

Euclidean Steiner Trees Optimal with Respect to Swapping 4-point Subtrees

Doreen A. Thomas · Jia F. Weng

Received: date / Accepted: date

Abstract The Steiner tree problem in Euclidean space E^3 asks for a minimum length network T , called a Euclidean Steiner Minimum Tree (ESMT), spanning a given set of points. This problem is NP-hard and the hardness is inherently due to the number of feasible topologies (underlying graph structure of T) which increases exponentially as the number of given points increases. Planarity is a very strong condition that gives a big difference between the ESMT problem in the Euclidean plane E^2 and in Euclidean d -space $E^d (d \geq 3)$: The ESMT problem in the plane is practically solvable whereas the ESMT problem in d -space is really intractable. The simplest tree rearrangement technique is to repeatedly replace a subtree spanning 4 nodes in T with another subtree spanning the same 4 nodes. This technique is referred to as the *Swapping 4-point Topology/Tree (S4pT)* technique in this paper. An indicator (or quasi-indicator) of T plays a similar role in the optimization of the length $L(T)$ of T in the discrete topology space (the underlying graph structure of T) to the derivative of a differentiable function which indicates a fastest direction of descent. T will be called S4pT-optimal if it is optimal with respect to swapping 4-point subtrees. In this paper we first make a complete analysis of 4-point trees in Euclidean space exploring all possible types of 4-point trees and their properties. We review some known indicators of 4-point ESMTs in E^2 , and give some simple geometric proofs of these indicators. Then, we translate these indicators to E^3 , producing 8 quasi-indicators in E^3 using computational experiments, the best quasi-indicator ρ_{osr} is sifted with an effectiveness for randomly generated 4-point sets as high as 98.62%. Finally we show how ρ_{osr} is used to find an S4pT-optimal ESMT on 14 probability vectors in $4d$ -space with a detailed example.

Keywords Steiner point, Steiner minimal tree, quasi-indicator, computational experiments.

D. A. Thomas

Department of Mechanical Engineering, The University of Melbourne, VIC 3010, Australia
E-mail: doreen.thomas@unimelb.edu.au

J. F. Weng E-mail: jfweng@unimelb.edu.au

1 Introduction and Motivation

The *Steiner tree problem* asks for a minimum length network T on a given set V of n points in a metric space. The network $T = T(V)$ is called a *Steiner minimum tree* (SMT). In particular, if the tree is in Euclidean d -space E^d then it is called a Euclidean Steiner minimum tree (ESMT). The length $L_{T(V)}$ (or $L(T)$) of T depends on the given point set (the *configuration*) V , as well as the *topology* τ , which is the underlying graph structure of T . To shorten the length $L_T = L_{T(V)}$, some additional points s_1, s_2, \dots may be added to V , which are called *Steiner points*, while the points in V are called *terminals*. Note that the location of the Steiner points $s_j = s_j(t)$ in T depends on the topology τ . In general, a tree topology with all points having degree at most three is called a *Steiner topology*. In this paper the word ‘Steiner’ is often abbreviated as ‘S-’, e.g. *S-point*, *S-tree*, *S-topology*. A Steiner topology, or a tree, is *full* if all terminals are of degree one. As a result all Steiner points in a full Steiner topology are of degree three. In a full topology/tree an edge incident to a terminal is called a *terminal edge* while all other edges are *interior edges*. If some edges in a topology/tree are shrunk to points, then the topology/tree is called *degenerate* [11,12]. Since the topology τ_1, τ_2, \dots is a discrete variable in the topology space \mathcal{T} while the locations of Steiner points s_j ($j = 1, 2, \dots$) are continuous variables in E^d , the Steiner tree problem in E^d is a mixed type of optimization problem and can be formally defined as an unconstrained optimization problem as follows: Given a point set V in E^d , find a network $T(V)$ whose length is

$$\min_{\tau_i \in \mathcal{T}} \min_{s_j(\tau_i) \in E^d} L_{T(V)}(\tau_i, s_j(\tau_i)). \quad (1)$$

There is also a graph version of the Steiner tree problem where an underlying graph \mathcal{G} is given and all the given points and Steiner points are nodes of the graph \mathcal{G} . Both versions of the Steiner tree problem have been proved to be NP-hard [13,9]. The Steiner tree problem has a wide range of applications such as computational phylogenetic analysis [3,26,27]; the structure of molecules (modeling and folding of molecular structure) [15,21]; in communication networks [4] and in the design of underground mining networks [2].

A network is *S-point optimal* if the locations of S-points are optimal for a fixed topology, *i.e.* optimal over all networks with the same topology, while a network is *S-topology optimal* if its length is S-point optimal over all feasible topologies. The two levels of optimization are clearly seen in the objective function (1). The NP-hardness of the ESMT problem is inherently due to the number of feasible topologies, *i.e.*, the size of \mathcal{T} increases exponentially with the number of given points. Because of the convexity of the Euclidean metric, once a topology is fixed, locating all the Steiner points in T (S-point optimization) is a convex optimization problem which can be solved by a convex optimization program, *e.g.*, CVX [10]. Therefore, the ESMT problem is NP-hard essentially on the S-topology level. The exponential increase of the size of \mathcal{T} does not mean that the ESMT problem is intractable in all cases. In fact, planarity is a very strong condition that gives a big difference between the ESMT problem in the Euclidean plane and in Euclidean d -space ($d \geq 3$).

- At the S-point optimization level, the length of a full Steiner tree L_T in the plane has a closed form which can be derived from the Melzak Euclidean construction [14] or from the hexagonal coordinate method [29, 24]. On the other hand, it has been proved that if $d \geq 3$ and if V has at least 4 terminals, then L_T does not have a closed form, that is, L_T cannot be computed by radicals (*algebraically unsolvable*) [19] and has to be computed by numerical iteration.
- At the S-topology optimization level, the divergence between the planar and spatial Euclidean Steiner tree problem is clearly shown in the existing algorithms for the ESMT problem. For the planar case, the program GeoSteiner can solve an ESMT problem for more than 10,000 terminals [1] while the best program for finding ESMTs in d -space, to the authors' knowledge, is either the original program proposed by W. Smith [20] or its improvement [6], which can only solve the ESMT problem in d -space for up to 10 terminals.

From the above facts we can conclude that the ESMT problem in the plane is practically solvable whereas the ESMT problem in d -space ($d \geq 3$) is really intractable. Hence the question arises: can we find an algorithm for constructing sub-optimal ESMTs on the S-topology level? Here 'sub-optimal' means locally optimal in the topology space \mathcal{T} . This problem is the focus of this paper.

Note that there is only one S-topology where there is a single S-point, consequently an S-topology with two S-points becomes the 'cell' of S-topologies. As we know there are 3 different S-trees, say $T_i^s, i = 1, 2, 3$ on four points. Hence, the simplest topology rearrangement technique for an S-tree T spanning n (≥ 4) terminals is to repeatedly replace a sub-tree T_i^s spanning 4 nodes in T with another sub-tree T_j^s ($j \neq i$) spanning the same 4 nodes. This technique is well known as *Nearest Neighbor Interchange (NNI)* and has been widely applied in computational phylogenetic analysis [22, 8, 23]. However, to the authors' knowledge, nobody has used the NNI technique in order to find Steiner minimal trees in Euclidean d -space. For simplicity, in this paper an S-topology or an S-tree on four points will be referred to as a *4-point topology* or *4-point tree*, and the interior edge of the 4-point topology or tree is referred as the *mid-topology* or *mid-edge* of the tree, respectively. This basic topology rearrangement technique NNI will be referred to as the *Swapping 4-point Topology/Tree (S4pT)* technique in this paper. An ESMT T is called *S4pT-optimal* if it cannot be improved by the S4pT technique, that is, all 4-point subtrees in T are optimal. The S4pT technique can be used to find sub-optimal ESMTs in E^d ($d \geq 3$) due to the following fact [25]:

Proposition 1 *Any two full Steiner topologies in \mathcal{T} can be changed, each to other, by consecutively swapping 4-point sub-topologies.*

This paper has three parts. In the first part (Section 2) we make a complete analysis of 4-point trees in Euclidean space in order to explore all the possible (degenerate or non-degenerate) types of 4-point trees, their properties and relations, and to indicate how to locate the Steiner points in each type of 4-point tree. Because the length of locally minimal 4-point trees does not have a closed form in E^d ($d \geq 3$), a new optimization search technique, called *quasi-indicator*, was introduced [30]. Roughly speaking, an indicator is a function that is simple to compute and that indicates the optimal solution to the original optimization problem while a quasi-indicator is an indicator that is valid in almost all instances (*i.e.*, valid with a high probability) of the optimization problem. For a differentiable objective function $f(X)$, the derivative $f'(X)$ indicates the direction of fastest descent. However,

the topology τ is a discrete variable for $L(T)$. In this case an indicator plays a similar role to the derivative $f'(X)$ for the objective $f(X)$, and *S4pT-optimal is locally optimal in the topology space T* . In the second part (Section 3) we first review some known indicators of 4-point minimal trees in the Euclidean plane, and give some simple geometric proofs of these indicators. Then, we translate these indicators to Euclidean space E^3 , producing 8 quasi-indicators for 4-point minimum trees in E^3 . Finally, we design computational experiments on these quasi-indicators that extend (both in the number of quasi-indicators and in the size of experimental data) our previous experiments given in [30]. From the computational experiments the best quasi-indicator ρ_{osr} is sifted whose effectiveness for randomly generated 4-point sets is as high as 98.62%. In the last part (Section 4) we show how ρ_{osr} is used for finding an S4pT-optimal ESMT on 14 probability vectors in $4d$ -space in a detailed example.

