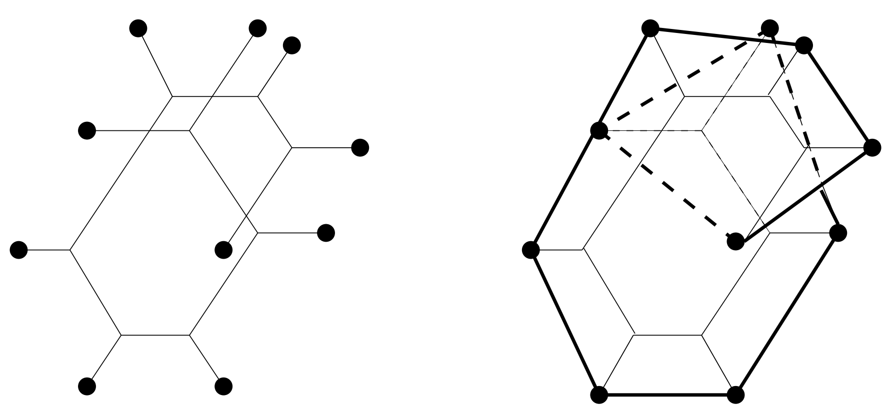

Fig. 8. A monotone path with and without self-crossing.

denote the set of all edge-length vectors y such that x together with $C(t;x)$ can be smoothly moved to y by varing the edge-lengths of all triangles which consist of the characteristic area. Clearly, for any $y \in X(t;x)$, $C(t;y)$ also has the properties (R1), (R2) and (R3). If none of the triangles in $C(t;y)$ is degenerated, then y must be an interior point of $X(t;x)$. Thus, for every boundary point y of $X(t;x)$, $C(t;y)$ must have a degenerated triangle. This means that this triangle has either an angle of 180^o or an edge of length zero. Look at back the triangles $\triangle AS_1S_2$, $\cdots$, $\triangle AS_kB$. In each of them, every angle other than the angle at A is at most 120^o. So, only the angle at A may vary to 180^o. This fact implies that for a boundary point y of $X(t;x)$, one of the following two cases has to occur:

(B1) y has at least one zero-component.

(B2) $t(y)$ has a regular point lies on the path from the regular point to an adjacent regular point. (For example, in Figure 9, (a) is not in Case (B2) because the regular point and the path which seem to overlap are in different layers. But, (b) is in Case (B2).)

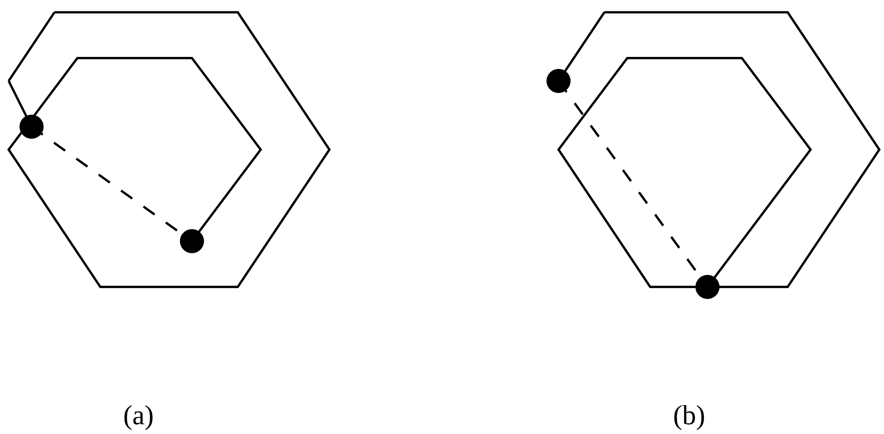

Fig. 9. (b) is in Case (B2) but (a) is not.

A spanning tree on $P(t;x)$ is called an *inner spanning tree* with respect to $t(x)$ and a characteristic area $C(t;x)$ if it lies inside of $C(t;x)$. Let $I(t;x)$ denote the set of inner spanning tree topologies. We will prove the following.

Theorem 8.1 *For every full Steiner tree topology t and any characteristic area $C(t;x)$,*

$$\min_{x\in\bar{X}_t} \max_{s\in I(t;x)} f_{t,s}(x) \geq 0.$$

Theorem 8.1 is equivalent to

$$L_S(P(t;x)) \geq \frac{\sqrt{3}}{2} L_N(P(t;x))$$

where $L_N(P(t;x))$ is the length of the minimum inner spanning tree with respect to $t(x)$. Since $L_N(P(t;x)) \geq L_M(P(t;x))$, Gilbert-Pollak conjecture is a consequence of Theorem 8.1.

Define

$$g_t(x) = \max_{s \in I(t;x)} f_{t,s}(x)$$

and

$$M(t;x) = \{i \in I(t;x) \mid f_{t,s}(x) = g_t(x)\}.$$

To use Theorem 4.3, we need to prove the following.

