SEPARATING POINTS BY AXIS-PARALLEL LINES

GRUIA CĂLINESCU

Computer Science, Illinois Institute of Technology
Chicago, IL 60616, USA

Email: calinesc@cs.iit.edu

ADRIAN DUMITRESCU

Computer Science, University of Wisconsin–Milwaukee
3200 N. Cramer Street, Milwaukee, WI 53211, USA

Email: ad@cs.uwm.edu

HOWARD KARLOFF

AT&T Labs–Research, 180 Park Ave.
Florham Park, NJ 07932, USA

Email: howard@research.att.com

PENG-JUN WAN

Computer Science, Illinois Institute of Technology
Chicago, IL 60616, USA

Email: wan@cs.iit.edu

Received (received date)
Revised (revised date)
Communicated by (Name)

We study the problem of separating n points in the plane, no two of which have the same x- or y-coordinate, using a minimum number of vertical and horizontal lines avoiding the points, so that each cell of the subdivision contains at most one point. Extending previous NP-hardness results due to Freimer *et al.* we prove that this problem and some variants of it are APX-hard. We give a 2-approximation algorithm for this problem, and a d-approximation algorithm for the d-dimensional variant, in which the points are to be separated using axis-parallel hyperplanes. To this end, we reduce the point separation problem to the rectangle stabbing problem studied by Gaur *et al.* Their approximation algorithm uses LP-rounding. We present an alternative LP-rounding procedure which also works for the rectangle stabbing problem. We show that the integrality ratio of the LP is exactly 2.

Keywords: Point separation; approximation algorithm; LP-rounding; integrality gap.

1. Introduction

Let P be a set of n points in the plane, no two of which have the same x- or y-coordinate. We consider the problem of finding a minimum set of axis-parallel lines that do not pass through any of the given points, such that each cell of the resulting subdivision contains at most one point. In other words, for each pair of points there is a line in our set which separates the two points. We refer to this problem as the *separation problem* SEPARATION. Its natural extension to higher dimensions, called the *multi-modal sensor allocation problem* in [11], asks for a minimum cardinality set of axis-parallel hyperplanes which separate n given points. It has applications to fault-tolerant multi-modal sensor fusion in the context of embedded sensor networks [11]. The problem appears to be closely related to other problems of separating points or hitting objects studied in the computational geometry literature [1], [3], [4], [8], [9], [10], [12].

The point separation problem appears to have been studied for the first time by Freimer, Mitchell and Piatko [5], under the name *point shattering problem*; they considered both the general case — when the points can be separated by arbitrary lines, and the special case — when only axis-parallel lines are used. They have shown that both variants are NP-hard, and have left the problem of obtaining good approximation algorithms as open for further research [5].

Our paper is organized as follows. In Section 2 we present two LP-based approximation algorithms with ratio 2 in the plane[a], respectively d in $\mathbb{R}^d$: the first is obtained by casting the separation problem as a special case of the rectangle stabbing problem [7], [8]. The second uses a different rounding procedure. We show that the second algorithm also works for the rectangle stabbing problem, with the same ratio, 2.

In Section 2.1, we show that, for any $\epsilon > 0$, there are examples in the plane having integrality ratio at least $2 - \epsilon$ for SEPARATION, and hence also for RECTANGLE STABBING[b]. Since the integrality ratio is 2, it means one cannot prove a constant approximation ratio less than 2 based only on the value of the linear program as a lower bound on the optimum value.

In Section 3, we show (under standard assumptions) that SEPARATION is in fact hard to approximate beyond a certain threshold (see Theorem 3).

A natural variant of the above point separation problem is a *colored* version: the points are colored, and one has to find a minimum set of axis-parallel lines, such that the set of points in each cell of the resulting subdivision, if nonempty, is monochromatic. Clearly having each point colored by a different color is equivalent to the original problem. Thus when the numbers of colors is part of the input this problem is also NP-hard. We prove that it remains so for any number k of colors, $k \geq 2$. This version also extends to higher dimensions, as the original problem does. Both our algorithms can be used to obtain a 2-approximate solution for the colored version in the plane, or d-approximate solutions for the colored version in $\mathbb{R}^d$.

[a] An approximation algorithm with ratio r outputs a separating set of lines of size at most $r \cdot OPT$, where OPT is the size of an optimal separating set.

[b] The integrality ratio (gap) of a minimization integer program is the supremum over instances of the ratio of the value of the integer program to the value of its linear program relaxation.

2. Algorithms for Separation

In this section we prove

Theorem 1. *There exists a 2-approximation algorithm for* SEPARATION.

Without loss of generality, we can restrict the set of vertical or horizontal separating lines to a set $\mathcal{L}$ of $2(n-1)$ *canonical lines*, one for each pair of horizontally consecutive points, and one for each pair of vertically consecutive points (say, at the average coordinate value of two consecutive points).

We first give two lower bounds on OPT, the size of an optimal solution. Consider the complete geometric graph $G = (V, E)$ whose vertex set is the set P of n points. We say that two edges of G are *independent* if there is no vertical or horizontal line that intersects both in their interior. Let I be a maximum independent set of edges of G. Then clearly $OPT \geq |I|$, since each edge of I requires a distinct separating line.

Write $l = OPT$. The maximum number of cells induced by l horizontal and vertical lines is attained when the lines are divided evenly into vertical and horizontal. Since each point requires a distinct cell of the arrangement of l lines, we have $(\lfloor l/2 \rfloor + 1)(\lceil l/2 \rceil + 1) \geq n$, which implies that for all sets of n points,

$$OPT \geq \lceil 2\sqrt{n} \rceil - 2.$$

In the *rectangle stabbing problem* [7], [8], we are given a set of (nondegenerate) axis-parallel rectangles in the plane, with the objective of stabbing all the rectangles with the minimum number of axis-parallel lines (a rectangle is said to be stabbed by line ℓ if ℓ intersects its interior). Gaur, Ibaraki and Krishnamurti have recently given a 2-approximation algorithm for this problem [7].

