Approximation Algorithms for Euclidean Group TSP

Khaled Elbassioni, Aleksei V. Fishkin, Nabil H. Mustafa, and René Sitters

Max-Planck-Institut für Informatik, Saarbrücken, Germany
Email: {elbassio,avf,nnmustafa,sitters}@mpi-sb.mpg.de

Abstract. In the Euclidean group Traveling Salesman Problem (TSP), we are given a set of points \( P \) in the plane and a set of \( m \) connected regions, each containing at least one point of \( P \). We want to find a tour of minimum length that visits at least one point in each region. This unifies the TSP with Neighborhoods and the Group Steiner Tree problem. We give a \((9.1 + 1)\)-approximation algorithm for the case when the regions are disjoint \( \alpha \)-fat objects with possibly varying size. This considerably improves the best results known, in this case, for both the group Steiner tree problem and the TSP with Neighborhoods problem. We also give the first \( O(1) \)-approximation algorithm for the problem with intersecting regions.

1 Introduction

A salesman wants to meet a set of potential buyers. Each buyer indicates a set of potential locations where he or she can meet the buyer. The salesman would like to minimize the total length of the tour required to meet all the potential buyers. How to construct such a tour? This problem is a generalization of the classical Traveling Salesman Problem (TSP), and hence is NP-hard. More formally, the problem we study can be stated as follows.

Euclidean Group TSP. Given a set \( P \) of points in the Euclidean plane, and \( m \) subsets \( \{S_1, \ldots, S_m\} \) of \( P \), one has to construct a tour on a set \( P' \subseteq P \) such that \( P' \) contains at least one point from each subset (group) \( S_i \). The objective is to minimize the length of the tour.

This model unifies two important separate bodies of research – the Group Steiner Tree problem, and the Euclidean TSP with Neighborhoods problem, which arise in VLSI design [RW90], and routing-related applications [Mit00]. We describe these problems in more detail.

Group Steiner Tree. Given a graph \( G \) on \( n \) vertices with non-negative weights on the edges, and \( m \) subsets of vertices, the group Steiner tree problem calls for computing a sub-tree of \( G \) which contains at least one vertex from each subset (group), and whose total length is minimized.
Slavik [Sha97] presented an $O(k)$-approximation algorithm for the Group
Steiner Tree Problem in the metric case, where $k$ is the maximum group size.
Using probabilistic tree embeddings together with randomized rounding, Garg
et al. [GKR00] obtained a randomized $O((\log^3 n \log m)\text{-approximation algorithm for the general case. On the negative side, it was observed (see e.g. [Sha96]) that the problem is a direct generalization of the set covering problem, and hence is NP-hard to approximate within a factor of $o(\log m)$.

A natural question is whether the above approximation ratio can be improved, if we consider restricted versions of the problem. For example, when the metric is Euclidean and each group is induced by a geometric object, such as a disk in the plane.

**Euclidean TSP with Neighborhoods (TSPN).** Given $m$ connected geometric regions $S_1, \ldots, S_m$ in the Euclidean plane, find a minimum length tour that hits each region at least once.

The TSPN is a generalization of the classical Euclidean TSP which is known to be NP-hard [GGJ76,Pap77]. A 3/2-approximation algorithm, that works for any metric, was given by Christofides [Chr76]. Arora and Mitchell [Aro98,Mit99] independently obtained a $(1 + \varepsilon)$-approximation algorithm.

The TSPN was first studied by Arkin and Hassin [AH94]. They presented constant-factor approximations for the case where the geometric regions are translates of disjoint convex polygons, and for disjoint unit disks. For the general problem, Mata and Mitchell [MM95], and later Gudmundsson and Lévêque [GL00], gave an $O(\log m)$-approximation algorithm. Dumitrescu and Mitchell [DM03] gave an $O(1)$-approximation algorithm for intersecting unit disks. For disjoint varying-sized convex fat regions de Berg et al. [dBGK+ar] present a sophisticated algorithm with approximation ratio $12000\alpha^3$, where $\alpha$ is a measure of fatness of the regions. Their algorithm uses Slavik’s algorithm [Sha97] for the Group Steiner tree problem, mentioned above. On the hardness side, Safra
and Schwartz [SS03] showed that approximating Euclidean TSPN within $(2-\varepsilon)$ is NP-hard.

