comput. complex. **14** (2005), 281 – 307
1016-3328/05/040281–27
DOI 10.1007/s00037-005-0200-3

computational complexity

ON THE COMPLEXITY OF APPROXIMATING TSP WITH NEIGHBORHOODS AND RELATED PROBLEMS

Shmuel Safra and Oded Schwartz

Abstract. We prove that various geometric covering problems related to the Traveling Salesman Problem cannot be efficiently approximated to within any constant factor unless P = NP. This includes the Group-Traveling Salesman Problem (TSP with Neighborhoods) in the Euclidean plane, the Group-Steiner-Tree in the Euclidean plane and the Minimum Watchman Tour and Minimum Watchman Path in 3-D. Some inapproximability factors are also shown for special cases of the above problems, where the size of the sets is bounded. Group-TSP and Group-Steiner-Tree where each neighborhood is connected are also considered. It is shown that approximating these variants to within any constant factor smaller than 2 is NP-hard.
For the Group-Traveling Salesman and Group-Steiner-Tree Problems in dimension d, we show an inapproximability factor of $O(\log^{(d-1)/d} n)$ under a plausible conjecture regarding the hardness of Hyper-Graph Vertex-Cover.

Keywords. NP-optimization problems, approximation, hardness of approximation, inapproximability, TSP, TSP with neighborhoods.
Subject classification. 68Q17, 68Q25.

1. Introduction

The Traveling Salesman Problem (TSP) is a classical problem in combinatorial optimization, and has been studied extensively in many forms. It is the problem of a traveling salesman who has to visit n locations, returning eventually to the starting point. The goal may be to minimize the total distance traversed, driving time, or money spent on toll roads, where the cost (in terms of length units, time units, money or other) is given by an $n \times n$ matrix of non-negative weights. In the geometric TSP, the matrix represents distances in a Euclidean space. In other certain natural instances (e.g., time and money), while weights might not agree with a Euclidean metric, they still obey the triangle inequality, namely the cost of traversing from a to b is not higher than the cost of traversing from a to b via other points. Formally, the Traveling Salesman Problem can be defined as follows: given a set P of points in a metric space, find a traversal of shortest length visiting each point of P and returning to the starting point.

TSP in the plane. Finding the optimal solution of a given instance of TSP with triangle inequality is NP-hard, as obtained by a simple reduction from the Hamilton-Cycle problem. Even in the special case where the matrix represents distances between points in the Euclidean plane, it is also proved to be NP-hard (Garey *et al.* 1976; Papadimitriou 1977). The latter problem has a polynomial time approximation scheme (PTAS)—that is, for any $\varepsilon > 0$, there exists a polynomial time algorithm which guarantees an approximation of size at most $1+\varepsilon$ times the optimal solution (Arora 1998; Mitchell 1999). This, however, is not the case for the non-geometric variants.

Triangle inequality. In the general case, approximating TSP to within any constant factor is NP-hard (again, by a simple reduction from the Hamilton-Cycle problem). When only the triangle inequality is ensured, the best known algorithm gives a $\frac{3}{2}$ approximation ratio if weights are symmetric (Christofides 1976). If weights can be asymmetric (that is, the cost from a to b is not necessarily the same as the cost from b to a), the best known approximation ratio is $O(\log n)$ (Frieze *et al.* 1982). Although the asymmetric case may seem unnatural having the Euclidean metric intuition in mind, when weights represent measures other than length, or for example when the lengths are of one-way roads, the asymmetric formulation is natural.

In regard to the hardness of approximation, Papadimitriou & Vempala (2000) gave evidence that unless P = NP, the symmetric case cannot be efficiently approximated to within a factor smaller than $\frac{220}{219}$, and the asymmetric case to within a factor smaller than $\frac{117}{116}$. For bounded metrics Engebretsen & Karpinski (2001) showed hardness of approximation factors of $\frac{131}{130}$ and $\frac{174}{173}$ respectively.

Group-TSP. A natural generalization of this problem is the Group-TSP (G-TSP), known also by the names of the One-of-a-Set-TSP, TSP with neighborhoods and the Errand Scheduling problem. A traveling salesman has to meet n customers. Each of them is willing to meet the salesman in specified locations (referred to as a *region*). For instances in which each region contains exactly one point, this becomes the TSP problem. For instances in which all edges are of weight 1, this becomes the Hitting-Sets (or Set-Cover) problem. Another natural illustration of the G-TSP is the Errand Scheduling Problem as described in Slavik (1997). A list of n jobs (or errands) to be carried out is given, each of which can be performed in a few locations. The objective is to find a close tour of minimal length such that all jobs can be performed. That is, for every job on the list, there is at least one location on the tour, at which the job can be performed (it is allowed to perform more than one job at a single location). If every job can be performed in at most k locations, then we call this problem *k-G-TSP*. k-G-TSP (with symmetric weights) can be approximated to within $3k/2$ (Slavik 1997). This algorithm generalizes the $3/2$ approximation ratio of Christofides (1976) for $k \geq 1$.

As G-TSP (with triangle inequality) is a generalization of both TSP and Set-Cover, an inapproximability factor for any of those two problems holds for the G-TSP. Thus, by Feige (1998), Lund & Yannakakis (1994) and Raz & Safra (1997) G-TSP is hard to approximate to within a logarithmic factor. However, this is not trivially true for the geometric variant of G-TSP.

