Insight into the computation of Steiner minimal trees in Euclidean space of general dimension

Marcia Fampa
Universidade Federal do Rio de Janeiro
email: fampa@cos.ufrj.br

Abstract

We present well known properties related to the topology of Steiner minimal trees and to the geometric position of Steiner points, and investigate their application in the main exact algorithms that have been proposed for the Euclidean Steiner problem. We discuss the difficulty in the application of properties that were very successfully applied to solve the problem in the plane, when the dimension of the space increases, and point out that the application of some geometric conditions for Steiner points is hindered when the well known implicit enumeration scheme proposed by Smith in 1992 is considered. Finally, we mention mathematical-optimization methods as a direction to explore in the search for good formulations of inequalities that would allow the application of these restrictive geometric conditions.

Keywords: Steiner minimal tree; Euclidean Steiner problem; MINLP

1 Introduction

On the d-dimensional Euclidean Steiner Problem (ESP), we are given a set $\mathcal{N} := \{t_1, \ldots, t_n\}$ of n points in $\mathbb{R}^d$ and look for the shortest network connecting them. The given points are called terminals, and the network that solves the problem consists of a tree, known as the Steiner Minimal Tree (SMT). The nodes of an SMT are the terminals and possibly additional points, called Steiner points, and the length of the tree is measured by the sum of the Euclidean distances between adjacent nodes. If no Steiner points were allowed in the solution, the ESP would coincide with the problem of finding the minimum spanning tree of $\mathcal{N}$, therefore, the possibility of adding the Steiner points to shorten the tree is what makes the ESP a big challenge for optimizers, particularly in dimensions greater than 2.

The ESP was first studied in the plane ($d = 2$), and was proved to be NP-hard [15]. In [27], the authors show that the problem is NP-hard, even in the particular case where the terminals lie on two parallel lines in the plane. Nevertheless, when $d = 2$, it was possible to solve the problem to optimality for a very large number of terminals in reasonable time, using the well known algorithm GeoSteiner [35, 36, 33, 34]. The generalization of the problem for higher dimensions was first proposed by Bopp in 1879 [2], and instances of the problem with more than 18 terminals were never solved to proven optimality in dimensions greater than 2. The success of the GeoSteiner algorithm is greatly related to the application of well known geometric properties of SMTs, which cannot be so efficiently applied in higher dimensions. Most algorithms presented in the literature to solve the ESP in dimensions higher than 2, are based on an implicit enumeration scheme proposed by Smith in 1992 [29]. Some improvements have been proposed to Simth's algorithm in an attempt to increase the number of nodes pruned during the enumeration [10, 31, 14]. In [31], Laarhoven and Anstreicher observed that Smith's algorithm does not apply any geometric property of SMTs to solve the ESP, and proposed some modifications to the algorithm. However, due to the enumeration scheme developed by Smith, the application of the properties is very limited. For an overview of exact algorithms for the d-dimensional ESP, we refer the reader to [11].

The design of VLSI circuits has been presented as an application for the ESP in the plane [38, 19, 21]. When $d > 2$, applications on phylogenetics and on the study of molecular structure have also been discussed [3, 5, 4, 28, 30]. However, the mathematical challenge that the problem represents, and how its solution can contribute to the solution of other difficult optimization problems with similar geometric characterization, have certainly been an important motivation for the interest of researches on the subject as well.

We present in this paper, some well known properties of SMTs and discuss how they were applied in the main exact algorithms for the ESP. We investigate how these properties were successfully applied to solve the problem in the plane and identify the reasons that prevent their application in higher dimensions. Although the increase of the dimension of the Euclidean space from 2 to 3 already hinders the application of some procedures in GeoSteiner, some of the geometric conditions on Steiner points cannot be applied in the Smith algorithm because of the specific enumeration scheme applied. Our goal in this paper is to present some insight into the difficulties in the application of well known properties to the solution of the ESP in dimensions higher than 2, and to point directions that would allow their application with the use of mathematical-optimization methods.

In Section 2, we start the paper with an overview of the interesting historical background of the ESP, which relates to a simplified version of the problem studied by famous mathematicians in the 17^{th} century. In Section 3, we discuss how the work of these mathematicians made it possible to identify several properties of SMTs, which are the core of the algorithms used today to solve the ESP. In Section 4, we present the main algorithms from the literature for the ESP. We discuss how the application of the well known properties of SMTs to the solution of the ESP, was so effective in the plane, and also talk about the difficulty of applying them to Smith's enumeration scheme. In Section 5, we discuss about what have been done concerning the formulation of the problem as a mixed-integer nonlinear program, and the application of branch-and-bound algorithms to solve it. In Section 6, we indicate how some of the geometric conditions of SMTs could be used in dimensions higher than 2 in the execution of branch-and-cut algorithms, considering mathematical programming formulations of the ESP from the literature. In Section 7, we conclude with some final remarks.

2 The historical background of the ESP

The ESP has an interesting history that traces back to an ancient problem that consists of a challenge proposed by Pierre de Fermat in the 17^{th} century. Given three points a, b and c, in the plane, the problem was to find a fourth point such that the sum of its distance to the three given points was at minimum. In 1640, Evangelista Torricelli constructed a geometric solution for the problem, where he showed that the three circles circumscribing the equilateral triangles constructed on the sides of and outside the triangle abc intersect in the point that is sought, known as the Torricelli point. Seven years later, Bonaventura Cavalieri also showed that the line segments connecting the three given points to the Torricelli point make 120 degrees with each other. In 1750, Thomas Simpson proved another important geometric result in his book *Doctrine and Application of Fluxions*. He considered again the three equilateral triangles used in the solution of Torricelli, and showed that the three lines joining the outside vertices of the triangles with the opposite vertices of triangle abc intersect in the Torricelli point. These lines are known as the Simpson lines. The solutions of Torricelli and Cavalieri, and the Simpson lines, are depicted in Figure 1.

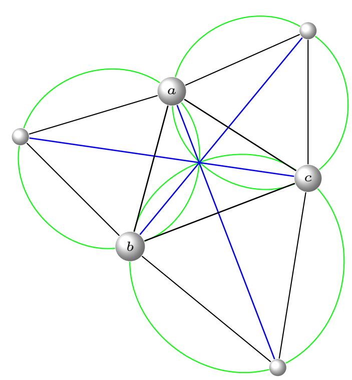


Figure 1: Solutions of Torricelli and Cavalieri to Fermat's challenge.

It was only about two centuries after Fermat proposed his challenge, that the problem was completely solved. In 1834, Franz Heinen showed that if the triangle abc has one angle greater than or equal to 120 degrees, then the minimizing point that solves Fermat problem is the vertex of the obtuse angle. Heinen also proved that the lengths of the Simpson lines are all equal, and are equivalent to the sum of the distances from Torricelli point to the points a, b, and c.

In 1941, the following two generalizations of the Fermat problem were presented in the popular book *What is Mathematics?*, published by Richard Courant and Herbert Robbins [7].

- Find a point such that the sum of the distances from the point to n given points is minimal.
- Find a shortest network interconnecting n given points in the Euclidean plane.

Among others, a famous mathematician who investigated these problems, was Jakob Steiner. For this reason, they ended up presented in the book as the Steiner problems. Nowadays, the first problem is usually referred to as the Fermat problem and the second remained known as the Steiner problem. The term Euclidean was introduced later to distinguish the versions of the problem, where different metrics are used to measure the length of the network [1, 38]. For more details on the history of the problem, we refer the reader to the comprehensive book *The Steiner Tree Problem* by Hwang, Richads and Winter [17].

3 Definitions and well known properties of SMTs

The solution of the 17^{th} century challenge of Fermat, made it possible to identify several properties of SMTs, which were very important to the development of algorithms for the ESP. A straightforward conclusion from Cavalieri's result is that no two edges can meet at a point in an SMT with an angle less than 120 degrees, as in this case the tree could be shortened by the inclusion of a Steiner point connected to their three incident points. The position of this Steiner point would be determined by the position of the three other points, as shown in Figure 1. As a consequence, we have that no node in an SMT can have degree more than 3. It is also clear that Steiner points cannot have degree 1 or 2, as in these cases the tree would be shortened if they were deleted from the tree together with their incident edges, and the two neighbors of the Steiner point were connected to each other, if its degree was 2. Finally, it is easy to verify that in an SMT, a Steiner point lies in the plane defined by its three neighbors, otherwise the tree could also be shortened by its projection onto this plane. The well known *degree condition* and *angle condition* for SMTs refer to these observations. These conditions and other results presented in this section can be found in [17]. They are valid for SMTs not only in dimension 2, but also in higher dimensions.

Theorem 1. [Degree condition] In an SMT, a terminal has degree between 1 and 3, and a Steiner point has degree equal to 3.

Based on the *degree condition*, we define a Steiner tree.

Definition 2. [Steiner Tree (ST)] A Steiner Tree (ST) for n terminals is a tree connecting all the terminals and possibly Steiner points, where the degrees of all nodes satisfy the *degree condition*.

Clearly, an SMT is an ST. Also, as a consequence of the *degree condition*, we can determine the maximum number of Steiner points in an SMT and define a full Steiner tree.