By viewing each geometric region $S_i$ as an infinite set of points, the TSPN becomes a special case of the Euclidean Group TSP. In this context, it is natural to study the Euclidean Group TSP in which the point sets are separated by geometric objects, such as disks or more generally fat objects (see Figure 1).

**Our Results.** As mentioned above, our problem relates to both the Group Steiner Tree and the TSP with neighborhoods problems. Our algorithms improve and give new results in both areas.

In Section 3 we consider the problem with disjoint regions and give a $(9.1\alpha + 1)$-approximation algorithm for the problem where groups are enclosed by non-intersecting $\alpha$-fat objects of arbitrary size. This improves the previous best result on the TSP with neighborhoods problem in several ways. First, we dramatically reduce the previous best approximation factor of $12,000\alpha^3$ [dBGK+ar]. Second, our groups are arbitrary point sets $S_1, \ldots, S_m$, separated by fat objects in the Euclidean plane, whereas previous results only deal with the (continuous) case where each set $S_i$ is the (infinite) set of points belonging to an object. We also do not require the objects to be convex. Furthermore, our algorithm yields an $O(\alpha/\sqrt{d})$-approximate solution for any dimension $d$. In contrast, it was shown in [SS03] that TSP with neighborhoods in $\mathbb{R}^d$ is unlikely to be approximable within $O(\log^{1/2} n)$, and thus, for $d \geq 3$, it is unlikely that there is a approximation factor independent of both $n$ and $\alpha$.

In Section 4 we consider the variant of the problem in which the instances are defined by $m$ sets of points in the Euclidean plane and a set of $m$ disks. Disks may intersect and each disk defines a group consisting of the enclosed points. Clearly, the intersection between disks admits much more complicated systems of subsets than what we can get from the non-intersecting sets. We present an $O(1)$-approximation algorithm for this problem.

## 2 Preliminaries

We consider instances of the Euclidean Group TSP in which the sets $S_i$ are contained in geometric regions in the plane. Formally, we define a *region* as a connected and closed subset of the Euclidean plane $\mathbb{R}^2$. An instance of our problem is given by a finite set of points $P$ and $m$ subsets $S_1, \ldots, S_m$ of $P$. The subsets have the property that there exist regions $O_1, O_2, \ldots, O_m$ such that $S_i = P \cap O_i$. A solution is given by a tour (spanning tree) on a subset $P'$ of $P$ such that $P' \cap S_i \neq \emptyset$. The objective is to minimize the length of the tour. We denote an optimal tour (spanning tree) by $\text{OPT}$, and its length by $|\text{OPT}|$.

In this paper we restrict to so called *fat* regions. The definition of fatness we use here was introduced by Van der Stappen [Sta94] and used by De Berg et al. [dBGK+ar] in their paper on the TSP with neighborhoods problem.
**Definition 1.** An object \( O \subseteq \mathbb{R}^2 \) is said to be \( \alpha \)-fat if for any disk \( \Theta \) which does not fully contain \( O \) and whose center lies in \( O \), the area of the intersection of \( O \) and \( \Theta \) is at least \( 1/\alpha \) times the area of \( \Theta \).

Notice for example that the plane \( \mathbb{R}^2 \) has fatness 1, a halfspace has fatness 2 and a disk has fatness 4. We define the **size** of an object as the diameter of its smallest enclosing disk.

**Lemma 1 (Packing Lemma).** The length of the shortest path connecting \( c \) disjoint \( \alpha \)-fat objects in \( \mathbb{R}^2 \) is at least \( (c/\alpha - 1)\pi S/4 \), where \( S \) is the size of the smallest object.

**Proof.** Consider a path \( T \) that connects the \( c \) objects and let the center of a disk with diameter \( S \) follow this path. At the point where the path touches a certain object, the disk intersects its boundary and hence at least an \( 1/\alpha \) fraction of the disk at that point intersects the object. The total area covered by the moving disk must be at least \( c/\alpha \) times the area \( \pi S^2/4 \) of the disk. On the other hand, it is easy to see that the total area covered by a disk that follows a continuous path \( T \) in \( \mathbb{R}^2 \) is at most \( \pi S^2/4 + S|T| \). Combining the upper and lower bound on the area we get,

\[
\frac{c \pi S^2}{(4 \alpha)} \leq \pi S^2/4 + S|T| \Rightarrow |T| \geq (c/\alpha - 1) \pi S/4.
\]

\( \square \)

### 3 Varying-sized Objects

In this section, we consider the case where the object \( O_1, \ldots, O_m \) are pairwise disjoint. Further, we assume that all objects have fatness at most \( \alpha \). However, we do not put any restriction on the size of the objects, i.e., we do not assume that objects have comparable sizes. The problem reduces to geometric TSP when each set \( S_i \) consists of a single point.