G-TSP in the plane. This problem was first studied by Arkin & Hassin (1994) who gave a constant approximation ratio algorithm for it where the regions (or neighborhoods) are well behaved in some form (e.g. consist of disks, parallel segments of equal length or translates of a convex region). Mata & Mitchell (1995) and Gudmundsson & Levcopoulos (1999) showed an $O(\log n)$ approximation ratio for arbitrary (possibly overlapping) polygonal regions. A constant factor approximation algorithm for the case where neighborhoods are disjoint convex fat objects was suggested by de Berg *et al.* (2002). Recently Dumitrescu & Mitchell (2003) gave a constant factor approximation algorithm for the case of arbitrary connected (i.e. path-wise connected) neighborhoods having comparable diameter, and a PTAS for the special case of pairwise disjoint unit disk neighborhoods. Note that the term "connected" refers to the geometric properties of each object forming a neighborhood, and not to the underlying weighted graph, which is of course full and thus connected, but in a different sense.

The best previously known approximation hardness for this problem is $\frac{391}{390} - \varepsilon \approx 1.003$ (de Berg *et al.* 2002). We improve this result to any constant factor, and give better results under a stronger complexity assumption.

Steiner Tree. Another related problem is the minimum Steiner spanning tree problem, or Steiner Tree problem (ST). A *Steiner tree* of S is a tree whose nodes contain the given set S. The nodes of the tree that are not the points of S are called *Steiner points*. A minimum spanning tree can be found in polynomial time. In contrast, finding a minimum *Steiner* spanning tree is an NP-hard problem. It remains NP-hard even in the Euclidean case (Garey *et al.* 1977), though a PTAS exists for this variant (Arora 1998; Mitchell 1999).

Group Steiner Tree. The Steiner tree notion can be generalized similarly to the generalization of TSP to G-TSP. In the Group Steiner Tree Problem (G-ST) (also known as the Class Steiner Problem, Tree Cover Problem and One-of-a-Set Steiner Problem) we are given an undirected graph with edge weights and subsets of the vertices. The objective is to find a minimum weighted tree having at least one vertex of each subset.

As G-ST is another generalization of set-cover (even when the weight function obeys the triangle inequality) any approximation hardness factor for set-cover applies to G-ST (Ihler 1992). Thus, by Feige (1998), Lund & Yannakakis (1994) and Raz & Safra (1997), G-ST is hard to approximate within a logarithmic factor. Halperin & Krauthgamer (2003) showed that approximating G-ST to within a poly-logarithmic factor is hard (their complexity assumption is stronger than $\mathrm{P} \neq \mathrm{NP}$). As in G-TSP, this is not trivially true for the geometric domain.

Slavik (1997) gave an $O(\log n)$ approximation algorithm for a restricted case of this problem and a $2k$-approximation algorithm for the variant in which sets are of size at most k. For sets of unbounded size, no constant approximation algorithm is known, even under Euclidean constraint (Mitchell 2000). If the weight function obeys the Euclidean metric in the plane, then, for some restricted variant of the problem, there is a polynomial time algorithm which approximates it within some (large) constant (a corollary of de Berg *et al.* 2002).

Minimum Watchman Tour and Minimum Watchman Path. The Minimum Watchman Tour (WT) and Minimum Watchman Path (WP) are the problems of a watchman who must have a view of n objects, while also trying to minimize the length of the tour (or path). These problems were extensively studied, and given some approximation algorithms as well as solving algorithms for special instances of the problem (Carlsson *et al.* 1999; Chin & Ntafos 1988; Gewali & Ntafos 1998; Mata & Mitchell 1995; Nilsson & Wood 1990; Xue-Hou *et al.* 1993).

1.1. Our results. We show that G-TSP in 2-D, G-ST in 2-D, WT in 3-D and WP in 3-D are all NP-hard to approximate to within any constant factor. This resolves a few open problems presented by Mitchell (2000) (open problems 21, 30 and problem 27—disconnected part). For dimension d, and under a plausible conjecture regarding the hardness of approximation of hypergraph vertex-cover, the hardness of approximation factor for G-TSP and G-ST becomes $O(\log^{(d-1)/d} n)$. These problems can be categorized according to three important parameters. One is the dimension of the domain; the second is whether each subset (region, neighborhood) is connected; and the third is whether sets are pairwise disjoint. For the G-TSP and G-ST problems in 2-D our results hold only if sets are allowed to be disconnected (but hold even for pairwise disjoint sets). If each set is connected (but sets are allowed to coincide), we show an inapproximability factor of $2-\varepsilon$ for both problems[1] (this result first appeared in Schwartz 2002). To achieve this we use an adaptation of a technique from de Berg *et al.* (2002). In 3-D our results hold for all parameter settings, that is, even when each set is connected and all sets are pairwise disjoint. We also show inapproximability factors of $\sqrt{k-1}/2\sqrt[4]{3}-\varepsilon$ and $\sqrt{k-1}/\sqrt{2}\sqrt[4]{3}-\varepsilon$ for the k-G-ST and k-G-TSP, respectively. The following table summarizes the main results for G-TSP and G-ST:

G-TSP and G-ST				
Dimension	2-D		3-D or more	
Pairwise disjoint sets	Yes	No	Yes	No
Connected sets	-	$2-\varepsilon$	$\forall c$	$\forall c$
Disconnected sets	$\forall c$	$\forall c$	$\forall c$	$\forall c$

Table 1.1: Inapproximability factors. The $\forall c$ indicates inapproximability *for every constant factor*.

1.2. Outline. We first prove our main theorem and show the approximation hardness factor for G-ST (Section 2). We extend this to hold for various parameter settings (pairwise disjoint regions in Section 2.1 and connected regions in 3-D in Section 2.2). We then deduce the same for G-TSP (Section 2.3). Variants of small regions are considered (Section 2.4). The hardness of G-TSP is then shown to hold for WT and WP problems in 3-D (Section 2.5). The proofs for G-TSP and G-ST in the plane with each region connected are given in Section 3. This is followed by a discussion (Section 4).

