# APPROXIMATION ALGORITHMS FOR HITTING OBJECTS WITH STRAIGHT LINES

Refael Hassin* and Nimrod Megiddo†

Revised June 1989

**Abstract.** In the hitting set problem one is given $m$ subsets of a finite set $N$ and one has to find an $X \subset N$ of minimum cardinality that "hits" (intersects) all of them. The problem is NP-hard. It is not known whether there exists a polynomial-time approximation algorithm for the hitting set problem with a finite performance ratio. Special cases of the hitting set problem are described for which finite performance ratios are guaranteed. These problems arise in a geometric setting. We consider special cases of the following problem: Given $n$ compact subsets of $R^d$, find a set of straight lines of minimum cardinality so that each of the given subsets is hit by at least one line. The algorithms are based on several techniques of representing objects by points, not necessarily points on the objects, and solving (in some cases, only approximately) the problem of hitting the representative points. Finite performance ratios are obtained when the dimension, the number of types of sets to be hit and the number of directions of the hitting lines are bounded.

## 1. Introduction

In the *hitting set* problem [4] one is given $m$ subsets $R_i \subset N = \{1, \cdots, n\}$ $(i = 1, \cdots, m)$, and one has to find an $X \subset N$ of minimum cardinality so that $X \cap R_i \neq \emptyset$ $(i = 1, \cdots, m)$ . The hitting set problem is "isomorphic" to the *set covering* problem where, given subsets $C_j \subset M = \{1, \cdots, m\}$ $(j = 1, \cdots, n)$ , one is asked to find a set of indices $X \subset N$ of minimum cardinality such that $\bigcup_{j \in X} C_j = M$. The sense of the isomorphism is that the formulations of these problems as integer linear programming problems are identical. The problems are famous as being NP-complete and thus substantial effort

*Department of Statistics, School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel.

†IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120-6099, and School of Mathematical Sciences, Tel Aviv University, Tel Aviv, Israel.

has been made in the direction of solving them approximately (see for example [2; 5; 6; 7]).

An approximation algorithm in the context of the hitting set problem is one that delivers a set $X \subset N$ not necessarily of minimum cardinality. The performance of an approximation algorithm for a minimization problem is sometimes evaluated by the supremum (over problem instances) of the ratio of the solution value delivered by the algorithm to the optimal solution value (the *performance ratio*). It is not known whether there exists a polynomial-time approximation algorithm for the hitting set problem with a finite upper bound on its performance ratio. Such finite performance ratios are known only for special cases. For example, if the cardinalities of the sets $R_i$ are bounded by $d$ then a performance ratio of at most $d$ is guaranteed [5; 1]. If the cardinalities of the sets $C_j$ (in the set covering interpretation) are bounded by $k$, a performance ratio of at most $H(k) = \sum_{i=1}^{k} 1/i$ is guaranteed by the greedy heuristic [7] (see also [2; 8]).

In this paper we describe special cases of the hitting set problem for which finite performance ratios are guaranteed. These problems arise in a geometric setting. The problems we consider are special cases of the following: Given $n$ compact subsets of $R^d$, find a set of straight lines of minimum cardinality so that each of the given subsets is hit by at least one line. In a crude form this problem is a hitting set problem with an infinite matrix but can easily be reduced to one with a finite matrix of dimensions polynomial in the length of a certain description of the subsets. We restrict attention to hitting lines with their slopes restricted to a finite set, mostly lines parallel to the coordinate axes. We obtain finite performance ratios if we bound the dimension of the space and the types of sets.

**Remark 1.1.** It is worth mentioning that the greedy algorithm does not provide a finite approximation ratio even in the problem of minimum cover of points in the plane by lines parallel to the coordinate axes. The following example is a realization of an example given in [7]. Let $k$ be an integer. We construct $k$ disjoint sets $S_j$ $(j = 1, \cdots, k)$ consisting of $k!$ points each (see Figure 1). The set $S_j$ is a union of $k!/j$ disjoint sets $S_j^{(i)}$ $(i = 1, \cdots, k!/j)$ consisting of $j$ points each. Let $H(j)$ denote $\sum_{q=1}^{j} 1/q$. The optimal solution of this problem consists of the $k!$ horizontal lines $L_r = \{y = r\}$, $(r = 1, \cdots, k!)$ (see Figure 1(a)). The greedy heuristic may pick the $k!H(k)$ vertical lines $M_r = \{x = r\}$, $(r = 1, \cdots, k!H(k))$ (see Figure 1(b)).