There exists a simple \((m-1)\)-approximation algorithm that we denote by **Greedy**. We define the **distance** between a point \( p \) and a set \( X \) as \( d(p,X) = \min_{x \in X} d(p,x) \).

**Algorithm Greedy:**

1. Pick the points \( p_i \in S_i \) \((i = 1 \ldots m)\) that minimize \( \sum_{j=2}^{m} d(p_1,p_j) \).
2. For all \( j \geq 2 \), select twice the edge \((p_1,p_j)\) and construct a tour by short cutting the edges.

**Lemma 2.** **Algorithm Greedy** gives an \((m-1)\)-approximate solution for Group TSP.

**Proof.** Any TSP-tour contains two edge disjoint paths from \( S_1 \) to \( S_i \) for all \( i \in \{2, \ldots, m\} \). Therefore, \((m - 1)|\text{Opt}| \geq 2 \sum_{i=2}^{m} d(p_1,p_i)\), which is at most the length of the tour constructed by the algorithm. \( \square \)
By \( \delta_i \) we denote the **diameter** of the point set \( S_i \), i.e. the largest distance between any two points in \( S_i \). Notice that \( \delta_i \) is at most the size of its enclosing object. The following is an immediate corollary of Lemma 1.

**Corollary 1.** The length of the shortest path connecting \( c \) of the given sets is at least \( (c/\alpha - 1)\pi \delta / 4 \), where \( \delta \) is the minimum of the diameters of the sets.

**Algorithm A:**

1. Order the point sets by their diameter \( \delta_1 \leq \delta_2 \leq \ldots \leq \delta_m \). Pick any \( p_1 \) in \( S_1 \).
2. For \( i = 2 \) up to \( m \) pick the point \( p_i \) in \( S_i \) that minimizes \( d(p_i, \{ p_1, \ldots, p_{i-1} \}) \), i.e. pick the point that is closest to the already chosen points.
3. Construct a \( (1 + \varepsilon) \)-approximate TSP tour \( T \) on this set of \( m \) points.
4. Output the minimum of \( T \) and the tour constructed by algorithm **Greedy**.

The second step can be done efficiently for any \( \varepsilon > 0 \) using techniques from [Aro98] and [Mit99].

**Theorem 1.** Algorithm A gives a \( (1 + \varepsilon)(9.1 \alpha + 1) \)-approximate solution for the group TSP with non-intersecting \( \alpha \)-fat neighborhoods.

**Proof.** We assume \( m - 1 > 9.1 \alpha + 1 \) since we can use **Greedy** for smaller values of \( m \). Denote the set of points chosen by A as \( F' = \{ p_1, \ldots, p_m \} \). Let \( p'_i \in \{ p_1, \ldots, p_{i-1} \} \) be the point at minimum distance from \( p_i \) and denote the distance by \( x_i \).

Consider some optimal solution \( OPT \) and fix an orientation of this tour. We choose some number \( c \in \{ 1, \ldots, m \} \) and define \( T_i \) as the part of this directed tour that connects exactly \( c \) sets and starts from the point in \( S_i \). Let \( t_i \) be the length of path \( T_i \).

We choose \( c = \lfloor \alpha(4/\pi + 1) \rfloor \). Notice that \( c \in \{ 1, \ldots, m \} \) is satisfied by the assumption in the first line of the proof. (By choosing \( c \) a bit smaller we can get a marginal improvement of the approximation ratio. Since this would make the proof more complicated we omit this here.) Consider some \( i \in \{ 1, \ldots, m \} \) and let \( S_{h(i)} \) be a set with smallest diameter among those from the \( c \) sets on the path \( T_i \). Then, by Corollary 1 and the choice of \( c \) we have

\[
t_i \geq (c/\alpha - 1)\pi \delta_{h(i)}/4 \geq \delta_{h(i)}.
\]

Since \( S_i \) is on this path \( T_i \) and we ordered the sets by their diameter we have \( 1 \leq h(i) \leq i \). We distinguish two cases.