2 A Complete Analysis of 4-Point Trees in Euclidean Space

2.1 Types of 4-point trees

As a network optimization problem the Steiner tree problem can be studied in any ℓ_p , $p \geq 1$, space. In the Euclidean plane for a fixed Steiner topology the Steiner tree $T(V)$, for any size of V , can be constructed in $O(n)$ time either by the Melzak construction [14] or by the hexagonal coordinate method [29]. Both of the methods are based on locating a non-degenerate Steiner point s spanning three given points, say a, b, c . The basic property of this degree-3 Steiner point s is the so-called 120° condition (see below Corollary 1). Let e_{ab} be the third vertex of the equilateral triangle based on the line segment ab and lying on the opposite side of c . This point is called an *E-point* of $\triangle abc$ with respect to ab . The important properties of E-points are given in the following proposition.

Proposition 2 Suppose e_{ab} is the E-point of $\triangle abc$ with respect to ab .

- (1) The line ce_{ab} (called a *Simpson line*) intersects ab at a point j_{ab} (the *Simpson intersection* on ab) [19].
- (2) The Steiner point s for $\triangle abc$ lies on ce_{ab} . Hence, s can easily be computed as the intersection of two Simpson lines [25].

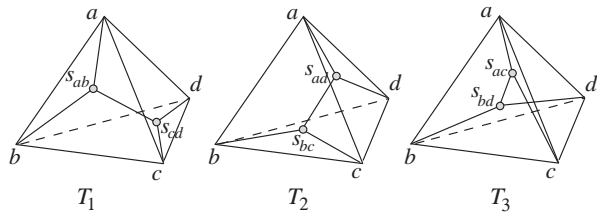


Fig. 1 Three full Steiner topologies on four points.

A Steiner point s may be *degenerate*. There are two types of degeneracies.

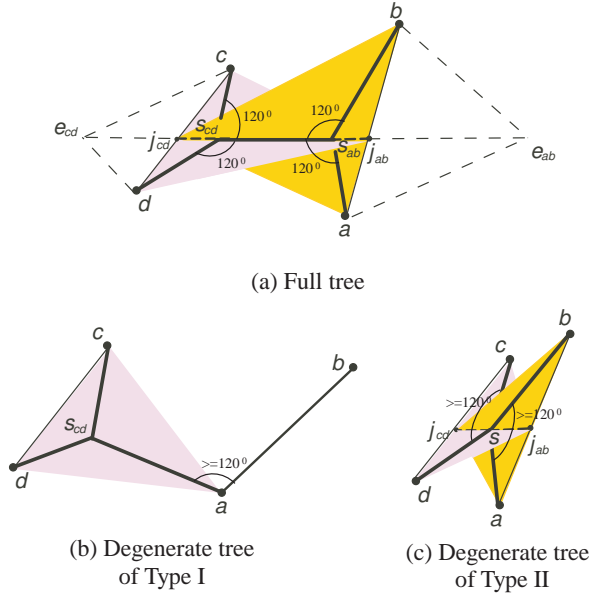


Fig. 2 Three types of 4-point trees.

1. **(Type I)** s collapses into a neighboring terminal p . In this case p becomes a terminal of degree greater than one and the tree becomes a union of full trees.

2. **(Type II)** Two adjacent Steiner point s_1, s_2 collapse into each other. In this case the new point $s = s_1 = s_2$ becomes a degree-4 point and the topology τ is not actually a Steiner topology but a *degenerate Steiner topology*. However, for convenience we still refer to s as a Steiner point and τ as a Steiner topology.

For the rest of this paper suppose V is a point set consisting of four points a, b, c, d in Euclidean space. For the spatial set V there exist three full Steiner topologies, which are denoted by τ_i , and the corresponding ESMTs are denoted by $T_i = T(V, \tau_i)$, $i = 1, 2, 3$ as shown in Fig. 1. If a Steiner point joins two terminals, say p, q , then it is denoted by s_{pq} . Similarly, let m_{pq} denote the midpoint of a line segment pq . Moreover, similar to the notation for Steiner points and midpoints, the three trees T_i can be directly denoted as $T_{ab-cd}, T_{ad-bc}, T_{ac-bd}$ as shown in Fig. 1. If the tree currently studied is T_{ab-cd} , then the other two trees T_{ad-bc}, T_{ac-bd} are referred to as trees *associated with* T_{ab-cd} .

2.2 Locating Steiner points

Now we discuss how to locate the Steiner points in each of the three types of 4-point trees in space.

(i) A full 4-point tree.

As stated in Section 1, there are no closed forms for computing the two Steiner points. However, we can alternately, and iteratively, locate the Steiner points using Proposition 2. Take $T_1 = T_{ab-cd}$ as an example. We can locate the Simpson intersections j_{ab}, j_{cd} (Fig. 2(a)) alternately and the procedure is convergent because

of the convexity of the tree length. Once j_{ab}, j_{cd} are determined then s_{ab}, s_{cd} can be located using the 120° condition. The reader can refer to [19] for the detailed iteration formulae.

(ii) A degenerate tree of Type I.

Suppose ab in T_1 is tilted around a away from the Simpson intersection j_{cd} , then the Steiner point s_{ab} collapses into terminal a , and T becomes a union of a full Steiner tree and a line segment ab (Fig. 2(b)). In this case s_{cd} can be determined directly using the three terminals a, c, d and we need only to check and to confirm that $\angle bas_{cd} \geq 120^\circ$. (Note that a degenerate tree of Type I can contain not only one but two degenerate Steiner points, *i.e.*, not only s_{ab} but also s_{cd} may collapse into an adjacent terminal.)

(iii) A degenerate tree of Type II.

Suppose ab in T_1 moves toward cd along $s_{ab}s_{cd}$, then the two Steiner points s_{ab}, s_{cd} will become closer and closer. When ab and cd are close enough the Steiner points will collapse into one (Fig. 2(c)). Such a degree-4 Steiner point s can be determined using the *variational method*, that is, s can be determined by the variations (directional derivatives) of the lengths of edges joining s . (In E^3 three variations in three independent directions are needed.) Suppose the endpoint p of a line segment pq is perturbed in direction \vec{pr} with q fixed. Then the directional derivative of the length $L(pq)$ is [18]

$$\dot{L}(pq) = -\cos(\angle qpr) = \frac{|qp|^2 + |pr|^2 - |qr|^2}{2|qp||pr|} \stackrel{\text{def}}{=} f_{\text{var}}(qpr). \quad (2)$$

In principle we can use any three independent directions to locate a degree-4 Steiner point s but the simplest set is the three axis directions. Let $s = [x, y, z]$, $X = [1, 0, 0]$, $Y = [0, 1, 0]$, $Z = [0, 0, 1]$. Then x, y, z are determined by the variations of the lengths of as, bs, cs and ds :

$$\begin{aligned} f_{\text{var}}(asX) + f_{\text{var}}(bsX) + f_{\text{var}}(csX) + f_{\text{var}}(dsX) &= 0, \\ f_{\text{var}}(asY) + f_{\text{var}}(bsY) + f_{\text{var}}(csY) + f_{\text{var}}(dsY) &= 0, \\ f_{\text{var}}(asZ) + f_{\text{var}}(bsZ) + f_{\text{var}}(csZ) + f_{\text{var}}(dsZ) &= 0. \end{aligned} \quad (3)$$

Remark 1 Note that like the Steiner points in a full 4-point tree, this system is also algebraically unsolvable, and an iteration algorithm is needed to solve the system of Equations (3). Note also that the initial solution $s_0 = [x_0, y_0, z_0]$ should lie inside the tetrahedron $abcd$, *e.g.*, $s_0 = (a + b + c + d)/4$, otherwise the iteration procedure may converge to a point outside the tetrahedron and the point is not the required degree-4 Steiner point.

An important fact in the variational argument is given by the following lemma [18].

Lemma 1 *Suppose two edges ps and qs meet at s with angle $\alpha < 180^\circ$ and s is perturbed in $\angle psq$. Then, the variation of $L(ps) + L(qs)$ is minimized when s moves along the bisector of the angle. Consequently, the variation of $L(ps) + L(qs)$ is maximized when s moves along the opposite direction to the bisector of the angle.*

Corollary 1 (120° condition) *Any Steiner point s of degree 3 in an ESMT T in Euclidean space E^d ($d \geq 2$) lies in the triangle composed of its three adjacent points and all three angles at s are 120° . It follows that if T spans 3 points a, b, c in E^d then T lies in the plane determined by $\triangle abc$.*

Corollary 2 (splitting angle) *If an angle pqr is strictly less than 120° , then $|pq| + |qr| > |ps| + |qs| + |rs|$ where s is the S -point for triangle pqr .*

Theorem 1 *If an S -tree T contains a Steiner point s whose degree is no less than 4, then T cannot be globally optimal.*

Proof Suppose s spans at least 4 nodes in T , let T_s be the subtree consisting of the edges linking s . Then, there are at least two edges of s , say sa, sb such that $\angle asb$ is strictly less than 120° . Hence, T as well as T_s is not optimal by the above corollary. ■

The variational argument also gives the following properties of degree-4 Steiner points. Let \vec{sx} denote an arrow with the direction from s to x .