Let us first see how the separation problem can be cast as a rectangle stabbing problem. For each pair of points $u, v \in P$, consider the rectangle R_{uv} whose diagonal is uv. Then separating all the points in P is equivalent to stabbing all rectangles R_{uv}, with $u, v \in P$. Note also that it is enough to restrict ourselves to empty rectangles, i.e., those that do not contain other points of P: stabbing all empty rectangles R_{uv} guarantees that all rectangles are stabbed. However, in general this restriction may be not significant, as it is easy to construct examples with $\Omega(n^2)$ empty rectangles determined by the n points.

Let $\mathcal{R}$ be the collection of rectangles in the rectangle stabbing problem. A set $\mathcal{L}$ of canonical lines is selected first, as in the separation problem. The natural IP (integer program) with variables α_L, for $L \in \mathcal{L}$, is

$$\text{minimize} \sum_{L \in \mathcal{L}} \alpha_L$$

$$\text{subject to} \sum_{L \text{ stabs } R} \alpha_L \geq 1 \quad \forall R \in \mathcal{R} \tag{1}$$

$$\alpha_L \in \{0, 1\} \quad \forall L \in \mathcal{L}. \tag{2}$$

The linear programming relaxation of IP is obtained by replacing the constraints (2) by

$$\alpha_L \geq 0 \quad \forall L \in \mathcal{L}.$$

Denote by LP the value of the above linear program. The algorithm of Gaur *et al.* solves the linear program and classifies rectangles as horizontal or vertical (with ties broken arbitrarily), depending on whether

$$\sum_{\text{horizontal } L:\ L \text{ stabs } R} \alpha_L \geq \frac{1}{2} \qquad \text{or} \qquad \sum_{\text{vertical } L:\ L \text{ stabs } R} \alpha_L \geq \frac{1}{2}.$$

It then solves optimally the problem of stabbing the horizontal rectangles with vertical lines, by solving the corresponding linear programs LP_H and LP_V. The solutions of these two linear programs are integral, a property that follows from the total unimodularity of their constraint matrices. Putting together the two sets of lines results in a 2-approximation algorithm, using again the total unimodularity property. Instead of solving LP_H and LP_V, one can solve directly the corresponding stabbing problems using the greedy algorithm, since these become interval stabbing problems on the line.

The formulation of the integer and linear programs for the separation problem is analogous. The IP with variables α_L, for $L \in \mathcal{L}$, is

$$\text{minimize} \sum_{L \in \mathcal{L}} \alpha_L$$

$$\text{subject to} \quad \sum_{L \text{ separates } uv} \alpha_L \geq 1 \quad \forall u, v \in P, u \neq v, \tag{3}$$

$$\alpha_L \in \{0, 1\} \quad \forall L \in \mathcal{L}. \tag{4}$$

The linear programming relaxation of IP is obtained by replacing the constraints (4) by

$$\alpha_L \geq 0 \quad \forall L \in \mathcal{L}.$$

The 2-approximate solution is obtained in the same way.

We now provide a new, conceptually simpler, LP-based algorithm, that only solves the linear program above and directly rounds the solution. Sort the horizontal lines $L_1, L_2, \ldots, L_{n-1}$ in order of their y-coordinates. Pick line L_j if and only if the interval $\left(\sum_{i=1}^{j-1} \alpha_{L_i}, \sum_{i=1}^{j} \alpha_{L_i}\right]$ contains a multiple of 0.5. There are at most $2\sum_{i=1}^{n-1} \alpha_{L_i}$ multiples of 0.5 in the interval $(0, \sum_{i=1}^{n-1} \alpha_{L_i}]$ and therefore the number of horizontal lines picked does not exceed $2\sum_{i=1}^{n-1} \alpha_{L_i}$. Apply a similar procedure to the vertical lines $\bar{L}_1, \bar{L}_2, \ldots, \bar{L}_{n-1}$ sorted in order of their x-coordinates. Hence the number of lines picked cannot exceed twice the value of the LP.

Now we show that we obtain a valid integral solution. Let P and Q be two points and let i_P (i_Q, respectively) be the index in the sorted order of the first horizontal line after P (Q, respectively) with the convention that if P has the highest y-coordinate, then $i_P = n$. Similarly, we define j_P and j_Q in reference to vertical lines.

Assume $i_P < i_Q$ and $j_P < j_Q$—the other three cases are symmetric. Constraint (3) gives $\sum_{k=i_P}^{i_Q-1} \alpha_{L_k} + \sum_{k=j_P}^{j_Q-1} \alpha_{\bar{L}_k} \geq 1$, and therefore $\sum_{k=i_P}^{i_Q-1} \alpha_{L_k} \geq \frac{1}{2}$ or $\sum_{k=j_P}^{j_Q-1} \alpha_{\bar{L}_k} \geq \frac{1}{2}$. Assume the first inequality holds, the other case being symmetric. Then there is a multiple of 0.5 in the interval $\left(\sum_{k=1}^{i_P-1} \alpha_{L_k}, \sum_{k=1}^{i_Q-1} \alpha_{L_k}\right]$ and therefore one of the lines $L_{i_P}, L_{i_P+1}, \ldots, L_{i_Q-1}$ is selected by the algorithm and separates P and Q.

Since $LP \leq OPT$, the approximation ratio is at most 2. It is easy to see that this algorithm works for the rectangle stabbing problem as well, with the same ratio of 2.

We finally remark that both algorithms can be used to solve the colored version of the separation problem in the plane with the same ratio of 2: write constraints only for the set of bichromatic edges, i.e., those whose endpoints have different colors.

2.1. *Integrality Ratio*

The main result of this section is that the integrality ratio is exactly 2. As a warm-up we show (Lemma 1) an infinite sequence of simple examples in the plane having integrality ratio 3/2, for both the rectangle stabbing and the separation problem. It is enough to do this for Separation (as a special case of the rectangle stabbing problem).

