Guarding galleries in which every point can see a large area

Gil Kalai

Jiří Matoušek*

Department of Applied Mathematics
Charles University
Malostranské nám. 25, 118 00 Praha 1
Czech Republic

Abstract

We prove that if X is a compact simply connected set in the plane of Lebesgue measure 1, such that any point $x \in X$ sees a part of X of measure at least ε, then one can choose a set G of at most $const\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}$ points in X such that any point of X is seen by some point of G. More generally, if for any k points in X there is a point seeing at least 3 of them, then all points of X can be seen from at most $O(k^{-???})$ points.

1 Introduction

A *gallery* is a compact set in the plane, X. A point $x \in X$ *sees* a point $y \in X$ if the segment xy is fully contained in X. A subset $G \subseteq X$ *guards* X if each point of X is seen by at least one point of G. For a point $x \in X$, denote by $V(x)$ the *visible region* of x in X, that is, the set of all points $y \in X$ seen by X.

Theorem 1 *Let X be a simply connected gallery of Lebesgue measure $\lambda^2(X) = 1$, and let $\varepsilon > 0$ be a real number such that $\lambda^2(V(x)) \geq \varepsilon$ for all $x \in X$. Then X can be guarded by at most const $\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}$ points. More generally, if X has h holes, and $\lambda^2(V(x)) \geq \varepsilon$ for all $x \in X$, then X can be guarded by $C(h)\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}$ points.*

This answers a problem of Raghavan [?], who asked whether a gallery satisfying the assumptions of the first part of the theorem can be guarded by $f(\varepsilon)$ points, for some function f. The proof uses concepts and results of Vapnik and Chervonenkis [11] and Haussler and Welzl [4] (VC-dimension and ε-nets). For a set with holes (i.e. for $h > 0$), we use a Ramsey-type argument which was suggested by Nešetřil in a slightly different context. The proof is also related to the so-called *visibility graphs*; see a remark in section 2.

The following is a more general result for simply connected art galleries:

Theorem 2 *Let X be a simply connected gallery and k an integer such that among any k points in X, there are some 3 which can be seen from a single point. Then X can be guarded by at most $k^{???}$ points.*

The proof uses a technique developed by Alon and Kleitman [1], employing a fractional Helly theorem and a linear separation theorem.

There exist galleries (with a large number of holes) in which each point sees a constant fraction of the area of the gallery (1/10, say) but arbitrarily many points are needed to guard the whole gallery. IS THIS KNOWN, OR SHOULD WE GIVE AN EXAMPLE????

*Research supported by Czech Republic Grant GAČR 201/94/2167 and Charles University grants No. 351 and 361. Part of the work was done while the author was visiting the Hebrew University, Jerusalem.

2 VC-dimension of galleries

First we recall definitions and results from [11], [4]. Let $\mathcal{S}$ be a set system on a set X. We say that a subset $A \subseteq X$ is *shattered* (by $\mathcal{S}$) if every possible subset of A can be obtained as the intersection of some $S \in \mathcal{S}$ with A. The *VC-dimension* of $\mathcal{S}$ is the supremum of the sizes of all finite shattered subsets of X.

Let μ be a probability measure on X such that all sets of $\mathcal{S}$ are measurable. A set $N \subseteq X$ is called an ε-net for $\mathcal{S}$ (with respect to μ) if it intersects each $S \in \mathcal{S}$ with $\mu(S) \geq \varepsilon$ ($\varepsilon > 0$ is a real number). Haussler and Welzl [4], extending ideas of Vapnik and Chervoneniks [11], proved the following:

Theorem 3 *Let X be a set, μ a probability measure on X, and $\mathcal{S}$ a system of measurable sets on X of VC-dimension at most d. Then for any $\varepsilon \in (0,1)$ there exists an ε-net for $\mathcal{S}$ (with respect to μ) of size at most $C(d)\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}$, where the number $C(d)$ depends on d only.*

Let us remark that [4] proves this result for the special case when X is finite and μ is the uniform distribution on X. However, the same proof goes through almost literally for an arbitrary probabilistic measure (see also [11] for a proof of a related result for general probability measures). Another fact we need is as follows:

Lemma 4 [8] [9] [11] *If $\mathcal{S}$ is a set system of VC-dimension d on an n-point set then*

$$|\mathcal{S}| \leq \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d}.$$

In particular, if d is fixed, $|\mathcal{S}|$ is bounded by a fixed polynomial in n.

