Geometric set systems

Jiří Matoušek*

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

Abstract

Let X be a finite point set in the plane. We consider the set system on X whose sets are all intersections of X with a halfplane. Similarly one can investigate set systems defined on point sets in higher-dimensional spaces by other classes of simple geometric figures (simplices, balls, ellipsoids, etc.). It turns out that simple combinatorial properties of such set systems (most notably the *Vapnik-Chervonenkis dimension* and related concepts of *shatter functions*) play an important role in several areas of mathematics and theoretical computer science. Here we concentrate on applications in discrepancy theory, in combinatorial geometry, in derandomization of geometric algorithms, and in geometric range searching. We believe that the described tools might be useful in other areas of mathematics too.

1 Introduction

For a set system $\mathcal{S} \subseteq 2^X$ on an arbitrary ground set X and for $A \subseteq X$, we write $\mathcal{S}|_A = \{S \cap A; S \in \mathcal{S}\}$ for the *set system induced by* $\mathcal{S}$ *on* A (or the *trace* of $\mathcal{S}$ on A). Let $\mathcal{H}$ denote the system of all closed halfplanes in the plane, and let $\mathcal{T}$ be the system of all triangles in the plane. For a finite set $A \subset \mathbb{R}^2$, $\mathcal{H}|_A$ is thus the the system of all subsets of A that can be "cut off" by a halfplane. We will be interested in combinatorial properties set systems of this type. They are far from being understood. For instance, if we ask for the maximum possible number of sets of size exactly k in $\mathcal{H}|_A$ for an n-point set A, we get the notoriously difficult *k-set problem* of combinatorial geometry ([PSS92], [ABFK92], [DE94]). On the other hand, many interesting results can be derived from quite simple combinatorial properties of such set systems; one

*Part of this survey was written while the author was visiting ETH Zürich, whose support is gratefully acknowledged. Also supported by Czech Republic Grant GAČR 0194/1996 and by Charles University grants No. 193,194/1996.

such important property is the so-called Vapnik-Chervonenkis dimension (or VC-dimension for short).

The *VC-dimension* is defined for any set system $\mathcal{S} \subseteq 2^X$ on an arbitrary set X. It is the supremum of the sizes of all shattered subsets $A \subseteq X$; here A is called *shattered* if $\mathcal{S}|_A = 2^A$, i.e. for any $B \subseteq A$ there exists a set $S \in \mathcal{S}$ such that $B = A \cap S$.

For example, it is not difficult to check that the VC-dimension of the set system $\mathcal{H}$ is 3 (no 4-point set can be shattered). Determining the VC-dimension of $\mathcal{T}$ exactly requires some work, but using simple tools presented below it is easily seen that this dimension is bounded by a constant. Similarly set systems defined by other simple geometric figures in a Euclidean space have typically a bounded VC-dimension (a precise formulation will be given later). On the other hand, the system of all convex sets in the plane, say, has an infinite VC-dimension.

The notion now commonly called VC-dimension has been introduced by Vapnik and Chervonenkis [VC71][1]. Numerous applications and extensions of the VC-dimension concept have been developed in statistics (in the theory of so-called empirical processes; some relevant references are [Vap82], [Dud84], [GZ84], [Dud85], [AT89], [Pol90]), in learning theory (where VC-dimension is one of the main concepts; e.g., [BEHW89], [AB92], [Hau92],[KW93], [ABCBH93], [DHS94]), but also for example in program testing [RV]. In combinatorics of hypergraphs, set systems of VC-dimension d can be viewed as a class of hypergraphs with a certain forbidden subhypergraph (the complete hypergraph on $d + 1$ points), which puts this topic into a broader context of extremal hypergraph theory (see for instance [Fra83], [WF94], [DSW94]). Here we do not consider these areas. This survey is mainly focused on the directions of author's own work; we review some general results of combinatorial nature about set systems of bounded VC-dimension, and present applications in geometric discrepancy theory, in combinatorial geometry, and in computational geometry. We mention also more geometric notions and results (which have no good analogue for arbitrary set systems of bounded VC-dimension), namely *cuttings* (sec. 4) and *simplicial partitions* (sec. 5).

2 Set systems of bounded VC-dimension

Shatter functions. These are parameters of a set system related to VC-dimension but often giving more information and easier to work with. The *primal shatter function* of a set system $(X, \mathcal{S})$ is a function, denoted by $\pi_{\mathcal{S}}$,

[1] Under different names, this also appears in other papers ([Sau72],[She72]) but the work [VC71] was probably the most influential for the subsequent developments.

whose value at m ($m = 0, 1, 2, \ldots, |X|$) is defined by

$$\pi_{\mathcal{S}}(m) = \max_{A \subseteq X,\ |A|=m} \left| \mathcal{S}|_A \right|.$$

In words, it $\pi_{\mathcal{S}}(m)$ is the maximum possible number of distinct intersections of the sets of $\mathcal{S}$ with an m-point subset of X. Yet another way to understand the primal shatter function is via the incidence matrix, M, of the set system $\mathcal{S}$ (with rows indexed by sets of $\mathcal{S}$ and columns indexed by points of X): $\pi_{\mathcal{S}}(m)$ is the maximum number of distinct rows appearing in any m-column submatrix of M. An alternative definition of VC-dimension of $\mathcal{S}$ is then $\sup\{d;\ \pi_{\mathcal{S}}(d) = 2^d\}$.

The following simple but basic result bounds the primal shatter function in terms of VC-dimension:

2.1 Lemma *(Vapnik and Chervonenkis [VC71]; Sauer [Sau72]; Shelah [She72]) For any set system $\mathcal{S}$ of VC-dimension at most d, we have $\pi_{\mathcal{S}}(m) \leq \Phi_d(m)$, where $\Phi_d(m) = \binom{m}{0} + \binom{m}{1} + \cdots + \binom{m}{d}$, and this bound is tight in the worst case.*

Hence the behavior of the primal shatter function is fairly restricted: Either it is 2^m for all m (the case of infinite VC-dimension) or it is bounded by a fixed polynomial (the case of finite VC-dimension). Set systems of bounded VC-dimension can thus be characterized as ones with "hereditarily polynomial" number of sets.

An easy but important consequence of Lemma 2.1 is that if we take several set systems $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_k$ on X, each of bounded VC-dimension, and form a new set system $\mathcal{S}$ from the $\mathcal{S}_i$'s by a fixed set-theoretic formula (such as $\mathcal{S} = \{(S_1 \cup S_2) \setminus S_3;\ S_1 \in \mathcal{S}_1, S_2 \in \mathcal{S}_2, S_3 \in \mathcal{S}_3\}$, such an $\mathcal{S}$ has a bounded VC-dimension again (because its primal shatter function is polynomially bounded).

It may often happen (e.g., in many geometric situations) that the primal shatter function is actually considerably smaller than the bound implied by Lemma 2.1. For instance, the set system $\mathcal{H}$ of all halfplanes has VC-dimension 3 but $\pi_{\mathcal{H}}$ is only quadratic. More generally, the set system $\mathcal{H}_d$ of all halfspaces in $\mathbb{R}^d$ has primal shatter function of the order m^d (for a fixed d). A yet more general result is the following:

2.2 Theorem *Let $f(x_1, x_2, \ldots, x_d, a_1, a_2, \ldots, a_p)$ be a fixed real polynomial in $d+p$ variables, and for $a = (a_1, \ldots, a_p) \in \mathbb{R}^p$, let $S_f(a)$ be the set of all points $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ such that $f(x_1, \ldots, x_d, a_1, \ldots, a_p) \geq 0$. Finally let $\mathcal{S}_f = \{S_f(a);\ a \in \mathbb{R}^p\}$. Then $\pi_{\mathcal{S}_f}(m) = O(m^p)$, where the constant of proportionality depends on p, d, and the degree of f.*

Theorem 2.2 is a consequence of results in real algebraic geometry on the number of sign patterns of a system of polynomials ([Ole51], [Mil64], [Tho65]; for recent more precise bounds see also [BPR]). We may thus say that the primal shatter function mainly depends on the number of “degrees of freedom” of the surface delimiting the sets.

Together with the above remark on set systems defined by a fixed set-theoretic formula from other set systems, Theorem 2.2 implies that any set system in $\mathbb{R}^d$ whose each set is defined by at most k polynomial inequalities of maximum degree D has VC-dimension bounded in terms of d, k, D. This subsumes many (but not all) interesting examples with bounded VC-dimension. On the other hand, precise estimates for the VC-dimension may be difficult in such cases.

Next, we define the dual shatter function. This is just the primal shatter function of the set system “dual” to $\mathcal{S}$, whose incidence matrix arises by transposing the incidence matrix of $\mathcal{S}$ and deleting multiple rows. Explicitly, the *dual shatter function* of a set system $(X, \mathcal{S})$ is a function, denoted by $\pi^*_{\mathcal{S}}$, whose value at m is the maximum number of equivalence classes on X defined by an m-element subfamily $\mathcal{A} \subseteq \mathcal{S}$, where two points $x, y \in X$ are *equivalent with respect to* $\mathcal{A}$ if x belongs to the same sets of $\mathcal{A}$ as y does.

As observed by Assouad [Ass83], the set system dual to a set system of VC-dimension d has VC-dimension at most $2^{d+1} - 1$. (This is not difficult to see if we adopt the incidence matrix point of view: From a $2^d \times d$ zero-one matrix containing all possible row patterns, we can select a $\lfloor \log_2 d \rfloor \times d$ submatrix containing all possible column patterns.) Consequently, the dual shatter function is polynomially bounded iff the primal one is.