Lemma 8.2 *For every interior point x of $\bar{X}_t$, there is a neighborhood of x such that for any y in the neighborhood, $M(t;y) \subseteq M(t;x)$.*

Proof. First, we show that for any $m \in M(t;x)$ there exists a neighborhood U of x such that for any $y \in U$, m is in $I(t;y)$. For contradiction, suppose such a neighborhood does not exist. Then there is a sequence of points y_k converging to x such that $m \notin I(t;y_k)$. Thus every $m(t;y_k)$ has at least one edge not in the characteristic area $C(t;y_k)$. Since the number of edges is finite, there exists a subsequence of $m(t;y_k)$ each of which contains an edge not in $C(t;x)$, but these edges converge to an edge AB in $m(t;x)$. It is easy to see that AB is on the boundary of the area $C(t;x)$ and that A and B are not adjacent. (An edge between two adjacent regular points always lies in the characteristic area.) Since all vertices in an inner spanning tree lie on the boundary of $C(t;x)$, there is a regular point lying in the interior of the segment AB, contradicting the minimality of $m(t;x)$.

Now, we prove the lemma by contradiction. Suppose that there is a sequence of points y_k converging to x such that for each y_k, a spanning tree topology m_k exists such that $m_k \in M(t;y_k) \setminus M(t;x)$. Since the number of spanning tree topologies is finite, there is a subsequence of points $y_{k'}$ with the same $m_{k'}$, denoted by m. We can also assume that this subsequence lies inside of the neighborhood U of x. Thus, for every k', $\ell(m(t;y_{k'})) \leq \ell(m'(t;y_{k'}))$ for all $m' \in M(t;x)$ since $M(t;x) \subseteq I(t;y_{k'})$. Letting $k' \to \infty$, we obtain that $\ell(m(t;x)) \leq \ell(m'(t;x))$ for $m' \in M(t;x)$. Since $m \notin M(t;x)$, $m(t;x)$ must not be an inner spanning tree. It follows that there exists a neighborhood of x such that for any point y in the neighborhood, $m(t;y)$ is not an inner spanning tree for $t(y)$, contradicting the existence of the subsequence of points $y_{k'}$. □

An immediate consequence of Lemma 8.2 is that $g_t(x)$ is continuous over interior of $\bar{X}_t$. Denote $F(t) = \min_{x \in X_t} g_t(x)$. By Theorem 4.3 and Lemmas 7.1 and 8.2, $F(t)$ is achieved at some critical point. Choose a full topology t^* such that $F(t^*) = \min_t F(t)$ where t is over all full Steiner tree topologies on n regular points. We prove Theorem 8.1 by contradiction. Suppose that Theorem 8.1 is false, i.e., $F(t^*) < 0$, and that n is the smallest number of regular points such that $F(t^*) < 0$. From now on, a point x in $\bar{X}_{t^*}$ is called a *minimum point* iff $g_{t^*}(x) = F(t^*)$.

Lemma 8.3 *Every minimum point is an interior point of $\bar{X}_{t^*}$.*

Proof. Suppose to the contrary that there exists a minimum point x on the boundary of $\bar{X}_{t^*}$. First, assume that (B1) occurs, that is, $t^*(x)$ have some edges vanished. If there is a vanished edge incident to a regular point, then $t^*(x)$ can be decomposed into several edge-disjoint smaller Steiner trees. Since every smaller Steiner tree has fewer regular points, we can apply Theorem 8.1 to them. Note that a union of inner spanning trees for the smaller Steiner trees is an inner spanning tree for $t^*(x)$. We find a contradiction to $F(t^*) < 0$ by summing all inequalities. So, every vanished edge is between two Steiner points. In this case, we can find a topology t satisfying the following conditions (Figure 10).

(1) Two regular points are adjacent in t iff they are adjacent in t^*.(t is called a *companion* of t^* when t satisfies this condition.)

(2) There is a tree T interconnecting the n points in $P(t^*; x)$, with the topology t and with length less than $\ell(t^*(x))$.

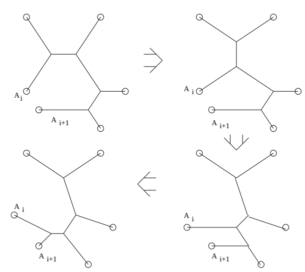

Fig. 10. A companion.

To do so, let us first note that

If the Steiner tree of topology t for $P(t; x)$ exists, then there exists a parameter vector y such that $P(t; y) = P(t^*; x)$. Let $h = 1/\ell(t(y))$. Since $\ell(t(y)) \leq \ell(T) < \ell(t^*(x)) = 1$, $h > 1$. Note that $t(hy)$ is similar to $t(y)$. Hence

$$\begin{aligned} f_{t,s}(hy) &= 1 - (\sqrt{3}/2)\ell(s(t; hy)) \\ &= 1 - (\sqrt{3}/2)h\ell(s(t; hy)) \\ &= 1 - (\sqrt{3}/2)h\ell(s(t*; x)) \\ &< g_{t^*}(x) \\ &= F(t^*) \end{aligned}$$

where s is a minimum spanning tree topology for the point sets $P(t; hy)$ and $P(t^*, x)$. Since $hy \in \bar{X}_t$, we have $F(t) \leq g_t(hy) < F(t^*)$, contradicting the minimality of $F(t^*)$.

