On the Generalized Minimum Spanning Tree in the Euclidean Plane

Homa Ataei Kachooei* Mansoor Davoodi† Dena Tayebi ‡

Abstract

We aim at finding a minimum spanning tree consisting of exactly one point per cluster, for a set of n points in the plane partitioned into $k < n$ clusters. We show that this problem is NP-complete even if every cluster contains two points with equal y coordinates. Further, we show that this problem does not have an FPTAS unless $P = NP$.

Keywords: Minimum Spanning Tree, NP-completeness, Approximation Algorithm

1 Introduction

In this paper, we study Generalized Minimum Spanning Tree (GMST) problem in the plane. First we review the GMST problem on graphs. Given a connected undirected graph $G = (V, E)$ in which nodes are partitioned into k clusters:

$$V = V_1 \cup V_2 \cup .... \cup V_k \qquad \forall i \neq j,\ \ V_i \cap V_j = \emptyset, \quad (1)$$

where V_i is a subset of nodes. The GMST problem is to find a Minimum Spanning Tree (MST) which consists of exactly one point from each cluster.

The GMST problem on graphs was proved NP-hard by a reduction from the Vertex Cover problem [9]. It was also proved that the GMST problem on graphs cannot be approximated within any constant factor [10]. Furthermore when $G = (V, E)$ is a tree, the problem is NP-hard [10]. However, there is an approximation algorithm for the GMST problem when the cluster size is bounded by a constant ρ. In this case, the GMST problem can be approximated to within 2ρ [11].

A geometric version of the GMST problem was studied with grid clustering. In this version, the graph was considered complete and the nodes are placed in a $(r \times l)$-grid and weight of each edge is the distance between two nodes. The nodes belonging to a cell of the grid make a cluster. A PTAS for the GMST problem with grid clustering was proposed in [3]. Moreover, a $(1 + 4\sqrt{2} + \epsilon)$-approximation algorithm was presented for the GMST problem with the grid clustering [1]. The existence of a PTAS shows that the GMST problem with grid clustering is easier than the GMST problem. An alternative version of the GMST problem focuses on finding the MST which consists of at least one point per cluster [7]. The NP-completeness of this version was shown in [7] including the case where each cluster contains three points. Additionally, it was shown that this version cannot be approximated within any constant factor [12]. Also it was shown that this version of the GMST problem with grid clustering is strongly NP-hard even if non-empty grid cells are connected and each grid cell (cluster) contains at most two points [5]. Further, a $(r \times l)$-grid $(r \leq l)$ was used in [5] and a dynamic programming algorithm was presented which solves this version in $O(l\rho^{6r}2^{34r^2}r^2)$ time. Note that if r or l are bounded, this algorithm is polynomial. Feremans et al. [5] using the dynamic programming algorithm, presented a PTAS when all non-empty grid cells are connected and the number of non-empty grid cells is superlinear in r and l.

The Class Steiner Tree (CST) problem (known also as Group Steiner Tree (GST) problem) is similar to the GMST problem. A short review of the CST problem is provided in the following.

Given a connected undirected graph $G = (V, E)$ in which the nodes are partitioned into disjoint sets, such that:

$$V = S \cup R_1 \cup ... \cup R_k, \quad (2)$$

where R_i is a required class for $i = 1, 2, ..., k$ and S is Steiner class, the CST problem tries to find an MST which include at least one node per required class. The CST problem was proved to be NP-hard even if there is no Steiner node, the weight of all edges are unit and the nodes degree are less than or equal to three [6]. Finally, the CST problem cannot be approximated within any constant factor even for trees without Steiner node and unit edges [6].

As pointed out before, the focus of this paper is on studying the GMST problem in the plane (not graph). Hence, for a given set P containing n points in the plane which is partitioned into k clusters:

$$P = P_1 \cup P_2 \cup .... \cup P_k \qquad \forall i \neq j,\ \ P_i \cap P_j = \emptyset, \quad (3)$$

the GMST problem in the plane tries to find an MST which consists of exactly one point from each cluster.