Geometrically, the dual shatter function mainly depends on the space dimension. More precisely, *if f and $\mathcal{S}_f$ have the same meaning as in Theorem 2.2 then $\pi^*_{\mathcal{S}_f}(m) = O(m^d)$, where the constant of proportionality depends on p, d, and the degree of f*. This is actually quite intuitive: Given m sets from $\mathcal{S}_f$, their bounding surfaces partition $\mathbb{R}^d$ into $O(m^d)$ cells, and these cells correspond to the equivalence classes from the definition of the dual shatter function.

More examples. Let X be a compact and simply connected set in the plane. For a point $x \in X$, define $V(x) \subseteq X$ as the set of all points y visible from x within X, i.e. such that the segment xy is completely contained in X. The set system $\{V(x); x \in X\}$ is a nontrivial example of a geometrically defined set system of bounded VC-dimension [KM95a], of a different type than the geometric examples considered above. The current best estimates for the maximum possible VC-dimension of such a set system are 6 from below and 23 from above [Val95]. If we allow a bounded number, h, of holes in X, the VC-dimension is still bounded by a function of h.

For other interesting and nontrivial bounds on VC-dimension see Karpinski and Macintyre [KM95b].

The examples mentioned so far were all of geometric nature. A very non-geometric example (in some sense) a set system of VC-dimension 2 is the system of lines of a finite projective plane; some its properties in this context have been investigated by Alon et al. [AHW87].

Pseudohyperplanes and oriented matroids. We have just remarked that a finite projective plane provides a "non-geometric" example of a set system of bounded VC-dimension. Can some geometrically defined set systems be combinatorially distinguished from other set systems with the same VC-dimension (and shatter functions, say)? Partial positive results are known for the case of set systems defined by cells in hyperplane arrangements (which are closely related, via geometric duality, to set systems defined by halfspaces on finite point sets in $\mathbb{R}^d$). Let H be a collection of n hyperplanes in $\mathbb{R}^d$ in general position. These hyperplanes cut $\mathbb{R}^d$ into a collection C of d-dimensional open cells. Suppose that for every hyperplane $h \in H$, one of the halfspaces bounded by h is distinguished as positive. Then each cell $c \in C$ defines a subset of H, namely the set of hyperplanes for which c lies in their positive halfspace. All such sets together define a set system $\mathcal{C}$ on H which has VC-dimension d. The number of sets in $\mathcal{C}$ is $\Phi_d(n)$, i.e. it attains the bound in Lemma 2.1. As is well known in the theory of oriented matroids (see e.g., [BVS+92]), there is no good way of combinatorially distinguishing arrangements of hyperplanes from arrangements of *pseudohyperplanes* (these are possibly curved surfaces such that the arrangement of any $d+2$ of them is topologically equivalent to an arrangement of $d+2$ hyperplanes, but globally an equivalent arrangement need not be realizable by hyperplanes). Also, a cleaner picture is obtained by passing from an "affine" to a "projective" case; in our situation this means considering an arrangement of $(d-1)$-dimensional (pseudo)spheres on the d-dimensional sphere. For set systems corresponding to simple pseudospherical arrangements (i.e. with the pseudospheres being in general position), Gärtner and Welzl [GW94] found the following nice characterization: a set system $(X, \mathcal{S})$ comes from a simple arrangement of pseudospheres on S^d iff it has VC-dimension d, it is closed on complements ($S \in \mathcal{S} \Rightarrow X \setminus S \in \mathcal{S}$), and $|\mathcal{S}| = 2\Phi_{d-1}(|X|-1)$ (which is the maximum number of sets a set system of VC-dimension d closed on complements may have). This paper also contains a number of other characterizations and connections to the oriented matroids theory.

Sampling properties. If X is equipped by a probability measure μ and $\mathcal{S}$ is a system of μ-measurable sets on X, we are interested in finite samples $A \subseteq X$ which approximate μ on the sets of $\mathcal{S}$ in a suitable sense. The weakest such notion is an ε-net ([HW87]; not to be confused with the synonymous notion for metric spaces), which is a set intersecting all "μ-big" sets of $\mathcal{S}$: A set $N \subseteq X$ (not necessarily one of the sets of $\mathcal{S}$) is called an *ε-net for* $(X, \mathcal{S}, \mu)$ ($\varepsilon \in [0,1]$ is a real number) if any set in $S \in \mathcal{S}$ with $\mu(S) > \varepsilon$ intersects N. The most important case is that of X finite and μ uniform; then N is required to intersect each set $S \in \mathcal{S}$ with $|S| > \varepsilon|X|$ (in the sequel, if we speak about ε-nets and similar for a set system on a finite set, we always mean the uniform measure unless stated otherwise).

A straightforward probabilistic argument shows that if $\mathcal{S}$ is finite, then a subset of $(1/\varepsilon)\ln|\mathcal{S}|$ points randomly chosen from X (according to μ) is an ε-net with a positive probability. A key result of Haussler and Welzl [HW87] says that if $\mathcal{S}$ has a bounded VC-dimension d, then ε-nets exist of size depending only on d and ε. To make the formulas look simpler, let us write $1/r$ instead of ε.

2.3 Theorem (Epsilon-net theorem) *(Haussler and Welzl [HW87], Komlós et al. [KPW92]) Let $N(d,r)$ be the minimum number such that any set system of VC-dimension d admits an $(1/r)$-net of size at most $N(d,r)$. For any $d \geq 2$ there exists an $r_0 > 0$ such that for all $r > r_0$, $(d-2)r\ln r \leq N(d,r) \leq dr\ln r$.*

(The case $d = 1$ is special; set systems of VC-dimension 1 can be completely described and have $(1/r)$-nets of size $O(r)$.) Both the upper and lower bounds on $N(d,r)$ are proved by a probabilistic argument; for the upper bound, it turns out that a random sample of $dr\ln r$ points of X is a $(1/r)$-net with high probability. (A proof of a slightly weaker upper bound is sketched in section 3.) Since the lower bound uses a random set system, a challenging and probably very difficult open problem is whether some improvement in the ε-net size is possible for set systems defined geometrically. For instance, does the set system $\mathcal{T}|_A$ defined by triangles on a finite set A in the plane admit $(1/r)$-nets of size $O(r)$?

A stronger requirement on a sample than being an ε-net is to be an ε-approximation. A subset $A \subseteq X$ is an *ε-approximation* for $(X,\mathcal{S},\mu)$ provided that

$$\left|\frac{|A\cap S|}{|A|} - \mu(S)\right| \leq \varepsilon$$

for every set $S \in \mathcal{S}$. Clearly, an ε-approximation is also an ε-net, but not conversely. The ε-approximation was the notion of a "good" sample considered by Vapnik and Chervonenkis [VC71]. They showed that for a set system of VC-dimension bounded by a constant d, a random sample A of size $Cr^2\log r$ with a suitable constant $C = C(d)$ is a $(1/r)$-approximation for $(X,\mathcal{S},\mu)$ with high probability. The size of the $(1/r)$-approximation obtained in this way is thus roughly quadratic compared to a $(1/r)$-net. Although a random sample of fewer than $const.r^2\log r$ points typically is not a $(1/r)$-approximation, $(1/r)$-approximations of a somewhat smaller size can be constructed by other methods (based on a connection of ε-approximations with discrepancy) [MWW93]: If the primal shatter function is bounded by Cm^d for some fixed C,d $(d > 1)$, then there exist $(1/r)$-approximations of size $O(r^{2-2/(d+1)})$ ([MWW93],[Mat95b]). Similarly if $\pi^*_{\mathcal{S}}(m) \leq Cm^d$ for all m, $(1/r)$-approximations of size $O(r^{2-2/(d+1)}(\log r)^{1-1/(d+1)})$ exist.

Ball packing. Consider a set system $(X, \mathcal{S})$ of VC-dimension d on an n-point set X. Define a metric on $\mathcal{S}$: The distance of two sets $S_1, S_2 \in \mathcal{S}$ is the cardinality of their symmetric difference $S_1 \triangle S_2$. An interesting and quite useful fact is that as far as packing of balls of some given radius is concerned, this metric space behaves similarly as the d-dimensional Euclidean space. Precisely:

2.4 Lemma (Packing lemma) *(Haussler [Hau95]) Let $(X, \mathcal{S})$ have VC-dimension d, and let $k \geq 1$ be an integer. Let $\mathcal{P} \subseteq \mathcal{S}$ be a set such that any two distinct sets of $\mathcal{P}$ have distance at least k. Then $|\mathcal{P}| \leq (cn/(k+d))^d$ for some constant c.*

This means that for a fixed d, at most $O((n/\rho)^d)$ disjoint balls of radius ρ can be packed into $\mathcal{S}$ (which is an analogy to packing Euclidean balls of radius ρ into a d-cube of volume n, say). In fact, it is not difficult to verify that Haussler's proof also implies that if the primal shatter function $\pi_{\mathcal{S}}(m)$ is at most Cm^d for all m and for some constants $C, d > 1$ then the $O((n/\rho)^d)$ bound for the number of disjoint ρ-balls remains valid.

