Can Visibility Graphs Be Represented Compactly?*

Pankaj K. Agarwal† Noga Alon‡ Boris Aronov§ Subhash Suri¶

March 16, 1993

Abstract

We consider the problem of representing the visibility graph of line segments as a union of cliques and bipartite cliques. Given a graph G, a family $\mathcal{G} = \{G_1, G_2, \ldots, G_k\}$ is called a *clique cover* of G if (i) each G_i is a clique or a bipartite clique, and (ii) the union of G_i is G. The size of the clique cover $\mathcal{G}$ is defined as $\sum_{i=1}^{k} n_i$, where n_i is the number of vertices in G_i. Our main result is that there exist visibility graphs of n nonintersecting line segments in the plane whose smallest clique cover has size $\Omega(n^2/\log^2 n)$. An upper bound of $O(n^2/\log n)$ on the clique cover follows from a well-known result in extremal graph theory. On the other hand, we show that the visibility graph of a simple polygon always admits a clique cover of size $O(n \log^3 n)$, and that there are simple polygons whose visibility graphs require a clique cover of size $\Omega(n \log n)$.

*The work by the first author was supported by National Science Foundation Grant CCR-91-06514. The work by the second author was supported by a USA-Israeli BSF grant. The work by the third author was supported by National Science Foundation Grant CCR-92-11541.

†Department of Computer Science, Box 90129, Duke University, Durham, NC 27708-0129 USA

‡Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel and Bellcore, Morristown, NJ 07960, USA

§Computer Science Department, Polytechnic University, Six MetroTech Center, Brooklyn, NY 11201 USA

¶Bell Communications Research, 445 South Street, Morristown, NJ 07960 USA

1 Introduction

Given a set S of n nonintersecting line segments in the plane, its visibility graph $G(S)$ has the endpoints of S as vertices and pairs of mutually visible endpoints as edges. (Two points in the plane are visible, with respect to S, if the open line segment joining them does not intersect any segment of S.) The number of edges of $G(S)$ may range from linear to quadratic in n, as shown in Figure 1.

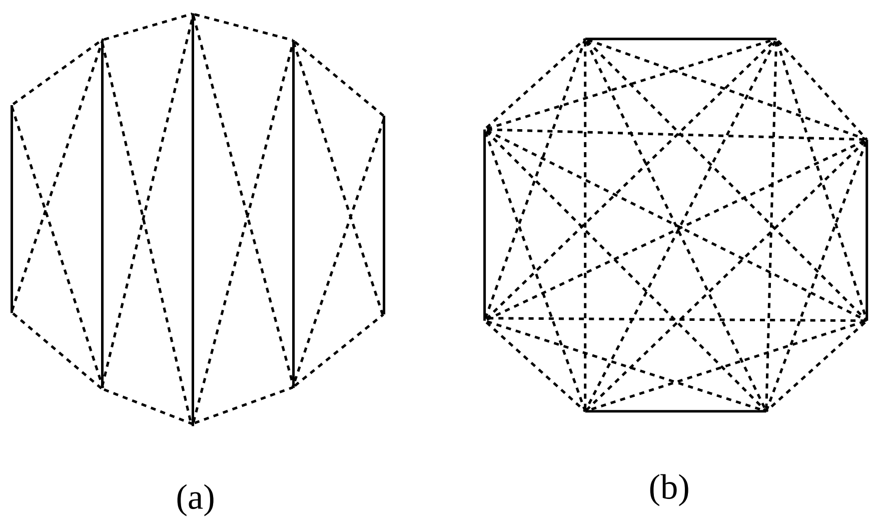

Figure 1: Two extreme visibility graphs. The example in (a) has a linear number of edges, and the one in (b) has a quadratic number of edges.

In this paper, we consider the problem of representing a visibility graph compactly. Our motivation stems from the example in Figure 1(b), where the visibility graph has a quadratic number of edges, but we can represent it implicitly by storing only the vertices. Similarly, a complete bipartite visibility graph can also be represented compactly by storing its two vertex classes. The idea of representing a visibility graph as a union of cliques or bipartite cliques has the advantage that each component is particularly simple. We discuss some algorithmic implications of our compact representation in Section 1.3. Let us first define our model of the compact representation more formally.

1.1 The Model

Let S be a set of line segments in the plane, where no two segments intersect except possibly at endpoints. Let $V(S)$ denote the set of endpoints in S. We say that two points are *mutually visible* if the open segment connecting them does not intersect the closure of any segments of S; however, it is convenient to assume that the endpoints of the same segment are visible to each other. This visibility relation induces a *visibility graph* $G = G(S)$ with vertices $V(S)$ and edges $E(S)$. Let $\mathcal{G} = \{G_1, G_2, \ldots, G_k\}$ be a family of subgraphs of G. We say that $\mathcal{G}$ is a *clique cover* of $G(S)$ if the following conditions hold:

1. Each G_i is a clique or a bipartite clique, and

2. $E(S) = E_1 \cup E_2 \cup \cdots \cup E_k$, where E_i denotes the set of edges in G_i.[1]

Since each G_i is a clique or a bipartite clique, it can be represented compactly in $O(|V_i|)$ space, where V_i is the vertex set of G_i. Let $f(S, \mathcal{G})$ denote the *size* of the clique cover $\mathcal{G}$:

$$f(S, \mathcal{G}) \;=\; \sum_{i=1}^{k} |V_i|,$$

and let $f(S)$ denote the size of a smallest clique cover of $G(S)$; that is, $f(S) = \min_{\mathcal{G}} f(S, \mathcal{G})$. Finally, define

$$f(n) \;=\; \max_{S} \; f(S),$$

where the maximum is taken over all sets S of n nonintersecting line segments in the plane. In order to be able to consider graphs of varying densities, we also define

$$g(n, e) \;=\; \max_{\substack{|S|=n \\ |E(S)|=e}} \; f(S).$$

We establish nearly tight upper and lower bounds on the quantities $f(n)$ and $g(n, e)$.

1.2 The Summary of Results

The main result of our paper is that the smallest clique cover of a visibility graph has size $\Omega(n^2/\log^2 n)$ in the worst case.[2] Thus,

$$f(n) \;=\; \Omega\left(\frac{n^2}{\log^2 n}\right). \tag{1}$$

Roughly speaking, we show that there are visibility graphs with a quadratic number of edges that do not contain a large bipartite clique. Thus, in the worst-case, the best representation of a visibility graph by cliques and bipartite cliques can save at most a factor of $O(\log^2 n)$ over an explicit representation. This result is also close to the best possible — any graph on n vertices has a clique cover of size $O(n^2/\log n)$ [16].