We prove that the GMST problem in the plane is NP-complete even if each cluster contains two points with equal y coordinates. We further prove that this problem

*ataei.homa@gmail.com

†Institute for Advanced Studies in Basic Sciences (IASBS), Zanjan, Iran, mdmonfared@iasbs.ac.ir

‡denatayebi@yahoo.com

does not have an FPTAS unless $P = NP$. This version is a simple case of the GMST problem.

2 NP-completeness of the GMST Problem in the Euclidean Plane

By a reduction from the planar 3SAT problem we prove that the GMST problem in the Euclidean plane is NP-complete.

Consider the 3SAT problem where $C = \{c_1, c_2, ..., c_m\}$ is the set of clauses and $V = \{v_1, v_2, ..., v_n\}$ is the set of variables. Create a graph $G = (U, E)$ for every instance of the 3SAT problem such that:

$$U = C \cup V, \tag{4}$$

and

$$E = E_1 \cup E_2, \tag{5}$$

where E_1 and E_2 are:

$$E_1 = \{(c_i, v_j) \mid v_j \in c_i \ \ or \ \ \overline{v}_j \in c_i\} \tag{6}$$

and

$$E_2 = \{(v_j, v_{j+1}) \mid 1 \leq j < n\} \cup \{(v_n, v_1)\}. \tag{7}$$

The set of all edges in E_2 is called *spinal path* [4]. There is a node for each variable and for each clause in the graph G, resulting in $|U| = |C| + |V|$. Draw one edge between a variable node and a clause node in G, if and only if the clause contains a literal of the variable. The planar 3SAT problem includes all instances of the 3SAT problem with similar planar graphs. The planar 3SAT is proved to be NP-complete by a reduction from the 3SAT problem [8].

Theorem 1 *The GMST problem in the plane is NP-complete. This claim is true even under the constraint that every cluster contains two points with equal y coordinates. Also the GMST problem does not have an FPTAS unless $P = NP$.*

Proof. We utilize a similar approach to that of Theorem 7 in [4] for demonstrating the correctness of the theorem. As such, We perform the proof by using a reduction from the planar 3SAT problem for the GMST problem.

Every instance of the planar 3SAT problem should be converted to an instance of the GMST problem in the plane. First, we design two gadgets for the variables and clauses called *variable gadget* and *clause gadget*, respectively. The design of these gadgets is based on some clusters of points where every cluster consists of a pair of points. The designed gadgets in the graph of ϕ are replaced as follows: if a node in the graph of ϕ is corresponding to a variable in ϕ, it is replaced by a variable gadget and if it is corresponding to a clause in ϕ, it is replaced by a clause gadget. Consequently, we replace all the nodes in the graph of ϕ with the gadgets. One should ensure that the number of clusters (pairs of pionts) in this reduction is polynomially bounded in the size of ϕ. Therefor, we use a special drawing graph called *orthogonally drawing* [2]. In the orthogonally drawing each node is shown with a rectangle and each edge is denoted by a sequence of vertical and horizontal segments. The orthogonally drawing provide a practical mean to draw the graph in the defined space. To continue, we need the Theorem 2 from [2]. The theorem is provided below.

Theorem 2 *[12, Theorem 4] Let H be a simple graph without nodes of degree ≤ 1, where n is the number of nodes and m is the number of edges. Then H has an orthogonally drawing in an $(\frac{m+n}{2} \times \frac{m+n}{2})$-grid with one bend per edge. The box size of each node v is at most $\frac{deg(v)}{2} \times \frac{deg(v)}{2}$. It can be found in $O(m)$ time.*

At this stage, we want to convert the planar 3SAT instance to a GMST instance. Therefor we first draw the orthogonally drawing and then, replace the variable gadgets with the variables and the clause gadgets with the clauses.

