

Quasilinear Approximation Scheme for Steiner Multi Cycle in the Euclidean plane^{1,2}

Carla N. Lintzmayer

Center for Mathematics, Computing and Cognition. Federal University of ABC. Santo André, São Paulo, Brazil

Flávio K. Miyazawa Phablo F. S. Moura³ Eduardo C. Xavier

Institute of Computing. University of Campinas. Campinas, São Paulo, Brazil

Abstract

We propose a randomized approximation scheme for the Euclidean Steiner Multi Cycle problem which runs in quasilinear time. In this problem, we are given a set of n pairs of points (terminals) $\mathcal{T} = \{\{t_i, t'_i\}\}_{i=1}^n$ in the Euclidean plane, and the objective is to find a collection of cycles of minimum cost such that t_i and t'_i belong to a same cycle, for each $i \in \{1, \dots, n\}$. This problem extends the Steiner Cycle problem in the same way the Steiner Forest extends the Steiner Tree problem. Additionally, it has applications on routing problems with pickup and delivery locations.

Keywords: Euclidean plane, geometry, Steiner, approximation algorithm, quasilinear.

1 Introduction

The Steiner Multi Cycle Problem was proposed by Pereira et al., in the context of a routing problem with pickup and delivery locations, as a natural generalization of the Steiner Cycle problem [4]. In this problem we are given a complete weighted graph G which respects the triangle inequality and a collection of terminal sets $\{T_1, \dots, T_k\}$ such that $T_i \subseteq V(G)$ for every $i \in [k] := \{1, \dots, k\}$, and these terminal sets are pairwise disjoint. The problem is to find a set of vertex disjoint

¹ The authors thank the São Paulo Research Foundation, FAPESP, (grants #2015/11937-9, #2016/01860-1, #2016/23552-7, #2017/22611-2, and #2016/21250-3), the Brazilian National Council for Scientific and Technological Development, CNPq, (grants #314366/2018-0, #425340/2016-3, and #304856/2017-7). This research was partially financed by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil (CAPES), Finance Code 001.

² Emails: carla.negri@ufabc.edu.br, fk@ic.unicamp.br, phablo@ic.unicamp.br, eduardo@ic.unicamp.br

³ Corresponding author.

cycles of minimum cost such that, for each $i \in [k]$, all vertices of T_i belong to a same cycle. One may show that this reduces to the problem where each terminal set contains exactly two terminals. In this paper, we address the problem restricted to the Euclidean plane.

The Euclidean distance of a line segment $\ell = \overline{xy}$ between points x and y is denoted $\text{length}(\ell)$. We say that ℓ is a line segment with endpoints x and y . If \mathcal{L} is a multiset of line segments, then we define $\nu(\mathcal{L})$ as the set of all endpoints of line segments in \mathcal{L} , and $\delta_{\mathcal{L}}(x)$ as the number of line segments in \mathcal{L} with endpoint at x . Additionally, the *length* of \mathcal{L} , denoted by $\text{length}(\mathcal{L})$, is $\sum_{\ell \in \mathcal{L}} \text{length}(\ell)$.

Problem 1.1 EUCLIDEAN STEINER MULTI CYCLE PROBLEM (ESMC)

INSTANCE: A set $\mathcal{T} = \{\{t_i, t'_i\}\}_{i=1}^n$ of n pairs of points (a.k.a. terminals) in \mathbb{R}^2 with Euclidean distances.

FIND: A multiset \mathcal{L} of line segments such that $\delta_{\mathcal{L}}(x)$ is even for every $x \in \nu(\mathcal{L})$, every terminal t is an endpoint of some line in \mathcal{L} , and the pair of terminals t_i and t'_i are connected by some cycle in \mathcal{L} , for every $i \in [n]$.

GOAL: Minimize $\text{length}(\mathcal{L})$.

The Steiner Multi Cycle Problem (SMCP) is a generalization of the Steiner Cycle Problem (SCP) in the same way that the Steiner Forest generalizes the Steiner Tree problem. The SCP was introduced by Salazar-González who proposed a polyhedral study of integer linear programming formulations of the problem [5]. Steinová proved that the general problem in directed graphs does not admit an approximation algorithm with polynomial ratio in the input size, unless $\mathcal{P} = \mathcal{NP}$ [6]. She also presented a $(3/2)$ -approximation algorithm for the case in which the input graph is undirected and metric. Pereira et al. addressed the metric version of the SMCP [4]. They designed a 4-approximation algorithm and several heuristics for it.

In this work, we propose a randomized approximation scheme to ESMC which runs in $\mathcal{O}(n \log^{\kappa} n)$ time, where n denotes the number of pairs of terminals of the input and κ is a constant. To the best of our knowledge, this is the first polynomial-time approximation scheme designed for a variant of the Steiner Cycle Problem. Our algorithm combines some techniques introduced for the Euclidean TSP by Arora [1] and for the Euclidean Steiner Forest by Borradaile et al. [3]. Despite the similarity between ESMC and the Euclidean Steiner Forest problem, many additional complications arise when dealing with cycles instead of trees. In the proposed algorithm, we introduce new ideas and make several adaptations in the strategy presented by Borradaile et al. [3]. In the dynamic programming table, we use a different description of feasible solutions by incorporating an order in which portals are visited by the cycle. We also ensure that points in the solution being constructed have even degree so that cycles covering the points can be constructed.