It is instructive to prove a weaker bound, namely $O((n/\rho)^d \log^d(n/\rho))$, using ε-nets. So let d be a constant, and let $\mathcal{P} \subseteq \mathcal{S}$ satisfy $|S_1 \triangle S_2| > 2\rho$ for all $S_1 \neq S_2 \in \mathcal{P}$. Consider the set system $\mathcal{D} = \{S_1 \triangle S_2;\ S_1, S_2 \in \mathcal{S}\}$. Set $r = n/2\rho$, and fix a $(1/r)$-net N of size $O(r \log r)$ for $\mathcal{D}$ ($\mathcal{D}$ has a bounded VC-dimension since its primal shatter function is polynomially bounded). Whenever the symmetric difference of any two sets $S_1, S_2 \in \mathcal{S}$ has more than $n/r = 2\rho$ elements then it contains a point of N. In particular, we get $S_1 \cap N \neq S_2 \cap N$ for any two distinct sets $S_1, S_2 \in \mathcal{P}$. Therefore the set system induced by $\mathcal{S}$ on N has at least $|\mathcal{P}|$ elements, and so we get $|\mathcal{P}| \leq \pi_{\mathcal{S}}(|N|) = O((n/\rho)^d \log^d(n/\rho))$ as claimed.

Haussler's proof of Lemma 2.4 was simplified by Chazelle [unpublished note].

Spanning trees with low crossing number. Let $(X, \mathcal{S})$ be a set system, and let G be a (simple, undirected) graph with vertex set X. We say that a set $S \in \mathcal{S}$ *crosses* an edge $\{u, v\}$ of G if $|S \cap \{u, v\}| = 1$. The *crossing number of G with respect to the set S* is the number of edges of G crossed by S, and the *crossing number of G* is the maximum of crossing numbers of G with respect to all sets of $\mathcal{S}$. The following theorem is essentially due to Welzl [Wel88]. (That paper has a somewhat weaker bound. The presented bound arises by plugging Haussler's lemma 2.4 into Welzl's original proof; see also [CW89] for arguments proving particular geometric cases without Haussler's lemma.)

2.5 Theorem *Let $\mathcal{S}$ be a set system on an n-point set X, with $\pi^*_{\mathcal{S}}(m) \leq Cm^d$ for all m, where $C, d > 1$ are constants. Then there exists a spanning tree T with vertex set X whose crossing number is at most $C_1 n^{1-1/d}$, where $C_1 = C_1(C, d)$ is another constant. (In fact, T can be chosen as a path.)*

For a proof, we need

2.6 Lemma (Short edge lemma) *Let $\mathcal{S}$ be a set system as in Theorem 2.5. Then for any set (or multiset[2]) $Q \subseteq \mathcal{S}$, points $x, y \in X$ exist such that the edge $\{x, y\}$ is crossed by at most $C_2\,|Q|/n^{1/d}$ sets of Q, C_2 a suitable constant.*

Lemma 2.6 follows by applying the Packing lemma 2.4 on the set system dual to (X, Q).

We sketch the proof of Theorem 2.5. Imagine that we tried to construct a spanning tree T on X as required in Theorem 2.5 by the greedy (Kruskal's) algorithm. i.e. by selecting edges into T one by one and always choosing an edge of the smallest weight connecting distinct components of the already constructed part of T, where the weight of an edge is the number of sets of $\mathcal{S}$ crossing it. Having selected i edges, we have $n - i$ components in the graph constructed so far. Let X_i be a set containing one point from each component; applying Lemma 2.6 with $X = X_i$ and $Q = \mathcal{S}|_{X_i}$ we get that there exists an edge of weight $O(|\mathcal{S}|/(n-i)^{1/d})$ connecting two distinct components. A simple calculation shows that this algorithm produces a spanning tree which is good on the average, i.e. an average set of $\mathcal{S}$ crosses the right number of edges; however, it might happen that few exceptional sets of $\mathcal{S}$ could cross many more edges.

To circumvent this, the edge selection strategy is improved to penalize sets of $\mathcal{S}$ which already cross many of the edges selected so far (this "reweighing strategy" is useful in several proofs in different areas [Lit88],[Cla93]). Specifically, start with all sets of $\mathcal{S}$ having weight 1, and having selected a new edge e_i in ith step, double the current weight of all sets of $\mathcal{S}$ crossing e_i. In each step, we select the lightest edge connecting two distinct components, where the weight of an edge e is the sum of current weights of sets crossing e.

We need to bound the crossing number, κ, of the resulting spanning tree, T. Let $w_i(S)$ be the weight of a set $S \in \mathcal{S}$ after i edges have been selected, and put $W_i = \sum_{S \in \mathcal{S}} w_i(S)$. We have $W_{n-1} \geq \max_{S \in \mathcal{S}} w_{n-1}(S) = 2^\kappa$. On the other hand, from Lemma 2.6 (with Q containing $w_i(S)$ copies of each $S \in \mathcal{S}$) one can derive $W_{i+1} \leq W_i + O(W_i/(n-i)^{1/d})$. Calculation then gives $\kappa = O(\log|\mathcal{S}| + n^{1-1/d})$. Since the dual shatter function of $\mathcal{S}$ is polynomially bounded, $\mathcal{S}$ has a constant-bounded VC-dimension and hence $|\mathcal{S}|$ is polynomial in n. This concludes a proof sketch for Theorem 2.5.

3 Bounds for geometric discrepancy

The notion of *uniformly distributed sequences* and *uniformly distributed sets* is important in many branches of mathematics (measure theory, ergodic theory,

[2]Multiset means that Q may contain several copies of the same set $S \in \mathcal{S}$. The cardinality of Q is counted with these multiplicities.

diophantine approximation theory, statistics, numerical integration). Together with a theory of uniform distribution, questions about *irregularities of distribution* have been studied, for instance: How uniformly can n points be placed in the unit cube, with respect to a given collection of simple geometric shapes (such as axis-parallel boxes, balls, convex sets etc.). Discrepancy theory is a rich discipline today, with lots of difficult open problems), and here we can only touch it (more comprehensive sources are [BC87], [BS95]; discrepancy from a combinatorial point of view is treated in some chapters of [Spe87], [AS93], [PA95]).

The whole discrepancy theory, or theory of irregularities of distribution, starts with the following Van Aardenne-Ehrenfest theorem (proving a conjecture of Van der Corput; see [BC87] for references on earlier work in discrepancy theory): For each infinite sequence of real numbers in the interval $[0,1]$ and for any $k > 0$ there exists an initial segment $(x_1, \ldots, x_n)$ of the considered sequence and a subinterval $(\alpha, \beta) \subseteq [0,1]$ such that the number of elements of $\{x_1, x_2, \ldots, x_n\}$ belonging to (α, β) deviates from $n(\beta - \alpha)$ (i.e. the expected number for a total uniformity) by at least k. This result has been improved several times. Schmidt proved that the right order of magnitude of k is $const. \log n$. The one-dimensional formulation for initial segments of sequences is equivalent to a two-dimensional formulation with a discrepancy of an n-point set with respect to axis-parallel rectangles.

Discrepancy with respect to Lebesgue measure. A general definition of discrepancy of a point set in the d-dimensional unit cube $U = [0,1]^d$ can be formulated as follows (numerous variations have been considered in the literature, though): Let $\mathcal{S}$ a system of (measurable) subsets of $\mathbb{R}^d$, and A an n-point set in U. The *discrepancy of A with respect to Lebesgue measure on $\mathcal{S}$* is

$$D(A, \mathcal{S}) = \max_{S \in \mathcal{S}} \left| n.\lambda^d(S \cap U) - |A \cap S| \right|,$$

where λ^d stands for the d-dimensional Lebesgue measure. Further we let $D(n, \mathcal{S}) = \min_{A \subset U;\, |A|=n} D(A, \mathcal{S})$. By comparing the definition of $D(A, \mathcal{S})$ and the definition of an ε-approximation given earlier, we see that $D(A, \mathcal{S}) \le \Delta$ holds iff A is a (Δ/n)-approximation for the set system $\mathcal{S}|_U$ with the measure μ being λ^d restricted to U.

The above mentioned result of Schmidt is then equivalent to a statement about discrepancy of axis-parallel rectangles: $D(n, \mathcal{R}_2) \ge c \log n$, where $c > 0$ is a constant and $\mathcal{R}_d$ denotes the system of all axis-parallel boxes in $\mathbb{R}^d$ (an apparently very difficult open problem is to determine the asymptotics of $D(n, \mathcal{R}_d)$ for $d \ge 3$). In subsequent works (for enstance, of Schmidt, Roth, Halász, Beck, Alexander) the discrepancy with respect to Lebesgue measure has been studied in higher dimensions and for other classes of shapes. Surprisingly, while the discrepancy for axis-parallel rectangles is of the order $\log n$, the discrepancy for circular discs or for arbitrarily rotated rectangles is of the order (approximately) $n^{1/4}$.

Combinatorial discrepancy. Also a purely combinatorial notion of discrepancy for set systems has been investigated in the literature. Let $(X, \mathcal{S})$ be a finite set system and $\chi : X \to \{-1, +1\}$ a mapping; in this context it is called a *coloring* of X. For a set $A \subseteq X$, we write $\chi(A) = \sum_{x \in A} \chi(x)$. The *discrepancy* of χ for $\mathcal{S}$ is defined by $\operatorname{disc}(\mathcal{S}, \chi) = \max_{S \in \mathcal{S}} |\chi(S)|$, and the *discrepancy* of $\mathcal{S}$ is $\operatorname{disc}(\mathcal{S}) = \min\{\operatorname{disc}(\mathcal{S}, \chi);\ \chi : X \to \{-1, +1\}\}$. Intuitively, we want a coloring such that each set has nearly equal number of pluses and minuses.

If A is a set of n points in $\mathbb{R}^d$ and $\mathcal{S}$ a system of subsets of $\mathbb{R}^d$, we may consider the discrepancy of the set system induced on A by $\mathcal{S}$. Such discrepancy is sometimes called the *red-blue discrepancy* or *combinatorial discrepancy* of A, in order to distinguish it from the "continuous" discrepancy with respect to Lebesgue measure.

