Approximations and Lower Bounds for the Length of Minimal Euclidean Steiner Trees

J.H. Rubinstein*

Department of Mathematics and Statistics

University of Melbourne

VIC 3010 Australia

rubin@ms.unimelb.edu.au

J. Weng*

ARC Special Research Center for Ultra-Broadband Information Networks (CUBIN)

Department of Electrical and Electronic Engineering

University of Melbourne

VIC 3010 Australia

weng@ee.unimelb.edu.au

N. Wormald†

Department of Combinatorics and Optimization

University of Waterloo

Waterloo ON N2L 3G1, Canada

nwormald@uwaterloo.ca

Abstract

We give a new lower bound on the length of the minimal Steiner tree with a given topology joining given terminals in Euclidean space, in terms of toroidal images. The lower bound is equal to the length when the topology is full. We use the lower bound to prove bounds on the "error" e in the length of an approximate Steiner tree, in terms of the maximum deviation d of an interior angle of the tree from 120°. Such bounds are useful for validating algorithms computing minimal Steiner trees. In addition we give a number of examples illustrating features of the relationship between e and d, and make a conjecture which, if true, would somewhat strengthen our bounds on the error.

*Research supported by the Australian Research Council

†Research supported by the Australian Research Council and the Canada Research Chairs Program. Research partly carried out while the author was in the Department of Mathematics and Statistics, University of Melbourne

1 Introduction

We consider Steiner trees in 3-dimensional Euclidean space $\mathbb{E}^3$. Given a set of points in $\mathbb{E}^3$ which we call *terminals*, a Steiner tree is an embedding of a tree containing the terminals in which each edge is a straight line segment, and each nonterminal vertex, called a *Steiner point*, has degree 3 and the three incident edges make three 120° angles. Basically everything we deal with applies also to $\mathbb{E}^d$ where $d \geq 3$, but for simplicity we restrict the discussion to the case $d = 3$. A Steiner tree is called *full* if all the terminals are leaves, i.e. vertices of degree 1. Otherwise, as commonly done in [5], any Steiner tree can be decomposed at the terminals of degree > 1 into full components.

The minimum length network joining a set of terminals is a Steiner tree [5], and we refer to the problem of finding a minimum length network as the *Steiner tree problem.* The *performance ratio* of an approximation algorithm for the Steiner tree problem is the largest ratio (over all sets of terminals) of the length of the tree found by the algorithm to the length of the minimum Steiner tree, which is the minimum length network. Zelikovsky [8] and others have designed a number of approximation algorithms for the Steiner tree problem for which performance ratios have been given. In particular, Arora [1] finds a polynomial-time approximation for the Steiner tree problem whose result can be guaranteed (regardless of the dimension) to be within a factor $1 + \epsilon$ of the optimum for any fixed $\epsilon > 0$. However, Arora's approach has not yet led to a practical algorithm for reasonably small ϵ (even $\epsilon = 0.1$ say).

It is easy to give an upper bound on the length of a minimum Steiner tree, since the length of any network, in particular any tree, connecting the terminals provides such a bound. However, finding lower bounds is not so easy. In this paper we generalise an observation of Smith [7] for four terminals, to give a general lower bound for arbitrary sets of terminals, in terms of toroidal images. We then use this to give more specific lower bounds based on angles created by the edges of an approximating tree.

After we obtained the results on toroidal images in this paper (in particular Theorems 3.1 and 3.2), they were communicated to the authors of [3], who made some use of them (though our priority might not otherwise be clear, because [3] happened to appear first).

Minimum networks connecting terminals in Euclidean space have a significant feature, which was observed from the very beginning: not only (as observed above) do all angles at Steiner points equal 120° (and every Steiner point has degree exactly 3), but, moreover, this angle condition uniquely characterises the minimum tree connecting the terminals with any given topology (provided all terminals are leaves of the tree). This property has been used to design approximations, e.g. see Chang [4] and Beasley [2]. In particular, Smith [7] gave an approximation algorithm for minimum Steiner trees in d-dimensional Euclidean space which is practical for small sets of terminals and small d. However, no guaranteed performance ratio at all has been published for such algorithms.

Specifically, in d-dimensional Euclidean space the following criterion was used in [7] for when a full tree T is a close enough approximation to a Steiner tree with the same topology: for some predetermined $\epsilon > 0$, all angles formed by the edges at the Steiner points of T are at least $120° - \epsilon$. (Note that *some* angle formed by three edges meeting at a point must be at most 120°.)

In this paper we address, in particular, the unresolved question of bounding the error in an approximate Steiner tree (with a given topology) in terms of the maximum deviation ϵ from 120° at the angles formed by edges. We conjecture that, for ϵ sufficiently small, the performance ratio is at most $1 + C\epsilon$ for some absolute constant C, uniformly over all sets of terminals and all numbers of terminals. This statement is made precise in Section 2. Failing to prove this, we are interested in bounding the performance ratio, ρ, as a function of ϵ and n. For given n, it is desirable to find the behaviour of $\rho - 1$ as $\epsilon \to 0$, since an algorithm designer may decrease ϵ in order to improve the ratio. Our best result essentially achieves the bound $\rho - 1 = O(\epsilon \log n)$ for sufficiently small ϵ (see Corollary 2.1). This problem concerns an arbitrary number of angles and so we do not attempt to replace the constant implicit in $O(\cdot)$ by an explicit constant, as the best possible result would be complicated. Our results resemble one of the main tools of Lin and Han [6], who characterise an approximate solution to another geometric problem (of finding the distance between two ellipsoids) in terms of just two angles between vectors.

A key feature of our approach is useful in other ways. We determine, for any set $\mathcal{T}$ of terminals and given topology of a Steiner tree on those terminals, a set $S_{\mathcal{T}}$ of points in $\mathbb{E}^3$ (which can be specified as a toroidal image) such that the length of the minimum Steiner tree with the given terminals and topology is equal to the maximum distance between a specified terminal and all the points in the set.

Smith [7] gave a proof of this result in the case of four terminals. One can also construct two toroidal images and obtain the length of the minimum Steiner tree as the maximum distance between points in the two image sets. One immediate corollary of our results is that one can compute a lower bound on the length of a Steiner tree (with given terminals and topology) by computing the distance between the specified terminal and any point in $S_{\mathcal{T}}$. In the case of two toroidal images, one can similarly obtain a lower bound as the distance between any two points in the two images. On the other hand, one easily gets an upper bound by computing the length of any tree containing all the terminals (with, presumably but not necessarily, angles close to 120°).