If the Steiner tree of topology t for $P(t^*; x)$ does not exist, then we cannot use the above argument directly since $g_t(y)$ is undefined. (Remember that $F(t^*)$ is the minimum over all full Steiner topologies. So even though T is a shorter tree, there is no contradiction to the minimality of $F(t^*)$.) Now, we consider any tree of topology t. Such a tree can be determined by edge lengths and angles at every Steiner point. Write the lengths into a length vector y and the angles into an angle vector θ. Denote such a tree by $t(y, \theta)$. Two regular points is said to be adjacent in $t(y, \theta)$ if in a Steiner tree of topology t, the corresponding two regular points are adjacent. Constructing the characteristic area for $t(y, \theta)$ by connecting every pair of adjacent regular points, we can define an inner spanning tree and a minimum inner spanning tree for $t(y,\theta)$ in a similar way. Let $L_N(t;y,\theta)$ denote the length of a minimum inner spanning tree for $t(y,\theta)$. We can also show the continuity of $L_N(t;y,\theta)$. Restrict all angles to be between 0^o and 360^o and the sum of any three angles at the same Steiner point to equal 360^o. Let Y_t be the set of vectors (y,θ) with the described restrictions on θ and the restrictions $\sum y_i = 1$ and $y \geq 0$. Then Y_t is compact. So, the function h defined by $h_t(y,\theta) = 1 - (\sqrt{3}/2)L_t(y,\theta)$ reaches its minimum in Y_t. We denote this minimum value by $H(t)$. By an argument similar to that in the last paragraph, we can prove that $H(t) < F(t^*)$. Thus, $H(t) < F(t)$.

Suppose that $h_t(y,\theta) = H(t)$. If all components of θ equal 120^o, then $t(y,\theta) = t(y)$ and $y \in \bar{X}_t$. Thus, $F(t) \leq h_t(y,\theta) = H(t)$, a contradiction. Therefore, θ must have a component less than 120^o. Note that for an angle that is less than 120^o in $t(y,\theta)$, at least one edge of the angle must be vanished, for otherwise, we can shorten the tree without changing the topology. Thus, $t(y,\theta)$ contains vanished edges. If there exists a vanished edge incident to a regular point, we decompose $h(y,\theta)$ and find a full topology t' with fewer regular points such that $H(t') < 0$. If there exists a vanished edge between two Steiner points, then we can find a new companion t' of t such that $H(t') < H(t)$. Repeating the above argument, we will obtain infinitely many full topologies with at most n regular points, contradicting the finiteness of the number of topologies. Therefore, (B1) cannot occur.

Now, assume that (B2) occurs. So, $t(x)$ (in its characteristic area) has a regular point touching an edge or another regular point. In the former case, we can decompose $t(x)$ at the touching point to obtain two trees each with less than n regular points. In the latter case, we can reduce the number of regular points by one. In either case, an contradiction is achieved by an argument similar to the one used at the beginning of this proof. □

8.2. Critical Structure

In this subsection, we want to determine the geometric structure of $P(t^*, x)$ for every interior minimum point x in $\bar{X}_{t^*}$. For simplicity of notation, we use t for t^* in this subsection and the next subsection.

Let $\Gamma(t;x)$ denote the union of minimum inner spanning trees for $P(t;x)$. Let us first show some properties of $\Gamma(t;x)$.

Lemma 8.4 *Two minimum inner spanning trees can never cross, i.e., edges meet only at vertices.*

Proof. Suppose that AB and CD are two edges crossing at the point E (Figure 11) and they belong to two minimum inner spanning trees T and T', respectively. Without loss of generality, assume that EA is the shortest one among the four segments EA, EB, EC and ED. Removing the edge CD from the tree T', the remaider has two connected components containing C and D, respectively. Without loss of generality, assume that A and C are in the same component. Note that $\ell(AD) < \ell(EA) + \ell(ED) \leq \ell(CD)$. If AD lies in the characteristic area, then connecting the two components by AD results in a shorter inner spanning tree, contradicting the minimality of T'. If AD does not lie in the characteristic area, there must exist some regular points lying inside of the triangle EAD. Consider

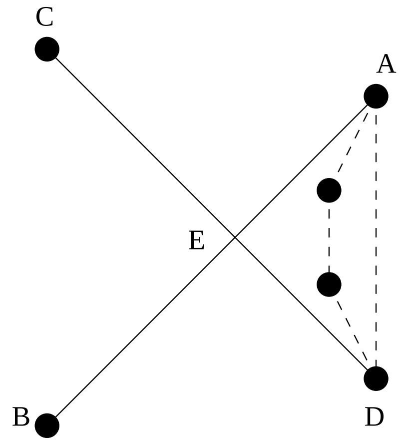

Fig. 11. AB and CD cross each other.

the convex hull of those regular points and two points A and D. The boundary of the convex hull other than the edge AD must lie in the characteristic area. This boundary contains a path from A to D. In this path there exists two adjacent vertices which belong to different connected components of $T' \setminus \{CD\}$. Connecting two such adjacent vertices also results in an inner spanning tree shorter than T', a contradiction. □