[1]Recently Kindler (2004) showed how to combine our reductions to G-TSP and G-ST, achieving an inapproximability factor of $2+\frac{1}{\pi}-\varepsilon$ for the G-TSP with connected regions.

1.3. Preliminaries. In order to prove inapproximability of a minimization problem, one usually defines a corresponding gap problem.

Definition 1.1. Let A be a minimization problem. gap-A-$[a, b]$ is the following decision problem: Given an input instance, decide whether

- there exists a solution of size at most a, or
- every solution of the given instance is of size larger than b.

If the size of the solution lies between these values, then any output suffices.

Clearly, for any minimization problem A, if gap-A-$[a, b]$ is NP-hard, then it is NP-hard to approximate A to within any factor smaller than b/a.

Our main result in this paper is derived by a reduction from the hyper-graph vertex-cover problem. A *hyper-graph* $G = (V, E)$ is a set of vertices V, and a family E of subsets of V, called edges. It is called k*-uniform* if all edges $e \in E$ are of size k, that is, $E \subseteq \binom{V}{k}$. A *vertex-cover* of a hyper-graph $G = (V, E)$ is a subset $U \subseteq V$ that "hits" every edge in G, that is, $e \cap U \neq \emptyset$ for all $e \in E$.

Definition 1.2. The *Ek-Vertex-Cover problem* is, given a k-uniform graph $G = (V, E)$, to find a minimum size vertex-cover U.

For $k = 2$ this is the vertex-cover problem on conventional graphs (VC). To prove the approximation hardness result of G-ST (for any constant factor) we use the following approximation hardness of hyper-graph vertex-cover:

Theorem 1.3 (Dinur *et al.* 2003). *For $k > 4$, Gap-Ek-Vertex-Cover-$[\frac{n}{k-1-\varepsilon}, (1-\varepsilon)n]$ is NP-hard.*

2. Group Steiner Tree and Group TSP in the plane

Definition 2.1 (G-ST). We are given $X = (P, R)$: a set P of n points in the plane, and a family R of subsets of P. A solution to X is a tree T such that every set $r \in R$ has at least one point in the tree, that is, $r \cap T \neq \emptyset$ for all $r \in R$. The size (length) of a solution T is the sum of the lengths of all its edges. The objective is to find a solution T of minimal length.

Let us now prove the main result:

Theorem 2.2. *G-ST is NP-hard to approximate to within any constant factor.*

Proof. The proof is by reduction from vertex-cover in hyper-graphs to G-ST. The reduction generates an instance X of G-ST such that the size of its minimum tree T is related to the size of the minimum vertex-cover U of the input graph $G = (V, E)$. Therefore, an approximation for T would imply an approximation for U, and hence the inapproximability factor known for Gap-Ek-Vertex-Cover yields an inapproximability factor for G-ST.

2.0.1. The construction. Given a k-uniform hyper-graph $G = (V, E)$ with $|V| = n$ vertices (we can assume that $\sqrt{n}$ and $\sqrt{n/k}$ are integers), we embed it in the plane to construct X. All the regions are subsets of points of a single square of the $\sqrt{n} \times \sqrt{n}$ section of the grid. Each point represents an arbitrary vertex of G and each region stands for an edge of G. Formally,

$$P = \{p_{v_i} \mid v_i \in V\}, \quad p_{v_i} = (i \bmod \sqrt{n}, \lfloor i/\sqrt{n} \rfloor).$$

We now define the set R of regions. For every $e \in E$ we define the region r_e to be the collection of k points on the grid, the vertices of the edge e:

$$R = \{r_e \mid e \in E\}, \qquad r_e = \{p_v \mid v \in e\}.$$

Claim 2.3 (Soundness). *If every vertex-cover U of G is of size at least $(1-\varepsilon)n$ then every solution T for X is of size at least $(1 - \varepsilon)n/2$.*

Proof. Inspect a Steiner tree that covers $(1-\varepsilon)n$ of the grid points. Relate each point on the (segments of the) tree to the nearest covered grid point. As every covered grid point is connected to the tree, the total length related to each covered grid point is at least $1/2$. Thus the size of the tree is at least $\frac{1}{2}(1 - \varepsilon)n$. □

Lemma 2.4 (Completeness). *If there is a vertex cover U of G of size at most n/t then there is a solution T for X of size at most $3n/\sqrt{t}$.*

Proof. We define $T_N(U)$, the natural tree according to a vertex-cover U of G, as follows (see Figure 2.1). We take a vertical segment along the first column of points, and horizontal segments along every d^{th} row, where $d = \sqrt{t}$. For every point p_v of $v \in U$ which is not already covered by the tree, we add a segment from it to the closest point q_v on any of the horizontal segments.

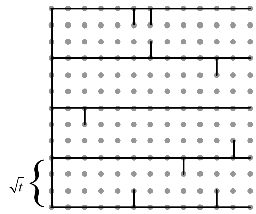

Figure 2.1: The natural tree.

Definition 2.5. The *natural tree* $T_N(U)$ of a subset $U \subseteq V$ is the polygon consisting of the following segments:

$$T_N(U) = \{((0,0),(0,\sqrt{n}))\} \cup \{((0,(i-1)\cdot d),(\sqrt{n},(i-1)\cdot d))\}_{i\in[\sqrt{n}/d]} \cup \{(p_v,q_v)\}_{v\in U}$$

where

$$q_{v_i} = \left(i \bmod \sqrt{n}, d\cdot \left\lfloor \frac{1}{d}\left\lfloor \frac{i}{\sqrt{n}}\right\rfloor + \frac{1}{2}\right\rfloor\right).$$

Thus, the natural tree contains $\sqrt{n/t}+1$ horizontal segments of length $\sqrt{n}$ each, a vertical segment of length $\sqrt{n}$, and at most n/t segments, each of length not more than $\sqrt{t}/2$. Therefore

$$|T_N(U)| \le \left(\sqrt{\frac{n}{t}}+2\right)\sqrt{n} + \frac{n}{t}\cdot\frac{\sqrt{t}}{2} < \frac{3n}{\sqrt{t}}. \qquad \square$$

Applying the soundness claim and Lemma 2.4 to Theorem 1.3 we conclude that for sufficiently large k, Gap-G-ST-$[\frac{3n}{\sqrt{k-1-\varepsilon}}, \frac{1}{2}(1-\varepsilon)n]$ is NP-hard. Thus, as k can be arbitrarily large, it is NP-hard to approximate G-ST to within any constant factor.