Our results on angles suggest a very convenient way of calculating a lower bound which could be used in conjunction with Smith's algorithm, simply by obtaining an upper bound on the constant implicit in $O(\cdot)$ mentioned above. Our results on toroidal images also suggest that other direct approaches may be possible, similar to that in [6].

2 Preliminary results on angles

We start with some definitions and notation. We will use T to denote a tree, which can be full or non full. The terminals of T will be denoted by N and $|N|$ will denote the number of terminals. The topology of T will be denoted by $\mathcal{T}$ and $L(T)$ will denote the length of T.

Define an *ϵ-approximate Steiner tree* for a given set of terminals to be a tree whose vertices include the terminals, possibly as well as other vertices of degree 3, whose edges are straight line segments, and such that all angles between edges meeting at a vertex are within ϵ° of 120°. The non-terminal vertices are called *pseudo-Steiner* points. A *cherry* of such a tree is a pair of terminals incident with a common pseudo-Steiner point. Note that every full tree with at least four vertices must have at least two cherries. (This can be proved by considering the tree obtained by deleting all leaves: it must have at least two leaves, which give the required cherries.)

The set of full ϵ-approximate Steiner trees with n terminals, is denoted by $\mathcal{A}_\epsilon(n)$. Let us define two functions of interest. First, given a full tree T embedded in $\mathbb{E}^3$, let $S(T)$ denote the minimum length tree with the same terminals and topology as T (permitting degeneracies, which means that some edges may have zero length). Define

$$F(\epsilon, n) = \sup_{T \in A_\epsilon(n)} \frac{L(T)}{L(S(T))} - 1.$$

Perhaps the first thing to do is check that this quantity is finite. This is easy to verify.

Lemma 2.1 *For all $n \geq 2$ and all $\epsilon < 120°$, $F(\epsilon, n) \leq K^{n-1}$ for some $K = K(\epsilon)$.*

Proof We define a function $C = C(\epsilon)$ below, and prove by induction that if n terminals are in a region of diameter D, then the length of an ϵ-approximate tree T is bounded above by a function $f(n, D)$, where $f(n, D) = 2(n-1)D(C+1)^{n-2}$, $n \geq 2$. Clearly, the length of the minimum Steiner tree S is at least half of the minimum possible value of D. (This follows since all terminals are within distance $L(S)$ of one terminal.) Combining these will then imply the lemma.

For $n = 2$, the bound $f(n, D) = D$ is immediate. For general $n > 2$, take any cherry of T. Its terminals have distance at most D apart, and it is easy to see that the pseudo-Steiner point s in the cherry must have distance at most CD from each of the terminals, where $C = C(\epsilon)$ depends only on ϵ. Hence, deleting the cherry except for s, which becomes a terminal, we obtain a new ϵ-approximate Steiner tree on $n-1$ terminals, and $L(T) \leq f(n-1, (C+1)D) + 2CD$ by induction on n. The inductive claim follows, since $f(n-1, (C+1)D) + 2CD = f(n, D) + 2CD - 2D(C+1)^{n-2} \leq f(n, D)$ for $n \geq 3$. ∎

We next give an example showing that for large ϵ, there is no upper bound on $F(\epsilon, n)$ which is independent of n.

Lemma 2.2 *For $\epsilon > 60°$, there exists a constant $c(\epsilon) > 0$ such that $F(\epsilon, n) > n^c$ for infinitely many n.*

Proof Let U_r denote the sphere of radius r centred at the origin. For $k \geq 2$ we construct a set of $n = 2^k$ terminals on U_1, and an ϵ-approximate Steiner tree T with these terminals such that $L(T) \geq (2+\delta)^{k-1}$ for some $\delta = \delta(\epsilon) > 0$.

We describe the last steps of the construction first: we first give the idea of how to locate the terminals on U_1, given the locations of the Steiner points of the cherries. Given any point s on U_2, it is easy to verify that s is a pseudo-Steiner point in a cherry with terminals on U_1, where the angle between the edges at s is precisely $60°$, and so is the angle they make to the tangent plane to U_2 at s. The lines to the terminals from s are actually tangent to U_1. For $\epsilon > 60°$, the angle between the edges at s can be made equal to $120 - \epsilon°$ by moving to $U_{2+\delta}$ for some $\delta = \delta(\epsilon) > 0$ whilst the angles to the tangent plane at s become larger. Hence, given any 2^{k-1} points on $U_{2+\delta}$, these can be made pseudo-Steiner points in cherries which can appear in an ϵ-approximate Steiner tree with 2^k terminals in U_1. The condition on angles to the tangent planes permits us to iterate this construction, we may construct two trees, each with 2^{k-1} terminals, starting with any two desired points on $U_{(2+\delta)^{k-1}}$. By choosing these two points sufficiently close together, an ϵ-approximate Steiner tree can be completed. Its length is at least twice the distance from U_1 to $U_{(2+\delta)^{k-1}}$, whilst the minimum Steiner tree has length at most $n = 2^k$ (by simply joining each terminal on U_1 to the origin). The lemma follows. ∎

For algorithms seeking approximately minimal trees, one searches potential topologies of the minimum Steiner tree and looks at ϵ-approximate trees. So in fact the main question about approximation needs to be asked about trees with the same topology as the minimum Steiner tree. Define $\overline{A}_\epsilon(n)$ to be the set of trees T in $A_\epsilon(n)$ such that the minimum Steiner tree for the terminals of T has the same topology as T, and put

$$\overline{F}(\epsilon, n) = \sup_{T \in \overline{A}_\epsilon(n)} \frac{L(T)}{L(S(T))} - 1.$$

We immediately have

$$\overline{F}(\epsilon, n) \leq F(\epsilon, n). \tag{1}$$

Despite the more practical relevance of $\overline{F}(\epsilon, n)$, we have no results about it other than via $F(\epsilon, n)$.

Our emphasis (with a basis in practical considerations) will be to determine the behaviour of $F(\epsilon, n)$ or $\overline{F}(\epsilon, n)$ for very small ϵ. This is because an algorithm of successive approximations can cause the angle error to be as small as we like for any given n. When we make assumptions like $\epsilon < n^{-3}$ for convenience, we believe the results still give a good idea of the behaviour for large n even when the bound on ϵ is relaxed considerably.

We next give a lower bound on $F(\epsilon, n)$ valid for all ϵ, including small ϵ, unlike Lemma 2.2. Note that $\Omega(x)$ denotes a function bounded below by some positive constant times x.