Lemma 1. *The integrality ratio of the linear program is* $3/2$ *on a set of examples with arbitrarily large optimal values of the integer program.*

Proof. Consider the five-point configuration in Fig. 1 (left), that we call an X.

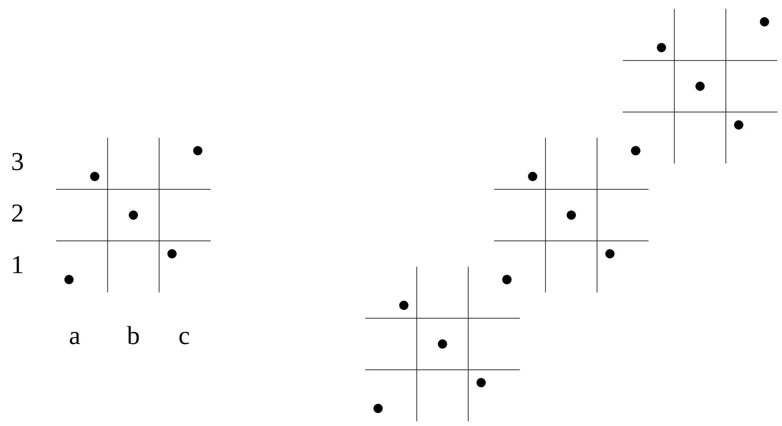

Fig. 1. A class of examples with integrality ratio 3/2.

The points can be fractionally separated with weights 1/2 on each of the four canonical lines shown in the figure. Thus $LP \leq 4/2 = 2$. Using the trivial lower bound (1) (or by inspection) gives $OPT \geq \lceil 2\sqrt{5} \rceil - 2 = 3$, and it is easy to see that this is tight.

By repeating the X diagonally k times, such that two adjacent X's share one point, we obtain a configuration with $4k+1$ points, as in Fig. 1 (right), for $k=3$. One can think of the points as being placed on an (infinite) chessboard. Observe that in each row or column of the board the points have increasing x- and y-coordinates. Again, the points can be fractionally separated with weights $1/2$ on each of the canonical lines shown in the figure. Thus $LP \leq 4k/2 = 2k$. To separate the points of each X requires three lines, and since the points have increasing x- and y-coordinates in each row or column, no line used to separate one X is of any help in separating other X's; thus $OPT \geq 3k$. It is easy to see that $3k$ lines are also enough, and the lemma follows. □

We now state and prove the main result of this section. Let Z^*_{LP} be the optimal value of the LP relaxation and Z^*_{IP} be the optimal value of the IP. Note that the proof of Theorem 1 gives $Z^*_{IP} \leq 2Z^*_{LP}$. We have

Theorem 2. *For every $\epsilon > 0$ there is an instance of* Separation *such that $Z^*_{IP} \geq (2-\epsilon)Z^*_{LP}$, where both Z^*_{LP} and Z^*_{IP} can be arbitrarily large.*

Proof. Let $\epsilon > 0$. Using a probabilistic argument we show that there are instances such that

$$Z^*_{LP} \leq (2 + \frac{1}{2}\epsilon)q \tag{5}$$

$$Z^*_{IP} \geq (4 - \frac{1}{2}\epsilon)q \tag{6}$$

for all sufficiently large integers q. As $4 - \frac{1}{2}\epsilon > (2-\epsilon)(2+\frac{1}{2}\epsilon)$, and since $Z^*_{IP} \leq 2Z^*_{LP}$, the two inequalities above imply the theorem.

We fix a parameter $k > 200/\epsilon$. Let $q \geq k$ be sufficiently large. Our instances have points in $[0, q+1) \times [0, q+1)$. There are $n = \lceil q^{5/4} \rceil$ pairs of points P_i and Q_i (so the number of points is $2n$, not n) obtained as follows: independently and uniformly at random pick x_{P_i} and y_{P_i} to be multiples of $1/k$ in $[0, q)$. Add $1/(2k) + 1/(3ki)$ to both x_{P_i} and y_{P_i}. Also, for every i, independently choose l_i uniformly at random from the set $\{1/k, 2/k, \ldots, (k-1)/k\}$ and set $x_{Q_i} = x_{P_i} + l_i$ and $y_{Q_i} = y_{P_i} + (1 - l_i)$. It is easy to see that no two points have the same x-coordinate and no two points have the same y-coordinate.

Now we construct the LP solution. Sort the $2n$ points by x-coordinate; they define exactly $2n-1$ canonical vertical lines. If two consecutive points in the sorted order above have x-coordinates $x' < x''$, the variable in the LP associated with vertical line L has value $\alpha_L = x'' - x'$. Similarly, using y-coordinates, we define a fraction α_L for every canonical horizontal line L.

In addition, we sometimes increase the fractions to give a valid LP solution as described below. Notice that the initial values of α_L ensure that the separating constraints (3) are satisfied for u, v if there is an $i \in \{1, 2, ..., n\}$ such that $u = P_i, v = Q_i$. However, for points from different pairs some constraints might be violated. Call a pair (i, j) of indexes, $1 \leq i < j \leq n$, *bad* if any of $||P_i - P_j||_1, ||P_i - Q_j||_1, ||Q_i - P_j||_1, ||Q_i - Q_j||_1$ is less than 1. For every bad pair of indexes (i, j), we increase α_L to 1 for three vertical lines. These three vertical lines are chosen from $\mathcal{L}$ such that any two horizontally consecutive points from $\{P_i, Q_i, P_j, Q_j\}$ are separated by one of the three lines, and we obtain a valid LP solution.