These two types of geometric discrepancy are closely related, as was observed by Beck [Bec84]; roughly speaking, lower bounds for the continuous discrepancy yield lower bounds for the red-blue discrepancy and upper bounds for the red-blue discrepancy give upper bound for the continuous discrepancy. One possible quantitative formulation is the following:

3.1 Proposition *Let $\mathcal{S}$ be a system of measurable sets in $\mathbb{R}^d$ and let be k, n natural numbers. Moreover, assume that there is an $S_0 \in \mathcal{S}$ containing the whole unit cube U. Then $D(n, \mathcal{S}) \leq \frac{1}{2^k} D(2^k n, \mathcal{S}) + \sum_{i=1}^{k} \frac{\operatorname{disc}(2^i n, \mathcal{S})}{2^{i-1}}$.*

The proof idea is to take a set A_0 with $2^k n$ points and with $D(A_0, \mathcal{S})$ as small as possible. Then we fix a ± 1 coloring χ_0 of A_0 with minimum discrepancy, we take the smaller of the color classes $\chi_0^{-1}(1)$, $\chi_0^{-1}(-1)$ and we add a small number of points to it so that we obtain a set A_1 of precisely $|A_0|/2$ points. This "halving step" is repeated k times, and calculation gives the claimed inequality. Beck used Proposition 3.1 (with $k \approx \log n$) to derive $\operatorname{disc}(n, \mathcal{R}_2) \geq c' \log n$ from the above mentioned Schmidt's result $D(n, \mathcal{R}_2) \geq c \log n$ (determining the red-blue discrepancy of axis-parallel rectangles is known as Tusnády's problem and it is not satisfactorily solved yet).

Discrepancy from shatter functions. Lower bounds for geometric discrepancy are proved, with few exceptions, in the continuous setting, and they employ analytic methods like orthogonal functions, Fourier transform, etc. (see [BC87], [Ale90]). Upper bounds seem to go in two main directions: For axis-parallel boxes, various explicit constructions based on number theory, finite fields, and so on are known; they are studied in a great detail also from a practical point of view, since such sets with a low discrepancy turned out to be very important for higher-dimensional numerical integration and computer simulations ([Nie92]). On the other hand, upper bounds for sets such as disks or rectangles with rotation allowed have been obtained via semi-random combinatorial constructions. A large part of known upper bounds in this direction is subsumed by the following theorem bounding discrepancy in terms of shatter functions.

3.2 Theorem *Let $\mathcal{S}$ be a set system on an n-point set X, and let C and $d > 1$ be constants.*

(i) [Mat95b] *If the primal shatter function satisfies $\pi_{\mathcal{S}}(m) \leq Cm^d$ for all $m = 1, 2, \ldots, |X|$, then*

$$\mathrm{disc}(\mathcal{S}) = O\left(n^{\frac{1}{2}-\frac{1}{2d}}\right) .$$

(ii) [MWW93] *If the dual shatter function satisfies $\pi^*_{\mathcal{S}}(m) \leq Cm^d$ for all $m = 1, 1, \ldots, |\mathcal{S}|$, then*

$$\mathrm{disc}(\mathcal{S}) = O\left(n^{\frac{1}{2}-\frac{1}{2d}}\sqrt{\log n}\right) .$$

As a consequence of (i), the red-blue discrepancy of an n-point set in $\mathbb{R}^d$ with respect to halfspaces is $O(n^{1/2-1/2d})$, which asymptotically matches a lower bound due to Alexander [Ale90]; hence (i) is asymptotically sharp. On the other hand, for the set system of all circular discs in the plane, say, the primal shatter function is cubic, while the dual shatter function is quadratic, and hence a better upper bound, of $O(n^{1/4}\sqrt{\log n})$, is obtained from (ii). It is not known whether this upper bound for the discrepancy of disks can be improved, but the bound of Theorem 3.2(ii) in general cannot be improved (a proof will be sketched below).

Dual shatter function bound. To understand the connection of shatter functions to discrepancy, it is perhaps best to begin with the following simple probabilistic lemma, showing that if we have polynomially many sets on an n-point set, then the discrepancy is at most about $\sqrt{n}$:

3.3 Lemma (Random coloring lemma) *Let $\mathcal{S}$ be a set system on an n-point set X, such that all sets of $\mathcal{S}$ have size at most k. For a random coloring $\chi : X \to \{+1, -1\}$, $\mathrm{disc}(\mathcal{S}, \chi) \leq \sqrt{2k \ln(4|\mathcal{S}|)}$ holds with probability at least $\frac{1}{2}$.*

Indeed, for any set $S \in \mathcal{S}$ we have $\Pr(|\chi(S)| > \lambda\sqrt{k}\,) < 2e^{-\lambda^2/2}$ ($\lambda > 0$ a parameter) by the simplest Chernoff-type tail estimate, and setting $\lambda = \sqrt{2\ln(4|\mathcal{S}|)}$ does the job. There exist set systems with k sets of size k and with discrepancy $const.\sqrt{k}$, hence the bound in the Random coloring lemma cannot be much improved without further assumptions on the set system. The point of Theorem 3.2 is that a substantial improvement is possible if the shatter functions are polynomially bounded.

From Theorem 2.5 and Lemma 3.3, the "dual" discrepancy bound in Theorem 3.2(ii) follows quite simply. Namely, given X and $\mathcal{S}$ as in Theorem 3.2(ii), choose a spanning path with crossing number $O(n^{1-1/d})$ on X.

For simplicity, let $|X|$ be even. Delete every second edge of this path, which leaves us with a matching $M = \{\{u_1, v_1\}, \{u_2, v_2\}, \ldots, \{u_k, v_k\}\}$ (a set of pairwise vertex-disjoint edges) covering all vertices. Define a random coloring $\chi : X \to \{+1, -1\}$, by coloring the points $u_1, \ldots, u_k$ randomly and independently, and by setting $\chi(v_i) = -\chi(u_i)$ for all i.

Look at a fixed set $S \in \mathcal{S}$, and classify the edges of M into two types: those with both points inside S or both points outside S, and those crossed by S. The edges of the former type contribute 0 to $\chi(S)$. The contributions of edges of the latter type to $\chi(S)$ are, by definition of χ, independent random variables attaining values $+1$ and -1 with equal probability. The number of these variables is $O(n^{1-1/d})$. Thus, the situation is as if we had a random coloring of $|\mathcal{S}|$ sets of size $O(n^{1-1/d})$ each, and the Random coloring lemma 3.3 implies that $\mathrm{disc}(\mathcal{S}, \chi) = O(n^{1/2-1/2d}\sqrt{\log n})$ with a positive probability.

Before turning to the "primal" discrepancy bound, let us remark that Lemma 3.3 also gives a simple proof of existence of $(1/r)$-nets of size $O(r \log r)$ for set systems of constant-bounded VC-dimension (a rough form Theorem 2.3). The idea is similar as in the proof of Proposition 3.1. For simplicity, we consider the case of a finite X with the uniform measure. Let the primal shatter function of $\mathcal{S}$ be bounded by a fixed polynomial. Form a new system $\mathcal{S}' = \mathcal{S} \cup \{X\}$. By the Random coloring lemma 3.3, we know that any subsystem of $\mathcal{S}'$ induced by an m-point set has discrepancy $O(\sqrt{m \log m})$. Start with a ± 1 coloring of X with discrepancy $\Delta_1 \leq C\sqrt{n \log n}$, take the smaller color class and add some points to get a set X_1 of $n_1 = \lfloor n/2 \rfloor$ points, color it with discrepancy at most $\Delta_2 = C\sqrt{n_1 \log n_1}$, obtain X_2, etc. Iterate as long as it can be guaranteed that the current set, X_i, intersects all sets of $\mathcal{S}$ of size $> n/r$. Take the last such X_i for the $(1/r)$-net. Calculation shows that this last X_i has size $O(r \log r)$.

As was mentioned above, the $\sqrt{\log n}$ factor cannot in general be removed from the bound in Theorem 3.2(ii), at least for $d = 2, 3$. The probabilistic construction showing it [Mat94] starts with some set system $\mathcal{S}_0$ on an n-point set with at least $n^{1+\delta}$ sets for some fixed $\delta > 0$ such that the sets in $\mathcal{S}_0$ have size roughly $n^{1-1/d}$, and any d of them intersect in at most a bounded number, $t = t(d)$, points. The set system $\mathcal{S}$ serving as a lower-bound example for Theorem 3.2(ii) is $\mathcal{S} = \{R(S);\, S \in \mathcal{S}_0\}$, where $R(S)$ denotes a random subset of S (the random selections are mutually independent). It is straightforward to show that such an $\mathcal{S}$ always satisfies $\pi^*_{\mathcal{S}}(m) = O(m^d)$. If we set $\Delta = cn^{1/2-1/2d}\sqrt{\log n}$ for a sufficiently small constant c, it can be calculated that for any fixed coloring χ, $\Pr[\mathrm{disc}(\mathcal{S}, \chi) \leq \Delta] < 2^{-n}$, and hence there exists a specific $\mathcal{S}$ for which none of the 2^n possible colorings gives discrepancy at most Δ.