Lemma 8.5 *Every polygon of $\Gamma(t;x)$ has at least 2 equal longest edges.*

Proof. Suppose to the contrary that $\Gamma(t;x)$ has a polygon Q with the unique longest edge e. Let m be the minimum inner spanning tree containing e. For every edge e' of Q not in m, the union of m and e' contains a cycle. If this cycle contains e, then $(m \setminus \{e\}) \cup \{e'\}$ is an inner spanning tree shorter than m, a contradiction. Thus, such a cycle does not contain e. Hence, for every e' in Q not in m, m has a path connecting two endpoints of e'. These paths and e form a cycle in m, a contradiction. □

Lemma 8.6 *Let A, B and C be three regular points. Suppose that all three edges AB, BC and CA lie in $C(t;x)$. If the edge AB is in $\Gamma(t;x)$, then*

$$\ell(AB) \leq \max(\ell(AC), \ell(BC)).$$

Moreover, if AB is in $\Gamma(t;x)$ and

$$\ell(AB) \geq \max(\ell(AC), \ell(BC)),$$

then either BC or CA is in $\Gamma(t;x)$ and also has the same length as AB.

Proof. To prove the first half, for contradiction, suppose that

$$\ell(AB) > \max(\ell(AC), \ell(BC)).$$

Removal AB from the minimum spanning tree results in two connected components containing A and B, respectively. C is in one of the components. Thus, adding AC or BC would result in a spanning tree shorter than the minimum spanning tree, a contradiction.

The second half can be proved in a similar way. □

Note that the characteristic area of $t(x)$ is bounded by a polygon of n edges. Partitioning the area into $n-2$ triangles by adding $n-3$ edges, we will obtain a network with n vertices and $2n-3$ edges. This network will be called a triangulation of $C(t;x)$. Let us first ignore the full Steiner tree $t(x)$ and consider the relationship between the vertex set and the length of edges. Note that in the previous discussion, when we say that a set P of points is given, we really mean that the distance between every two points in the set is given, that is, relative positions between those points have been given. With this understanding, we make the following observations.

(1) The vertex set (i.e., the set of regular points, $P(t;x)$) can be determined by $2n-3$ edge lengths of the network.

(2) The $2n-3$ edge-lengths are independent variables, that is, the network could vary by changing any edge-length and fixing all others as long as in each triangle, the triangular inequality is preserved.

Note that every $\Gamma(t;x)$ can be embedded in a triangulation of $C(t;x)$. Thus, all edges in $\Gamma(t;x)$ are independent.

A $\Gamma(t;x)$ is said to have a *critical structure* if $\Gamma(t;x)$ partitions $C(t;x)$ into $n-2$ equilateral triangles. Such a structure has the property that any perturbation would change the set of topologies of minimum inner spanning tree. The following lemma shows that every minimum point has $\Gamma(t;x)$ with a critical structure.

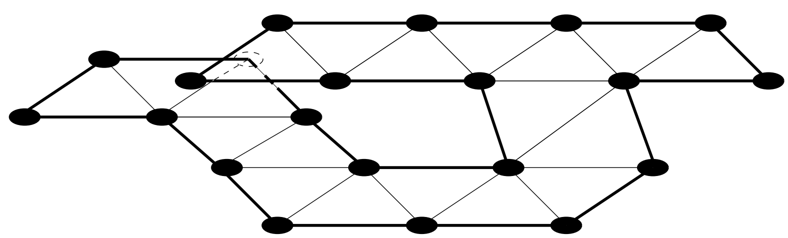

Fig. 12. A critical structure.

Lemma 8.7 *If x^* is a minimum point, then $\Gamma(t;x^*)$ divides $C(t;x^*)$ into $2n-3$ equilateral triangles (Figure 12).*

Proof. First, let us embed $\Gamma(t;x^*)$ into a triangulation of $C(t;x^*)$. If the lemma is false then one of the following must occur:

(a) There is an edge in the triangulation which does not belong to $\Gamma(t;x^*)$.

(b) No edge in the triangulation does not belong to $\Gamma(t;x^*)$. But, $\Gamma(t;x^*)$ has a nonequilateral triangle.

We will show that in each case, the number of minimum spanning trees can be increased, i.e., we can find another minimum point y such that $M(t;x^*) \subset M(t;y)$ and $M(t;x^*) \neq M(t;y)$.

First, assume that (a) occurs. Let ℓ' be the length of the logest edge which is in the triangulation but is not in $\Gamma(t;x^*)$. We shrink all longest edges and keep other edge-lengths until a new minimum spanning tree is produced. Let ℓ'' be the length of the longest edge at the last minute during the shrinking. Note that the triangular inequality is always preserved in every triangle if shrinking happens to all logest edges in the triangle or shrinking happens to the shortest edge in an isosceles. The latter is guaranteed by Lemma 8.6. Thus, during the shrinking from ℓ' to ℓ'', we do not need to worry about the condition on the triangular inequality.