The problem of hitting objects in the plane with a minimum number of straight lines has a military application. In many cases when a bomber attempts to destroy targets on the ground, protected by anti-aircraft missiles, it has to spend as little time as possible close to the targets. Thus, careful planning of an air raid on a multi-target site (for example, a clsuter of fuel tanks) calls for a minimum number of times a bomber has to fly across the site. Moreover, each pass has to be carried out as fast as possible, hence for each dive into the site there exists a straight line (a "stick") along which targets are destroyed. Another application (in three dimensions) is in medicine where radiotherapy is administered by inserting a minimum number of radioactive needles into a certain area of the body so as to achieve a required level of radiation (see [11]

To conclude this introduction we mention another application of a special case of our problem. Consider a system consisting of $m$ components $c_1, \cdots, c_m$ and a schedule which tells for each component $c_j$ the (finite) set of time intervals during which $c_j$ is busy. We need to inspect each component at least once during each time interval of its operation. An inspection may be carried out in either of the following ways. First, we may place a permanent monitor on a component at a total cost of $\alpha$ for all the checks on that component. Second, we may perform a check of the entire system at a time point and then we incur a total cost of $\beta$ for checking all the components active at the time. The problem can be modeled as follows. For each component $c_j$ place the time intervals of $c_j$ on the line $L_j = \{(x, y) : y = j\}$ in the plane. We now have a collection of horizontal line segments in the plane. The problem is to find a collection of horizontal and vertical lines, so that each segment is hit by at least one of the lines, minimizing $\alpha$ times the number of horizontal lines plus $\beta$ times the number of vertical ones. We later show that this problem is NP-hard even when all the intervals have unit length and endpoints of integral coordinates. Among other results, we provide a polynomial-time approximation algorithm for the latter with a performance ratio of at most 2.

In Section 2 we discuss various problems of hitting line segments by vertical and horizontal lines in the plane. In Section 3 we consider lines of restricted slopes. In Section 4 we discuss hitting general compact sets in the plane. In Section 5 we discuss extensions to higher dimensions. In Section 6 we prove NP-hardness of special cases of the problems considered here.

## 2. Hitting line segments in the plane

In this section we discuss approximation algorithms for the following problem:

**Problem 2.1. Hitting segments in the plane.** Given a set $\Sigma = \{\sigma_1, \cdots, \sigma_n\}$ of vertical and horizontal line segments in the plane, find the smallest number $k$ and a set $\{\ell_1, \cdots, \ell_k\}$ of straight lines parallel to the axes, so that each $\sigma \in \Sigma$ is hit (intersected) by at least one of the lines $\ell_i$.

Problem 2.1 is NP-hard even when all the line segments are horizontal with length 1 and their endpoints have integral coordinates (see Proposition 6.2 below). Thus, we are interested here in approximation algorithms. The following proposition follows immediately from the famous König-Egerváry theorem [3] and the fact the max-flow problem can be solved in polynomial-time:

**Proposition 2.2.** *The problem of hitting points in the plane with a minimum number of lines parallel to the axes can be solved in polynomial time.*

*Proof:* Consider a graph $G$ whose vertices correspond to all the straight lines passing through the given points and parallel to the axes. Let the edges of $G$ correspond to the input points, so that a horizontal line and a vertical line that intersect at an input point correspond to adjacent vertices in $G$. Obviously, $G$ is bipartite. It is easy to see that Problem 2.1 is equivalent to finding a minimum vertex cover in $G$. It is well known (see [3]) that the latter can be solved in polynomial time as a maximum flow problem. ∎

It is interesting to note that the same problem in $R^3$ is NP-hard (see Proposition 6.4 below).

We start with the case of line segments of equal lengths. For several of the propositions below we omit the proofs. As a representative case we choose the following:

**Proposition 2.3.** *There is a polynomial-time approximation algorithm with a performance ratio of at most* $2 - \frac{1}{K+1}$ *for the case where* $\Sigma$ *is a set of horizontal line segments of length* $K$ *whose endpoints have integral* $x$*-coordinates.*

*Proof:* Consider a line segment $[(a,b),(a+K,b)]$ where $a$ is an integer. Obviously, the numbers $a, a+1, \cdots, a+K$ have all the possible residues modulo $K+1$. For every $j$ ($0 \le j \le K$) consider the representation of segments $\sigma$ by points $p_j = p_j(\sigma)$ ($p_j(\sigma) \in \sigma$) where the $x$-coordinate of $p_j$ is congruent to $j \pmod{K+1}$. Let $\Sigma'_j = \{p_j(\sigma) : \sigma \in \Sigma\}$ ($j = 0, \cdots, K$). The approximation algorithm solves the point cover problem with respect to each of the sets $\Sigma'_j$ and picks the one with the smallest number of lines. We will prove that this number is at most $2 - \frac{1}{K+1}$ times the number of lines in an optimal solution of the segment hitting problem. Consider an optimal solution $L = \{\ell_1, \cdots, \ell_k\}$ for $\Sigma$. Without loss of generality assume that for each vertical line $\ell_i = \{x = \xi_i\}$, $\xi_i$ is an integer. Denote by $k_j$ the number of vertical lines in $L$ such that $\xi_i \equiv j \pmod{K+1}$ ($j = 0, \cdots, K$). Consider the point cover problem with the set $\Sigma'_j$. We claim that the number of lines in an optimal solution of the latter is at most $2k - k_j - \bar{k}$, where $\bar{k}$ is the number of horizontal lines in $L$. To prove this claim, consider first a vertical line $\ell_i$ such that $\xi_i \equiv j \pmod{K+1}$. Obviously, if a segment $\sigma$ is hit by $\ell_i$ then $p_j(\sigma)$ lies on $\ell_i$. If $\xi_i \not\equiv j \pmod{K+1}$ then the $x$-coordinates of point $p_j(\sigma)$ for segments $\sigma$ which are hit by $\ell_i$ have at most two distinct values. Thus, we can replace $\ell_i$ by two vertical lines $\ell'_i$ and $\ell''_i$ which cover all these points $p_j(\sigma)$. Also, a horizontal line obviously covers the representative point of each (horizontal) segment hit by it. Finally, since $\sum_j k_j = k - \bar{k}$, we have $(\bar{k} + \max_j k_j)/k \ge \frac{1}{K+1}$. ∎

