On a nonconvex MINLP formulation of the Euclidean Steiner tree problems in n-space

Claudia D'Ambrosio[1], Marcia Fampa[2], Jon Lee[3], and Stefan Vigerske[4]

[1] LIX CNRS (UMR7161), École Polytechnique 91128 Palaiseau Cedex, France. dambrosio@lix.polytechnique.fr

[2] Instituto de Matemática and COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil. fampa@cos.ufrj.br

[3] IOE Department, University of Michigan. Ann Arbor, Michigan, USA. jonxlee@umich.edu

[4] ZIB (Konrad-Zuse-Zentrum für Informationstechnik Berlin), Berlin, Germany. vigerske@zib.de

Abstract. The Euclidean Steiner Tree Problem in dimension greater than 2 is notoriously difficult. Successful methods for exact solution are *not* based on mathematical-optimization — rather, they involve very sophisticated enumeration. There are two types of mathematical-optimization formulations in the literature, and it is an understatement to say that neither scales well enough to be useful. We focus on a known nonconvex MINLP formulation. Our goal is to make some first steps in improving the formulation so that large instances may eventually be amenable to solution by a spatial branch-and-bound algorithm. Along the way, we developed a new feature which we incorporated into the global-optimization solver `SCIP` and made accessible via the modeling language `AMPL`, for handling piecewise-smooth univariate functions that are globally concave.

1 Introduction

The Euclidean Steiner tree problem (ESTP) in $\mathbb{R}^n$ is: Given a set of finite points in $\mathbb{R}^n$, find a tree of minimal Euclidean length spanning these points, using or not additional points. Original points are *terminals* and additional nodes in the tree are *Steiner points*. The ESTP is NP-Hard [9], and interest in the problem stems from both the mathematical challenge and its potential applications (e.g., communications, infrastructure networks). In biology, [3] gives an application of the ESTP to phylogenetic analysis (i.e., the construction of evolutionary trees).

Basic properties of an optimal solution, called a *Steiner minimal tree (SMT)*, are: (*i*) A Steiner point in an SMT has degree 3; a Steiner point and its adjacent nodes lie in a plane, and the angles between the edges connecting the point to its adjacent nodes are 120 degrees. (*ii*) A terminal in an SMT has degree between 1 and 3. (*iii*) An SMT on p terminals has at most $p-2$ Steiner points (see [12,5])).

The *topology* of a Steiner tree is the tree for which we have fixed the number of Steiner points and the edges between all points, but not the position of the Steiner points. A topology is a *Steiner topology* if each Steiner point has degree

3 and each terminal has degree 3 or less. A Steiner topology with p terminals is a *full Steiner topology* if there are $p-2$ Steiner points and each terminal has degree 1. A *full Steiner tree* is a Steiner tree corresponding to a full Steiner topology. A Steiner tree corresponding to some topology, but with certain edges shrunk to zero length, is *degenerate*. Any SMT with a non-full Steiner topology can be associated with a full Steiner topology for which the tree is degenerate.

Many papers have addressed exact solution in $\mathbb{R}^2$, and impressive results were obtained with the GeoSteiner algorithm [18]. But these algorithms cannot be applied when $n \geq 3$, and only a few papers have considered exact solution in this case. [11] proposed solving the problem in $\mathbb{R}^n$ by enumerating all Steiner topologies and computing a min-length tree associated with each topology, which in practice, can only solve very small instances because of the fast growth of the number of topologies as p increases. A branch-and-bound (b&b) algorithm for finding SMTs in $\mathbb{R}^n$ was proposed by Smith [14]. He presented a scheme for implicitly enumerating all full Steiner topologies on a given set of terminals, and he gave computational results sufficient to disprove for all $3 \leq n \leq 9$, an important conjecture of Gilbert and Pollak on the "Steiner ratio". Fampa and Anstreicher [6] used Smiths's enumeration scheme and proposed a conic formulation for the problem of locating the Steiner points for a given topology, to obtain a lower bound on the min tree length and to implement a "strong branching" technique. [15] presented geometric conditions that are satisfied by Steiner trees with a full topology and applied those conditions to eliminate candidate topologies in Smith's scheme. The best computational results for the ESTP for $n \geq 3$ are presented in these two last papers.

None of the works mentioned above have considered a math-programming formulation for the ESTP, which was presented only in [13] and [7]. [13] formulated the ESTP as a non-convex mixed-integer nonlinear programming (MINLP) problem and proposed a b&b algorithm using Lagrangian dual bounds. [7] presented a convex MINLP formulation that could be implemented in a b&b algorithm using bounds computable from conic problems. Both formulations use 0/1 variables to indicate whether the edge connecting two nodes is present in a Steiner topology. The presence of these 0/1 variables leads to a natural branching scheme, however neither [13] nor [7] present computational results.

We did some preliminary experiments with the nonconvex model. We solved some randomly generated instances using `SCIP` [1,16], and the results are dismal. Two difficulties observed have motivated this research: the weakness of the lower bounds given by the relaxations, and non-differentiability at points where the solution degenerates. In what follows, we investigate strategies to deal with these difficulties: (i) the use of approximate differentiable functions for the Euclidean norm, and (ii) nonconvex cuts based on geometric considerations.

2 A Nonconvex MINLP formulation

[13] formulates the ESTP as a nonconvex MINLP problem, first defining a special graph $G = (V, E)$. Let $P := \{1, 2, ..., p\}$ be the indices associated with the given

terminals $a^1, a^2, ..., a^p$ and $S := \{p+1, p+2, ..., 2p-2\}$ be the indices associated with the Steiner points $x^{p+1}, x^{p+2}, ..., x^{2p-2}$. Let $V = P \cup S$. Denote by $[i,j]$ an edge of G, with $i, j \in V$ such that $i < j$. Define $E := E_1 \cup E_2$, where $E_1 := \{[i,j] \ : \ i \in P, j \in S\}$ and $E_2 := \{[i,j] \ : \ i \in S, j \in S\}$. Define a 0/1 y_{ij} for each edge $[i,j] \in E$, where $y_{ij} = 1$ if the edge $[i,j]$ is present in the SMT and 0 otherwise. The ESTP is then formulated as