Now, for each $\ell \in [\ell', \ell'']$, denote by $\bar{P}(\ell)$ the corresponding set of regular points. Then $P(\ell') = P(t;x^*)$. Consider the set L of all $\ell \in [\ell', \ell'']$ satisfying the condition that there is a minimum point y in $\bar{X}_t$ such that $\bar{P}(\ell) = P(t;y)$. Since $\ell' \in L$, L is nonempty. Moreover, L is a closed set since all minimum points form a closed set. Now, consider the minimal element ℓ^* of L. We may assume $\ell^* > \ell''$ for if $\ell^* = \ell''$, then y meets the requirement already. Suppose $\bar{P}(\ell^*) = P(t;y)$. Then for any $m \in M(t;x^*)$, $\ell(m(t;y)) = \ell(m(t;x^*))$. Since both x^* and y are minimum points, we have $g_t(x^*) = g_t(y)$, that is, the length of a minimum inner spanning tree for $P(t;x^*)$ equals that for $P(t;y)$. Hence $M(t;x^*) \subseteq M(t;y)$. However, x is a critical point. Thus, $M(t;x^*) = M(t;y)$. By Lemma 8.3, y is an interior point of $\bar{X}_t$. This means that there exists a neighborhood of ℓ^* such that for ℓ in it, the Steiner tree of topology t exists for the point set $\bar{P}(\ell)$. Thus, there exists $\ell'' < \ell < \ell^*$ such that $\bar{P}(\ell) = P(t;z)$ for some vector z (not-necessarily in $\bar{X}_t$ but $hz \in \bar{X}_t$ for some $h > 0$). Since $\ell(m(t;x))$ is continuous with respect to x, there is a neighborhood of y such that for every point y' in the neighborhood, $M(t;y') \subseteq M(t;y)$. So, ℓ can be chosen to make z satisfy $M(t;z) \subseteq M(t;y)$, too. Note that $M(t;x^*) = M(t;y)$ and for every $m \in M(t;x^*)$, $\ell(m(t;z)) = \ell(m(t;x^*))$. It follows that for every $m \in M(t;x^*)$, $m(t;z)$ is a minimum inner spanning tree for $P(t;z)$. Thus, $M(t;z) = M(t;x^*)$ and $g_t(x^*) = g_t(z)$. Suppose $hz \in X$ where h is a positive number. By the second half of Theorem 4.3, $g_t(x^*) = g_t(hz) = hg_t(z)$. So, $h = 1$, i.e., $z \in X$. Hence, z is a minimum point, contradicting the minimality of ℓ^*.

In case (b), we can give a similar proof by increasing the length of all shortest edges in $\Gamma(t;x^*)$. □

8.3. Hexagonal Trees

In this subsection, we prove $g_t(x^*) \geq 0$ where x^* is a minimum point. To do this, we begin with studying a different kind of trees. A tree in $C(t;x^*)$ is called a *hexagonal tree* if every edge of the tree is parallel to some edge in $\Gamma(t;x^*)$. The shortest hexagonal tree interconnecting the point set P is called a *minimum hexagonal tree* on P. Let $L_h(P)$ denote the length of the minimum hexagonal tree on P. The following relation was discovered by Weng [25].

Lemma 8.8 $L_s(P) \geq (\sqrt{3}/2)L_h(P)$

Proof. First, we note that if a triangle ABC has the angle at A not less than 120^o, then $\ell(BC) \geq (\sqrt{3}/2)(\ell(AB) + \ell(AC))$. Now, each edge of a Steiner minimum tree can be replaced by two edges meeting at an angle of 120^o and parallel to the given directions. Therefore, the lemma holds. □

A point on a hexagonal tree but not in P is called a *junction* if the point is incident to at least three lines. A hexagonal tree for n points is said to be *full* if all regular points are leaves. Any hexagonal tree can be decomposed into a union of edge-disjoint smaller full hexagonal trees. Such a smaller full hexagonal tree will be said to be a full *component* of the hexagonal tree.

In the hexagonal tree, an edge is referred to as a path between two vertices (regular points or junctions). Thus, an edge can contain several straight segments. An edge is called a *straight edge* if it contains only one straight segment, and is called a *nonstraight edge* otherwise. Any two segments adjacent to each other in an nonstraight edge meet at an angle of 120^o since if they meet at an angle of 60^o then we can shorten the edge easily.

In any minimum hexagonal tree T, an edge with more than two straight segments can be replaced by an edge with at most two segment. To see this, consider a nonstraight edge e in a minimum hexagonal tree T. Suppose A and B are two endpoints of e. Then all shortest hexagonal paths from A to B form a parallelogram (see Figure 13). This parallelogram must lie in $C(t; x^*)$. For otherwise, the part of

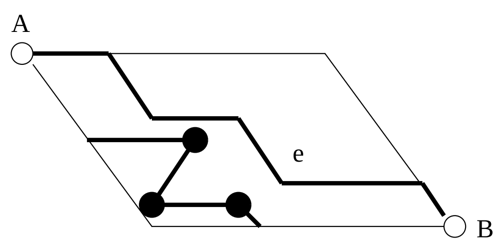

Fig. 13. The parallelogram.

this parallelogram which is inside of $C(t; x^*)$ must contain a piece of the boundary of $C(t; x^*)$. This boundary has to have at least two consecutive segments in the different directions in order to pass through the parallelogram without crossing e. The common endpoint of the two segment is a regular point lying in the parallelogram. Consider all such regular points and all shortest hexagonal paths from A to B in $C(t; x^*)$. One of the paths must pass through one of the regular points, say C (see Figure 13). Replace e by this path and delete an edge incident to C. This would result in a shorter hexagonal tree, contradicting the minimality of T. Now, since the parallelogram lies in $C(t; x^*)$, we can use a path with at most two straight segments to replace e.