**Remark 2.4.** If under the conditions of Proposition 2.3 we allow both horizontal and vertical segments then a performance ratio of at most $3 - \frac{1}{K+1}$ can be guaranteed by the same algorithm.

**Problem 2.5.** Find a minimum hitting set as in Problem 2.1, given that all the horizontal segments have the same length and all the vertical ones have the same length.

As already mentioned, Problem 2.5 is NP-hard. However, we have

**Proposition 2.6.** *Problem 2.5 has a polynomial-time approximation algorithm whose performance ratio is at most* 3*.*

**Remark 2.7.** If in Proposition 2.6 $\Sigma$ contains only horizontal segments then a performance ratio of 2 can be guaranteed since in that case a vertical line from the optimal solution can be *replaced* by two lines, whereas no horizontal lines have to be added. By symmetry, the same argument holds if all the segments are vertical.

**Remark 2.8.** We note that our approximation algorithms have the same guarantees of performance ratios in a more general case where hitting lines in different directions have different costs and one is asked to minimize the total cost of hitting lines. The validity of this claim follows from the fact that in the proof we replace a line of the optimal solution by parallel lines for the solution of the approximating problems.

Our next development is an algorithm for segments of various lengths. For this case the finite guaranteed performance ratio depends on an upper bound on the length of segments.

**Proposition 2.9.** *If* $\Sigma$ *is a set of integral points and vertical and horizontal line segments of lengths between* 1 *and* $K$ *then an approximate solution for Problem 2.1 with a performance ratio of at most* $2\lceil \log_2 K \rceil + 3$ *can be found in polynomial time.*

*Proof:* Without loss of generality assume $K = 2^m$ for some integer $m$. The idea of the algorithm again is to represent each line segment $\sigma \in \Sigma$ by a point $p = p(\sigma)$ and solve the problem of minimum cover of the set of points $\Sigma' = \{p(\sigma) : \sigma \in \Sigma\}$. Here we choose the representative as follows. Let $\sigma = [(a,b),(a+\delta,b)]$ $(1 \leq \delta \leq K)$ be a horizontal segment. The representation of vertical segments is analogous. We represent $\sigma$ by a point $p(\sigma) = (q,b)$ where $a \leq q \leq a+\delta$ and $q \equiv 0 \pmod{2^\nu}$ with the largest possible integer $\nu \leq m$. Note that we may have both $a \equiv 0 \pmod{K}$ and $a+\delta \equiv 0 \pmod{K}$ in which case we fix $q = a$. Otherwise the number $q$ is uniquely defined by the above condition. In particular, if $\sigma$ has length 1 and has endpoints with integral

$x$-coordinates then we choose the one with the even value; if the line segment contains a unique point with an integral $x$-coordinate then this point is chosen as $q$, and finally if the segment consists of a unique point then this point is chosen as $q$. Consider an optimal solution $L = \{\ell_1, \cdots, \ell_k\}$ for $\Sigma$. Suppose $\ell_i = \{(x, y) : x = \xi_i\}$ is a vertical line and consider any horizontal segment $\sigma$ hit by it. It is easy to see that if $p(\sigma) = (q, b)$ and $q \equiv 0 \pmod{2^\nu}$ then there is no $q' \equiv 0 \pmod{2^\nu}$ between $\xi_i$ and $q$ (including $q' = \xi_i$). This implies that the number of distinct $q$'s for such segments is at most $m + 1$ on either side of $\ell_i$. The choice $q = \xi_i$ may also be necessary. This implies that the number of lines in an optimal solution of the point cover problem with the set $\Sigma'$ is at most $2(m + 1) + 1$ times the optimum of the segment hitting problem. ∎

## 3. Hitting with lines of a finite number of slopes

One possible extension of the results of Section 2 in the plane is to consider hitting problems where the hitting lines are not restricted to be parallel to the axes. For example, we may consider the case where these lines must have their slopes taken from a given finite set of rational numbers $S = \{r_1, \cdots, r_s\}$. However, let us first comment on the case with no restrictions at all:

**Problem 3.1.** Given a collection $\Sigma = \{\sigma_1, \cdots, \sigma_n\}$ of horizontal unit segments in $R^2$, find the smallest number $k$ and a set $\{\ell_1, \cdots, \ell_k\}$ of straight lines so that each $\sigma_j$ is hit by at least one of the lines $\ell_i$.