2.1. Group Steiner Tree—pairwise disjoint regions. We now show that the factors for G-ST hold even when regions are pairwise disjoint.

Claim 2.6. *The approximation hardness factors for G-ST (Theorem 2.2) apply even when regions are pairwise disjoint.*

Proof (sketch). We extend the reduction, such that the generated instance X contains only pairwise disjoint sets. For every $e \in E$ and $v \in e$, we redefine the point $p_v \in r_e^{\mathrm{out}}$: we replace p_v with a set of new points $p_{v,e}$, shifted to a distance of at most ε' of the original location (where $\varepsilon' > 0$ is arbitrarily small), such that for any $e_i, e_j \in E$ there are no joint points of the regions r_{e_i}, r_{e_j}. Clearly this does not affect any of the previous proofs. ■

2.2. Group Steiner Tree—connected regions in 3-D

CLAIM 2.7. *The approximation hardness factors for G-ST (Theorem 2.2) apply in 3-D even when regions are pairwise disjoint, and each region is connected.*

PROOF (sketch). The above reduction generates a two-dimensional instance in which the regions are not connected. One can show that the approximation hardness result applies for connected regions in dimension three or higher. This is achieved by adding segments to connect the points of every region in such a way that the optimal solution T does not change. For each relevant point p in the plane (i.e. a point which is part of a region) we add a segment from p to a matching point on an enlarged copy of the given plane instance, in which distances are multiplied by a large factor (say, n^4), located parallel to the original instance, at great height (say, n^4) above it. We call that segment the *up-going segment of* p (see Figure 2.2).

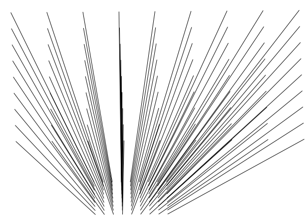

Figure 2.2: 2-D to 3-D connected.

Each region r_e will be redefined to contain the up-going segments of all $p \in r$. We will now describe new segments which connect the different segments of a region; these new segments will be called the *connecting segments*. Any two up-going segments of the same region are connected using a new segment, parallel to the plane. The new segments are added at each consecutive multiple of n^2 starting from height n^3. They are added in a way which ensures that none of the points on them are at a distance n^2 or less from other segments. To this end, we may, should the need arise, use three segments rather than one, in order to connect two up-going segments. Each region is once again redefined to contain both its up-going segments and its connecting segments. Note that if the original regions were pairwise disjoint then the new regions are as well. Trivially, an optimal solution T never contains points outside the original plane. ∎

2.3. Group TSP in the plane. We next show that the hardness factor shown for G-ST holds for G-TSP with the same parameter settings.

Definition 2.8 (G-TSP). In G-TSP in the plane we are given $X = (P, R)$: a set P of points in the plane, and a family R of n subsets of P. A solution to the G-TSP is a traversal T such that every set (region) $r \in R$ has at least one point in the traversal, that is, $r \cap T \neq \emptyset$ for all $r \in R$. The objective is to find a solution T of minimal length.

Claim 2.9. *If there exists an efficient α-approximation for G-TSP then an efficient 2α-approximation for G-ST also exists.*

Proof. The size of a minimum tree of G-ST, $T^*_{\text{G-ST}}$, is smaller than the size of a minimum tour of G-TSP, $T^*_{\text{G-TSP}}$, of the same given instance (as one can take a tour and transform it to a tree by removing one of its edges). On the other hand, $T^*_{\text{G-ST}}$ is at most $2T^*_{\text{G-TSP}}$ (as one can take two copies of the same tree to have a tour). Thus

$$T^*_{\text{G-ST}} < T^*_{\text{G-TSP}} \leq 2T^*_{\text{G-ST}}. \qquad \square$$

Therefore, by this claim and Theorem 2.2 we conclude:

Corollary 2.10. *G-TSP in the plane is NP-hard to approximate to within any constant factor.*

This holds for the same parameter settings as in G-ST (i.e. disconnected sets for 2-D, and both connected and disconnected for 3-D and higher dimensions).

2.4. Small regions in the plane—Group TSP and Group Steiner Tree. Let k-G-ST and k-G-TSP be the variants of G-ST and G-TSP respectively, where each region is comprised of k points. We use a triangular grid (see Figure 2.3) instead of the grid used in the original proof (of Theorem 2.2) to achieve better constants of inapproximability for these variants.

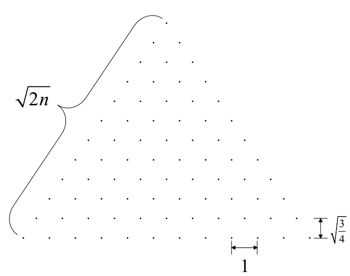

Figure 2.3: The triangular construction.