Lemma 2.3 *$F(\epsilon, n) \geq \Omega(\epsilon)$ for all n.*

Proof We give an example with a full topology on n terminals, though of course for non-full trees, the case $n = 3$ suffices to prove the result. Take T_0 to be a centrally symmetrical Steiner tree in the x, y-plane with all edge lengths 1, and with all terminals having paths of length k to the central Steiner point. (Thus many terminals will actually coincide, but a tiny perturbation permits us to ignore this.) Take T to be a tree with the same terminals as T_0, but with all pseudo-Steiner points lying above the plane so that they project onto the Steiner points of T_0, and with maximum angle errors (all angles $120° - \epsilon$). This can be done very symmetrically, so that the gradients (in the z direction) of the edges from the central pseudo-Steiner point s are all equal (and roughly $c\sqrt{\epsilon}$), and all pseudo-Steiner points of given distance from s have the same z coordinate. (Note that if the upper edge enters a true Steiner point with downward gradient α, where α is small, and the other two edges are placed symmetrically, then their downward gradients are approximately $\alpha/2$. So the symmetrical placement can only be done so that all edges have gradients $O(\sqrt{\epsilon})$.) Hence $L(T) - L(T_0) = O(\epsilon L(T_0))$. ∎

Returning to the consideration of upper bounds, we next give an example to demonstrate that, in a certain sense, the upper bound in the proof in Lemma 2.1 does not give an entirely misleading picture of ϵ-approximate Steiner trees. We show that for arbitrarily small ϵ, there exist ϵ-approximate Steiner trees in which the distance from some of the pseudo-Steiner points to the convex hull of the terminals is an arbitrarily large multiple of the diameter of that convex hull. This shows how little physical similarity there may be between the minimum Steiner tree T_0 and an ϵ-approximate Steiner tree T with the same set of terminals. This does not yield a lower bound, but has an implication that certain approaches for upper bounds will not work, or will be very difficult to carry out. For, if ϵ is much bigger than $\frac{1}{\ln n}$, it is clear from the following example that there is no strong relation between the length, position or angle of an edge of T and the corresponding edge of T_0.

Example For all $\epsilon > 0$ sufficiently small there is some $C > 0$ such that we can construct for all n an ϵ-approximate tree with n terminals, all located within a unit ball centred at the origin, and such that one of the pseudo-Steiner points has distance at least $n^{C\epsilon}$ from the origin. This is in the spirit of the proof of Lemma 2.2. Begin with a point s of distance R from the origin (R to be determined), and construct three points s_i ($1 \leq i \leq 3$) such that the angles which the edges $s_i s$ make at s are within ϵ of $120°$ and $|P_0 s_i| = R(1 - \epsilon/2)$ for each i, where P_0 denotes the origin. (This can be done by taking the three edges pointing inwards from the surface of the sphere through s_0 centred at P_0, and choosing the s_i to be the closest points to P_0 on these lines. In fact $\epsilon/2$ can be increased to some constant times $\sqrt{\epsilon}$, but this is not needed.) Then branch again at each s_i, again choosing edges entering the sphere through s_i centred at P_0, with the entering angles at least $\epsilon/2$. After i branchings, the distance from the branch points to P_0 is $R(1 - \epsilon/2)^i$. When this quantity is less than 1, stop and make these points terminals. The number of terminals is $n = 3 \times 2^{i-1}$. We may choose $R(1 - \epsilon/2)^i = 1$, so that $\ln R = -i \ln(1 - \epsilon/2) > C' i\epsilon > C\epsilon \ln n$, as required.

We conjecture that the error in an ϵ-approximate Steiner tree cannot be substantially larger than that given in Lemma 2.3, when ϵ is sufficiently small.

Conjecture 2.1 *There exists $\epsilon_0 > 0$ such that uniformly for all $0 < \epsilon < \epsilon_0$ and all n,*

(a) $F(\epsilon, n) = O(\epsilon)$,

(b) $\overline{F}(\epsilon, n) = O(\epsilon)$.

(c) For any $d > 3$, the analogues of (a) and (b) hold for trees in $\mathbb{E}^d$.

Note that (a) implies (b). The separate statement (b) is included in case (a) is false, since (b) is important for algorithmic purposes. To express it in words, part (a) says that there exists a constant c such that for $\epsilon > 0$ sufficiently small the following is true. Let T_0 be a Steiner tree for a set of n terminals (not necessarily the minimum tree for those terminals). Let T be an ϵ-approximate tree with the same topology as T_0 and the same terminals. Then $L(T) - L(T_0)$ is at most $c\epsilon L(T_0)$.

As a first step we will prove the following. Although it gives a weaker bound than our next result, it requires less conditions on ϵ and has a much simpler proof.

Theorem 2.1 *Let $\epsilon_0 < 120$. Uniformly for all $\epsilon < \epsilon_0$, $F(\epsilon, n) \leq O(n^2\sqrt{\epsilon})$.*

Then in Section 4 we will prove the following result, which is much sharper for small ϵ.

Theorem 2.2 *Uniformly for all $\epsilon < n^{-2}$, $F(\epsilon, n) \leq O(\epsilon(\log n + n^3\epsilon))$.*

Corollary 2.1 *Uniformly for all $\epsilon < n^{-3}\log n$, $F(\epsilon, n) \leq O(\epsilon \log n)$.*

So we have come within a factor of $\log n$ of Conjecture 2.1, provided ϵ is sufficiently small compared with n.

Note 1. It is easy to get planar examples with "error" of the order of $n\epsilon^2 L$ for large n, though the conjecture would imply that this would not occur if ϵ is much bigger than $1/n$.

Note 2. Our proofs immediately give the same results for d dimensions, $d \geq 4$. For $d = 2$ the same method gives a more precise result, but for this there are presumably other methods available as well.

Proof of Theorem 2.1 Start with an ϵ-approximate tree T with $\epsilon < \epsilon_0$. At each pseudo-Steiner vertex in turn, make the angles precisely equal to 120°, by rotating the three branches of the tree at the vertex, keeping the whole of each branch rigid so that the angles at all other vertices are fixed. To effect this change, the edges at the vertex only have to move through angles $O(\sqrt{\epsilon})$. Each rotation moves all terminals a distance at most $O(\sqrt{\epsilon})L(T)$. After proceeding through all pseudo-Steiner vertices, we have a true Steiner tree T', of the same length as the original tree T, but in which each terminal has moved distance at most $O(n\sqrt{\epsilon})L(T)$. If we now move these terminals back to their original positions, with the natural induced motion of the Steiner tree, preserving its 120° angles, the total change in length of the tree is at most $O(n^2\sqrt{\epsilon})L(T)$. (If the topology of the Steiner tree remains nondegenerate, this comes from a standard first derivative argument; see Hwang et al. [5]. If it becomes degenerate, we can permit "negative length" edges in a standard way to preserve the argument.) So the Steiner tree for these terminals, with the same topology as T, is $O(n^2\sqrt{\epsilon})L(T)$ shorter than T' and hence T. ∎

3 Toroidal images

Given n terminals in $\mathbb{R}^3$ and a choice of full topology, we can generate a toroidal image, by generalising the planar procedure of Melzak (see Hwang et al. [5]). In $\mathbb{R}^3$, the procedure consists of successive iterations of a step in which a cherry of the topology is chosen, and the two leaf terminals are "merged" into a new terminal on the equilateral circle, i.e. the set of positions of the third point of an equilateral triangle with two vertices at the cherry points. In the given full topology, the Steiner point incident with the two cherry vertices disappears and its neighbour is joined in the topology to the new terminal on the equilateral circle. We will call this new terminal an equilateral point; it corresponds to the Melzak or Simpson point for Steiner trees in the plane.