Our proof of Theorem 1 is based on the following:

Proposition 5 *Let $X \subseteq \mathbb{R}^2$ be compact and simply connected. Then the VC-dimension of the set system $\mathcal{V}(X) = \{V(x);\, x \in X\}$ is bounded by a constant. More generally, if X has at most h holes, then the VC-dimension of $\mathcal{V}(X)$ is bounded by a function of h.*

Proof of Theorem 1. Consider a gallery X as in Theorem 1. Since each $V(x)$ has Lebesgue measure at least ε, an ε-net for the set system $\mathcal{V}(X)$ (with respect to Lebesgue measure) intersects each $V(x)$, and thus guards X. By Theorem 3 and Proposition 5, an ε-net of size $O_h(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon})$ exists in this situation. $\Box$

Let us introduce some terminology concerning visibility graphs. If A is a subset of X, we define the *visibility graph of A in X*, denoted by $VG_X(A)$, as the graph with vertex set A and with two distinct points u, v forming an edge iff they see each other (within X). For two sets $A, B \subseteq X$, we define the *bipartite visibility graph of A, B in X*, denoted by $BVG_X(A, B)$, as the bipartite graph (A, B, E) (A, B are the color classes and $E \subseteq A \times B$ is the edge set), where $(a, b) \in E$ iff a and b see each other.

Proof of Proposition 5. Let d be a sufficiently large number, and suppose that there exists a d-point set $A \subseteq X$ shattered by $\mathcal{V}(X)$. This means that for each subset $S \subseteq A$ there exists a point $y_S \in X$ which sees all points of S but no point of $A \setminus S$; put $B = \{y_S;\, S \subseteq A\}$. In such a situation, we say that A is *shattered by* B.

Consider the bipartite visibility $BVG_X(A,B)$. For later use we note that if $G=(U,V,E)$ is any fixed bipartite graph, A is sufficiently large, and is shattered by B, then $BVG_X(A,B)$ contains an isomorphic copy of G as an induced subgraph (with vertices of U mapped into A and vertices of V mapped into B).

Starting with A,B as above, we find a smaller shattered set is a special position. Draw a line thru each pair of points of A. The arrangement of these at most $\binom{d}{2}$ lines has $O(d^4)$ cells (vertices, edges, and open convex polygons), so there is one such cell containing a subset $B' \subseteq B$ of at least $2^d/O(d^4)$ points of B. These points correspond to subsets of A, so they define a set system $\mathcal{S}_1$ on A. If d_1, the VC-dimension of $\mathcal{S}_1$, were bounded by a constant independent of d, then the number of sets in $\mathcal{S}_1$ would grow at most polynomially with d (by Lemma 4), but we know it grows exonentially, hence d_1 grows to infinity with $d\to\infty$. Thus, we may assume that some subset $A_1 \subseteq A$ is shattered by a subset $B_1 \subseteq B'$, with $d_1=|A_1|$ large.

By the observation made in the beginning of the proof, we know that the bipartite visibility graph of A_1 and B_1 contains any prescribed bipartite induced subgraph (up to some size). In particular, we can select subsets $B_2 \subseteq B_1$ and $A_2 \subseteq A_1$ such that $d_2=|B_2|$ is large, $|A_2|=2^{d_2}$ and B_2 is shattered by A_2 (so we reverse the sides; d_2 can be chosen $\lfloor \log_2 d_1 \rfloor$ in this situation — this is an observation due to Assouad [2]).

Next, we repeat the procedure from the first step of the proof, this time selecting a set $B_3 \subseteq B_2$ of size d_3 (still sufficiently large), and $A_3 \subseteq A_2$, such that B_3 is shattered by A_3 and A_3 lies in a single cell of the arrangement of all lines defined by pairs of points of B_3. This cell must be 2-dimensional (if it were an edge, we would get that all the points of A_3 and B_3 are collinear, which is impossible), so no line determined by two points of A_3 intersects $\mathrm{conv}(B_3)$, and vice versa (in particular, $\mathrm{conv}(A_3)\cap\mathrm{conv}(B_3)=\emptyset$). Hence each point of B_3 sees all points of A_3 within an angle smaller than π, and in the same clockwise angular order; let $\leq_A$ be this linear order the points of A_3. Similarly we have the common counterclockwise angular order, $\leq_B$, of points of B_3 around any point of A_3.

Let us consider the case of X simply connected. Here it suffices to have $d_3=5$. We put $B'=B_3$, and for each $b\in B'$ we consider the point $a=a(b)\in A_3$ which sees all points of B' but b. Let these 5 points form a set $A' \subset A_3$.

