

On the approximability of covering points by lines and related problems*

Adrian Dumitrescu[†]

Minghui Jiang[‡]

June 16, 2015

Abstract

Given a set P of n points in the plane, COVERING POINTS BY LINES is the problem of finding a minimum-cardinality set \mathcal{L} of lines such that every point $p \in P$ is incident to some line $\ell \in \mathcal{L}$. As a geometric variant of SET COVER, COVERING POINTS BY LINES is still NP-hard. Moreover, it has been proved to be APX-hard, and hence does not admit any polynomial-time approximation scheme unless $P = NP$. In contrast to the small constant approximation lower bound implied by APX-hardness, the current best approximation ratio for COVERING POINTS BY LINES is still $O(\log n)$, namely the ratio achieved by the greedy algorithm for SET COVER.

In this paper, we give a lower bound of $\Omega(\log n)$ on the approximation ratio of the greedy algorithm for COVERING POINTS BY LINES. We also study several related problems including MAXIMUM POINT COVERAGE BY LINES, MINIMUM-LINK COVERING TOUR, MINIMUM-LINK SPANNING TOUR, and MIN-MAX-TURN HAMILTONIAN TOUR. We show that all these problems are either APX-hard or at least NP-hard. In particular, our proof of APX-hardness of MIN-MAX-TURN HAMILTONIAN TOUR sheds light on the difficulty of BOUNDED-TURN-MINIMUM-LENGTH HAMILTONIAN TOUR, a problem proposed by Aggarwal et al. at SODA 1997.

Keywords: Geometric set cover and maximum coverage, Geometric tours, NP-hardness, APX-hardness.

1 Introduction

Given a set U of n elements and a family \mathcal{F} of m subsets of U , SET COVER is the problem of finding a minimum-cardinality subfamily $\mathcal{F}' \subseteq \mathcal{F}$ whose union is U . It is well-known that SET COVER can be approximated within $H_n \leq \ln n + 1$ [29, 34, 17], where H_n is the n th harmonic number, by a simple greedy algorithm that repeatedly selects a set that covers the most remaining elements; a more refined analysis [43] shows that the approximation ratio of the greedy algorithm is in fact $\ln n - \ln \ln n + \Theta(1)$. On the other hand, SET COVER cannot be approximated in polynomial time within $c \log n$ for some constant $c > 0$ unless $P = NP$ [42], and within $(1 - \varepsilon) \ln n$ for any $\varepsilon > 0$ unless $NP \subset \text{TIME}(n^{O(\log \log n)})$ [20]; see also [4, 35].

The first problem that we study in this paper is a geometric variant of SET COVER. Given a set P of n points in the plane, COVERING POINTS BY LINES is the problem of finding a minimum-cardinality set \mathcal{L} of lines such that every point $p \in P$ is in some line $\ell \in \mathcal{L}$. (Without loss of generality, we can assume that $n \geq 2$ and that the lines in \mathcal{L} are selected from the set \mathcal{L}_P of at most $\binom{n}{2}$ lines with at least two points of P in each line.)

*Supported in part by NSF grant DMS-1001667 awarded to the first author. The work of the second author was performed during his sabbatical leave in Fall 2012.

[†]Department of Computer Science, University of Wisconsin–Milwaukee, USA. Email: dimitres@uwm.edu.

[‡]Department of Computer Science, Utah State University, Logan, USA. Email: mjiang@cc.usu.edu.

As a restricted version of SET COVER, COVERING POINTS BY LINES may appear as a much easier problem. Indeed, in terms of parameterized complexity, SET COVER is clearly W[2]-hard when the parameter is the number k of sets in the solution (as easily seen by a reduction from the canonical W[2]-hard problem k -DOMINATING SET), while COVERING POINTS BY LINES admits very simple FPT algorithms based on standard techniques in parameterized complexity such as bounded search tree and kernelization [32]; see also [24, 45, 30]. In terms of approximability, however, the current best approximation ratio for COVERING POINTS BY LINES is still the same $O(\log n)$ upper bound for SET COVER achieved by the greedy algorithm. No matching lower bounds are known for COVERING POINTS BY LINES, although it has been proved to be NP-hard [36] and even APX-hard [12, 31].

Let X be a finite set and let \mathcal{R} be a system of subsets of an underlying set $U \supseteq X$. The pair (X, \mathcal{R}) is usually called a *range space* in a geometric setting. The *VC-dimension* of a range space (X, \mathcal{R}) is the maximum cardinality of a subset A of X , so that $|\{r \cap A : r \in \mathcal{R}\}| = 2^{|A|}$ [40, Ch. 15]. Observe that any range space (P, \mathcal{L}) of points P and lines \mathcal{L} in the plane as well as its dual range space have VC-dimension at most 2. Thus the well-known method for approximating geometric variants of HITTING SET and SET COVER based on ε -nets [25, 13, 19] yields an $O(\log \text{OPT})$ approximation ratio for COVERING POINTS BY LINES, where OPT denotes the size of the optimal solution. This approximation ratio is still $O(\log n)$ in the worst case. Although the method of ε -nets has been used successfully in obtaining better approximation ratios for geometric SET COVER when the sets are “fat” geometric objects [18], Alon’s superlinear lower bound on the minimum size of an ε -net for points and lines in the plane [3] rules out the possibility of a constant approximation for COVERING POINTS BY LINES by this method; see also [41].

The APX-hardness of COVERING POINTS BY LINES implies a constant lower bound on the approximation ratio, in particular, the problem does not admit any polynomial-time approximation scheme (PTAS), i.e., a $(1 + \epsilon)$ -approximation in $n^{O(1)}$ time for fixed ϵ , unless $P = NP$; moreover the problem does not admit any quasi-polynomial-time approximation scheme (QPTAS), i.e., a $(1 + \epsilon)$ -approximation in $2^{\text{poly} \log(n/\epsilon)}$ time, unless $NP \subseteq \text{DTIME}(2^{\text{poly} \log n})$. See [16, 38, 39] for other APX-hardness and PTAS/QPTAS results for related geometric set cover problems. One may observe that both the local search technique [38] and the separator technique [39] for the PTAS/QPTAS results require certain properties of the system of geometric sets, which lines in the plane do not have.

In this paper, we give an asymptotically tight lower bound on the approximation ratio of the greedy algorithm for COVERING POINTS BY LINES:

Theorem 1. *The approximation ratio of the greedy algorithm for COVERING POINTS BY LINES is $\Omega(\log n)$.*

A problem closely related to SET COVER is the following. Given a set U of n elements, a family \mathcal{F} of m subsets of U , and a number k , MAXIMUM COVERAGE is the problem of finding a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ of k subsets whose union has the maximum cardinality. In the setting of COVERING POINTS BY LINES, given a set P of n points in the plane and a number k , MAXIMUM POINT COVERAGE BY LINES is the problem of finding k lines that cover the maximum number of points in P . For the general MAXIMUM COVERAGE problem, the greedy algorithm that repeatedly selects a set that covers the most remaining elements achieves an approximation ratio of $1 - 1/e = 0.632\dots$ [26, Section 3.9]; this is also the current best approximation ratio for MAXIMUM POINT COVERAGE BY LINES. On the other hand, MAXIMUM COVERAGE cannot be approximated better than $1 - 1/e + \varepsilon$ for any $\varepsilon > 0$ unless $P = NP$ [20], while MAXIMUM POINT COVERAGE BY LINES is only known to be NP-hard as implied by the NP-hardness of COVERING POINTS BY LINES [36]. We show that MAXIMUM POINT COVERAGE BY LINES is APX-hard, complementing the APX-hardness of COVERING POINTS BY LINES¹:

¹We prove the APX-hardness of both problems by the same construction. These results were obtained unaware of the previous APX-hardness results of Brodén et al. [12] and Kumar et al. [31] for COVERING POINTS BY LINES. In retrospect, we note that the construction in our proof is the exact dual of the construction in [12]: we cover points by lines; they cover lines by points. In order to accommodate both problems, however, our core argument (Lemma 1) is more refined.

Theorem 2. MAXIMUM POINT COVERAGE BY LINES is APX-hard. This holds even if no four of the given points are collinear.

Theorem 3 (Brodén et al. [12] and Kumar et al. [31]). COVERING POINTS BY LINES is APX-hard [12, 31]. This holds even if no four of the given points are collinear [12].

Instead of using lines, we can cover the points using a polygonal chain of line segments. Given a set P of n points in the plane, a *covering tour* is a closed chain of segments that cover all n points in P , and a *spanning tour* is a covering tour in which the endpoints of all segments are points in P . The problem MINIMUM-LINK COVERING TOUR (respectively, MINIMUM-LINK SPANNING TOUR) aims at finding a covering tour (respectively, *spanning tour*) with the minimum number of links (i.e., segments). Stein and Wagner [44] presented an $O(\log n)$ approximation algorithm for MINIMUM-LINK COVERING TOUR. We note here that a similar method yields an $O(\log n)$ approximation algorithm for the MINIMUM-LINK SPANNING TOUR problem. From the other direction, Arkin et al. [6] proved that MINIMUM-LINK COVERING TOUR is NP-hard. The problem MINIMUM-LINK SPANNING TOUR has not been studied before in terms of its computational complexity, although it is a natural variant of MINIMUM-LINK COVERING TOUR. See also [5, 7, 28] for other related problems.

Strengthening the NP-hardness result of Arkin et al. [6], our next theorem shows that MINIMUM-LINK COVERING TOUR is in fact APX-hard:

Theorem 4. MINIMUM-LINK COVERING TOUR is APX-hard. This holds even if no four of the given points are collinear.

We also show that MINIMUM-LINK SPANNING TOUR is NP-hard:

Theorem 5. MINIMUM-LINK SPANNING TOUR is NP-hard. This holds even if no four of the given points are collinear.

Given a set P of n points in the plane, a *Hamiltonian tour* is a closed polygonal chain of *exactly* n segments whose n endpoints along the chain are a circular permutation of the n points in P . Note that every Hamiltonian tour is a spanning tour, but not vice versa, although every spanning tour can be transformed into a Hamiltonian tour by subdividing some segments into chains of shorter collinear segments.