Claim 2.11 (Soundness). *If any vertex-cover U of G is of size at least an, then any solution is of size at least $(1-\varepsilon)an$ for the k-G-TSP and of size at least $\frac{1}{2}(1-\varepsilon)an$ for the k-G-ST.*

Proof. Let T be a solution of X. Let U be a set of vertices that correspond to points of the grid, covered by T. Clearly T is a solution only if U is a vertex-cover of G, hence $|U| \geq an$. Consider a circle of radius $1/2-\varepsilon/2$ around each covered grid point. All these circles are pairwise disjoint.

For the case of k-G-TSP, each circle contains at least two legs of the path, each of length at least $\frac{1}{2}-\frac{\varepsilon}{2}$, hence the total length of T is

$$|T| \geq an(1-\varepsilon).$$

For the case of k-G-ST, each circle contains a portion of the tree of length at least $1/2-\varepsilon/2$, hence the total length of T is

$$|T| \geq an\left(\frac{1}{2}-\frac{\varepsilon}{2}\right). \qquad \square$$

We next give the completeness claims for both problems:

Claim 2.12 (Completeness—k-G-ST). *If U is a minimum vertex-cover of G, then there exists a solution of X (the k-G-ST generated instance) of size at most*

$$\frac{n}{d}+\sqrt{2n}+|U|\cdot\frac{\sqrt{3}}{2}\cdot\frac{d}{2}$$

for any parameter $d, 0<d<\sqrt{n}$.

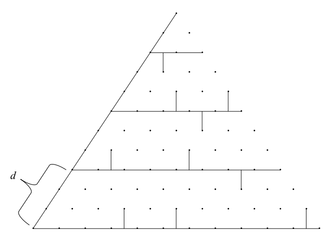

Figure 2.4: The natural tree.

Proof. The proof is by inspecting the natural tree $T_N(U)$ of a minimum vertex-cover U (as demonstrated in Figure 2.4). $T_N(U)$ contains horizontal segments of decreasing length, every d^{th} row, starting with the bottom row.

It contains another segment of length $\sqrt{2n}$ which connects all leftmost points of the rows; and at most $|U|$ vertical segments, connecting each point of the vertex-cover U to the nearest horizontal segment (of length at most $\frac{\sqrt{3}}{2} \cdot \frac{d}{2}$ each).

The total length of the horizontal segments is

$$\sum_{i=0}^{\sqrt{2n}/d} (\sqrt{2n} - i \cdot d) = \frac{n}{d}.$$

Therefore, the total length of $T_N(U)$ is

$$|T_N(U)| \leq \frac{n}{d} + \sqrt{2n} + |U| \cdot \frac{\sqrt{3}}{2} \cdot \frac{d}{2}. \qquad \square$$

For $|U| = \frac{n}{k-1-\varepsilon}$ and $d = \sqrt{\frac{4(k-1-\varepsilon)}{\sqrt{3}}}$ we conclude that

$$T_N(U) = \frac{\sqrt[4]{3}\, n}{\sqrt{k-1-\varepsilon}} + \sqrt{2n}.$$

Therefore by applying the known hardness of Hyper-Graph Vertex-Cover (Theorem 1.3) to the soundness claim (Claim 2.11) and the above completeness claim (Claim 2.12) we deduce the following:

COROLLARY 2.13. *For $k > 4$, k-G-ST is hard to approximate to within*

$$\frac{\sqrt{k-1}}{2\sqrt[4]{3}} - \varepsilon.$$

CLAIM 2.14 (Completeness: k-G-TSP). *If U is a minimum vertex-cover of G, then there exists a solution of X (the k-G-TSP generated instance) of size at most*

$$\frac{n}{d} + 2\sqrt{2n} + |U| \cdot \frac{\sqrt{3}}{2} \cdot d$$

for any parameter $d, 0 < d < \sqrt{n}$.

PROOF. The proof is by inspecting the natural tour $T_N(U)$ of a minimum vertex-cover U (as demonstrated in Figure 2.5). $T_N(U)$ contains horizontal segments, as in the natural tree, of total length n/d (see Claim 2.12). It has another segment of length $\sqrt{2n}$, connecting the rightmost and topmost points of the grid; $\sqrt{2n}/d$ segments of length d (from a horizontal segment to

Proof (sketch; see Figure 2.6). We show a gap preserving reduction from an instance X of G-TSP in the plane with pairwise disjoint (disconnected) regions to an instance Y of WT. The idea is to use opaque objects. Each object conceals one point in it, but has a few holes. The point can only be seen through the openings. A watchman who wants to see all concealed points may visit one opening of each such object.

We can use such an object to represent a region (composed of a finite number of points) of G-TSP: each point of the region is represented by a hole in the object. We do this as follows.

Each region is represented by an octopus-shaped figure (see Figure 2.6). This object is comprised of hollow cylinders. If a region contains k points then its corresponding object is composed of k cylinders. The bottom part of each cylinder is positioned on the corresponding point. All k cylinders converge at the top. The concealed point is located at that junction.

Thus a watchman of Y has to visit one opening of each object, which immediately translates to a solution for X. Therefore by Corollary 2.10 we obtain the same hardness factors for WT in 3-D. ■

Claim 2.21. *If there exists an efficient α-approximation for WP then an efficient 2α-approximation for WT also exists.*

Proof. The size of a minimum path of WP, T^*_{WP}, is smaller than the size of a minimum tour of WT, T^*_{WT}, of the same given instance (as one can take a tour and transform it to a path by removing one of its edges). On the other hand, T^*_{WP} is at most $2T^*_{\mathrm{WP}}$ (as one can take two copies of the same path to have a tour). Thus

$$T^*_{\mathrm{WP}} < T^*_{\mathrm{WT}} \leq 2T^*_{\mathrm{WP}}. \qquad \square$$

Thus the hardness factors of WT apply for WP too.