From now on, we make the convention that any edge in a minimum hexagonal has at most two straight segment. In addition, when we talk about an edge of a junction, its *first segment* is the segment incident to the junction. The other segment, if it exists, is the *second segment* of the edge. Note that the junctions as shown in Figure 14 can result in a shorter tree. Thus, those kinds of junctions cannot exist in a minimum hexagonal tree.

Let T be a minimum hexagonal tree for the point set P with the maximum number of full components.

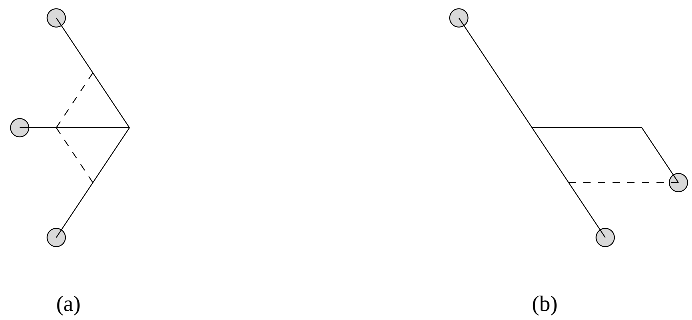

Fig. 14. Junctions like these cannot exist.

Lemma 8.9 *T can be chosen to have the properties that every junction of degree three in T has at most one nonstraight edge.*

Proof. First, consider a junction of degree three has two nonstraight edges. Then these two edges have segments in the same direction. Flip the edges if necessary to line up these two segments, then the second segments of these two edges as well as the first segment of the third edge are three segments each lying completely on one side of the line just constructed. Therefore one side has the majority of the three segments and we can move the line to decrease the number of nonstraight edges (see Figure 15).

For a junction of degree more than three, the proof is similar (see Figure 15). □

Now, we complete our proof for Theorem 8.1 by proving the following lemma.

Lemma 8.10 *Let T be a minimum hexagonal tree for the point set P with the maximum number of full hexagonal subtrees and the property in Lemma 8.9. Then T is a minimum inner spanning tree.*

Proof. Suppose that the lemma is false. Then T has a full component T' with at least one junction. Suppose that T' interconnects a subset P' of P. Clearly, T' has a junction J adjacent to two regular points A and B. (Otherwise, T' contains a cycle.)

Let us first consider the case that both edges AJ and JB are straight. If AJ and JB are in different directions then J is a regular point. Hence, they are in the same direction. Let C be the third vertex adjacent to J. First, we can assume that JC is straight for if JC is not straight, we can replace it by a straight edge without increasing the length and the number of full components (Figure 16 (a)).

Since C being a regular point implies J being a regular point, we see that C is a junction. We will show that one of the following occurs:

(a) J is a regular point.

(b) C can be moved further away from J.

Since the latter movement cannot last forever, J is a regular point which contradicts the definition of the junction.

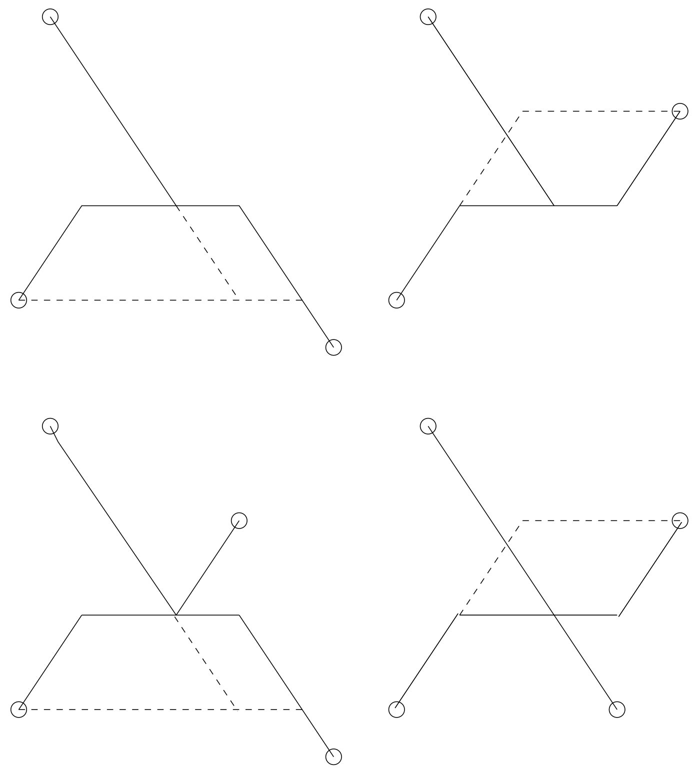

Fig. 15. Decrease the number of nonstraight edges.