Theorem 3. In an SMT for n terminals, there exist at most $n-2$ Steiner points.

Proof. Let us consider that the SMT has k Steiner points, and therefore $n+k-1$ edges. Since each Steiner point has 3 incident edges and each terminal has at least 1, the total number of edges must be at least $(3k+n)/2$. Then $n+k-1 \geq (3k+n)/2 \quad \Rightarrow \quad k \leq n-2$. □

Definition 4. [Full Steiner Tree (FST)] A Full Steiner Tree (FST) for n terminals is an ST, with $n-2$ Steiner points.

The following corollaries show important results that can be directly obtained from the proof of Theorem 3.

Corollary 5. In an FST, all terminals have degree equal to 1.

Corollary 6. In an FST for $n \geq 4$ terminals, there exist at least 2 Steiner points which are each adjacent to 2 terminals.

We note that the results above are related only to the connections between the nodes of STs, still disregarding the position of the Steiner points. These connections define the topology of the tree, as stated below.

Definition 7. [Steiner topology and full Steiner topology]

- A Steiner topology $\mathcal{T}$ of an ST T for $\mathcal{N}$ is a specification of the interconnections in T, disregarding the positions of its nodes.
- The topology of an FST is said to be a full Steiner topology.

Definition 8. [Relative Minimal Tree (RMT)] The shortest ST for a given Steiner topology $\mathcal{T}$, is called a Relatively Minimal Tree (RMT) for $\mathcal{T}$.

Theorem 9. The RMT for a given Steiner topology $\mathcal{T}$ always exists and is unique. Its Steiner points may overlap with terminals and with other Steiner points because of zero length edges.

Definition 10. [Degenerated topology and degenerated full Steiner topology]

- If Steiner points overlap with terminals or with other Steiner points in the RMT for a given topology $\mathcal{T}$, the topology $\mathcal{T}'$ of the resulting RMT is said to be degenerated from $\mathcal{T}$.

$\mathcal{T}'$ is obtained from $\mathcal{T}$, by eliminating all the edges with zero length on the RMT, and also eliminating one incident Steiner point to each one of these edges. Considering (s, v) as a zero length edge and s as the eliminated Steiner point, we also eliminate from $\mathcal{T}$, the other two edges incident to s, and finally connect v to the two other neighbors of s.

- If $\mathcal{T}'$ is degenerated from a full Steiner topology $\mathcal{T}$, then $\mathcal{T}'$ is said to be a degenerated full Steiner topology.

Now we present well known properties of SMTs that are also related to the geometric position of the nodes, starting with the *angle condition* mentioned above. These properties have been used very effectively in the development of algorithms for the ESP in the plane.

Theorem 11. **[Angle condition]** In an SMT, a Steiner point and its 3 adjacent nodes lie in a plane, and the angles between the edges connecting the Steiner point to its adjacent nodes are all 120 degrees.

From the *angle condition*, we also have the following result.

Corollary 12. In an SMT, there are no crossing edges.

The following straightforward result determines a bound on the length of an edge incident to a terminal.

Definition 13. **[Shortest distance from terminal t_i (η_i)]** The shortest distance from terminal t_i, denoted by η_i, is the distance from t_i to the nearest other terminal.

Lemma 14. In an SMT, the length of the edge incident to terminal t_i is at most η_i.

The forbidden connection determined in Lemma 14 is depicted in Figure 2.

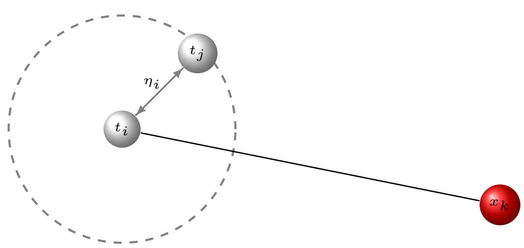


Figure 2: Forbidden connection between t_i and x_k (Lemma 14).

The lune and the bottleneck properties presented below, are also a useful results to restrict the position of the Steiner points in SMTs.

Lemma 15. **[Lune property]** Let u and v be two adjacent nodes of an SMT. Let $L(u, v)$ be the region consisting of all points p satisfying

$$\|p - u\| < \|v - u\| \quad \text{and} \quad \|p - v\| < \|u - v\|.$$

$L(u, v)$ is the lune-shaped intersection of the circles of radius $\|u - v\|$, centered on u and v. No other node of the SMT can lie in $L(u, v)$.

Corollary 16. If there are terminals t_i, t_j, t_k such that t_k is in the interior of $L(t_i, t_j)$, then t_i and t_j cannot be connected in an SMT.

Definition 17. [**Bottleneck distance** (β_{ij})] The bottleneck distance corresponding to terminals t_i and t_j of $\mathcal{N}$, denoted by β_{ij}, is the length of the longest edge on the path connecting t_i and t_j in the minimum spanning tree of $\mathcal{N}$.

Lemma 18. [**Bottleneck property**] No edge on a path between a pair of terminals t_i and t_j in an SMT can be longer than the bottleneck distance β_{ij}.

4 Exact algorithms

4.1 ESP in the plane

The most successful algorithm for the ESP in the plane is the GeoSteiner algorithm [35, 36, 33, 34]. It has solved instances of the problem with thousands of terminals and has its foundations on the first finite algorithm for the ESP, proposed by Melzak in 1961 [25] and simplified later by Cockayne [6], who made the following important remark, illustrated in Figure 3.

Remark 19. An SMT can be uniquely partitioned into edge disjoint trees, each with a with full Steiner topology.

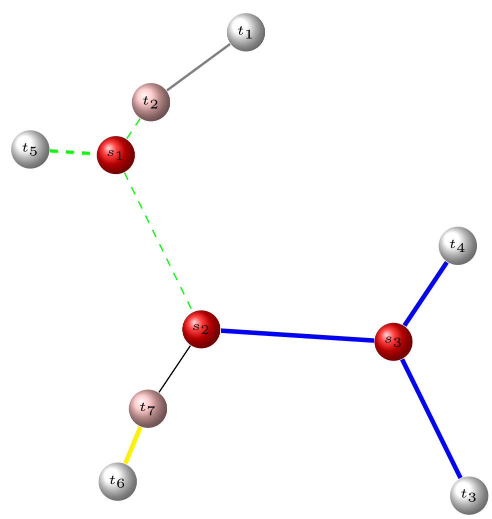


Figure 3: Decomposition of an SMT into FSTs.

Using the observation above, Melzak's algorithm computes an SMT for a set of terminals $\mathcal{N}$, by computing and saving RMTs for subsets of $\mathcal{N}$, if they are FSTs. In the end the stored FSTs are concatenated to generate STs for $\mathcal{N}$, and the shortest one is given as an SMT. The core of Melzak's algorithm is the so-called Melzak FST procedure. For a given full Steiner topology for a subset of $\mathcal{N}$, the Melzak FST procedure, either computes the RMT with the full Steiner topology, or returns "0", in case such RMT does not exist.

The Melzak FST procedure is based on the geometric solutions to the challenge of Fermat that were proposed by Torricelli, Cavalieri and Simpson and are illustrated in Figure 1. It consists of two phases, the merging phase, where the number of terminals in a given full Steiner topology $\mathcal{T}$ for a subset of the terminals $\bar{\mathcal{N}} \subseteq \mathcal{N}$ is reduced to two, and the reconstruction phase, where the FST is constructed by iteratively fixing the position of the Steiner points.

Considering that in any full Steiner topology (assuming $n \geq 4$), there exists one Steiner point that is connected to two terminals (Corollary 6), the merging procedure initiates with the selection of two terminals t_i and t_j that are connected to the same Steiner point. Denote this Steiner point by s_k, and the third neighbor of s_k, by v. Let now e_{ij} be the third vertex of the equilateral triangle with other vertices t_i and t_j. The point e_{ij} is called the equilateral point related to t_i and t_j. In the plane, there are only two possible locations for e_{ij}, and suppose one of them was chosen. Once the position of the equilateral point e_{ij} is selected, this point is connected to v, and t_i, t_j, s_k, and their incident edges are eliminated from $\mathcal{T}$. The equilateral point e_{ij} is considered as a terminal on the resulting topology, which is therefore, a full Steiner topology with one less terminal and one less Steiner point. The procedure is repeated with the resulting topology, selecting again a Steiner point connected to two terminals. The merging phase ends when the topology $\mathcal{T}$ has been transformed into a full Steiner topology with two terminals and no Steiner points. In the end, every terminal in $\bar{\mathcal{N}}$, except at most one, has been replaced.

The constructing phase starts considering the FST with the topology $\mathcal{T}$ obtained at the end of the merging phase. Note that, as $\mathcal{T}$ has two terminals and no Steiner point, the position of the nodes are fixed in the FST. At each iteration of the reconstruction phase, an edge (e_{ij}, v) of the FST, which is incident to an equilateral point e_{ij}, is selected. Let t_i and t_j be the two terminals corresponding to e_{ij}. If the edge (e_{ij}, v) does not intercept the $120°$ arc $t_i t_j$, then the reconstruction phase stops returning "0", otherwise the point of interception, denoted by s_k, is connected to t_i, t_j and v, and the equilateral point e_{ij} and its incident edge (e_{ij}, v) are removed from the FST (see Figure 4). Note that the resulting tree is also an FST. The reconstruction phase stops when either no intersection point exists or the constructed FST contains all terminals in $\bar{\mathcal{N}}$.