$$\text{(MMX)} \ \min \sum_{[i,j]\in E_1} \|a^i - x^j\| y_{ij} + \sum_{[i,j]\in E_2} \|x^i - x^j\| y_{ij}, \tag{1}$$

$$\sum_{j\in S} y_{ij} = 1, \quad \text{for } i \in P, \tag{2}$$

$$\sum_{i\in P} y_{ij} + \sum_{k<j, k\in S} y_{kj} + \sum_{k>j, k\in S} y_{jk} = 3, \quad \text{for } j \in S, \tag{3}$$

$$\sum_{k<j, k\in S} y_{kj} = 1, \quad \text{for } j \in S - \{p+1\}, \tag{4}$$

$$y_{ij} \in \{0,1\}, \ [i,j] \in E, \qquad x^i \in \mathbb{R}^n, \ i \in S, \tag{5}$$

where $\|v\| := \sqrt{\sum_{l=1}^n v_l^2}$ is the Euclidean norm of $v \in \mathbb{R}^n$. The constraints model a full Steiner topology for p given terminals in $\mathbb{R}^n$. Constraints (2) enforce that the degree of each terminal node is equal to 1. Constraints (3) enforce that the degree of each Steiner point is equal to 3, and constraints (4) eliminate cycles. Every full Steiner tree corresponds to a feasible solution of the formulation.

We aim to solve MMX using a spatial branch-and-bound (sbb) algorithm, as implemented in `SCIP`. This is not nearly straightforward, as we have to deal with non-differentiability of the distance function and with poor bounds that arise. In what follows, we propose approaches for handling these difficulties.

3 Dealing with the non-differentiability

The continuous relaxation of MMX is a nonconvex NLP problem. Convergence of most NLP solvers (e.g. `Ipopt` [17]) requires that functions be twice continuously differentiable. This is not the case for MMX due to the non-differentiability of the Euclidean norm when the solution degenerates (i.e., when the norm is zero); and it is easy to see examples where the optimal solution does degenerate. In water-network optimization, [2] smooths away non-differentiability at zero of another function (modeling the pressure drop due to friction of the turbulent flow of water in a pipe). There are different ways that we can deal with the non-differentiability that we face. Let $w(x^i - x^j) := \|x^i - x^j\|^2$, so that $\|x^i - x^j\| = \sqrt{w(x^i - x^j)}$. In this way, we can focus on $\sqrt{\cdot}$, the source of the non-differentiability.

3.1 Implicit square roots

One possibility is to introduce an auxiliary variable z, and use the additional inequality $-z^2 + w \leq 0$ and the nonnegativity constraint $z \geq 0$. In this way, any optimal solution will have $z = \sqrt{w}$. This looks possibly attractive, but the overhead of so many additional nonnegatively-constrained variables and the difficulty of many additional nonconvex constraints is prohibitive.

On the other hand, while nonconvex, these functions $-z^2+w$ manifest themselves as $-z_{ij}^2+\sum_{l=1}^n v_l^2$, with $v = x^i - x^j$. The nonconvexity in $-z_{ij}^2+\sum_{l=1}^n v_l^2$ is isolated to $-z_{ij}^2$, so it may be possible to adapt the techniques of [4] (exploiting concave separability), though we have to deal with the multiplication of distance variables z_{ij} by 0/1 variables y_{ij} in the objective function of MMX.

Another view is that $-z^2 + w \leq 0$ is equivalent to $\sqrt{w} \leq z$, which is manifested as the second-order cone constraint $\|x^i - x^j\|^2 \leq z_{ij}$. We can try to exploit methods for handling such constraints, though we have to deal with the multiplication of distance variables z_{ij} by 0/1 variables y_{ij} in the objective function.

3.2 Shifting

A simple fix is to approximate $\sqrt{w}$ by $h(w) := \sqrt{w+\delta} - \sqrt{\delta}$ for some small $\delta > 0$. Then we underestimate all positive distances (via the triangle inequality). Because our objective function is increasing in each distance, we get a *relaxation* of MMX. But a strong downside is that we underestimate *all* positive distances, and the error in each distance calculation rapidly approaches $\sqrt{\delta}$ as w increases.

3.3 Linear extrapolation

The following approximation, depending on the choice of a $\sigma > 0$, was proposed in [10] to avoid non-differentiability of the Euclidean-distance function for the "traveling-salesman problem with neighborhoods."

$$l(w) := \begin{cases} \sqrt{w}, & \text{if } w \geq \sigma^2; \\ \frac{\sigma}{2} + \frac{1}{2\sigma} w, & \text{if } w \leq \sigma^2. \end{cases}$$

The function l is well-defined at $w = \sigma^2$. It is analytic except when $w = \sigma^2$, and in this case, it is still differentiable once. In fact, l simply uses the tangent at $w = \sigma^2$ of the graph of the strictly concave function $\sqrt{w}$ to overestimate $\sqrt{w}$ on $[0, \sigma^2)$. We already see that because l is not twice continuously differentiable, we should not expect good behavior from most NLP solvers. A strong shortcoming for our context is that l miscalculates zero distances; that is, $l(0) = \sigma/2$, while obviously $\sqrt{0} = 0$. Because $l(w)$ is an upperbound on $\sqrt{w}$, and because our objective function is increasing in distances, using the approximation l, we do *not* get a relaxation of MMX. Moreover, for degenerate Steiner trees, we will systematically overestimate distances that should be zero.

3.4 Smooth under-estimation

We propose another piecewise smoothing, using a particular homogeneous cubic depending on the choice of a $\lambda > 0$, that has very nice properties:

$$c(w) := \begin{cases} \sqrt{w}, & \text{if } w \geq \lambda^2; \\ \frac{15}{8\lambda} w - \frac{5}{4\lambda^3} w^2 + \frac{3}{8\lambda^5} w^3, & \text{if } w \leq \lambda^2. \end{cases}$$

We have depicted all of these smoothings in Fig. 1.