To make the proof work we still need to construct a suitable $\mathcal{S}_0$. Unfortunately, this is equivalent to an old open problem of constructing certain asymptotically extremal bipartite graphs without a $K_{d,t}$ subgraph (the complete bipartite graph on d and t vertices), and no such construction is known for a general d. For $d = 3$, one can take $X = F \times F \times F$ (where F is a finite field) and define $\mathcal{S}_0$ as the set of graphs of $n^{1+\delta}$ random bivariate polynomials of a suitable constant degree D over F; applying Bézout's theorem, one can show that $\mathcal{S}_0$ has the required properties with high probability (for $d = 2$, a similar but simpler construction works). For $d \geq 4$, the construction of a suitable $\mathcal{S}_0$ is not known (this problem is not solved by a recent result of Kollár et al. [KRS96] since we need a bipartite graph where one class has significantly more vertices than the other class).

Primal shatter function bound. The first step towards a proof of of Theorem 3.2(i) is Beck's "Partial coloring lemma" [Bec81]. By a *partial coloring* of a set X we mean any mapping $\chi : X \to \{-1, +1, 0\}$. One possible formulation is the following:

3.4 Lemma (Partial coloring lemma) *Let $\mathcal{F}$, $\mathcal{M}$ be set systems*[3] *on an n-point set X, $|\mathcal{M}| > 1$, such that $|M| \leq k$ for every $M \in \mathcal{M}$ and*

$$\prod_{F \in \mathcal{F}} (|F| + 1) \leq 2^{(n-1)/5}. \tag{1}$$

Then there exists a partial coloring $\chi : X \to \{-1, 0, +1\}$, such that the value of χ is nonzero for at least $n/10$ elements of X, $\chi(F) = 0$ for every $F \in \mathcal{F}$ and $|\chi(M)| \leq \sqrt{2k \ln(4|\mathcal{M}|)}$ for every $M \in \mathcal{M}$.

Intuitively, the situation is as follows. We have the "few" sets of $\mathcal{F}$, for which we insist that the discrepancy of χ is 0. Each such $F \in \mathcal{F}$ thus puts one condition on χ. It seems plausible that if we do not put too many conditions, then a χ randomly selected among the ones satisfying these conditions will still be "random enough" to behave as a true random coloring on the sets of $\mathcal{M}$. In the lemma, we claim something weaker, however: Instead of a "true" coloring $\chi : X \to \{+1, -1\}$ we obtain a partial coloring χ, which is only guaranteed to be nonzero at a constant fraction of points. If we want to get a full coloring, the lemma has to be applied iteratively: Color 10% of points, restrict the set system on the remaining points, color 10% of them, etc.; the primal shatter function condition is well-suited for such an iterative method since it is inherited to subsets.

To prove the Lemma, let $\mathcal{C}_1$ be the set of all colorings χ with $\mathrm{disc}(\mathcal{M}, \chi) \leq \sqrt{2k \ln(4|\mathcal{M}|)}$. By the Random coloring lemma 3.3, we have $|\mathcal{C}_1| \geq 2^{n-1}$. Define a mapping $v : \mathcal{C}_1 \to \mathbb{Z}^{|\mathcal{F}|}$ by setting $v(\chi) = (\chi(F);\ F \in \mathcal{F})$. Counting shows that there exists a value $v_0 \in \mathbb{Z}^{|\mathcal{F}|}$ with $\mathcal{C}_2 = v^{-1}(v_0)$ big, and hence $\mathcal{C}_2$ contains two colorings χ_1, χ_2 differing in at least 10% of components. The desired partial coloring is then defined as $\chi = (\chi_1 - \chi_2)/2$.

Let us remark that this proof (a combination of probabilistic method with the pigeonhole principle) is nonconstructive, and it is a challenging open

[3] $\mathcal{F}$ for "few" sets, $\mathcal{M}$ for "many" sets.

problem to find an algorithm (randomized or deterministic) for computing a partial coloring χ as in the Lemma in polynomial time.

To apply Lemma 3.4 for bounding the discrepancy of a set system $\mathcal{S}$ with a bounded primal shatter function, we need to find suitable set system $\mathcal{F}$ with not too many sets, such that for any $S \in \mathcal{S}$ there exists an $F = F_S \in \mathcal{F}$ with $|S \triangle F_S|$ small, smaller than some parameter k. Then we set $\mathcal{M} = \{S \setminus F_S;\, S \in \mathcal{S}\} \cup \{F_S \setminus S;\, S \in \mathcal{S}\}$ and apply Lemma 3.4; for the resulting partial coloring we have, for each $S \in \mathcal{S}$, $|\chi(S)| \le |\chi(F_S)| + |\chi(S \setminus F_S)| + |\chi(F_S \setminus S)| = O(\sqrt{k \log n})$. Since we have the Packing lemma 2.4 at our disposal, we can simply choose $\mathcal{F}$ as a maximal collection of sets in $\mathcal{S}$ such that any two have symmetric difference larger than k. Then $|\mathcal{F}| = O((n/k)^d)$, and we have to set k in such a way that (1) holds. Calculation leads to the bound $\mathrm{disc}(\mathcal{S}) = O(n^{1/2-1/2d}(\log n)^{1/2+1/2d})$.

This is nearly the correct bound; the next improvement is based on an idea of Spencer [Spe85]. Namely, the set system $\mathcal{F}$ need not be colored perfectly (with zero discrepancy); it suffices to color it with about the same discrepancy as the sets of $\mathcal{M}$, and this observation allows us to weaken the condition (1) somewhat. Technically, one incorporates this observation into the proof via estimating the entropy of a suitable vector function of a random coloring χ (see [AS93][4]). As a final trick, instead of just two auxiliary set systems, $\mathcal{F}$ and $\mathcal{M}$, we define a whole sequence of such set systems, say $\mathcal{S}_0, \mathcal{S}_1, \ldots, \mathcal{S}_q$ with $q \approx \log n$. The $\mathcal{S}_i$ has roughly 2^{id} sets of size at most $n/2^i$, and each set of $\mathcal{S}$ can be written as $S = (\ldots((((A_1 \cup A_2) \setminus B_2) \cup A_3) \setminus B_3) \cup \ldots \cup A_q) \setminus B_q$, where $A_i, B_i \in \mathcal{S}_i$, the unions are disjoint and the subtracted sets B_i are fully contained in the sets they are subtracted from. For each $\mathcal{S}_i$, we then prescribe a suitable discrepancy bound Δ_i, and show that there exists a partial coloring under which each $\mathcal{S}_i$ has discrepancy at most Δ_i; see [Mat95b] for details.

Another problem in discrepancy theory concerns the discrepancy of the set system formed by all arithmetic progressions on the set $\{1, 2, \ldots, n\}$. Roth [Rot64] proved a lower bound of $const.n^{1/4}$. The upper bound was improved several times (e.g, by Beck [Bec81] to $O(n^{1/4}\log^3 n)$), and [MS96] proves a tight upper bound of $O(n^{1/4})$ using a part of the technique from the proof of Theorem 3.2(ii). The first partial coloring for the set system of arithmetic progressions is constructed in nearly the same way as for a set system with a quadratic primal shatter function, but for the subsequent iterations one has to work somewhat more since here the situation is not hereditary (the structure of subsystems induced by various subsets may be more complicated).

Another application of similar methods led to improved upper bounds for the following discrepancy-type problem: For a given $\varepsilon > 0$, approximate the unit ball B in $\mathbb{R}^d$ (considering d fixed) with error at most ε by a convex polytope A of the form $A = \{x_1 + x_2 + \cdots + x_n;\, x_1 \in I_1, \ldots, x_n \in I_n\}$, where $I_1, \ldots, I_n$ are segments in $\mathbb{R}^d$ (such a polytope is called a *zonotope*), with the number

[4]Results similar to Spencer's [Spe85] were also obtained by Gluskin [Glu89] via Minkowski's lattice point theorem and by Giannopoulos [Gia95] using the Gaussian measure.

n of summands as small as possible. This problem has several interpretations; one of them deals with a numerical integration of a special class of functions over the unit sphere S^d, and another can be formulated as a "tomography" problem. Also generalizations have been considered where one approximates a general zonoid (a convex body approximable by zonotopes arbitrarily precisely) by a zonotope with possibly few summands. Lower bounds for this problem have been established by Bourgain, Lindenstrauss, and Milman [BLM89] (using harmonic analysis). Nearly tight upper bounds for the ball approximation were given by Bourgain and Lindenstrauss [BL88],[BL93], and an application of the methods described above led to tightening these bounds and generalizing them for arbitrary zonoids [Mat96b].

4 Derandomizing geometric algorithms

The topics of this and the subsequent section (geometric range searching) belong to the field of *computational geometry*. This is a branch of theoretical computer science considering the design and analysis of efficient algorithms for computing with configurations of simple geometric objects in a Euclidean space. The space dimension is usually considered a constant (many problems are studied mainly in the plane or in the 3-dimensional space); this distinguishes computational geometry e.g., from geometric aspects of combinatorial optimization (where the dimension is usually comparable to the number of objects).

Computational geometry as a subject constituted sometimes around the year 1980. In the period until, say, 1986, a number of basic algorithmic problems were solved (particularly in the plane), and also it was clearly demonstrated that for good algorithmic results one needs to understand the combinatorial properties of geometric configurations. Since that time, combinatorial and computational geometry influence each other substantially. The status of computational geometry around that time is captured in monographs [PS85], [Ede87].

The algorithms described in these books are, with minor exceptions, *deterministic*. A subsequent more extensive introduction of *randomized algorithms* into computational geometry worked almost as a miraculous medicine. Methods elaborated in the period (approximately) 1986–1989 provide, relatively easily, efficient probabilistic algorithms for problems which seemed hopelessly difficult for previous methods. Even for problems with previously known optimal deterministic algorithms are the probabilistic algorithms simpler, practically more efficient and easier to implement in most cases. A monograph picturing computational geometry in this new look is [Mul94].