Once there are two possible locations for each equilateral point considered in Melzak FST procedure, a backtracking was initially proposed to determine the correct choice. The algorithm was greatly improved by Hwang in [18], who presented a linear time implementation of Melzak FST procedure, which is capable of determining the correct location of the equilateral points without backtracking, by applying the result from Corollary 12 (see also [17] for more details on the linear time implementation of Melzak FST procedure).

Still based on Remark 19, Winter [35] proposed an algorithm for the ESP,

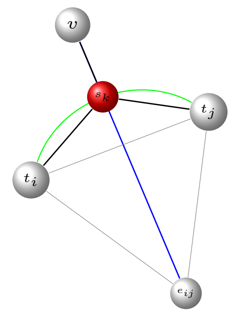


Figure 4: Illustration of Melzak FST procedure.

which considers an enumeration scheme to generate full Steiner topologies for each subset of $\mathcal{N}$. Pruning strategies are applied to discard subsets for which no STs exists, and the Melzak FST procedure is applied to remaining full Steiner topologies. Finally, all possible concatenations of the FSTs computed are performed in the search for the SMT.

In [32], Warme observed that the concatenation of FSTs can be formulated as a minimum spanning tree problem for a hypergraph with terminals as vertices and subsets of terminals spanned by FSTs as hyperedges. This problem can be solved by a branch-and-cut algorithm.

The GeoSteiner algorithm is based on Winter's algorithm and very successfully applies the branch-and-cut for the concatenation of FSTs. The branch-and-cut algorithm can also be used in higher dimensions. However, the success of GeoSteiner is highly related to the fact that a very large number of FSTs are discarded by the pruning strategies, and the concatenation phase starts with a very reduced number of FSTs stored, to be considered for the concatenation. The criteria used to discard FSTs, on the other hand, are mainly restricted to be applied in the plane, as the existence of only two possible locations for equilateral points has an important role in the algorithm.

4.2 ESP in higher dimensions

In 1968, Gilbert and Pollak [16] proposed the first solution strategy to solve the ESP in general dimension d, based on the result of the following theorem.

Theorem 20. Given an SMT, its topology is either a full Steiner topology or a degenerated full Steiner topology.

Proof. It suffices to construct a full Steiner topology $\mathcal{T}$, for which the RMT for $\mathcal{T}$ is the given SMT. For that, we first include in the SMT one Steiner point, for each terminal with degree 3, disconnect the terminal and any 2 of its neighbors, and then connect the 3 disconnected nodes to the new Steiner point. We then include in the resulting tree, one Steiner point for each terminal with degree two, disconnect the terminal and its 2 neighbors, and connect the 3 disconnected nodes to the new Steiner points (Figure 5). The resulting tree has a full Steiner topology $\mathcal{T}$, and clearly, the RMT for $\mathcal{T}$ is the given SMT. $\square$

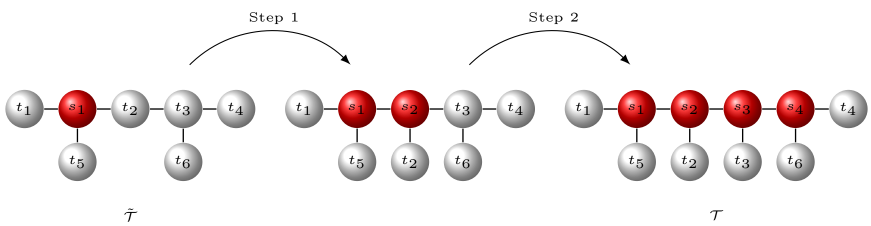


Figure 5: Obtaining a full Steiner topology $\mathcal{T}$ from topology $\tilde{\mathcal{T}}$.

The result presented by Gilbert and Pollak is illustrated in Figure 6, which shows in part (a) an SMT with a full Steiner topology and in part (b) an SMT with a topology degenerated from a full Steiner topology, where the dashed edges has zero length on the optimized tree.

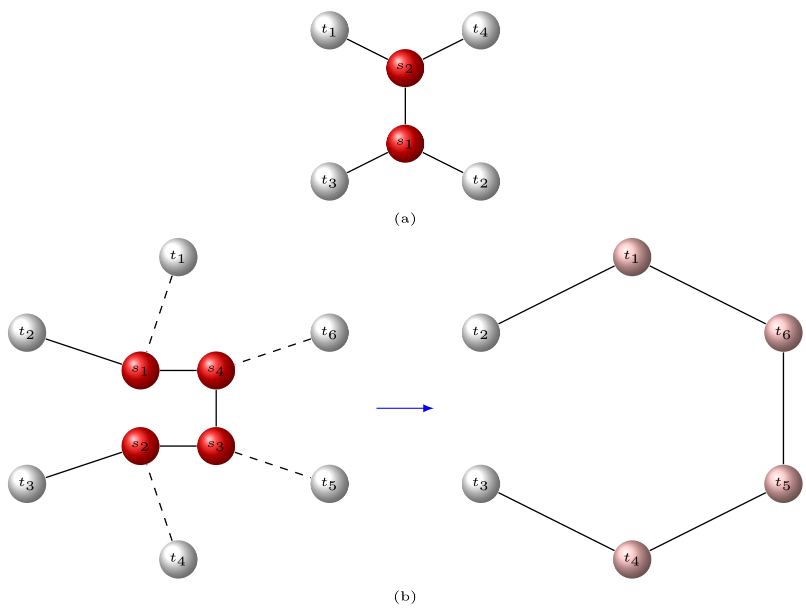


Figure 6: Illustration of Theorem 20.

The idea of Gilbert and Pollak's procedure is simply to enumerate all the full Steiner topologies for a given set $\mathcal{N}$ of n terminals and then compute the RMT for each one of them. We note that finding the RMT for a given topology is a convex optimization problem known in the literature as the minimization of the sum of Euclidean norms. The problem can be formulated as a second-order cone program (SOCP) and can be efficiently solved by interior-point algorithms [37]. Nevertheless, the combinatorial part of this algorithm, related to the enumeration of full Steiner topologies, is extremely heavy. The number of full Steiner topologies for n terminals is given by $f(n) := 1 \cdot 3 \cdot 5 \cdot 7 \cdots (2n-5) = (2n-5)!!$. This number has a super exponential growth with n, and for $n = 12$, for example, we already have more than 650 million full Steiner topologies to evaluate.

In 1992, Smith [29] proposed an algorithm aiming at turning Gilbert and Pollak's approach into a more practical method. Smith's algorithm implicitly enumerates the full Steiner topologies on an enumeration tree where nodes are pruned by bound. The algorithm orders the terminals in $\mathcal{N}$, and associates to each node of the enumeration tree, a full Steiner topology for a subset of $\mathcal{N}$. The enumeration tree has $n-2$ levels. At level 1, the root node corresponds to the unique full Steiner topology for the first 3 nodes in $\mathcal{N}$. Then, at each level down on the enumeration tree, the next terminal in $\mathcal{N}$ in the ordered sequence and a Steiner point are added to the topology corresponding to the parent node.

If no pruning occurs, the enumeration tree will have $1 \cdot 3 \cdot 5 \cdot 7 \cdots 2i-1$ nodes at level i, each corresponding to a distinct full Steiner topology for the first $i+2$ terminals in $\mathcal{N}$. Each node at level $i \in \{1, \dots, n-2\}$ of the enumeration tree corresponds to a topology with $i+2$ terminals, i Steiner points, and $2i+1$ edges. If a node at level $i < n-2$ is not pruned by bound, it will have $2i+1$ children, each child corresponding to an edge of its topology $\mathcal{T}$.

The children are obtained through a merging operation, which includes a terminal and a Steiner point on the given full Steiner topology $\mathcal{T}$. For the merging, an edge e of $\mathcal{T}$ is selected. The edge e is removed from $\mathcal{T}$, and the 2 disconnected nodes are then connected to the new Steiner point. Finally the new terminal is also connected to the new Steiner point, forming a new full Steiner topology. Figure 7 depicts the full Steiner topologies at the first two levels of Smith's enumeration tree.

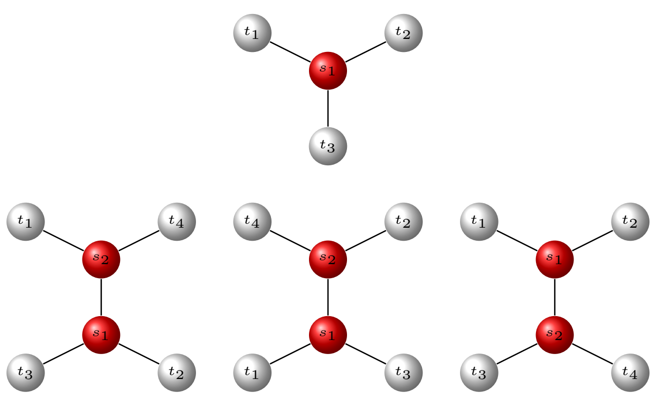


Figure 7: Full Steiner topologies at the first two levels of Smith's enumeration tree.