The remainder of this paper is organized as follows. Section 2 presents some initial and important definitions which will be needed throughout the text. It also gives an overview of how our algorithm works. Section 3 specifies some changes we have to perform on the input instance. Section 4 is dedicated to prove the existence of an ε -approximate solution for ESMC which has low complexity, where ε is a

constant in the open interval $(0, 1)$. Section 5 shows the details of an algorithm for computing such a solution.

2 Preliminaries and an overview

As the algorithms of Arora [1] and Borradaile et al. [3], ours also depends on recursively partitioning the space enclosing terminals with lines. The proposed algorithm is a dynamic programming which finds an optimal solution that crosses each of such lines a constant amount of times and only at specified points called portals. This partition, called *dissection*, works as follows.

Initially, we have a *root dissection square*, which is a bounding box containing all points of the input instance. The length of the side of this root dissection square is L , which is a power of two and it is at least twice the length of the side of the smallest square enclosing all terminals of the input. We assume this root square is initially positioned with its bottom left corner at position $(0, 0)$ in the plane. Hence, every terminal is in a position inside the square defined by points $(0, 0)$ and $(L/2, L/2)$.

Then we randomly translate the root square, moving its bottom left corner to position $(-a, -b)$, where a and b are chosen uniformly at random from the range $[0, L/2)$. This is called a *random shift*. The root square and the lines around it are considered to be at *depth zero*. We start the dissection of this root square by partitioning it into four squares of same area, which is done with a horizontal and a vertical *dissection lines* that cut the root square. These lines and the four squares created by them are thus at *depth one*. For each of the *dissection squares* created, this process of partition is repeated until there is only one terminal in each dissection square (or multiple terminals, but all with the same coordinates).

We say that the four squares created by dissecting a given square R are *children* of R and that R is their *parent*. If a dissection square R encloses another dissection square R' , then R' is a *descendant* of R . Note that the depth of any square is one plus the depth of its parent. This nested structure of squares forms the *quadtrees*. The depth of a dissection line ℓ , denoted by $\text{depth}(\ell)$, is defined as the lowest depth of dissection squares that ℓ separates. Observe that the squares that are leaves of the quadtree have depth at most $\mathcal{O}(\log L)$.

Consider an integer grid of granularity one in the Euclidean plane, that is, a grid that divides the plane into cells of side length one and whose lines cross at integer points. Let ℓ be a grid line. Since there are $L/2$ values for the vertical shift and $L/2$ values for the horizontal shift and 2^{i-1} of these will make $\text{depth}(\ell) = i$, we have that $\mathbb{P}[\text{depth}(\ell) = i] = 2^i/L$.

Borradaile et al. [3] mentioned that, for the Steiner Forest problem, it was necessary to further subdivide each dissection square in order to obtain a polynomial-size memoization table in the dynamic programming algorithm. Instead of pairing terminals to portals, which would require exponential time, the dynamic programming pairs cells to portals, where a cell can contain more than one terminal. As we face the same problem here, each dissection square is further subdivided into a regular

grid with $B \times B$ cells, where $B = \mathcal{O}(\varepsilon^{-1})$ (to be precisely defined later on) is an integer power of two and $\varepsilon \in (0, 1)$ is a constant.

The solution constructed by our algorithm crosses the border of each dissection square only at designated points of the dissection lines that are called *portals*, which is also a strategy to reduce the complexity of the dynamic programming. The dynamic program starts building solutions for the leaves of the quadtree. In a bottom-up fashion, the solutions of the four children dissection squares are combined to build a solution of their parent square. Thus, the amount of time required by the algorithm depends on the depth of the tree, which is $\mathcal{O}(\log L)$. However, the size of the smallest square enclosing all terminals of the input may be arbitrarily larger than the value of an optimal solution, which implies arbitrarily larger than n . Another important concern is not to have terminals lying on the dissection lines, where the portals are.

To overcome these issues, we transform the original input \mathcal{T} , by first producing a partition $\{\mathcal{T}_i\}_{i \in [k]}$ of \mathcal{T} such that each \mathcal{T}_i has a bounding box of length that depends on $|\mathcal{T}_i|$. We also show that the optimal solution for \mathcal{T} can be built from a simple union of optimal solutions for each \mathcal{T}_i . Then we discretize the instance by scaling it and rounding the coordinates of the terminals to centers of a grid over the plane. Details of this transformation are given on Section 3. Let us denote by $\mathcal{T}'' = \{\mathcal{T}_i''\}_{i \in [k]}$ the transformed input.

Our algorithm builds up a table \mathcal{M}_R for each dissection square R . The table is indexed by the *valid configurations* of R , which encodes subpartitions of cells and portals of R . We describe this concept in Section 5.

Let π be a valid configuration of R . The algorithm keeps in $\mathcal{M}_R[\pi]$ the cost of a minimum-length solution which is *compatible* (to be defined in Section 5) with configuration π and which *conforms* (to be defined in Section 4) to square R . A configuration π is compatible with a solution \mathcal{L} if the line segments in \mathcal{L} match the partition of cells and portals described by π . The conformance definition is the one needed to keep a solution feasible and simple (i.e. low complexity): the line segments form cycles, connect pairs of terminals, and they only cross the border of the square at (a constant amount of) portals.

To write in an entry $\mathcal{M}_R[\pi]$, the algorithm accesses the tables already built for the children dissection squares of R . Among all possible configurations of the four children, it considers only some of them, which are what we called *consistent* with π . Essentially, the consistency guarantees that conforming solutions which are compatible with these children configurations can be combined to build up a feasible conforming solution for R which is compatible with π .