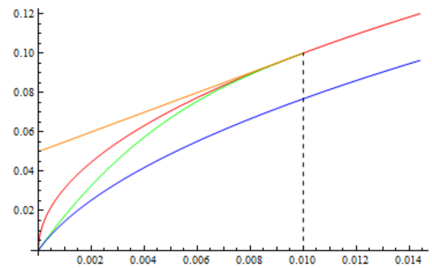

Fig. 1: Behavior of all smoothings ($\lambda^2 = \sigma^2 = 0.01$):

- the red curve is the true $\sqrt{\cdot}$ function.
- the "smooth underestimation" c, which we advocate, follows the cubic green curve below $w = 0.01$.
- the "linear extrapolation" l follows the orange line below $w = 0.01$.
- both piecewise-defined functions c and l follow the true $\sqrt{\cdot}$ function above $w = 0.01$.
- the "shift" h (with $\delta = (4\lambda/15)^2$ chosen so that it has the same derivative as c does at 0) follows the consistent underestimate given by the blue curve.

The next result makes a very strong case for the smoothing c.

Thm. 1.

1. *$c(w)$ agrees with $\sqrt{w}$ in value at $w = 0$;*
2. *$c(w)$ agrees with $\sqrt{w}$ in value, derivative and second derivative at $w = \lambda^2$; hence c is twice continuously differentiable;*
3. *c is strictly concave on $[0, +\infty]$ (just like $\sqrt{\cdot}$);*
4. *c is strictly increasing on $[0, +\infty]$ (just like $\sqrt{\cdot}$); consequently, $c(\|x^i - x^j\|^2)$ is quasiconvex (just like $\|x^i - x^j\|$);*
5. *$\sqrt{w} - c(w) \geq 0$ on $[0, +\infty]$;*
6. *For all $\lambda > 0$, $\max\{(\sqrt{w} - c(w))/\lambda \ : \ w \in [0, +\infty]\}$ is the real root γ of $-168750 + 1050625x + 996300x^2 + 236196x^3$, which is approximately 0.141106.*

Because distances only appear in the objective, and in a very simple manner, the approximation with c is a *relaxation* (due to (5) of Thm. 1). Meaning that the objective value of a global optimum using distance approximation is a lower bound on the true global optimum. And plugging the obtained solution into the true objective function gives an upper bound on the value of a true global optimum. So, in the end we get a solution and a bound on how close to optimal we can certify it to be. To be precise, for any $\lambda > 0$, let MMX(λ) denote MMX with all square roots in the norms replaced by the function c.

Cor. 2. *The optimal value of MMX is between the optimal value of MMX(λ) and the optimal value of MMX(λ) plus $\lambda\gamma(2p - 3)$ ($\gamma \approx 0.141106$).*

We note that the upper bound of Cor. 2 is a very pessimistic worst case — only achievable when the optimal tree has a full Steiner topology and all edges are very short. If such were the case, certainly λ should be decreased. Furthermore, we emphasize that if the length of every edge in the SMT that solves MMX(λ) is either zero (degenerate case) or greater than or equal to λ, then the optimal values of MMX and MMX(λ) are the same.

For $\delta = (4\lambda/15)^2$, we have $c'(0) = h'(0)$. That is, at this value of δ we can expect that c and h will have the same numerical-stability properties near 0.

Prop. 3. *For $\delta = (4\lambda/15)^2$, we have $h(w) < c(w)$ on $(0, +\infty)$. Moreover $c(w) - h(w)$ is strictly increasing on $[0, +\infty)$.*

Hence the relaxation provided using c is always at least as strong as the relaxation provided using h. In fact, strictly stronger in any realistic case (specifically, when there is more than 1 terminal). We make a few further observations about c:

Choosing λ. Note that $c'(0) = \frac{15}{8\lambda}$, so we should not be too aggressive in picking λ extremely small. But a large derivative at 0 is the price that we pay for getting everything else; in particular, any concave function f that agrees with $\sqrt{w}$ at $w = 0$ and $w = \lambda^2$ has $f'(0) \geq \frac{1}{\lambda}$. By Cor. 2, choosing λ to be around $2p-1$ would seem to make sense, thus guaranteeing that our optimal solution of MMX(λ) is within a universal additive constant of the optimal value of MMX. If the points are quite close together, either they should be scaled up or λ can be decreased.

Secant lower bound. Owing to the strict concavity, in the context of an sbb algorithm, c is always best lower bounded by a secant on *any* subinterval. Of course, it can be an issue whether a sbb solver can recognize and exploit this. In general, such solvers (e.g., `Baron` and `Couenne`) do not yet support general piecewise-smooth concave functions, while conceptually they could *and should*. However, we implemented a feature in `SCIP` (version 3.2) that allowed us to solve the proposed relaxation. In particular, by using the modeling-language `AMPL` interface and its general suffix facility (see [8]), we are able to specify to `SCIP` to treat a piecewise-defined constraint as globally concave. With this new `SCIP` feature we were able to solve the secant relaxation — see Section 5.

4 Tightening relaxations

4.1 Integer extreme points

Here, we examine the continuous relaxation of the feasible region of MMX, just in the space of the y-variables. That is, the set of $y_{ij} \in [0, 1]$ satisfying (2-4). The coefficient matrix of the system (2-4) is *not* totally unimodular (TU). However, for each $j \in S - \{p + 1\}$, we can subtract the equation of (4) from the corresponding equation of (3) to arrive at the system:

$$\sum_{j \in S} y_{ij} = 1, \quad \text{for } i \in P, \tag{6}$$

$$\sum_{i \in P} y_{ij} + \sum_{k>j, k \in S} y_{jk} = \begin{cases} 3, \text{ for } j = p+1, \\ 2, \text{ for } j \in S - \{p+1\}, \end{cases} \tag{7}$$

$$\sum_{k<j, k \in S} y_{kj} = 1, \quad \text{for } j \in S - \{p+1\}. \tag{8}$$

This resulting system is the set of constraints for the ordinary formulation of a *bipartite* 0/1 "b-matching" problem. Such a formulation has a TU constraint matrix. So we immediately have the following theorem (also observed in [13, §4.3, pp. 217-9] via a much more complicated and less revealing proof) and corollaries:

Thm. 4. *The set of $y_{ij} \in [0,1]$ satisfying (2-4) has integer extreme points.*

Cor. 5. *No valid linear inequality in the y-variables alone can improve the linear relaxation of the equations (2-4) describing full Steiner topologies.*

This is not to say that *optimality*-based (linear) inequalities in the y-variables alone cannot be derived (see §4.2.2).