It is possible to verify that, if at each node on levels $1, \dots, n-3$ of the enumeration tree, we apply the merging, selecting all the edges of the corresponding topology, one each time, the nodes at level i will correspond to all distinct full Steiner topologies for the first $i + 2$ terminals, and each topology is generated only once.

For each node created on the enumeration tree, the RMT for the corresponding full Steiner topology for a subset of the terminals in $\mathcal{N}$ is computed. If the size of the RMT is not smaller than a given upper bound for the size of the SMT for $\mathcal{N}$, then the node is pruned, as it was proved by Smith that the merging operation cannot decrease the size of the RMT. The main drawback of Smith's algorithm is this weak pruning criterion. As the number of full Steiner topologies grows very fast with the number of terminals, it would be important to have a strong pruning criterion, that could prune nodes before too deep in the enumeration tree, to avoid the computation RMTs corresponding to a huge number of topologies. However, we observe that in general, only deep down in the enumeration tree, the number of terminals already added to the partially constructed full Steiner topology is enough to generate an RMT larger than an upper bound for an SMT for all terminals in $\mathcal{N}$. Consequently, the enumeration tree ends up having a very large number of nodes.

To mitigate this drawback, some research has been done in an attempt to increase the number of nodes pruned in Smith's algorithm. Fampa and Anstreicher [10] showed that the order in which the terminals are added to the tree does not need to be fixed to guarantee that all full Steiner topologies are enumerated by Smith's procedure. Thereby, they proposed a strong branching procedure using dual information given by the solution of the sub-problems, formulated as SOCPs. The idea of the strong branching is to invest some computation effort in the selection of the next terminal to add to a partially constructed topology, in an attempt to increase the size of the corresponding RMTs faster and to prune more nodes in consequence. Later, Laarhoven and Anstreicher [31] propose to use an alternative ordering of the terminals based on their distance from their centroid, with the first terminal being the farthest away and the final terminal being the closest. They claim that the new ordering leads to a faster computation than the strong branching and also decreases the number of nodes in Smith's enumeration tree. The authors also observe that geometric conditions that have been so successfully applied to prune topologies in the solution of the problem in the plane, were not considered at all in the algorithm proposed by Smith. They present a modification on Smith's algorithm applying the so called fathoming by geometry criterion, based on the following two lemmas.

Lemma 21. Consider terminals t_i and t_j. Let η_i (and η_j) be the distance from t_i (and from t_j), to the nearest other terminal in $\mathcal{N}$. Then in an SMT with a full Steiner topology, t_i and t_j may be connected to a common Steiner point only if

$$\|t_i - t_j\| \leq \eta_i + \eta_j. \tag{1}$$

The forbidden connection determined in Lemma 21 is depicted in Figure 8.

Considering that the 3 incident edges to a Steiner point make 120 degrees with each other in an SMT (*angle condition*), the result above was tightened in

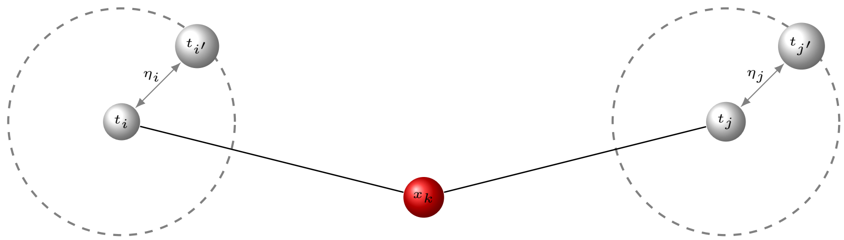


Figure 8: Forbidden path between t_i and t_j (Lemma 21).

[31], by replacing the inequality (1) by

$$\|t_i - t_j\| \leq \sqrt{\eta_i^2 + \eta_j^2 + \eta_i \eta_j}. \tag{2}$$

Considering then, the *bottleneck property*, the result above was extended as it follows.

Lemma 22. Consider terminals t_i and t_j. Let η_i (and η_j) be the distance from t_i (and from t_j) to the nearest other terminal in $\mathcal{N}$, and β_{ij} be bottleneck distance corresponding to t_i and t_j. Then in an SMT with a full Steiner topology, t_i and t_j may be connected by a path having two or fewer Steiner points only if

$$\|t_i - t_j\| \leq \eta_i + \eta_j + \beta_{ij}. \tag{3}$$

Finally, considering again the *angle condition*, the result above was tightened in [31], by replacing the inequality (3) by

$$\|t_i - t_j\| \leq \sqrt{(\eta_i + \eta_j)^2 + \beta_{ij}^2 + (\eta_i + \eta_j)\beta_{ij}}. \tag{4}$$

We should note here that (2) and (4) are mentioned as the base of a pruning criterion for full Steiner topologies on Smith's enumeration scheme in [31]. However, the RMTs corresponding to those topologies do not satisfy the *angle condition* in cases where the topology degenerates and Steiner points become incident to zero length edges. Therefore, the tighter results based on both inequalities (2) and (4) cannot be applied to prune topologies in Smith's algorithm and should not replace (1) and (3) in the pruning criteria.

The following simple example illustrates the incorrect pruning derived from inequality (2).

Example 23. Consider 3 terminals on a line. Smith's procedure solves the problem, considering the unique full Steiner topology $\mathcal{T}$, where the 3 terminals are connected to a Steiner point. The topology of the corresponding RMT is degenerated from $\mathcal{T}$, as shown in Figure 9. Note that a pruning criterion based on inequality (2) would suggest that $\mathcal{T}$ should be disregarded, once $\|t_1 - t_3\| = \eta_1 + \eta_3 > \sqrt{\eta_1^2 + \eta_3^2 + \eta_1\eta_3}$.

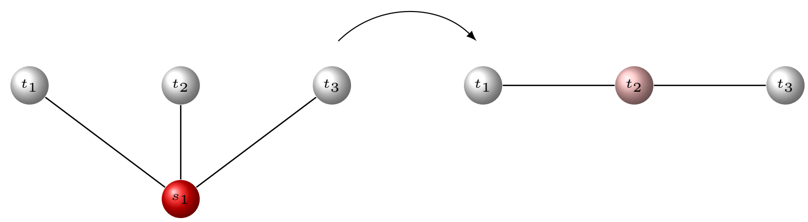


Figure 9: Application of Smith's algorithm to three terminals on a line.

In [14], the authors also propose a small modification on Smith's algorithm based on a best-first search strategy and on an alternative ordering of the terminals. In their version of the algorithm, terminals t_1, t_2 and t_3 maximize the sum of their pairwise distances, and the terminal t_k, for $k = 4, \dots, n$ is farthest away from $t_1, \dots, t_{k-1}$.

Moreover, they propose a new algorithm for the ESP based on the decomposition of full Steiner topologies described in the following lemma.

Lemma 24. Let $\mathcal{T}_1$, $\mathcal{T}_2$, and $\mathcal{T}_3$ be the edge disjoint topologies obtained by excluding a Steiner point s and its incident edges, from a full Steiner topology on n given terminals. It is possible to select s such that each one of the three topologies obtained has at most $\lfloor n/2 \rfloor$ terminals.

The topologies $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_3$ described in Lemma 24 are called branches in [14], and the new algorithm proposed is called the branch enumeration algorithm. Extending the ideas used in the GeoSteiner algorithm to concatenate FSTs on subsets of $\mathcal{N}$, the branch enumeration algorithm concatenates branches that are not pruned. The number of branches that need to be generated by the algorithm reduces with the application of the result in Lemma 24. The pruning criteria for the remaining branches are based on an upper bound UB on the length of the SMT for $\mathcal{N}$, and on the extension of the result in Lemma 22, described next.

Lemma 25. Let θ be the number of edges in the path between two given terminals t_i and t_j in an SMT for $\mathcal{N}$. Let β_{ij} be the *bottleneck distance* corresponding to this pair of terminals. Then

$$\theta \geq \|t_i - t_j\| / \beta_{ij}.$$

The pruning by bound in the branch enumeration algorithm is an expensive procedure. A branch $\mathcal{T}_k$ that contains the terminals in a subset $\mathcal{N}_k$ of $\mathcal{N}$ is pruned by bound if the length of the RMT for a full Steiner topology obtained from $\mathcal{T}_k$, by eliminating a Steiner point from it and connecting its neighbors, is not smaller than the difference between UB and the length of the SMT for the terminals in a subset of $\mathcal{N} \setminus \mathcal{N}_k$. These SMTs are previously computed for all subsets of $\mathcal{N}$ with up to 8 (or $\lfloor n/2 \rfloor$) terminals by the modified version of Smith's algorithm proposed in [14]. Therefore, to prune a branch, besides computing these SMTs, it is necessary to compute the RMT of the full Steiner topology that corresponds to the branch. Unlike what happens in the GeoSteiner algorithm, there is no way of interrupting the computation of the RMTs for identifying that the branch can be pruned, before the computation is finished. Furthermore, neither the RMTs computed for branches not pruned, nor the SMTs previously computed, are used after the concatenation of the branches, when another RMT, for the concatenated topology, needs to be computed. All branches with k terminals, for $k = 1, \ldots, \lfloor n/2 \rfloor$, that were not pruned by the result in Lemma 25, are considered for this process.