**Remark 3.2.** The NP-hardness of Problem 3.1 can be proven with the techniques of [10; 9]. For an approximate solution, we might represent each segment $\sigma \in \Sigma$ by a point $p(\sigma)$ whose $x$-coordinate is an integer, and solve the point cover problem on the set $\Sigma' = \{p(\sigma) : \sigma \in \Sigma\}$. Unfortunately, the latter is also NP-hard [10]. Furthermore, we do not know whether the point cover problem with no restrictions on the slopes of the hitting lines has a polynomial-time approximation algorithm with a finite performance ratio.

In view of Remark 3.2, we now consider the case where the slopes of the hitting lines are restricted to a given finite set of integer numbers. We note that a problem with rational slopes can be reduced to one with integral slopes.

**Problem 3.3.** Given a collection $\Sigma = \{\sigma_1, \cdots, \sigma_n\}$ of horizontal unit segments in $R^2$ with integral endpoints and a finite set of integers $S = \{K_1, \cdots, K_s\}$ (including the possibility of $K_1 = \infty$), find the smallest number $k$ and a set $\{\ell_1, \cdots, \ell_k\}$ of straight lines whose slopes are in $S$, so that each $\sigma_j$ is hit by at least one of the lines $\ell_i$.

**Remark 3.4.** Under the conditions of Problem 3.3, the corresponding point cover problem is NP-hard for $s \geq 3$. The proof of this fact is similar to the one we give in Proposition 6.4 for hitting in $R^3$. Alternatively, the problem of hitting points in $R^3$ can be reduced by a careful projection to the point cover in $R^2$ with lines of three slopes. Fortunately, the point cover problem has a polynomial-time approximation algorithm with performance ratio $s$. To see this, we recall a result from [5] and [1]. As an integer linear programming problem, the set covering problem is the following:

$$\begin{aligned} \text{Minimize} \quad & e^T x \\ \text{subject to} \quad & Ax \geq e \\ & x \geq 0 \end{aligned}$$

where $A$ is a zero-one matrix and $e = (1, \cdots, 1)^T$. Although the problem is NP-complete, it is possible to find in linear time [1] a feasible solution $z$ with $e^T z$ not greater than $s$ times the optimum, where $s$ is the largest number of 1's in any row of $A$ (in [5] the same level of approximation is obtained by rounding up an optimal solution of the linear programming relaxation of the problem). The proof is quite simple and we indicate it here for the sake of completeness. The algorithm selects a maximal set (with respect to inclusion) of mutually orthogonal rows of $a_{i_1}, \cdots, a_{i_m}$ of $A$. In the terminology of the set hitting problem (see Section 1) this corresponds to a maximal collection of disjoint subsets $R_i$. Obviously, for any feasible solution $x$, $e^T x \geq m$ since we have to hit $m$ disjoint sets. If we use all the members of the union of these sets $R_i$ then we get a hitting set since the collection is maximal. Specifically, if we set $x_j = 1$ for every $j$ such that $a_{i_\nu j} = 1$ for some $1 \leq \nu \leq m$ and $x_j = 0$ otherwise, then $x$ is a feasible solution and $e^T x \leq sm$.

**Proposition 3.5.** *There exists a polynomial-time approximation algorithm for Problem 3.3 with a performance ratio of at most* $(2 \sum \lceil \log_2 |K_j| \rceil + 3)s$ *where the sum is taken over the finite nonzero members of* $S$.

*Proof:* If a line $\ell$ of finite slope $K_i$ hits a certain set of horizontal unit segments, then there exists such a line which hits the same set of segments where at least one of the segments is hit at an endpoint. It follows that each segment is hit at a point where the $x$-coordinate is of the form $\frac{q}{|K_i|}$ for some integer $q$. By this argument, there exists an optimal solution to the Problem 3.3 where each segment is hit at an $x$-coordinate of the form $\frac{q}{K}$ where $K$ is the l.c.m. (least common multiple) of the finite nonzero members of $S$. We now apply the idea of the proof of Proposition 2.9. Each segment will be represented by a point of the form $\frac{q}{K}$ where $q$ is chosen to be congruent to $0 \pmod{2^\nu}$ with the largest possible $\nu$. The problem of covering the representative points can be solved approximately in polynomial time, with a performance ratio of $s$ (see Remark 3.4). The result then follows by the arguments in the proof of Proposition 2.9. ∎

## 4. Hitting general sets in the plane

We now consider the problem of hitting general sets in the plane. However, we make certain simplifying assumptions about the sets.