Cor. 6. *Given a globally-optimal solution of the continuous relaxation of MMX, in polynomial time we can calculate a globally-optimal solution of MMX.*

The relevant b-matching problem is depicted in Fig. 2. There is a node for every constraint of (6-8) and an edge for every variable. To the right of each node is its required degree. All possible edges exist between nodes at levels (6) and (7); i.e., those node sets induce a complete bipartite graph. All possible edges extending "down to the right" exist between nodes at levels (7) and (8); i.e., between nodes at level (7) and all nodes of greater number at level (8). A feasible choice of edges (selecting a full Steiner topology) is one that meets the degree requirements. E.g., we can see that we must choose the edge between node $p+1$ of level (7) and node $p+2$ of level (8).

4.2 Geometric cuts

Based on various geometric considerations concerning optimal solutions, we can derive several families of valid inequalities seeking to improve relaxations of our formulation. Some of these are nonlinear, and it is an ongoing challenge to take advantage of them computationally. Until we address it in §4.2.3, *assume that all norms are taken exactly — not smoothed.*

Let η_i be the distance from terminal a^i to the nearest other terminal; i.e.,

$$\eta_i := \min_{j \in P,\, j \neq i} \{\|a^i - a^j\|\}, \ \forall\ i \in P. \tag{9}$$

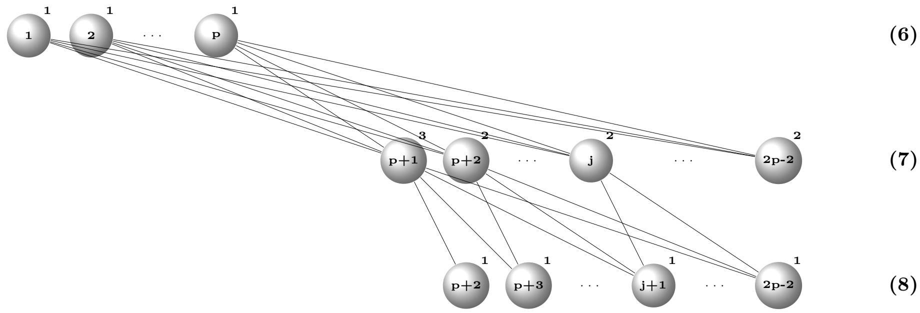

Fig. 2: The bipartite b-matching model for selecting a full Steiner topology

4.2.1 Non-combinatorial cuts.

Thm. 7. *For all $n \geq 2$, we have*

$$y_{ik}\left(\|x^k - a^i\|\right) \leq \eta_i, \ \forall\ i \in P,\ k \in S. \tag{10}$$

Lem. 8. *Among triangles with edge lengths a, b, c and corresponding angles x,y,z, with c and z fixed, the one maximizing $a + b$ is isosceles (that is $a = b$, $x = y$).*

Thm. 9. *For all $n \geq 2$, we have*

$$y_{ik}y_{jk}\left(\|x^k - a^i\| + \|x^k - a^j\|\right) \leq \tfrac{2}{\sqrt{3}}\|a^i - a^j\|, \ \forall\ i,j \in P,\ i < j,\ k \in S. \tag{11}$$

In computations, we treat the bilinear term $y_{ik}y_{jk}$ by replacing it with a variable y_{ijk} and using the standard McCormick inequalities.

Another way to try and use the same principle is as follows:

Thm. 10. *For all $n \geq 2$, we have*

$$y_{ik}y_{jk}\|x^k - a^i\|\|x^k - a^j\|\left(\|x^k - a^i\| - \tfrac{1}{\sqrt{3}}\|a^i - a^j\|\right) = 0, \ \forall\ i,j \in P,\ i < j,\ k \in S. \tag{12}$$

Thms. 9 and 10 easily extends to the case where the Steiner point x^k is adjacent to only 1 terminal in the SMT and also to the case where x^k is not adjacent to any terminal. In the such cases, we have

$$y_{ik}y_{kl}\left(\|x^k - a^i\| + \|x^k - x^l\|\right) \leq \tfrac{2}{\sqrt{3}}\|a^i - x^l\|, \ \forall\ i \in P,\ k,l \in S,\ k < l; \tag{13}$$

$$y_{kl}y_{km}\left(\|x^k - x^l\| + \|x^k - x^m\|\right) \leq \tfrac{2}{\sqrt{3}}\|x^l - x^m\|, \ \forall\ k,l,m \in S,\ k < l < m. \tag{14}$$

Via another geometric principle, we have the following result.

Thm. 11. *For $n = 3$ and $i, j \in P,\ i < j,\ k, l \in S, k < l$, we have*

$$y_{ik}y_{jk}y_{kl} \cdot \det\begin{bmatrix} a^i & a^j & x^k & x^l \\ 1 & 1 & 1 & 1 \end{bmatrix} = 0. \tag{15}$$

We note that the determinant of the 4×4 matrix in Thm. 11 is quadratic in the x-variables — actually bilinear between x^k and x^l. In computations, we treat the trilinear term $y_{ik}y_{jk}y_{kl}$ by replacing it with a variable y_{ijkl} and using the standard inequalities

$$y_{ijkl} \le y_{ik}, \quad y_{ijkl} \le y_{jk}, \quad y_{ijkl} \le y_{kl},$$
$$y_{ik} + y_{jk} + y_{kl} \le 2 + y_{ijkl}.$$

Thm. 11 easily extends to dimensions $n > 3$.

Thm. 12. *For $n \ge 3$, $i,j \in P$, $i < j$, $k,l \in S, k < l$, and for every 3×4 submatrix $B = [b^i, b^j, \xi^k, \xi^l]$ of the $n \times 4$ matrix $\left[a^i, a^j, x^k, x^l\right]$, we have*

$$y_{ik} y_{jk} y_{kl} \cdot \det \begin{bmatrix} b^i & b^j & \xi^k & \xi^l \\ 1 & 1 & 1 & 1 \end{bmatrix} = 0. \tag{16}$$

4.2.2 Combinatorial cuts.

In this section, we introduce some valid linear inequalities satisfied by optimal topologies of the ESTP.