Our proof of the lower bound in Eq. (1) uses a non-constructive, probabilistic argument. By a constructive method, we can prove a slightly weaker result, namely, $f(n) = \Omega(n^{3/2})$. Specifically, we construct a set of n disjoint line segments whose visibility graph G has $e = O(n^{3/2})$ edges, G has a vertex-induced subgraph G' also with $\Theta(e)$ edges, and G' does not contain a $K_{2,2}$. This construction actually shows that $g(n, e) = \Theta(e)$ whenever $e = O(n^{3/2})$. Our probabilistic construction gives the general lower bound $g(n, e) = \Omega(n + e/\log^2 n)$ for

[1] In some applications, a proper partition of the edges may be desired; however, since we are primarily interested in a lower bound, we work with the weaker model allowing overlap.

[2] All logarithms in our paper are to the base 2.

all $e = O(n^2)$. These results imply that virtually no compaction is possible in the worst case, no matter how dense G is.

Finally, we establish a positive result for the visibility graph of segments forming the boundary of a simple polygon. We show that the visibility graph of a simple polygon on n vertices always admits a clique cover of size $O(n \log^3 n)$. We also show that the clique cover size is at least $\Omega(n \log n)$ for the visibility graph of certain simple polygons.

1.3 The Motivation

A compact representation of the visibility graph appears to be the key to deriving efficient algorithms for several visibility-related problems. We use the following three problems to illustrate this point.

1. [*Size of a visibility graph.*] Given a set S of nonintersecting line segments in the plane, count the number of edges in its visibility graph $G(S)$.
2. [*The biggest stick or diagonal.*] Given a bounded polygonal region with holes, compute the longest segment ("stick") that can be placed inside the region. The longest diagonal problem requires that both endpoints of the stick be vertices of the polygonal region.
3. [*Range-limited visibility graph.*] Given a set S of nonintersecting line segments in the plane, compute all edges of $G(S)$ of length at most one.

The problems 1–3 can be easily solved in $O(n^2)$ time, by explicitly computing the visibility graph [12, 17]. Whether they can be solved in $o(n^2)$ time remains an open problem. Interestingly enough, though, all three problems can be solved in substantially better time for a simple polygon. In particular, the number of edges in the visibility graph of a simple polygon can be computed in time $O(n \log^2 n)$ [1], the biggest stick can be computed in time $O(n^{8/5+\varepsilon})$ for any $\varepsilon > 0$ [1], the biggest diagonal can be computed in time $O(n \log^3 n)$ [3], and the range-limited visibility graph can be computed in time $O(n^{4/3+\varepsilon} + k)$, where k is the output size.

A common element of all these algorithms is that they implicitly depend on the fact that the visibility graph of a simple polygon admits a small clique cover, which also can be computed efficiently. Moreover, these algorithms can be generalized to a collection of (nonintersecting) segments *provided* that the visibility graph of the segments has a small, and efficiently computable, clique cover. Our main result, $f(n) = \Omega(n^2/\log^2 n)$, suggests that the existing algorithms or their variants are not likely to yield $O(n^{2-\delta})$-time algorithms for these problems, for any $\delta > 0$. Although we are unable to prove that a lower bound on $f(n)$ implies a similar lower bound for the time complexity of problems 1–3, we believe it to be the case.

Our paper contains four sections. In Section 2, we present our main result: a lower bound on the clique cover size of the visibility graph of segments in the plane. In Section 3, we give an algorithm for computing a small clique cover of the visibility graph of a polygon and a lower bound on the worst-case size of such a cover. We close in Section 4 with some discussion of the possible implication of our results.

2 A Lower Bound on Compact Representation

In this section, we prove a lower bound on the function $f(n)$. We give two different proofs; both proofs use essentially the same construction, however, one is constructive while the other is probabilistic.

2.1 The Construction

Our construction uses three sets A, B, C of points and segments, arranged along three vertical lines, as shown in Figure 2. A and C consist of uniformly spaced points along the lines $x = 1$ and $x = 3$, respectively. The middle set B has point-sized "holes" along the line $x = 2$. The holes are created by placing open line segments end-to-end along the line. Specifically, to create holes at points $b_1, b_2, \ldots, b_m$, where $b_j = (2, i_j)$, we use open segments $(b_{-\infty}, b_1), (b_1, b_2), (b_2, b_3), \ldots, (b_{m-1}, b_m), (b_m, b_\infty)$, where $b_{-\infty} = (2, -\infty)$ and $b_\infty = (2, \infty)$.

Remark: The construction outlined above is quite degenerate: it uses point-sized segments and holes; all segments are contained in three parallel lines. We will use this simpler form for our proofs since it best illustrates the main idea of the construction. At the end of this section, we discuss how to convert our construction into a non-degenerate one, in which all segments have finite lengths, every pair of segments is separated by a finite distance, and no three endpoints are collinear.

The sets A and C consist of uniformly spaced lattice points on the lines $x = 1$ and $x = 3$, respectively, with y-coordinates between 1 and $3n$. The middle set B consists of holes at some subset of the points $(2, i)$, where $n+1 \leq i \leq 2n$. Let p_y be the y-coordinate of a point p, and define $P_y = \{p_y \mid p \in P\}$ for a set of points P. We put $A_y = C_y = \{1, 2, \ldots, 3n\}$ and leave $B_y \subseteq \{n+1, n+2, \ldots, 2n\}$ unspecified. Define

$$\begin{aligned} A &= \{(1, i) \mid i \in A_y\}, \\ B &= \{(2, i) \mid i \in B_y\}, \\ C &= \{(3, i) \mid i \in C_y\}, \\ \\ S &= A \cup B \cup C. \end{aligned}$$

Slightly abusing the notation, we will let B denote both the set of holes as well as the set of segments that are used to create these holes. We will argue that, for an appropriate choice of the set B_y, the visibility graph of S has clique cover size $\Omega(n^2/\log^2 n)$. We begin with some definitions.

Definition: Given two sets of numbers X, Y and a number z, define $z+X = \{z+x \mid x \in X\}$, $2X = \{2x \mid x \in X\}$, and $X+Y = \{x+y \mid x \in X \text{ and } y \in Y\}$. We say that (X, Y) satisfies the *sum condition* (with respect to B_y) if $X + Y \subseteq 2B_y$.

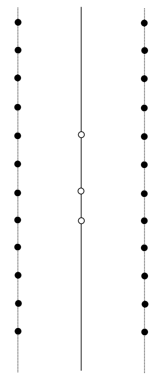

Figure 2: Sketch of the lower bound construction.

The following lemma is straightforward.