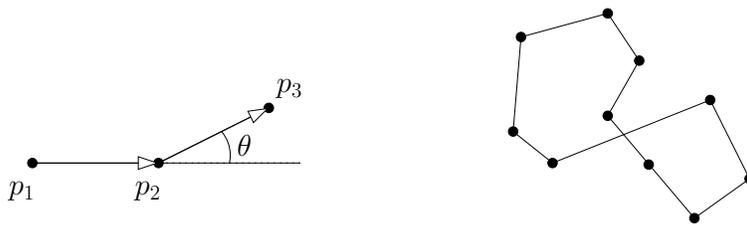


Figure 1: Left: turning angle at p_2 . Right: an obtuse tour of 10 points.

When three points p_1, p_2, p_3 are traversed in this order in a Hamiltonian tour, the *turning angle* at p_2 , denoted by $\text{turn}(p_1, p_2, p_3)$, is equal to $\pi - \angle p_1 p_2 p_3$, where $\angle p_1 p_2 p_3 \in [0, \pi]$; see Fig. 1. Note that the turning angle belongs to $[0, \pi]$, regardless of the direction of the turn (left or right). A tour or path with each turning angle in $[0, \pi/2]$ is called *obtuse*.

In the EUCLIDEAN TRAVELING SALESMAN PROBLEM (ETSP), given a set P of n points in the plane, one seeks a shortest Hamiltonian tour that visits each point. However, frequently other parameters are of interest, such as in motion planning, where small turning angles are desired. For example, an aircraft or a boat moving at high speed, required to pass through a set of given locations, cannot make sharp turns in its motion [1, 2, 10, 22, 27, 33]. A rough approximation of a smooth trajectory is provided by paths or tours

that are obtuse. However, not all point sets admit obtuse tours or even obtuse paths. For instance, some point sets require turning angles at least $5\pi/6$ in any Hamiltonian path [21]. Moreover, certain point sets (e.g., collinear) require the maximum turning angle possible, namely π , in any Hamiltonian tour.

Aggarwal et al. [1] have studied the following variant of angle-TSP, which we refer to as MIN-SUM-TURN HAMILTONIAN TOUR: Given n points in the plane, compute a Hamiltonian tour of the points that minimizes the total turning angle. The total turning angle of a tour is the sum of the turning angles at each of the n points. They proved that this problem is NP-hard and gave a polynomial-time algorithm with approximation ratio $O(\log n)$. They also suggested another natural variant of the basic angle-TSP problem, where the maximum turning angle at a vertex is bounded and the goal is to minimize the length measure.

Here we study the computational complexity of the following two variants of the angle-TSP problem. The first variant naturally presents itself, however it does not appear to have been previously studied. The second variant was proposed by Aggarwal et al. [1].

- (I) MIN-MAX-TURN HAMILTONIAN TOUR: Given n points in the plane, compute a Hamiltonian tour that minimizes the maximum turning angle.
- (II) BOUNDED-TURN-MINIMUM-LENGTH HAMILTONIAN TOUR: Given n points in the plane and an angle $\alpha \in [0, \pi]$, compute a Hamiltonian tour with each turning angle at most α , if it exists, that has the minimum length.

We have the following two results for the two variants of angle-TSP:

Theorem 6. MIN-MAX-TURN HAMILTONIAN TOUR is APX-hard.

Theorem 7. BOUNDED-TURN-MINIMUM-LENGTH HAMILTONIAN TOUR is NP-hard.

2 An $\Omega(\log n)$ lower bound on the approximation ratio of the greedy algorithm

In this section we prove Theorem 1. Our construction is inspired by a construction of Brimkov et al. [11] for the related problem COVERING SEGMENTS BY POINTS, which is in turn inspired by a classic lower bound construction for VERTEX COVER. This construction shows that there exist graphs with n vertices on which the greedy algorithm for VERTEX COVER achieves a ratio of $\Omega(\log n)$. The idea of the construction is perhaps folklore [37].

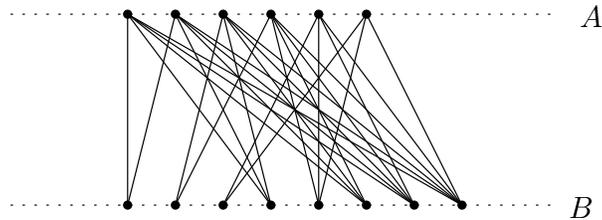


Figure 2: A lower bound construction for VERTEX COVER and COVERING SEGMENTS BY POINTS ($k = 6$ in this example).

Let $G_k = (A \cup B, E)$ be a bipartite graph where A is a set of k vertices, B is a set of $\sum_{i=2}^k \lfloor \frac{k}{i} \rfloor$ vertices partitioned into $k - 1$ subsets B_2, B_3, \dots, B_k , and E is a set of $\sum_{i=2}^k \lfloor \frac{k}{i} \rfloor \cdot i$ edges. For $2 \leq i \leq k$, each subset B_i contains $\lfloor \frac{k}{i} \rfloor$ vertices which are connected to $\lfloor \frac{k}{i} \rfloor \cdot i$ vertices in A , with each vertex in B_i connected to exactly i distinct vertices in A . Refer to Fig. 2 for an illustration of the graph G_k with $k = 6$.

Execute the greedy algorithm for VERTEX COVER on the bipartite graph G_k . In each step of the algorithm, after a vertex of the maximum degree is selected, the vertex and its incident edges are removed from

the graph. The crucial observation here is that before each selection, the degree of each vertex in A is at most the number of subsets B_i that are not empty, while the degree of each vertex in a non-empty subset B_i is exactly i . Thus the vertex of maximum degree selected in each step is always from a non-empty subset B_i with the maximum index i . A simple induction shows that the greedy algorithm always selects vertices from B_k, B_{k-1}, \dots, B_2 , in this order, and stops when all vertices in B are selected. On the other hand, the set of vertices in A clearly covers all edges too. Thus the approximation ratio of the greedy algorithm for Vertex Cover is at least

$$\begin{aligned} \frac{|B|}{|A|} &= \frac{\sum_{i=2}^k \lfloor \frac{k}{i} \rfloor}{k} = \frac{(\sum_{i=1}^k \lfloor \frac{k}{i} \rfloor) - k}{k} \geq \frac{(\sum_{i=1}^k \frac{k}{i}) - 2k}{k} \\ &= \sum_{i=1}^k \frac{1}{i} - 2 \geq \int_1^{k+1} \frac{1}{x} dx - 2 = \ln(k+1) - 2, \end{aligned}$$

which is $\Omega(\log n)$, where $n = \Theta(k \log k)$ is the number of vertices of G_k .

We now relate VERTEX COVER to a geometric problem, COVERING SEGMENTS BY POINTS: Given a set \mathcal{S} of n line segments in the plane, find a set P of points of minimum size such that each segment in \mathcal{S} contains at least one point in P . To adapt the construction for VERTEX COVER to COVERING SEGMENTS BY POINTS, Brimkov et al. [11] place the vertices in A and B in two parallel lines, with unit distance between consecutive vertices in each line, and with the vertices in each subset B_i placed consecutively, as illustrated in Fig. 2. Each edge in G_k corresponds to a line segment in \mathcal{S} with the two vertices as the endpoints. Without loss of generality, each point in P is either a vertex in A or B in one of the two parallel lines, or the intersection of two or more segments in \mathcal{S} between the two parallel lines. Observe that during the execution of the greedy algorithm, each intersection between (but in neither of) the two parallel lines is incident to at most one segment from the subset of segments incident to the vertices in B_i , $2 \leq i \leq k$; similar to the vertices in A , these intersections are never selected by the greedy algorithm. Thus the greedy algorithm still selects the vertices in B to cover the segments, and its approximation ratio is still $\Omega(\log n)$ by the same analysis.

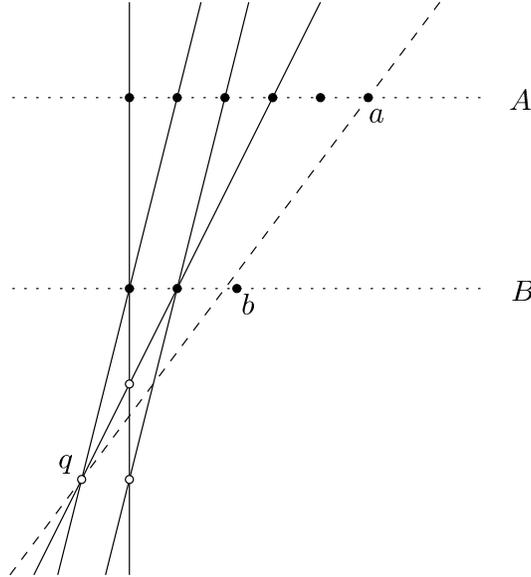


Figure 3: Adapting the lower bound construction for COVERING SEGMENTS BY POINTS to COVERING LINES BY POINTS.

We next adapt this construction further to the problem COVERING LINES BY POINTS: given a set \mathcal{L} of n lines in the plane, find a set P of points of the minimum cardinality such that each line in \mathcal{L} contains at

least one point in P . Since COVERING LINES BY POINTS and our original problem COVERING POINTS BY LINES are exact duals of each other, any lower bound we obtain for the former is also a lower bound for the latter.

Refer to Fig. 3. The straightforward part of the adaptation simply extends each segment in \mathcal{S} to a line in \mathcal{L} . This leads to more intersections, however, above and below the two parallel lines. As in the construction for COVERING SEGMENTS BY POINTS, we place the vertices in A evenly in the top line, with unit distance between consecutive points. For the vertices in B , however, we place them *almost* evenly in the bottom line, with *near-unit* distance between consecutive points (for convenience), such that the following property is satisfied:

P1: Any intersection of the lines in \mathcal{L} , if it is not a vertex in A or B in one of the two parallel lines, it is incident to exactly two lines in \mathcal{L} .