Thm. 13. *For $n \ge 2$ and $i,j \in P$, $i < j$, we have*

$$\text{If } \|a^i - a^j\| > \eta_i + \eta_j, \text{ then } y_{ik} + y_{jk} \le 1, \ \forall\ k \in S. \tag{17}$$

Considering the valid inequalities in (17), we note that the inequalities (11) can only be active for $i,j \in P$, $i < j$, $k \in S$, such that $\|a^i - a^j\| \le \eta_i + \eta_j$. Therefore, only such valid inequalities should be included in MMX. Furthermore, the result in Thm. 13 extends to the more general case addressed in Thm. 14.

Thm. 14. *Let H be a graph with vertex set P, and such that i and j are adjacent if $\|a^i - a^j\| > \eta_i + \eta_j$. Then for each $k \in S$, the set of $i \in P$ such that $y_{ik} = 1$ in an optimal solution is a stable set of H. Therefore, every valid linear inequality $\sum_{i \in P} \alpha_i y_i \le \rho$ for the stable set polytope of H yields a valid linear inequality $\sum_{i \in P} \alpha_i y_{ik} \le \rho$ for each $k \in S$.*

Let T be a min-length spanning tree on terminals a^i, $i \in P$. For $i,j \in P$, let

$$\beta_{ij} := \text{length of the longest edge on the path between } a^i \text{ and } a^j \text{ in } T. \tag{18}$$

The proof of the following well-known lemma can be found for example in [12]. We use it to prove the validity of cuts presented in Cor. 16.

Lem. 15. *An SMT contains no edge of length greater than β_{ij} on the unique path between a^i and a^j, for all $i,j \in P$.*

Cor. 16. *For $n \ge 2$ and $i,j \in P$, $i \ne j$, we have*

$$\text{If } \|a^i - a^j\| > \eta_i + \eta_j + \beta_{ij}, \text{ then } y_{ik} + y_{kl} + y_{jl} \le 2, \ \forall\ k,l \in S, \ k < l. \tag{19}$$

4.2.3 Smoothing in the context of cuts.

In the definition and derivation of all cuts, we assumed that norms are taken exactly — not smoothed. Now, we confront the issue that we prefer to work with smoothed norms in the context of mathematical optimization.

First of all, for any norm that just involves data and no variables, we do *not* smooth. This pertains to all occurrences of $\|a^i - a^j\|$, and hence also all occurrences of the parameters defined in (9) and (18). Any valid equation or inequality based *only* on such use of norms (or not based on norms at all) is valid for the original problem. So, these equations and inequalities do not exclude optimal solutions of the original problem, and so these solutions are candidate solutions to the problem where distances are smoothly underestimated (employing h or c), possibly with lower objective value. Therefore, any such equation or inequality is valid for the problem with distances smoothly underestimated. This applies to (15), (16), (17), inequalities based on Thm. 14, and (19).

For (10) and (11), norms involving variables occur only on the low side of the inequalities, so smooth underestimation of these norms, employing h or c, keeps them valid.

Inequalities (13,14) also contain norms involving variables on the high side of those inequalities. For those norms, we should replace them with smooth *overestimates*. For example, we could use an "overestimating shift" $\hat{h}(w) := \sqrt{w+\delta}$, or the linear extrapolation l (possibly with a different breakpoint). In fact, by choosing the breakpoint for l at $\sigma^2 := (4\lambda/15)^2$, where λ^2 is the breakpoint for c, we get $c'(0) = l'(0)$ $(= 15/8\lambda)$. That is, if we can numerically tolerate a breakpoint for the underestimate c at $w = \lambda^2$, then we can equally tolerate a breakpoint for the overestimate l at $w = \sigma^2 = (4\lambda/15)^2$. While global sbb solvers do not yet support piecewise functions, at the modeling level we could utilize tangents of the concave $\sqrt{\cdot}$ at a few values of w greater than $(4\lambda/15)^2$. I.e., choose some values $\sigma_r^2 > \cdots > \sigma_1^2 > \sigma_0^2 := (4\lambda/15)^2$, let $\tau_{\sigma_i}(w) := \frac{\sigma_i}{2} + \frac{1}{2\sigma_i}w$, and we can instantiate (13) and (14) $r+1$ times each, replacing the $\sqrt{\cdot}$ implicit on the high side of these inequalities with τ_{σ_i}, for $i = 0, 1, \ldots, r$.

Finally, we have the *equation* (12). The norms $\|x^k - a^i\|$ and $\|x^k - a^j\|$ can be smoothed in any way that correctly evaluates the norm at 0, and the equation remains valid. So employing h or c leaves the equation valid. The norm involving x^k in the multiplicand $\left(\|x^k - a^i\| - \|a^i - a^j\|/\sqrt{3}\right)$ is thornier. One way to address it is to simply replace it with $\left(\|x^k - a^i\|^2 - \|a^i - a^j\|^2/3\right)$, with these (smooth) squared norms calculated exactly.

5 Experiments

In this section, we demonstrate the impact of our cuts on the solution of the ESTP, with results on 25 instances in $\mathbb{R}^3$, where the terminals are randomly distributed in the cube $[0, 10]^3$. For each value of p from 4 to 8, we created five instances. We show the effect of the cuts by solving the instances a few times with the sbb code `SCIP`, adding different cuts each time to MMX. In our experiments, we considered cuts (10,11, 13–15, 17). We compare the performance of `SCIP` on eight models, running with a time limit of 2 hours. The first model is MMX, with no cuts, the following six are MMX with the independent addition of cuts, and the last model is MMX with the addition of the three most effective cuts on the previous experiments.