If \( h(i) = i \), meaning that \( S_i \) has smallest diameter, then by (1) we have

\[
t_i \geq \delta_i.
\]

Otherwise, if \( h(i) < i \), then we argue as follows. Since the algorithm picked point \( p_i \) we know that the distance from any point in \( S_i \) to the point \( p_{h(i)} \) (which is chosen before \( p_i \)) is at least \( x_i \). Hence, the distance from any point in \( S_i \) to any
point in $S_{h(i)}$ is at least $x_i - \delta_{h(i)}$, implying $t_i \geq x_i - \delta_{h(i)}$. Together with (1) this yields

$$t_i \geq \max\{\delta_{h(i)}, x_i - \delta_{h(i)}\} \geq x_i/2. \quad (3)$$

We will construct a tour on the set of points $P' = \{p_1, \ldots, p_m\}$ chosen by the algorithm, using the bounds (2) and (3). Let $H$ be the set of indices $i$ for which $t_i \geq \delta_i$ and let $\text{OPT}_H$ be the smallest tour through the points $\{p_i | i \in H\}$. Clearly,

$$|\text{OPT}_H| \leq |\text{OPT}| + 2 \sum_{i \in H} \delta_i \leq |\text{OPT}| + 2 \sum_{i \in H} t_i.$$

Let $\tilde{H} = \{1, \ldots, m\} \setminus H$. Then, by (3) we know that for any $i \in \tilde{H}$, the length of the edge $(p_i, p'_i)$ equals $x_i \leq 2t_i$. We add this edge twice for any $i \in \tilde{H}$ to the tour $\text{OPT}_H$. Clearly, the resulting graph is Eulerian. Moreover, it is connected since $p_i \in H$, and for any $i$ we have $p'_i = p_j$ for some $j < i$. The total length of the Eulerian graph is

$$|\text{OPT}_{\tilde{H}}| + 2 \sum_{i \in \tilde{H}} d(p_i, p'_i) \leq |\text{OPT}| + 2 \sum_{i \in H} t_i + 2 \sum_{i \in \tilde{H}} 2t_i \leq |\text{OPT}| + 4 \sum_{i = 1}^m t_i.$$

When we take the sum over all $t_i$ then every edge is counted $c - 1$ times, implying $(c - 1)|\text{OPT}| = \sum_{i = 1}^m t_i$. Substituting the value of $c$ we conclude that the tour given by our algorithm has length at most

$$(1 + \epsilon)(1 + 4(c - 1))|\text{OPT}| < (1 + \epsilon)(9.093\alpha + 1)|\text{OPT}|.$$

The algorithm and proof apply directly to the Euclidean TSP with Neighborhoods problem under the weak assumption that, given the points $\{p_1, \ldots, p_{i-1}\}$, we can efficiently find the point $p_i$ in the infinite set of points $S_i$ that minimizes $d(p_i, \{p_1, \ldots, p_{i-1}\})$.

**Corollary 2.** Algorithm A gives a $(1 + \epsilon)(9.1\alpha + 1)$-approximate solution for the TSP with neighborhoods problem.

### 3.1 Higher dimensions

The generalization of the definitions and lemma’s of the previous section to higher dimensions is straightforward.

**Definition 2.** An object $O \subseteq \mathbb{R}^d$ is said to be $\alpha$-fat if for any $d$-dimensional sphere $D$ which does not fully contain $O$ and whose center lies in $O$, the volume of the intersection of $O$ and $D$ is at least $1/\alpha$ times the volume of $D$.

We denote the volume of a $d$-dimensional sphere with radius $r$ by $V_d(r)$.

**Lemma 3.** If the center of a $d$-dimensional sphere with radius $r$ follows a path $T$ in $\mathbb{R}^d$, then the volume covered by the sphere is at most $|T|V_{d-1}(r) + V_d(r)$.
Lemma 4. The length of the shortest path connecting $c$ disjoint $\alpha$-fat objects is at least $(c/\alpha - 1) V_d(r)/V_{d-1}(r)$, where $r$ is half the size of the smallest object.