To ensure this property, we place the vertices in B incrementally as follows. Let B' be the subset of vertices in B that have been placed, and let \mathcal{L}' be the subset of lines in \mathcal{L} incident to B' . Let Q be the set of points that are intersections of the lines in \mathcal{L}' but are not vertices in A or B . For each point q in Q , and for each vertex a in A , mark the intersection of the bottom line and the line through q and a . Place the next vertex b in B in the bottom line to avoid such marks.

Due to the property P1, the greedy algorithm selects vertices in B_k, \dots, B_3 as before. Then, to cover the $2|B_2|$ lines incident to B_2 , it may select intersections not in the two parallel lines, but the number of points it selects is at least $2|B_2|/2 = |B_2|$ since these lines are in general position. Consequently, the same $\Omega(\log n)$ lower bound follows and this completes the proof of Theorem 1.

3 APX-hardness of MAXIMUM POINT COVERAGE BY LINES and COVERING POINTS BY LINES

In this section we prove Theorems 2 and 3. Given a set V of n variables and a set C of m clauses, where each variable has exactly p literals (in p different clauses) and each clause is the disjunction of exactly q literals (of q different variables), Ep -OCC-MAX- Eq -SAT is the problem of finding an assignment of the variables in V that satisfies the maximum number of clauses in C . Note that $pn = qm$. Berman and Karpinski [8] showed that even the simplest version of this problem, E3-OCC-MAX-E2-SAT, is APX-hard, and moreover this holds even if the three literals of each variable are neither all positive nor all negative; see also [9] for the current best approximation lower bounds for the many variants of Ep -OCC-MAX- Eq -SAT and related problems. We prove that both COVERING POINTS BY LINES and MAXIMUM POINT COVERAGE BY LINES are APX-hard by two gap-preserving reductions from E3-OCC-MAX-E2-SAT (Lemmas 5 and 6, respectively).

Let (V, C) be an instance of E3-OCC-MAX-E2-SAT, where V is a set of n variables v_i , $1 \leq i \leq n$, and C is a set of m clauses c_j , $1 \leq j \leq m$. We construct a set P of $4n + m$ points, including four variable points $v_i^1, v_i^2, v_i^3, v_i^4$ for each variable v_i , $1 \leq i \leq n$, and one clause point c_j^* for each clause c_j , $1 \leq j \leq m$. Assume that the three literals of each variable are neither all positive nor all negative. Then each variable has either two positive literals and one negative literal, or two negative literals and one positive literal. We place the point set P in the plane (an example appears in Fig. 4) such that no line goes through more than two points in P except in the following two cases:

1. If a variable v_i has two positive literals in c_r and c_s , respectively, and has one negative literal in c_t , then c_r^*, v_i^1, v_i^2 are collinear, c_s^*, v_i^3, v_i^4 are collinear, and c_t^*, v_i^1, v_i^3 are collinear.
2. If a variable v_i has two negative literals in c_r and c_s , respectively, and has one positive literal in c_t , then c_r^*, v_i^1, v_i^3 are collinear, c_s^*, v_i^2, v_i^4 are collinear, and c_t^*, v_i^1, v_i^2 are collinear.

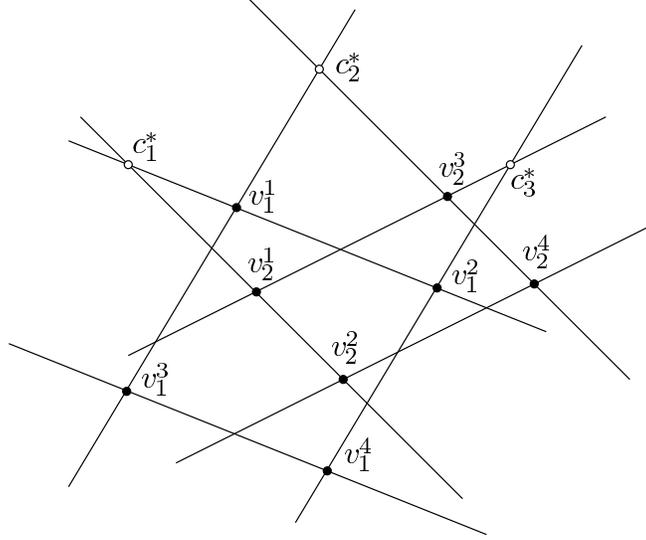


Figure 4: The set P of $4 \times 2 + 3 = 11$ points corresponding to the E3-OCC-MAX-E2-SAT instance (V, C) where $c_1 = v_1 \vee v_2$, $c_2 = \bar{v}_1 \vee v_2$, $c_3 = \bar{v}_1 \vee \bar{v}_2$.

For any set \mathcal{L} of lines, let $\mathcal{L}_i \subseteq \mathcal{L}$ denote the subset of lines incident to the four variable points $v_i^1, v_i^2, v_i^3, v_i^4$ of the variable v_i . For each variable v_i , and for each pair of indices $\{r, s\} \subset \{1, 2, 3, 4\}$, let $e_i^{r,s}$ denote the line through the two points v_i^r and v_i^s . We say that a set \mathcal{L} of lines is *canonical* if for each variable v_i , $|\mathcal{L}_i| \leq 2$ and $\mathcal{L}_i \subseteq \{e_i^{1,2}, e_i^{3,4}, e_i^{1,3}, e_i^{2,4}\}$, and moreover, if $|\mathcal{L}_i| = 2$, then \mathcal{L}_i is either $\{e_i^{1,2}, e_i^{3,4}\}$ or $\{e_i^{1,3}, e_i^{2,4}\}$. The following lemma is used by both reductions:

Lemma 1. *Any set \mathcal{L} of k lines that cover x points in P can be transformed into a canonical set of at most k lines that cover at least x (possibly different) points in P .*

Proof. Consider an arbitrary function $f : P \rightarrow \mathcal{L}$ that maps each point $p \in P$ to a line $\ell \in \mathcal{L}$ incident to p ; if p is not incident to any line in \mathcal{L} , then p is unmapped, i.e., $f(p)$ is undefined. For each variable v_i , let $\mathcal{L}_i^f \subseteq \mathcal{L}$ denote the subset of lines to which f maps the four variable points $v_i^1, v_i^2, v_i^3, v_i^4$. Clearly, $|\mathcal{L}_i^f| \leq 4$. Note that $\mathcal{L}_i^f \subseteq \mathcal{L}_i$ but in general \mathcal{L}_i^f is not necessarily the same as \mathcal{L}_i , because a variable point of v_i may be incident to multiple lines in \mathcal{L} and is mapped by f to at most one of them. In the following, we will transform \mathcal{L} and update f accordingly, until $\mathcal{L}_i^f = \mathcal{L}_i$ for all variables v_i and \mathcal{L} is canonical. Initially, every covered point is mapped to some line. During the transformation, we maintain the invariant that every mapped point is covered by some line (but not necessarily every covered point is mapped to some line) and the number of mapped points is non-decreasing.

Categorize each line $\ell \in \mathcal{L}$ in one of three types according to f : if there are two variable points of the same variable mapped to ℓ , then ℓ is of type 2; if there are no variable points mapped to ℓ , then ℓ is of type 0; otherwise, ℓ is of type 1. Note that each type-1 line either has only one variable point mapped to it, or has two variable points of different variables mapped to it.

In the first step, we transform \mathcal{L} until $|\mathcal{L}_i^f| \leq 2$ for each variable v_i . If $|\mathcal{L}_i^f| > 2$ for some variable v_i , then \mathcal{L}_i^f must include at least two lines of type 1. Note that each type-1 line in \mathcal{L}_i^f has exactly one variable point of v_i and at most one other point (either some clause point or a variable point of some other variable) mapped to it. As long as \mathcal{L}_i^f includes two type-1 lines, we replace them by a line of type 2 (through the two variable points of v_i previously mapped to the two type-1 lines) and at most one other line of type 0 or 1 (through the at most two other points, if any, previously mapped to the two type-1 lines), and then update the function f accordingly (so that the points previously mapped to the two type-1 lines are mapped to the lines that replace

them). This replacement reduces $|\mathcal{L}_i^f|$ by 1 (because the other line, if any, does not have any variable points of v_i mapped to it), and does not increase $|\mathcal{L}_j^f|$ for any $j \neq i$. Repeat such replacement whenever applicable. Eventually we have $|\mathcal{L}_i^f| \leq 2$ for every variable v_i .

In the second step, we transform \mathcal{L} until no lines of type 0 are incident to variable points. Consider any line ℓ of type 0. If ℓ is incident to two clause points, then by construction it is not incident to any variable point. Otherwise ℓ is incident to at most one clause point, and hence can be rotated, if necessary, to avoid all variable points. Note that the function f and the subsets \mathcal{L}_i^f are not changed during this step.

In the third step, we transform \mathcal{L} until it contains no lines of type 1. Consider any line ℓ of type 1 in \mathcal{L}_i^f , with a variable point of v_i , say v_i^r , mapped to it. Since $|\mathcal{L}_i^f| \leq 2$, there is at most one other line besides ℓ in \mathcal{L}_i^f , with at most two other variable points of v_i mapped to it. It follows that at least one of the four variable points of v_i , say v_i^s , is not mapped to any line in \mathcal{L} . Replace ℓ by the line $e_i^{r,s}$, and update the function f accordingly (first unmap the at most two points previously mapped to ℓ , including v_i^r , then map both v_i^r and v_i^s to the line $e_i^{r,s}$ of type 2).

In the fourth step, we transform \mathcal{L} by considering two cases for each line of type 2:

1. A line ℓ of type 2 is in \mathcal{L}_i^f with $|\mathcal{L}_i^f| = 1$. Then ℓ is the only line with variable points of v_i mapped to it, and $\ell \in \{e_i^{1,2}, e_i^{3,4}, e_i^{1,3}, e_i^{2,4}, e_i^{1,4}, e_i^{2,3}\}$. If $\ell \in \{e_i^{1,4}, e_i^{2,3}\}$, we replace ℓ by any line in $\{e_i^{1,2}, e_i^{3,4}, e_i^{1,3}, e_i^{2,4}\}$.
2. Two lines ℓ_1 and ℓ_2 of type 2 are in \mathcal{L}_i^f with $|\mathcal{L}_i^f| = 2$. Then $\{\ell_1, \ell_2\}$ must be either $\{e_i^{1,2}, e_i^{3,4}\}$, or $\{e_i^{1,3}, e_i^{2,4}\}$, or $\{e_i^{1,4}, e_i^{2,3}\}$. If the two lines are $\{e_i^{1,4}, e_i^{2,3}\}$, we replace them by either $\{e_i^{1,2}, e_i^{3,4}\}$ or $\{e_i^{1,3}, e_i^{2,4}\}$, arbitrarily.