Summarizing our results, we have that adding the cuts have significant effect on improving the lower bounds and decreasing the running time. The three classes of cuts, (10), (14) and (17) together improve the lower bound computed in the time limit, in the only two instances still not solved to optimality in 51% and 52%, and decreased the overall running time in 61% on average. Three extra instances, all with $p = 8$, were solved to optimality in the time limit, after the addition of the cuts. In Table 1, we show for each value of p, the average percentage improvements on the running time of each model tested, when compared to MMX (100(Time(MMX) − Time(Model))/Time(MMX)). (Negative values indicate worse times with the addition of the cuts)

p	MMX+						
	(10)	(11)	(13)	(14)	(15)	(17)	(10,14,17)
4	51	2	20	0	-21	24	41
5	44	-10	21	0	-82	4	56
6	92	72	91	88	-462	93	93
7	81	-274	48	56	-654	76	84
8	29	0	0	6	0	15	32

Table 1: Average % improvements on running time compared to MMX

Although these are still preliminary results, we see that our cuts can potentially improve the quality of the lower bounds computed by SCIP. The two other cuts proposed (16,19) have a positive effect in instances with more terminals or in higher dimensions. Finally, we see that using several families of cuts together can bring further improvements.

6 Conclusions

The ESTP is very difficult to solve for $n > 2$, and the application of sbb solvers to test instances using the MMX formulation points out considerable drawbacks, concerned with the non-differentiability of the Euclidean norm, and also to the extreme weakness of the lower bounds given by relaxations. MMX in its original form, with today's sbb solvers, leads to dismal results.

We presented different approximations for the square-roots — the source of our non-differentiability, which can be judiciously applied in the context of valid inequalities. In particular, we introduced a smooth underestimation function, and we established several very appealing properties. We implemented this smoothing with a new feature of SCIP that we developed. This feature could be specialized to automatically smooth roots and other power functions.

To improve the quality of lower bounds computed in an sbb algorithm, we presented a variety of valid inequalities, based on known geometric properties of a SMT. Preliminary numerical experiments demonstrates the potential of these cuts in improving lower bounds. Many of these cuts are nonconvex, and it is an interesting research direction to apply similar ideas to other nonconvex gloabl-optimization models. We demonstrated that the performance of the MMX model can be significantly improved, though it is still not the best method for ESTP with $n > 2$. So more work is needed to make the MMX approach competitive.

Acknowledgments

J. Lee was partially supported by NSF grant CMMI–1160915 and ONR grant N00014-14-1-0315, and Laboratoire d'Informatique de l'École Polytechnique. M. Fampa was partially supported by CNPq and FAPERJ.

References

1. T. Achterberg. SCIP: Solving constraint integer programs. *Mathematical Programming Computation*, 1(1):1–41, 2009.
2. C. Bragalli, C. D'Ambrosio, J. Lee, A. Lodi, and P. Toth. On the optimal design of water distribution networks. *Optim. Eng.*, 13(2):219–246, 2012.
3. L. Cavalli-Sforza and A. Edwards. Phylogenetic analysis: Models and estimation procedures. *Evolution*, 21:550–570, 1967.
4. C. D'Ambrosio, J. Lee, and A. Wächter. A global-optimization algorithm for mixed-integer nonlinear programs having separable non-convexity. In *Algorithms—ESA 2009*, volume 5757 of *LNCS*, pages 107–118. Springer, Berlin, 2009.
5. D. Du and X. Hu. *Steiner tree problems in computer communication networks*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2008.
6. 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.
7. M. Fampa and N. Maculan. Using a conic formulation for finding Steiner minimal trees. *Numerical Algorithms*, 35(2-4):315–330, 2004.
8. R. Fourer, D. M. Gay, and B. W. Kernighan. *AMPL: A Modeling Language for Mathematical Programming*. Duxbury Press, Nov. 2002.
9. M. Garey, R. Graham, and D. Johnson. The complexity of computing Steiner minimal trees. *SIAM J. Applied Mathematics*, 1977.
10. I. Gentilini, F. Margot, and K. Shimada. The travelling salesman problem with neighbourhoods: MINLP solution. *Opt. Meth. and Soft.*, 28(2):364–378, 2013.
11. E. Gilbert and H. Pollack. Steiner minimal trees. *SIAM J. Applied Math.*, 1968.
12. F. Hwang, D. Richards, and W. Winter. *The Steiner tree problem*, volume Ann. of Disc. Math. 53. Elsevier, Amsterdam, 1992.
13. N. Maculan, P. Michelon, and A. Xavier. The Euclidean Steiner tree problem in R^n: A mathematical programming formulation. *Ann. OR*, 96(1-4):209–220, 2000.
14. W. D. Smith. How to find Steiner minimal trees in Euclidean d-space. *Algorithmica*, 7(1-6):137–177, 1992.
15. J. W. Van Laarhoven and K. M. Anstreicher. Geometric conditions for Euclidean Steiner trees in R^d. *Comput. Geom.*, 46(5):520–531, 2013.
16. S. Vigerske. *Decomposition of multistage stochastic programs and a constraint integer programming approach to MINLP*. PhD thesis, Humboldt-U. Berlin, 2013.
17. A. Wächter and L. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale NLP. *Math. Prog.*, 106:25–57, 2006.
18. D. Warme, P. Winter, and M. Zachariasen. Exact algorithms for plane Steiner tree problems: a computational study. *Advances in Steiner Trees*, pages 81–116, 1998.

Appendix 1: Proofs

Proof of Thm. 1:

Proof. (1) is trivial. (2) is verified by simple calculus. For (3), we can verify that $c''(w) < 0$ for $w < (10/9)\lambda^2$. For (4), because c is strictly concave, for $w < \lambda^2$ we have $c'(w) > c'(\lambda^2) = \frac{1}{2\lambda} > 0$. Consequently, $c(\|x^i - x^j\|^2)$ is the composition of the increasing function c with the (quasi)convex function $\|x^i - x^j\|^2$, so it is quasiconvex.

For (5), we reparameterize with $t := \sqrt{w}$, letting $\tilde{c}(\sqrt{w}) := c(w)$; so we just need to show that the polynomial $t - \tilde{c}(t)$ is nonnegative on $[0, \lambda]$. It is easy to check that

$$t - \tilde{c}(t) = t(\lambda - t)^3 \left(\frac{8\lambda^2 + 9\lambda t + 3t^2}{8\lambda^5} \right),$$

and each of these factors is obviously nonnegative for $t \in [0, \lambda]$, and hence so is $t - \tilde{c}(t)$.