Since we have 5 points on each side, we may choose a $b\in B'$ such that b is neither the first nor the last point of B' in the $\leq_B$ ordering, and at the same time $a(b)\in A'$ is not the first or last point in the $\leq_A$ ordering of A'. We get a situation as in Fig. 1, namely that b sees both the predecessor a' and the successor a'' of $a(b)$, and $a=a(b)$ sees both the predecessor b' and the successor b'' of b. It is easy to check that the segments ab' and $a'b$ intersect as shown (because of the restrictions on the relative position of A' and B'), and similarly for the segments $a''b$ and ab''. These four segments are contained in X, and since X is simply connected, also the shaded region bounded by the segments must be a part of X, hence a and b see each other — a contradiction.

Next, let X have h holes. We use a Ramsey-type result of Nešetřil and Rödl [5]. An *ordered bipartite graph* is a bipartite graph (U,V,E) with some linear orderings on U and on V. (Nešetřil [private communication] earlier suggested this kind of approach for exhibiting a forbidden induced bipartite subgraph of visibility graphs of simple polygons; see a remark in the end of this section.)

Theorem 6 *[5] Let (U,V,E) be a fixed ordered bipartite graph. There exists a bipartite graph (R,S,F) such that for any linear orders $\leq_R$ on R and $\leq_S$ on S, the corresponding*

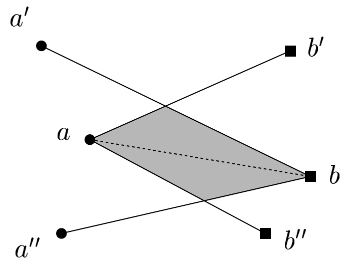

Figure 1: A contradiction to invisibility of a and b.

ordered bipartite graph contains an ordered induced copy of (U, V, E) *(i.e. the induced embedding sends* U *into* R *and* V *into* S *in an order-preserving manner).*

In our situation, we let (U, V, E) be the ordered bipartite graph with $U = (u_0, u_1, \ldots, u_{3h+2})$, $V = (v_0, \ldots, v_{3h+2})$, where the subgraph on each two triples $(u_{3i}, u_{3i+1}, u_{3i+2})$ and $(v_{3i}, v_{3i+1}, v_{3i+2})$ $(i = 0, 1, \ldots, h)$ is as the one in Fig. 1, i.e. u_{3i+1} is connected to v_{3i}, v_{3i+2} but not to v_{3i+1}, v_{3i+1} is connected to u_{3i} and u_{3i+2}, and the remaining edges do not matter. We choose d_3 so large that the bipartite visibility graph $BVG_X(A_3, B_3)$ contains an induced copy (non-ordered) of the graph (R, S, F) constructed for (U, V, E) as in Theorem 6. In the drawing of $BVG_X(A_3, B_3)$, we then obtain $h + 1$ situations as in Fig. 1, with the $h + 1$ shaded regions being pairwise disjoint. At most h of these regions may contain a hole of X, and the remaining one gives a contradiction to the supposed invisibility as for the simply connected case. $\square$

For simply connected galleries, our proof yields a bound of roughly 10^{12} for the VC-dimension. The best lower bound we know is 5, and it seems reasonable to conjecture that the VC-dimension is in fact a small number, perhaps 6. An art gallery with a 5-point shattered subset is shown in Fig 2. The shattered set is indicated by dots numbered 1 thru 5. Crosses mark points which see only certain subsets; e.g., a cross labeled by 134 sees points 1,3, and 4 only. All types of subsets up to symmetry are shown, with an exception of the empty set (for which we can always make a tiny niche somewhere in the wall, so that a point there sees no-one).

Remark. In her thesis [3], Everett asked whether there is a bipartite graph which cannot occur as an induced subgraph of the visibility graph of a simple polygon (i.e. a graph of the form $VG_X(A)$, where X is a simple polygon and A is the set of its vertices). Shermer [10] exhibited such a forbidden bipartite subgraph; in fact he constructed a bipartite graph on 30 vertices which cannot occur as $VG_X(A)$ for any simply connected X and any $A \subseteq X$. For the VC-dimension result we need more — namely a bipartite graph which cannot occur as $BVG_X(A, B)$, i.e. it is not a visibility graph even if we add any subset of edges on A and on B. Perhaps Shermer's method can be extended to provide a better numerical bound on the VC-dimension in the simply connected case.