Theorem 2 *Suppose T_s spanning four distinct points a, b, c, d is an S -point optimal tree with s as a degree-4 S -point. The 6 sides of the tetrahedron $abcd$ form 3 pairs of opposite sides (ab, cd) , (ac, bd) , (ad, bc) , and the four terminal edges as, bs, cs, ds form 3 pairs of opposite angles $(\angle asb, \angle csd)$, $(\angle asc, \angle bsd)$, $(\angle asd, \angle bsc)$ subtending the aforementioned 3 pairs of opposite sides, see (Fig. 3).*

(1) *The bisectors of two opposite angles in each pair lie in a line but in opposite directions.*

(2) *The opposite angles in each pair of angles are equal.*

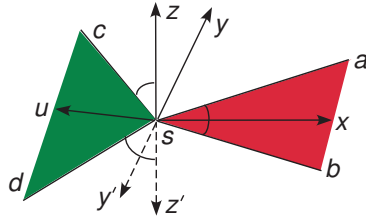


Fig. 3 A 4-point tree with a degree-4 Steiner point.

Proof From the 120° condition we can assume that not all four points lie in a plane and s is not lying on any surface of the tetrahedron $abcd$. Moreover, it is sufficient that we analyze one pair of opposite angles, say $(\angle asb, \angle csd)$. Let sx be the bisector of $\angle asb$ and su be the bisector of $\angle csd$ (see Fig. 3). Let sx, sy, sz be the Cartesian axes.

(1) If $\angle csz < \angle dsz'$ where sz' is the opposite direction of sz , then moving s along sz results in $f_{\text{var}}(asz) = f_{\text{var}}(bsz) = 0$, $f_{\text{var}}(csz)$ is negative and $|f_{\text{var}}(csz)| > |f_{\text{var}}(dsz)|$. Hence, the total variation of $L(T)$ is negative, contradicting the T_s is S -point optimality. On the other hand, if $\angle csz > \angle dsz'$, then moving s along z' results in a similar contradiction. This proves $\angle csz = \angle dsz'$. Considering $\angle csy$ and $\angle dsy'$, the same argument leads to $\angle csy = \angle dsy'$ where sy' is the opposite direction of sy . Now, both $\angle csz = \angle dsz'$ and $\angle csy = \angle dsy'$ make the bisector su of $\angle csd$ lies in the same line of sx but in the opposite direction.

(2) Suppose $\angle asb \neq \angle csd$, say $\angle asb < \angle csd$, then moving s along sx reduces the length $L(T_s)$, again by the variational argument. Hence, $\angle asb = \angle csd$. ■

2.3 Two informative examples

Now we give two examples to show that all three types of trees may exist for a point set V and the statements in Theorems 1 and 2 are verified in these examples.

Example 1

$$a = [0.92181, 0.93547, 0.057891], b = [0.73821, 0.9169, 0.35287],$$

$$c = [0.19627, 0.41027, 0.81317], d = [0.40571, 0.89365, 0.0098613].$$

Then $T_1 = T_{ab-cd}$, $T_2 = T_{ad-bc}$, $T_3 = T_{ac-bd}$ are full, degenerate of Type I, and degenerate of Type II, respectively (Fig. 4). The lengths of these trees as well as the coordinates of Steiner points and angles at Steiner points are given in Table 1.

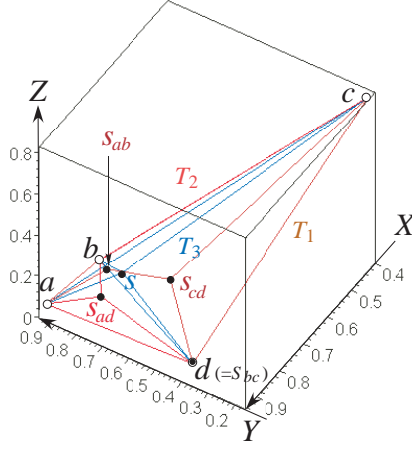


Fig. 4 An example of 3 types of trees.

Table 1 Example 1: A configuration where the three different types of trees exist.

	T_1 (full)	T_2 (degenerate of Type I)	T_3 (degenerate of Type II)
L_{T_i}	1.638106774	1.641205510	1.6834288903
$s_1(x)$.7216366352	.7306753086	.8820988691
$s_1(y)$.9024587443	.9183281456	.6834288903
$s_1(z)$.2994282599	.1798839567	.2794060773
$s_2(x)$.5467595627	$s_{bc} = b$	$s = s_1 = s_2$
$s_2(y)$.8351639592		
$s_2(z)$.2692329913		
angle	all 120°	$\angle s_{ad}bc = 120.29678609^\circ$ all 120° at s_{ad}	$\angle asb = \angle csd = 155.7797741^\circ$ $\angle asd = \angle bsc = 92.63874341^\circ$ $\angle asc = \angle bsd = 92.40633766^\circ$

Remark 2 Degenerate trees of Type II cannot be SMTs as proved in Theorem 1, whereas a degenerate trees of Type I can be an SMT. In Example 1, suppose

terminal c undergoes a small perturbation so that its x -coordinate increases by 0.03, then the three types of topology do not change but $L_{T_1} = 1.624807351$ and $L_{T_2} = 1.622907280$. Hence, T_2 , a degenerate tree of Type I, becomes the SMT for $abcd$.

Remark 3 Theorem 1 prevents all three trees being degenerate trees of Type II. But two trees being degenerate trees of Type II is possible as shown in the example below.

Example 2

$$a = [.28440859, .58279168, .43290660], \quad b = [.46922429, .42349626, .22594987], \\ c = [.64781123, .51551175, .57980687], \quad d = [.98833494, .33395148, .76036501].$$

Then both T_1, T_2 are degenerate trees of Type II but T_3 as a degenerate tree of Type I is optimal. The lengths of these trees as well as the coordinates of Steiner points and angles at Steiner points are given in Table 2.

Table 2 Example 2: two degenerate trees of Type II and one degenerate tree of Type I.

	$T_1 (= T_2)$	T_3
L_{T_i}	1.402384883	1.312319415
$s_1(x)$.30778963	$s_{ac} = a$
$s_1(y)$.548465793	
$s_1(z)$.440453663	
$s_2(x)$	$s = s_1 = s_2$.46703175
$s_2(y)$.46532777
$s_2(z)$.35292462
angle	$\angle asb = 121.03873^\circ$ $\angle asd = 141.10041^\circ$ $\angle asc = 72.91050^\circ$	$\angle s_{bd}ac = 134.16087^\circ$ all 120° at s_{bd}

3 Finding Optimal 4-Point Trees in Euclidean Space

3.1 Equivalence, indicators and quasi-indicators in optimization problems

Suppose in an optimization problem the objective function is $F(V, \tau)$ where τ is a variable and V is a parameter in domain Ω . Often $F(V, \tau)$ is complicated and its optimal value is hard to found. To reduce the hardness of optimization problems, recently a concept *equivalence* was introduced [30] and here we describe a brief introduction to this concept. Suppose there is a different objective function $f(V, \tau)$ with the same constraints such that $F(V, \tau)$ and $f(V, \tau)$ have the same optimal point t^* , i.e., both functions achieve their minimum values at t^* . Then, we say that the two functions are *equivalent* with respect to optimization although their optimal values may not be equal, $F(V, t^*) \neq f(V, t^*)$. If such an equivalence does not always hold, then we can ask in how many cases does the equivalence hold. That is, define the probability

$$q(f, F) \stackrel{\text{def}}{=} Pr \left(t^f - t^F = 0 \mid V \in \Omega \right),$$

where t^f, t^F are the optimal points of $f(V, \tau), F(V, \tau)$ respectively. We call $q(f, F)$ the *effectiveness index* for equivalence of the objective function pair $(F(V, \tau), f(V, \tau))$. If $q(f, F) = 1$, then $f(V, \tau)$ is called an *indicator* of the original objective function $F(V, \tau)$. However, if $q(f, F) \neq 1$ but $q(f, F)$ is very close to one and if $f(V, \tau)$ is simpler to optimize than $F(V, \tau)$, more precisely, if the computational complexity of $f(V, \tau)$ is lower than $F(V, \tau)$, then in most practical cases $f(V, \tau)$ can be used to replace $F(V, \tau)$ in order to find the optimal point t^* and thus to solve the original optimization problem. In that case $f(V, \tau)$ is a *quasi-indicator* of $F(V, \tau)$ (or *q-indicator* for short).

3.2 Indicators of optimal 4-point trees in the Euclidean plane

In this section we consider the planar case for the indicators of optimal 4-point trees. If all four points of V lie in a plane, then at most two of three full Steiner topologies can be realized as Steiner trees. Suppose the four points in the plane form a quadrilateral $abcd$, then the trees $T_1 = T_{ab-cd}, T_2 = T_{ac-bd}$ can exist no matter what type they are. There are several functions equivalent to identify the lengths of optimal 4-point trees in the plane, and we call the equivalences rules for identifying optimal 4-point trees. The first two rules (diagonal angle rule and mid-edge rule) were found by Pollak [17] and Ollerenshaw [16] independently in the same year. Instead of the original long complicated proofs we give two simple geometrical proofs. Let o be the intersection of two diagonals ac and bd . Let $\phi(V, \tau_i)$ be the angles at o subtended by the opposite sides of $abcd$ in $T_i, i = 1, 2$, i.e., $\phi(V, \tau_1) = \angle aob, \phi(V, \tau_2) = \angle aod$ (Fig. 5).