After each replacement, we update f accordingly. This completes the transformation of \mathcal{L} into canonical form. \square

For the reduction to MAXIMUM POINT COVERAGE BY LINES, we have the following lemma about the construction:

Lemma 2. *There exists an assignment of the variables in V that satisfies at least w clauses in C if and only if there exists a set of $2n$ lines that cover at least $4n + w$ points in P .*

Proof. We first prove the direct implication. Let $g : V \rightarrow \{\text{true}, \text{false}\}$ be an assignment that satisfies at least w clauses in C . For each variable $v_i \in V$, $1 \leq i \leq n$, select two lines: if $g(v_i) = \text{true}$, select the line through v_i^1 and v_i^2 and the line through v_i^3 and v_i^4 ; if $g(v_i) = \text{false}$, select the line through v_i^1 and v_i^3 and the line through v_i^2 and v_i^4 . By construction, these $2n$ lines cover not only all $4n$ variable points, but also the at least w clause points for the satisfied clauses.

We next prove the reverse implication. Let \mathcal{L} be a set of $2n$ lines that cover at least $4n + w$ points in P . We will construct an assignment of the variables in V that satisfies at least w clauses in C . By Lemma 1, we can assume that \mathcal{L} is canonical. Consider any line $\ell \in \mathcal{L}$. If ℓ is not incident to any variable point, then it is incident to at most two clause points, and can be replaced by a line through two variable points of some variable while keeping \mathcal{L} in canonical form. Repeat such replacement whenever applicable. Eventually \mathcal{L} includes exactly $2n$ lines incident to all $4n$ variable points and at least w clause points. Compose an assignment $g : V \rightarrow \{\text{true}, \text{false}\}$ by setting $g(v_i)$ to true if $\mathcal{L}_i = \{e_i^{1,2}, e_i^{3,4}\}$ and to false if $\mathcal{L}_i = \{e_i^{1,3}, e_i^{2,4}\}$. Then by construction g satisfies at least w clauses. \square

For the reduction to COVERING POINTS BY LINES, we have the following two lemmas analogous to Lemma 2 about the construction:

Lemma 3. *If there exists an assignment of the variables in V that satisfies at least w clauses in C , then there exists a set of at most $2n + \lceil (m - w)/2 \rceil$ lines that cover all points in P .*

Proof. Let $g : V \rightarrow \{\text{true}, \text{false}\}$ be an assignment that satisfies at least w clauses in C . For each variable $v_i \in V$, $1 \leq i \leq n$, select two lines: if $g(v_i) = \text{true}$, select the line through v_i^1 and v_i^2 and the line through v_i^3 and v_i^4 ; if $g(v_i) = \text{false}$, select the line through v_i^1 and v_i^3 and the line through v_i^2 and v_i^4 . By construction, these $2n$ lines cover not only all $4n$ variable points, but also the at least w clause points for the satisfied clauses. To cover the remaining at most $m - w$ clause points for the unsatisfied clauses, we pair them up arbitrarily and use at most $\lceil (m - w)/2 \rceil$ additional lines. \square

Lemma 4. *If there exists a set of at most $2n + \lfloor (m - w)/2 \rfloor$ lines that cover all points in P , then there exists an assignment of the variables in V that satisfies at least w clauses in C .*

Proof. Let \mathcal{L} be a set of at most $2n + \lfloor (m - w)/2 \rfloor$ lines that cover all points in P . We will construct an assignment of the variables in V that satisfies at least w clauses in C . By Lemma 1, we can assume that \mathcal{L} is canonical. Since all points are covered, this requires that $|\mathcal{L}_i| = 2$ for each $i = 1, \dots, n$, thus \mathcal{L} includes exactly $2n$ lines incident to the variable points. These $2n$ lines must cover at least w clause points because the other at most $\lfloor (m - w)/2 \rfloor$ lines in \mathcal{L} can cover at most $m - w$ clause points. Compose an assignment $g : V \rightarrow \{\text{true}, \text{false}\}$ by setting $g(v_i)$ to true if $\mathcal{L}_i = \{e_i^{1,2}, e_i^{3,4}\}$ and to false if $\mathcal{L}_i = \{e_i^{1,3}, e_i^{2,4}\}$. Then by construction g satisfies at least w clauses. \square

The following lemma implies that MAXIMUM POINT COVERAGE BY LINES is APX-hard:

Lemma 5. *For any ε , $0 < \varepsilon < \frac{1}{5}$, if MAXIMUM POINT COVERAGE BY LINES admits a polynomial-time approximation algorithm with ratio $1 - \varepsilon$, then E3-OCC-MAX-E2-SAT admits a polynomial-time approximation algorithm with ratio $1 - 5\varepsilon$.*

Proof. Let (V, C) be an instance of E3-OCC-MAX-E2-SAT with n variables and m clauses, where $3n = 2m$. Consider the following algorithm: first construct a set P of points from (V, C) (refer back to Fig. 4) and set k (the number of lines) to $2n$; then run the $(1 - \varepsilon)$ -approximation algorithm for MAXIMUM POINT COVERAGE BY LINES on the instance (P, k) to obtain a set \mathcal{L} of lines, and finally compose an assignment $g : V \rightarrow \{\text{true}, \text{false}\}$ as in the reverse implication of Lemma 2. The algorithm can clearly be implemented in polynomial time. It remains to analyze its approximation ratio.

Let x^* be the maximum number of points in P that can be covered by any set of $k = 2n$ lines. Clearly, $x^* \leq 4n + m$. Let w^* be the maximum number of clauses in C that can be satisfied by any assignment of V . Observe that $w^* \geq \frac{3}{4}m$ since a random assignment of each variable independently to either true or false with equal probability $\frac{1}{2}$ satisfies each disjunctive clause of two literals with probability $1 - (\frac{1}{2})^2 = \frac{3}{4}$. Recall that $3n = 2m$. Thus we have

$$x^* \leq 4n + m = \frac{11}{3}m \leq \frac{11}{3} \frac{4}{3}w^* = \frac{44}{9}w^* < 5w^*. \quad (1)$$

Let x' be the number of points in P covered by the lines in \mathcal{L} . Let w' be the number of clauses in C satisfied by the assignment g . Lemma 2 implies that $w^* = x^* - 4n$ and by the reverse implication in Lemma 2 we have $w' \geq x' - 4n$. Thus $w^* - w' \leq x^* - x'$. It then follows from (1) that

$$\frac{w^* - w'}{w^*} \leq 5 \frac{x^* - x'}{x^*}.$$

The $(1 - \varepsilon)$ -approximation algorithm for MAXIMUM POINT COVERAGE BY LINES guarantees the relative error bound $(x^* - x')/x^* \leq \varepsilon$. So we have $(w^* - w')/w^* \leq 5\varepsilon$ and hence $w'/w^* \geq 1 - 5\varepsilon$, as desired. \square

The following lemma, analogous to Lemma 5, implies that COVERING POINTS BY LINES is APX-hard:

Lemma 6. *For any ε , $0 < \varepsilon < \frac{1}{10}$, if COVERING POINTS BY LINES admits a polynomial-time approximation algorithm with ratio $1 + \varepsilon$, then E3-OCC-MAX-E2-SAT admits a polynomial-time approximation algorithm with ratio $1 - 10\varepsilon$.*

Proof. Let (V, C) be an instance of E3-OCC-MAX-E2-SAT with n variables and m clauses, where $3n = 2m$. Let w^* be the maximum number of clauses in C that can be satisfied by any assignment of V . We have $w^* \geq \frac{3}{4}m$ since a random assignment of each variable independently to either true or false with equal probability $\frac{1}{2}$ satisfies each disjunctive clause of two literals with probability $1 - (\frac{1}{2})^2 = \frac{3}{4}$. Without loss of generality, we assume that $w^* \geq 1/\varepsilon$, since otherwise the instance (V, C) would have size $O(1/\varepsilon)$ and would admit a straightforward brute-force algorithm running in $2^{O(1/\varepsilon)}$ time, which is constant time for any fixed $\varepsilon > 0$.

Under the assumption that $w^* \geq 1/\varepsilon$, we have the following algorithm for E3-OCC-MAX-E2-SAT: first construct a set P of points from (V, C) (refer back to Fig. 4), then run the $(1 + \varepsilon)$ -approximation algorithm for COVERING POINTS BY LINES on P to obtain a set \mathcal{L} of lines, and finally compose an assignment $g : V \rightarrow \{\text{true}, \text{false}\}$ as in the proof of Lemma 4. The algorithm can clearly be implemented in polynomial time. It remains to analyze its approximation ratio.