2.1 Variable Gadget

For each variable in the planar 3SAT instance, we design a gadget which consists of k cluster, such that k is an even number and $4 \leq k \leq 2c + 4$, where c is the number of clauses that include this variable. Consider two variables i and j such that $1 \leq i \leq k$. Variable i is the number of the clusters and j is an index of i which is equivalent to the number of points in the cluster. Because there are only two points in every cluster, we set $j = 0$ or $j = 1$. The $j = 0$ and $j = 1$ cases correspond to the first and second points in the cluster, respectively. If the coordinates of the first point in the first cluster is denoted by (x, y), then for every i and j where $1 \leq i \leq k - 1$ and $j = 0, 1$, coordinates of point i_j are

$$(x + 2(i - 1) + j, y + ((i + 1) \ mod \ 2)).$$

This way, $k-1$ clusters are being embedded in the plane and the coordinates of the points in the cluster k are $(x - \sqrt{5}, y)$ and $(x + 2(k-1) - 1 + \sqrt{5}, y)$. Consequently, the placement of these points in the plane leads to a structure which is a part of the variable gadget. We denote this structure by A.

Structure A with $k = 8$ is depicted in Figure 1. Here, the location of the points in the plane is not important but the distance between them is an issue.

The proof of the Theorem 1 and the design for the variable gadget are not completed yet, we need Lemma 3 in that regard.

Figure 1: Structure A. The segment between two points shows that they are in a cluster. Also the right side and the left side points are in a cluster.

Lemma 3 *In the structure A with k clusters, there are two different choices from clusters which lead to the optimal solution of the GMST problem. In one of these solutions the right side points choose in all clusters and in other solution the left side points choose in all clusters. The weight of the Euclidean MST (EMST) for these solutions is* $\sqrt{5}(k-1)$. *Further, the weight of EMST in every other selection of points except the aforementioned two, is at leas 0.1 more than the weight of an optimal solution.*

Proof. See Appendix at the end of the article. □

With regard to this Lemma, there are two possible solutions for the GMST problem for the structure A. We assume where the right side points are selected from all clusters to be equivalent to the case for which the variable is True. Moreover we assume where the left side points are selected from all clusters to be equivalent to the case for which the variable is False.

These assumptions along with Lemma 3, provide the necessary requirements to complete the design of the variable gadgets. Suppose that the number of clauses containing this variable is equal to c. For each of these clauses, we posit a fixed point at a distance of 2 units to one of the right side points in the direction of y. Similarly, for each clause containing a variable negation, a fixed point is placed at a distance of 2 units of one of the left side points in the direction of y.

Because of these clauses are located above or below this variable, we locate mentioned points above or below the gadget. Additionally, we locate two points for spinal path connections such that they don't have any effect on choosing the right or the left side points. Therefore, these points should be selected in a way that they have the same distances from the nearest right and left side points in the structure A. So, the location of these points are selected as $(x+0.5, y-1)$ and $(x+2(k-2)+0.5, y-1)$. Indeed, these are the endpoints of the edges that establish the spinal path connections.

Figure 2 is an example of a variable gadget. This Figure illustrates the variable such as z attending in three clauses which $\overline{z}$ comes in one clause and z in the other two.

2.2 Clause Gadget

Consider a sequence of points which are located in unit distances along a line. A clause gadget is formed from three of the mentioned sequences that collide at a point

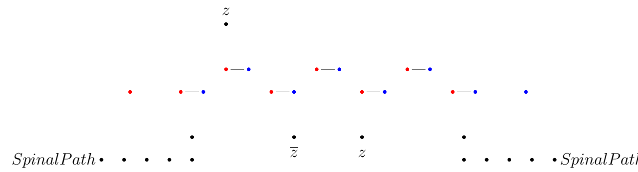


Figure 2: An example of the variable gadget

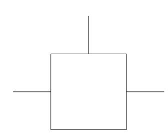
[4]. Figure 3a is related to a node which is corresponding to a clause in the orthogonally drawing. If we replace this node with a clause gadget Figure 3b is obtained.

2.3 Reduction

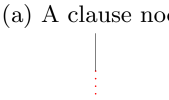
We scale up orthogonally drawing with factor of 2 and replace the graph nodes of the orthogonally drawing with the introduced gadgets. This is also the case for the graph edges which are exchanged with a sequence of points located in a unit distances along the edges.

As pointed out before, in the orthogonally drawing, each node is a box with the maximum size of $\frac{deg(v)}{2} \times \frac{deg(v)}{2}$. Designed gadget may not fit into the intended box in the orthogonally drawing, but size of this gadget is at most $(4(deg(v)-1)+2\sqrt{5}+1)\times 5$ which is bounded in the size of the box.