Lemma 2.1 *Two points a and c, where $a \in A$ and $c \in C$, are mutually visible if and only if $a_y + c_y \in 2B_y$. Two subsets $P \subseteq A$ and $Q \subseteq C$ induce a bipartite clique in $G(S)$ if and only if (P_y, Q_y) satisfies the sum condition with respect to B_y.*

Lemma 2.2 *In the visibility graph of S, $|E(G)| = \Theta(n\,|B|)$. The number of visibility graph edges between A and C is also $\Theta(n\,|B|)$.*

Proof: Let us first count the number of visible pairs (a, c), where $a \in A$ and $c \in C$. A pair (a, c) is visible *through* the hole at b, for $b \in B$, if and only if $a_y + c_y = 2b_y$. Since $a_y, c_y \in \{1, 2, \ldots, 3n\}$ and $b_y \in \{n+1, \ldots, 2n\}$, there are at least $2n$ and at most $3n$ solutions to the equation $a_y + c_y = 2b_y$, for a fixed b_y. Thus, each hole $b \in B$ creates $\Theta(n)$ visibility edges between A and C, and so the total number of visible pairs of the form (a, c) is $\Theta(n|B|)$.

Next, due to collinearity of A (resp. B and C), the number of visible pairs among points of A (resp. B and C) is linear. Finally,

$$\begin{aligned} |((A \cup C) \times B) \cap E(S)| &\leq |A \cup C| \cdot |B| \\ &= O(n\,|B|)\,. \end{aligned}$$

This completes the proof of the lemma. □

The main idea behind our lower bound argument is to show that there exist large sets B that preclude all but small bipartite cliques. In particular, we show that there are sets B, with $|B| = \Theta(n)$, such that there is no bipartite clique $K_{p,q}$ between A and C with $\min\{p, q\} \geq c \log^2 n$, where c is an absolute constant. In the remainder of this section, we concentrate primarily on the subgraph induced by $A \cup C$.

Definition: We say that a set B has the property $L(m,d)$, where $m = m(n)$ and $d = d(n)$, if the following conditions are satisfied:

1. $|B_y| = \Theta(m)$, with $B_y \subseteq \{n+1, n+2, \ldots, 2n\}$,
2. For every pair (P_y, Q_y), with $P \subseteq A$ and $Q \subseteq C$, that satisfies the sum condition with respect to B_y, we have $\min\{|P|, |Q|\} \leq d$.

The following theorem relates property $L(m,d)$ to the size of a clique cover.

Theorem 2.3 *The existence of a set B with property $L(m,d)$ implies that $f(n) = \Omega(nm/d)$.*

Proof: The collinearity of points in A and C implies that the visibility subgraph induced by $A \cup C$ cannot have a clique of size greater than four. Thus, it suffices to consider only the bipartite cliques. Let $G_1, \ldots G_k$ be a clique cover of $G(S)$, and let E_i denote the set of edges in G_i. If $G_i \equiv K_{p,q}$, we put $w(G_i) = (p+q)/pq$. Next, for an edge $e \in E(S)$, let $w(e) = \sum_{e \in E_i} w(G_i)$. Since $\min\{p, q\} \leq d$, we have $w(G_i) \geq 1/d$, and therefore $w(e) \geq 1/d$, for every $e \in E(S)$. Finally,

$$f(S) \;=\; \sum_{e \in E(S)} w(e) \;\geq\; \frac{|E(S)|}{d} \;=\; \Omega(mn/d),$$

where the last inequality follows from Lemma 2.2 and the fact that $|B| = \Theta(m)$. □

If a set B satisfies property $L(m,d)$, then trivially a subset $B' \subseteq B$ satisfies property $L(m', d)$, where $m' = |B'|$, giving the following corollary of the above theorem.

Corollary 2.4 *The existence of a set B with property $L(m,d)$ implies that $g(n,e) = \Omega(n + e/d)$, for any $n \leq e \leq mn$.*

The key remaining step in the proof is to show that there exists a set B with property $L(m,d)$ where m is large and d is small. The next two sections address this problem. In Section 2.2, we give a construction of a set B with property $L(\sqrt{n}, 2)$, which implies $f(n) = \Omega(n^{3/2})$, and $g(n,e) = \Theta(e)$ whenever $e = O(n^{3/2})$. In Section 2.3, we give a probabilistic proof for the existence of a set B with property $L(\Theta(n), O(\log^2 n))$, which gives a near-quadratic lower bound for $f(n)$.

2.2 A Constructive Lower Bound

We employ the following result of Erdős and Turán [10], proved independently by Singer [15]. For the sake of completeness, we include the proof given in [10].

Lemma 2.5 (Erdős-Turán [10]) *Given any integer $m > 0$, let*

$$T(m) = \{\sigma_1, \sigma_2, \ldots, \sigma_t\} \subseteq \{1, \ldots, m\}$$

be a largest-cardinality set such that $\sigma_i + \sigma_j \neq \sigma_{i'} + \sigma_{j'}$ whenever $\{i,j\} \neq \{i', j'\}$. Then, $t = \Theta(\sqrt{m})$.

Proof: It is clear that a larger set with this property does not exist: the numbers $|\sigma_j - \sigma_i|$ must be different, for all $1 \leq i < j \leq t$, and therefore $\binom{t}{2} < m$. We now exhibit a set of $t = \Omega(\sqrt{m})$ numbers with the required property. Pick a prime number p, where $1 < p \leq \lfloor\sqrt{m/2}\rfloor$. Given an integer i, for $1 \leq i < p$, define (i^2) to be the smallest positive integer u satisfying $i^2 \equiv u \pmod{p}$, where $1 \leq u < p$. Define a sequence of numbers

$$\sigma_i \;=\; 2pi + (i^2), \quad \text{for } 1 \leq i < p\,. \tag{2}$$

It is easily checked that $\sigma_i \leq 2p^2$, and $\sigma_i < \sigma_j$ for $i < j$. We claim that $\sigma_i + \sigma_j \neq \sigma_k + \sigma_l$ whenever $\{i,j\} \neq \{k,l\}$. To prove the claim, we observe that if $\sigma_i + \sigma_j = \sigma_k + \sigma_l$, then Eq. (2) implies

$$i + j = k + l \quad \text{and} \quad i^2 + j^2 \equiv k^2 + l^2 \pmod{p}. \tag{3}$$

Thus, $i - k = l - j$ and $i^2 - k^2 \equiv l^2 - j^2 \pmod{p}$. Since $\{i,j\} \neq \{k,l\}$, we have $i-k, l-j \neq 0$, which implies that $i + k \equiv j + l \pmod{p}$. But, then Eq. (3) implies that $i = l$ and $j = k$, which contradicts our assumption that $\{i,j\} \neq \{k,l\}$.