The volume of a $d$-dimensional sphere with radius $r$ is

$$V_d(r) = \frac{\pi^{d/2} r^d}{\Gamma\left(\frac{d+1}{2}\right)}$$

where $\Gamma$ is the well-known gamma function. For $d \geq 3$ we get

$$\frac{V_d(r)}{V_{d-1}(r)} = r\sqrt{\frac{\pi}{\Gamma\left(\frac{(d+2)/2}{2}\right)}} > r\sqrt{\frac{\pi d}{2}}.$$

For small values of $K$ we can simply get a $K$-approximation as described in the previous section. If we choose $c$ such that $(c/\alpha - 1)\sqrt{\pi d/2} = 1$, then the proof of Theorem 1 applies here without any adjustment.

Theorem 2. Algorithm $\mathcal{A}$ is an $O(\alpha/\sqrt{d})$-approximation algorithm for the TSP in $\mathbb{R}^d$ with non-intersecting $\alpha$-fat neighborhoods.

Notice that the approximation factor decreases in the dimension for constant $\alpha$. However, for bounded objects, $\alpha$ grows exponentially in $d$. For example, $\alpha = 2^d$ for a $d$-dimensional sphere.

Safra and Schwartz [SS03] showed that TSP with neighborhoods in $\mathbb{R}^3$ is unlikely to be approximable within $O(\log^{1/2} n)$. Hence, there is little hope to improve our result for $d \geq 3$ to a ratio independent of the fatness $\alpha$.

4 Intersecting Objects

In this section, we consider the case when the objects defining the sets $S_1, \ldots, S_m$ are intersecting disks of the same radius $r$. We denote these disks by $D = \{D_1, \ldots, D_m\}$, and their centers by $c_i$. Then $S_i = P \cap D_i$, and assume that $P = S_1 \cup \ldots \cup S_m$.

A subset $P' \subseteq P$ is called a hitting pointset for $D$ if $P' \cap S_i \neq \emptyset$ for $i = 1, \ldots, m$ and a minimal hitting pointset if for every $x \in P'$ there exists an $i \in [m]$ such that $(P' \setminus \{x\}) \cap S_i = \emptyset$. A minimal hitting set can be found by the natural greedy algorithm: Set $P' = P$, and keep deleting points from $P'$ as long as it is still a hitting set. A (square) box $B$ is called a covering box for the set of disks $D$ if it contains a hitting pointset for $D$, and a minimum covering box if it has the smallest size amongst all such covering boxes. Since a minimum covering box is determined by two points of $P$ on its boundary, there are only $O(n^2)$ such candidates. By enumerating over all such boxes, and verifying if they contain a hitting set, one can compute a minimum covering box.

Consider the following algorithm for the Group TSP problem on the sets $S_1, \ldots, S_m$:

Algorithm $B$: 


(1) Compute a minimum covering box $B$ of $\mathcal{D}$.
(2) Find a minimal hitting pointset $P^* \subseteq P$ for $\mathcal{D}$ inside $B$.
(3) Compute a $(1 + \epsilon)$-approximate TSP tour on $P^*$.

The last step can be done efficiently for any $\epsilon > 0$ using techniques from [Arc98] and [Mit99]. To analyze the performance of the algorithm, we need the following lemma.

**Lemma 5.** Let $B$ be a box of diameter $L$ that contains $P = \{p_1, \ldots, p_n\}$. Let $\mathcal{D} = \{D_1, \ldots, D_n\}$ be a collection of disks of radius $r$, such that (i) each point $p \in P$ is contained in exactly one disk $D(p) \in \mathcal{D}$ (ii) each disk $D$ contains exactly one point of $P$. Then there exists a tour $T$ on $P$ with length at most $f(L/r)L$, where $f(L/r)$ is defined in (5). In particular, $f(6) \leq 113$.

**Proof.** Fix constants $s, k > 0$. Denote the center of disk $D_i$ by $c_i$. Set $s$ equally-spaced identical ‘direction’ points on each circle, say $a_1, \ldots, a_s$. Partition each disk $D_i$ into $s$ identical cones $A_1, \ldots, A_s$ where $A_j(D_i) = \{q \in D_i : \pi/s \leq \angle c_iq c_ja_j \leq \pi/s\}$. Denote by $\hat{a}_j$ the direction of the tangent to the circles at point $a_j$. See Figure 2(a).

Now partition $P$ into $s$ subsets $P_1, \ldots, P_s$, where $P_j$ contains all the points $p \in P$ that lie in the cone $A_j(D(p))$ of the disk containing $p$. We will first construct a path on all the points in each $P_j$ separately, and then connect these paths together.