So far we converted each planar 3SAT instance to a GMST instance. The next step at this point is to show that every solution of the GMST problem is a solution of the planar 3SAT problem. We used the orthogonally drawing which is drawn in a $(\frac{m+n}{2} \times \frac{m+n}{2})$-grid, and the drawing is polynomially bounded in the size of the planar 3SAT instance. Hence the number of used fixed points in this reduction is polynomially bounded in the size of the planar 3SAT instance. These fixed points have a unique MST with constant weight. Therefor,the MST obtained from the gadgets themselves and their connections determine the weight of the MST. Sum of the MST weight of the connection between the spinal path and the gadgets, and the MST weight of the fixed points is denoted by W_{edges}. Also the MST weight of the



(a) A clause node



(b) A clause gadget

Figure 3: Replacing a clause node in the orthogonally drawing with a clause gadget

gadgets and gadgets connections to edges is denoted by $W_{clusters}$. So the total weight of MST is sum of W_{edges} and $W_{clusters}$:

$$W_{total} = W_{edges} + W_{clusters}. \tag{8}$$

In the optimal solution of MST, the connectivity should be reached by minimum cost. W_{edges} has a constant value, no matter which points are selected. The connections between spinal paths and variable gadgets have constant weight and these connections cause all variable gadgets to be connected to each other. Now we investigate connection of the edges to the variable gadget. In the optimal solution of the GMST problem each clause gadgets should be connected to a variable gadget and the cost of each connection is 2 units. So the total cost of these connections is twice the number of all clauses. If only the right or the left side points are selected in each gadget, the weight of the obtained MST can be calculated as :

$$W_{clusters} = (\sqrt{5}(R - n)) + 2c + 2n\sqrt{4 + (0.5)^2}, \tag{9}$$

where n is the number of variables, c is the number of clauses and R is the number of all clusters. In this case, each gadget is just connected to a clause containing the variable or its negation. This means there is a True assignment for the planar 3SAT problem. If $W_{clusters}$ is more than this value, there is at least one variable gadget which is connected to a clauses containing the variable and a clause containing variable negation. This means there is no True assignment for the planar 3SAT problem.

Now we show that the GMST problem does not have an FPTAS unless $P \neq NP$. Consider the existence of an FPTAS for the GMST problem. Given a planar 3SAT instance, we build the GMST problem input as explained previously and calculate W_{total}. We determine the ϵ value such that $\epsilon < \frac{0.1}{W_{total}}$. So a $(1+\epsilon)$-approximation solution for the GMST problem can be used to verify whether there is a True assignment for the planar 3SAT problem or not. Since the planar 3SAT is NP-Hard, we can conclude there is no FPTAS for the GMST problem unless $P \neq NP$. $\square$

3 Discussion

The maximization version of this problem has not been studied yet. In this version, given a set P consisting points in the plane which is partitioned into k clusters.

$$P = P_1 \cup P_2 \cup .... \cup P_k \qquad \forall i \neq j, \;\; P_i \cap P_j = \emptyset. \tag{10}$$

The GMST problem in the plane is to find maximum MST which consists of exactly one point from each cluster. Complexity of this problem and whether it has approximation algorithm with constant factor, would be studied in future.

References

[1] B. Bhattacharya, A. Ćustić, A. Rafiey, A. Rafiey, and V. Sokol. Approximation algorithms for generalized mst and tsp in grid clusters. In *Combinatorial Optimization and Applications*, pages 110–125. Springer, 2015.

[2] T. C. Biedl and M. Kaufmann. Area-efficient static and incremental graph drawings. In *European Symposium on Algorithms*, pages 37–52. Springer, 1997.

[3] F. Corinne, G. Alexander, et al. An approximation scheme for the generalized geometric minimum spanning tree problem with grid clustering. Technical report, 2004.

[4] R. Dorrigiv, R. Fraser, M. He, S. Kamali, A. Kawamura, A. López-Ortiz, and D. Seco. On minimum-and maximum-weight minimum spanning trees with neighborhoods. *Theory of Computing Systems*, 56(1):220–250, 2015.