Thus the value of the LP solution is at most $2(q+1)+3b$, where b is the number of bad pairs of indexes. The probability of a pair being bad is at most $(5 \cdot 5)/q^2$, as the pair can be bad only if $||\lfloor x_{P_i} \rfloor - \lfloor x_{P_j} \rfloor|| \leq 2$ and $||\lfloor y_{P_i} \rfloor - \lfloor y_{P_j} \rfloor|| \leq 2$, and the random variables $\lfloor x_{P_i} \rfloor$, $\lfloor y_{P_i} \rfloor$, $\lfloor x_{P_j} \rfloor$, and $\lfloor y_{P_j} \rfloor$ are independent and uniformly distributed in the set $\{0, 1, \ldots, q-1\}$. Thus the expected value of b satisfies $E[b] \leq n^2 25/q^2 \leq 50q^{1/2}$. By Markov's inequality, with probability at least 1/2 we have $b \leq 100q^{1/2}$ and in this case we have $Z^*_{LP} \leq 2q + 2 + 3(100q^{1/2}) = 2q + 2 + 300\sqrt{q}$. For sufficiently large q we have $2 + 300\sqrt{q} \leq (\epsilon/2)q$, and then Equation (5) holds with probability at least 1/2.

Consider now the potential integral solutions (a *potential integral solution* is a set of horizontal and vertical lines) of value (i.e., size) less than $(4 - \frac{1}{2}\epsilon)q$. In fact, we consider only integral solutions required to separate only P_i from Q_i for every $i = 1, 2, \ldots, q$, and show that there is a configuration of points from our probability space such that all such potential integral solutions fail to separate at least one such pair (P_i, Q_i).

We can assume without loss of generality that the lines used by such integral solutions have coordinates $j \cdot (1/k)$, j being a positive integer, since any line with coordinate in the interval $((j-1) \cdot (1/k), j \cdot (1/k))$ can be replaced by one with coordinate $j \cdot (1/k)$ and all the previously separated pairs (P_i, Q_i) are still separated. Moreover we assume $j \leq kq$, as cutting at a coordinate larger than q is not needed, since the largest possible x_{P_i} or y_{P_i} is $(qk-1)/k + 1/(2k) + 1/(3k) < q$. There are in total at most $2kq$ such lines (both vertical and horizontal), and thus the total number of potential integer solutions of value at most $4q - 1$ is at most

$$\sum_{i=0}^{4q-1} \binom{2kq}{i} \leq 4q(2kq)^{4q-1} \leq (4kq)^{4q} = e^{4q\ln(4kq)}. \tag{7}$$

Let us fix now a potential integral solution of size $r \leq (4 - \frac{1}{2}\epsilon)q$. If $r < q$, we add more lines (at coordinates $j \cdot (1/k)$ for some j's) to the solution until $r \geq q$. Let r_1 be the number of vertical lines and r_2 be the number of horizontal lines used; $r = r_1 + r_2$. Together with the four lines, horizontal and vertical, at coordinates 0 and q, the lines of the solution divide the $[0, q] \times [0, q]$ square into rectangles. Let t be the number of rectangles and note that

$$t \leq (r_1 + 1)(r_2 + 1) \leq \frac{1}{4}(r+2)^2, \tag{8}$$

where the second inequality follows from $(r_1 + 1) + (r_2 + 1) = r + 2$.

Label the rectangles $R_1, R_2, ..., R_t$ in some order. For $i = 1, 2, \ldots, t$, let $\bar{w}_i$ be the (horizontal) width of R_i and $\bar{h}_i$ be its (vertical) height. The sum of the $\bar{w}_i$'s and $\bar{h}_i$'s is at least $q(r-2)$, as every line used (except those at coordinate q) by the integral solution has length q and contributes to either the $\bar{w}_i$'s of the rectangles above it, if the line is horizontal, or to the $\bar{h}_i$'s of the rectangles to its right, if the line is vertical.

Let $w_i = k\bar{w}_i \in \mathbb{N}$ and $h_i = k\bar{h}_i \in \mathbb{N}$. Thus

$$\sum_{i=1}^{t}(h_i + w_i) = k\sum_{i=1}^{t}(\bar{h}_i + \bar{w}_i) \geq k[q(r-2)]. \tag{9}$$

We will need later the following inequality:

$$kq(r-2) - (k+1)\frac{1}{4}(r+2)^2 \geq qr, \tag{10}$$

which we now prove for sufficiently large q based on the facts that $k > 200/\epsilon$ and $q \leq r \leq (4 - \frac{1}{2}\epsilon)q$. Indeed, (10) is equivalent to

$$q(kr - 2k - r) \geq (k+1)\frac{1}{4}(r+2)^2. \tag{11}$$

As $k > 4$ and $r \geq q$ can be assumed to be large, and using $r \leq (4 - \frac{1}{2}\epsilon)q$, it is enough to show that

$$r(kr - 2k - r) \geq (1 - \epsilon/8)(k+1)(r+2)^2. \tag{12}$$

On the left-hand side of (12) the coefficient of r^2 is $k - 1 > k + 1 - k\epsilon/8 - \epsilon/8 = (k+1)(1-\epsilon/8)$ (using $k > 200/\epsilon$), which is the coefficient of r^2 on the right-hand side. Thus for sufficiently large r, (12) holds, implying (10).

Claim 2.1. *The total number of possible placements for the pair (P_j, Q_j) in which both P_j and Q_j are in the same rectangle R_i for some $i \in \{1, 2, \ldots, t\}$ is at least $\sum_{i=1}^{t}(w_i + h_i - k - 1)$.*

Proof. It is enough to consider only rectangles satisfying $w_i + h_i > k+1$, and from now one we discuss only such rectangles. Every i such that $w_i + h_i > k+1$ can be classified into exactly one of these four sets:

A: Those with $1 < w_i \leq k$ and $1 < h_i \leq k$.
B: Those with $w_i > k$ and $h_i \leq k$.
C: Those with $w_i \leq k$ and $h_i > k$.
D: Those with $w_i > k$ and $h_i > k$.

Note that rectangles R_i with $w_i = 1$ and $w_i + h_i > k+1$ are in C and rectangles with $h_i = 1$ and $w_i + h_i > k+1$ are in B.