Finally, we prove several results to show that our algorithm indeed builds a feasible solution for \mathcal{T}'' and is a $(1 + \varepsilon)$ -approximation for ESMC. One of the main results is that this dynamic programming indeed finds the value of the optimal solution for \mathcal{T}'' which is conforming to the root dissection square (Theorem 5.3). The other main result is that there exists a solution for \mathcal{T}'' which is conforming to the root square and whose total length is not so distant (at most a factor of ε) from an optimal solution for \mathcal{T} (Theorem 4.3).

3 Restricted problem

Let \mathcal{R} denote the set of all dissection squares. Let \mathcal{C}_R , \mathcal{T}_R , and \mathcal{P}_R be the sets of cells, terminals, and portals, respectively, of the dissection square R , for each $R \in \mathcal{R}$. To simplify notation, we prefer to write \mathcal{C}_* , \mathcal{T}_* , and \mathcal{P}_* when we refer to the sets of cells, terminals, and portals of a dissection square R_* . Let $\text{OPT}(\mathcal{T})$ denote an optimal solution and, as an abuse of notation, the cost of an optimal solution for instance \mathcal{T} .

The first step is to transform the instance into one that gives us a guarantee on the size of the root dissection square. We now describe such transformation and then we show that it is possible to build back a solution for the original instance with a small extra cost. Our algorithm will thus work on such transformed instance.

Let \mathcal{T} denote an instance of ESMC. For every subset $S \subseteq \mathcal{T}$ of pairs of terminals, let $\mu(S) = \{t_i, t'_i : \{t_i, t'_i\} \in S\}$ be the set of underlying terminals, and let $\text{diam}(S)$ be the maximum Euclidean distance between any pair of points in S , that is, the *diameter* of S . Let $G(\mathcal{T})$ denote the graph with vertex set $\mu(\mathcal{T})$ and edge set \mathcal{T} .

Let $\text{dist}(\mathcal{T}) = \max\{\text{diam}(V(H)) : H \text{ is a component of } G(\mathcal{T})\}$, that is, the maximum distance between any pair of terminals that must be connected. Observe that every feasible solution must connect the terminals belonging to a same component of $G(\mathcal{T})$. On the other hand, one may easily construct a feasible solution by connecting each pair of terminals $\{t, t'\} \in \mathcal{T}$ with two copies of the line segment tt' . Note that every line segment in this solution has length at most $\text{dist}(\mathcal{T})$. Hence, it follows that

$$\text{dist}(\mathcal{T}) \leq \text{OPT}(\mathcal{T}) \leq 2|\mathcal{T}| \text{dist}(\mathcal{T}) . \quad (1)$$

The first step of the transformation is to partition the set \mathcal{T} of terminals, which is performed by PARTITION, given in Algorithm 1. Proposition 3.1 shows that the union of optimal solutions for each element of such a partition gives an optimal solution for the whole instance \mathcal{T} . This partition allows us to bound the diameter of each part of the partition considering the number of pairs of terminals it contains and the maximum distance between such terminals. This bound is shown in Proposition 3.2.

Proposition 3.1 *The optimal solution for \mathcal{T} is the union of optimal solutions for each element in the partition produced by Algorithm 1 on input (\mathcal{T}, T) , where T is a minimum-length spanning tree of $\mu(\mathcal{T})$.*

Proposition 3.2 *Algorithm 1 produces a partition $\{\mathcal{T}_i\}_{i \in [k]}$ of \mathcal{T} such that, for each $i \in [k]$, $\text{diam}(\mu(\mathcal{T}_i)) \leq 4|\mathcal{T}_i|^2 \text{dist}(\mathcal{T}_i)$. Moreover, it runs in $\mathcal{O}(n \log n)$ time, where $n = |\mathcal{T}|$.*

Proof. *Sketch.* The stop condition in line 2 guarantees that, for each $i \in [k]$, the points in $\mu(\mathcal{T}_i)$ admit a spanning tree in which every line segment has length at most $2|\mathcal{T}_i| \text{dist}(\mathcal{T}_i)$. This means that there exists a path of length at most $4|\mathcal{T}_i|^2 \text{dist}(\mathcal{T}_i)$ between any pair of points in $\mu(\mathcal{T}_i)$. We show that the partition procedure described in Algorithm 1 can be implemented in $\mathcal{O}(n \log n)$ time, where $n = |\mathcal{T}|$. \square

Algorithm 1: PARTITION

Input: Set \mathcal{T} of pairs of terminals and a minimum-length spanning tree T of the set of points $\mu(\mathcal{T})$

Output: Partition of \mathcal{T}

```

1 Let  $e$  be a longest edge of  $T$ 
2 if  $\text{length}(e) > 2|\mathcal{T}|\text{dist}(\mathcal{T})$  then
3   | Let  $T_1$  and  $T_2$  be the components of  $T - e$ 
4   | Let  $\mathcal{T}_i$  be the subset of pairs of terminals in  $T_i$  for  $i \in \{1, 2\}$ 
5   | /*  $\{\mathcal{T}_1, \mathcal{T}_2\}$  is indeed a partition of  $\mathcal{T}$  */
5   |  $P \leftarrow \text{PARTITION}(\mathcal{T}_1, T_1) \cup \text{PARTITION}(\mathcal{T}_2, T_2)$ 
6 else
7   |  $P \leftarrow \{\mathcal{T}\}$ 
8 return  $P$ 

```

Now that we have a partition $\{\mathcal{T}_i\}_{i \in [k]}$ of the instance \mathcal{T} , the idea is that our algorithm finds approximate solutions to each subset \mathcal{T}_i and then compute the union of the solutions. Since each element of the partition is not particularly different from any other element, we shall assume from now on that we are dealing with one of such subsets. For the sake of simplicity, we keep using \mathcal{T} as the input instance (it represents one of the elements of the partition). Remark that by Proposition 3.2 we have

$$\text{diam}(\mu(\mathcal{T})) \leq 4n^2 \text{dist}(\mathcal{T}) . \quad (2)$$

The second step to transform the instance is the discretization. Herein, we consider that there is an integer grid of granularity one in the Euclidean plane, that is, a grid that divides the plane into cells of side length one and whose lines cross at integer points. The idea is to move the terminals, which can have any coordinates in the original instance, into centers of grid cells.