**Problem 4.1.** Given a collection $\Gamma$ of compact and connected sets $C_1, \cdots, C_n \subset R^2$ and a finite set of integers $S = \{r_1, \cdots, r_s\} \subset (-\infty, \infty]$, find the smallest number $k$ and a set $\{\ell_1, \cdots, \ell_k\}$ of straight lines whose slopes are in $S$, so that each $C_j$ is hit by at least one of the lines $\ell_i$.

**Remark 4.2.** The complexity of Problem 4.1 depends of course on the description of the sets $C_j$. We are interested in special cases where the sets $C_j$ are copies (more accurately, translates) of a fixed finite number of types. For example, consider a special case where each $C_j$ is either a disk of radius 1 or a square of area 1 whose edges are parallel to the coordinate axes. In principle, for every fixed finite set of types an approximation algorithm with a guaranteed finite performance ratio can be designed along the lines presented above. We first construct a grid $G$ of points in the plane so that each $C_j$ contains at least one point of $G$. We then represent each $C_j$ by a grid point $p(C_j)$ and solve the point cover problem with the set $\Gamma' = \{p(C_j) : j = 1, \cdots, n\}$. The grid has to be chosen so as to satisfy the following conditions. There should exist a finite number $K$ so that for every line $\ell$ (whose slope is in $S$) there is a set of lines $\ell'_1, \cdots, \ell'_K$ parallel to $\ell$, so that $p(C_j) \in \bigcup_i \ell'_i$ for every $C_j$ such that $C_j \cap \ell \neq \emptyset$.

A more careful consideration reveals a simplification of the solution. Given the set of slopes $S$ (see Problem 4.1), let us define for any set $C \subset R^2$ a "closure" $\bar{C}$ with respect to $S$ as follows. For every $\nu$, $1 \le \nu \le s$, let $v^\nu \in R^2$ denote a vector in the direction corresponding to the slope $r_\nu$. Denote $a_\nu = \inf\{x^T v^\nu : x \in C\}$ and $b_\nu = \sup\{x^T v^\nu : x \in C\}$. Now,

$$\bar{C} = \{x \in R^2 : a_\nu \le x^T v^\nu \le b_\nu \ , \nu = 1, \cdots, s\} \ .$$

Note that $\bar{C}$ is a convex polygon with at most $2s$ edges. Figure 2 depicts a set $S$ of directions, a set $C$ and the closure $\bar{C}$ relative to $S$.

**Proposition 4.3.** *Problem 4.1 with sets $C_1, \cdots, C_n$ is equivalent to Problem 4.1 with sets $\bar{C}_1, \cdots, \bar{C}_n$.*

We omit the easy proof. As an example of an application of Proposition 4.3, consider the problem of hitting $n$ given unit disks in the plane with a minimum number of lines parallel to the axes. It follows that this problem is equivalent to hitting unit squares whose edges are parallel to the axes. It can be easily seen, using the techniques presented so far, that the latter has a polynomial-time approximation algorithm with a performance ratio of at most 2. In fact, hitting $n$ identical rectangles (whose edges are parallel to the axes) by lines parallel to the axes has a performance ratio of 2. In view of Proposition 4.3, we have the following:

**Proposition 4.4.** *The problem of hitting $n$ copies (translates) of a compact connected set $B \subset R^2$ with a minimum number of lines parallel to the axes has a polynomial-time approximation algorithm with a performance ratio of at most 2.*

A generalization of the latter to the case of more than one type is obvious:

**Proposition 4.5.** *The problem of hitting $n$ sets, each of which is a copy (translate) of one of $p$ compact connected sets $B_1, \cdots, B_p \subset R^2$ with a minimum number of lines parallel to the axes has a polynomial-time approximation algorithm with a performance ratio of at most $2p$.*

*Proof:* First, we can replace each of the given sets by a rectangle and these rectangles will be of at most $p$ distinct types. For each type we construct a grid of points from which representatives of rectangles of the type are chosen. A line $\ell$ in an optimal solution of the problem can be replaced by at most two lines per type so that the resulting $2p$ lines cover all the representatives of rectangles covered by $\ell$. ∎

**Remark 4.6.** In Proposition 4.5 we can allow the hitting lines to have any two distinct slopes and not necessarily the ones parallel to the axes. If the number of permissible slopes is greater than 2 then we still get a finite performance ratio in polynomial time for any fixed set of rational slopes. However, the ratio itself depends on the set of slopes and there does not seem to exist a uniform bound independent of the slopes themselves.

Yet another generalization is to disconnected sets. A finite performance ratio can be achieved in polynomial time for any fixed set of rational slopes and a fixed set of types. A more explicit result can be proven in terms of the number of connected components in the case of two slopes.