First fix a rectangle R_i from A. Recall that $w_i + h_i > k+1$, $1 < w_i \leq k$, and $1 < h_i \leq k$, i.e., the width $\bar{w}_i$ and height $\bar{h}_i$ of R_i are at most 1. (Informally, this is the "general" case.) We claim that at least $w_i + h_i - k - 1$ of the potential placements of (P_j, Q_j) result in both P_j's and Q_j's being in R_i. Indeed, place P_j in the $1/k \times 1/k$ square sharing the lower left corner with the lower left corner of R_i (such a placement exists for every j since $1/(2k) + 1/(3kj) < 1/k$). Then Q_j with $x_{Q_j} = x_{P_j} + (w_i - 1)/k$ and $y_{Q_j} = y_{P_j} + (k - (w_i - 1))/k$ (indeed $(x_{Q_j} - x_{P_j}) + (y_{Q_j} - y_{P_j}) = 1$) is also in R_i, as $h_i - 1 \geq k - (w_i - 1)$. Furthermore, P_j and Q_j are in the rectangle as well if they are both translated upward by a/k, where a is any integer in the set $\{1, 2, \ldots, w_i + h_i - k - 2\}$, as $a + [k - (w_i - 1)] \leq h_i - 1$. In total, we have found $w_i + h_i - k - 1$ placements.

Consider two consecutive vertical lines of the potential integral solution at horizontal distance greater than 1, and let G be the set of $i \in A \cup B \cup C \cup D$ such that R_i borders *both* these lines. All such rectangles R_i have the same $w_i > k$, which we denote by w. All such rectangles are in $B \cup D$. Let Z_G be the set of potential placements of (P_j, Q_j) with P_j and Q_j both inside some rectangle R_i with $i \in G$ and having $x_{Q_j} - x_{P_j} = (k-1)/k$.

We now prove that there are at least $\sum_{i \in G}(h_i - 1)(w_i - k + 1)$ such placements of (P_j, Q_j). Indeed, if the lower left corner of R_i has coordinates (x_i, y_i), then for all integers a, b satisfying $0 \leq a < w_i - k + 1$ and $0 \leq b < h_i - 1$, placing P_j in the $1/k \times 1/k$ square with lower left corner at $(x_i + a/k, y_i + b/k)$ results in P_j's and Q_j's being in the rectangle R_i, as the reader can verify by adding and comparing numbers. It follows that $|Z_G| \geq \sum_{i \in G}(h_i - 1)(w - k + 1)$.

Using $\sum_{i \in G} h_i = kq$, we have $|Z_G| \geq (w - k + 1)(kq - |G|)$. As this potential solution has $r \leq (4 - \frac{1}{2}\epsilon)q \leq 4q - 1$ horizontal lines, $|G| \leq 4q$, and therefore $|Z_G| \geq (w-k+1)q(k-4) = (w-k-1)q(k-4) + 2q(k-4)$. Since $k > 8$, we obtain

$$|Z_G| \geq 4q(w - k - 1) + kq \geq (w - k - 1)|G| + \sum_{i \in G} h_i = \sum_{i \in G}(w_i + h_i - k - 1). \quad (13)$$

Notice that each rectangle of $B \cup D$ appears for some two consecutive vertical lines at horizontal distance exceeding 1.

Consider now two consecutive horizontal lines of the potential integral solution at vertical distance greater than 1, and let G be the set of $i \in A \cup B \cup C \cup D$ such that R_i borders *both* these lines. All such rectangles R_i have the same $h_i > k$, which we denote by h. All such rectangles are in $C \cup D$. Let Z_G be the the set of potential placements of (P_j, Q_j) with P_j and Q_j both inside some rectangle R_i with $i \in G$ and having $y_{Q_j} - y_{P_j} = (k-1)/k$. Analogously to the above argument, we have

$$|Z_G| \geq \sum_{i \in G}(w_i + h_i - k - 1). \quad (14)$$

Notice that each rectangle of $C \cup D$ appears for some two consecutive horizontal lines at vertical distance exceeding 1.

For rectangles in D, the two sets of placements given above are disjoint: in the first set, $x_{Q_j} - x_{P_j} = (k-1)/k$, and in the second, $x_{Q_j} - x_{P_j} = 1/k$.

Since each rectangle of B appears exactly once for some consecutive vertical pair of lines, each rectangle of C appears exactly once for some consecutive horizontal pair of lines, and each rectangle of D appears exactly once in both, the total number of placements we have found is at least

$$\sum_{i \in A}(w_i + h_i - k - 1) +$$
$$\left[\sum_{i \in B}(w_i + h_i - k - 1) + \sum_{i \in C}(w_i + h_i - k - 1) + 2\sum_{i \in D}(w_i + h_i - k - 1)\right],$$

thus completing the proof of Claim 2.1 (since $\sum_{i \in D}(w_i + h_i - k - 1) \geq 0$). □

We continue with the proof of Theorem 2. From the previous claim, the total number of placements for the pair (P_j, Q_j), where both P_j and Q_j are in the same rectangle of the potential integral solution, is at least

$$\sum_{i=1}^{t}(w_i + h_i - k - 1) \geq kq(r-2) - (k+1)t$$
$$\geq kq(r-2) - (k+1)\left[\frac{1}{4}(r+2)^2\right] \geq qr \geq q^2,$$

where the first inequality follows from Equation (9), the second inequality follows from Equation (8), the third inequality from Equation (10), and the last inequality from our assumption that $r \geq q$.