The new cherry can be chosen in any manner at each step, but we may fix any sequence of choices by referring to the original topology. For simplicity of discussion, at first we restrict to the case that there is a reserved terminal t_f, called the *final* terminal, that is never chosen in a cherry (and hence remains till there is only one other terminal). The procedure is called "unfolding" the original topology with respect to the sequence of cherries (where each cherry, after the first, exists in an intermediate topology). At the end, the topology has been unfolded to an interval, which joins the final terminal with the last equilateral point chosen.

Points on the equilateral circle can be parametrised by circular coordinates. Unfolding a topology with respect to a given sequence of cherries, using all possible points on the equilateral circle in each step of the iterative process produces a map ϕ from T^{n-2} into $\mathbb{R}^3$. (T^{n-2} is the Cartesian product of $n-2$ circles, i.e an $(n-2)$-dimensional torus). Namely, a given choice of $n-2$ circular coordinates corresponds to choices of points on each of the equilateral circles, and then ϕ gives the position of the point on the final equilateral circle. We will refer to $\phi(T^{n-2})$ as the toroidal image (with respect to a given sequence of cherries and final terminal).

Recall that the topology $\mathcal{T}'$ of a tree T' is called a degenerate of a full topology $\mathcal{T}$ if it can be obtained by collapsing some edges to zero length, so that the vertices at the ends of the collapsed edges merge. By convention, $\mathcal{T}$ is a degenerate of $\mathcal{T}$, i.e no edges might be collapsed.

Theorem 3.1 *For any full topology $\mathcal{T}$ and given set of terminals, the greatest distance between the final terminal and the toroidal image (with respect to any final terminal and sequence of cherries) is a lower bound on the length of every (Steiner) tree whose topology is a degenerate of $\mathcal{T}$.*

Proof Let $f(\mathcal{T}, \mathbf{s})$ denote the greatest distance between the final terminal and the toroidal image generated with respect to a sequence of cherries $\mathbf{s}$. We have to show $L(T_0) \geq f(\mathcal{T}, \mathbf{s})$, and we do this for all trees T_0 which are degenerates of $\mathcal{T}$ (regardless of their angles).

This is by induction on the number of terminals. It is clearly true for two terminals. So take an approximate Steiner tree T_0 with at least three terminals, and the specified final terminal t_f and sequences of cherries for the topology $\mathcal{T}$. Then there exists a cherry in the sequence; take the first, with terminals say t_0 and t_1 at a pseudo-Steiner point s in the tree T_0. Let s' be the third neighbour of s in T_0, which can be assumed to also be a pseudo-Steiner point. Let m be the (or any) point on the equilateral circle defined by t_0 and t_1 and involved in the maximum of the distance between t_f and the toroidal image. Obtain T_1 from T_0 by replacing the cherry by m; i.e. delete the edges incident with s and join m directly to the third neighbour of s in T_0 by a straight line. Then $L(T_1) \leq L(T_0)$. (In fact if T_0 is a full Steiner tree, then the lengths are equal if m is chosen as the furthest point on the equilateral circle from s', and it is easy to see from this that the inequality holds in all other cases. Note also that the inequality $L(T_1) \leq L(T_0)$ relies on the fact that for any point s in space, the distance to a vertex of an equilateral triangle is at most the sum of the distances to the other two vertices of this triangle.)

If $\mathbf{s}$ denotes the specified sequences of cherries, and $\mathbf{s}'$ is the same with the cherry above omitted, we now have

$$L(T_0) \geq L(T_1) \geq f(\mathcal{T}_1, \mathbf{s}') = f(\mathcal{T}, \mathbf{s})$$

where the middle step is by induction and the last is by choice of m. (Note that $\mathcal{T}_1$ denotes the topology of the tree T_1.) The theorem follows. ∎

This theorem is actually all we need for purposes of the following section. However, there are a number of other nice observations we would like to make.

If there exists a full Steiner tree T_0 with the given terminals, then the unfolding procedure can be performed so that at each step, the Melzak point is chosen on the extension of the third edge e of T_0 that meets the Steiner point of the chosen cherry. (The first two edges being the ones from the terminals of the cherry.) The fact that the extension of the edge passes through the circle follows from the well-known properties of Steiner trees in $\mathbb{R}^2$ [5], noting that we may focus on just three points here and the 3-point problem has a planar minimum tree. Furthermore, if the point chosen on the circle is the farthest one from the Steiner point of the cherry, then the length of the Steiner tree on the new set of terminals is the same as the old one (again, appealing to the known planar result). By induction, the original Steiner tree has the same length as the final interval joining the final terminal and the toroidal image. Combining this observation with Theorem 3.1 immediately gives the following.

Corollary 3.1 *If a full Steiner tree exists for a set of terminals, then the length of the Steiner tree is equal to the maximum distance between the final terminal and the toroidal image in an unfolding with respect to any sequence of cherries.*

When the procedure is performed in the above way, using a full Steiner tree T_0, with respect to any sequence of cherries, we call it a *proper unfolding*. Note that this is unique because the Steiner tree is unique.

We may generalise these results as follows. We may use any edge to split the given topology into two components and, loosely speaking, unfold each separately. To be precise, consider a full Steiner topology $\mathcal{T}$ for a set of n terminals with a distinguished edge e in $\mathcal{T}$. Obtain two topologies $\mathcal{T}_1$ and $\mathcal{T}_2$ by deleting e. (The ordering of these can be arbitrary.) To $\mathcal{T}_1$ we can reinstate e and add its incident vertex from $\mathcal{T}_2$, which we may call t_1^*, an artificial new terminal (which can be placed anywhere). Similarly, add e and another artificial terminal t_2^* to $\mathcal{T}_2$. The resulting topologies $\mathcal{T}_1^*$ and $\mathcal{T}_2^*$ can be unfolded with respect to any sequences of cherries, with t_1^* and t_2^* the final terminals respectively. The possible positions of the last Melzak points chosen (in the two cases) constitute two toroidal images of $\phi_{e,1}(T^p)$ and $\phi_{e,2}(T^q)$ mapped into $\mathbb{R}^3$, where $p+q=n-2$. Finally, $\mathcal{T}$ is unfolded with respect to e (and the choice of cherries) by choosing one point from each of these toroidal images.