Fix any set $P_j$. Partition $B$ into $k$ strips of equal width (each of width at most $L/k$), in the direction $\hat{a}_j$. W.l.o.g., assume $\hat{a}_j$ is horizontal (one can always rotate everything to get this). The situation is shown in Figure 2(b). Let $P_j^i$ be the set of points belonging to the $i$-th strip. Assume $P_j^i$ is sorted along the $x$-coordinates. Now construct a path $M$ by connecting all the points in $P_j^i$ in this linear sequence.

![Figure 2](image.png)

*Fig. 2. (a) Partitioning disks into cones. (b) Partitioning $B$ into $k$ strips in direction $\hat{a}_2$. The longest path is $M$ (bold) if each angle is at most $\alpha$.**
Claim. The path $M$ constructed above has length at most $L/\cos \alpha$, where

$$\alpha \leq \max \left\{ \sin^{-1} \frac{L}{r_k}, \frac{\pi}{2s} + \frac{\pi}{4} - \frac{1}{2} \sin^{-1} \left( \cos \frac{\pi}{s} - \frac{L}{r_k} \right) \right\} \quad (4)$$

Proof. Let $P_j^i = (r_1, \ldots, r_m)$. We first bound the angle that each $r_r r_{r+1}$ and $r_{r-1} r_r$ makes with the horizontal line passing through $r_r$. The case of $r_r r_{r-1}$ is symmetric, so we only consider the first case. One can assume that $r_{r+1}$ is below $r_r$ (otherwise, consider the angle made by $r_{r+1} r_r$ with line passing through $r_{r+1}$; both angles are the same and now $r_r$ lies below $r_{r+1}$).

We would like to place points $r_r$ and $r_{r+1}$ such that the angle of the edge $r_r r_{r+1}$ with the horizontal is maximized. By assumption, $r_r$ and $r_{r+1}$ are contained in two disks of radius $r$ containing only their respective points. It is not hard to see that there are two possible worst-case choices, as shown in Figure 3. The first case is when $r_r$ is the center of its enclosing disk $D(r_r)$, and $r_{r+1}$ lies (almost) on the boundary of $D(r_r)$, see Figure 3(a). Clearly, the required angle $\alpha \leq \sin^{-1} \frac{L}{r_k}$. The second case is shown in Figure 3(b). By elementary geometry, we get

$$a = r \cos \frac{\pi}{s}, \quad b = a - \frac{L}{r_k} \quad \sin c = \frac{b}{r}$$

from which one can derive that

$$\alpha \leq \frac{\pi}{2s} + \frac{\pi}{4} - \frac{1}{2} \sin^{-1} \frac{b}{r}.$$ 

Therefore, the maximum angle that any edge $r_r r_{r+1}$ makes with the horizontal is given by Equation 4. Finally, the weight of $M$ is maximized if each edge makes the maximum angle (at most $\alpha$) with the horizontal, as shown in Figure 2 (b). Then $M \leq L/\cos \alpha$, as required.

The above claim bounds the weight of the spanning path (constructed by simply sorting the points by $x$-coordinates and connecting them in this order) of points lying in cone $A_j$ and the $i$-th strip. There are at most $sk$ such paths,
and they can be connected together into a tour $T$ at the cost of additional total length of $sk \cdot L$. Therefore the constructed tour has a total length of at most $L \cdot f(L/r)$, where

$$f(\frac{L}{r}) = \min_{s,k} \{ sk(1 + \sec \alpha(s,k,L/r)) \},$$

and where $\alpha(s,k,L/r) = \max\{ \sin^{-1} \frac{L}{2r}, \frac{\pi}{2} - \frac{1}{2} \sin^{-1}(\cos \frac{\pi}{r} - \frac{L}{2r}) \}$. In particular, with $L/r = 6$, and setting $k = 6.5$ and $s = 4.5$, the tour has length at most $113L$.