As there are in total $(kq)(kq)(k-1) < k^3 q^2$ ways to select the coordinates of the pair (P_j, Q_j), we obtain that the probability that P_j, Q_j are separated by the given collection of lines, i.e., do not fall together in the same rectangle given by the potential integral solution, is at most $1 - q^2/(k^3 q^2) \leq e^{-1/k^3}$. Given that there are $n = \lceil q^{5/4} \rceil$ pairs, we obtain that the probability that this fixed integral solution is valid for a set of points, i.e., P_j is separated from Q_j for all j, is at most

$$(e^{-1/k^3})^n = e^{-\lceil q^{5/4} \rceil / k^3}. \tag{15}$$

Given that the total number of potential integral solutions is bounded by $e^{4q\ln(4kq)}$ (Equation (7)), for q so large that $q^{5/4}/k^3 > 4q\ln(4kq) + 1$, the probability that some pair is not separated by any potential integral solution of cost at most $(4 - \frac{1}{2}\epsilon)q$ is strictly bigger than $1 - 1/e$. Hence, the probability that (6) holds exceeds $1 - 1/e$. We showed earlier that (5) holds with probability at least $1/2$. Because $(1 - 1/e) + 1/2 > 1$, it follows that there is a placement of points satisfying both (5) and (6).

3. Hardness Results

In this section we prove:

Theorem 3. Separation *is APX-hard, that is: assuming $P \neq NP$, there is an absolute constant $\epsilon_S > 0$ such that no polynomial-time algorithm has approximation ratio at most $1 + \epsilon_S$.*

The decision version of Separation has been shown to be NP-complete [5]. Our APX-hardness reduction is similar to that in [5] and is inspired by the reduction from Proposition 6.2 of [8], which uses the satisfiability problem 3-Sat.

The *maximum* 3-*satisfiability* problem Max-3Sat is that of finding, in a 3CNF Boolean formula (in which each clause has exactly three literals), a truth value assignment which satisfies the maximum number of clauses. For each fixed k, define Max-3Sat(k) to be the restriction of Max-3Sat to Boolean formulae in which each variable occurs at most k times. Theorem 4 below is immediate from Theorems 29.7, 29.11, and Corollary 29.8 in [13].

Theorem 4. [13] *Assuming P $\neq$ NP, there is an absolute constant $\epsilon_M > 0$ such that no polynomial time algorithm for* Max-3Sat(5) *satisfies at least $(1-\epsilon_M)m$ clauses for every formula ϕ with m clauses which is satisfiable.*

To prove the approximation hardness stated in Theorem 3, we use the following reduction from Max-3Sat(k) to Separation. The input to 3-Sat is a Boolean formula ϕ in 3CNF form. Let ϕ have n variables and m clauses. The reduction constructs a set P_ϕ of $7n + 11m + 2$ points in the plane, no two of which have the same x- or y-coordinate. The construction is illustrated in Figure 2 for $\phi = (t + y + \overline{z})(\overline{x} + \overline{y} + z)(\overline{x} + \overline{y} + \overline{z})$. Here $n = 4$ and $m = 3$; the three clauses are denoted C_1, C_2, C_3.

There are three types of points: *variable* points, *clause* points and *control* points. The control points come in pairs, have increasing y-coordinates when scanned from left to right, and are denoted $q_1, \ldots, q_{4n+2m+2}$. For $1 \le i \le n+1$, the pair q_{2i-1}, q_{2i} "forces" a horizontal line (which is more useful than the vertical line separating the pair), and for $n + 2 \le i \le 2n + m + 1$, the pair q_{2i-1}, q_{2i} "forces" a vertical line. We call these lines *grid lines*, and we denote by h the lowest horizontal grid line. There are three variable points for each variable, and nine clause points for each clause. The nine points of each clause C are made up of six points that appear in the rows of the variables that appear in C (above the horizontal line h), and three points below h. We have a pair of points in the grid cell given by each variable-clause pair (x, C), where the variable x appears in C; thus six points per clause above line h. The three points of each variable require two separating lines. Every optimal solution can be assumed to use exactly one vertical line, as one vertical line also separates two control points and a second one is not needed. The choice of the higher (resp., lower) horizontal line corresponds to setting the variable to true (resp., false). If x appears unnegated in C, the pair of points is separated by the higher horizontal line, whereas if x appears negated in C, the pair of points is separated by the lower horizontal line.

The first $4n + 4$ control points form spine 1, and the $3m$ clause points below h form spine 2. The segments $q_{2i+1}q_{2i+2}$, for $i = 0, \ldots, 2n + m$, are called *control edges*. The segments $q_{2i}q_{2i+1}$, for $i = 1, \ldots, n$, and $i = n + 2, \ldots, 2n + 1$, are called *variable edges*. The segments $q_{2i}q_{2i+1}$, for $i = 2n + 2, \ldots, 2n + m$, are called *clause edges*. We denote by a, b, c, d the four canonical vertical lines which could be used