**Proposition 4.7.** *Let $B \subset R^2$ be any compact set consisting of $q$ connected components. The problem of hitting $n$ copies (translates) of $B$ with a minimum number of lines parallel to the axes has a polynomial-time approximation algorithm with a performance ratio of at most $2q$.*

*Proof:* In view of Proposition 4.5, we can assume without loss of generality that each connected component of $B$ is a rectangle whose edges are parallel to the axes. Our approximation algorithm considers only two of the rectangles, namely, $B^x$ whose $x$-dimension is maximal and $B^y$ whose $y$-dimension is maximal. For simplicity, suppose the units on the axes are equal to these maxima, respectively. Now, for each copy $C_j$ of $B$, consider the corresponding copies of the rectangles $B^x$ and $B^y$ and denote them by $C_j^x$ and $C_j^y$, respectively. It is easy to see that for every $C_j$ there exists at least one grid point $p = p(C_j) = (\xi_j, \eta_j)$ such that the line $\{x = \xi_j\}$ intersects $C_j^x$ and the line $\{y = \eta_j\}$ intersects $C_j^y$. (If there is more than one such point choose the point with smaller coordinates). Figure 3 depicts a set $C$ consisting of three disjoint rectangles and its representative point. Note that here is another case where the representative points do not necessarily belong to the sets they represent. We now solve the point cover problem on the set of points $p(C_j)$. We claim that this algorithm has a performance ratio of at most $2q$. The proof is as follows. Consider a line $\ell$ from an optimal solution. Suppose $\ell$ is vertical (the case of a horizontal line is analogous). For each set $C_j$ which is hit by $\ell$, there is a certain connected component $C_j'$ of $C_j$ hit by $\ell$. We classify these sets $C_j$ according to the connected component hit by $\ell$. For all the sets in the same class there exist at most two lines parallel to $\ell$ which cover all the representative points of members of the class. The number of classes is at most $q$. ∎

## 5. Hitting in higher dimensions

Our approximation algorithms for hitting certain objects in the plane are based on the fact that minimum cover of points in the plane (by lines parallel to the axes) can be found in polynomial time. It is possible to extend the arguments relating the problems of hitting representative points and hitting objects to higher dimensions. However, unfortunately, the problem of hitting points in $R^3$ by lines parallel to the axes is NP-complete. We prove this fact in Section 6. On the other hand, the point cover problem can be formulated as a set covering problem (see Remark 3.4) and for every fixed dimension the number of 1's per row is fixed. This implies the existence of polynomial-time approximation algorithm with finite performance ratios for the point cover problem, on which we can base such algorithms for hitting more general sets. The first claim is:

**Proposition 5.1.** *There is a linear-time approximation algorithm for the problem of minimum cover of points in $R^d$ by lines parallel to the axes with a performance ratio of at most $d$.*

*Proof:* See Remark 3.4 for the argument of the proof. ∎

Proposition 5.1 suggests the development of approximation algorithms for more general sets based on choosing representative points for the sets. In view of the observations in Section 4, rectangular boxes play the key role. Thus, we consider the following problem:

**Problem 5.2. Hitting boxes.** Given a set $\Pi = \{\pi_1, \cdots, \pi_n\}$ of identical rectangular boxes in $R^d$ with edges parallel to the axes, find the smallest number $k$ and a set $\{\ell_1, \cdots, \ell_k\}$ of straight lines parallel to the axes, so that each $\pi \in \Pi$ is hit by at least one of the lines $\ell_i$.

**Proposition 5.3.** *There is a polynomial-time algorithm for Problem 5.2 with a performance ratio of at most $d\ 2^{d-1}$.*

*Proof:* Without loss of generality, assume the boxes are unit cubes and consider the grid of integral points in $R^d$. For every cube $\pi \in \Pi$, let $p(\pi) \in \pi$ denote a grid point. If there is more than one grid point in $\pi$ then choose the one which is smallest in the partial order on $R^d$ (there is a unique minimum since the cubes are parallel to the axes). Now, obtain an approximate solution for the point cover problem on the set $\Pi' = \{p(\pi) : \pi \in \Pi\}$ by solving it as a set covering problem with $d$ 1's per row. The set of lines obtained in this way is a feasible solution for Problem 5.2. We claim that this solution has a performance ratio of at most $d\ 2^{d-1}$. This follows from the fact that every line $\ell$ in a feasible solution of Problem 5.2 can be replaced by $2^{d-1}$ lines $\ell^1, \cdots, \ell^{2^{d-1}}$ parallel to $\ell$ so that $p(\pi)$ is on one of these lines provided $\pi \cap \ell \neq \emptyset$. Specifically, if the line $\ell$ has the form $\ell = \{x \in R^d : x_i = \xi_i\ ,\ i = 1, \cdots, d\ ,\ i \neq j\}$, the $2^{d-1}$ lines we refer to have the form $\ell' = \{x \in R^d : x_i = \xi_i'\ , i = 1, \cdots, d\ ,\ i \neq j\}$, where $\xi_i' \in \{\lceil \xi_i \rceil, \lceil \xi_i \rceil - 1\}$. ∎

As in Sections 3 and 4, it is possible to extend this result to more general, not necessarily connected, compact sets and to directions other than parallel to the axes.