For (6), it is very convenient to observe that we can reparameterize by $f := t/\lambda$, and we then restrict our attention to $f \in [0, 1]$. Then

$$(t - \tilde{c}(t))/\lambda = f - 15f^2/8 + 5f^4/4 - 3f^6/8.$$

Taking the derivative with respect to f, we find that we get one real root:

$$\frac{1}{9} \left(-6 + \sqrt[3]{15\left(9 - \sqrt{66}\right)} + \sqrt[3]{15\left(9 + \sqrt{66}\right)} \right)$$

(approximately 0.301825). Plugging this exact expression of the root back into $f - 15f^2/8 + 5f^4/4 - 3f^6/8$ and using `Mathematica` (to calculate the minimal polynomial of the exact expression of the resulting algebraic number), we get the result. □

Proof of Prop. 3:

Proof. Clearly the result is true for $w \geq \lambda^2$, so we confine out attention to $0 < w < \lambda^2$. Let $t := w/\lambda^2$, so $0 < t < 1$. Then it is easy to check that

$$h(t\lambda^2) = \lambda \left(\sqrt{t + \left(\frac{4}{15}\right)^2} - \frac{4}{15} \right),$$

and

$$c(t\lambda^2) = \lambda \left(\frac{15}{8}t - \frac{5}{4}t^2 + \frac{3}{8}t^3 \right).$$

Therefore, we have

$$c(t\lambda^2) - h(t\lambda^2) = \lambda \left(\frac{15}{8}t - \frac{5}{4}t^2 + \frac{3}{8}t^3 - \sqrt{t + \left(\frac{4}{15}\right)^2} + \frac{4}{15} \right).$$

So, it remains to demonstrate that

$$\frac{15}{8}t - \frac{5}{4}t^2 + \frac{3}{8}t^3 - \sqrt{t + \left(\frac{4}{15}\right)^2} + \frac{4}{15} \tag{20}$$

is positive — we see that the result does not depend on the choice of λ. Now, we resort to `Mathematica` to check that (20) has no zero on $(0,1)$. Seeing that (20) is positive at $t=1$ and continuous on $(0,1]$, this implies that it is positive on all of $(0,1)$.

Furthermore, checking that the derivative of (20) is positive at $t=1$ and has no zero on $(0,1)$, we can conclude that $c-h$ is strictly increasing on $[0,+\infty)$. □

<u>Proof of Thm. 7</u>:

Proof. Suppose $y_{ik}=1$ and $\|x^k - a^i\| > \eta_i$. Then it is possible to construct a tree with smaller length, by disconnecting a^i and x^k, and connecting a^i to the nearest terminal. □

<u>Proof of Lem. 8</u>:

Proof. Using the law of sines, we have

$$a = \sin x \ (c/\sin z);$$
$$b = \sin y \ (c/\sin z).$$

So, $a+b = (c/\sin z)(\sin x + \sin y)$.

Differentiating with respect to x (c and z are fixed, and y is an implicit function of z; that is, $y = 180 - x - z$), we get $(c/\sin z)(\cos x - \cos y)$ (note the use of the implicit chain rule dealing with y). Setting this last expression to 0, we get that the only stationary point has $\cos x = \cos y$, so $x = y = (180-z)/2$. The second derivative is $(c/\sin z)(-\sin x - \sin y)$, which is negative for $x \in [0,180]$ ($y = 180 - x - z$), so $x = y = (180-z)/2$ is indeed the maximizer. We have then that is the optimizing triangle is isosceles. □

<u>Proof of Thm. 9</u>:

Proof. First, we consider the validity when the Steiner point x^k does *not* degenerate to either a^i or a^j. Then we know that the angle between the edges connecting the Steiner point x^k to the terminals a^i or a^j is 120 degrees. Applying Lem. 8, with $z=120$ and $c = \|a^i - a^j\|$, and using a bit of plane geometry, we see that the choice of x^k maximizing $\|x^k - a^i\| + \|x^k - a^j\|$ has

$$\|x^k - a^i\| = \|x^k - a^j\| = \|a^i - a^j\|/\sqrt{3}.$$

So we can easily see that the inequality is valid (and even tight in this case).

Next, suppose, without loss of generality, that the Steiner point x^k degenerates to a^i. Then $\|x^k - a^i\| + \|x^k - a^j\| = \|x^k - a^j\|$, and the inequality of the theorem holds because $1 \le 2/\sqrt{3}$. □

Proof of Thm. 10:

Proof. The equation enforces that when terminals a^i and a^j are connected to the same Steiner point x^k, and if x^k does not degenerate to either of these two terminals, then $\|a^i - a^j\|/\|x^k - a^i\| = \sqrt{3}$ — that is, the angle of the edges connecting the terminals to the Steiner point is 120 degrees. □

Proof of Thm. 11:

Proof. When a Steiner point x^k is connected to a pair of terminals a^i and a^j and Steiner point x^l, all 4 points must lie in the same plane. The 4 points a^i, a^j, x^k, x^l are affinely dependent — i.e., lie on a plane – precisely when the determinant is zero. □

Proof of Thm. 13:

Proof. We have $\|x^k - a^i\| + \|x^k - a^j\| \ge \|a^i - a^j\|$. Suppose $\|a^i - a^j\| > \eta_i + \eta_j$ and $y_{ik} = y_{jk} = 1$. Then it is possible to construct a tree with smaller length, by removing the connections between a^i and x^k and also between a^j and x^k, and adding to the tree the connections between a^i and the nearest terminal to it, and also between a^j and the nearest terminal to it. □

Proof of Cor. 16:

Proof. The result comes directly from Thm. 7 and Lem. 15, which imply that there are at least 3 Steiner points (or 4 edges) in the path between a^i and a^j in a SMT, if $\|a^i - a^j\| > \eta_i + \eta_j + \beta_{ij}$. □

Appendix 2: Additional details about our experiments

All experiments were run on a 64-bit Intel(R) Xeon(R) CPU E5-2609 2 processor running at 2.50GHz with 10240 KB cache.

In Table 2, we show the number of cuts that we utilized for our instances.