In connection to this development, a question arose naturally, namely to what extent are the probabilistic algorithms more powerful than the deterministic ones. In other words: Can one eliminate the use of randomness from a given probabilistic algorithm, possibly preserving the asymptotic efficiency or making only a little worse? This *derandomization* was of course investigated, approximately in the same period, also in other areas of theoretical computer science. In computational geometry, specific techniques have been obtained, which yield, for most of the randomized algorithms studied there, deterministic algorithms with the same or only slightly larger asymptotic complexity. A detailed survey on this subject is [Mat96a], and here we restrict ourselves to few remarks.

A randomized algorithm can be viewed as a deterministic algorithm which is allowed to read a finite sequence x of random bits from some special device. Usually, we know that for an overwhelming majority of sequences $x \in \{0,1\}^n$, the algorithm works well, and the task of derandomization is to compute at least one good x deterministically. Two general techniques for this purpose go by the names "method of conditional probabilities" and "(approximate) k-wise independence". The former technique sets the bits of x one by one, depending on the value of suitable conditional expectations; it can be viewed as a "binary search" in the probability space $\{0,1\}^n$. The latter technique tries to replace the (exponentially large) probability space $\{0,1\}^n$ of random sequences by a suitable smaller probability space Ω, such that the algorithm is still guaranteed to work well on average if its random bit sequence x is chosen from Ω, but Ω is small enough so that it can be searched exhaustively. Both these general ideas have many variations and subtleties in applications (see e.g., [MR95] for references).

In computational geometry, most algorithms can be easily derandomized by the method of conditional probabilities in its basic form; the running time increases by a factor which is polynomial but usually quite large [CF90]. Methods specific to computational geometry yield much better asymptotic complexities.

Let us consider a specific algorithmic problem: P is a given set of n points in the plane, L is a given set of n lines in the plane, and we want to decide whether any of the points of P lies on any of the lines of L (this is so-called *Hopcroft's problem*; its algorithmic complexity is not known, but it is suspected to be of the order $n^{4/3}$). The known asymptotically efficient algorithms for this problem all use the following strategy ("geometric divide-and-conquer"). They divide the plane into some suitable number, m, of regions $\Delta_1, \Delta_2, \ldots, \Delta_m$ (usually the Δ_i are triangles, possibly unbounded ones). Let P_i be the subset of P lying in Δ_i and L_i the set of lines of L intersecting Δ_i. In this way, the original problem is subdivided into m subproblems (it suffices to check whether for any i, a point of P_i lies on a line of L_i), and these subproblems are then solved recursively or by another method. We will not go into any further details of the algorithm here; we concentrate on the part constructing the regions Δ_i. For algorithm's efficiency it is crucial that the sets L_i be possibly small. To avoid some technicalities, we let L_i be only the lines intersecting the *interior* of Δ_i (we suppose that the lines only touching the boundary of Δ_i are somehow handled separately). We introduce the following definition: We say that regions $\Delta_1, \ldots, \Delta_m$ form a $(1/r)$-*cutting for* L if they cover the whole plane and the interior of each Δ_i is intersected by at most $|L|/r$ lines of L.

One of good constructions of a $(1/r)$-cutting by random sampling is the following [HW87], [Cla87]. We fix a suitable number s and choose a random sample $S \subset L$ of s lines from L. The lines of S divide the plane into convex polygons. We further subdivide these polygons into triangles; an easy calculation shows that we get $O(s^2)$ triangles in total, and these will be our Δ_i's. Next, we consider the set system $\mathcal{L}$ on L, consisting of all sets of the form L_T, where T is some triangle and L_T is the set of lines of L intersecting the interior of T. The set system $\mathcal{L}$ has a constant-bounded VC-dimension, and hence our random sample S of s lines is, with high probability, an ε-net for $\mathcal{L}$ by Theorem 2.3, with $\varepsilon = C \log s/s$ for an appropriate constant C. This means that any open triangle intersected by no line of S is intersected by $O((n/s)\log s)$ lines of L only. In particular, since the interiors of the Δ_i's are not intersected by lines of S, we get $|L_i| = O((n/s)\log s)$ for all i. Therefore, we get a $(1/r)$-cutting for L, with $r = const.s/\log s$, consisting of $O(r^2 \log^2 r)$ triangles. It turns out that this construction is nearly best possible (up to the logarithmic factor; see below).

One way of making the just described construction deterministic is to compute an ε-net for the set system $\mathcal{L}$ deterministically. An algorithm for deterministic computation of ε-nets was given in [Mat95a]; the key point is that one actually computes with ε-approximations (which have some more pleasant algebraic properties than ε-nets) and in the end, the desired ε-net is computed from an ε-approximation. The algorithm works for arbitrary set system $\mathcal{S}$ of bounded VC-dimension on a finite set X, provided that the set system $\mathcal{S}$ is given by a *subsystem oracle*. This means that given a subset $A \subseteq X$, we can compute the system $\mathcal{S}|_A$ (in the form of an incidence matrix) in time $O(|A|^{d+1})$, where d is a constant called the *dimension* of the subsystem oracle. The current most efficient version of the algorithm was given in [BCM93]. For a fixed oracle dimension d, it can compute a $(1/r)$-net for $\mathcal{S}$ of size $O(r \log r)$ in time $O(|X| r^d \log^d r)$; in particular, if r is a constant, a $(1/r)$-net is found in time linear in the cardinality of X. This result allows one to make most of divide-and-conquer algorithms in computational geometry deterministic with only an $O(n^\delta)$ loss in asymptotic efficiency, where $\delta > 0$ is an arbitrarily small constant.

It can also be used to derandomize the above probabilistic construction of a $(1/r)$-cutting for a given set of n lines. However, for the construction of $(1/r)$-cuttings in the plane and in higher dimensions, special methods have been developed to make it still more efficient. Let us recall that the above-sketched randomized construction gave a $(1/r)$-cutting consisting of $O(r^2 \log^2 r)$ triangles. It is not difficult to show that a $(1/r)$-cutting for n lines in the plane has to consist of at least $const.r^2$ regions (count the total number of polygons into which the lines of L divide the plane, and compare it with the maximum number of polygons incident to a single region of a $(1/r)$-cutting). On the other hand, $(1/r)$-cuttings of size $O(r^2)$ actually exist [CF90]. Let us formulate a

d-dimensional version of this result:

4.1 Theorem (Cutting lemma) *Given a set H of n hyperplanes in $\mathbb{R}^d$ (d considered constant) and a parameter $r > 1$, there exist a $(1/r)$-cutting of size $O(r^d)$ for H, that is, a collection $\Delta_1, \ldots \Delta_m$ of (possibly unbounded) simplices covering $\mathbb{R}^d$, such that $m = O(r^d)$ and the interior of each Δ_i is intersected by at most n/r hyperplanes of H [CF90]. Such a $(1/r)$-cutting can be deterministically computed in time $O(nr^{d-1})$ [Cha93], and if $r < n^\alpha$ for a suitable $\alpha = \alpha(d) > 0$ then even in $O(n \log r)$ time [Mat92].*

The $(1/r)$-cuttings have found also various non-algorithmic applications; one is in discrepancy theory [Mat96b], few others will be mentioned in section 6.

5 Geometric range searching

Here we consider a specific class of algorithmic problems, which is quite important for the design of various geometric algorithms.

As a prototype question, consider the following *triangle range counting* problem. Let P be a given n-point set in the plane. We want an algorithm which, given a triangle T, determines the number of points of P lying in T quickly. The point set P is given in advance, and we can prepare some auxiliary information about it and store it in a suitable data structure. Then we will be repeatedly given triangles T as queries. Assuming that the number of queries will be large, it will be advantageous to invest some computing time into building the data structure (this phase is called the *preprocessing*) if this makes the query answering faster.

This problem can of course be generalized in many directions. We may consider a higher-dimensional space $\mathbb{R}^d$ instead of the plane, and use sets from some class $\mathcal{S}$ of simple geometric shapes in $\mathbb{R}^d$ instead of triangles as queries (natural choices are axis-parallel boxes, simplices, balls, etc.). Also the type of a query may be different from point counting: We may require a list of all points of P lying in a query range, or we can only ask if the query range contains any point of P at all, or each point of P can be assigned some real weight and we can be interested in the maximum weight of a point in a given range, etc.

In general, we assume that every point $p \in P$ is assigned a weight $w(p) \in \Sigma$, where $(\Sigma, +)$ is some commutative semigroup (common to all the points). The objective of a query is to find the sum of weights of all points of P lying in a given range $S \in \mathcal{S}$, that is, $\sum_{p \in S \cap P} w(p)$. For example, for counting queries, $(\Sigma, +)$ will be the natural numbers with addition, and all weights will be equal to 1. For queries on maximum weight, the appropriate semigroup will be the real numbers with the operation of taking a maximum of two numbers.

It turns out that for many classes $\mathcal{S}$ of ranges (e.g., all the examples named above), one can achieve a query time $O(\log n)$, where n is the number of points in P. In most cases, however, this fast query answering requires a huge amount of memory and preprocessing time. The problem is thus to achieve a reasonable compromise between storage and preprocessing on one side and query time on the other side. Mathematically most interesting and most studied was the case when we insist that the data structure is stored in memory space $O(n)$ or close to $O(n)$, and we ask what can be saved in the query time compared to the trivial $O(n)$ solution (store the points without any preprocessing and, given a query range, examine each point in turn). From a practical point of view, unfortunately, it seems that except for several simple low-dimensional situations ($d = 1, 2$, say), the results are disappointing — the possible savings in query time with a linear space are meager. Theoretically, though, the problem led to nice mathematical developments, mainly in attempts to prove lower bounds.