3. The case of connected regions

In the following proofs, we show that G-ST and G-TSP in the plane where *every region is connected* are NP-hard to approximate to within any constant factor smaller than 2. This extends a known factor of $\frac{391}{390} - \varepsilon \approx 1.003$ by de Berg *et al.* (2002) for G-TSP. These reductions are similar to theirs but differ in the problem we reduce from (vertex-cover on hyper-graphs instead of vertex-cover on graphs of bounded degree) and in some parts of the gadgets constructed.

3.1. Group TSP—connected regions in the plane. We consider the 2-D variant of G-TSP in which each set is connected.

For every edge $e \in E$ we have a region r_e^{out} composed of points on the outer circle relating to the vertices of e:

$$r_e^{\mathrm{out}} = \{q_v \mid v \in e\}, \qquad R_E = \{r_e^{\mathrm{out}} \mid e \in E\}.$$

One can easily amend each of the disconnected regions (which are all in R_E) to be connected without affecting the following proof. For details see the last part of this section.

Proof's idea. We are next going to show that the most efficient way to traverse X is by traversing all points on the inner circle (say counterclockwise), detouring to visit the closest points on the outer circle, for every point that corresponds to a vertex in the minimum vertex-cover of G (see Figure 3.1). More formally, we have

DEFINITION 3.2. The *natural tour* $T_N(U)$ of a subset $U \subseteq V$ is the closed polygon consisting of the following segments:

$$\begin{gathered} T_N(U) = T_{\mathrm{in}} \cup T_{\mathrm{out}}, \\ T_{\mathrm{in}} = \{(p_{v,j+1}, p_{v,j+2}) \mid v \in V,\, j \in [l-2]\} \cup \{(p_{v_i,l}, p_{v_{i+1 \bmod n},1}) \mid i \in [n]\}, \\ T_{\mathrm{out}} = \{(p_{v,i}, q_v) \mid v \in U,\, i \in [2]\}. \end{gathered}$$

Let us consider the length $|T_N(U)|$ of this tour. The natural tour $T_N(U)$ consists of $nl - |U|$ segments of size $\varepsilon = 1/nl$ (on the inner circle), $|U|$ segments for the detourings of size $1/2n$, and $|U|$ segments of size in the range $(1/2n, 1/2n + \varepsilon)$. Thus

$$1 + \frac{|U|}{n}(1-\delta) \leq T_N(U) \leq 1 + \frac{|U|}{n}$$

(for some $0 < \delta < 1/l$). The exact length of $T_N(U)$ can be computed, but it is not important for our purpose. Thus, by the upper bound on $|T_N(U)|$ we have:

CLAIM 3.3 (Completeness). *If there is a vertex-cover U of G of size bn, then there is a solution of X of length at most $1 + b$.*

We next show

CLAIM 3.4 (Soundness). *If any vertex-cover U of G is of size at least an, then any solution of X is of length at least $1 + a - 3/l$.*

Proof. Let T be a solution of X. Clearly T covers all points of P_{inner} (otherwise it is not a solution for X). Let U be the set of vertices that correspond to points on the outer circle, visited by T:

$$U = \{v \mid q_v \in T \cap P_{\text{outer}}\}.$$

Clearly T is a solution only if U is a vertex-cover of G, hence $|U| \geq an$. Consider a circle of radius $1/2n - \varepsilon$ around each covered point of the edge regions q_v ($v \in U$). All these circles are pairwise disjoint (as the distance between two points of the edge regions is at least $1/n$). Each of them contains at least two legs of the path, each of length at least $1/2n - \varepsilon$. In addition the tour visits all the points of the vertex regions, and at least $nl - 3n$ of them are at distance of at least ε from any of the above circles. Thus the in-going path to at least $nl - 3n$ extra points is of length at least ε each. Hence the total length of T is

$$|T| \geq |U| \cdot 2 \cdot \left(\frac{1}{2n} - \varepsilon\right) + (nl - 3n)\varepsilon = 1 + a - 2a\varepsilon - \frac{3}{l} \geq 1 + a - \frac{3}{l}. \qquad \square$$

Hence by the soundness and completeness claims, and fixing l to be sufficiently large we have the following:

Lemma 3.5. *If Gap-Ek-Vertex-Cover-$[b, a]$ is NP-hard then for any $\varepsilon' > 0$, it is NP-hard to approximate G-TSP in the plane with connected regions to within $\frac{1+a}{1+b} - \varepsilon'$.*

Plugging in the known gap for vertex-cover in hyper-graphs (Theorem 1.3) we conclude that G-TSP is NP-hard to approximate to within

$$\frac{1 + 1 - \varepsilon''}{1 + \frac{1}{k-1-\varepsilon''}} - \varepsilon',$$

hence, for arbitrarily small $\varepsilon > 0$ and for sufficiently large k, G-TSP is NP-hard to approximate to within $2 - \varepsilon$, even if each region is connected. $\square$

3.1.1. Making each region connected. To make each region $r_e \in R_E$ connected (see Figure 3.2), we add segments connecting each of the points on the outer circle to the closest point on a concentric circle (the *outmost circle*, C), of radius ρ_{outmost} suitably large (say, n); namely, for each $q_v \in P_{\text{outer}}$, $q_v = (\rho + 1/2n, \alpha)$ we add the segment

$$l_v = \left[\left(\rho + \frac{1}{2n}, \alpha\right), (\rho_{\text{outmost}}, \alpha)\right].$$

CLAIM 3.8. *The size of the natural tree is* $|T_N(U)| = n - 1/l + |U|$.

PROOF. $T_N(U)$ contains a horizontal segment of length $n - 1/l$ and $|U|$ vertical segments of length 1 each. □

CLAIM 3.9. *A natural tree* $T_N(U)$ *is a solution of* X *if and only if* U *is a vertex-cover of* G.