Let k^* be the minimum cardinality of any set of lines that cover P . It is easy to see that $k^* \leq 2n + m$. Recall that $w^* \geq \frac{3}{4}m$ and $3n = 2m$. Thus we have

$$k^* \leq 2n + m \leq 3m \leq 4w^*. \quad (2)$$

Lemma 3 implies that $k^* \leq 2n + \lceil (m - w^*)/2 \rceil$. It follows that $k^* \leq 2n + (m - w^* + 1)/2$ and hence

$$w^* \leq 2(2n - k^*) + m + 1. \quad (3)$$

Let k' be the number of lines in \mathcal{L} . Let w' be the number of clauses in C that are satisfied by the assignment g . Put $w := m - 1 - 2(k' - 2n)$. Then

$$k' = 2n + (m - 1 - w)/2 = 2n + \lfloor (m - w)/2 \rfloor. \quad (4)$$

Note that $w := m - 1 - 2(k' - 2n)$ is the smallest integer (there are two such integers) satisfying the equation $k' = 2n + \lfloor (m - w)/2 \rfloor$. If $w' < w$, then by (the contrapositive of) Lemma 4 we would have $k' > 2n + \lfloor (m - w)/2 \rfloor$, which contradicts (4). So we must have $w' \geq w$, that is,

$$w' \geq m - 1 - 2(k' - 2n). \quad (5)$$

From (3) and (5), we have

$$w^* - w' \leq 2(2n - k^*) + m + 1 - m + 1 + 2(k' - 2n) = 2(k' - k^*) + 2,$$

and hence

$$\frac{w^* - w'}{w^*} \leq \frac{2(k' - k^*)}{w^*} + \frac{2}{w^*} \leq \frac{8(k' - k^*)}{k^*} + \frac{2}{w^*},$$

where the second inequality follows from (2). The $(1 + \varepsilon)$ -approximation algorithm for COVERING POINTS BY LINES guarantees the relative error bound $(k' - k^*)/k^* \leq \varepsilon$. Recall our assumption that $w^* \geq 1/\varepsilon$. Consequently we have $(w^* - w')/w^* \leq 8\varepsilon + 2\varepsilon = 10\varepsilon$ and hence $w'/w^* \geq 1 - 10\varepsilon$, as desired. \square

Remark. For simplicity, we did not attempt obtaining the best multiplicative constant factors of ε in the previous two lemmas. Those expressions can be improved.

4 APX-hardness of MINIMUM-LINK COVERING TOUR

In this section we prove Theorem 4. We show that MINIMUM-LINK COVERING TOUR is APX-hard by a gap-preserving reduction from COVERING POINTS BY LINES², which was proved to be APX-hard in Theorem 3.

Let P be a set of n points for the problem COVERING POINTS BY LINES. We will construct a set Q of $3n$ points for the problem MINIMUM-LINK COVERING TOUR, such that P can be covered by k lines if and only if Q admits a covering tour with $3k$ segments.

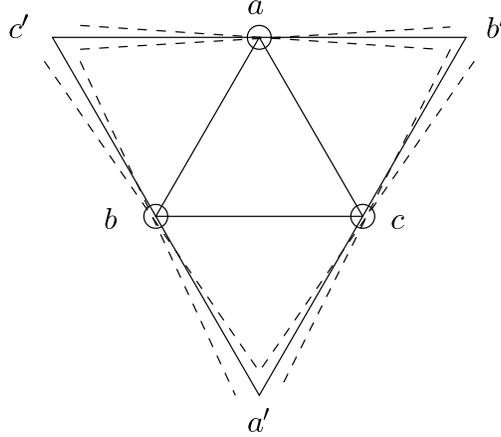


Figure 5: The construction of Q from P .

Refer to Fig. 5. By an affine transformation, we first transform P into a set P' of n points such that (i) P' is enclosed in a circle of some small radius r , say, $r = 1/100$; (ii) the angle between any two lines ℓ_1 and ℓ_2 , each incident to at least two points in P' , is at most some small angle θ , say, 1° . Now take an equilateral triangle abc of side length 1 inscribed in an equilateral triangle $a'b'c'$ of side length 2, where the three vertices of the smaller triangle are the midpoints of the three edges of the larger triangle. The point set Q is the union of three rotated copies of P' that we refer to as the three clusters, one cluster near each vertex of abc , such that the circle of radius r enclosing each cluster is centered at the vertex, and all lines through at least two points in the cluster are at angles at most θ from the edge of $a'b'c'$ that contains the vertex.

Lemma 7. *There exists a set of k lines that cover all points in P if and only if there exists a covering tour with $3k$ segments for Q .*

Proof. We first prove the direct implication. Let \mathcal{L} be a set of k lines that cover all points in P . Then by the affine transformation, we have a set \mathcal{L}' of k lines that cover all points in P' , and the $3k$ lines in the three copies of \mathcal{L}' corresponding to the three copies of P' cover all points in Q . These $3k$ lines can obviously be linked into a covering tour with $3k$ segments, where any three consecutive segments are from three different clusters, and the turns between consecutive segments are near the vertices of $a'b'c'$.

We next prove the reverse implication. Let \mathcal{T} be a covering tour with $3k$ segments for Q , and let $\mathcal{L}_{\mathcal{T}}$ be the set of at most $3k$ lines supporting the $3k$ segments in \mathcal{T} . A line is an *intra-cluster* line if the points in Q that are covered by it, if any, are all from the same cluster; it is an *inter-cluster* line otherwise. By construction, each inter-cluster line covers exactly two points, from two different clusters. Let l_1 (respectively, l_2) be the number of intra-cluster (respectively, inter-cluster) lines in $\mathcal{L}_{\mathcal{T}}$; then $l_1 + l_2 \leq 3k$. Let n_a, n_b, n_c be the numbers of

²Arkin et al. [6] proved the NP-hardness of MINIMUM-LINK COVERING TOUR by a reduction from the same problem COVERING POINTS BY LINES, but since their reduction is not gap-preserving, their proof does not immediately imply the APX-hardness of MINIMUM-LINK COVERING TOUR even if COVERING POINTS BY LINES was known to be APX-hard. It is quite likely, however, that their construction can be combined with our construction in the proof of Theorem 3 to obtain a gap-preserving reduction directly from E3-OCC-MAX-E2-SAT to MINIMUM-LINK COVERING TOUR.

points in clusters near a, b, c , respectively, that are covered by the inter-cluster lines; then $n_a + n_b + n_c = 2l_2$. Since any two points in the same cluster can be covered by some intra-cluster line, these $n_a + n_b + n_c$ points can be covered by at most $\lceil n_a/2 \rceil + \lceil n_b/2 \rceil + \lceil n_c/2 \rceil$ intra-cluster lines (instead of l_2 inter-cluster lines). Since the sum of the three numbers n_a, n_b, n_c is even, we have either two of them odd and one even, or all three of them even. Thus we have $\lceil n_a/2 \rceil + \lceil n_b/2 \rceil + \lceil n_c/2 \rceil \leq (n_a + n_b + n_c)/2 + 1 = l_2 + 1$. It follows that Q can be covered by at most $l_1 + (l_2 + 1) \leq 3k + 1$ intra-cluster lines, and hence at least one of the three copies of P' can be covered by at most $\lfloor (3k + 1)/3 \rfloor = k$ lines. The corresponding k lines obtained by reversing the affine transformation cover all points in P . \square

From the above lemma, we can easily prove the following lemma similar to Lemmas 5 and 6, which implies the APX-hardness of MINIMUM-LINK COVERING TOUR:

Lemma 8. *For any $\rho \geq 1$, if MINIMUM-LINK COVERING TOUR admits a polynomial-time approximation algorithm with ratio ρ , then COVERING POINTS BY LINES admits a polynomial-time approximation algorithm with ratio ρ .*

5 NP-hardness of MINIMUM-LINK SPANNING TOUR

In this section we prove Theorem 5. We show that MINIMUM-LINK SPANNING TOUR is NP-hard by a reduction from a variant of the NP-hard problem HAMILTONIAN CIRCUIT IN CUBIC GRAPHS [23], in which the input consists of not only a cubic graph G but also some edge e of G that is required to be part of the Hamiltonian circuit. A simple Turing reduction shows that this variant is still NP-hard.

Let G be a cubic graph with n vertices and m edges, where $3n = 2m$. Let $e = \{s, t\}$ be the edge of G that is required to be part of the circuit. We first obtain a graph G' from G , by removing the edge e , and then adding two dummy vertices s' and t' with two new edges $\{s, s'\}$ and $\{t, t'\}$. Then there exists a Hamiltonian circuit in G containing the edge e if and only if there exists a Hamiltonian path in G' from s' to t' . Observe that G' has exactly $n + 2$ vertices and exactly $m + 1$ edges, where $n + 2 < m + 1$, and moreover every vertex except s' and t' has degree 3.

We next construct a set P of $n + 2 + m + 1$ points, one vertex point for each vertex, and one edge point for each edge in G' . The $n + 2$ vertex points are in some arbitrary convex position, say, on a circle. The $m + 1$ edge points are in the interior of the convex hull of the $n + 2$ vertex points. Moreover, for each edge $e = \{u, v\}$, the edge point of e is in the line through the two vertex points of u and v (i.e., in the interior of the segment uv), and is not in any other line containing more than two points in P .

The reduction can clearly be implemented in polynomial time. Then the following lemma establishes the NP-hardness of MINIMUM-LINK SPANNING TOUR:

Lemma 9. *There exists a Hamiltonian path in G' from s' to t' if and only if there exists a spanning tour with $m + 2$ segments for P .*

Proof. We first prove the direct implication. Let H be a Hamiltonian path in G' from s' to t' . We will construct a spanning tour with $m + 2$ segments for P . Corresponding to the Hamiltonian path H that visits all $n + 2$ vertices in G' using $n + 1$ edges, there is a polygonal chain of $n + 1$ segments that connect the $n + 2$ vertex points in P in the same order, which also cover $n + 1$ edge points. The chain can be extended to visit the remaining $(m + 1) - (n + 1) = m - n$ edge points in any order with $m - n$ segments, and finally closed into a tour with another segment. The total number of segments is $(n + 1) + (m - n) + 1 = m + 2$.

We next prove the reverse implication. Let \mathcal{T} be a spanning tour with $m + 2$ segments for P . We will find a Hamiltonian path in G' from s' to t' . We first transform \mathcal{T} , without increasing the number of segments, into a canonical spanning tour that visits each vertex point exactly once.

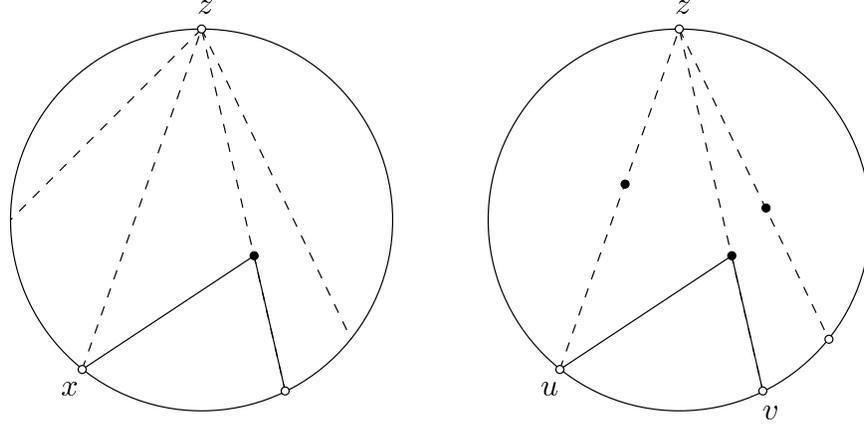


Figure 6: Transforming \mathcal{T} into a canonical spanning tour that visits each vertex point exactly once.