First we scale the input instance by a factor of $(80n\sqrt{2})/(\varepsilon \text{dist}(\mathcal{T}))$ obtaining \mathcal{T}' , which is called *scaled* instance. By the same argument given in Equation (1) and considering the factor above, we now have

$$\text{OPT}(\mathcal{T}') = \text{OPT}(\mathcal{T}) \frac{80n\sqrt{2}}{\varepsilon \text{dist}(\mathcal{T})} \geq \frac{80n\sqrt{2}}{\varepsilon} . \quad (3)$$

Furthermore, using the inequality from Equation (2) we have

$$\text{diam}(\mu(\mathcal{T}')) = \text{diam}(\mu(\mathcal{T})) \frac{80n\sqrt{2}}{\varepsilon \text{dist}(\mathcal{T})} \leq 4n^2 \text{dist}(\mathcal{T}) \frac{80n\sqrt{2}}{\varepsilon \text{dist}(\mathcal{T})} = \frac{320n^3\sqrt{2}}{\varepsilon} . \quad (4)$$

The discretization ends with a rounding on the position of each terminal to the nearest center of a grid cell. Let \mathcal{T}'' be the instance obtained after rounding the terminals' positions of instance \mathcal{T}' . We call \mathcal{T}'' the *transformed* instance, which is generated after scaling and rounding the points in $\mu(\mathcal{T})$. Since our algorithm works on the transformed instance, Lemma 3.3 guarantees that we can construct a

solution to the original instance from the transformed one with a minor additional cost.

Lemma 3.3 *There exists a solution to \mathcal{T}' derived from any solution for \mathcal{T}'' with additional cost of at most $(\varepsilon/40) \text{OPT}(\mathcal{T}')$.*

As mentioned before, our algorithm works on the transformed instance. The goal is to find a near-optimal solution \mathcal{L}'' for \mathcal{T}'' such that the number of times that \mathcal{L}'' intersects the grid lines is bounded by the length of \mathcal{L}'' itself, which is in turn bounded by a constant times the value of an optimal solution for \mathcal{T}' .

Lemma 3.4 *There exists a solution \mathcal{L}'' to \mathcal{T}'' such that*

$$\text{length}(\mathcal{L}'') \leq \left(1 + \frac{\varepsilon}{40}\right) \text{OPT}(\mathcal{T}') \quad \text{and} \quad \sum_{e \in \mathcal{L}''} \sum_{\ell \in \mathcal{G}} |e \cap \ell| \leq 3 \text{OPT}(\mathcal{T}') ,$$

where \mathcal{G} is the set of lines in the integer grid of granularity one over the Euclidean plane.

Our algorithm will actually find an approximation for the solution mentioned in Lemma 3.4.

Recall that L denotes the length of the side of the root dissection square. We next give an upper bound for the value of L in terms of n . First note that a square of side length $\text{diam}(\mu(\mathcal{T}'))$ is the smallest that encloses all terminals of \mathcal{T}' . Thus, we define L as the smallest integer power of two which is at least $2 \text{diam}(\mu(\mathcal{T}'))$. It follows from Equation (4) that $L \leq (1280n^3\sqrt{2})/\varepsilon$.

Let A be the smallest integer power of two which is at least $30\varepsilon^{-1} \log L$. For each horizontal (resp. vertical) dissection line ℓ , the portals on ℓ are exactly the points in ℓ with x -coordinates (y -coordinates) which are integer multiples of $L/(A2^{\text{depth}(\ell)})$.

Recall that every square of depth i is bounded by two dissection lines of depth i and two dissection lines of depth smaller than i . Moreover, the side of a dissection square R of depth i has length $L/2^i$. Since portals are placed at integer multiples of $L/(A2^i)$ in a dissection line of depth i , there is at most A portals on each side of A . Therefore, it holds that every dissection square has at most $4A$ portals on its boundary. One may also prove that, for every dissection square R , all corners of R are portals, except for the points which are corners of the root dissection square.

4 Approximate conforming solutions

Let R be a dissection square where ∂R denotes its border and let \mathcal{L} be a finite multiset of line segments in R . Let us denote by $\nu(\mathcal{L})$ the set of all endpoints of line segments in \mathcal{L} . For each point $x \in \nu(\mathcal{L})$, $\delta_{\mathcal{L}}(x)$ denotes the number of line segments in \mathcal{L} that contain x . Additionally, let $D = \mathcal{O}(1/\varepsilon)$ be a constant (to be defined in Theorem 4.1).