For definiteness, let us consider triangle range searching with $O(n)$ memory space, where the points of P have weights from some semigroup S, and the query is to find the sum of weights of points in a given triangle. The most interesting aspects of known results already appear in this simple situation. Moreover, various more complicated-looking range searching problems can be often reduced to searching with simplices by a suitable transformation. A more comprehensive information can be found, e.g., in [Mat95c], [Mul94], [PA95].

Lower bounds. As shown by Chazelle [Cha89], an algorithm for answering queries with triangles and working for arbitrary semigroup weights must essentially operate as follows: In the preprocessing phase, weights are pre-computed for a suitable collection $\mathcal{C} = \{C_1, C_2, \ldots, C_m\}$ of subsets of the set P; here m is a lower bound on the memory space used by the algorithm. When answering a query with a triangle T, the set $P \cap T$ is expressed as a disjoint union of some number, k, of sets of the system $\mathcal{C}$, and the total weight of points lying in T is computed by summing the weights of these k sets of $\mathcal{C}$. The maximum value of k (over all query triangles) is a lower bound for the worst-case time complexity of a query. (This is admittedly inexact, since a precise statement is somewhat technical.) The sets in $\mathcal{C}$ are called the *canonical sets* (for the particular algorithm and set P).

Chazelle [Cha89] shows (among others) that if we allow only $O(n)$ canonical subsets in $\mathcal{C}$, then the number of sets in the decomposition, k, must be at least $\Omega(\sqrt{n})$ for some query triangle. The proof requires a construction of a suitable point set P and a suitable finite set $\mathcal{T}_0$ of triangles, such that the bipartite incidence graph between P and $\mathcal{T}_0$ has sufficiently many edges, but it contains no large complete bipartite subgraphs (which means that no large canonical set $C \in \mathcal{C}$ can contribute to the decomposition of many triangles from $\mathcal{T}_0$). In Chazelle's proof, P is chosen as a uniformly distributed set in the unit square, and $\mathcal{T}_0$ is chosen as a random set of strips of width approximately $n^{-1/2}$ intersecting the unit square.

For dimension $d > 2$, the expected generalization would be that with $O(n)$ canonical sets, we need at least about $n^{1-1/d}$ sets in the decomposition.

Chazelle's proof only gives an $n^{1-1/d}/\log n$ lower bound, and the possibility of improving the lower bound using the same method has an interesting relation to a generalization of a famous problem of combinatorial geometry, the so-called Heilbronn problem. The Heilbronn problem itself can be formulated as follows: *For a set $P \subset \mathbb{R}^2$, let $a(P)$ denote the area of a smallest triangle with vertices in P. What is the asymptotic behavior of the function $a(n) = \sup\{a(P);\ P \subset [0,1]^2, |P| = n\}$?* Despite a considerable attention paid to this nice problem (see [KSP82], [Rot76] for a survey and references), a complete solution still seems to be quite far. The following is a generalization related to Chazelle's lower bound proof method: *For a set $P \subset \mathbb{R}^d$, denote by $a_d(P,j)$ the smallest volume of the convex hull of a j-tuple of points of P. What is the asymptotic behavior of the function $a_d(n,j) = \sup\{a_d(P,j);\ P \subset [0,1]^d, |P| = n\}$?* Chazelle has shown that for a suitably chosen set P in the unit cube the volume of the convex hull of each j-tuple is at least proportional to j/n for any $j \geq c \log n$, with a sufficiently large constant c. In other words,

$$a_d(n,j) = \Omega(j/n) \tag{2}$$

for $j \geq c \log n$. This is essentially a result about uniform distribution of the set P, saying that no j-tuple is too clustered (in the sense of volume). If (2) could also be proved for smaller j, an improvement of the lower bound for decomposing sets in triangles would follow immediately. Such an improvement may not be easy. Known results for the Heilbronn problem imply that (2) is false for $d = 2$, $j = 3$ (which may contradict intuition somewhat). However, it is not known that (2) could not hold for larger but constant j.

An algorithm for triangle range searching with $O(n)$ space and $O(\sqrt{n})$, which matches Chazelle's lower bound, was given in [Mat93], and no asymptotically better algorithm is known even for various special cases of the problem. On the other hand, the situation is not as satisfactory as it might seem, since the lower bound was shown under the condition that subtraction of point weights is not allowed in the algorithm. This is a somewhat unnatural restriction (say for the case of counting the points in a query range), although it is not known how to actually use weight subtraction to get a faster algorithm. Hence it would be desirable to have lower bounds also for stronger models of computation. Some progress was made by Chazelle again [Cha94]; he shows that n queries together require at least $\Omega(n \log n)$ operations even if subtractions are allowed. His proof uses eigenvalue estimates for a suitable matrix based on known lower bounds for the combinatorial discrepancy of rectangles. Other papers considering lower bounds for geometric range searching and leading to very interesting geometric problems are [BCP93] and [Eri96].

Upper bounds. Let us begin with a simple but nontrivial one-dimensional range-searching example, namely with intervals on the real line. Let $P \subset \mathbb{R}$ be an n-point set, and let $\mathcal{I}$ be the family of all intervals in $\mathbb{R}$. For the purpose of canonical decompositions of sets of the form $P \cap I$, $I \in \mathcal{I}$, we can as well assume that $P = \{0, 1, \ldots, n-1\}$. Then a suitable system of canonical sets (subintervals of P, in this case) is $\{\{q2^k, q2^k+1, \ldots, (q+1)2^k-1\};\ k = 0, 1, \ldots, \lfloor \log_2 n \rfloor, q = 0, 1, \lfloor n/2^k \rfloor\}$. This set system appears very often in mathematics and in computer science under various names and disguises (for instance, it corresponds to nodes of a complete binary tree with leaves $0, 1, \ldots, n-1$). It has $O(n)$ sets, and it is easy to see that any interval on P can be partitioned into $O(\log n)$ canonical sets.

Let us now consider range searching with triangles in the plane. Spanning trees with low crossing number were originally invented exactly for this purpose [Wel88]. Since the dual shatter function for triangles in the plane is quadratic, Theorem 2.5 guarantees the existence of a spanning path with crossing number $O(\sqrt{n})$ on our given n-point set P. Geometrically, the crossing number condition means that the boundary of a query triangle T cuts this path into $O(\sqrt{n})$ pieces, so that each piece lies either completely inside T or completely outside T. Hence the set $P \cap T$ can be expressed as a disjoint union of $O(\sqrt{n})$ intervals along the path, and this reduces the triangle range searching problem into $O(\sqrt{n})$ one-dimensional queries with intervals on the path. These one-dimensional queries can be handled in a manner sketched above; hence we obtain a decomposition of $P \cap T$ into $O(\sqrt{n} \log n)$ canonical sets. To make this method into an actual algorithm, we have to specify how to find the edges of the spanning path crossed by the query triangle. In the plane, this algorithmic problem can be solved satisfactorily, as well as an analogous problem for tetrahedra in dimension 3, but in dimension 4 and higher it is not known how to use spanning trees with low crossing numbers efficiently in an algorithm (although they still provide good decompositions into canonical sets).

Let us sketch one more approach to triangle range searching; we demonstrate it on the simpler case of halfplane searching (presenting the geometric ideas only and omitting all algorithmic details). The approach was suggested by Willard [Wil82]; it was the first known algorithm for triangle range searching with a near-linear space and a sublinear query time. A geometric construction of the following type is required: Given an n-point set $P \subset \mathbb{R}^2$, partition the plane into some number, a, of regions, in such a way that each region contains roughly n/a points of P, and no line intersects more than b of the regions, for some $b < a$. The simplest such construction is to take two lines such that each of the four "quadrants" R_1, R_2, R_3, R_4 determined by them contains about $n/4$ points of P (the first line can be chosen as any line slicing P into two equal-size parts, and the second line is then found using the so-called *ham-sandwich theorem*). Here we have $a = 4$, $b = 3$. Let $P_i = P \cap R_i$ be the points in the ith region. We declare the four sets $P_1, \ldots, P_4$ as canonical sets, and we apply the same construction recursively for each P_i, until one-point sets are reached. This procedure produces $O(n)$ canonical sets in total.

We claim that for any halfplane H, the set $H \cap P$ can be expressed as a disjoint union of $O(n^\alpha)$ canonical sets thus defined, where $\alpha = \log_4 3 \approx 0.774$.

Indeed, since the boundary of H crosses only 3 of the 4 regions R_i, one of the sets P_i lies either completely outside H (then we may ignore it) or completely within H, and then we use it as one of the canonical sets in the decomposition. If $f(n)$ denotes the maximum number of canonical sets in a decomposition used by this method for an n-point set P, we get the recurrence $f(n) \leq 1+3f(n/4)$ (ignoring some rounding to integers), and this leads to the claimed bound. In general, having a construction with a regions, at most b of them intersected by a line, we get a bound of $O(n^\alpha)$ with $\alpha = \log_a b$. To get a good construction of this type by ad hoc methods was not easy, and the exponent has been improved many times (we remark that the paper [HW87] which introduced the key notion of ε-nets was one of the attempts in this direction!). A near-optimal and relatively simple method of this type was given in [Mat92]. We formulate the main geometric result for an arbitrary dimension d. First, a definition: A *simplicial partition* of a finite set $P \subset \mathbb{R}^d$ is a collection $\Pi = \{(P_1, \Delta_1), \ldots, (P_m, \Delta_m)\}$, where the P_i are disjoint subsets of P (called the *classes*) forming a partition of P, and each Δ_i is a simplex containing the set P_i.