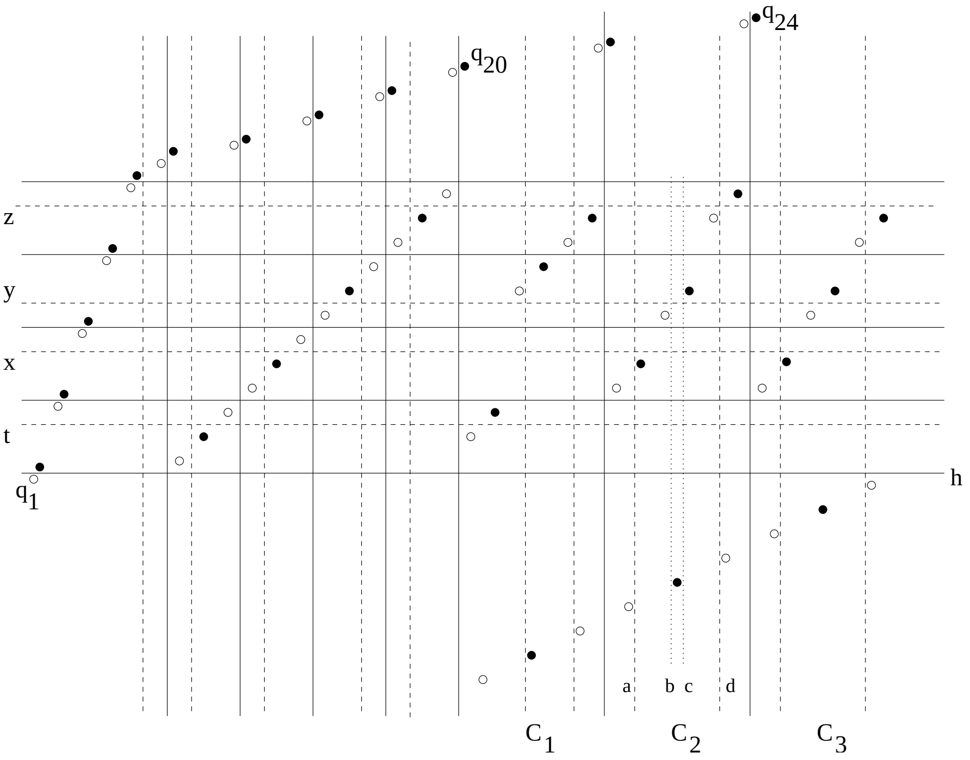

Fig. 2. The point set P_ϕ corresponding to $\phi = (t + y + \overline{z})(\overline{x} + \overline{y} + z)(\overline{x} + \overline{y} + \overline{z})$. The solution (i.e., set of separating lines) corresponding to the truth assignment $t = 1, x = 1, y = 0, z = 1$ is shown; the grid lines are solid, while the other separating lines are dashed. (The colors of the points only have meaning when discussing the colored version of the separation problem at the end of Section 3.)

to separate the three pairs and the triplet of a clause. They are shown in the figure for the clause C_2.

Clearly, constructing P_ϕ can be accomplished in polynomial time. We first determine the number of lines used when the input Boolean formula is satisfiable.

Claim 3.1. *If ϕ is satisfiable then P_ϕ can be separated using $4n + 3m + 2$ lines.*

Proof. Let τ be an assignment which satisfies ϕ. Use the $(n+1)+(n+1)+(m-1) = 2n+m+1$ grid lines to separate the pairs of control points q_{2i-1}, q_{2i}; add a vertical line to separate $q_{2n+2}q_{2n+3}$. We have thus used $2n+m+2$ lines so far. If a variable is set true by τ, use the higher of the two horizontal lines for that variable; otherwise use the lower horizontal line. For each variable, add a vertical line which separates the remaining pair of points. These lines also cut all variable edges. Thus, using $2n$ more lines, all variable points are separated; this yields $4n + m + 2$ lines so far.

Note now that for each clause, at least one of the three pairs of points above

h must be already separated, otherwise by construction, all literals in that clause would be set to false and the clause would not be satisfied, a contradiction. One can now check that the remaining two pairs of points above h and the three points below h can be separated using exactly two vertical lines per clause (at least two such lines are necessary to separate the three points below h). Overall, $4n+m+2+2m = 4n+3m+2$ lines have been used. □

Note that separating the points of spine 1 requires at least $4n+3$ lines. Similarly, at least $3m-1$ lines are necessary to separate the points of spine 2. Moreover, none of these lines can be shared, so at least $4n+3m+2$ lines are necessary to separate P_ϕ. Denote by $p = 4n+3m+2$ the exact number of lines needed to separate P_ϕ, when ϕ is satisfiable.

Assume that there exists a polynomial-time approximation algorithm for SEPARATION with performance ratio at most $1+\epsilon$ for some $\epsilon > 0$. The assumed algorithm gives a solution (set of lines) S having at most $(1+\epsilon)p$ lines. We first transform S to S' without any increase in cost, where S' is a solution that fulfills the following two conditions: (*i*) S' contains the grid lines, and (*ii*) S' uses exactly two vertical lines per clause (i.e., for separating its nine clause points).

To achieve (*i*), switch any of the vertical lines cutting the first $n+1$ control edges to horizontal ones, and any of the horizontal lines cutting the other $n+m$ control edges to vertical ones; note that the result is still a solution (i.e., separates the points). Similarly, switch any of the vertical lines cutting the first n variable edges to horizontal ones, and any of the horizontal lines cutting the remaining variable edges to vertical ones; note that the result is still a solution in which the triplet of each variable is separated by at least one horizontal and at least one vertical line.

We further transform the solution so as to satisfy (*ii*). We observe that at most ϵp clauses are separated vertically by three vertical lines (while each other uses exactly two vertical lines, the minimum required), otherwise one could separate P_ϕ with fewer than p lines, a contradiction. For each such clause, switch one of the three vertical lines to horizontal, so that the resulting three lines still separate the nine points of the clause. There are four cases, two of which are symmetric. If the three lines are a, b, c, switch b; if the three lines are a, b, d, switch b, etc.

We call S' the resulting solution. Note that at most ϵp variables are cut twice horizontally, as p lines are needed just to separate the points of the two spines, and a second horizontal line cutting a variable does not help with separating the points of the spines. We now construct a truth value assignment τ: for each variable, if it is cut horizontally by the higher line, set it to true, if it is cut horizontally by the lower line, set it to false, and if it is cut horizontally by two lines set it arbitrarily (say, to true). The at-most-ϵp variables that are cut twice horizontally appear in at most $5\epsilon p$ clauses (cf. the definition of MAX-3SAT(5)). Let C be any of the remaining clauses. We claim that τ makes C true. One of the three pairs of points of C above h must be separated by a horizontal line (otherwise only two vertical lines would separate the three pairs above h, a contradiction). By construction, the literal corresponding to the pair of points that is cut horizontally is true, hence C is true.

Therefore, the number of satisfied clauses is at least $m - 5\epsilon p$. Since we can assume that $m \geq n/3 + 1$, we have $m - 5\epsilon p \geq m - 5\epsilon(12m + 3m) = (1 - 75\epsilon)m$. Setting $\epsilon_M = 75\epsilon$, the result follows from Theorem 4. That is, we can take $\epsilon_S = \epsilon_M/75$, and the proof of Theorem 3 is complete.