5 Mathematical programming formulations

Although some work has been done to increase the number of nodes pruned in the implicit enumeration of full Steiner topologies proposed by Smith, the pruning criteria are still very weak when compared to the ones that make the GeoSteiner algorithm so successful. Certainly, when we increase the dimension of the problem from 2 to 3, the core of Melzak FST algorithm, which is the computation of equilateral points, becomes a problem, as the number of possible locations for an equilateral point for a given pair of terminals goes from 2 to infinity. As a consequence, several geometric restrictions for the position of Steiner points that are so effective to prune topologies in the plane cannot be extended to higher dimensions. Nevertheless, there is another difficulty in using the geometric restrictions for Steiner points to prune topologies, which is specifically related to the enumeration scheme proposed by Smith. The difficulty comes from the fact that locations of Steiner points become more restricted as their connections become determined in the topology, and on Smith's scheme, the connections between nodes are only totally defined in the last level of the enumeration tree. We note that any edge that exists in a given topology during the enumeration may cease to exist in its descendant topologies. The complete lack of information about the final connections in the topology during the enumeration is a big obstacle to the restriction of the location of Steiner points. For example, the application of well known properties, as the *lune property*, and also of the more recent results presented in [31], derived from the Voronoi diagram associated with the set of terminals, becomes very restricted.

A possible way of eliminating this drawback, which is intrinsic to Smith's enumeration scheme, is to enumerate full Steiner topologies using a mathematical programming formulation for the ESP, where binary variables indicate whether or not a given connection in the tree exists. During an implicit enumeration of topologies by a branch-and-bound algorithm, once the value of a binary variable is fixed at a node of the enumeration tree, it will be fixed on all of its descendant nodes. Therefore, once an edge is set to exist in the topology, it will exist in all descendants. Another advantage of the application of a branch-and-bound algorithm to a mathematical programming formulation for the ESP is that, unlike what we have in Smith's enumeration, the formulation considered from the root node of the enumeration tree already takes into account all the terminals in $\mathcal{N}$. Therefore, once we can develop strong convex relaxations for the formulation, we have better chances of obtaining stronger lower bounds from the solution of sub-problems in higher levels of the enumeration tree.

The first mathematical programming formulation for the ESP was presented by Maculan, Michelon and Xavier in [22] and consists of a nonconvex mixed-integer nonlinear programming (MINLP) formulation. The authors consider a graph $G(V, E)$, where the set of vertices V contains the n given terminals indexed from 1 to n, and $n-2$ Steiner points indexed from $n+1$ to $2n-2$. The set of edges E connects each terminal to all Steiner points, and all Steiner points to each other. More specifically, the authors define $N := \{1, \ldots, n\}$, $S := \{n+1, \ldots, 2n-2\}$, $E_1 := \{(i,j) | i \in N, j \in S\}$ and $E_2 := \{(i,j) | i, j \in S,\ i < j\}$, $V := N \cup S$, and $E := E_1 \cup E_2$. They define a binary variable y_{ij} for each edge $(i,j) \in E$ and a variable $x_k \in \mathbb{R}^d$ for each $k \in S$, which represents the location of the Steiner point indexed by k. The objective of the formulation (MMX) presented below, represents the length of the solution tree and is modeled by a nonconvex function. The linear constraints of (MMX) model the set of full Steiner topologies for $\mathcal{N}$.

$$
\begin{array}{lll}
\text{(MMX)} \min & \displaystyle\sum_{(i,j)\in E_1} \|t_i - x_j\| y_{ij} + \sum_{(i,j)\in E_2} \|x_i - x_j\| y_{ij}, & \\
\text{s.t.:} & \displaystyle\sum_{j \in S} y_{ij} = 1, & i \in N, \\
& \displaystyle\sum_{i\in N} y_{ij} + \sum_{k<j, k\in S} y_{kj} + \sum_{k>j, k\in S} y_{jk} = 3, & j \in S, \\
& \displaystyle\sum_{k<j, k \in S} y_{kj} = 1, & j \in S - \{n+1\}, \\
& y_{ij} \in \{0,1\}, & (i,j) \in E, \\
& x_i \in \mathbb{R}^d, & i \in S.
\end{array}
$$

The authors in [22] show that all (and only) full Steiner topologies for $\mathcal{N}$ are represented by a feasible solution of (MMX). They also proved that the feasible set of the continuous relaxation of (MMX), obtained by replacing the integrality constraints by $y_{ij} \in [0,1]$, for all $(i,j) \in E$, represents the convex hull of the set of full Steiner topologies for $\mathcal{N}$. An alternative proof for this result was also given later in [8]. As done in Smith's algorithm, only full Steiner topologies are enumerated by (MMX), and the result in Theorem 20 guarantees that its optimal solution is an SMT.

The solution of (MMX) is a big challenge for global-optimization solvers. In [8], D'Ambrosio, Fampa, Lee and Vigerske identify the non-differentiability of the objective function at points where the topologies degenerate, as one of the difficulties faced by the solvers, and propose a piecewise-defined approximate differentiable function for the Euclidean norm. We refer the reader interested in the subject also to [20].

Similarly to Smith's original enumeration procedure, the formulation (MMX) does not consider any geometric condition to restrict the position of Steiner points. To strengthen the formulation, the authors in [8] present valid nonconvex inequalities, which are derived from the well known properties of SMTs. The inequalities added to the formulation are described on the following four theorems from [8]. For the proof of these theorems we refer the reader to [9].

Theorem 26 uses the result in Lemma 14.

Theorem 26. For all $n \geq 3$, we have

$$y_{ik}\|x_k - t_i\| \leq \eta_i,$$

for all $i \in N,\ k \in S$, where η_i is the distance from t_i to the nearest other terminal point.

Theorem 27 is based on the *angle condition*, which was proven to bound the sum of the distances between a Steiner point and two of its neighbors, even if the topologies degenerate. Figure 10 illustrates the result in the theorem, showing the isosceles triangle for which the sum of the distances is maximum.

Theorem 27. For all $n \geq 3$, we have

$$y_{ik}y_{jk}\left(\|x_k - t_i\| + \|x_k - t_j\|\right) \leq 2\|t_i - t_j\|/\sqrt{3}, \quad (5)$$

for all $i, j \in N,\ i < j,\ k \in S$.

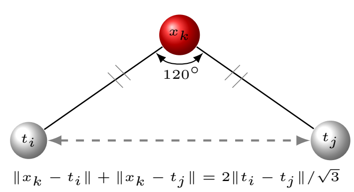


Figure 10: Illustration of Theorem 27.

Theorem 27 was also extended to the case where the Steiner point x_k is adjacent to only one terminal in the SMT and also to the case where x_k is not adjacent to any terminal. In these cases, inequalities 5 are replaced respectively by

$$y_{ik}y_{kl}\left(\|x_k - t_i\| + \|x_k - x_l\|\right) \leq 2\|t_i - x_l\|/\sqrt{3}, \quad (6)$$

for all $i \in N,\ k, l \in S,\ k < l$, and

$$y_{kl}y_{km}\left(\|x_k - x_l\| + \|x_k - x_m\|\right) \leq 2\|x_l - x_m\|/\sqrt{3}, \quad (7)$$

for all $k, l, m \in S,\ k < l < m$.

Theorem 28 is based on the location of a Steiner point and its neighbors in the same plane when $d = 3$. The results is also extended in [8] for higher dimensions.

Theorem 28. For $d = 3$ and $i, j \in N,\ i < j,\ k, l \in S, k < l$, we have

$$y_{ik} y_{jk} y_{kl} \cdot \det \begin{bmatrix} t_i & t_j & x_k & x_l \\ 1 & 1 & 1 & 1 \end{bmatrix} = 0. \tag{8}$$

Finally, Theorem 29 uses the result in Lemma 21.

Theorem 29. For $i, j \in N,\ i < j$, such that $\|t_i - t_j\| > \eta_i + \eta_j$,

$$y_{ik} + y_{jk} \leq 1,$$

for all $k \in S$.

Even with the improvements proposed in [8], the solution of the ESP using a nonconvex MINLP formulation has led to poor numerical results. Besides the bad quality of lower bounds given by the solution of sub-problems in the spacial branch-and-bound algorithm, a drawback observed with the use of global solvers is that they do not explore a nice characteristic of (MMX), which is the fact that model becomes convex, once the binary variables are all fixed, and continue the spacial branching on the continuous variables that represent the location of the Steiner points.