Theorem 3 (Diagonal angle rule) *If T_1 and T_2 both exist and are both full, then $L_{T_1}(V, \tau_1) \leq L_{T_2}(V, \tau_2)$ if and only if $\phi(V, \tau_1) \leq \phi(V, \tau_2)$.*

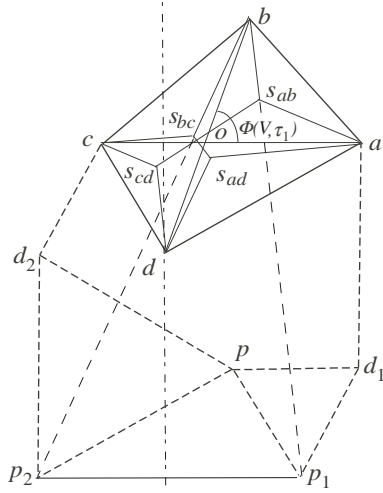


Fig. 5 Proof of the diagonal angle rule.

Proof By the Melzak construction let d_1 (d_2) be the E-point with respect to ad (cd , respectively), let p_1 (p_2) be the E-point with respect to ad_2 (cd_1 , respectively) (Fig. 5). Note that the Melzak construction can be regarded as a process of extending a tree by rotation. For T_1 , first the broken line cda rotates 60° anti-clockwise around a , resulting in pd_1a , then the line segment pd_1 rotates 60° anti-clockwise around d_1 , resulting in p_1d_1 and $|bp_1| = L_{T_1}$ (Fig. 5). Similarly for T_2 , first the broken line cda rotates 60° clockwise around c , resulting in pd_2c , then the line segment pd_2 rotates 60° clockwise around d_2 , resulting in p_2d_2 and $|p_2b| = L_{T_2}$. Because of the symmetry of the rotations, $p_1p_2 \parallel ac$, $|p_1p_2| = |ac|$, and d lies on the bisector of p_1p_2 . This implies that $L_{T_1}(V, \tau_1) \leq L_{T_2}(V, \tau_2)$ if and only if b lies on the same side of the bisector of p_1p_2 as a , i.e., if and only if $\phi(V, \tau_1) \leq \phi(V, \tau_2)$. ■

Let $f_{ie}(V, \tau_i)$ be the length of the mid-edge in T_i , i.e., $f_{ie}(V, \tau_1) = |s_{ab}s_{cd}|$, $f_{ie}(V, \tau_2) = |s_{ad}s_{bc}|$.

Theorem 4 (Interior edge rule) *If T_1 and T_2 both exist and are both full, then $L_{T_1} \leq L_{T_2}$ if and only if $f_{ie}(V, \tau_1) \geq f_{ie}(V, \tau_2)$.*

Du et. al. [5] gave a simple proof of Ollerenshaw's mid-edge rule using a transformation of the points in V . Essentially, their proof relies on an easily proved lemma.

Lemma 2 *Let a triangle have 3 edges u, v, w and the angle between u and v be 120° . Then*

$$|u| + |v|/2 = \sqrt{4w^2 - 3v^2}/2$$

is a decreasing function of v .

Let $f_{te}(V, \tau_i)$ be the sum of the lengths of the four terminal edges for the Steiner topology τ_i . That is,

$$f_{te}(V, \tau_1) = |as_{ab}| + |s_{ab}b| + |s_{cd}c| + |s_{cd}d|, f_{te}(V, \tau_2) = |as_{ad}| + |s_{ad}d| + |bs_{bc}| + |s_{bc}c|.$$

By the above lemma the mid-edge rule can be geometrically re-proved as follows.

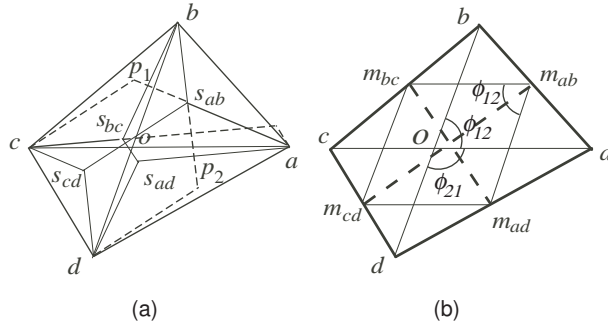


Fig. 6 Proof of (a) the mid-edge rule and (b) the midpoint distance rule.

Proof For T_1 let e_{ie} be the length of the mid-edge $s_{ab}s_{cd}$ and let e_a be the length of terminal edge as_{ab} . Similarly denote the lengths of other three terminal edge as e_b, e_c, e_d . Let as_{ab} extend to p_1 such that $cp_1 \parallel s_{cd}s_{ab}$ as shown in Fig. 6(a). Similarly, extending bs_{ab} to p_2 such that $dp_2 \parallel s_{cd}s_{ab}$. Then by Lemma 2

$$\begin{aligned} L_{T_1}(V, \tau_1) &= f_{te}(V, \tau_1) + f_{ie}(V, \tau_1) = e_a + e_c + e_{ie}/2 + e_b + e_d + e_{ie}/2 \\ &= |ap_1| + \frac{|cp_1|}{2} + |bp_2| + \frac{|dp_2|}{2} = \frac{\sqrt{(-3e_{ie}^2 + 4|ac|^2)}}{2} + \frac{\sqrt{(-3e_{ie}^2 + 4|bd|^2)}}{2}, \end{aligned} \quad (4)$$

which is a decreasing function of the length of its interior edge. The theorem is proved. \blacksquare

Because $L_{T_i}(V, \tau_i) = f_{te}(V, \tau_i) + f_{ie}(V, \tau_i)$ the following sum of terminal edges rule is equivalent to the mid-edge rule.

Corollary 3 (Sum of terminal edges rule) *If T_1 and T_2 both exist and are both full, then $L_{T_1} \leq L_{T_2}$ if and only if $f_{te}(V, \tau_1) \leq f_{te}(V, \tau_2)$.*

Re-denote the diagonal angle $\phi(V, \tau_i)$ as ϕ_{ij} , the angle subtending T_i with respect to T_j , that is, (Fig. 6(b))

$$\phi(V, \tau_1) = \phi_{12} = \angle aob, \quad \phi(V, \tau_2) = \phi_{21} = \angle aod = 180^\circ - \angle aob.$$

Let $f_{md}(V, \tau_i)$ be the midpoint distance between two opposite sides, i.e., $f_{md}(V, \tau_1) = |m_{ab}m_{cd}|$, $f_{md}(V, \tau_2) = |m_{ad}m_{bc}|$. Recently, the following midpoint distance rule was found [30].

Theorem 5 (Midpoint distance rule) *If T_1 and T_2 both exist and are both full, then $L_{T_1} \leq L_{T_2}$ if and only if $f_{md}(V, \tau_1) \geq f_{md}(V, \tau_2)$.*

Let $f_{os}(V, \tau_i)$ be the sum of the lengths of opposite sides of T_i and $f_{os^2}(V, \tau_i)$ be the sum of the squared lengths of opposite sides of T_i , that is,

$$f_{os}(V, \tau_1) = |ab| + |cd|, f_{os}(V, \tau_2) = |ad| + |bc|,$$

$$f_{os^2}(V, \tau_1) = |ab|^2 + |cd|^2, f_{os^2}(V, \tau_2) = |ad|^2 + |bc|^2.$$

In a private correspondence with the second author, Booth, R. S. gave the following rule.

Theorem 6 (Sum of squared lengths of opposite sides rule) *If T_1 and T_2 both exist and are both full, then $L_{T_1} \leq L_{T_2}$ if and only if $f_{os^2}(V, \tau_1) \leq f_{os^2}(V, \tau_2)$.*

Proof Let the diagonals bd and ac meet at o . By the cosine law

$$\begin{aligned} |ab|^2 &= |ao|^2 + |bo|^2 - 2\cos(\phi(V, \tau_1)) |ao||bo|, \quad |cd|^2 = |co|^2 + |do|^2 - 2\cos(\phi(V, \tau_1)) |co||do|, \\ |ad|^2 &= |ao|^2 + |do|^2 - 2\cos(\phi(V, \tau_2)) |ao||do|, \quad |bc|^2 = |bo|^2 + |co|^2 - 2\cos(\phi(V, \tau_2)) |bo||co|. \end{aligned} \quad (5)$$

Because $\phi(V, \tau_1) + \phi(V, \tau_2) = 180^\circ$ the theorem holds by Theorem 3. \blacksquare

Using f_{te}, f_{ie} and f_{os^2}, f_{md} we can define two ratios

$$\rho_{ter} \stackrel{\text{def}}{=} \frac{f_{te}}{f_{ie}}, \quad \rho_{os^2r} \stackrel{\text{def}}{=} \frac{f_{os^2}}{f_{md}^2},$$

called *sum of terminal edges ratio* and *sum of square opposite sides ratio*, respectively. From Theorems 3, 4, 6 and 5 we obtain the following corollary immediately.