It is a well-known fact of number theory that there exists a prime number between m and $2m$, for any $m \geq 1$; see for instance [13]. Thus, we can always find a prime p with $\lfloor\sqrt{m/8}\rfloor < p \leq \lfloor\sqrt{m/2}\rfloor$. This completes the proof that $|T(m)| = \Theta(\sqrt{m})$. □

The preceding proof also gives an $O(m)$ time algorithm for constructing a set $T(m)$ with $|T(m)| = \Theta(\sqrt{m})$. In order to construct a set B with property $L(\sqrt{n}, 2)$, we pick B_y as the shifted set $T(n)$:

$$B_y = \{n + i \mid i \in T(n)\}.$$

Since the shift does not affect the sum property, the implication of Lemma 2.5 continues to hold. We show that the set B so obtained has property $L(\sqrt{n}, 2)$. The first condition, namely, $|B_y| = \Theta(\sqrt{n})$, is clearly satisfied. To prove the second condition, we assume for the sake of a contradiction that there exist distinct $a_1, a_2 \in A_y$ and $c_1, c_2 \in C_y$ such that $(a_i + c_j)/2 \in B_y$, for all $i, j \in \{1, 2\}$. Let $b_{ij} = (a_i + c_j)/2$, where $1 \leq i, j \leq 2$. Then, we have

$$b_{11} + b_{22} = b_{12} + b_{21} = (a_1 + a_2 + c_1 + c_2)/2.$$

By Lemma 2.5, we have either $b_{11} = b_{12}$ and $b_{22} = b_{21}$, or $b_{11} = b_{21}$ and $b_{22} = b_{12}$. In either case, we arrive at the conclusion that either $a_1 = a_2$ or $c_1 = c_2$, which contradicts the assumption that a_1, a_2 and c_1, c_2 are distinct. We have established the following key lemma.

Lemma 2.6 *One can construct a set B with property $L(\sqrt{n}, 2)$.*

Theorem 2.7 *One can construct a set S of n disjoint segments in the plane such that $|E(S)| = \Theta(n^{3/2})$ and the minimum clique cover size of $G(S)$ is also $\Theta(n^{3/2})$. This implies that $f(n) = \Omega(n^{3/2})$.*

Corollary 2.8 *$g(n,e) = \Theta(e)$ for $e = O(n^{3/2})$.*

2.3 A Probabilistic Lower Bound

We prove the existence of a set $B \subseteq \{n+1, \ldots, 2n\}$ with property $L(\Omega(n),\ \log^2 n)$, thus establishing the lower bound $f(n) = \Omega(n^2/\log^2 n)$. Our proof uses a probabilistic argument. To simplify the notation, we omit all floor and ceiling signs whenever they are not essential, and assume that n is sufficiently large.

Let $N = \{1, 2, \ldots, 3n\}$, and let p be a small absolute constant, to be fixed later. Let Z be a random subset of N obtained by choosing each element of N randomly and independently with probability p. The cardinality of the set $Z \cap \{n+1, \ldots, 2n\}$ is a Binomial random variable with parameters n and p. By the standard estimates for Binomial distributions (see for instance [2, Appendix A]), $|Z| \geq np/2$ with high probability; high probability means "with probability approaching 1 as n goes to infinity". Our proof hinges on the following crucial claim:

With high probability, there do not exist subsets $S, T \subseteq N$ such that

(i) $|S|, |T| \geq \log^2 n$, and

(ii) $S + T \subseteq 2Z$.

The claim implies that the set B obtained from $B_y = Z \cap \{n+1, \ldots, 2n\}$ has property $L(np/2,\ \log^2 n)$, with high probability. We now proceed to prove this claim.

Consider two arbitrary sets $S, T \subseteq N$, with $|S| = |T| = d$. We always have the following bounds for $|S+T|$:

$$2d - 1 \leq |S+T| \leq d^2.$$

Let N_m denote the number of ordered pairs (S,T) satisfying $|S+T| = m$, where $2d-1 \leq m \leq d^2$. Let E denote the expected number of pairs (S,T) with $S + T \subseteq 2Z$. Since the elements of Z are chosen independently, we have the following upper bound on E:

$$E \leq \sum_{m=2d-1}^{d^2} N_m p^m. \tag{4}$$

Our goal is to show that this expectation is $o(1)$ for large n, provided p is a sufficiently small constant. This is shown by proving that $|S+T|$ is sufficiently large for most of the pairs S, T that satisfy the above properties. The crucial lemma is the following.

Lemma 2.9 *For all m, $2d-1 \leq m \leq d^2$,*

$$N_m \leq \frac{1}{d!^2}(3n)^{2\sqrt{m}} \sum_{i=0}^{2\sqrt{m}} \binom{2d-2\sqrt{m}}{i} (3n)^i (2m)^{2d-2\sqrt{m}-i} \tag{5}$$

$$\leq \left(\frac{4em}{d}\right)^{2d} (3n)^{4\sqrt{m}}. \tag{6}$$

Proof: Clearly, N_m is the number of ordered pairs of *ordered* sets $S = \{s_1, \ldots, s_d\}$ and $T = \{t_1, \ldots, t_d\}$ of distinct elements of N satisfying $|S + T| = m$, divided by $(d!)^2$. To estimate this number, it is convenient to choose the members of S and T sequentially, alternating between S and T. For each i, $1 \le i \le d$, define $S_i = \{s_1, \ldots, s_i\}$ and $T_i = \{t_1, \ldots, t_i\}$. Put $S' = S_{\sqrt{m}}$ and $T' = T_{\sqrt{m}}$. For $i > \sqrt{m}$, we call s_i *enlarging* if

$$|(s_i + T') \cap (S_{i-1} + T_{i-1})| \ \le\ \sqrt{m}/2.$$

Similarly, t_i is called enlarging if $|(t_i + S') \cap (S_i + T_{i-1})| \le \sqrt{m}/2$. Observe that if s_i is enlarging then

$$|S_i + T_{i-1}| - |S_{i-1} + T_{i-1}| \ \ge\ \sqrt{m}/2$$

and an analogous statement holds for a enlarging t_i. Since $|S + T| = m$, there are at most $2\sqrt{m}$ enlarging elements in $S \cup T$. The proof depends on the observation that the number of ways to choose a non-enlarging s_i is at most $2|S_{i-1} + T_{i-1}| \le 2m$, for any fixed i; a similar statement holds for a non-enlarging t_i. This follows because if s_i is chosen uniformly at random among the $3n$ members of N then the expected value of $|(s_i + T') \cap (S_{i-1} + T_{i-1})|$ is at most

$$\frac{|T'| \cdot |S_{i-1} + T_{i-1}|}{3n} = \frac{\sqrt{m} \cdot |S_{i-1} + T_{i-1}|}{3n}.$$

By Markov's Inequality, the probability that the cardinality of this intersection exceeds $\sqrt{m}/2$ is smaller than $2 \cdot |S_{i-1} + T_{i-1}|/3n \ \le\ 2m/3n$. Thus, the number of ways to choose a non-enlarging s_i is at most $2m$.