In [13], Fampa and Maculan, proposed the convex MINLP formulation (FM) for the ESP, by the introduction of a "big-M" parameter in formulation (MMX) that bounds the distances between nodes in the SMT and transforms the non-convex continuous relaxation of (MMX) into an SOCP.

$$\begin{array}{lrl}
\text{(FM) min} & \displaystyle\sum_{(i,j)\in E} d_{ij}, & \\
\text{s.t.:} & d_{ij} \geq ||t_i - x_j|| - M(1 - y_{ij}), & (i,j) \in E_1, \\
& d_{ij} \geq ||x_i - x_j|| - M(1 - y_{ij}), & (i,j) \in E_2, \\
& \displaystyle\sum_{j \in S} y_{ij} = 1, & i \in N, \\
& \displaystyle\sum_{i \in N} y_{ij} + \sum_{k<j, k\in S} y_{kj} + \sum_{k>j, k\in S} y_{jk} = 3, & j \in S, \\
& \displaystyle\sum_{i<j, i\in S} y_{ij} = 1, & j \in S \ - \ \{n+1\}, \\
& y_{ij} \in \{0,1\},\ d_{ij} \in \mathbb{R}_+, & (i,j) \in E, \\
& x_i \in \mathbb{R}^d, & i \in S.
\end{array}$$

Although the application of a branch-and-bound algorithm to the convex MINLP problem (FM), is much more efficient than the application of global solvers to (MMX), in [12], Fampa, Lee and Melo identify two weakness of (FM), which are the presence of a large number of isomorphic full Steiner topologies in the feasible set, and also the bad quality of the lower bounds given by the solution of the sub-problems, weakened by the "big-M" parameter. The authors proposed a specialized branch-and-bound procedure based on formulation (FM), called SAMBA. An important feature of the algorithm is related to the pruning of sub-problems where the binary variables already fixed at 1 correspond to edges on topologies that should not be evaluated, either because they are isomorphic to another topology in the feasible set, or because they do not satisfy the results in Lemmas 21 and 22.

Improved results were presented for SAMBA, when compared to the original application of convex MINLP solvers to (FM). Nevertheless, we observe that some well known properties of SMTs could still be applied in the algorithm.

6 Ongoing research and future directions to explore

In this section, we explore possible strategies to increase the number of nodes pruned when a branch-and-cut algorithm is applied to (FM).

We start presenting Lemma 30, which gives a new bound for the number of edges in the path between two terminals in an SMT. It is straightforward to verify that the *bottleneck distance* corresponding to a pair of terminals t_i and t_j cannot be smaller than the distance from t_i, or from t_j, to the nearest other terminal, i.e., $\beta_{ij} \geq \max\{\eta_i, \eta_j\}$. Therefore if $\|t_i - t_j\| \leq \eta_i + \eta_j$, then $\|t_i - t_j\| \leq 2\beta_{ij}$, or equivalently $\|t_i - t_j\|/\beta_{ij} \leq 2$. In this case, Lemma 25 only guarantees that there are at least 2 edges in the path between t_i and t_j, which is a redundant information, as we are considering only full Steiner topologies. In case $\|t_i - t_j\| > \eta_i + \eta_j$, on the other hand, we present a tighter bound for the number of edges in the path between the terminals.

Lemma 30. Let θ be the number of edges in the path between two given terminals t_i and t_j in an SMT for $\mathcal{N}$, such that $\|t_i - t_j\| \geq \eta_i + \eta_j$, where $\eta_{i(j)}$ is the distance from $t_{i(j)}$ to the nearest other terminal. Let β_{ij} be the *bottleneck distance* corresponding to this pair of terminals. Then

$$\theta \geq \lceil (\|t_i - t_j\| - \eta_i - \eta_j)/\beta_{ij} \rceil + 2.$$

Proof. Let $(t_i, u_1, u_2, \ldots, u_k, t_j)$ be the path between t_i and t_j. Note that the nodes u_ℓ, for $\ell = 1, \ldots, k$, may be Steiner points or terminals, as we do not restrict the result to SMTs with full Steiner topologies. Also note that the length of the edges (t_i, u_1) and (u_k, t_j) cannot be greater than η_i and η_j, respectively. Therefore the distance between u_1 and u_k is at least $\|t_i - t_j\| - \eta_i - \eta_j$. As each edge in the path between u_1 and u_k cannot be longer than β_{ij}, the result follows. □

The following example illustrates the result in Lemma 30.

Example 31. Consider 5 terminals t_1, t_2, t_3, t_4, t_5 on a line, such that $\|t_1 - t_5\| = \tau$, as represented in Figure 11. The minimum spanning tree of the terminals contains edges (t_1, t_2),(t_2, t_3),(t_3, t_4),(t_4, t_5) with respective lengths 0.1τ,

0.5τ, 0.3τ, and 0.1τ. In this case, $\beta_{15} = 0.5\tau$. Lemma 25 determines that the number of edges in the path between t_1 and t_5 in an SMT for the terminals is at least $\tau/(0.5\tau) = 2$, while Lemma 30 determines that it is at least $\lceil(\tau - 0.1\tau - 0.1\tau)/(0.5\tau)\rceil + 2 = 4$.

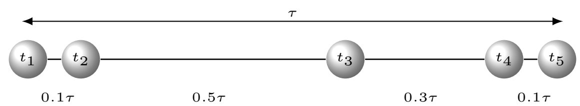


Figure 11: Illustration of Example 31.

As done for SAMBA, we consider an enumeration scheme for full Steiner topologies with several enumeration trees, each one with its own root node. In each root node, all variables y_{ij}, for $(i,j) \in E_2$, are already fixed, and each partial integer solution corresponds to the topology of a spanning tree on the $n-2$ Steiner points. As mentioned in [12], the number of non-isomorphic topologies of spanning trees with all nodes of degree 3 or less, connecting the $n-2$ Steiner points, is much smaller than the number of full Steiner topologies for n terminals. For $n = 10$, for example, these numbers are 11 and $2,027,025$ respectively, and the 11 topologies of non-isomorphic spanning trees on the Steiner points could be computed in less than 0.01 seconds. For this reason, we focus on new strategies of pruning, considering that all variables y_{ij}, for $(i,j) \in E_2$, are fixed at the sub-problems on the branch-and-bound enumeration. This means that all non-isomorphic topologies of spanning trees on the Steiner points only, with degree three or less, are computed and stored in advance.

Unlike what was done in SAMBA, however, we do not consider that all the non-isomorphic full Steiner topologies for n terminals are stored, as this restricts the application of the algorithm to a small number of terminals. As mentioned in [12], it would be interesting then to check isomorphism between the full Steiner topologies constructed, during the execution of the branch-and-bound. This procedure is one of ours current topics of research, which explores the symmetry group of the problem and is based on the ideas of isomorphism pruning and orbital branching, proposed in [23, 24, 26].

We consider an ordering $t_1, t_2, \ldots, t_n$ in which the terminals are connected to the partially constructed Steiner tree. At each level down on the enumeration tree, the next terminal ℓ in the sequence is connected, meaning that a variable $y_{\ell k}$ is fixed at 1, for some $(\ell, k) \in E_1$. In each node at this level of the enumeration tree, a different Steiner point k is chosen for the connection, such that its degree is still less than 3. Therefore, the children of a given node in the enumeration tree represent all possible ways of connecting the next terminal to partially constructed Steiner tree.

We propose then the following pruning strategies, based on Lemma 30, on the well known geometric properties of SMTs, and also on results from the literature previously described.

At a given node of the branch-and-bound enumeration tree, suppose that terminal t_i is connected to Steiner point x_k, for some $k \in S$, i.e., y_{ik} is fixed at 1. We consider the unique path between between t_i and t_j on the partially constructed tree, for each $j = 1, \dots, i-1$ and let E_{ij} be the set of all edges on this path, except the edges incident to the two terminals. We note that, as the spanning trees on the Steiner points are computed a priori on the branch-and-bound algorithm, the path between each pair of Steiner points can also be computed a priori. Therefore, the set E_{ij} can be quickly determined during the execution of the algorithm, once the Steiner points to which terminals t_i and t_j are connected, are specified.

We first verify if the number of edges in the path connecting t_i and t_j violates the upper bound given by Lemma 30, for any $j = 1, \dots, i-1$. If it does, the node is pruned, otherwise, we add the following constraints to the sub-problem, based on the results of Lemmas 14 and 18.

$$\|t_i - x_k\| \leq \eta_i, \tag{9}$$

$$\|x_l - x_{l'}\| \leq \beta_{ij}, \text{ for all pair } l, l', \text{ such that } (x_l, x_{l'}) \in E_{ij}. \tag{10}$$

Furthermore, if there is another terminal t_q also connected to x_k, i.e., if $y_{ik} + y_{qk} = 2$, for some $q \in \{1, \dots, i-1\}$, then we also add the following constraint to the sub-problem, based on Theorem 27, from [8, 9].