## 6. NP-completeness results

In this section we prove NP-hardness of two of the problems discussed in the paper.

**Problem 6.1. Minimum hitting of horizontal unit segments.** Given are $n$ pairs $(a_i, b_i)$ $(i = 1, \cdots, n)$ of integers and an integer $k$. Recognize whether there exist in the plane $k$ straight lines $\ell_1, \cdots, \ell_k$ parallel to the axes with the property that each line segment $[(a_i, b_i), (a_i + 1, b_i)]$ is hit by at least one of the lines.

**Proposition 6.2.** *The problem of minimum hitting of horizontal unit segments is NP-complete.*

*Proof:* We prove the proposition by reduction from 3-satisfiability where each clause contains precisely three distinct variables. Given a set of clauses $E_j = x_j \vee y_j \vee z_j$ $(j = 1, \cdots, m)$ where $\{x_j, y_j, z_j\} \subset \{u_1, \bar{u}_1, \cdots, u_n, \bar{u}_n\}$, we represent each variable $u_i$

by six segments as follows. For simplicity, let a line segment $[(a,b),(a+1,b)]$ be denoted $[a,b]$. The collection of six segments representing a single variable looks like the set

$$S \quad = \quad \{[1,2],[2,4],[3,3],[5,2],[6,1],[7,3]\} \; .$$

It is easy to check that three lines are both necessary and sufficient to intersect all the members of $S$. Moreover, any set of three such lines contains exactly one of lines $\{y=2\}$ and $\{y=3\}$. The assignment of a truth-value to the variable corresponds the choice of one of these lines. The actual set $S_i$ representing $u_i$ is a translate of the set $S$, namely, $S_i = S + (8i, 4i)$ . Note that there is no vertical or horizontal line that intersects members of more than one of the sets $S_i$. Thus, the presence of the line $\{y = 4i+2\}$ corresponds to $u_i$ being true whereas the presence of $\{y = 4i+3\}$ means $u_i$ is false. A clause $E_j$ is represented by a set of five segments consisting of a pair $T_j = \{[8n+6j+2, -2j-1], [8n+6j+4, -2j-2]\}$ and a triple $R_j$ which depends on the structure of $E_j$ as we specify below. Note that there is no vertical or horizontal line that intersects members of more than one of the pairs $T_j$. The triple $R_j$ is defined as follows. Each literal of $E_j$ contributes one segment. The $x$-coordinates of the left-hand endpoints of the segments corresponding to $x_j, y_j$ and $z_j$ are $8n+6j+1, 8n+6j+3$ and $8n+6j+5$, respectively. If a literal equals $u_i$ then the $y$-coordinate of the corresponding segment is equal to $4i+2$. If it equals $\bar{u}_i$ then this coordinate is equal to $4i+3$. Figure 4 depicts a clause $E_j = u_1 \vee u_2 \vee \bar{u}_3$. It is easy to check that three lines are both necessary and sufficient for hitting the five segments in $T_j \cup R_j$. Moreover, for any proper subset $R'_j \subset R_j$, two lines are both necessary and sufficient for hitting the segments in $T_j \cup R'_j$. We leave it to the reader to verify that the set of clauses $E_1, \cdots, E_m$ is satisfiable if and only if the set of segments we have constructed can be covered in the sense discussed above by $3n+2m$ vertical or horizontal lines. ∎

**Problem 6.3. Point cover in $R^3$.** Given a set $S$ of $n$ points $(x_i, y_i, z_i)$ $(i = 1, \cdots, n)$ and an integer $k$, recognize whether there exist $k$ lines parallel to the axes whose union contains $S$.

**Proposition 6.4.** *The problem of point cover in $R^3$ is NP-complete.*

*Proof:* A line parallel to an axis in $R^3$ can be represented by a triple of one of the following forms: $(a,b,*)$, $(a,*,b)$ and $(*,a,b)$, so that a point $(x,y,z)$ can be covered only by one of the lines $(x,y,*)$, $(x,*,z)$ and $(*,y,z)$. The problem is obviously in NP. We prove the proposition by reduction from 3-satisfiability. Given a set of clauses $E_j = x_j \vee y_j \vee z_j$ $(j = 1, \cdots, m)$ where $\{x_j, y_j, z_j\} \subset \{u_1, \bar{u}_1, \cdots, u_n, \bar{u}_n\}$, we represent each variable $u_i$ by a set $S_i$ of an even number of points. All the sets $S_i$ have a cyclic structure like the following set:

$$S = \{(a_1,b_1,c_1),(a_1,b_1,c_2),(a_1,b_2,c_2),(a_2,b_2,c_2),(a_2,b_2,c_3),(a_2,b_3,c_3),\cdots$$
$$\cdots,(a_{\ell+1},b_{\ell+1},c_{\ell+1}) = (a_1,b_1,c_1)\} \; ,$$

where $\ell \geq 4m$ is even and the cardinality of $S$ is $3\ell$. The set $S$ can be covered by $1.5\ell$ lines. There are exactly two sets of $1.5\ell$ lines that cover the members of $S$:

$$T = \{(a_1, b_1, *), (*, b_2, c_2), (a_2, *, c_3), (a_3, b_3, *), (*, b_4, c_4), (a_4, *, c_5), \cdots, (a_{\ell-1}, *, c_1)\}$$

and

$$F = \{(a_1, *, c_2), (a_2, b_2, *), (*, b_3, c_3), (a_3, *, c_4), (a_4, b_4, *), (*, b_5, c_5), \cdots, (*, b_1, c_1)\} \ .$$