Corollary 4 *If T_1 and T_2 both exist and are both full, then $L_{T_1} \leq L_{T_2}$ if and only if $\rho_{ter}(V, \tau_1) \leq \rho_{ter}(V, \tau_2)$, or if and only if $\rho_{os^2r}(V, \tau_1) \leq \rho_{os^2r}(V, \tau_2)$.*

3.3 Quasi-indicators of optimal 4-point trees in Euclidean space

For the 4-point Steiner tree problem in the Euclidean plane E^2 , all 7 functions $\phi, -f_{ie}, -f_{md}, f_{te}, f_{os^2}, \rho_{ter}, \rho_{os^2r}$ are equivalent to the tree length $L_T(V, \tau)$ with respect to the variable (topology) τ and to the parameter (4-point sets) V in the domain (Euclidean plane) E^2 . However, when the 4-point Steiner tree problem moves from E^2 to E^3 , we must compare not only two but three trees T_1, T_2, T_3 . Suppose T_1 is optimal, then the equivalence can be restated as

$$L_{T_1} = \min_{i=1,2,3} L_{T_i} \text{ if and only if } f(\tau_1) = \min_{i=1,2,3} f(\tau_i), \quad (6)$$

where f stands for any function equivalent to L_T . It is easily seen that ϕ needs to be redefined in space (see below) and it turns out that the other 6 equivalent functions mentioned above are no longer equivalent to L_T in E^3 . Moreover, even in the plane the assumptions for the equivalence (both T_1, T_2 exist and are full) are very strong assumptions and do not always hold. For instance, as shown in Example 2, an SMT may be a degenerate tree of Type I, *i.e.*, non-full. These facts stimulated the authors of [30] to investigate whether efficient q-indicators exist for 4-point trees in Euclidean space E^3 . For randomly generated 4-point sets they tested f_{ie}, f_{md} and a function defined as

$$\rho_{osr} \stackrel{\text{def}}{=} \frac{f_{os}}{f_{md}},$$

which is called the *sum of opposite sides ratio* in this paper (but called *midpoint distance ratio* in [30]). That is,

$$\rho_{osr}(\tau_1) = \frac{|ab| + |cd|}{|m_{ab}m_{cd}|}, \quad \rho_{osr}(\tau_2) = \frac{|ad| + |bc|}{|m_{ad}m_{bc}|}, \quad \rho_{osr}(\tau_3) = \frac{|ac| + |bd|}{|m_{ac}m_{bd}|}.$$

Note that the definition of ρ_{osr} is similar to ρ_{os^2r} , however, whereas ρ_{os^2r} is an indicator for planar 4-point trees, ρ_{osr} is not. Here is a simple counterexample. Let a, b, c, d be 4 terminals in the plane, and the diagonals satisfy $ac \perp bd$. Suppose $|ao| = |do| = 1, |bo| = |co| = \varepsilon \ll 1$. Then it is easy to see that T_1 and T_2 both exist and are both full, by the diagonal angle rule $L_{T_1} = L_{T_2} = \sqrt{2}(1 + \sqrt{3})(1 + \varepsilon)/2$ but $f_{os}(V, \tau_1) = 2\sqrt{1 + \varepsilon^2} > \sqrt{2}(1 + \varepsilon) = f_{os}(V, \tau_2)$.

We can also compare ρ_{osr} with f_{ie} . For T_1 , if the midpoint line $m_{ab}m_{cd}$ is regarded as a representation of the mid-edge $s_{ab}s_{cd}$, then the corresponding representations of the four terminal edges should be $am_{ab}, bm_{ab}, cm_{cd}, dm_{cd}$, and the

sum of their lengths is $|ab| + |cd| = f_{os}(V, \tau_1)$. Hence, ρ_{osr} can be regarded as a deformation of f_{ie} . However, although f_{ie} is an indicator for planar 4-point trees, ρ_{osr} is not. Despite ρ_{osr} not being an indicator like ρ_{os^2r} and f_{ie} in the Euclidean plane, it was interesting to find by computational experiments [30] that ρ_{osr} is the best q-indicator in E^3 . Below we extend (both in the number of tested q-indicators and in the size of experimental data) the computational experiments on q-indicators to confirm this result.

Before describing the new computational experiments we need to first precisely redefine the diagonal angle ϕ for spatial 4-point trees. In E^3 we need to make comparisons between 3 pairs of trees $(T_1, T_2), (T_2, T_3), (T_1, T_3)$. Note that any two lines in space are parallel to a plane and we call the angle between their projections on the plane the *intersecting angle* of the two spatial lines. For a tetrahedron $abcd$ in space, there are 3 pairs of opposite sides and 6 midpoints. To compare T_1 and T_2 , look at the plane that is parallel to their diagonals ac and bd , i.e., the opposite sides of the third tree T_3 . Without loss of generality we may assume that the plane is the XY -plane, ac lies on the x -axis and the z -axis is the normal perpendicular to both ac and bd as shown in Fig. 7. It is easy to see that the four midpoints

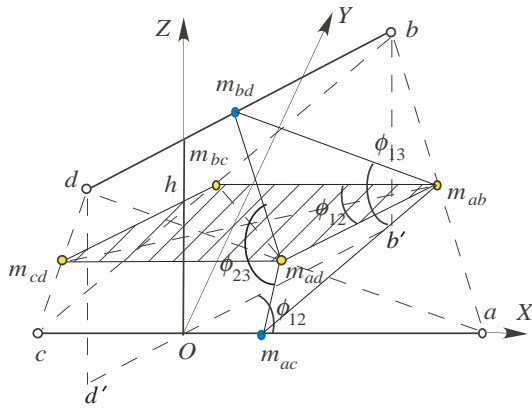


Fig. 7 Diagonal angles in space.

$m_{ab}, m_{cd}, m_{bc}, m_{ad}$ constitute a parallelogram lying on a plane that is parallel to the XY -plane.

Theorem 7 A tetrahedron $abcd$ in space has three parallelograms $m_{ab}m_{bc}m_{cd}m_{ad}$, $m_{ab}m_{bd}m_{cd}m_{ac}$, $m_{ad}m_{bd}m_{bc}m_{ac}$ that are parallel to three pairs of opposite sides $(ac, bd), (bc, ad), (ab, cd)$ respectively. Let $\phi_{12}, \phi_{13}, \phi_{23}$ be the three angles of the parallelograms as shown in Fig. 7. Then we have

(1) The three angles $\phi_{12}, \phi_{13}, \phi_{23}$ equal the intersecting angles of the three pairs of opposite sides $(ac, bd), (bc, ad), (ab, cd)$ respectively.

(2) For ϕ_{12}

$$\phi_{12} \leq 90^\circ \Leftrightarrow |ab|^2 + |cd|^2 \leq |ad|^2 + |bc|^2 \quad (7)$$

$$\phi_{12} \leq 90^\circ \Leftrightarrow |m_{ab}m_{cd}| \geq |m_{ad}m_{bc}| \quad (8)$$

$$\phi_{12} \leq 90^\circ \Leftrightarrow \frac{|ab|^2 + |cd|^2}{|m_{ab}m_{cd}|} \leq \frac{|ad|^2 + |bc|^2}{|m_{ad}m_{bc}|}. \quad (9)$$

Similar statements hold for ϕ_{13}, ϕ_{23} .

Proof We need only to prove the statements for ϕ_{12} and the tree pair (T_1, T_2) . The other statements hold similarly.

(1) The intersecting angle of ac and bd is the angle formed by ac and $b'd'$, the projection of bd on the XY -plane. Since $m_{ab}m_{bc} \parallel ac$, $m_{ab}m_{ad} \parallel bd$, $\phi_{12} = \angle m_{ad}m_{ab}m_{bc}$ is the intersecting angle of ac and bd as in the planar case (Fig. 6(b)).

(2) Let h be the distance between ac and bd . Since bb' is perpendicular to the parallelogram $m_{ab}m_{bc}m_{cd}m_{ad}$,

$$|ab|^2 = |ab'|^2 + |bb'|^2 = |ab'|^2 + h^2 = |ao|^2 + |b'o|^2 - 2\cos(\phi_{12}) + h^2.$$

We have similar equations for $|cd|^2, |ad|^2$ and $|bc|^2$ as in Equation (5). As a common term h^2 is added in each expression, Equivalence (7) holds. Using a similar argument and from the planar case, we have Equivalence (8) by Theorem 5 and Equivalence (9) by Corollary (4). ■

Since $\phi_{21} = 180^\circ - \phi_{12}, \phi_{31} = 180^\circ - \phi_{13}, \phi_{32} = 180^\circ - \phi_{23}$, in terms of $\phi_{12}, \phi_{13}, \phi_{23}$ the logical statement in Theorem 3 needs to be replaced with a combination of three logical statements for 4-point minimal trees in E^3

$$\begin{cases} (L_{T_1} \leq L_{T_2}) \& (L_{T_1} \leq L_{T_3}) & \text{if and only if } (\phi_{12} \leq 90^\circ) \& (\phi_{13} \leq 90^\circ), \text{ or} \\ (L_{T_2} \leq L_{T_1}) \& (L_{T_2} \leq L_{T_3}) & \text{if and only if } (\phi_{12} \geq 90^\circ) \& (\phi_{23} \leq 90^\circ), \text{ or} \\ (L_{T_3} \leq L_{T_1}) \& (L_{T_3} \leq L_{T_2}) & \text{if and only if } (\phi_{13} \geq 90^\circ) \& (\phi_{23} \geq 90^\circ). \end{cases} \quad (10)$$

This combination of logical statements is denoted by f_ϕ .