In Table 3, we compare the performance of `SCIP` on eight models, running with a time limit of 2 hours. The first model is MMX, with no cuts, and the following six are model MMX with the independent addition of cuts (10), (11), 13, 14, 15, and (17), and the last model is MMX with the simultaneous addition of the cuts (10), (14) and (17). The first and second columns in Table 3 correspond to the number of terminals p and to the number of the instance, respectively. The other columns report, for each model, the value of the percentage duality gap (Gap) (100(UB-LB)/UB), and the running time in seconds (Time).

p	inst	(10)	(11)	(13)	(14)	(15)	(17)
4	1	8	12	4	0	6	0
4	2	8	12	4	0	6	6
4	3	8	12	4	0	6	0
4	4	8	12	4	0	6	2
4	5	8	12	4	0	6	0
5	1	15	30	15	1	30	6
5	2	15	30	15	1	30	0
5	3	15	30	15	1	30	6
5	4	15	30	15	1	30	0
5	5	15	30	15	1	30	6
6	1	24	60	36	4	90	8
6	2	24	60	36	4	90	8
6	3	24	60	36	4	90	20
6	4	24	60	36	4	90	16
6	5	24	60	36	4	90	32
7	1	35	105	70	10	210	60
7	2	35	105	70	10	210	50
7	3	35	105	70	10	210	10
7	4	35	105	70	10	210	40
7	5	35	105	70	10	210	40
8	1	48	168	120	20	420	42
8	2	48	168	120	20	420	54
8	3	48	168	120	20	420	66
8	4	48	168	120	20	420	90
8	5	48	168	120	20	420	84

Table 2: Number of cuts of each type utilized for each instance

		MMX		**MMX+**													
				(10)		(11)		(13)		(14)		(15)		(17)		**(10,14,17)**	
p	Inst	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time
4	1	0.00	0.39	0.00	0.24	0.00	0.22	0.00	0.11	0.00	0.38	0.00	0.42	0.00	0.38	0.00	0.23
4	2	0.00	0.32	0.00	0.10	0.00	0.38	0.00	0.38	0.00	0.33	0.00	0.41	0.00	0.21	0.00	0.25
4	3	0.00	0.37	0.00	0.17	0.00	0.39	0.00	0.11	0.00	0.36	0.00	0.37	0.00	0.37	0.00	0.16
4	4	0.00	0.37	0.00	0.08	0.00	0.40	0.00	0.36	0.00	0.37	0.00	0.38	0.00	0.10	0.00	0.11
4	5	0.00	0.20	0.00	0.22	0.00	0.22	0.00	0.36	0.00	0.21	0.00	0.42	0.00	0.20	0.00	0.22
5	1	0.00	0.81	0.00	0.71	0.00	1.94	0.00	1.58	0.00	0.61	0.00	3.76	0.00	0.64	0.00	0.57
5	2	0.00	1.54	0.00	0.99	0.00	1.66	0.00	0.76	0.00	1.69	0.00	1.38	0.00	1.57	0.00	0.93
5	3	0.00	1.04	0.00	0.94	0.00	1.33	0.00	1.25	0.00	0.84	0.00	2.23	0.00	0.84	0.00	0.50
5	4	0.00	1.00	0.00	0.66	0.00	1.67	0.00	0.87	0.00	1.23	0.00	4.83	0.00	1.01	0.00	0.88
5	5	0.00	2.93	0.00	0.83	0.00	1.46	0.00	1.35	0.00	2.95	0.00	1.11	0.00	3.00	0.00	0.34
6	1	0.00	11.81	0.00	9.64	0.00	29.26	0.00	9.62	0.00	11.83	0.00	1356.32	0.00	7.10	0.00	10.07
6	2	0.00	501.66	0.00	13.00	0.00	29.75	0.00	12.59	0.00	20.59	0.00	112.06	0.00	16.21	0.00	12.83
6	3	0.00	14.03	0.00	7.81	0.00	31.97	0.00	8.50	0.00	10.15	0.00	88.30	0.00	7.71	0.00	5.64
6	4	0.00	10.16	0.00	6.65	0.00	26.78	0.00	10.48	0.00	8.68	0.00	1481.49	0.00	6.21	0.00	5.93
6	5	0.00	14.73	0.00	5.41	0.00	35.11	0.00	7.94	0.00	13.77	0.00	66.91	0.00	4.15	0.00	3.60
7	1	0.00	201.24	0.00	95.05	0.00	1545.95	0.00	259.33	0.00	316.73	72.46	7200.00	0.00	74.63	0.00	44.12
7	2	0.00	187.95	0.00	81.60	0.00	5758.72	0.00	853.23	0.00	189.53	68.86	7200.00	0.00	99.00	0.00	64.86
7	3	0.00	1209.66	0.00	417.65	0.00	1396.01	0.00	649.26	0.00	496.32	74.69	7200.00	0.00	397.51	0.00	380.30
7	4	0.00	379.91	0.00	146.05	0.00	2215.55	0.00	254.50	0.00	236.76	82.91	7200.00	0.00	102.33	0.00	102.95
7	5	0.00	2795.34	0.00	171.68	0.00	6941.09	0.00	452.47	0.00	848.04	100.00	7200.01	0.00	472.54	0.00	174.64
8	1	78.00	7200.00	55.31	7200.00	100.00	7200.01	66.87	7200.00	55.92	7200.00	100.00	7200.04	59.21	7200.01	37.18	7200.00
8	2	82.78	7200.00	56.16	7200.00	100.00	7200.01	70.60	7200.00	76.36	7200.00	100.00	7200.00	83.60	7200.01	39.96	7200.00
8	3	54.14	7200.00	0.00	4819.10	100.00	7200.00	61.22	7200.00	47.47	7200.01	100.00	7200.01	56.40	7200.00	0.00	4238.69
8	4	45.19	7200.01	0.00	2371.79	100.00	7200.00	37.99	7200.00	0.00	4950.73	100.00	7200.02	0.00	1978.05	0.00	1369.08
8	5	100.00	7200.00	0.00	4058.00	100.00	7200.01	51.76	7200.00	100.00	7200.00	100.00	7200.01	84.72	7200.00	0.00	4353.07

Table 3: Effect of adding cuts to MMX