PROOF. If U is a vertex-cover of G then for every $e \in E$ there is at least one vertex $v \in e \cap U$, therefore each of the regions r_e^{top} and r_e^{bottom} is covered at least once, at the points q_v^{top} and q_v^{bottom} respectively (and trivially, in any natural tree, all regions of R_V are covered).

If U is not a vertex-cover of G then there is an edge $e \in E$ such that $e \cap U = \emptyset$, thus no point of r_e^{top} (and r_e^{bottom}) is covered, and $T_N(U)$ is not a solution of X. □

We therefore have:

CLAIM 3.10 (Completeness). *If there is a vertex-cover* U *of* G *of size* bn, *then there is a solution of* X *of length at most* $n(1+b) - 1/l$.

CLAIM 3.11 (Soundness). *If any vertex-cover* U *of* G *is of size at least* an, *then any solution of* X *is of length at least* $n(1+a) - (1+2an)/l$.

PROOF. Let T be a solution of X. Clearly T covers all points of P_{inner} (otherwise it is not a solution for X). Let U^{top} be the set of vertices that correspond to points on the top segment covered by T:

$$U_{\text{top}} = \{v \mid q_v^{\text{top}} \in T \cap P_{\text{top}}\},$$

and similarly

$$U_{\text{bottom}} = \{v \mid q_v^{\text{bottom}} \in T \cap P_{\text{bottom}}\}.$$

Clearly T is a solution only if each of U_{top} and U_{bottom} is a vertex-cover of G. Hence $|U_{\text{top}}| \geq an$ and $|U_{\text{bottom}}| \geq an$. Consider a circle of radius $1/2 - 1/l$ around each point covered of the top and bottom segments. All these circles are pairwise disjoint. Each of them contains a part of the tree of length at least $1/2 - \varepsilon$. In addition the tree covers all the points of the vertex regions. Thus we account for extra length of at least $n - 1/l$, and the total length of T is

$$|T| \geq |U_{\text{top}}| \cdot \left(\frac{1}{2} - \frac{1}{l}\right) + |U_{\text{bottom}}| \cdot \left(\frac{1}{2} - \frac{1}{l}\right) + n - \frac{1}{l} \geq n + an - \frac{1+2an}{l}. \quad \square$$

4. Discussion

We have shown that G-TSP, G-ST, WP and WT cannot be efficiently approximated to within any constant factor unless P = NP. In this respect Group-TSP and Group-ST seem to behave more like the Set-Cover problems, rather than the geometric-TSP and geometric-Steiner tree problems.

These reductions illustrate *the importance of gap location*; the approximation hardness result for hyper-graph vertex-cover (see Dinur *et al.* 2003) is weaker than that of Feige (1998), in the sense that the gap ratio is smaller (but works, of course, for the bounded variant). However, their gap location, namely, their almost perfect soundness (Dinur *et al.* 2003, Lemma 4.3), is a powerful tool (see for example Petrank 1994). In the reductions shown here this aspect plays an essential role. We conjecture that the two properties can be combined:

CONJECTURE 4.1. *Gap-Hyper-Graph-Vertex-Cover-*$[O(n/\log n), (1-\varepsilon)n]$ *is intractable.*

Using the exact same reductions, this will extend the known approximation hardness factors of G-TSP, G-ST, WT and WP, as follows:

COROLLARY 4.2. *If Conjecture* 4.1 *is correct then approximating G-TSP in the plane and G-ST in the plane to within* $O(\log^{1/2} n)$ *is intractable, and approximating WT and WP in 3-D to within* $O(\log^{1/2} n)$ *is also intractable.*

This conjecture also implies the following corollary.

COROLLARY 4.3. *If Conjecture* 4.1 *is correct then approximating G-TSP in dimension* $d > 1$ *and G-ST in dimension* $d > 1$ *to within* $O(\log^{(d-1)/d} n)$ *is hard.*

PROOF (sketch). By the simplest generalization of the proof for disconnected sets in 2-D. All the regions are subsets of a section of the d-dimensional grid. The section is a single d-dimensional cube, with sides of length $\sqrt[d]{n}$. ∎

Table 4.1 summarizes our result, assuming correctness of Conjecture 4.1.

An interesting open problem is whether the square root loss of the approximation hardness factor in the 2-D variant is merely a fault of this reduction or is intrinsic to the plane version of these problems; i.e., is there an approximation with a ratio smaller than $\ln n$ for the plane variants? Are there approximations to the G-TSP and G-ST that perform better in the plane variants than Slavik's (1997) approximations for these problems with triangle inequality only? Does

G-TSP and G-ST						
Dimension	2-D		3-D		d	
Pairwise disjoint sets	Yes	No	Yes	No	Yes	No
Connected sets	-	$2-\varepsilon$	$\log^{\frac{1}{2}} n$	$\log^{\frac{1}{2}} n$	$\log^{\frac{d-2}{d-1}} n$	$\log^{\frac{d-2}{d-1}} n$
Disconnected sets	$\log^{\frac{1}{2}} n$	$\log^{\frac{1}{2}} n$	$\log^{\frac{2}{3}} n$	$\log^{\frac{2}{3}} n$	$\log^{\frac{d-1}{d}} n$	$\log^{\frac{d-1}{d}} n$

Table 4.1

higher dimension in these problems impose an increase in complexity? Other open problems remain for various parameter settings. A most basic variant of G-TSP and G-ST, namely in 2-D, where every region is connected, and regions are pairwise disjoint, remains open, as well as the WT and WP in 2-D (open problem 29 of Mitchell 2000).

Acknowledgements

Many thanks to Shakhar Smorodinsky, who first brought these problems to our attention; to Matthew J. Katz, for his stimulating lecture on his results; and to Vera Asodi, Guy Kindler and Manor Mendel for their sound advice and insightful comments.