To establish the bound in Eq. (5), observe that there are less than $(3n)^{2\sqrt{m}}$ choices for the (ordered) sets T' and S'. Among the remaining $2d - 2\sqrt{m}$ (ordered) elements of $S \cup T$ there are $i \ \le\ 2\sqrt{m}$ enlarging choices. There are $\binom{2d-2\sqrt{m}}{i}$ ways to choose the i steps when an enlarging element is picked, and each enlarging element can be chosen in at most $3n$ ways (trivially). Each of the $2d - 2\sqrt{m} - i$ non-enlarging elements can be chosen in at most $2m$ different ways, by the above observation. This completes the proof of Eq. (5). The bound in Eq. (6) follows from the observations that $1/d! \le (e/d)^d$, and that

$$\sum_{i=0}^{2\sqrt{m}} \binom{2d - 2\sqrt{m}}{i} (3n)^i (2m)^{2d-2\sqrt{m}-i} \ \le\ 2^{2d}(3n)^{2\sqrt{m}}(2m)^{2d}.$$

□

Lemma 2.10 *There exists an absolute positive constant p_0 such that for $p \le p_0$ the probability that $S + T \subseteq 2Z$ for some subsets $S, T \subset N$ is $o(1)$ (as n tends to infinity), where $|S| = |T| = d = \log^2 n$.*

Proof: By Lemma 2.9, the expectation E of the number of pairs (S, T) with $S + T \subset 2Z$ is at most

$$\sum_{m=2d-1}^{d^2} N_m p^m \ \le\ \sum_{m=2d-1}^{d^2} \left(\frac{4em}{d}\right)^{2d} (3n)^{4\sqrt{m}} p^m.$$

Put $R_m = \left(\frac{4em}{d}\right)^{2d}(3n)^{4\sqrt{m}}p^m$. Then

$$\log R_m = 2d\log\left(\frac{4em}{d}\right) + 4\sqrt{m}\log(3n) - m\log\left(\frac{1}{p}\right). \tag{7}$$

Without attempting to optimize the constants, we prove the corollary for p_0 defined by $\log(1/p_0) = 32e$. If $p \le p_0$, then $\log(1/p) \ge 32e$. For each $m \ge 2d-1 \ge d = \log^2 n$, we have

$$4\sqrt{m}\log(3n) \le 8m \tag{8}$$

and, since $\log x < x$ for $x \ge 4$,

$$2d\log\left(\frac{4em}{d}\right) \le 2d\frac{4em}{d} = 8em \tag{9}$$

Substituting Eqs. (8) and (9) in Eq. (7), we get $\log R_m \le -16em$ for each admissible m. Thus,

$$E \le \sum_{m=2d-1}^{d^2} R_m \le \sum_{m=2d-1}^{d^2} 2^{-16em} = o(1),$$

provided n is sufficiently large. Thus, the probability that there are S and T, with $|S| = |T| = d$ and $S + T \subset 2Z$, is $o(1)$. This completes the proof. □

Lemma 2.10 and Theorem 2.3 together imply the following theorem.

Theorem 2.11 *There is a set S of n disjoint line segments in the plane whose visibility graph has $\Theta(n^2)$ edges and the smallest clique cover of the visibility graph of S has size $\Omega(n^2/\log^2 n)$. Thus, $f(n) = \Omega(n^2/\log^2 n)$.*

Remark. The proof of Lemma 2.10 can be modified to show that there also exist sets with property $L(\Omega(n^{1-\delta}), O(1/\delta^2))$, for any fixed $0 < \delta < 1$. The modification sets $p = 1/n^\delta$ and $d = c/\delta^2$, where c is an appropriate constant independent of δ. Substituting these values in Eq. (7) shows that $\log R_m \le -c'\delta\log n$, for some constant $c' > 0$. Thus, the expected number of (S,T) pairs with $S + T \subseteq 2Z$ is $E \le \sum_{m=2d-1}^{d^2} 2^{-c'\delta\log n}$, which is $o(1)$. This gives the following corollary of Theorem 2.11.

Corollary 2.12 *$g(n,e)$ is $O(e)$ and $\Omega(n + e/\log^2 n)$, for any $e \ge n$. If $e = O(n^{2-\delta})$, for any constant $0 < \delta < 1$, then $g(n,e) = \Omega(e)$; the constant of proportionality depends on δ.*

In fact, we can prove the following theorem, which is a slightly stronger version of Lemma 2.10.

Theorem 2.13 *Let Z be a random subset of $N = \{1, 2, \ldots, 3n\}$ obtained by choosing each $a \in N$ randomly and independently with probability p (where p is* any *constant, $0 < p < 1$). Then, with high probability, there are no subsets S and T of N, $|S| = |T| = c(p)\log^2 n$, with $S + T \subset 2Z$.*

The proof of Theorem 2.13 depends on the following lemma, which can be proved by a simple greedy argument; we omit the proof for lack of space.

Lemma 2.14 *For every* $0.65 < \epsilon < 1$, *there is a* $\delta = \delta(\epsilon) > 0$ *such that, for any two subsets* $S, T \subset N$ *of size* d *and* $g \le \delta\sqrt{d}$, *there exist* $S' \subset S$ *and* $T' \subset T$ *with* $|S'| = |T'| = g$ *and* $|S' + T'| \ge (1-\epsilon)g^2$.

Proof: Our proof is by induction on g. Assuming that d is sufficiently large, the claim obviously holds for $g \le 10$. Inductively assume that the lemma holds for all $g \le g_0$, and consider $g = g_0 + 1$. By the induction hypothesis, there exist $S'' \subset S$ and $T'' \subset T$, with $|S''| = |T''| = g_0$, such that $|S'' + T''| \ge (1-\epsilon)g_0^2$.

We claim that there is an element $x \in S - S''$ such that

$$\left|(x + T'') - (S'' + T'')\right| \;\ge\; 2(1-\epsilon)g_0 + 1. \tag{10}$$

Indeed, assuming otherwise implies that, for all $x \in S - S''$,

$$\left|(x + T'') \cap (S'' + T'')\right| \;\ge\; g_0 - 2(1-\epsilon)g_0 - 1 \;\ge\; g_0(2\epsilon - 1.3),$$

since $g_0 \ge 10$. A simple counting argument shows that there exist a $y \in T''$ and a subset $\tilde{S} \subset S - S''$ of size $(2\epsilon - 1.3)(d - g_0)$ such that $y + \tilde{S} \subseteq S'' + T''$. This implies that

$$(2\epsilon - 1.3)(d - g_0) \;=\; |y + \tilde{S}| \;\le\; |S'' + T''| \;\le\; g_0^2\,. \tag{11}$$

But $(2\epsilon - 1.3)(d - g_0) > g_0^2$ if $g_0 \le \delta\sqrt{d}$, where $\delta = \delta(\epsilon)$ is an appropriate constant, thereby contradicting Eq. (11).