Suppose that a vertex point z is visited twice in \mathcal{T} . Then there are two pairs of consecutive segments with turns at z . Refer to Fig. 6. If, out of these four segments incident to z , there is a segment connecting some point x directly to z with no other point of P in the interior of the segment, then we can take a shortcut (as in Fig. 6 left) in the pair of segments including xz to skip a visit to z . Otherwise, each of the four segments must connect z to the vertex point of a neighbor of the vertex of z in G' , going through the corresponding edge point. Recall that every vertex in G' has degree at most 3. By the pigeonhole principle, at least two of these four segments must be the same segment, say, uz . If the two copies of uz are consecutive and form a turn at z in \mathcal{T} , then we can shorten both of them to skip a visit to z . Otherwise, one copy of uz must form a turn at z with some other segment, say vz , and again we can take a shortcut (as in Fig. 6 right) to skip a visit of z . Observe that in both cases, \mathcal{T} remains a spanning tour after the transformation.

Now observe that every edge point is either in the interior of some segment between two vertex points, or at a turning point between two consecutive segments. Associate a cost of 1 with each edge point. For each edge point, charge its cost to the segments that contain it: if it is in the interior of one segment, charge 1 to the segment; if it is at a turning point between two consecutive segments, charge $1/2$ to each segment. Observe that

- each segment between two vertex points is charged 1 if the segment contains an edge point, and is charged 0 otherwise;
- each segment between two edge points is charged $1/2 + 1/2 = 1$;
- each segment between a vertex point and an edge point (there must be at least two such segments along the tour since there are more edge points than vertex points) is charged exactly $1/2$.

Since the number of edge points in P is $m + 1$ and the number of segments in \mathcal{T} is $m + 2$, we must have exactly 2 segments charged $1/2$ each, and exactly m segments charged 1 each, so that $m + 1 = 2 \cdot 1/2 + m \cdot 1$. It follows that (i) there are exactly two segments between vertex points and edge points, and (ii) there is no segment connecting two vertex points and containing no edge point. Condition (i) implies that the segments between vertex points are consecutive in the tour. Condition (ii) implies that these consecutive segments correspond to a Hamiltonian path in G' . Finally, this Hamiltonian path must have s' and t' as the two ends because each of them has exactly one neighbor. \square

6 APX-hardness of MIN-MAX-TURN HAMILTONIAN TOUR and NP-hardness of BOUNDED-TURN-MINIMUM-LENGTH HAMILTONIAN TOUR

In this section we prove Theorems 6 and 7. We first show that MIN-MAX-TURN HAMILTONIAN TOUR is APX-hard by a gap-preserving reduction from COVERING POINTS BY LINES, which was proved to be APX-hard in Theorem 3. Let P be a set of n points for the problem COVERING POINTS BY LINES. We will construct a set Q of points for the problem MIN-MAX-TURN HAMILTONIAN TOUR.

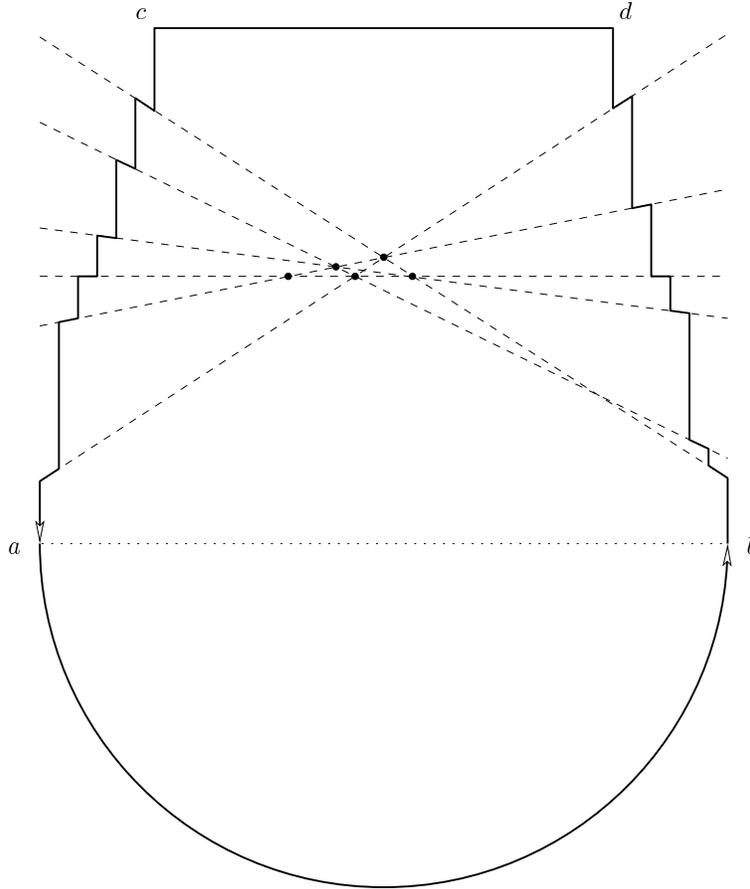


Figure 7: The construction of the point set Q for MIN-MAX-TURN HAMILTONIAN TOUR ($s = 4 \times 6 + 3 = 27$ in this example).

Let \mathcal{L}_P be the set of at most $\binom{n}{2}$ lines determined by P , where each line in \mathcal{L}_P goes through at least two points in P (we assume without loss of generality that no line in \mathcal{L}_P is vertical). Let α_P be the minimum turning angle determined by any three non-collinear points in P ; we will show later in Lemma 12 that $1/\alpha_P$ may be assumed to be polynomial in n . Let $\alpha := \alpha_P/n^\mu$ for some suitable constant $\mu \geq 1$.

Refer to Fig. 7. We first construct a simple closed curve γ that is composed by a half-circle and a polygonal chain joined at their endpoints a and b . The polygonal chain consists of $s := 2|\mathcal{L}_P| + 2(|\mathcal{L}_P| + 1) + 1 = O(n^2)$ segments, including 2 segments from each line in \mathcal{L}_P , $2(|\mathcal{L}_P| + 1)$ vertical segments, and 1 horizontal segment cd . Observe that γ is *monotone* in the horizontal direction, in the sense that the intersection of every vertical line with the region enclosed by γ is a single line segment.

Refer to Fig. 8. We next transform γ into a simple closed curve $\tilde{\gamma}$ that is not only monotone but also *smooth*, by smoothing the corner between every pair of consecutive segments in the chain into a small circular arc tangent to both segments. Then $\tilde{\gamma}$ is an alternating cycle of s segments (including the horizontal segment

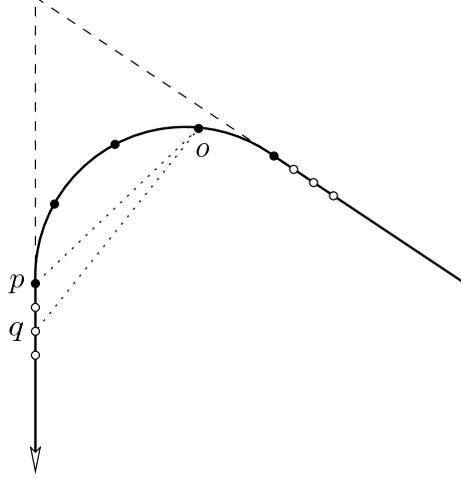


Figure 8: Smoothing the corner between two consecutive segments into a circular arc tangent to both segments ($t = 4$ and $n = 3$ in this example).

cd) and s circular arcs (including the half-circle with diameter ab).

Put $t := \lceil 9n^2\pi/\alpha \rceil$. The point set Q consists of all n points in P and $m := s(t + 1 + 2n)$ points from the curve $\tilde{\gamma}$. Take $t + 1$ points (including the two endpoints) from each of the s circular arcs, which divide any such arc into t sub-arcs with the same central angle at most $\theta := \pi/t \leq \alpha/(9n^2)$. Take n more points from each of the two segments adjacent to the circular arc, near the shared endpoints, such that the following θ -property is satisfied: the angle $\angle poq$, where p is an endpoint of the arc, o is any of the other t points in the arc, and q is any of the n points in the segment containing p , is at most θ . Observe that the total number of points is $s(t + 1 + 2n)$.

We have the following two lemmas about the construction:

Lemma 10. *If there exists a set of k lines that cover all points in P , for some $k \leq n$, then there exists a Hamiltonian tour with maximum turning angle at most $(k + 2)\theta$ for Q .*

Proof. Let $\mathcal{L} = \{\ell_1, \dots, \ell_k\}$ be a set of k lines that cover all points in P ; without loss of generality, $\mathcal{L} \subseteq \mathcal{L}_P$. We will construct a Hamiltonian tour with maximum turning angle at most $(k + 2)\theta$ for Q . Index the m points in Q taken from the curve $\tilde{\gamma}$ by the circular order of their locations along the curve: q_1, \dots, q_m . Assign each point a color in $[0, k]$: q_i has color $i \bmod (k + 1)$, $1 \leq i \leq m$.

The tour consists of $k + 1$ rounds. In the i th round, $1 \leq i \leq k$, the tour follows the half-circle from a to b and continues along the chain from b to d until it reaches the line ℓ_i , then takes a shortcut along ℓ_i from right to left and continues along the chain from c to a until it reaches the half-circle again; the tour visits each point of color i in or below the line ℓ_i while following the curve, and visits each point in P that is covered by the line ℓ_i (if the point was not visited in previous rounds) while taking the shortcut. In the last round, the tour follows the curve entirely to visit each point of color 0, and points of other colors not visited in previous rounds due to the shortcuts.