We say that \mathcal{L} *connects* two points x, x' of R if x and x' belong to the same connected component of \mathcal{L} . Moreover, \mathcal{L} *conforms* to R if

- (i) $\mathcal{L} \cap \partial R$ contains at most $4(D + 1)$ connected components;
- (ii) every connected component of $\mathcal{L} \cap \partial R$ contains a portal of R ;
- (iii) for every cell $x \in \mathcal{C}_R$, there exists at most one component of \mathcal{L} that intersects x and also intersects the boundary of R ;
- (iv) for every portal $p \in \mathcal{P}_R$, we have $\delta_{\mathcal{L}}(p) \leq 2$;
- (v) for every point $s \in \nu(\mathcal{L})$ that is not a portal of R , $\delta_{\mathcal{L}}(s)$ is even; and
- (vi) if a terminal t in the square R is not connected to its mate by \mathcal{L} , then t is connected to the boundary of R by \mathcal{L} .

An example of a conforming multiset of line segments is depicted in Figure (1a).

We say that \mathcal{L} *recursively conforms* to R if it conforms to R and to all dissection squares that are descendant of R . If \mathcal{L} recursively conforms to the root dissection square, then \mathcal{L} is said to be *conforming*.

In what follows, we prove the existence of a conforming solution for ESMC of cost at most $(1 + \varepsilon) \text{OPT}(\mathcal{T})$. We start from a feasible solution ensured by Lemma 3.4, then we transform it to satisfy all properties of a conforming solution. In expectation over the random shifts of the dissection lines, the transformations made on the initial solution increase its length only by a small constant factor of the optimal length.

Theorem 4.1 *Let \mathcal{T}'' be a transformed instance of the original instance \mathcal{T} of the EUCLIDEAN STEINER MULTI CYCLE problem. There exists a feasible solution \mathcal{L} for \mathcal{T}'' that has length at most $(1 + 99\varepsilon/120) \text{OPT}(\mathcal{T})$, in expectation over the random shift of the bounding box, and such that \mathcal{L} satisfies the following properties for each dissection square R :*

- (i) *for each side S of R , $\mathcal{L} \cap S$ has at most D noncorner components where $D = 60\varepsilon^{-1}$;*
- (ii) *every connected component of $\mathcal{L} \cap \partial R$ contains a portal of R ;*
- (iii) *for every cell $x \in \mathcal{C}_R$, there exists at most one component of \mathcal{L} that intersects x and also intersects the boundary of R ; and*
- (iv) *if a terminal t in R is not connected to its mate by $\mathcal{L} \cap R$, then t is connected to the boundary of R by \mathcal{L} .*

Proof. *Sketch.* Let \mathcal{L}'' be the feasible solution for \mathcal{T}'' given by Lemma 3.4. Note that \mathcal{L}'' trivially satisfies property (iv) since it is a feasible solution for \mathcal{T}'' . We can show how to transform \mathcal{L}'' into a feasible solution \mathcal{L} which satisfies properties (i), (ii), and (iii). In all these transformations, we only augment already existing components of \mathcal{L}'' by adding sets of line segments that form a cycle. Thus, one may easily verify that property (iv) is preserved in \mathcal{L} and $\delta_{\mathcal{L}}(s)$ is even, for every point $s \in \nu(\mathcal{L})$. Therefore, \mathcal{L} is a solution for \mathcal{T}'' which satisfies properties (i)–(iv).

As a consequence of the small augment on the length of the solution (details omitted), we have $\text{length}(\mathcal{L}) - \text{length}(\mathcal{L}'') \leq (3\varepsilon/5) \text{OPT}(\mathcal{T})$. By combining the last inequality with the bound given in Lemma 3.4, we obtain $\text{length}(\mathcal{L}) \leq (1 +$

$$\varepsilon/40) \text{OPT}(\mathcal{T}) + (3\varepsilon/5) \text{OPT}(\mathcal{T}) \leq (1 + 25\varepsilon/40) \text{OPT}(\mathcal{T}). \quad \square$$

Corollary 4.2 *Let \mathcal{T}'' be a transformed instance of the original instance \mathcal{T} of the EUCLIDEAN STEINER MULTI CYCLE problem. There exists a conforming solution for \mathcal{T}'' that has length at most $(1 + 99\varepsilon/120) \text{OPT}(\mathcal{T})$, in expectation over the random shift of the bounding box.*

Proof. Let \mathcal{L} be the solution given in Theorem 4.1. Note that portals of the dissection squares may be visited multiple times by \mathcal{L} . Following the same ideas of shortcut solution described by Arora in [1], we can obtain from \mathcal{L} a conforming solution to \mathcal{T}'' which has length at most $\text{length}(\mathcal{L}) \leq (1 + 99\varepsilon/120) \text{OPT}(\mathcal{T})$. \square

Theorem 4.3 *Let \mathcal{T} be an instance of the EUCLIDEAN STEINER MULTI CYCLE. There exists a conforming solution for \mathcal{T} with length at most $(1 + \varepsilon) \text{OPT}(\mathcal{T})$, in expectation over the random shift of the bounding box.*

Proof. Let \mathcal{L} be the solution to the transformed version \mathcal{T}'' guaranteed by Corollary 4.2. By Lemma 3.3, we obtain from \mathcal{L} a solution \mathcal{L}' to \mathcal{T} such that $\text{length}(\mathcal{L}') \leq \text{length}(\mathcal{L}) + (\varepsilon/40) \text{OPT}(\mathcal{T}) \leq (1 + 102\varepsilon/120) \text{OPT}(\mathcal{T})$. \square

5 The algorithm

Let S be a set and let $\mathcal{P}(S)$ denote the power set of S , that is, the set of all subsets of S . A set π is said to be a *subpartition* of S if $\pi \subseteq \mathcal{P}(S)$, and the elements in π are pairwise disjoint. Let π and π' be two subpartitions of S . We say that π' is a *coarsening* of π if, for every $P' \in \pi'$, there exists $\mathcal{S} \subseteq \pi$ such that $P' = \bigcup_{P \in \mathcal{S}} P$. Finally, for every element $s \in S$, $\pi(s)$ denotes the element in π which contains s if there is one, or \emptyset otherwise.