Thus, we can find an element $x \in S - S''$ that satisfies Eq. (10). By setting $S' = S'' \cup \{x\}$, we obtain

$$|S' + T''| \;\ge\; (1-\epsilon)g_0^2 + 2(1-\epsilon)g_0 + 1 \;\ge\; (1-\epsilon)(g_0+1)^2\,.$$

Since $|S' + T''|$ is already at least $(1-\epsilon)(g_0+1)^2$, we can choose an arbitrary element $x \in T - T''$ and set $T' = T'' \cup \{x\}$. This completes the proof of the lemma. □

Proof of Theorem 2.13: Let $\epsilon = 3/4$. If $S + T \subseteq 2Z$, for some $|S| = |T| = c\log^2 n$, then by Lemma 2.14 there exist $S' \subset S$ and $T' \subset T$, with $|S'| = |T'| = g = c'\log n$ and $|S' + T'| \ge g^2/4$, where $c' = \delta(3/4)\sqrt{c}$ is a constant depending only on δ and c. But the expected number of such pairs (S', T') is at most

$$(3n)^{2g}p^{g^2/4} \;=\; \exp\left(2c'\log^2 n + 2c'\log 3\log n - \frac{c'^2\log^2 n \cdot \log(1/p)}{4}\right)$$

$$(3n)^{2g}p^{g^2/4} = \;=\; \exp\left(\log^2 n\left(2c' + \frac{2c'\log 3}{\log n} - \frac{(c')^2\log(1/p)}{4}\right)\right)$$

which is $o(1)$ for any $p < 1$ provided that c (and thus c') is sufficiently large. Thus, with high probability, $2Z$ contains no such $S' + T'$, and hence no such $S + T$, completing the proof. □

Although the above proof is shorter and gives a slightly better estimate, we believe that the proof in Lemma 2.10 may eventually lead to an asymptotically better estimate.

Remark. The pseudo-random properties of Paley graphs (see [2] for the definition) suggest that the following *explicit* construction of a subset $Z \subset \{n+1, \ldots, 2n\}$ may satisfy property $L(\Omega(n), \log^{O(1)} n)$. Let q be the smallest prime larger than $2n$, and let Z be the set of all $i, n+1 \leq i \leq 2n$, that are quadratic residues modulo q. It is easy to see that $|Z| = (1+o(1))n/2$. By applying known estimates for character sums it can be shown that, for every $k \leq \log n/4$, if two subsets $S, T \subseteq \{1, \ldots, 3n\}$, with $|S| = k$, satisfy $S + T \subseteq 2Z$, then $|T| \leq (1+o(1))q/2^k = O(n/2^k)$. One can also show that Z satisfies property $L(\Omega(n), O(\sqrt{n}))$ (for instance, see [2], pp. 116–119). However, it is not known if Z satisfies property $L(\Omega(n), \log^{O(1)} n)$, although it seems plausible (but difficult, as a proof would have some far reaching number-theoretic consequences).

2.4 Removing Degeneracies from the Construction

A simple modification of our construction turns it into a non-degenerate configuration. In the modified version of our construction, the segments have finite lengths, they are separated by finite distances, and no three endpoints lie on a line. We first replace the open segments of B by a collection of slightly shorter closed segments separated by tiny but finite-length gaps; the segments still lie along the line $x = 2$. Clearly, this does not affect the visibility between A and C, it at most doubles the number of edges between B and $A \cup C$, and it introduces $\Theta(|B|)$ visibility edges between endpoints of B. Next, we replace the points of A by tiny horizontal segments whose left endpoints lie on a concave function of the y coordinate and whose right endpoints lie on a convex function, as shown in Figure 3. We apply a similar transformation to C. We then tilt each segment of A and C slightly in order to avoid horizontal collinearities; this is done in such a way that one endpoint of a segment does not block the visibility of the other endpoint. A similar tilt is applied also to the segments of B, ensuring that the number of visibility edges among endpoints of B does not exceed $O(|B|)$.

The final construction has at most twice as many endpoints and at most four times as many visibility edges as the original one. It is easily checked that it does not contain cliques or bipartite cliques that are much larger than those contained in the original construction. The use of segments instead of points necessarily means that certain bipartite cliques inadmissible in the original construction are possible in the modified construction; however, this only affect the constants in our theorems, not the asymptotic form of their expressions. For instance, while $K_{2,2}$ is not possible in the original (degenerate) construction used for the proof of Theorem 2.7, in the modified construction, we can only exclude $K_{4,4}$.

3 The Visibility Graph of a Simple Polygon

Consider a simple polygon P on n vertices, and let S denote the set of segments forming its boundary. We show in this section that $G(S)$ admits a compact representation; specifically,

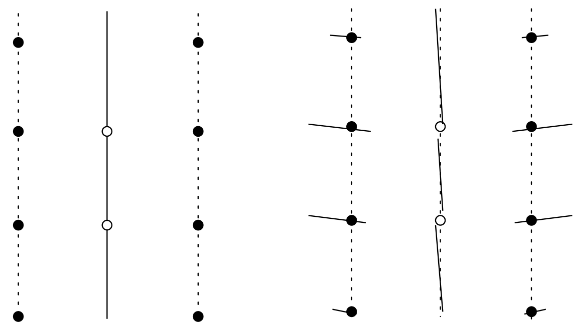

Figure 3: Removing degeneracies.

we produce a $O(n \log^3 n)$ size clique cover of $G(S)$. We also show that there are polygons whose visibility graph requires a clique cover of size $\Omega(n \log n)$.