Consider any three consecutive points p_1, p_2, p_3 in the tour. If the three points are all in some line ℓ_i supporting a shortcut, then obviously $\text{turn}(p_1, p_2, p_3) = 0$. Otherwise, the three points must come from some sub-curve of $\tilde{\gamma}$ consisting of a circular arc and an adjacent $\tilde{\gamma}$ segment. The number of sub-arcs of this arc that are between p_1 and p_3 is at most $2(k + 1)$; each such sub-arc contributes half of its central angle to the turning angle at p_2 . Taking into account the possibility that the three points are not all in the arc and using the θ -property, we have $\text{turn}(p_1, p_2, p_3) \leq 2(k + 1)\theta/2 + \theta = (k + 2)\theta$. \square

Lemma 11. *If there exists a Hamiltonian tour with maximum turning angle at most $k\theta$ for Q , for some $k \leq n$, then there exists a set of k lines that cover all points in P .*

Proof. Let τ be a Hamiltonian tour with maximum turning angle at most $k\theta$ for Q . We will find a set of k lines that cover all points in P .

Break the tour τ into rounds, such that each round consists of some points in the half-circle followed by some points not in the half-circle (i.e., in the chain or in P). When the angle α is sufficiently small (hence $k\theta \leq k\alpha/(9n^2)$ is even smaller), in each round the tour must visit some points in the half-circle in order of their x -coordinates, say, from left to right, then visit points near some corners of the chain, from right to left. While in the chain, the tour may take shortcuts between non-consecutive corners, but since the curve $\tilde{\gamma}$ is monotone, it can take at most one *crossing* shortcut from a corner on the bd side to a corner on the ac side. Only when taking such a crossing shortcut can the tour visit some points in P . Moreover, since $k\theta \leq k\alpha/(9n^2) < \alpha_P$, the points in P that are visited during each crossing shortcut must be collinear.

We next show that τ has at most k rounds. Index the $t + 1$ points in the half-circle from left to right by numbers from 0 to t , where a has index 0 and b has index t . Consider an arbitrary round. Let r be the number of points in the half-circle that are visited in this round. Let i_1, \dots, i_r be the indices of these points, from left to right. Then we must have $i_1 \leq 2k - 1$ because otherwise the turning angle at the point with index i_1 would be greater than $(i_1 + 1)\theta/2 > 2k\theta/2 = k\theta$. Similarly, we must have $i_r \geq t - (2k - 1)$, and $i_{j+2} - i_j \leq 2k$ for each j , $1 \leq j \leq r - 2$. Counting in pairs, we have $r \geq 2r_2$, where $r_2 = \lfloor \frac{t - (2k - 1) - (2k - 1)}{2k} \rfloor$. It follows that the number of rounds in τ is at most

$$\frac{t}{r} \leq \frac{t}{2 \lfloor \frac{t - (2k - 1) - (2k - 1)}{2k} \rfloor} \leq \frac{t}{2 \left(\frac{t - (2k - 1) - (2k - 1)}{2k} - 1 \right)} = \frac{t}{t - 6k + 2} k.$$

It is easy to check that $t \geq 9n^2 > 8k^2$ and hence $\frac{t}{t - 6k + 2} < \frac{k + 1}{k}$. Thus τ has at most k rounds. Finally, since each round has at most one crossing shortcut that can cover some collinear points in P , all points in P can be covered by k lines. \square

Unlike the reductions in previous sections, our reduction to MIN-MAX-TURN HAMILTONIAN TOUR is numerically sensitive because the construction depends on a small angle α_P and has points placed precisely on circular arcs. Even if we reduce from a restricted version of COVERING POINTS BY LINES, where the coordinates are polynomial in the number of points (it can be checked that our proof for the APX-hardness of COVERING POINTS BY LINES fulfills this restriction), it is still not immediately clear that the reduction is polynomial. To clarify this, we prove a property concerning lattice points in the next lemma, which implies that $1/\alpha_P$ may be assumed to be polynomial in the lattice size N :

Lemma 12. *Let a, b , and c be three non-collinear points in the $[0, N] \times [0, N]$ section of the integer lattice, where $N \geq 10$. Then the turning angle of the path (a, b, c) at point b is at least $1/(3N^2)$.*

Proof. Let β denote the turning angle. We can assume that $0 < \beta < \pi/2$, since otherwise the inequality holds. Since a, b, c are non-collinear lattice points, the tangent function of the turning angle can be expressed as:

$$\tan \beta = \tan(\beta_2 - \beta_1) = \frac{\tan \beta_2 - \tan \beta_1}{1 + \tan \beta_1 \tan \beta_2},$$

where $\tan \beta_1 = \frac{m_1}{n_1} < \frac{m_2}{n_2} = \tan \beta_2$, and m_1, m_2, n_1, n_2 are nonnegative integers less or equal to N . Write $t_1 = \tan \beta_1$ and $t_2 = \tan \beta_2$. We distinguish two cases depending on whether the product $t_1 t_2$ is smaller or larger than 1:

Case 1: $t_1 t_2 \leq 1$. We have

$$\tan \beta = \frac{t_2 - t_1}{1 + t_1 t_2} \geq \frac{t_2 - t_1}{2} = \frac{1}{2} \left(\frac{m_2}{n_2} - \frac{m_1}{n_1} \right) = \frac{m_2 n_1 - m_1 n_2}{2n_1 n_2} \geq \frac{1}{2n_1 n_2} \geq \frac{1}{2N^2}.$$

Case 2: $t_1 t_2 \geq 1$. We have

$$\begin{aligned} \tan \beta &= \frac{t_2 - t_1}{1 + t_1 t_2} \geq \frac{t_2 - t_1}{2t_1 t_2} = \frac{1}{2} \left(\frac{1}{t_1} - \frac{1}{t_2} \right) \\ &= \frac{1}{2} \left(\frac{n_1}{m_1} - \frac{n_2}{m_2} \right) = \frac{m_2 n_1 - m_1 n_2}{2m_1 m_2} \geq \frac{1}{2m_1 m_2} \geq \frac{1}{2N^2}. \end{aligned}$$

Since $N \geq 10$ was assumed, in both cases it follows that $\beta \geq 1/(3N^2)$, as required. \square

Having polynomial representation of the small angles and rational points on circular arcs is a non-trivial problem [15, 14]. Without delving too much into technical details such as Lemma 12, we claim that for any δ , $0 < \delta < 1$, the construction can use integers polynomial in n for the coordinates of all points in Q , such that the angle determined by any three points in Q deviates by a multiplicative factor at least $1 - \delta$ and at most $1 + \delta$. Consequently, the reduction is strongly polynomial, and we have the following approximate versions of Lemmas 10 and 11:

Lemma 10 (Approximate Version). If there exists a set of k lines that cover all points in P , for some $k \leq n$, then there exists a Hamiltonian tour with maximum turning angle at most $(1 + \delta)(k + 2)\theta$ for Q .

Lemma 11 (Approximate Version). If there exists a Hamiltonian tour with maximum turning angle at most $(1 - \delta)k\theta$ for Q , for some $k \leq n$, then there exists a set of k lines that cover all points in P .

The following lemma shows that MIN-MAX-TURN HAMILTONIAN TOUR is almost as hard to approximate as COVERING POINTS BY LINES:

Lemma 13. For any $\rho \geq 1$, if MIN-MAX-TURN HAMILTONIAN TOUR admits a polynomial-time approximation algorithm with ratio ρ , then COVERING POINTS BY LINES admits a polynomial-time approximation algorithm with ratio $(1 + \varepsilon)\rho$ for any $\varepsilon > 0$.

Proof. Let P be a set of n points for COVERING POINTS BY LINES. Let k^* be the minimum number of lines necessary for covering all points in P . Without loss of generality, we assume that $k^* \geq 6/\varepsilon$, since otherwise a brute-force algorithm can find k^* lines to cover P in $n^{O(1/\varepsilon)}$ time, which is polynomial time for any fixed $\varepsilon > 0$.

Let $\delta = \Theta(\varepsilon)$ such that $\frac{1+\delta}{1-\delta} = \frac{1+\varepsilon}{1+\varepsilon/2}$. Under the assumption that $k^* \geq 6/\varepsilon$, we have the following algorithm: first construct a point-set Q from P (refer to Fig. 7), and then run the ρ -approximation algorithm for MIN-MAX-TURN HAMILTONIAN TOUR on Q to obtain a tour of maximum turning angle at most $(1 - \delta)\kappa\theta$ for some $\kappa > 0$, and finally obtain a set \mathcal{L} of at most $k = \lceil \kappa \rceil$ lines that cover P by Lemma 11 (approximate version).

By Lemma 10 (approximate version), the minimum value of the maximum turning angle of a Hamiltonian tour for Q is at most $(1 + \delta)(k^* + 2)\theta$ (i.e., there exists a Hamiltonian tour with maximum turning angle bounded as such). The ρ -approximation algorithm for MIN-MAX-TURN HAMILTONIAN TOUR guarantees that $(1 - \delta)\kappa\theta \leq \rho(1 + \delta)(k^* + 2)\theta$ and hence $\kappa \leq \rho \frac{1+\delta}{1-\delta} (k^* + 2)$. By the assumption that $k^* \geq 6/\varepsilon$, we have $k = \lceil \kappa \rceil \leq \rho \frac{1+\delta}{1-\delta} (k^* + 2) + 1 \leq \rho \frac{1+\delta}{1-\delta} (k^* + 3) = \rho \frac{1+\delta}{1-\delta} (1 + 3/k^*) k^* \leq \rho \frac{1+\delta}{1-\delta} (1 + \varepsilon/2) k^* = \rho(1 + \varepsilon) k^*$. \square

Since COVERING POINTS BY LINES is APX-hard (Theorem 3), the above lemma implies that MIN-MAX-TURN HAMILTONIAN TOUR is APX-hard too. This completes the proof of Theorem 6.

Consider the decision versions of the two Hamiltonian Tour problems:

- (I) MIN-MAX-TURN HAMILTONIAN TOUR (Decision Problem): Given n points in the plane and an angle $\alpha \in [0, \pi]$, decide whether there exists a Hamiltonian tour with maximum turning angle at most α .