In the case that e is incident with a terminal t, this corresponds to unfolding $\mathcal{T}$ with t as the final terminal.

Theorem 3.2 *For any full topology $\mathcal{T}$ and given set of terminals, the greatest distance between the two toroidal images (with respect to any edge of $\mathcal{T}$ and sequences of cherries) is a lower bound on the length of every (Steiner) tree whose topology is a degenerate of $\mathcal{T}$, and is equal to the length of the full Steiner tree with topology $\mathcal{T}$ (if it exists).*

Proof This uses the proof of the previous theorem and corollary. The induction works using the first cherry in either of the sequences for the two subtopologies. The rest of the proof is virtually the same. ∎

We return now to the simpler unfoldings first described, in which the edge e is incident with a final terminal. It is interesting to consider the distance function d from the final terminal to the Melzak point chosen on the last equilateral circle. Note that d can be viewed as a map defined on T^{n-2}, by composing the usual distance function with ϕ. The following result is elementary.

Lemma 3.1 *The map ϕ is differentiable; so also is d, assuming that the final terminal is not on the toroidal image. If the final terminal does lie on the toroidal image, then d^2 is differentiable.*

We emphasise that for a given topology, there may be many possible sequences of choices of cherries during the unfolding procedure, each producing a different toroidal image. Lemma 3.1 refers to any one predeterimined sequence that has a final terminal.

Theorem 3.3 *In a proper unfolding, the last equilateral point is the unique furthest point on the toroidal image $\phi(T^{n-2})$ (with respect to the same sequence of cherries) from the final terminal. In particular, there is* **only one** *local maximum for distance on the toroidal image.*

Proof Observe that for a critical point of d, the equilateral point m chosen on the last equilateral circle C must be either closest or furthest away from the final terminal t_n. Moreover for a local maximum, the latter must be the point selected. (If the equilateral circle was equidistant from t_n, it is easy to see that no full tree can be constructed. So we do not need to worry about this degenerate case.)

Assume that C is obtained by rotating an equilateral triangle with two fixed vertices m' and t_{n-1}, where m' lies on the previous equilateral circle C', t_{n-1} was the other terminal of that cherry, and the third moving vertex m sweeps out the circle C. Also let s be the Steiner point between m' and t_{n-1}, for the three point tree m', t_{n-1}, t_n. Firstly, it is an elementary exercise to check that the plane through m', t_{n-1}, t_n meets C in the two points which are critical for d if all earlier equilateral points are held fixed, with the farthest point m from t_n being the correct choice for the equilateral point, required to merge the vertices m', t_{n-1}. Next, we claim that for a critical point of d, the line from s to m' must be orthogonal to C' at m'. We will show there are precisely two such choices for m' on C', giving a local maximum and minimum for d (again with the choices of earlier equilateral points held fixed).

We need to compute first variation of the length of a Steiner tree, when precisely one terminal moves. If θ is the angle of the direction of movement to the edge at that terminal, then the first variation is $-cos\theta$. So we observe that if the line sm' is not orthogonal to C' then the (directional) derivative of d is not zero as m' moves around C', so this is not a critical point. Consequently, the length of the 3-terminal Steiner tree m', t_{n-1}, t_n, s is the same as the distance $d(t_n, m)$, since the line $t_n m$ is orthogonal to the equilateral circle C and m is the furthest point away from t_n.

We need also to show that there are just two points m' on C' with the property that the line sm' is perpendicular to C'. But by replacing the terminals t_{n-1} and t_n by the appropriate equilateral point $\bar{m}$ using Melzak's procedure, we can find the distance d by measuring the distance function from $\bar{m}$ to C', which has two critical points. Notice again for a local maximum, we must choose the furthest point m' on C' away from $\bar{m}$. Moreover, from the comments before this proof, this is the correct choice to make $m'', t_{n-2}, t_{n-1}, t_n, s, s'$ a 4 point Steiner tree, where s' is the Steiner vertex at the cherry m'', t_{n-2}, C' is the equilateral circle obtained by rotating an equilateral triangle around the fixed vertices m'', t_{n-2} and m'' is the equilateral point chosen on the previous equilateral circle C''.

This argument iterates by looking back into the procedure to consider the previous equilateral circle, and so on. At each stage, we find that the Steiner tree constructed in the previous step is orthogonal to the relevant equilateral circle, and moreover, choosing the furthest point away from the point corresponding to $\bar{m}$ gives a local maximum of the length function and a Steiner tree with length $d(t_n, m)$. Iterating back to the first equilateral circle proves the theorem. ∎

Further consideration of the proof of the theorem shows that on each circle, there are only two choices of point which give a critical point for d and hence 2^{n-2} choices in total, assuming that no circle is equidistant from the point corresponding to $\bar{m}$ in any given step. Now it is well-known that for a Morse function on an $n-2$ dimensional torus, this is the smallest number of critical points. This observation may be quite useful in searching for the unique local (and global) maximum in the toroidal image.

We may generalise Theorem 3.3 to apply to the more general unfoldings.

Theorem 3.4 *If a full Steiner tree on n terminals is properly unfolded with respect to any edge into an interval joining points x, y on the two toroidal images, then x and y are furthest apart on the two images. Moreover the distance function between the two images has only one local maximum.*

Proof We follow the same strategy — in fact if one 'end' of the Steiner tree is unfolded into a point x on the image of T^p, we can keep this point fixed and apply the previous argument, as in Theorem 3.3. This shows that y is the furthest point on the image of T^q from x and the distance function from x to the image of T^q has a unique local maximum. One can then vary x, keeping y fixed. Of course, critical points for both x, y varying are the same as for varying each separately. (This is by the usual consideration of partial derivatives). A local maximum for x, y varying together must be a local maximum for each point varying separately so there is only one such choice of x, y. ∎

4 Proof of Theorem 2.2

We make use of the following.