3.4 Computational experiments for q-indicators of optimal 4-point trees in space

In our new computational experiments, the 8 functions

$$f_{te}, f_{ie}, \rho_{ter}, f_{os^2}, f_{md}, \rho_{os^2r}, \rho_{osr}, f_\phi$$

defined in Subsections 3.2 and 3.3 are tested for randomized 4-point trees in space. The following are taken into consideration in the design of the test.

1. Two data models were used. In the *cube model*, all four points were uniformly distributed in a unit cube. That is, all coordinates of points were randomly generated with uniform distribution in the interval $[0, 1]$. In the *tetrahedron model*, all four points were uniformly distributed in a unit tetrahedron. Let

$$a^* = [0, 0, 0], b^* = [1, 0, 0], c^* = \left[\frac{1}{2}, \frac{\sqrt{3}}{2}, 0 \right], d^* = \left[\frac{1}{2}, \frac{\sqrt{3}}{6}, \frac{\sqrt{2}}{\sqrt{3}} \right].$$

Then $a^*b^*c^*d^*$ is a regular tetrahedron with unit side length which is half of a skew unit cube. A random point in this model was

$$r_1 \cdot \overrightarrow{a^*b^*} + r_2 \cdot \overrightarrow{a^*c^*} + r_3 \cdot \overrightarrow{a^*d^*},$$

where r_1, r_2, r_3 were random numbers with uniform distribution in the interval $[0, 1]$ satisfying $r_1 + r_2 + r_3 \leq 1$.

2. For the cube model, 8 groups of point sets were generated such that each group contained 1000 random 4-point sets. For the tetrahedron model three groups of point sets were generated in the same way. Because we found that there was no essential difference between the two models we did not test more groups for the tetrahedron model. Therefore, 11000 random 4-point sets were tested in total.

3. The average effectiveness indices of these q-indicators were first computed for each group, and then the averages (arithmetic means) over all groups μ were computed.

The test results are listed in Table 3 followed by an analysis of the test results.

Table 3 Effectiveness indices of q-indicators.

model-group	f_{te}	f_{ie}	ρ_{ter}	f_{os^2}	f_{md}	ρ_{os^2r}	ρ_{osr}	f_ϕ
cube-1	0.992	0.987	0.987	0.973	0.973	0.973	0.977	0.973
-2	0.989	0.983	0.983	0.969	0.969	0.969	0.978	0.969
-3	0.994	0.988	0.988	0.972	0.972	0.972	0.982	0.972
-4	0.991	0.99	0.99	0.976	0.976	0.976	0.991	0.976
-5	0.992	0.991	0.991	0.977	0.977	0.977	0.993	0.977
-6	0.993	0.993	0.993	0.978	0.978	0.978	0.994	0.978
-7	0.995	0.994	0.994	0.976	0.976	0.976	0.991	0.976
-8	0.997	0.996	0.996	0.969	0.969	0.969	0.989	0.969
tetrahedron-9	0.990	0.988	0.988	0.975	0.975	0.975	0.987	0.975
-10	0.990	0.985	0.986	0.972	0.972	0.972	0.984	0.972
-11	0.991	0.983	0.984	0.970	0.970	0.970	0.982	0.970
(arithmetic mean) μ	0.9922	0.9889	0.9891	0.9734	0.9734	0.9734	0.9862	0.9734
(standard deviation) σ	0.0004	0.0018	0.0015	0.0010	0.0010	0.0010	0.0013	0.0010

1. First, it is clear that the effectiveness indices of f_{md} , f_{os^2} , ρ_{os^2r} and f_ϕ in each group are equal. Note that f_{md} , f_{os^2} , ρ_{os^2r} are equivalent to f_ϕ in the Euclidean plane (see Theorems 5, 6 and Corollary 4) and that their definitions are independent of the dimension, hence, they are still equivalent to f_ϕ given the logical statement about f_ϕ in Theorem 3 is extended to the logical statements 10. Hence, the effectiveness indices of the 4 quasi-indicators in space can be regarded as a numerical validation of Theorem 7.

2. The standard deviation is less than 0.18% for all tested q-indicators, in particular the standard deviation of f_{te} is only 0.04%. Hence, we can say that the tests are stable and reliable.

3. f_{te} has the highest effectiveness index followed by f_{ie} and ρ_{ter} . However, as we have to first find the Steiner points to compute these functions, their computational complexities (in terms of the number of arithmetic operations) are as high as the original 4-point tree problem and cannot be used as q-indicators in practice. Consequently, ρ_{osr} is the best practical q-indicator.

4. As argued above ρ_{osr} is not equivalent to L_T in the plane although ρ_{os^2r} is. However, the experiments show that in space ρ_{osr} has a higher effectiveness index than ρ_{os^2r} . This is surprising and difficult to explain.

5. The point sets are randomly generated and of all the 3×11000 trees about 64% trees are degenerate trees of Type II. We checked that none of these trees

with degree-4 Steiner points was optimal. This can be regarded as a numerical validation of Theorem 1.

Remark 4 Note that the coordinates of the midpoint of a line are a linear combination of the coordinates of the two endpoints of the line and ρ_{osr} depends only on the coordinates of the 4 terminals. Hence, we can reasonably claim that the opposite sides ratio is the best q-indicator for 4-point trees in any Euclidean d -space for $d \geq 3$.

Remark 5 A probability vector in Euclidean d -space is a vector whose sum of components is one. Hence, a probability vector space is a $d-1$ dimensional subspace in E^d . Hence, the Steiner tree problem in a probability vector space is no longer an unconstrained optimization problem but has a simple linear constraint: $\sum_{k=1}^d s_k = 1$ for any Steiner point $s = [s_1, s_2, \dots, s_d]$. Because this constraint does not affect ρ_{ter} we believe that ρ_{ter} is still the best q-indicator for 4-point trees in a probability vector space.

4 An Application of Quasi-indicators in S4pT Technique

As stated in the first section, Introduction and Motivation, the simplest tree rearrangement technique, called the *Swapping 4-point Topology/Tree (S4pT)* technique, repeatedly replaces a non-optimal 4-point subtree with the optimal 4-point tree on the same 4 nodes. In the last section we pointed out that computing the length of a 4-point tree is algebraically unsolvable and time-consuming compared with computing the q-indicator ρ_{osr} which is linear in the coordinates of the 4 nodes. We show the details of the S4pT technique in an example, next, and show how ρ_{osr} is used in the S4pT technique.

Example

Let Q be the set of 14 DNA sequences studied in [7] and [26], then the distributions of nucleotides A, G, C, T in these species comprise a set V of 14 probability vectors as shown in the following table:

Table 4 A set of probability vectors built from 14 DNA sequences.

	A	G	C	T
Marsupial Mole t_1	0.2350,	0.3026,	0.2368,	0.2256]
Wombat t_2	0.2199,	0.3158,	0.2425,	0.2218]
Rodent t_3	0.2105,	0.3139,	0.2462,	0.2293]
Elephant Shrew t_4	0.2030,	0.2989,	0.2707,	0.2274]
Elephant t_5	0.1917,	0.3120,	0.3064,	0.1898]
Whale t_6	0.1917,	0.3158,	0.2895,	0.2030]
Dolphin t_7	0.1974,	0.3101,	0.2951,	0.1974]
Pig t_8	0.2049,	0.3064,	0.3008,	0.1880]
Horse t_9	0.1880,	0.3196,	0.3045,	0.1880]
Bat t_{10}	0.1955,	0.3101,	0.2801,	0.2143]
Insectivore t_{11}	0.1861,	0.3045,	0.3026,	0.2068]
Human t_{12}	0.1842,	0.3365,	0.3026,	0.1767]
Sea Cow t_{13}	0.2049,	0.2989,	0.2876,	0.2087]
Hyrax t_{14}	0.2105,	0.3008,	0.2801,	0.2087]

The phylogenetic tree \mathcal{T} on Q , regardless of the method used in its construction, is an optimal full Steiner tree. Depending on the method, the data used for the construction is different. For example, the data used is the distances between the sequences if the construction method is the Distance Matrix method. In particular, if the distances are computed without a statistical model, then the distance is the Hamming distance, i.e. the l_1 metric. However, we can look at the problem using the Euclidean distance, i.e. l_2 distance. As mentioned in Section 1, since the size of V is greater than 10, branch-and-bound algorithms cannot solve the ESMT problem on V in a reasonable time. We now describe how to find an S4pT-optimal ESMT on the set V of 14 terminals t_i , $1 \leq i \leq 14$. Suppose the initial tree T^0 on V generated by a certain method is as shown in Fig. 8(a), the planar representation of the tree, and the edge lengths can be clearly seen in its ‘linearized’ image — Fig. 8(b). There are 11 interior edges in T^0 , consequently there are 11 4-point subtrees in T^0 . In Table 5 all sets of the endpoints (a, b, c, d) , the q-indicators ρ_{osr} of the current 4-point subtrees T_{ab-cd} in T^0 , and their associated other two subtrees are listed.