This research was supported by the Israeli Science Foundation (grant no. 230/02).

References

E. Arkin & R. Hassin (1994). Approximation algorithms for the geometric covering salesman problem. *Discrete Appl. Math.* **55**, 197–218.

S. Arora (1998). Polynomial time approximation schemes for Euclidean traveling salesman and other geometric problems. *J. ACM* **45**, 753–782.

M. de Berg, J. Gudmundsson, M. J. Katz, C. Levcopoulos, M. H. Overmars & A. F. van der Stappen (2002). TSP with neighborhoods of varying size. In *Annual European Symposium on Algorithms*, 187–199.

S. Carlsson, H. Jonsson & B. J. Nilsson (1999). Finding the shortest watchman route in a simple polygon. *Discrete Comput. Geom.* **22**, 377–402.

W. Chin & S. Ntafos (1988). Optimum watchman routes. *Inform. Process. Lett.* **28**, 39–44.

N. Christofides (1976). Worst-case analysis of a new heuristic for the traveling salesman problem. Technical report, Graduate School of Industrial Administration, Carnegy-Mellon Univ.

I. Dinur, V. Guruswami, S. Khot & O. Regev (2003). A new multilayered PCP and the hardness of hypergraph vertex cover. In *Proc. 35th ACM Symposium on Theory of Computing*, ACM Press, 595–601.

A. Dumitrescu & J. S. B. Mitchell (2003). Approximation algorithms for TSP with neighborhoods in the plane. *J. Algorithms* **48**, 135–159.

L. Engebretsen & M. Karpinski (2001). Approximation hardness of TSP with bounded metrics. In *Annual International Colloquium on Automata, Languages and Programming*, 201–212.

U. Feige (1998). A threshold of $\ln n$ for approximating set cover. *J. ACM* **45**, 634–652.

A. Frieze, G. Galbiati & F. Maffioli (1982). On the worst-case performance of some algorithms for the asymmetric travelling salesman problem. *Networks* **12**, 23–39.

M. R. Garey, R. L. Graham & D. S. Johnson (1976). Some NP-complete geometric problems. In *Conf. Record 8th Annual ACM Symposium on Theory of Computing* (Hershey, PA), 10–22.

M. R. Garey, R. L. Graham & D. S. Johnson (1977). The complexity of computing Steiner minimal trees. *SIAM J. Appl. Math.* **32**, 835–859.

L. Gewali & S. C. Ntafos (1998). Watchman routes in the presence of a pair of convex polygons. *Inform. Sci.* **105**, 123–149.

J. Gudmundsson & C. Levcopoulos (1999). A fast approximation algorithm for TSP with neighborhoods. *Nordic J. Comput.* **6**, 469–488.

E. Halperin & R. Krauthgamer (2003). Polylogarithmic inapproximability. In *Proc. 35th ACM Symposium on Theory of Computing*, ACM Press, 585–594.

J. Holmerin (2002). Vertex cover on 4-regular hyper-graphs is hard to approximate within $2 - \varepsilon$. In *ACM Symposium on Theory of Computing (STOC)*, 544–552.

E. Ihler (1992). The complexity of approximating the class Steiner tree problem. In *Graph-Theoretic Concepts in Computer Science*, G. Schmidt and R. Berghammer (eds.), Lecture Notes in Comput Sci. 570, Springer, 85–96.

G. Kindler (2004). Personal communication.

C. Lund & M. Yannakakis (1994). On the hardness of approximating minimization problems. *J. ACM* **41**, 960–981.

C. Mata & J. S. B. Mitchell (1995). Approximation algorithms for geometric tour and network design problems. In *Proc. 11th Annual Symposium on Computational Geometry*, ACM Press, 360–369.

J. S. B. Mitchell (1999). Guillotine subdivisions approximate polygonal subdivisions: a simple polynomial-time approximation scheme for geometric TSP, k-MST, and related problems. *SIAM J. Comput.* **28**, 1298–1309.

J. S. B. Mitchell (2000). *Geometric Shortest Paths and Network Optimization*. Elsevier Science, preliminary edition.

B. J. Nilsson & D. Wood (1990). Optimum watchmen in spiral polygons. In *CCCG: Canadian Conference in Computational Geometry*, 269–272.

C. H. Papadimitriou (1977). Euclidean TSP is NP-complete. *Theoret. Comput. Sci.* **4**, 237–244.

C. H. Papadimitriou & S. Vempala (2000). On the approximability of the traveling salesman problem (extended abstract). In *Proc. 32nd Annual ACM Symposium on Theory of Computing*, ACM Press, 126–133.

E. Petrank (1994). The hardness of approximation: gap location. *Comput. Complexity* **4**, 133–157.

R. Raz & S. Safra (1997). A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *Proc. 29th Annual ACM Symposium on Theory of Computing*, ACM Press, 475–484.

O. Schwartz (2002). On the hardness of approximating TSP with neighborhoods, group TSP and group Steiner tree. Master's thesis, Tel-Aviv Univ.

P. Slavik (1997). The errand scheduling problem. Technical Report 97-02, SUNY at Buffalo.

T. Xue-Hou, T. Hirata & Y. Inagaki (1993). An incremental algorithm for constructing shortest watchman routes. *Internat. J. Comput. Geom. Appl.* **3**, 351–365.

Manuscript received 24 December 2003

Shmuel Safra
School of Computer Science
Tel-Aviv University
Tel-Aviv 69978
ISRAEL
safra@math.tau.ac.il
www.cs.tau.ac.il/~safra

Oded Schwartz
School of Computer Science
Tel-Aviv University
Tel-Aviv 69978
ISRAEL
odedsc@tau.ac.il
www.cs.tau.ac.il/~odedsc