Lemma 4.1 *Consider points O and P in a plane Π, and define $Q(P)$ by the clockwise-oriented equilateral triangle OPQ. Let P move with velocity vector $\mathbf{u}$, where $\mathbf{u} \in \Pi$, and let $\mathbf{v}$ be the corresponding velocity vector of $Q(P)$. Then $\mathbf{v}$ is of the same magnitude as $\mathbf{u}$ and has direction $60°$ clockwise of that of $\mathbf{u}$.*

Proof This follows easily from the fact that the map $P \mapsto Q(P)$ is linear; it is clearly the $60°$ clockwise rotation around O. ∎

To prove Theorem 2.2, note that the definition of F involves full trees T. We can begin with any full ϵ-approximate pseudo-Steiner tree T_0 and select any terminal $\hat{t}$ of T_0, which we will call the *root* terminal.

We first give a brief sketch of this proof. By Theorem 3.1 $L(S(T_0))$ is at least the maximum distance of $\hat{t}$ from a point on the toroidal image generated by unfolding T, so that $\hat{t}$ is the final terminal, i.e T is unfolded to an edge ending at $\hat{t}$. Recall that any point on the toroidal image is generated by an iterated choice of equilateral points s_i^* on equilateral circles. These points can be viewed as lying on smaller toroidal images, starting at the first chosen cherry and working towards $\hat{t}$. We will use the positions of the pseudo-Steiner points of T_0 to iteratively choose such points s_i^*. In the process, we will also choose a point s_i' which lies close to s_i^* and *also* lies on the extension of an edge of T_0. At the end of this process, we will be able to compare the length of T_0 to the distance between $\hat{t}$ and a point on the toroidal image, which from above is a lower bound on $L(S(T_0))$.

This is an inductive argument. The first step, in which we consider the approximate error in a cherry, is a convenient simplified version of the general step, so we give it in detail. Then later, we compute the accumulation of errors caused by induction. Choose a cherry of the pseudo-Steiner tree, for which two non-root terminals t_1 and t_2 of T_0 are adjacent to a common pseudo-Steiner point s_1. (It is elementary to prove that such a cherry can be found, so long as there is at least one pseudo-Steiner point.) Let s_2 denote the other (third) vertex of T_0 adjacent to s_1. Put $L_i = |s_1 t_i|$ for $i = 1$ and 2, and without loss of generality assume $L_1 \geq L_2$. A key feature of the argument is the consideration of the plane Π determined by s_1, s_2 and t_1. In a Steiner tree, Π would also contain t_2. However, for an ϵ-approximate Steiner tree, such local planarity is lost. Define s_1' on the extension of $s_2 s_1$ so that

$$|s_1 s_1'| = L_1 + L_2.$$

See Figure 1.

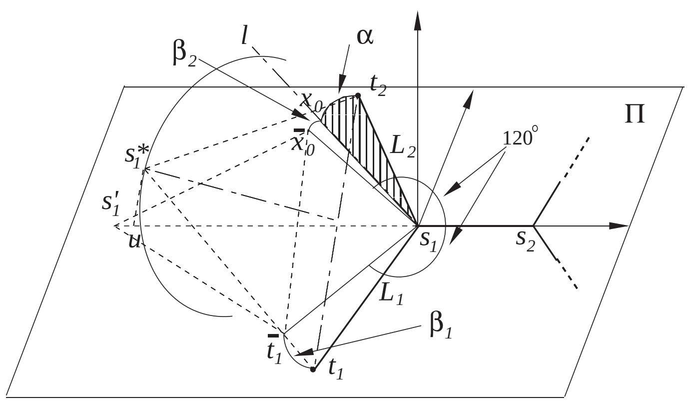

Figure 1: A cherry in T_0

We need to find an appropriate equilateral point s_1^* on the equilateral circle C for t_1 and t_2. The circle C is defined to be the set of all points P such that Pt_1t_2 forms an equilateral triangle. Let $u = u(s_1^*)$ denote the projection of s_1^* onto the line $s_1 s_1'$. Since the angles in T_0 at s_1 are within $\epsilon°$ of 120°, it is possible (shown next) to choose s_1^* so that

(i) $|us_1^*| = O(\epsilon L_1)$;

(ii) $|us_1'| = O(\epsilon^2 L_1 + \epsilon L_2)$.

(Recall $L_1 \geq L_2$.) We note that the particular point s_1^*, as defined below, lies in the plane Π. This is not required for the argument, but may help with visualisation. Here and in the rest of the proof, the constants implicit in the $O()$ are uniform for all ϵ sufficiently small.

We now show how to choose s_1^* to satisfy (i) and (ii). We define s_1^* to be the intersection of Π and C that is near s_1'. For later use, we define a function r as follows. For any point x near t_2 (where 'near' can be taken to mean at distance less than $|t_1t_2|/10$), define the point $r(x)$ to be that point in the plane Π which forms an equilateral triangle with t_1 and x, and is near to s_1'. (Note that there will be two such points — we are assuming ϵ is sufficiently small here, so the distance from t_2 to Π is less than $|t_1t_2|/10$ say.) Note that $s_1^* = r(t_2)$. If T_0 were a true Steiner tree, then $s_1' = s_1^*$.

Let l be the projection of the line s_1t_2 onto Π, and let x_0 be the point on l such that $|s_1x_0| = |s_1t_2| = L_2$. The angle $\alpha = \angle t_2s_1x_0$ will be called the *out-of-plane angle error* at s_1. (Note that this depends on the choice of which terminal is t_1 as opposed to t_2, but this does not matter.) It is clear from basic considerations of the angles at s_1, using the fact that T_0 is ϵ-approximate, that $\alpha = O(\sqrt{\epsilon})$. We next claim that

$$|r(x_0)r(t_2)| = O(\alpha^2 L_2). \tag{2}$$

This is easy to verify from the fact that the distance of t_2 from Π is $O(\alpha L_2)$ and that the distance from x_0 to the projection of t_2 onto Π is $O(\alpha^2 L_2)$.

If we rotate the edges s_1t_1 and s_1x_0 to $s_1\bar{t}_1$ and $s_1\bar{x}_0$, staying in the plane Π, so that the angles $s_2s_1\bar{t}_1$ and $s_2s_1\bar{x}_0$ are both exactly $120°$, then $\bar{x}_0\bar{t}_1s_1'$ becomes an equilateral triangle (by the definition of s_1' and the Melzak construction). Hence by Lemma 4.1 and noting $L_1 \geq L_2$, we get

$$|r(x_0)s_1'| \leq \beta(L_1 + L_2) = O(\beta L_1) \tag{3}$$

where $\beta = \angle t_1s_1\bar{t}_1 + \angle x_0s_1\bar{x}_0$, which we call the *in-plane angle error* at s_1.