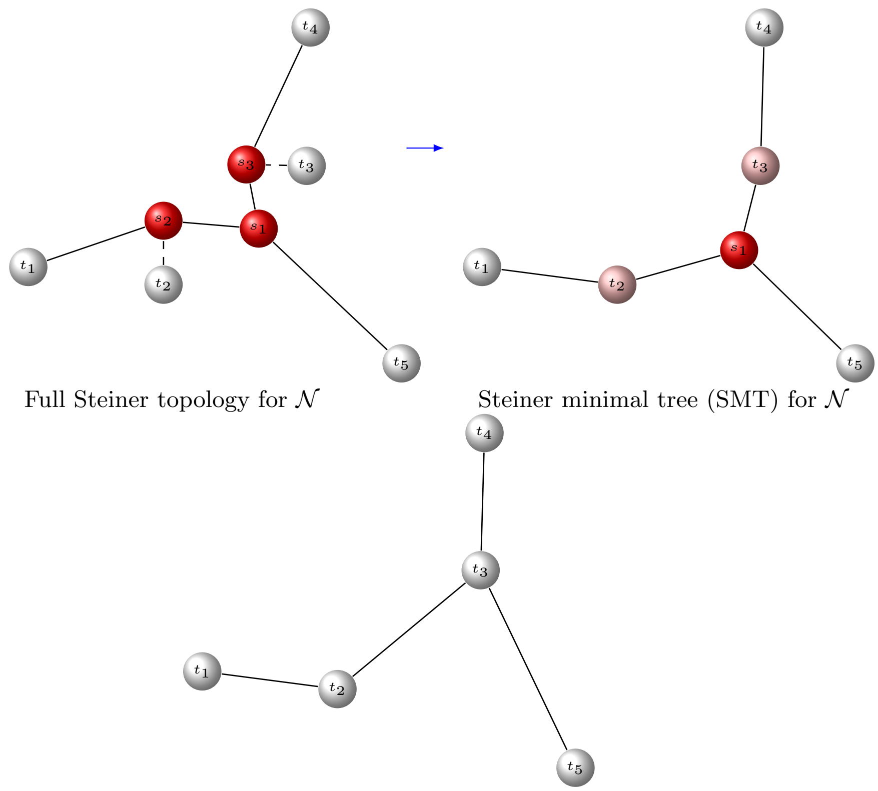
$$\|t_i - x_k\| + \|t_q - x_k\| \leq 2\|t_i - t_q\|/\sqrt{3}. \tag{11}$$

Another relevant topic is the order in which the terminals are connected to the tree. The modifications proposed in Smith's algorithm have shown the importance of ordering the terminals with the goal of increasing as fast as possible the lower bounds given by the solution of sub-problems. We propose an ordering for connecting the terminals in our branch-and-cut algorithm, with the purpose of making the cuts (9–11) effective as soon as possible.

We define the distance between two nodes in a tree as the number of edges in the path connecting the nodes. We compute the sum of the distances between each terminal and all the others in the minimum spanning tree for the terminals in $\mathcal{N}$. The criterion that we propose to order the terminals in our branch-and-cut algorithm, is the sum of distances from each terminal to all the others. The first terminal connected is the one for which this sum is the greater. To break ties, we select first, terminals for which the shortest distance to other terminals is minimum, which will restrict more the length of the edge incident to the terminal connected (inequality (9)). Once the terminal to be connected to the tree is selected, we also suggest an ordering to select the Steiner point to which the terminal is connected. For that, we compute the sum of the distances between each candidate Steiner point, and all terminals already connected to the tree. The first Steiner point selected is the one for which this sum is the greater.

The following example illustrates how the strategies proposed would affect the behavior of the branch-and-bound algorithm on an small instance in the plane.

Example 32. Let $\mathcal{N} = \{t_1, t_2, t_3, t_4, t_5\}$ be the set of 5 given terminals, and x_6, x_7, and x_8 represent 3 Steiner points. The SMT for the terminals, a full Steiner topology from which the topology of the SMT degenerates, and the minimal spanning tree for $\mathcal{N}$ are all depicted in Figure 12.



Full Steiner topology for $\mathcal{N}$ — Steiner minimal tree (SMT) for $\mathcal{N}$

Minimimum spanning tree (MST) for $\mathcal{N}$

Figure 12: SMT and MST for Example 32.

We consider that the connections between Steiner points are already determined at the beginning of the application of the branch-and-bound algorithm, by fixing $y_{6,7} = y_{6,8} = 1$.

In Figure 13, we present the plots of the graphs specified by the adjacency matrices defined by the values of the variables y_{ij}, for all $(i, j) \in E$, at the solutions of the continuous relaxation of (FM), with and without inequalities (9–11). The plots correspond to the nodes of the branch-and-bound enumeration tree in the branch determined when we fix the variables $y_{1,2} = 1$, $y_{4,3} = 1$, and $y_{5,1} = 1$, in this order, which was based on the criterion proposed above. The bounds given by the relaxations are also shown in the figure.

When the inequalities (9–11) are added to (FM), after fixing $y_{5,1} = 1$, the solution of the relaxation becomes equal to the SMT and the node is pruned, as the values of all variables become integer. If the inequalities are not added, on the other hand, the node can not be pruned.

In Example 32, the inequalities (9–11) added to (FM) are all convex inequalities. Additional geometric restrictions could be imposed to Steiner points, if nonconvex cuts were added. For example, considering inequalities (6,7) from [8], we first note that once terminal t_i is connected to Steiner point x_k, we could also add to the sub-problem the nonconvex cut

$$\|x_k - t_i\| + \|x_k - x_l\| \leq 2\|t_i - x_l\|/\sqrt{3}, \tag{12}$$

where x_k and x_l are connected in the spanning tree of the Steiner points.

Furthermore, once the connections between Steiner points are set in advance, the following nonconvex cuts could be included in the sub-problem already at the root node of the enumeration tree,

$$\|x_k - x_l\| + \|x_k - x_m\| \leq 2\|x_l - x_m\|/\sqrt{3}, \tag{13}$$

for all $k, l, m \in S,\ k < l < m$, such that x_k is connected to x_l and x_m.

To avoid the inclusion of nonconvex cuts to the sub-problems, we could replace the norms on the right-hand-side of inequalities (12,13), by the upper bound on the distance between two nodes of the SMT, represented by the "big-M" parameter in (FM). The same relaxations could be applied to nonconvex inequalities derived from the lune property. Nevertheless, the relaxations obtained with "big-M" are, in general, very week. Developing stronger relaxations for these nonconvex cuts is a topic for investigation.

7 Final remarks

The Euclidean Steiner Problem (ESP) in dimensions greater than 2 is an extremely difficult problem to solve in practice. All solution methods that have been proposed for the ESP involve the enumeration of full Steiner topologies on the given set of terminals $\mathcal{N}$, or on subsets of $\mathcal{N}$. As the number of full Steiner topologies has a super exponential growth with the number of terminals, it is essential to apply effective pruning criteria to be able to discard a large number of them. Several properties related to the topology of Steiner minimal trees and also to the geometric position of the Steiner points are known, and have been effectively used to solve problems in the plane. In this case, the GeoSteiner algorithm was able to solve instances of the problem with thousands of terminals. In higher dimensions, an enumeration scheme of full Steiner topologies proposed by Smith in 1992, has been considered in works with the best numerical results for the problem. Nevertheless, instances with more than 18 terminals could not be solved in dimensions higher than 2. Several geometric properties that were used in GeoSteiner for pruning, can only be applied in the plane. We note, however, that the enumeration scheme proposed by Smith restricts even more