- (II) **BOUNDED-TURN-MINIMUM-LENGTH HAMILTONIAN TOUR (Decision Problem):** Given n points in the plane, an angle $\alpha \in [0, \pi]$, and a positive number L , decide whether there exists a Hamiltonian tour with maximum turning angle at most α and length at most L .

Observe that the decision problem of **MIN-MAX-TURN HAMILTONIAN TOUR** is a special case of the decision problem of **BOUNDED-TURN-MINIMUM-LENGTH HAMILTONIAN TOUR**, with the parameter L set to some sufficiently large number, say, n times the diameter of the point set. Thus the APX-hardness (indeed NP-hardness suffices) of **MIN-MAX-TURN HAMILTONIAN TOUR** implies that **BOUNDED-TURN-MINIMUM-LENGTH HAMILTONIAN TOUR** is NP-hard. This completes the proof of Theorem 7.

Remark. While the decision problem of **BOUNDED-TURN-MINIMUM-LENGTH HAMILTONIAN TOUR** has both an *angle* constraint and a *length* constraint, it is interesting to note that our proof of its NP-hardness above (via the reduction from the decision problem of **MIN-MAX-TURN HAMILTONIAN TOUR**) effectively only uses the angle constraint. That is, the problem is already hard with the angle constraint alone. On the other hand, if the turning angle is unrestricted, i.e., if $\alpha = \pi$, then the problem **BOUNDED-TURN-MINIMUM-LENGTH HAMILTONIAN TOUR** is the same as the **EUCLIDEAN TRAVELING SALESMAN PROBLEM**, which is well known to be NP-hard with the length constraint alone. Our proof sheds light on a different aspect of the difficulty of the problem.

7 Concluding remarks

The obvious question left open by our work is whether **COVERING POINTS BY LINES** admits an approximation algorithm with constant ratio. Two other problems are finding approximation algorithms for **MIN-MAX-TURN HAMILTONIAN TOUR** and **BOUNDED-TURN-MINIMUM-LENGTH HAMILTONIAN TOUR**, respectively.

Acknowledgment. The authors are grateful to anonymous reviewers for bringing to our attention the previous APX-hardness results of Brodén et al. [12] and Kumar et al. [31], for suggesting the extension to lines—in our Theorem 1—of the result from [11] (Theorem 1, p. 1041) for line segments, and for pertinent advice on polynomial representation of points on circular arcs [15, 14].

References

- [1] A. Aggarwal, D. Coppersmith, S. Khanna, R. Motwani, and B. Schieber. The angular-metric traveling salesman problem. *SIAM Journal on Computing*, 29:697–711, 1999. An extended abstract in *Proceedings of the 8th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA’97)*, pages 221–229, 1997.
- [2] P. K. Agarwal, P. Raghavan, and H. Tamaki. Motion planning for a steering-constrained robot through moderate obstacles. In *Proceedings of the 27th Annual ACM Symposium on Theory of Computing (STOC’95)*, pages 343–352, 1995.
- [3] N. Alon. A non-linear lower bound for planar epsilon-nets. *Discrete & Computational Geometry*, 47:235–244, 2012.
- [4] N. Alon, D. Moshkovitz, and S. Safra. Algorithmic construction of sets for k -restrictions. *ACM Transactions on Algorithms*, 2:153–177, 2006.

- [5] E. M. Arkin, M. A. Bender, E. D. Demaine, S. P. Fekete, J. S. B. Mitchell, and S. Sethia. Optimal covering tours with turn costs. *SIAM Journal on Computing*, 35:531–566, 2005.
- [6] E. M. Arkin, J. S. B. Mitchell, and C. D. Piatko. Minimum-link watchman tours. *Information Processing Letters*, 86:203–207, 2003.
- [7] S. Bereg, P. Bose, A. Dumitrescu, F. Hurtado and P. Valtr. Traversing a set of points with a minimum number of turns, *Discrete & Computational Geometry*, 41:513–532, 2009.
- [8] P. Berman and M. Karpinski. On some tighter inapproximability results. *DIMACS Technical Report* 99-23, 1999.
- [9] P. Berman and M. Karpinski. Improved approximation lower bounds on small occurrence optimization. *Electronic Colloquium on Computational Complexity*, TR03-008, 2003.
- [10] J. Boissonat, A. Cérézo, and J. Leblond. Shortest paths of bounded curvature in the plane. *Journal of Intelligent and Robotic Systems: Theory and Applications*, 11:5–20, 1994.
- [11] V. E. Brimkov, A. Leach, J. Wu, and M. Mastroianni. Approximation algorithms for a geometric set cover problem. *Discrete Applied Mathematics*, 160:1039–1052, 2012.
- [12] B. Brodén, M. Hammar, and B. J. Nilsson. Guarding lines and 2-link polygons is APX-hard. In *Proceedings of the 13th Canadian Conference on Computational Geometry (CCCG’01)*, pages 45–48, 2001.
- [13] H. Brönnimann and M. T. Goodrich. Almost optimal set covers in finite VC-dimension. *Discrete & Computational Geometry*, 14:463–479, 1995.
- [14] C. Burnikel. Rational points on circles. Technical Report MPI-I-98-1-023, Max-Planck-Institut für Informatik, 1998.
- [15] J. Canny, B. Donald, E. K. Ressler. A rational rotation method for robust geometric algorithms. In *Proceedings of the 8th Annual Symposium on Computational Geometry (SOCG’92)*, pages 251–260, 1992.
- [16] T. M. Chan and E. Grant. Exact algorithms and APX-hardness results for geometric packing and covering problems. *Computational Geometry: Theory and Applications*, 47:112–124, 2014.
- [17] V. Chvátal. A greedy heuristic for the set-covering problem. *Mathematics of Operations Research*, 4:233–235, 1979.
- [18] K. L. Clarkson and K. Varadarajan. Improved approximation algorithms for geometric set cover. *Discrete & Computational Geometry*, 37:43–58, 2007.
- [19] G. Even, D. Rawitz, and S. Shahar. Hitting sets when the VC-dimension is small. *Information Processing Letters*, 95:358–362, 2005.
- [20] U. Feige. A threshold of $\ln n$ for approximating set cover. *Journal of the ACM*, 45:634–652, 1998.
- [21] S. P. Fekete and G. J. Woeginger. Angle-restricted tours in the plane. *Computational Geometry: Theory and Applications*, 8:195–218, 1997.
- [22] T. Frachard. Smooth trajectory planning for a car in a structured world. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 318–323, 1989.

- [23] M. R. Garey, D. S. Johnson, and L. Stockmeyer. Some simplified NP-complete problems. In *Proceedings of the 6th Annual ACM Symposium on Theory of Computing (STOC'74)*, pages 47–63, 1974.
- [24] M. Grantson and C. Levcopoulos. Covering a set of points with a minimum number of lines. In *Proceedings of the 22nd European Workshop on Computational Geometry*, pages 145–148, 2006.
- [25] D. Haussler and E. Welzl. ε -nets and simplex range queries. *Discrete & Computational Geometry*, 2:127–151, 2007.
- [26] D. S. Hochbaum. *Approximation Algorithms for NP-hard Problems*, PWS, 1997.
- [27] P. Jacobs and J. Canny. Planning smooth paths for mobile robots. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 2–7, 1989.
- [28] M. Jiang. On covering points with minimum turns. *International Journal of Computational Geometry & Applications*, 25:1–9, 2015.
- [29] D. S. Johnson. Approximation algorithms for combinatorial problems. *Journal of Computer and System Sciences*, 9:256–278, 1974.
- [30] S. Kratsch, G. Philip, and S. Ray. Point Line Cover: the easy kernel is essentially tight. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'04)*, pages 1596–1606, 2014.
- [31] V. S. A. Kumar, S. Arya, and H. Ramesh. Hardness of set cover with intersection 1. In *Proceedings of the 27th International Colloquium on Automata, Languages and Programming (ICALP'00)*, pages 624–635, 2000.
- [32] S. Langerman and P. Morin. Covering things with things. *Discrete & Computational Geometry*, 33:717–729, 2005.
- [33] J. Le Ny, E. Frazzoli, and E. Feron. The curvature-constrained traveling salesman problem for high point densities. In *Proceedings of the 46th IEEE Conference on Decision and Control*, pages 5985–5990, 2007.
- [34] L. Lovasz. On the ratio of optimal integral and fractional covers. *Discrete Mathematics*, 13:383–390, 1975.
- [35] C. Lund and M. Yannakakis. On the hardness of approximating minimization problems. *Journal of the ACM*, 41:960–981, 1994.
- [36] N. Megiddo and A. Tamir. On the complexity of locating linear facilities in the plane. *Operation Research Letters*, 1:194–197, 1982.
- [37] J. S. B. Mitchell. Personal communication, December 2013.
- [38] N. H. Mustafa and S. Ray. Improved results on geometric hitting set problems. *Discrete & Computational Geometry*, 44:883–895, 2010.
- [39] N. H. Mustafa, R. Raman, and S. Ray. QPTAS for geometric set-cover problems via optimal separators. arXiv:1403.0835, 2014.
- [40] J. Pach and P. K. Agarwal. *Combinatorial Geometry*, John Wiley, New York, 1995.
- [41] J. Pach and G. Tardos. Tight lower bounds for the size of epsilon-nets. *Journal of the American Mathematical Society*, 26:645–658, 2013.

- [42] R. Raz and S. Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing (STOC'97)*, pages 475–484, 1997.
- [43] P. Slavík. A tight analysis of the greedy algorithm for set cover. *Journal of Algorithms*, 25:237–254, 1997.
- [44] C. Stein and D. P. Wagner. Approximation algorithms for the minimum bends traveling salesman problem. In *Proceedings of the 8th International Conference on Integer Programming and Combinatorial Optimization (IPCO'01)*, pages 406–421, 2001.
- [45] J. Wang, W. Li, and J. Chen. A parameterized algorithm for the hyperplane-cover problem. *Theoretical Computer Science*, 411:4005–4009, 2010.