3 Fractional Helly and duality

The extra step we need for the proof of Theorem 2 is the following

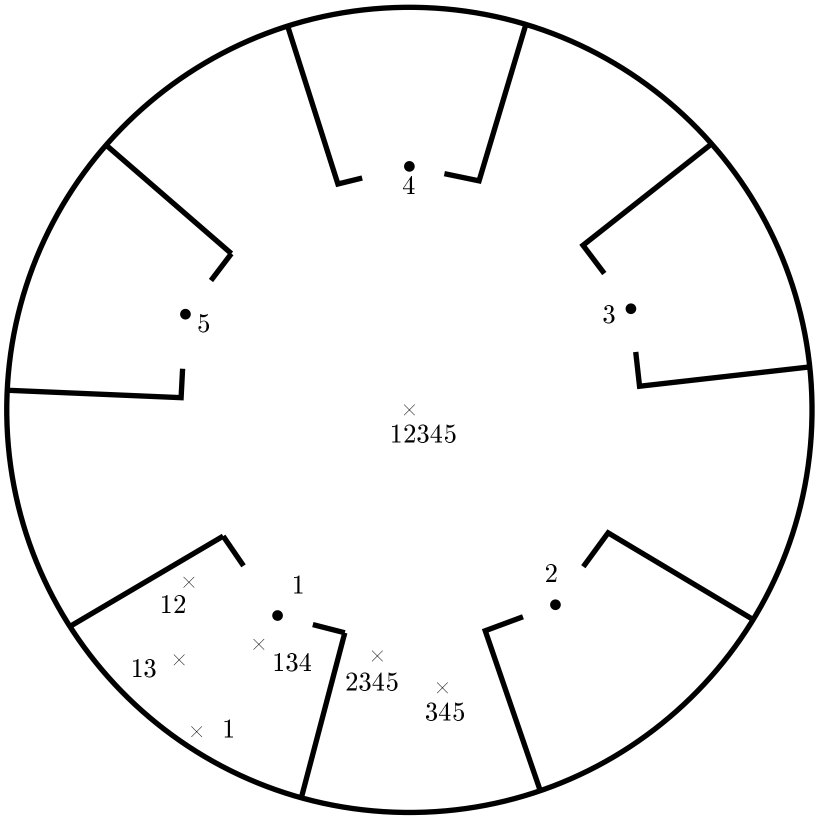

Figure 2: A 5-point shattered set.

Proposition 7 *If X and k are as in Theorem 2, then there exists a Borel measure on μ on X such that $\mu(V(x)) \geq \varepsilon = k???$ for all $x \in X$.*

Once this is proved, we can use the results of the preceding section, with the measure μ replacing the Lebesgue measure.

FRACTIONAL HELLY; WEIGHTED VERSION

The next step also follows the Alon-Kleitman proof. They use the duality theorem of linear programming (which is essentially the separation of disjoint convex sets in a finite-dimensional space by a hyperplane). We need to deal with an infinite family of sets, so we apply an infinite-dimensional separation theorem (alternatively, one could use linear programming duality and a limit argument), in a way suggested to us by E. Matoušková.

Lemma 8 *Let $\mathcal{F}$ be a family of closed sets in a compact Hausdorff space X, and let $\varepsilon \in (0,1]$ be a real number. Suppose that for any finite collection $F_1, F_2, \ldots, F_n$ of sets in $\mathcal{F}$ and any choice of nonnegative reals $t_1, \ldots, t_n$ with $t_1 + t_2 + \cdots + t_n = 1$ there exists a point $x \in X$ such that*

$$\sum_{i;\, x \in F_i} t_i \geq \varepsilon\,.$$

Then there exists a Borel measure, μ, on X, such that $\mu(F) \geq \varepsilon$ for all $F \in \mathcal{F}$.

Proof. Let $C(X)$ be the Banach space of all continuous real functions $X \to \mathbb{R}$ with the supremum norm (i.e. $\|f\| = \max_{x\in X} |f(x)|$). For each set $F \in \mathcal{F}$, define a set $D_F \subseteq C(X)$ as the set of all continuous functions $X \to [0,1]$ which are 1 at all points of F. Let $D \subseteq C(X)$ be the convex hull of $\bigcup_{F\in\mathcal{F}} D_F$.