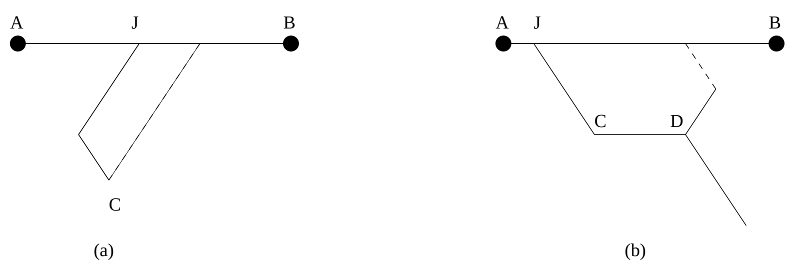

Fig. 16. (a) increases the number of full components; (b) shortens T'.

Let l be a line through C, parallel to AB. If C has a straight edge overlapping l on the right of C, then we go from C along the edges of T' to the left as far as possible. Suppose that we end at a point D. Then $\ell(CD) < \ell(JB)$ for if $\ell(CD) \geq \ell(JB)$, then JC can be moved to the right until J and B are identical so that the number of full components is increased. Since $\ell(CD) < \ell(JB)$, D cannot be a regular point. For otherwise, we can move JC to touch D which increases the number of full components. D cannot be a junction, neither. In fact, for otherwise, T' can be shorten (Figure 16 (b)). Thus, D is a corner of a nonstraight edge. A similar situation happens to the left hand side of C. Now, we can move C further away from J as shown in Figure 17. If C has no edge with segment overlapping l, then C can also be moved further away from J. This movement cannot happen forever. Finally, C becomes a regular point. It follows that J is a regular point.

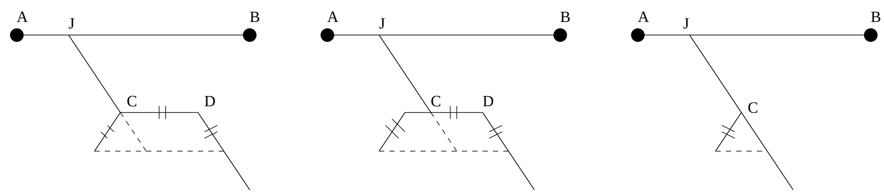

Fig. 17. C is moved further away from J.

Secondly, we consider the case that AJ is a straight edge and JB is a nonstraight edge with a segment in the same direction as AJ. Flip JB, if necessary, to line up the two first segments of AJ and JB. Let BD be the first segment of JB. Then D must be a regular point. If D is not identical to B, then we can shorten T by deleting an edge incident to D. If D is identical to B, then we go back to the first case.

Thirdly, if AJ is a straight edge and JB is a nonstraight edge without a segment in the same direction as AJ, then J can be moved either to A or to a regular point (Figure 18) which increases the number of full components.

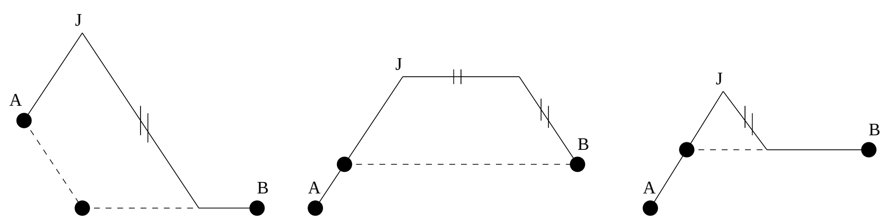

Fig. 18. J is moved to a regular point.

Since other cases are symmetric to the above three, the lemma is proved. □

By Lemmas 8.8 and 8.10, for any minimum point x^*,

$$L_S(P(t; x^*)) \geq \frac{\sqrt{3}}{2} L_N(P(t; x^*))$$

that is,

$$g_t(x^*) \geq 0.$$

It follows that $F(t) \geq 0$, contradicting the assumption that $F(t) < 0$. (Please note that this t is the t^* in Section 8.1.) Therefore, Theorem 8.1 is proved.

9. Discussion

Gao, Du and Graham [?] proved that in any normed plane, the Steiner ratio is at least 2/3. Their proof is also based on a minimax theorem in Section 4.

Graham and Hwang [19] conjectured that the Steiner ratio in n-dimensional rectilinear space is $n/(2n-1)$. The Graham-Hwang's conjecture can be easily transferred to a minimax problem meeting the condition requested by Theorem 1. For example, choose lengths of all straight segments of an interconnecting tree as parameters. When the graph structure of the tree is fixed, the set of original points can be determined by such segments-lengths, the total length of the tree is a linear function and the length of a spanning tree is also a linear function. Hence, their linear combination is a linear function which is certainly concave. However, it is hard to determine the critical structure according to this transformation. To explain the difficulty, we notice that in general the critical points could exist in both the boundary and interior of the polytope (see Theorem 1). In the proof of Gilbert-Pollak conjecture, a crucial fact is that only interior critical points need to be considered in a contradiction argument. The critical structure of interior critical points are relatively easy to be determined. However, for the current transformation for the Graham-Hwang conjecture, we have to consider some critical points on the boundary. It requires a new technique, either determine critical structure for such critical points or eliminate them from our consideration.