The two covers of the set $S_i$ correspond to the assignment of truth-values to the variable $u_i$. Thus, there are two sets, $T_i, F_i$, each consisting of $1.5\ell$ lines, that cover the members of $S_i$ and we will refer to them as the "true" and the "false" lines of $u_i$. We choose the numerical values defining the points in the sets $S_i$, $a_r^{(i)}, b_r^{(i)}, c_r^{(i)}$ ($r = 1, \cdots, \ell$, $i = 1, \cdots, n$) to be pairwise distinct with the following exceptions. We start the construction by representing each clause $E_j$ by a single point $p_j = (\alpha_j, \beta_j, \gamma_j)$. The coordinates of all these points will be pairwise distinct. We then choose some of the coordinates of points in the sets $S_i$ subject to the following requirement. The pair $(\alpha_j, \beta_j)$ will be associated with the literal $x_j$, $(\alpha_j, \gamma_j)$ with $y_j$, and $(\beta_j, \gamma_j)$ with $z_j$. If $u_i = x_j$ then we set $a_{4j-3}^{(i)} = \alpha_j$ and $b_{4j-3}^{(i)} = \beta_j$; if $\bar{u}_i = x_j$ then we set $a_{4j-2}^{(i)} = \alpha_j$ and $b_{4j-2}^{(i)} = \beta_j$. Similarly, if $u_i = y_j$ then $a_{4j-2}^{(i)} = \alpha_j$ and $c_{4j-1}^{(i)} = \gamma_j$; if $\bar{u}_i = y_j$ then $a_{4j-3}^{(i)} = \alpha_j$ and $c_{4j-2}^{(i)} = \gamma_j$. Finally, if $u_i = z_j$ then $b_{4j-2}^{(i)} = \beta_j$ and $c_{4j-2}^{(i)} = \gamma_j$; if $\bar{u}_i = z_j$ then $b_{4j-1}^{(i)} = \beta_j$ and $c_{4j-1}^{(i)} = \gamma_j$. It can be verified that no line parallel to one of the axes can cover more than one point in more than one of the sets $S_i$, but there exist such lines that cover two points in one of the $S_i$'s. Thus, any cover of the set $\{p_1, \cdots, p_m\} \cup \bigcup_i S_i$ requires $1.5\ell n$ lines. This number suffices if and only if there exists a satisfying assignment for $E_1 \wedge \cdots \wedge E_m$ since this is precisely the case where the points $p_1, \cdots, p_m$ can be covered by true or false lines of the variables. ∎

## References


[1] R. Bar-Yehuda and S. Even, "A linear-time approximation algorithm for the weighted vertex cover problem", *J. Algorithms* **2** (1981) 198-203.

[2] V. Chvátal, "A greedy heuristic for the set-covering problem", *Mathematics of Operations Research* **4** (1979) 233-235.

[3] L. R. Ford, Jr. and D. R. Fulkerson, *Flows in Networks*, Princeton University Press, Princeton, New Jersey, 1962.

[4] M. R. Garey and D. S. Johnson, *Computers and intractability: A Guide to the theory of NP-completeness*, W. H. Freeman, San Francisco, 1979.

[5] D. S. Hochbaum, "Approximation algorithms for the set covering and vertex cover problem", *SIAM J. Comput.* **11** (1982) 555-556.

[6] D. S. Hochbaum and W. Maass, "Approximation schemes for covering and packing problems in image processing and VLSI", *Journal of the Association for Computing Machinery* **32** (1985) 130-136.

[7] D. S. Johnson, "Approximation algorithms for combinatorial problems", *Journal of Computer and System Sciences* **9** (1974) 256-278.

[8] L. Lovász, "On the ratio of optimal integral and fractional covers", *Discrete Mathematics* **13** (1975) 383-390.

[9] N. Megiddo, "On the complexity of polyhedral separability", *Discrete & Computational Geometry* **3** (1988), 325-327.

[10] N. Megiddo and A. Tamir, "On the complexity of locating linear facilities in the plane", *Operations Research Letters* **1** (1982) 194-197.

[11] W. D. Renner, N. O. Pugh, D. B. Ross, R. E. Berg and D. C. Hall, "An algorithm for planning stereotactic brain implants", *International Journal of Radiation Oncology and Biological Physics* **13** (1987) 631-637.