We claim that D contains no function f with $\|f\| < \varepsilon$. Suppose the contrary; this means that there exists a finite convex combination $f = \sum_{i=1}^n t_i f_i$, where $t_i \geq 0$, $\sum t_i = 1$, and $f_i \in D_{F_i}$ for some $F_1, \ldots, F_n \in \mathcal{F}$, such that $\|f\| < \varepsilon$. Since each D_{F_i} is convex, we may assume that the F_i's are all distinct. Apply the assumption of the lemma on the collection $F_1, \ldots, F_n$ and the numbers $t_1, \ldots, t_n$ corresponding to this convex combination. This yields a point x with $\sum_{i;x\in F_i} t_i \geq \varepsilon$. If $x \in F_i$, then $f_i(x) = 1$ by the definition of D_{F_i}, so we get

$$\varepsilon > \|f\| \geq f(x) = \sum_{i=1}^n t_i f_i(x) \geq \sum_{i;x\in F_i} t_i \geq \varepsilon\,,$$

which is a contradiction.

The convex sets D and $\{f \in C(X);\, \|f\| < \varepsilon\}$ (the open ε-ball) are thus disjoint and the latter one has nonempty interior, hence there exists a hyperplane separating them, that is, a bounded linear functional $h : C(X) \to \mathbb{R}$ such that $h(f) \geq \varepsilon$ for $f \in D$, while $\mu(f) \leq \varepsilon$ for $\|f\| \leq \varepsilon$ (essentially by the Hahn-Banach theorem; see e.g. [7] for an appropriate version of the separation result and references for other results referred to in the rest of this proof). The latter condition gives $\|h\| \leq 1$, where $\|h\| = \sup\{h(f);\, f \in C(X), \|f\| = 1\}$.

By the Riesz Representation theorem, there exists a unique regular Borel signed measure ν on X such that $h(f) = \int_X f \mathrm{d}\nu$ for each $f \in C(X)$. We claim $\nu(F) \geq \varepsilon$ for all $F \in \mathcal{F}$. Indeed, if $\nu(F) < \varepsilon$, we could find a function $f \in D_F$ such that $h(f) < \varepsilon$ as well (choose a $G \supseteq F$ open with $\nu(G) < \varepsilon$, and then Tietze's theorem provides an $f \in C(X)$ which is 0 on $X \setminus G$ and 1 on F, so $h(f) = \int_X f \mathrm{d}\nu \leq \nu(G) < \varepsilon$).

The proof is finished by choosing μ as the variation of ν, i.e. the measure defined by $\mu(E) = \sup\{\sum_{i=1}^k \nu(E_i)\}$, where $E_1, \ldots, E_k$ are disjoint measurable subsets of a set $E \subseteq X$. We have $\mu(X) = \|h\| \leq 1$, and $\mu(F) \geq \nu(F) \geq \varepsilon$ for $F \in \mathcal{F}$. (Alternatively, we could add the set $\{f \in C(X);\, f(x) \geq \varepsilon\, \forall x \in X\}$ to D in the beginning; then the functional h provided by the separation is nonnegative and we get a measure right away.) □

4 A remark on greedy algorithm

One might suspect that under the conditions of Theorem 1, a guarding set of a size bounded in terms of ε could be obtained by a greedy algorithm: Select a guard which sees the maximum possible area, then select the second guard as one seeing the largest part of the area not seen by the first guard, etc. We present an example that this procedure might fail, i.e. select arbitrarily many guards for some simply connected galleries. An example of such a gallery is depicted in fig. 3.

The boundary of the gallery is drawn by a full line, the dotted lines are only auxiliary. The little spikes ("fins") $F_1, F_1', F_2, F_2', \ldots$ are chosen so that the area of F_i and F_i' is much larger than the area of F_{i+1} and F_{i+1}'. The guards placed at A and B suffice to guard all the gallery. However, the first greedily placed guard comes to the point G_1, where it sees both F_1 and F_1' and the largest possible piece of the other fins (all points of the gallery, except for the fins, see everything but possibly parts of the fins, and since F_1, F_1' dominate, we look for a position where both can be seen). Now only the shaded parts of the other fins remain