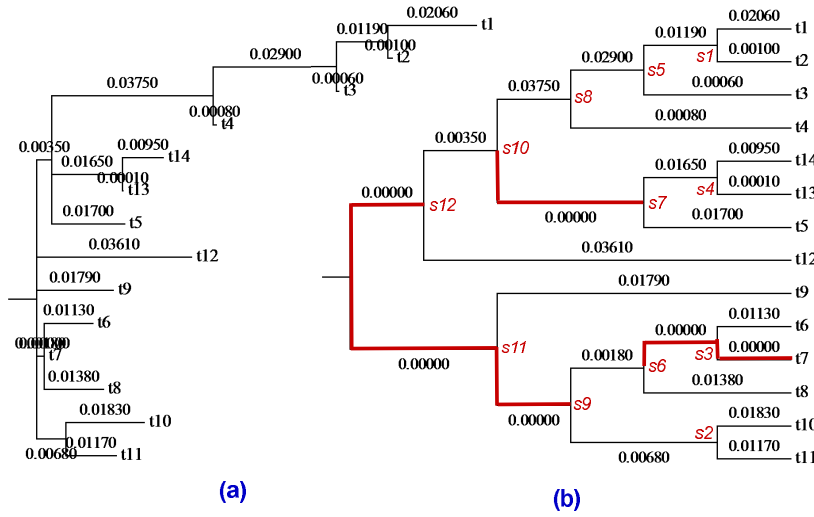


Fig. 8 Initial tree T^0 on 14 probability vectors.

Note that in each row in the table the minimum ρ_{osr} is in boldface and there are three 4-point subtrees in T^0 whose ρ_{osr} are not optimal. For example, the 4-point tree $T_{t_6 t_7 t_8 s_9}$ with mid-edge $s_3 s_6$ is non-optimal, and swapping the 4-point subtree $T_{t_6 t_7 t_8 s_9}$ with $T_{t_6 s_9 t_7 t_8}$ we can compose a new topology and construct a new tree T^1 on V that is shorter than T^0 .

There are some remarks to note on the application of ρ_{osr} to T^0 :

1. From Fig. 8(b) we can see that the edge lengths of the 4 mid-edges $s_3 s_6$, $s_7 s_{10}$, $s_9 s_{11}$ and $s_{11} s_{12}$ are all zero. These cases are all different and discussed in the points below:

Table 5 Quasi-indicators ρ_{osr} of 4-point subtrees in the initial tree T^0 .

mid-edge	endpoints (a, b, c, d)	$\rho_{\text{osr}}(T_{ab-cd})$	$\rho_{\text{osr}}(T_{ac-bd})$	$\rho_{\text{osr}}(T_{ad-bc})$
s_1s_5	(t_1, t_2, t_3, s_8)	1.7799	3.4611	3.4651
s_2s_9	$(t_{10}, t_{11}, s_6, s_{11})$	1.8658	2.8935	3.1202
s_3s_6	(t_6, t_7, t_8, s_9)	2.2287	9.3285	1.8504
s_4s_7	$(t_{13}, t_{14}, s_{10}, t_5)$	0.9738	8.9338	4.4759
s_5s_8	(s_1, t_3, t_4, s_{10})	1.1355	4.9441	4.4224
s_6s_9	(s_3, t_8, s_2, s_{11})	1.8631	2.3446	4.7653
s_7s_{10}	(t_5, s_4, s_{12}, s_8)	3.7907	1.0891	11.0271
s_8s_{10}	(s_5, t_4, s_{12}, s_7)	0.6873	6.6774	6.5972
s_9s_{11}	(t_9, s_{12}, s_2, s_6)	2.2983	2.7651	2.3624
$s_{10}s_{12}$	$(t_{12}, s_{11}, s_7, s_8)$	1.9332	2.3128	6.1184
$s_{11}s_{12}$	$(t_9, s_9, t_{12}, s_{10})$	6.4501	0.8584	4.8072

2. The 4-point tree $T_{t_5s_4-s_{12}s_8}$ is of Type II degenerate tree with a degree-4 S-point s_3s_6 . By Theorem 1 it is not optimal.
3. The 4-point tree $T_{t_6t_7-t_8s_9}$ is not optimal but of Type I since $s_3 = t_7$, i.e. the S-point s_3 collapses into a terminal t_7 .
4. From Table 5 we can see that the q-indicator ρ_{osr} detects the non-optimal mid-edge $s_{11}s_{12}$ but does not detect $(t_9, s_9, t_{12}, s_{10})$ the non-optimal mid-edge s_9s_{11} . The reason is the mid-edges s_9s_{11} and $s_{11}s_{12}$ are adjacent and three S-points s_9, s_{11}, s_{12} collapse together forming a degree-5 S-point. Hence, this case is not indicated by the q-indicator ρ_{osr} because it is not within the scope of q-indicators for 4-point trees.
5. An important fact is that the two 4-point trees $T_{t_6t_7-t_8s_9}$ and $T_{s_{12}s_4-t_5t_8}$ are disjoint. Hence we can apply the swapping technique to the two 4-point subtrees simultaneously which can speed up the convergency of tree length. In fact, we did so in the reconstruction of the S-tree from T^0 to T^1 .

After 10 swaps we obtained an S4pT-optimal ESMT on V . The lengths of the trees T^0, T^1, \dots form a decreasing curve as shown in Fig. 9.

The S4pT-optimal ESMT T^{10} is shown in Fig. 10. Table 6 lists the q-indicators ρ_{osr} of the current 4-point subtrees T_{ab-cd} in T^{10} , and their associated other two subtrees. From the table it is clearly shown that all 4-point subtrees in T^{10} are optimal.

Table 6 Quasi-indicators ρ_{osr} of 4-point subtrees in the S4pT-optimal tree T^{10} .

mid-edge	endpoints (a, b, c, d)	$\rho_{\text{osr}}(T_{ab-cd})$	$\rho_{\text{osr}}(T_{ac-bd})$	$\rho_{\text{osr}}(T_{ad-bc})$
s_1s_5	(t_1, t_2, t_3, s_8)	1.7897	3.4626	3.4666
s_2s_3	(s_4, t_{10}, t_6, s_9)	1.2444	9.6024	3.4121
$s_{12}s_6$	(s_7, t_8, s_9, t_7)	1.5000	3.1985	5.2415
s_4s_{10}	$(t_{13}, s_2, s_8, t_{14})$	2.2598	4.1401	2.2961
s_5s_8	(s_1, t_3, t_4, s_{10})	0.8791	8.8972	5.2739
s_9s_3	(s_2, t_6, s_6, t_{11})	2.1630	3.2996	3.0778
s_2s_4	$(t_{10}, s_3, s_{10}, t_{13})$	1.4013	3.1242	7.8159
s_8s_{10}	(s_5, t_4, s_4, t_{14})	1.2051	5.2194	3.6937
s_9s_6	$(s_{12}, t_7, s_3, t_{11})$	1.8105	4.3280	2.8122
$s_{12}s_7$	(t_8, s_6, s_{11}, t_5)	1.4188	5.6147	3.5784
$s_{11}s_7$	$(t_9, t_{12}, t_5, s_{12})$	1.4284	4.4996	3.4101

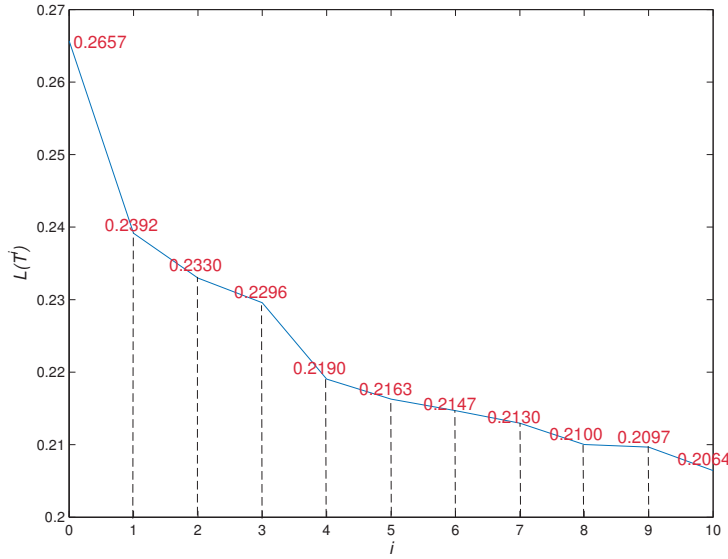


Fig. 9 Decreasing curve $L(T^i)$ $0 \leq i \leq 10$.