Pratt [21] consider the problem of determining the Steiner ratio on a sphere. He believe that the Steiner ratio on a sphere should also be $\frac{\sqrt{3}}{2}$ and proved that it is at most $\frac{\sqrt{3}}{2}$. The difficulty for solving Pratt's problem is at the job (1). In fact, if the transformation in the proof of Gilbert-Pollak conjecture can meet the condition in some minimax theorem similar to Theorem 1, then the rest proof can also be moved from there.

The k-Steiner ratio in a metric space is the largest lower bound for the ratio between the lengths of the Steiner minimum tree and the minimum k-size Steiner tree for the same given set of points where the minimum k-size Steiner tree (k-size ST) is the minimum length network interconnecting the given points and satisfying the condition that splitting the tree at each given point of degree more than one results subtrees of given points at most k. (The minimum 2-size Steiner tree is in fact the minimum spanning tree.) The significance of determining the k-Steiner ratio is stem from the study of polynomial-time heuristics for the Steiner minimum tree [1, 14, 23]. Du, Zhang, and Feng [14] conjectured that in the Euclidean plane, the 3-Steiner ratio is $(1+\sqrt{3})\sqrt{2}/(1+\sqrt{2}+\sqrt{3})$. The proof of this conjecture is blocked at the difficulty on characterizing the critical structure.

References

1. P. Berman and V. Ramaiyer, Improved approximations for the Steiner tree problem, preprint.
2. F.R.K. Chung and R.L. Graham, A new bound for Euclidean Steiner minimum trees, *Ann. N.Y. Acad. Sci.*, 440 (1985) 328-346.
3. F.R.K. Chung and F.K. Hwang, A lower bound for the Steiner tree problem, *SIAM J.Appl.Math.*, 34 (1978) 27-36.
4. H.T. Croft, K.J. Falconer, and R.K. Guy, *Unsolved Problems in Geometry*, (Springer-Verlag, New York, 1991) 107-110.

5. V.F. Dem'yanov and V.N. Malozemov, *Introduction to Minimax*, (Dover Publications, Inc., New York, 1974).
6. D.Z. Du, E.N. Yao, and F.K. Hwang, A short proof of a result of Pollak on Steiner minimal trees, *J. Combinatorial Theory, Ser. A*, 32 (1982) 396-400.
7. D.Z. Du and F.K. Hwang, A new bound for the Steiner ratio, *Trans. Amer. Math. Soc.* 278 (1983) 137-148.
8. D.Z. Du, F.K. Hwang, and E.Y. Yao, The Steiner ratio conjecture is true for five points, *J. Combinatorial Theory, Series A*, 38 (1985) 230-240.
9. D.-Z. Du and F.K. Hwang, The Steiner ratio conjecture of Gilbert-Pollak is true, *Proceedings of National Academy of Sciences*, 87 (1990) 9464-9466.
10. D.-Z. Du and F.K. Hwang, An approach for proving lower bounds: solution of Gilbert-Pollak's conjecture on Steiner ratio, *Proceedings 31th FOCS* (1990) 76-85.
11. D.-Z. Du and F.K. Hwang, A proof of Gilbert-Pollak conjecture on the Steiner ratio, *Algorithmica* 7 (1992) 121-135.
12. D.-Z. Du, B. Gao, R.L. Graham, Z.C. Liu, and P.J. Wan, Minimum Steiner trees in normed planes, manuscript.
13. D.-Z. Du and P.M. Pargalos, A minimax approach and its applications, manuscript.
14. D.Z. Du, Y. Zhang, and Q. Feng, On better heuristic for Euclidean Steiner minimum trees, *Proceedings 32nd FOCS* (1991).
15. D.-Z. Du, D.F. Hsu and K.-J Xu, Bounds on guillotine ratio, *Congressus Numerantium* 58 (1987) 313-318.
16. J. Friedel and P. Widmayer, A simple proof of the Steiner ratio conjecture for five points, *SIAM J. Appl. Math.* 49 (1989) 960-967.
17. 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.
18. E.N. Gilbert and H.O. Pollak, Steiner minimal trees, *SIAM J. Appl. Math.*, 16 (1968) 1-29.
19. R.L. Graham and F.K. Hwang, Remarks on Steiner minimal trees, *Bull. Inst. Math. Acad. Sinica*, 4 (1976) 177-182.
20. H.O. Pollak, Some remarks on the Steiner problem, *J.Combinatorial Theory, Ser.A*, 24 (1978) 278-295.
21. V. Pratt, Personal communication.
22. J.H. Rubinstein and D.A. Thomas, The Steiner ratio conjecture for six points, *J. of Combinatorial Theory, Series A*, 58 (1991) 54-77.
23. I. Stewart, Trees, telephones and tiles, *New Scientist* 16 (1991) 26-29.
24. A.Z. Zelikovsky, The 11/6-approximation algorithm for the S teiner problem on networks, manuscript.
25. S.Q. Zheng, Personal communication.