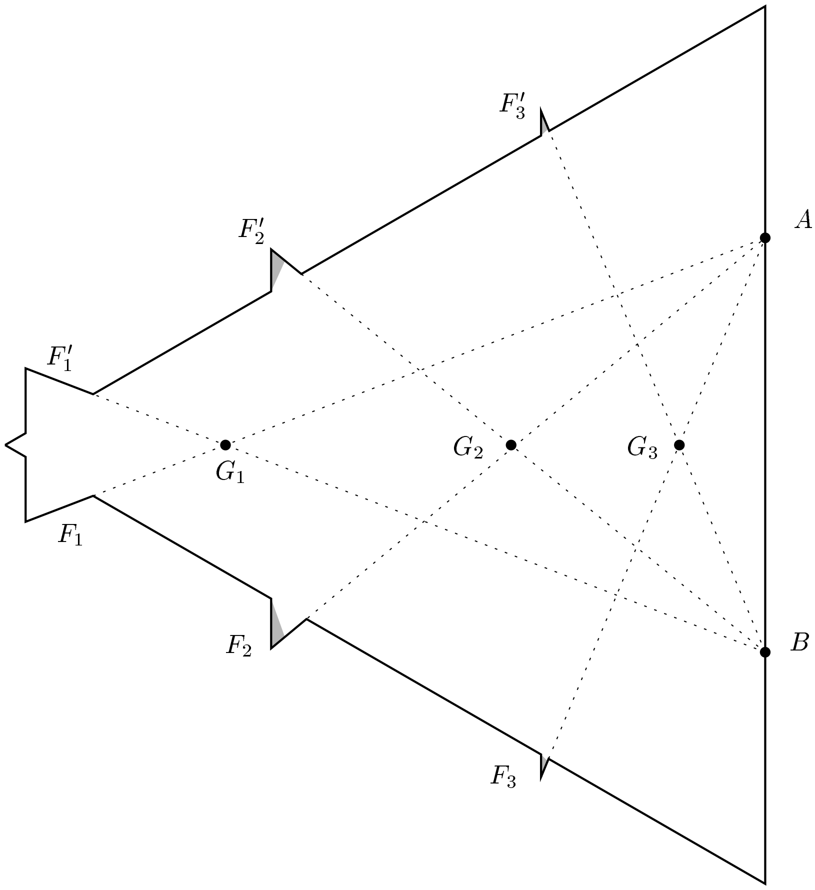

Figure 3: The greedy algorithm fails.

unguarded, with the dominating portion of the area being in F_2, F_2', so the next guard is place in G_2, etc.

The gallery in the figure requires 3 greedily placed guards. It is problematic to actually draw examples of this type forcing the greedy algorithm to place more guards, but the construction method is extended easily. Namely, we start with fins F_1, F_1' much smaller and much closer to the tip of the large triangle, and then we adjoin progressively smaller fins along the sides of the triangle, the next pair always coming to the right of the intersection of the lines connecting the previous pair to A and B (as in Fig. 3). Points in each fin F_i see the portion of the triangle above the horizontal level of A and to the right of the vertical level of the last fin, F_k. If F_k is placed sufficiently far from the vertical side of the triangle (i.e., if we start close enough to the tip with the first pair of fins), this represents a constant fraction of the area (we can get any fraction below $\frac{1}{2}$ by adjusting the proportions appropriately).

Acknowledgment. We thank Eva Matoušková for help with Lemma 8. Thanks go also to Nati Linial and Jaroslav Nešetřil for useful discussions, and to Hazel Everett for pointing out the reference [10].

References

[1] N. Alon and D. Kleitman. Piercing convex sets. *Adv. Math.*, 96(1):103, 1992.

[2] P. Assouad. Densite et dimension. *Ann.Inst.Fourier (Grenoble)*, 33:233–282, 1983.

[3] H. Everett. Visibility graph recognition. PhD. thesis, Univ. of Toronto, 1990. Tech. Report 231/90.

[4] D. Haussler and E. Welzl. Epsilon-nets and simplex range queries. *Discrete Comput. Geom.*, 2:127–151, 1987.

[5] J. Nešetřil and V. Rödl. A short proof of existence of restricted Ramsey graphs by means of a partite construction. *Combinatorica* 1,2:199-202, 1981.

[6] J. O'Rourke. Computational geometry column 18. *Internat. J. Comput. Geom. Appl.*, 3:107–113, 1993. Also in SIGACT News 24:1 (1993), 20–25.

[7] W. Rudin. *Functional Analysis* (2nd edition). McGraw-Hill, New York 1991.

[8] N. Sauer. On the density of families of sets. *Journal of Combinatorial Theory Ser. A*, 13:145–147, 1972.

[9] S. Shelah. A combinatorial problem, stability and order for models and theories in infinitary languages. *Pacific J. Math.* 41:247-261, 1972.

[10] T. Shermer. Several short results in the combinatorics of visibility. Technical Report 91-02, School of Computing Science, Simon Fraser University, British Columbia.

[11] V. N. Vapnik and A. Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.*, 16:264–280, 1971.