Consider now a dissection square R . Let $C_R : \mathcal{T}_R \rightarrow \mathcal{C}_R$ be a function such that $C_R(t)$ is the cell of R that contains t . To simplify, we use C_\star when referring to the function associated with dissection square R_\star .

A *configuration* of R is a tuple $(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)$ that satisfies

- (i) π_{IN} and π_{OUT} are subpartitions of $\mathcal{P}_R \cup \mathcal{C}_R$;
- (ii) π_{OUT} is a coarsening of π_{IN} ;
- (iii) each element of π_{IN} contains at least one portal and at least one cell of R ; and
- (iv) for each $Q \in \pi_{\text{IN}}$,

$$\theta(Q) = \left\{ p_1 \dots p_\ell : \{p_i\}_{i=1}^\ell \subseteq Q \text{ and the portals in } \{p_i\}_{i=1}^{\ell-1} \text{ are pairwise different if } \ell > 1 \right\},$$

that is, a set of sequences of portals in Q . Note that, if $\ell > 1$, then p_1 and p_ℓ may be the same portal.

A configuration $(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)$ of R is said to be *valid* if the following holds.

- (i) $|\pi_{\text{IN}}| \leq 4(D + 1)$, and π_{IN} contains at most $8(D + 1)$ portals;
- (ii) for every $Q \in \pi_{\text{IN}}$ and portal $p \in Q$, there is a single sequence in $\theta(Q)$ which contains p ;

- (iii) for every pair of mated terminals $\{t, t'\}$ such that $t \in \mathcal{T}_R$ and $t' \notin \mathcal{T}_R$, there is a single $Q \in \pi_{\text{IN}}$ such that $C_R(t) \in Q$. For every pair of mated terminals $\{t, t'\}$ such that $t, t' \in \mathcal{T}_R$, either there is $Q \in \pi_{\text{OUT}}$ such that $\{C_R(t), C_R(t')\} \subseteq Q$, or $\{C_R(t), C_R(t')\} \cap Q = \emptyset$ for all $Q \in \pi_{\text{IN}}$.

See Figure (1b) for an example of valid configuration.

Every dissection square has $4A$ portals and B^2 cells, so we can bound the amount of valid configurations.

Corollary 5.1 *There exist at most $(\varepsilon^{-2} \log n)^{\mathcal{O}(\varepsilon^{-2})}$ valid configurations.*

Let R and R' be different squares. Let $(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)$ and $(\pi'_{\text{IN}}, \pi'_{\text{OUT}}, \theta')$ be valid configurations of R and R' , respectively. We denote by $G(\pi_{\text{IN}}, \pi'_{\text{IN}})$ the *intersection graph* of π_{IN} and π'_{IN} constructed as follows. The vertex set of $G(\pi_{\text{IN}}, \pi'_{\text{IN}})$ is $\pi_{\text{IN}} \cup \pi'_{\text{IN}}$ and its edge set is $\{\{u, v\} : u \cap v \neq \emptyset\}$. Notice that only portals may be in the intersection of $u \in \pi_{\text{IN}}$ and $v \in \pi'_{\text{IN}}$. We define $\pi_{\text{IN}} \uplus \pi'_{\text{IN}} = \left\{ \bigcup_{v \in V(K)} v : K \text{ is a connected component of } G(\pi_{\text{IN}}, \pi'_{\text{IN}}) \right\}$. For any finite set of subpartitions $\{\pi^i\}_{i=1}^t$, $\biguplus_{i=1}^t \pi^i_{\text{IN}} = \pi^1_{\text{IN}} \uplus \dots \uplus \pi^t_{\text{IN}}$ is naturally defined.

Furthermore, we denote by $G(\theta)$ the *portal graph*, which has vertex set \mathcal{P}_R and edge set $\{\{u, v\} : \exists Q \in \pi_{\text{IN}} \text{ s.t. } uv \text{ is a subsequence of some sequence in } \theta(Q)\}$. Note that each element of $\theta(Q)$ defines a (closed or open) path in $G(\theta)$, for all $Q \in \pi_{\text{IN}}$.

Consider a dissection square R_0 and its child dissection squares R_1, R_2, R_3 , and R_4 . For each $i \in \{0, \dots, 4\}$, let $(\pi^i_{\text{IN}}, \pi^i_{\text{OUT}}, \theta^i)$ be a valid configuration for R_i . Moreover, let $\lambda = \biguplus_{i=1}^4 \pi^i_{\text{IN}}$ and let $G = \bigcup_{i=1}^4 G(\theta^i)$. The set of configurations $\{(\pi^i_{\text{IN}}, \pi^i_{\text{OUT}}, \theta^i)\}_{i=0}^4$ is said to be *consistent* if it satisfies all the following properties.