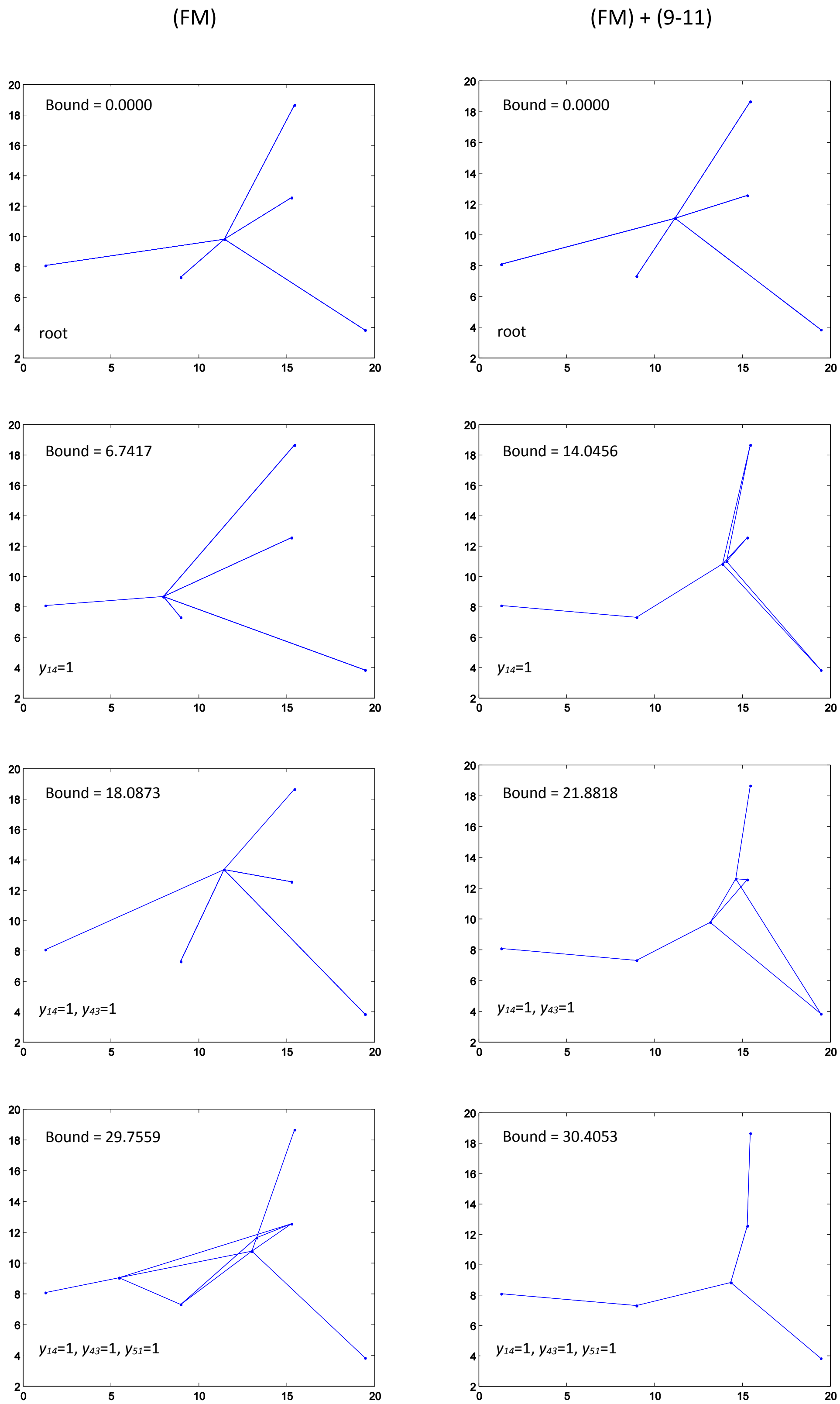


Figure 13: Branch-and-bound for Example 32.

the application of geometric properties, and some properties that could be applied in higher dimensions cannot be applied in Smith enumeration due to the lack of information about connections on the final constructed tree during the execution of the algorithm.

A possible way to apply geometric conditions for pruning in the enumeration of full Steiner topologies, is to use a mathematical programming formulation for the problem and apply to it a branch-and-bound or branch-and-cut algorithm. Results obtained so far with this methodology were still not able to improve the results presented by algorithms based on Smith's enumeration scheme. However, there is more space to apply the geometric properties of SMTs to prune topologies in branch-and-cut algorithm with the inclusion of valid inequalities to strengthen the relaxations of the sub-problems. Some of these inequalities are convex and can be immediately included in the model. Nevertheless, several restrictions on the positions of Steiner points are modeled as nonconvex cuts, and the development of good convex relaxations for these cuts is still a challenge.

References

[1] E. Althaus, J. Kupilas, and R. Naujoks. On the low–dimensional steiner minimum tree problem in hamming metric. *Theoretical Computer Science*, 505:2–10, 2013.

[2] K. Bopp. *Üeber das kürzeste Verbindungssystem zwischen vier Punkten.* PhD thesis, Universität Göttingen, 1879.

[3] M. Brazil, D. A. Thomas, B. K. Nielsen, P. Winter, C. Wulff-Nilsen, and M. Zachariasen. A novel approach to phylogenetic trees: d-dimensional geometric steiner trees. *Networks*, 53(2):104–111, 2009.

[4] L. L. Cavalli-Sforza and A. W. F. Edwards. Phylogenetic analysis: Models and estimation procedures. *Evolution*, 21:550–570, 1967.

[5] D. Cieslik. Shortest connectivity – introduction with applications in phylogeny. In *Combinatorial Optimization*, volume 17. Springer, New York, 2004.

[6] E. J. Cockayne. On the steiner problem. *Canad. Math. Bull.*, 10:431–450, 1967.

[7] R. Courant and H. Robbins. *What is Mathematics?* Oxford University Press, New York, 1941.

[8] C. D'Ambrosio, M. Fampa, J. Lee, and S. Vigerske. On a nonconvex MINLP formulation of the Euclidean Steiner tree problems in n-space. *LNCS (SEA 2015)*, 9125:122–133, 2015.

[9] C. D'Ambrosio, M. Fampa, J. Lee, and S. Vigerske. On a nonconvex MINLP formulation of the Euclidean Steiner tree problems in n-space: missing proofs. *Optimization Letters*, 2018. https://doi.org/10.1007/s11590-018-1295-1.

[10] M. Fampa and K. M. Anstreicher. An improved algorithm for computing Steiner minimal trees in Euclidean d-space. *Discrete Optimization*, 5(2):530–540, 2008.

[11] M. Fampa, J. Lee, and N. Maculan. An overview of exact algorithms for the euclidean steiner tree problem in n-space. *International Transactions in Operational Research*, 23:861–874, 2016.

[12] M. Fampa, J. Lee, and W. Melo. A specialized branch-and-bound algorithm for the Euclidean Steiner tree problem in n-space. *Computational Optimization and Applications*, 65:47–71, 2016.

[13] M. Fampa and N. Maculan. Using a conic formulation for finding Steiner minimal trees. *Numerical Algorithms*, 35(2-4):315–330, 2004.

[14] R. Fonseca, M. Brazil, P. Winter, and M. Zachariasen. Faster exact algorithm for computing Steiner trees in higher dimensional Euclidean spaces. *Presented at the 11th DIMACS Implementation Challenge Workshop*, 2014. http://dimacs11.cs.princeton.edu/workshop/FonsecaBrazilWinterZachariasen.pdf.

[15] R. Garey, M.R. Graham and D. Johnson. The complexity of computing Steiner minimal trees. *SIAM J. Applied Mathematics*, 32(4):835–859, 1977.

[16] E. Gilbert and H. Pollak. Steiner minimal trees. *SIAM J. Applied Math.*, 16:1–29, 1968.

[17] F. Hwang, D. Richards, and W. Winter. *The Steiner tree problem*, volume 53, Annals of Discrete Math. Elsevier, Amsterdam, 1992.

[18] F. K. Hwang. A linear time algorithm for full steiner trees. *Operations Research Letters*, 5:235–237, 1986.

[19] B. Korte, H. J. Promel, and A. Steger. Steiner trees in VLSI-layout. In B. Korte, L. Lovasz, H. J. Promel, and A. Schrijver, editors, *Paths, Flows, and VLSI-Layout*, pages 185–214. Springer-Verlag, Berlin, 1990.

[20] J. Lee and D. Skipper. Virtuous smoothing for global optimization. *Journal of Global Optimization*, 69:677–697, 2017.

[21] T. Lengauer. *Combinatorial Algorithms for Integrated Circuat Layout*. John Wiley & Sons, Chichester, England, 1990.

[22] N. Maculan, P. Michelon, and A. Xavier. The Euclidean Steiner tree problem in $\mathbb{R}^n$: A mathematical programming formulation. *Annals of Operations Research*, 96:209–220, 2000.

[23] F. Margot. Pruning by isomorphism in branch-and-cut. *Mathematical Programming*, 94:71–90, 2002.

[24] F. Margot. Exploiting orbits in symmetric ilp. *Mathematical Programming, Series B*, 98:3–21, 2003.

[25] Z. A. Melzak. On the problem of steiner. *Canad. Math. Bull.*, 4:143–148, 1961.

[26] J. Ostrowski, J. Linderoth, F. Rossi, and S. Smriglio. Orbital branching. *Mathematical Programming*, 126(1):147–178, 2011.

[27] J. Rubinstein, D. Thomas, and N. Wormald. Steiner trees for terminals constrained to curves. *Siam Journal on Discrete Mathematics*, 10(1):1–17, 1997.

[28] J. M. Smith, Y. Jang, and M. K. Kim. Steiner minimal trees, twist angles, and the protein folding problem. *PROTEINS: Structure, Function, and Bioinformatics*, 66(4):889–902, 2007.

[29] W. D. Smith. How to find Steiner minimal trees in Euclidean d-space. *Algorithmica*, 7(1-6):137–177, 1992.

[30] C. Stanton and J. M. Smith. Steiner trees and $3-d$ macromolecular conformation. *INFORMS Journal on Computing*, 16:470–485, 2004.

[31] J. W. Van Laarhoven and K. M. Anstreicher. Geometric conditions for Euclidean Steiner trees in $\mathbb{R}^d$. *Computational Geometry*, 46:520–531, 2013.

[32] D. M. Warme. A new exact algorithm for rectilinear steiner minimal trees. Technical report, System Simulation Solutions, Inc., Alexandria, VA 22314, USA, 1997.

[33] D. M. Warme, P. Winter, and M. Zachariasen. Exact algorithms for plane Steiner tree problems: a computational study. *Advances in Steiner Trees*, pages 81–116, 1998.

[34] D. M. Warme, P. Winter, and M. Zachariasen. Exact solutions to large-scale plane Steiner tree problems. *In: Proceedings of the Tenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 979–980, 1999.

[35] P. Winter. An algorithm for the steiner problem in the euclidean plane. *Networks*, 15:323–345, 1985.

[36] P. Winter and M. Zachariasen. Euclidean Steiner minimum trees: An improved exact algorithm. *Networks*, 30:149–166, 1997.

[37] G. Xue and Y. Ye. An efficient algorithm for minimizing a sum of Euclidean norms with applications. *Siam Journal of Optimization*, 7(4):1017–1036, 1997.

[38] M. Zachariasen and A. Rohe. Rectilinear group steiner trees and applications in vlsi design. *Mathematical Programming*, 94:407–433, 2003.