Clearly $\beta = O(\epsilon)$, which we will use shortly. Imagine moving x_0 directly to $\bar{x}_0$. Then from Lemma 3, its direction of movement is at approximately $60°$ to the line s_1s_1'. Hence, $r(x_0)$ moves at a direction within β of perpendicular to s_1s_1'. Applying the same observation to the movement of t_1 directly towards t_1', we obtain

$$|s_1r(x_0)| - |s_1s_1'| = O(\beta^2(L_1 + L_2)) = O(\beta^2 L_1). \tag{4}$$

In view of (3) (or arguing in a similar fashion), this is also a bound on the length of the projection of the interval $r(x_0)s_1'$ onto the line s_1s_1'. Hence, from (2) and noting $r(t_2) = s_1^*$,

$$|u(s_1^*)s_1'| = O(\alpha^2 L_2 + \beta^2 L_1), \tag{5}$$

whilst (2) and (3) combine directly to give

$$|s_1's_1^*| = O(\alpha^2 L_2 + \beta L_1). \tag{6}$$

The bounds (i) and (ii) follow from equations (6) and (5), since $\alpha = O(\sqrt{\epsilon})$ and $\beta = O(\epsilon)$.

Thus, s_1^* satisfies (i) and (ii). Now form a new tree T_1 by deleting the edges s_2s_1, s_1t_1 and s_1t_2 from T_0 and adding the edge $s_2s_1^*$. By (i) and (ii),

$$L(T_0) - L(T_1) = O(\epsilon^2 L_1 + \epsilon L_2), \tag{7}$$

and by (i), we have the bound

$$\angle s_1^*s_2s_1 = O(\epsilon). \tag{8}$$

The new tree T_1 has one less pseudo-Steiner point.

We proceed inductively: apply the above process to T_1 to obtain T_2 (with suitable constraints replacing (i) and (ii), as specified below), and so on, in the kth step working on a cherry of T_k. Each time a pseudo-Steiner vertex s of T_0 is replaced, the new vertex created is called s^*. The general argument will be similar to the one described above, except that now the angles of the edges at the pseudo-Steiner point are no longer within ϵ of $120°$, and the lengths of the edges do not represent precisely the length of T_0 in the corresponding branch. We need to quantify the build-up of these errors, so make the following definitions. For any terminal q of T_k, if q is a terminal of T_0, define $f(q) = g(q) = 0$. Otherwise, $q = s_j^*$ for some pseudo-Steiner point s_j of T_0. Letting s denote the vertex of T_k adjacent to q, define $f(q)$ to be the angle between sq and the edge ss_j of T_0. (For example, note that (8) says just $f(s_1^*) = O(\epsilon)$.) Also define $\bar{L}(sq)$ to be the total length of the branch of T_0 at s containing the vertex s_j, and let $g(q)$ be the absolute value of the difference between $|sq|$ and $\bar{L}(sq)$. See Figure 2. Note that $f(q)$ and $g(q)$ become defined either in T_0 if q is a terminal of it, or in some T_i when q first gets created. They have the same values in all trees T_i which contain q.

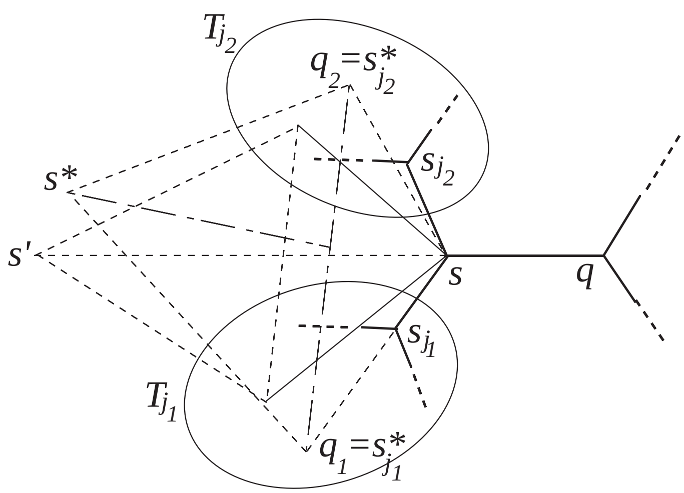

Figure 2: A cherry in T_k

Assume that after k iterations ($k \geq 1$), a pseudo-Steiner vertex s is adjacent to the two terminals q_1 and q_2 of T_k. That is, in T_0, the pseudo-Steiner vertex s is adjacent to points s_{j_1} and s_{j_2} such that for $i = 1$ and 2, either s_{j_i} is a terminal of T_0 or the branch of T_0 containing s_{j_i} has been replaced by an edge $sq_i = ss_{j_i}^*$.

For $i = 1$ and 2 let $L_i(s) = |sq_i|$. First suppose that q_1 and q_2 are terminals of T_0. Then the argument above applies, where $L_1 \leq L_1 + L_2$ and $L_2 = \min\{L_1, L_2\}$. So, to avoid making any assumption about the relative magnitudes of $L_1(s)$ and $L_1(s)$, we have as a result of (8)

$$f(s^*) = O(\epsilon(L_1(s) + L_2(s))),$$

and from (7)

$$g(s^*) = O(\epsilon^2(L_1(s) + L_2(s)) + \epsilon \min\{L_1(s), L_2(s)\}).$$

Now repeat the argument above in the case that q_1 and q_2 are not both terminals of T_0. The main difference is that now the edges at s can be made to have all angles within ϵ of 120° by rotating them through angles of $f(q_1)$ and $f(q_2)$ respectively, and we aim to bound $f(s^*)$ and $g(s^*)$. (The plane Π may be ambiguously defined if the angle errors have built up to the point that the three points determining it are collinear, but any choice will do the job in that case. The multiplicative error bouond is then constant, or larger.)

The effect of non-zero $f(q_1)$ and $f(q_2)$ is to increase the out-of-plane and in-plane angle errors α and β by up to $f(q_1)+f(q_2)$. (Actually, the out-of-plane and in-plane components of this error cannot both simultaneously be so large, and possibly analysing the split-up would improve the result, especially if the constraint $\epsilon < n^{-2}$ were relaxed.) Then we have

$$\alpha \leq O(\sqrt{\epsilon}) + f(q_1) + f(q_2), \qquad \beta \leq O(\epsilon) + f(q_1) + f(q_2) \tag{9}$$

and considering equations (5) and (6), whose derivation still applies in the current situation,

$$\begin{aligned} f(s^*) &\leq \frac{O(\alpha^2 \min\{L_1(s), L_2(s)\}) + \beta(L_1(s) + L_2(s))}{L_1(s) + L_2(s)} \\ &\leq O(\alpha^2) + \beta, &(10)\\ g(s^*) &\leq O\big(\alpha^2 \min\{L_1(s), L_2(s)\} + \beta^2(L_1(s) + L_2(s))\big) + g(q_1) + g(q_2) &(11) \end{aligned}$$

Define a function ν for any terminal q of T_k as follows. Let $\nu(q) = 1$ if q is a terminal of T_0. Otherwise, $q = s_j^*$ for some j, and define $\nu(q)$ to be the number of terminals of T_0 which lie in the component of $T_0 - s$ containing s_j.