- (i) π^0_{IN} is identical to the set obtained from λ by the following procedure. For each $P \in \lambda$, remove from P all portals in $(\bigcup_{i=1}^4 \mathcal{P}_i) \setminus \mathcal{P}_0$, and replace each cell in P by the corresponding parent cell of R_0 . Finally, remove from λ all elements that do not contain any portal of R_0 .
- (ii) For each $Q \in \lambda$ such that $Q \cap \mathcal{P}_0 = \emptyset$, each component of $G[Q \cap (\bigcup_{i=1}^4 \mathcal{P}_i)]$ has an Eulerian cycle.
- (iii) For each $Q \in \lambda$ such that $Q \cap \mathcal{P}_0 \neq \emptyset$, each component of $G[Q \cap (\bigcup_{i=1}^4 \mathcal{P}_i)]$ has an Eulerian path with both endpoints in \mathcal{P}_0 .
- (iv) For every $i \in \{1, 2, 3, 4\}$ and every $x, y \in \mathcal{P}_i \cup \mathcal{C}_i$, $\pi^i_{\text{OUT}}(x) = \pi^i_{\text{OUT}}(y)$ if and only if either $\lambda(x) = \lambda(y)$, or there are portals p and q of R_0 such that $\pi^0_{\text{OUT}}(p) = \pi^0_{\text{OUT}}(q)$, $\lambda(x) = \lambda(p)$ and $\lambda(y) = \lambda(q)$.
- (v) For every pair of mated terminals $\{t, t'\}$ such that $t \in \mathcal{T}_i$ and $t' \in \mathcal{T}_j$ with $i, j \in \{1, 2, 3, 4\}$, we have either $\lambda(C_i(t)) = \lambda(C_j(t'))$, or $\pi^0_{\text{OUT}}(C_i(t)) = \pi^0_{\text{OUT}}(C_j(t'))$.
- (vi) For every portal $p \in (\bigcup_{i=1}^4 \mathcal{P}_i) \setminus \mathcal{P}_0$, $d_G(p)$ is even, where $d_G(p)$ is the degree of p in G .
- (vii) For every portal $p \in \mathcal{P}_0$, $d_0(p) = d_G(p)$.
- (viii) There exists a partition of $E(G)$ into a set \mathcal{X} of edge-disjoint paths with the

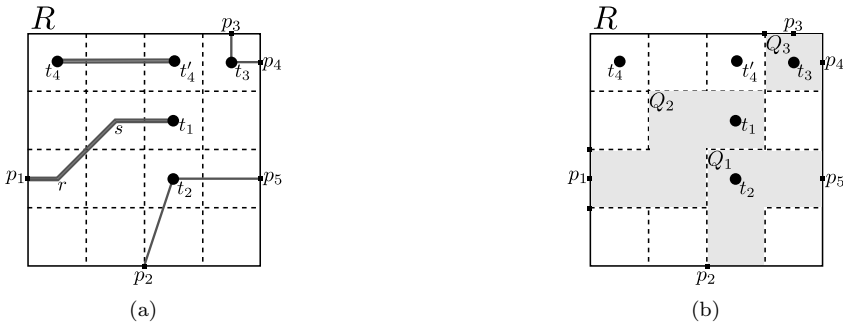


Fig. 1. A dissection square R with terminals t_i and portals p_j . Cells are illustrated by dotted lines. Figure 1(a) shows a multiset of line segments which conforms to R . Thicker lines represent two copies of the same line segments (e.g. between t_1 and p_1). Figure 1(b) shows a valid configuration $(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)$ for R . The terminals and portals of R are $\{t_1, t_2, t_3, t_4, t'_4\}$ and $\{p_1, p_2, p_3, p_4, p_5\}$, respectively. $\pi_{\text{IN}} = \{Q_1, Q_2, Q_3\}$, $\pi_{\text{OUT}} = \{Q_1 \cup Q_2, Q_3\}$, $\theta(Q_1) = \{p_2 p_5\}$, $\theta(Q_2) = \{p_1\}$, and $\theta(Q_3) = \{p_3 p_4\}$.

following property. K is a component of $G(\theta^0)$ if and only if there is a path $P \in \mathcal{X}$ such that K is identical to the graph obtained from P by lifting all edges incident to vertices in $V(P) \setminus \mathcal{P}_0$ and removing all vertices in $V(P) \setminus \mathcal{P}_0$.

A pseudocode of our algorithm is given in Algorithm 2. It finds the cost of an optimal conforming solution for the transformed instance. Proposition 5.2 and Theorem 5.3 are used to prove this claim.

Let \mathcal{L} and $(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)$ be a conforming solution and a valid configuration of R , respectively. Let $\Gamma(\mathcal{L})$ denote the set of all possible paths formed by line segments in \mathcal{L} . We define \mathcal{L}' as the set of line segments in \mathcal{L} which belong to components that contain some portal of R . We say that the portal graph $G(\theta)$ and \mathcal{L} are *coherent* if and only if there exists an injective function $f : E(G(\theta)) \rightarrow \Gamma(\mathcal{L}')$ such that (i) for each edge uv in $G(\theta)$, $f(uv)$ is a path of line segments with endpoints at the portals u and v , and (ii) for each $\ell \in \mathcal{L}'$, there exists a single $e \in E(G(\theta))$ such that $f(e)$ contains ℓ .

Recall that every component of $G(\theta)$ is a (open or closed) path. Hence, if \mathcal{L} and $G(\theta)$ are coherent, then the set of all components of $G(\theta)$ induces a partition of \mathcal{L}' into paths of line segments that link portals of R .

We say that \mathcal{L} and $(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)$ are *compatible* if and only if

- (i) for every component K of \mathcal{L} that intersects the boundary of R , there exists a single part $Q \in \pi_{\text{IN}}$ such that Q consists exactly of the cells and portals intersected by K in R ; and
- (ii) $G(\theta)$ and \mathcal{L} are coherent.