\hfill $\square$

**Fig. 4.** Constructing an approximate tour as in the proof of Theorem 3.

**Theorem 3.** Algorithm B is a $O(1)$-approximation algorithm for the Group TSP problem.

**Proof.** Since $P^*$ is a hitting set for $D$, it is enough to show that there exists a tour $T$ on the set $P^*$ whose total cost is within $O(1)$ of the optimum for $D$. \footnote{The constants in the approximation factor can be improved by a more complicated analysis, which we omit from this extended abstract.}

To $P^*$ we can associate a subset of the disks $D' \subseteq D$ with the property that $|P^* \cap D| = 1$ for all $D \in D'$ and $|\{ D \in D' : x \in D \}| = 1$ for all $x \in P^*$. The set $D'$ can be found as follows. By the minimality of $P^*$, for every point $x \in P^*$ there exists a disk $D(x) \in D$ such that $D \cap P^* = \{x\}$. Let $D = \{ D(x) : x \in P^* \}$.

For a disk $D \in D'$, let us denote by $x(D)$ the (unique) point of $P^*$ contained inside $D$. Let $I \subseteq D'$ be a maximal independent set of disks in $D'$, i.e. a maximal
collection of pairwise disjoint disks. If every maximal independent set in $D$ has size at most $2$, we assume that $I$ consists of two disks in $D$, the distance $\delta$ between which is the largest. Let $OPT_I$ be an optimal Group TSP tour on $I$, when we are allowed to use any of the points lying inside the disks of $I$, and let $OPT^I$ be an optimal TSP tour on the set of points $\{x(I) : I \in I\}$. Clearly, $|OPT^I| \leq |OPT|$. We consider three cases:

Case 1: $|I| \geq 3$: By the maximality of $I$, every disk in $D \setminus I$ must intersect some disk in $I$. Let $\{D_I \subseteq D \setminus I : I \in I\}$ be a partition of $D \setminus I$, such that, for $I \in I$, $D_I$ contains only disks intersecting $I$. For $I \in I$, let $OPT_I$ be an optimal TSP tour on the pointset $S_I = \{x(D) : D \in D_I\} \cup \{x(I)\}$ contained in the partition of disks intersecting $I$. To define the tour $T$, take the union $T'$ of $OPT_I$ and $\bigcup_{I \in I} OPT_I$ to obtain an Eulerian connected graph $T'$ on the pointset $P'$. Finally, we use short-cutting on an Eulerian tour in $T'$ to get a TSP tour $T$ on $P'$ (see Figure 4).

Claim. $|T| \leq (1 + (4 + 6f(6))/0.14)|OPT|$, where $f(\cdot)$ is defined by (5).

Proof. Note that, for $I \in I$, all the points of $S_I$ lie inside a box of diameter at most $6r$. Thus applying Lemma 5, we conclude that there is a tour on $S_I$ of size at most $f(6) \cdot 6r$. In particular, $|OPT_I| \leq 6f(6) \cdot r$. On the other hand, by Lemma 1, and using $\alpha = 4$ for disks, one can derive that $|OPT_I| \geq 0.14r|I|$. Furthermore, by connecting by a double edge, for each disk $I \in I$, the point picked by the optimal solution $OPT_I$ to the point $x(I)$, we can construct a TSP tour on the set $\{x(I) : I \in I\}$ of total length at most $OPT_I + 4r|I|$. Thus it follows that $|OPT^I| \leq |OPT_I| + 4r|I|$. Combining these together, we get

$$|T| \leq |OPT^I| + \sum_{I \in I} |OPT_I| \leq |OPT_I| + (4 + 6f(6))r|I|$$

$$\leq (1 + \frac{4 + 6f(6)}{0.14})|OPT_I| \leq (1 + \frac{4 + 6f(6)}{0.14})|OPT|,$$

and our claim follows.

Case 2: $|I| = 2$ and $\delta > 2r$: Note that $\delta$ is a lower bound on $|OPT_I|$. Then $r|I| \leq |OPT_I|$ and, similar to Case 1, we can construct a tour $T$ of length at most $(1 + (4 + 6f(6))/0.14)|OPT|$

Case 3: $|I| = 2$ and $\delta \leq 2r$, or $|I| = 1$: Let $L$ be the size of the minimum covering box $B$. Note that $OPT \geq L$ because (i) any box containing $OPT$ is a covering box, and (ii) the smallest such box has width or height at most $|OPT|$. It is easy to see, in this case, that all the points of $P'$ lie inside a box of size at most $6r$. In particular, this implies that $L \leq 6r$. By Lemma 5, we can construct a tour $T$ on the pointset $P'$ whose total length does not exceed $f(6) \cdot L \leq f(6)|OPT|$. 

References