3.1 An upper bound

In this section, we describe an algorithm for constructing a small clique cover of the visibility graph $G(S)$. Let $CH(P)$ denote the convex hull of P. The closure of $CH(P) \setminus P$ consists of a collection of simple polygons with disjoint interiors, called *pockets*. Each edge in $G(S)$ lies inside $CH(P)$ and does not cross any segment of S. Therefore it lies either in P or in one of the pockets of P. We will present an algorithm to compute a clique cover of the edges of $G(S)$ that lie inside P. A clique cover of other edges of $G(S)$ can be computed by repeating the same procedure for each pocket of P. Abusing the notation slightly, we will use $G(P)$ to denote the set of edges in $G(S)$ that lie inside P. Our construction is based on a divide-and-conquer approach. We partition P into two subpolygons P_1, P_2 by a diagonal e, such that each of the subpolygons has at most $2n/3$ vertices [4]. The edges of $G(P)$ can be partitioned into three subsets:

(i) E_{11}: an edge of $G(P)$ is in E_{11} if both of its endpoints lie in P_1,

(ii) E_{22}: an edge of $G(P)$ is in E_{22} if both of its endpoints lie in P_2,

(iii) E_{12}: an edge of $G(P)$ is in E_{12} if one of its endpoints lies in P_1 and the other in P_2.

We recursively compute clique covers of E_{11} and E_{22}. In the following, we describe a procedure for computing a clique cover for E_{12}.

Without loss of generality assume that e lies on the y-axis, and that the right (resp. left) side of e lies in P_1 (resp. P_2). Let ρ be a rightward directed ray emanating from e. Using a standard duality transformation, we can map the line supporting ρ to a point ρ^*. We will refer to the point ρ^* as the *dual* of ρ. We define a planar map M_1 in the dual plane as follows. Each face of M_1 is the set of points dual to the rays emanating from e and hitting first (the interior of) some fixed edge a of P_1 (i.e., the portion of ρ between e and a avoids the exterior of P_1). Every edge γ of M_1 is the locus of points dual to the rays that either hit a fixed vertex v of P_1, or touch a vertex v of P_1 before hitting an edge a of P_1. Let $v(\gamma)$ denote the vertex of P_1 that the rays corresponding to points on the edge γ intersect before crossing the boundary of P_1. By considering leftward-directed rays, define a similar map M_2 for P_2. By a result of Chazelle and Guibas [8], each M_i is a convex planar subdivision having $O(n)$ faces, edges, and vertices. Let Γ_1, Γ_2 denote the set of edges in M_1 and in M_2, respectively.

The intersection point of an edge $\gamma_1 \in \Gamma_1$ and an edge $\gamma_2 \in \Gamma_2$ is the dual of the line passing through $v(\gamma_1)$ and $v(\gamma_2)$. By construction, the edges γ_1 and γ_2 intersect if and only if the interior of the segment $\overline{v(\gamma_1)v(\gamma_2)}$ does not intersect the boundary of P, i.e., if and only if $(v(\gamma_1), v(\gamma_2))$ is a visibility edge of E_{12}. The problem of finding a small clique cover of E_{12} thus reduces to finding a small clique cover of the intersection graph G^* of $\Gamma_1 \cup \Gamma_2$ (i.e, the vertices of G^* are the segments of Γ_1 and Γ_2, and (γ_1, γ_2) is an edge in G^* if γ_1 and γ_2 intersect). Chazelle et al. [7] have presented an algorithm that can compute a clique cover of G^* of size $O(n \log^2 n)$. This immediately gives a clique cover of E_{12} of size $O(n \log^2 n)$. Let $S(n)$ denote the minimum clique cover size for the visibility graph of any simple polygon on n vertices. Then, the preceding discussion has shown that

$$S(n) \leq S(n_1) + S(n_2) + O(n \log^2 n),$$

where $n_1 + n_2 = n$ and $n_1, n_2 \leq 2n/3$. The solution to this recurrence is $S(n) = O(n \log^3 n)$. We apply the above procedure to all pockets of P, obtaining a clique cover of the entire visibility graph. Since the total number of vertices over all pockets is at most $2n$, we have established the following theorem.

Theorem 3.1 *Let S be a set of line segments forming the boundary of a simple polygon in the plane. Then, $f(S) = O(n \log^3 n)$.*

3.2 A lower bound

We (constructively) prove that there are simple polygons on n vertices whose visibility graphs require clique covers of size $\Omega(n \log n)$. The combinatorial lemma needed here follows from a result of Katona and Szemerédi [14], and our proof below applies their approach. Our lower bound construction uses a polygon P on $4n$ vertices, whose vertices are labeled $a_1, u_1, v_1, b_1, a_2, u_2, \ldots v_n, b_n$, in a counterclockwise order around the boundary. Let C_1, C_2, C_3 be three concentric circles of radii $1-\varepsilon, 1, 1+\varepsilon$, respectively, where ε is a sufficiently small positive number. In the polygon, the vertices $u_1, \ldots, u_n$ lie on circle C_1, the vertices $a_1, b_1, \ldots, a_n, b_n$ lie on circle C_2, and the vertices $v_1, \ldots, v_n$ lie on C_3, as shown in Figure 4.

Each 4-tuple a_i, u_i, b_i, v_i forms a sufficiently small convex quadrilateral so that the following conditions are satisfied:

1. b_i is not visible from a_i,

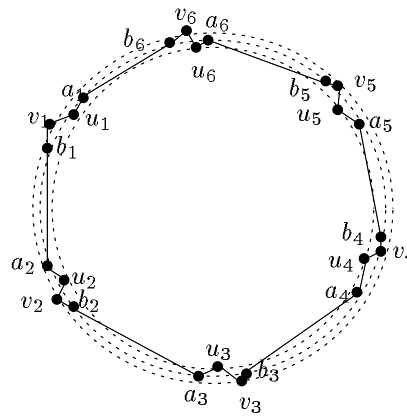

Figure 4: Lower bound construction

2. the line through a_i and u_i separates b_i, v_i from all other vertices of P, and a_i is visible to all b_j, $j \neq i$, and
3. the line through b_i and u_i separates a_i, v_i from all other vertices of P, and b_i is visible to all a_j, $j \neq i$.

Let S denote the set of edges of the polygon P. Let H denote the bipartite graph with vertices $\{a_1, \ldots, a_n\} \cup \{b_1, \ldots, b_n\}$ and edges $\{\,(a_i, b_j) \mid 1 \leq i \neq j \leq n\,\}$. Observe that H is the complete bipartite graph minus the matching $\{(a_i, b_i) \mid 1 \leq i \leq n\}$.

Lemma 3.2 $f(S) = \Omega(n \log n)$.

Proof: It is easily seen that $f(S)$ is at least as large as the size of a smallest clique cover of H. Indeed let $\mathcal{G}$ be a clique cover of $G(S)$. By deleting all u_i's and v_i's from each clique of $\mathcal{G}$, we obtain a clique cover of H. In the following, we prove a lower bound on the size of the smallest clique cover of H.