Proposition 5.2 *For every conforming solution \mathcal{L} to R , there exists a valid configuration $(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)$ of R which is compatible with \mathcal{L} .*

Theorem 5.3 *Let R be a dissection square. Algorithm 2 (i.e., SOLVE-ESMC) fills in a table \mathcal{M}_R such that, for every valid configuration $(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)$ of R , $\mathcal{M}_R[(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)]$ is the minimum length of a multiset of lines that is compatible with $(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)$ and recursively conforms to R .*

Algorithm 2: SOLVE-ESMC

Input: Dissection square R and \mathcal{T}_R

```

1 if  $R$  contains at most one terminal then
2   foreach valid configuration  $(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)$  of  $R$  do
3      $\mathcal{M}_R[(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)] \leftarrow 0$ 
4     foreach part  $Q \in \pi_{\text{IN}}$  do
5       if the cells in  $Q$  are connected and these cells contain the portals
        of  $Q$  then
6         Let  $\mathcal{L}_Q$  be a minimum-length multiset of lines contained in
          the cells of  $Q$  such that  $\mathcal{L}_Q$  is coherent with  $G(\theta)$ , contains
          the terminal (if any) in  $Q$ , and each cell in  $Q$  is crossed by
          at most one component of  $\mathcal{L}_Q$  that intersects the boundary
          of  $R$ . If there is no such a set, then
          define  $\text{length}(\mathcal{L}_Q) = +\infty$ .
7          $\mathcal{M}_R[(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)] \leftarrow \mathcal{M}_R[(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)] + \text{length}(\mathcal{L}_Q)$ 
8       else
9          $\mathcal{M}_R[(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)] \leftarrow +\infty$ 
10 else
11   Let  $R_1, R_2, R_3$  and  $R_4$  be the four child dissection squares of  $R$ 
12   SOLVE-ESMC( $R_i$ ) for every  $i \in \{1, 2, 3, 4\}$ 
13   foreach valid configuration  $(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)$  of  $R$  do
14      $\mathcal{M}_R[(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)] \leftarrow +\infty$ 
15     forall config.  $(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)$  of  $R$  and  $\{(\pi_{\text{IN}}^i, \pi_{\text{OUT}}^i, \theta^i)\}_{i=1}^4$  of  $\{R_i\}_{i=1}^4$  do
16       if  $\{(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)\} \cup \{(\pi_{\text{IN}}^i, \pi_{\text{OUT}}^i, \theta^i)\}_{i=1}^4$  is consistent then
17          $\mathcal{M}_R[(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)] \leftarrow$ 
           $\min \left\{ \mathcal{M}_R[(\pi_{\text{IN}}, \pi_{\text{OUT}}, \theta)], \sum_{i=1}^4 \mathcal{M}_i[(\pi_{\text{IN}}^i, \pi_{\text{OUT}}^i, \theta^i)] \right\}$ 

```

Proposition 5.4 *Let \mathcal{T} be an instance of ESMC with n pairs of terminals, and let R be a root dissection square enclosing the terminals of \mathcal{T} . Algorithm 2 (SOLVE-ESMC) on input (R, \mathcal{T}) runs in $\mathcal{O}(n \log^\kappa n)$ time, where $\kappa \in \mathcal{O}(\varepsilon^{-2})$.*

6 Conclusion and further research

Let us denote by \mathcal{T} be an instance of ESMC which contains n pairs of terminals, and let $\varepsilon \in (0, 1)$. We proved in Theorem 5.3 that the proposed algorithm finds a minimum-length solution of \mathcal{T} that is conforming with a given bounding box. In Proposition 5.4, we argued that such a solution can be found in $\mathcal{O}(n \log^\kappa n)$ time, where κ is a constant which only depends on ε . In Theorem 4.3, we showed the existence of a conforming solution for \mathcal{T} of length at most $(1 + \varepsilon) \text{OPT}(\mathcal{T})$ in expectation over the random shift of the bounding box. From the mentioned results and by Markov's inequality, we conclude that the solution computed for \mathcal{T}

has length at most $(1 + \varepsilon) \text{OPT}(\mathcal{T})$ with probability at least one half. Therefore, the proposed algorithm is a randomized approximation scheme for ESMC which runs in quasilinear time.

In 2011, Bateni et al. devised a polynomial-time approximation scheme (PTAS) for the Steiner Forest Problem in planar graphs [2]. Therefore, a natural direction for further research is to investigate the existence of a PTAS for the Steiner Multi Cycle Problem on planar graphs.

References

- [1] Arora, S., *Polynomial time approximation schemes for Euclidean traveling salesman and other geometric problems*, Journal of the ACM **45** (1998), pp. 753–782.
- [2] Bateni, M., M. Hajiaghayi and D. Marx, *Approximation schemes for Steiner forest on planar graphs and graphs of bounded treewidth*, Journal of the ACM **58** (2011), pp. 21:1–21:37.
- [3] Borradaile, G., P. N. Klein and C. Mathieu, *A polynomial-time approximation scheme for Euclidean Steiner forest*, ACM Transactions on Algorithms **11** (2015), pp. 19:1–19:20.
- [4] Pereira, V. N. G., M. C. S. Felice, P. H. D. B. Hokama and E. C. Xavier, *The Steiner multi cycle problem with applications to a collaborative truckload problem*, in: *17th International Symposium on Experimental Algorithms, (SEA'2018)*, 2018, pp. 26:1–26:13.
- [5] Salazar-González, J.-J., *The Steiner cycle polytope*, European Journal of Operational Research **147** (2003), pp. 671–679.
- [6] Steinová, M., *Approximability of the minimum Steiner cycle problem*, Computing and Informatics **29** (2012), pp. 1349–1357.