We now show from (10) by induction on k (or, if you prefer, on $\nu(s^*)$) that there exists a $C > 0$ (independent of n, T and ϵ) such that

$$f(s^*) < C\nu(s^*)\epsilon. \tag{12}$$

To establish this, observe firstly that for the initialisation, the statement can be extended to terminals s^* of T_0 and is then immediate. Next observe that the inductive assumptions $f(q_i) < C\nu(q_i)\epsilon$ ($i = 1$ and 2) imply using (9)

$$\alpha = O(\sqrt{\epsilon}) + O(n\epsilon) = O(\sqrt{\epsilon})$$

as $\epsilon < n^{-2}$, and so $\alpha^2 = O(\epsilon)$. Then (12) follows from (9), (10) and

$$\nu(q_1) + \nu(q_2) = \nu(s^*). \tag{13}$$

We now have $\beta = O(n\epsilon)$ from (9) and the fact that $\nu(s^*) < n$, and so from (11)

$$g(s^*) \leq O\big(\epsilon \min\{L_1(s), L_2(s)\} + n^2\epsilon^2(L_1(s) + L_2(s))\big) + g(q_1) + g(q_2). \tag{14}$$

We next prove by induction from (14) that for some absolute constant $C > 0$

$$g(s^*) \leq C(\epsilon \log \nu(s^*) + n^2(\nu(s^*) - 1)\epsilon^2)\bar{L}(s) \tag{15}$$

where

$$\bar{L}(s) = \bar{L}(sq_1) + \bar{L}(sq_2), \tag{16}$$

recalling that $\bar{L}(sq_i)$ denotes the length of the branch of T_0 at s corresponding to sq_i. Again, (15) can be extended so as to hold for terminals s^* of T_0 since then $g(s^*) = 0$. Since $\epsilon < 1/n^2$, (15) applied inductively gives $g(q_i) = O(\sqrt{\epsilon}\bar{L}(sq_i))$. Hence from the definition of g, $L_i(s) = \bar{L}(sq_i)(1 + O(\sqrt{\epsilon}))$, and so, noting from (14)

$$g(s^*) \leq O\big(\epsilon \min\{\bar{L}(sq_1), \bar{L}(sq_2)\} + n^2\epsilon^2\bar{L}(s)\big) + g(q_1) + g(q_2).$$

Hence, applying (15) inductively and using (13) and $\bar{L}(q_i) \leq \bar{L}(sq_i)$ (defining $\bar{L}(q_i) = 0$ if q_i is a terminal of T_0), we obtain

$$\begin{aligned}
g(s^*) \;&\leq\; O\big(\epsilon \min\{\bar{L}(sq_1), \bar{L}(sq_2)\} + n^2\epsilon^2\bar{L}(s)\big) + Cn^2(\nu(s^*) - 2)\epsilon^2\bar{L}(s) \\
&\quad + C\epsilon\sum_{i=1}^{2}\bar{L}(sq_i)\log\nu(q_i) \\
&\leq\; Cn^2(\nu(s^*) - 1)\epsilon^2\bar{L}(s) + \frac{1}{4}C\epsilon\min\{\bar{L}(sq_1), \bar{L}(sq_2)\} + C\epsilon\sum_{i=1}^{2}\bar{L}(sq_i)\log\nu(q_i)
\end{aligned}$$

upon choosing C to be four times the size of the constant implicit in the $O(\ )$. By (13) and (16), the latter can be written as

$$g(s^*) \leq Cn^2(\nu(s^*) - 1)\epsilon^2\bar{L}(s) + C\epsilon\bar{L}(s) \sup_{\substack{0\leq x\leq 1/2 \\ 1\leq y\leq \nu(s^*)-1}} \{x/4 + x\log y + (1 - x)\log(\nu(s^*) - y)\}$$

where x stands for $\bar{L}(sq_1)/\bar{L}(s)$ and y stands for $\nu(q_1)$, assuming without loss of generality that $\bar{L}(sq_1) \leq \bar{L}(sq_2)$. Fixing x in the last expression shows the maximum occurs for $y = 1$, and then the expression becomes

$$x/4 + (1 - x)\log(\nu(s^*) - y) \leq x/4 + (1 - x)\log\nu(s^*).$$

Since $\log\nu(s^*) \geq \log 2 > 1/4$, this last bound is maximised at $x = 0$ and so the supremum has value at most $\log\nu(s^*)$. From this, (15) follows.

We now apply (15) to the pseudo-Steiner point s adjacent to the root vertex $\hat{t}$, since for this s we have

$$g(s^*) = ||\hat{t}s^*| - \bar{L}(\hat{t}s^*)| = ||\hat{t}s^*| - L(T_0)|.$$

By Theorem 3.1, $L(S(T_0)) \geq |\hat{t}s^*|$. As $L(S(T_0)) \leq L(T_0)$, we obtain Theorem 2.2. ∎

Acknowledgement The authors would like to acknowledge Marcus Brazil and Doreen Thomas for useful discussions.

References

[1] S. Arora, Polynomial time approximation schemes for Euclidean TSP and other geometric problems, in *Proc. 37-th Ann. Symp. on Foundations of Computer Sciences*, pp. 2-13, 1996.

[2] E.J. Beasley, A heuristic for Euclidean and rectilinear Steiner problems, *European J. Oper. Research* **58** (1992), 299-327.

[3] M. Brazil, D.A. Thomas and J.F. Weng, Upper and Lower Bounds for the Lengths of Steiner Trees in 3-Space, *Geometriae Dedicata*, **109** (2004), 107-119.

[4] S.K. Chang, The generation of minimal trees with a Steiner topology, *J. ACM*, **19** (1992), 699-711.

[5] F. K. Hwang, D. S. Richards and P. Winter, *The Steiner Tree Problem; Annals of Discrete Mathematics 53*, Amsterdam, Elsevier Science Publishers B.V., 1992.

[6] A. Lin and S.-P. Han, On the distance between two ellipsoids, *SIAM J. Optimization* **13** (2002), 298–308.

[7] W. D. Smith, How to find Steiner minimal trees in Euclidean d-space, *Algorithmica* **7** (1992), 137–177.

[8] A.Z. Zelikovsky, An 11/6-approximation algorithm for the network Steiner problem, *Algorithmica*, **9** (1993), 463-470.