We can use the same reduction to show that the separation problem with colored points is APX-hard. The points are colored as in Figure 2. The 2-coloring used has the property that all the edges specified in the above proof are bichromatic. We thus have

Corollary 1. *The separation problem in the plane with colored points is APX-hard.*

4. Remarks

4.1. *A Dual Problem*

Our covering LP for the separation problem suggests the following dual *edge packing problem.* Given a (non-necessarily planar) graph $G = (V, E)$ with a straight-line embedding in the plane, find a maximum set of independent edges of G, where two edges are said to be *independent* if they cannot be stabbed by a common vertical or horizontal line. A 4-approximation algorithm of Bar-Yehuda *et al.* [2] for finding a maximum independent set of rectangles in the plane—where two rectangles are said to be *independent* if they cannot be stabbed by a common vertical or horizontal line—gives a 4-approximation for this problem, by considering the set of rectangles $\{R_{uv} \mid uv \in E\}$. They use rounding of the dual of the LP, and thus their result combined with [7] shows that the optimal rectangle packing and the optimal rectangle stabbing are within a constant factor of each other.

Even the simple case when $E(G)$ is the edge set of a convex polygon P does not seem trivial. A 1/2-approximation algorithm is the following: divide P into its upper and lower chains, U and L, respectively. Find an optimal solution for both U and L, and choose the one with the larger number of edges. Finding an optimal solution for U (or L) amounts to finding a maximal independent set of intervals on a line, and it is thus solvable in polynomial time. It is easy to see that the result is at least half of the optimal.

4.2. *Higher Dimensions*

Following [7], it is now straightforward to observe that both our algorithms yield a d-approximation for the separation problem in $\mathbb{R}^d$. This holds for the colored version as well. One has to replace $1/2$ with $1/d$ in the corresponding places. In the first phase, after solving the linear program, edges are classified into d types, depending on the coordinate for which the sum of fractional weights is at least $1/d$. In the second phase, the first algorithm solves d linear programs (as in [7]), or solves d interval stabbing problems on the line (as in Section 2). The second algorithm cycles through all coordinates and, for each coordinate, goes through the hyperplanes in order, and chooses a hyperplane if and only if the running sum interval for that hyperplane includes a multiple of $1/d$.

4.3. *Concluding Remarks*

Several interesting questions regarding the separation problem in the plane remain, such as: Is it possible to improve the approximation ratio? Do special cases, e.g., points in convex position, admit better approximation ratios, or even exact solutions? One can potentially strengthen the LP by adding constraints. For example, a "stronger" LP could also require that each triplet of points is fractionally separated by at least 2. However, our probabilistic construction from Theorem 2 has also a ratio of at least $2-\epsilon$ for the stronger LP. In the proof, one must define the "bad" pairs of indexes to be those with any of $||P_i - P_j||_1, ||P_i - Q_j||_1, ||Q_i - P_j||_1, ||Q_i - Q_j||_1$ less than 2. This will increase only by a constant factor the expected number of bad pairs, and the proof with the adjusted constants can be used.

References

1. E. Arkin, F. Hurtado, J. S. B. Mitchell, C. Seara, and S. Skiena, Some separability problems in the plane, *Proceedings of the 16th European Workshop on Computational Geometry*, 2000, 51–54.
2. R. Bar-Yehuda, M. Halldórsson, J. Naor, H. Shachnai, and I. Shapira, Scheduling split intervals, *Proceedings of the 13th ACM-SIAM Symposium on Discrete Algorithms* (SODA '02), January 2002, 732–741.
3. E. Demaine, J. Erickson, F. Hurtado, J. Iacono, S. Langerman, H. Meijer, M. Overmars, and S. Whitesides, Separating point sets in polygonal environments, *Proceedings of the 20th Annual Symposium on Computational Geometry* (SOCG '04), June 2004, 10–16.
4. O. Devillers, F. Hurtado, M. Mora, and C. Seara, Separating several point sets in the plane, *Proceedings of the 13th Canadian Conference on Computational Geometry* (CCCG '01), 2001, 81–84.
5. R. Freimer, J.S.B. Mitchell, and C. Piatko, On the complexity of shattering using arrangements. Extended abstract in *Proceedings of the 2nd Canadian Conference on Computational Geometry* (CCCG '90), 1990, 218-222. Full paper available as Technical Report 91-1197, Department of Computer Science, Cornell University, April, 1991.
6. M. Garey and D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness.*, W. H. Freeman and Company, 1979.
7. D. R. Gaur, T. Ibaraki, and R. Krishnamurti, Constant ratio approximation algorithm for the rectangle stabbing problem and the rectilinear partitioning problem, *Journal of Algorithms* **43** (2002), 138–152.
8. R. Hassin and N. Megiddo, Approximation algorithms for hitting objects with straight lines, *Discrete Applied Mathematics* **30(1)** (1991), 29–42.
9. F. Hurtado, M. Mora, P. A. Ramos, and C. Seara, Separability by two lines and by nearly straight polygonal chains, *Discrete Applied Mathematics* **144** (2004), 110–122.
10. F. Hurtado, M. Noy, P. A. Ramos, and C. Seara, Separating objects in the plane with wedges and strips, *Discrete Applied Mathematics* **109** (2001), 109–138.
11. F. Koushanfar, S. Slijepcevic, M. Potkonjak, and A. Sangiovanni-Vincentelli, Error-tolerant multi-modal sensor fusion, *IEEE CAS Workshop on Wireless Communication and Networking*, Pasadena, 2002.

12. M. Segal, On piercing of axis-parallel rectangles and rings, *International Journal of Computational Geometry & Applications*, **9(3)** (1999), 219–233.
13. V. Vazirani, *Approximation Algorithms*, Springer-Verlag, New York, 2001.