[5] C. Feremans, A. Grigoriev, and R. Sitters. The geometric generalized minimum spanning tree problem with grid clustering. *4OR: A Quarterly Journal of Operations Research*, 4(4):319–329, 2006.

[6] E. Ihler. Minimum rectilinear steiner trees for intervals on two parallel lines. In *International Workshop on Graph-Theoretic Concepts in Computer Science*, pages 123–134. Springer, 1992.

[7] E. Ihler, G. Reich, and P. Widmayer. On shortest networks for classes of points in the plane. In *Computational Geometry-Methods, Algorithms and Applications*, pages 103–111. Springer, 1991.

[8] D. Lichtenstein. Planar formulae and their uses. *SIAM journal on computing*, 11(2):329–343, 1982.

[9] Y.-S. Myung, C.-H. Lee, and D.-W. Tcha. On the generalized minimum spanning tree problem. *Networks*, 26(4):231–241, 1995.

[10] P. C. Pop. *The generalized minimum spanning tree problem*. Twente University Press, 2002.

[11] P. C. Pop, W. Kern, and G. Still. An approximation algorithm for the generalized minimum spanning tree problem with bounded cluster size. 2001.

[12] S. Safra and O. Schwartz. On the complexity of approximating tsp with neighborhoods and related problems. *computational complexity*, 14(4):281–307, 2006.

Appendix

We prove lemma 3 in this part.

Lemma 3. *Proof.* In the structure A with k clusters (pairs of points), there are two different choices from every cluster which lead to optimal solution of the GMST problem.

Proof 2.1.in the structure A, every cluster regardless of choice of points will be connected to the next cluster in solution of EMST. Also cluster k will be connected to the cluster 1 or $k-1$. We consider L and R symbols, which are equivalent to selection of the left side point and the right side point in a cluster, respectively. Any selection of points in the clusters are shown with sequence of these symbols. As an example, LRR sequence shows that the left side point is chosen in the first cluster and the right side point is chosen in the second and third clusters. Now, we show the optimal solutions are $LLL....L$ sequence or $RRR....R$ sequence. In these cases because in EMST every cluster connects to the next cluster, weight of optimal solution of EMST is $\sqrt{5}(k-1)$. We prove this claim for $LLL....L$ sequence, and for $RRR....R$ sequence will be proved similarly. In order to prove this claim, we assume the right side point is chosen in one of the clusters. It means we have $LL...LRL...L$. Consequently, in the cluster in which the right side point is chosen and in the next and the former cluster, EMST is as Figure 4a or Figure 4b and in both cases weight of EMST is $\sqrt{10}+\sqrt{2}$. Therefore weight of EMST in the entire structure is $((k-3)\sqrt{5}+\sqrt{10}+\sqrt{2})$ and it is $(\sqrt{10}+\sqrt{2}-2\sqrt{5}) \approx 0.1$ more than weight of EMST when the left side points are chosen in all clusters. When the right side points are chosen in more than one consecutive cluster it means we have $LL...LRR...RL...L$, yet the weight of EMST is $((k-3)\sqrt{5}+\sqrt{10}+\sqrt{2})$ which is $(\sqrt{10}+\sqrt{2}-2\sqrt{5}) \approx 0.1$ more than Weight of EMST when the left side points are chosen in all clusters. Consequently every time when the consecutive left side points are chosen, at least 0.1 is added to the weight of EMST.

As an example in $LL...LR...RL...LR...RL...LL$ sequence, the weight of EMST is 0.2 more than when all of the right side points are chosen in all clusters. Optimal solutions of structure A with $k=8$ is shown in Figure 5. □

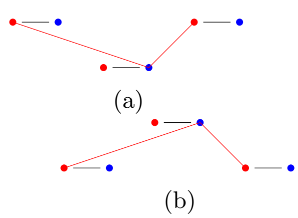


Figure 4: Part of EMST in $LL...LRL...L$ sequence

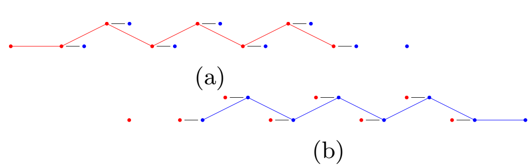


Figure 5: Optimal solutions of structure A with $k=8$.