Consider a collection of subgraphs that covers H. Since H is bipartite, every induced subgraph of H is also a bipartite graph. For each i, $1 \leq i \leq n$, let X_i denote the collection of subgraphs in our cover that contain a_i and let Y_i be the collection of the subgraphs that contain b_i. Observe that the size of the cover is precisely $\sum_{i=1}^{n}(|X_i| + |Y_i|)$. Any subgraph in the cover of H cannot contain both a_i and b_i, because (a_i, b_i) is not an edge in H. Consequently, X_i and Y_i are disjoint, for every i. Let us choose for each subgraph in our collection, randomly and independently, a color 0 or 1 with probability 1/2. Let E_i be the event that all the members of X_i received color 0 and all those in Y_i color 1. Then the probability of E_i is $2^{-(|X_i|+|Y_i|)}$. Also, the events E_i are pairwise disjoint. Indeed, assuming otherwise implies that there is coloring so that all the subgraphs in X_i are colored 0 and all those in Y_j are colored 1 for some $i \neq j$. But that means that X_i and Y_j are disjoint, which is false, as both of them contain the subgraph containing the edge (a_i, b_j). Therefore, the sum of the probabilities of the events E_i is at most 1, i.e.,

$$\sum_{i=1}^{n} \frac{1}{2^{|X_i|+|Y_i|}} \leq 1.$$

By the arithmetic-geometric inequality, the left-hand side divided by n is at least the n-th root of the product

$$\prod_{i=1}^{n} \frac{1}{2^{|X_i|+|Y_i|}} = 1/2^s,$$

where s denotes here the size of the cover. Thus $2^s \geq n^n$, implying the desired result. □

4 Discussion and Open Problems

We have considered the problem of representing the visibility graph of a set of nonintersecting line segments by cliques and bipartite cliques. We showed that there are families of n segments whose visibility graphs require clique covers of size $\Omega(n^2/\log^2 n)$ (Theorem 2.11). On the other hand, the visibility graph of a simple polygon can always be represented by a clique cover of size $O(n \log^3 n)$. Our investigation is motivated by the observation that the existing efficient algorithms for several visibility-related problems depend on the cover size of the visibility graph. We conjecture that our lower bound of $\Omega(n^2/\log^2 n)$ on the size of clique cover implies a similar lower bound on the time complexity of solving the problems 1–3 mentioned in the introduction.

The problems 1–3 stated in the introduction, and several other visibility-related problems, are instances of the following abstract problem. Let S be a set of n nonintersecting line segments in the plane, let $V(S)$ denote the endpoints of the segments in S, and let $E(S)$ denote the edges of the visibility graph $G(S)$. Consider a commutative semigroup $(C, +)$ and a weight function w from pairs of endpoints in $V(S)$ to C; that is, $w : V(S) \times V(S) \rightarrow C$. Consider the problem of computing the total weight on the edges of $E(S)$:

$$W(S) = \sum_{(p,q) \in E(S)} w(p, q)\,. \tag{12}$$

In this setting, for instance, the biggest diagonal problem can be formulated by taking the semigroup $(\Re, \max)$ and the Euclidean weight function; that is, $w(p, q)$ is the Euclidean distance between p and q. Other problems have similar formulations.

We believe that one can define a model for visibility-type problems along the lines of the *semigroup* model of computation used by Fredman [11] and Chazelle [5, 6, 9], which has been used successfully to prove lower bounds on range searching problems. In particular, one needs to formalize the cost of computing the weight W. It seems reasonable that, in the absence of additional assumptions, computing $W(T)$ for an arbitrary subset $T \subseteq S$ would require at least $|(V(T) \times V(T)) \cap E(S)|$ operations, that is, the time proportional to the size of the visibility graph induced by T. On the other hand, if the visibility graph induced by T is a clique or a bipartite clique, then the weight $W(T)$ can be computed with $O(|T|)$ operations. We would like to argue that, in some sense, these two extremes are the only cases, and that the cost of computing $W(T)$ is at least $\Omega(|T|)$ even when the graph induced by T is a clique or a bipartite clique. In that case, the results of this paper imply an almost quadratic lower bound for the abstract problem of computing $W(S)$. We leave it as an open problem to prove or disprove this claim.

Acknowledgment We would like to thank J. Pach and J. Spencer for helpful comments that led to an improved estimate in Theorem 2.11.

References

[1] P. Agarwal, M. Sharir, and S. Toledo, Applications of parametric searching in geometric optimization, to appear in *J. Algorithms*, 1993.

[2] N. Alon and J. H. Spencer, *The Probabilistic Method*, John Wiley and Sons, Inc., New York, 1991.

[3] A. Aggarwal and S. Suri, The biggest diagonal in a simple polygon, *Inf. Proc. Letters* 35 (1990), 13–18.

[4] B. Chazelle, A polygon cutting theorem, *Proc. 23rd IEEE Symp. Foundations of Computer Science*, 1982, pp. 339–349.

[5] B. Chazelle, Lower bounds for the orthogonal range searching: II. The arithmetic model, *J. ACM* 37 (1990), 439–463.

[6] B. Chazelle, Lower bounds on the complexity of polytope range searching, *J. Amer. Math. Soc.*, 2 (1989), 637–666.

[7] B. Chazelle, H. Edelsbrunner, L. Guibas and M. Sharir, Algorithms for bichromatic line segment problems and polyhedral terrains, to appear in *Algorithmica.*

[8] B. Chazelle and L. Guibas, Visibility and intersection problems in plane geometry, *Discrete Comput. Geom.* 4 (1989), 551–589.

[9] B. Chazelle and B. Rosenberg, The complexity of computing partial sums off-line, *Int. J. Computational Geometry and Applications* 1 (1991), 33–46.

[10] P. Erdős and P. Turán, On a problem of Sidon in additive number theory, and on some related problems, *Proc. London Math. Soc.* 16 (1941), 212–215.

[11] M. L. Fredman, A lower bound on the complexity of orthogonal range queries, *J. ACM* 28 (1981), 696–705.

[12] S. Ghosh and D. Mount, An output-sensitive algorithm for computing visibility graphs, *SIAM J. Comp.* 20 (1991), 888–910.

[13] G. Hardy and E. Wright, *An Introduction to the Theory of Numbers*, Oxford University Press, London, 1959.

[14] G. Katona and E. Szemerédi, On a problem in a graph theory, *Studia Scientiarum Mathematicarum Hungarica*, 2 (1967), 23–28.

[15] J. Singer, A theorem in finite projective geometry and some applications to number theory, *Trans. American Math. Soc.* 43 (1938), 377-385.

[16] Z. Tuza, Covering of graphs by complete bipartite subgraphs: complexity of $0-1$ matrices, *Combinatorica* 4 (1984), 111-116.

[17] E. Welzl, Constructing the visibility graph for n line segments in $O(n^2)$ time, *Inf. Proc. Letters* 20 (1985), 167–171.