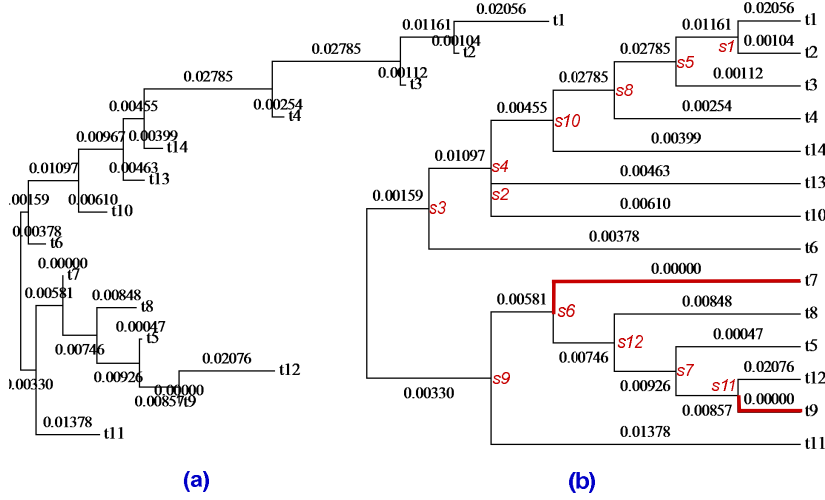


Fig. 10 S4pT-optimal tree T^{10} on 14 probability vectors.

Remark 6 From the planar tree T^{10} (Fig. 10) we can see there are two terminal edges t_7s_6 and t_9s_{11} of zero length, that is, the two terminals t_7, t_9 collapse into their adjacent S-points and the 4-point subtrees are of Degenerate Type I. As argued in Section 2, 4-point tree of Degenerate Type I can be optimal. Because the probability vectors in 4d-space have only three independent variables. Hence, taking the first three components as Cartesian coordinates, T^{10} can be represented

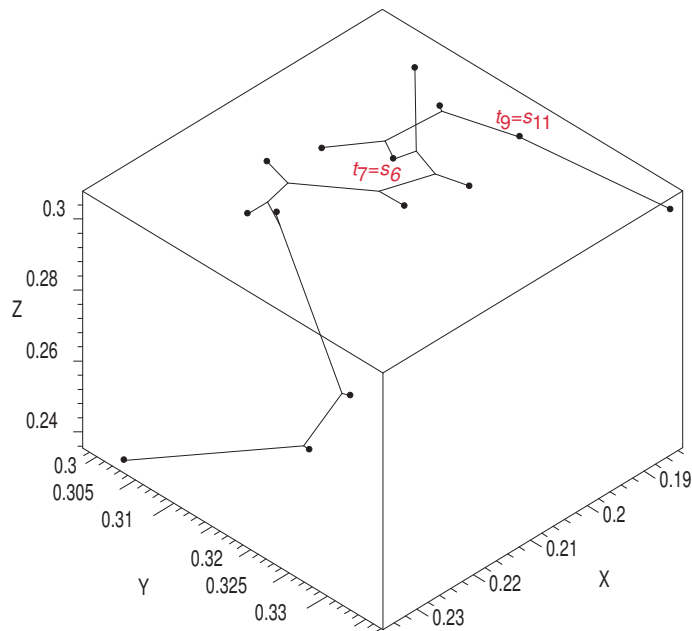


Fig. 11 Decreasing curve $L(T^i)$ $0 \leq i \leq 10$.

as a tree in Euclidean 3d-space. From the spacial representation of T^{10} (Fig. 10) it is clearly seen that the two terminals t_7, t_9 cut T^{10} into 3 full subtrees. The angles at $t_7 = s_6, t_9 = s_{11}$ are greater than 120° while all angles at other S-points are equal to 120° .

References

1. Badri, T. N.: Steiner minimal trees in three-dimensional Euclidean space. PhD Dissertation, University of Massachusetts, <http://scholarworks.umass.edu/dissertations/AAI3039336>, 2002.
2. Brazil, M., Grossmann, P. A., Lee, D. H., Rubinstein, J. H., Thomas, D. A., Wormald, N. C.: Decline design in underground mines using constrained path optimisation. Mining Technology: Trans. of the Institution of Mining and Metallurgy, Section A, **117**, 93–99 (2008)
3. Brazil, M., Thomas, D. A., Nielsen, B. K., Winter, P., Wulff-Nilsen, C., Zachariasen, M.: A novel approach to phylogenetic trees: d-dimensional geometric Steiner trees. Networks, **53**(2), 104–11 (2009)
4. Du, D-Z., Hu, X.: Steiner Tree Problems in Computer Communication Networks. World Scientific (2008)
5. Du, X., Du, D-Z., Gao, B., Que, L.: A simple proof for a result of Ollenschaw on Steiner trees, In: Du, D-Z., Sun, J. (eds.) Advances in Optimization and Approximation, pp. 68–71, (1994)
6. Fampa, M., Anstreicher, K. M.: An improved algorithm for computing Steiner minimal trees in Euclidean d -pace. Discrete Optimization, **5**(2), 530–540 (2008)
7. Ewens, W.J., Grant, G.R.: *Statistical Methods in Bioinformatics: An Introduction*, Springer, USA, 2005.
8. Felsenstein, J.: Inferring Phylogenies, Sinauer Associates Inc., 2004
9. Garey, M. R., Graham, R. L., Johnson, D. S.: The complexity of computing Steiner minimal trees, SIAM J. Appl. Math., **32**, 835–859 (1977)

10. Grant, M., Boyd, S.: CVX: Matlab software for disciplined convex programming, version 1.21. `./../cvx`, April 2011.
11. Gilbert, E. N., Pollak, H. O.: Steiner minimal trees, *SIAM J. Appl. Math.* **16**, 1–29 (1968)
12. Hwang, F. K., Richards, D. S., Winter, P.: The Steiner Tree Problem, *Annals of Discrete Mathematics* (53), Elsevier Science Publishers B.V., Amsterdam (1992)
13. Karp, R. M.: Ruducibility among combinatorial problems, In: Miller, R. E., Thatcher, J. W. (eds.) *Complexity of Computer Computations*, Plenum Press, New York, pp. 85–103 (1972)
14. Melzak, Z. A.: On the problem of Steiner, *Canad. Math. Bull.*, **4**, 143–148 (1961)
15. R. P. Mondaini: Euclidean full Steiner trees and the modelling of biomolecular structures, *Proceedings: Biomat 2006*, pp. 247–258.
16. Ollerenshaw, K.: Minimum networks linking four points in a plane, *Inst. Math. Appl.* **15**, 208–211 (1978)
17. Pollak, H. O.: Some remarks on the Steiner problem, *J. Comb. Theory, Ser. A*, **24**, 278–295 (1978)
18. Rubinstein, J. H., Thomas, D. A.: A variational approach to the Steiner network problem, *Annals of Operations Research*, **33**, 481–499 (1991)
19. Rubinstein, J. H., Thomas, D. A., Weng, J. F.: Minimum networks for four points in space, *Geometriae Dedicata*, **93**, 57–70 (2002)
20. Smith, W. D.: How to find Steiner minimal trees in Euclidean d -space. *Algorithmica*, **7**, 137–177 (1992)
21. Smith, J. M., Jang, Y., Kim, M. K.: Steiner minimal trees, twist angles, and the protein folding problem, *PROTEINS: Structure, Function, and Bioinformatics* **66**, 889–902 (2007)
22. Takahashi, K., Nei, M.: Efficiencies of Fast Algorithms of Phylogenetic Inference Under the Criteria of Maximum Parsimony, Minimum Evolution, and Maximum Likelihood When a Large Number of Sequences Are Used. *Mol. Biol. Evol.* **17**, 1251–1258 (2000)
23. Tria, F., Caglioti, E., Loreto, V., Pagnani, A.: A Stochastic Local Search Algorithm for Distance-Based Phylogeny Reconstruction. *Mol. Biol. Evol.* **27**, 2587–2595 (2010)
24. Weng, J. F.: Steiner Trees, Coordinate Systems and NP-hardness. In: Du, D.-Z., Smith, J. M., Rubinstein, J. H. (eds.) *Advances in Steiner Trees*, KluwerAcademic Publishers, pp. 63–80 (2000)
25. Weng, J. F.: Identifying Steiner minimal trees on four points in space. *Discrete Mathematics, Algorithms and Applications*, **1**, 401–411 (2009)
26. Weng, J. F., Mareels I., Thomas, D. A.: Probability Steiner trees and maximum parsimony in phylogenetic analysis. *Journal of Mathematical Biology*, **64**(7), 1225–1251 (2012)
27. Weng, J. F., Thomas, A., Mareels, I.: Maximum Parsimony, substitution Model and probability phylogenetic trees. *Journal of Computational Biology*, **18**, 67–80 (2011)
28. Weng, J. F.: Variational approach and Steiner minimal trees on four points, *Discrete Math.* **132**, 349–362 (1994)
29. Weng, J. F.: Generalized Steiner problem and hexagonal coordinate system (in Chinese), *Acta Math. Appl. Sinica*, **8**, 383–397 (1985)
30. Weng, J. F., Smith, J. M., Brazil, M., Thomas, D. A.: Equivalence, indicators, quasi-indicators and optimal Steiner topologies on four points in space, *Fundamenta Informaticae*, **84**, 135–149 (2008)



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Thomas, DA;Weng, JF

Title:

Euclidean Steiner trees optimal with respect to swapping 4-point subtrees

Date:

2014-04

Citation:

Thomas, D. A. & Weng, J. F. (2014). Euclidean Steiner trees optimal with respect to swapping 4-point subtrees. OPTIMIZATION LETTERS, 8 (4), pp.1337-1359. <https://doi.org/10.1007/s11590-013-0660-3>.

Persistent Link:

<http://hdl.handle.net/11343/283075>