5.1 Theorem *Let P be an n-point set in $\mathbb{R}^d$ ($d > 1$ considered as a constant), and let r be a parameter, $1 < r \ll n$. Then there exists a simplicial partition Π of P, whose classes have sizes between n/r and $2n/r$ (hence there are $O(r)$ classes), and such that any hyperplane crosses at most $O(r^{1-1/d})$ simplices of Π (a hyperplane h crosses a simplex Δ if Δ intersects both the open halfspaces bounded by h).*

In particular, for $d = 2$, we get a partition into about r subsets such that any line crosses at most $O(\sqrt{r})$ of their respective triangles. In the above notation, we thus have $a \approx r$, $b \approx const.\sqrt{r}$, so that $\log_a b \to \frac{1}{2}$ for $r \to \infty$, and we get close to the optimal $O(\sqrt{n})$ query time.

Theorem 5.1 is a geometric generalization of Theorem 2.5 on spanning trees with low crossing numbers. Indeed, Theorem 2.5 also provides a partition of the given set into pairs such that no set crosses too many pairs; in Theorem 5.1 we have chunks of about n/r points instead of pairs. Also the proof is quite similar to the proof of Theorem 2.5; while in the spanning tree construction, we select "short edges" one by one, in the construction of a simplicial partition one selects "small chunks" (ones crossed by few hyperplanes from a suitable "test set"). The existence of these small chunks at every step, i.e. an analogue of Lemma 2.6, is proved using $(1/r)$-cuttings (defined in the preceding section). As was shown by Alon et al. [AHW87], an analogue of Theorem 5.1 doesn't hold for general set systems of bounded VC-dimension; at least for r bounded by a constant no nontrivial "partition scheme" exists in general.

6 Combinatorial geometry applications

A large part of combinatorial geometry deals with properties of geometrically defined set systems. Here we select just few applications of the notions and tools discussed above. We refer e.g., to the book [PA95] for more material.

Pach [Pac91] applies spanning trees with a low crossing number to give a simple proof of a result of Aronov et al. [AEG$^+$91]: For any set of n points in the plane, one can choose $const.\sqrt{n}$ disjoint pairs of these points such that any two of the segments determined by these pairs intersect. It is a nice open problem whether the bound of $const.\sqrt{n}$ can be replaced by $const.n$.

Haussler and Welzl [HW87] introduced, besides ε-nets, also *weak ε-nets*. The following particular geometric case of this notion found surprising applications: Let $A \subset \mathbb{R}^d$ be a finite point set; a set $N \subset \mathbb{R}^d$ is a *weak ε-net for A with respect to convex sets* if $N \cap C \neq \emptyset$ for any convex set C with $|C \cap A| > \varepsilon|A|$. This looks similar to the definition of ε-net, but the difference is that here N may consist of any points of $\mathbb{R}^d$ and is not restricted to points of A (as would be the case for an ε-net). While ε-nets of a bounded size for convex sets do not exist in general, weak ε-nets do exist. The current best bounds for the size of a weak $(1/r)$-net for convex sets are $O(r^2)$ in the plane [ABFK92] and $O\left(r^d (\log r)^{b_d}\right)$ for a suitable constant b_d in $\mathbb{R}^d$, $d > 2$ [CEG$^+$95] (see also [BC95] for related results). The truth is probably much smaller, maybe around $r^{\lfloor d/2 \rfloor}$. Weak ε-nets for convex sets have some connections to work on the k-set problem mentioned in the introduction [ABFK92], and they have been applied by Alon and Kleitman [AK92] as one of three key ingredients in a beautiful solution to the Hadwiger-Debrunner (p, q)-problem.

The $(1/r)$-cuttings (defined in section 4) and various their generalizations have been used in problems of bounding the number of incidences, complexity of cells in arrangements, etc. A prototype problem of this kind is the following question: Consider a set P of n points and a set L of m lines in the plane. What is the maximum possible number of their *incidences*, i.e. pairs (p, ℓ) such that $p \in P$, $\ell \in L$ and $p \in \ell$? A tight upper bound, $O(n^{2/3}m^{2/3} + m + n)$, was proved by Szemerédi and Trotter [ST83]. A considerably simpler proof and various generalizations have been given by Clarkson et al. [CEG$^+$90] using methods related to $(1/r)$-cuttings; other similar results are e.g., in [EGS90], [CEGS89], [AEGS92]. The approach via $(1/r)$-cuttings is one of standard methods in proving combinatorial complexity bounds for geometric configurations by now. For numerous other methods and more information about the extensive realm of combinatorial complexity of arrangements in a fixed dimension we refer to a recent monograph [AS95]. Recently, a yet much simpler proof of the Szemerédi-Trotter theorem and of some related results was given by Székely [Szé95] by a different approach, but so far the approach via $(1/r)$-cuttings has its advantages (it generalizes to higher-dimensional questions, and it usually provides efficient algorithms).

Let us sketch a proof of the Szemerédi-Trotter bound using $(1/r)$-cuttings. Denote the number of incidences for specific sets L of lines and P of points by $I(L,P)$, and let $I(n,m)$ be the maximum of $I(L,P)$ over all choices of an n-element L and an m-element P. First consider the bipartite graph on vertex set $L \cup P$ with edges corresponding to incidences. No two lines share two points and hence this graph has no $K_{2,2}$ subgraph. Well-known extremal graph theory result then implies

$$I(n,m) = O(\min(n\sqrt{m}+m, m\sqrt{n}+n)). \tag{3}$$

This bound is already asymptotically tight if $m \geq n^2$ or $n \geq m^2$, i.e. if the number of lines and points is highly "unbalanced". Let us treat the "balanced" case; for simplicity let $m = n$ and let n-element sets L and P be given. With foresight, we set $r = n^{1/3}$ and by Theorem 4.1 we find a $(1/r)$-cutting $\Delta_1, \ldots, \Delta_t$ for L with $t = O(r^2) = O(n^{2/3})$. Let P_i denote the points of P lying in Δ_i except for its vertices, and let L_i be the lines of L intersecting the interior of Δ_i. We have

$$\begin{array}{rcll} I(L,P) & \leq & \sum_{i=1}^{t} I(L_i, P_i) & (A) \\ & & +I(n, 3t) & (B) \\ & & +I(3t, n)\,. & (B') \end{array}$$

The term (B) covers the incidences of points lying at vertices of some Δ_i. The term (B') covers all cases when a line of L contains an edge of a triangle Δ_i but does not intersect its interior. By (3), terms (B) and (B') are both $O(t\sqrt{n}+n) = O(n^{7/6})$. Concerning the main term (A), we use $|L_i| \leq n/r = n^{2/3}$ and $\sum_{i=1}^{t} |P_i| \leq 2n$ (since each point of P goes into at most two P_i). An average P_i has about $n^{1/3}$ points while L_i has about $n^{2/3}$ lines, so the subproblems are unbalanced enough for (3) to be tight. More precisely, applying (3) for each $I(L_i, P_i)$ we get $\sum_{i=1}^{t} |I(L_i, P_i)| \leq \sum_{i=1}^{t} O(|P_i| n^{1/3} + n^{2/3}) = O(n^{4/3})$, which proves the $n = m$ case of the Szemerédi-Trotter theorem.

References

[AB92] M. Anthony and N. Biggs. *Computational Learning Theory.* Cambridge University Press, Cambridge, 1992.

[ABCBH93] N. Alon, S. Ben-David, N. Cesa-Bianchi, and D. Haussler. Scale-sensitive dimension, uniform convergence and learnability. In *Proc. 34th IEEE Symposium on Foundations of Computer Science*, pages 292–300, 1993.

[ABFK92] N. Alon, I. Bárány, Z. Füredi, and D. Kleitman. Point selections and weak ε-nets for convex hulls. *Combin., Probab. Comput.*, 1(3):189–200, 1992.

[AEG+91] B. Aronov, P. Erdős, W. Goddard, D. J. Kleitman, M. Klugerman, J. Pach, and L. J. Schulman. Crossing families. In *Proc. 7th Annu. ACM Sympos. Comput. Geom.*, pages 351–356, 1991.

[AEGS92] B. Aronov, H. Edelsbrunner, L. J. Guibas, and M. Sharir. The number of edges of many faces in a line segment arrangement. *Combinatorica*, 12(3):261–274, 1992.

[AHW87] N. Alon, D. Haussler, and E. Welzl. Partitioning and geometric embedding of range spaces of finite Vapnik-Chervonenkis dimension. In *Proc. 3rd Annu. ACM Sympos. Comput. Geom.*, pages 331–340, 1987.

[AK92] N. Alon and D. Kleitman. Piercing convex sets and the hadwiger debrunner (p, q)-problem. *Adv. Math.*, 96(1):103–112, 1992.

[Ale90] R. Alexander. Geometric methods in the theory of uniform distribution. *Combinatorica*, 10(2):115–136, 1990.

[AS93] N. Alon and J. Spencer. *The probabilistic method*. J. Wiley & Sons, 1993.

[AS95] P. K. Agarwal and M. Sharir. *Davenport-Schinzel Sequences and Their Geometric Applications*. Cambridge University Press, Cambridge, UK, 1995.

[Ass83] P. Assouad. Densite et dimension. *Ann.Inst.Fourier (Grenoble)*, 33:233–282, 1983.

[AT89] K. S. Alexander and M. Talagrand. The law of the iterated logarithm for empirical processes on vapnik- cervonenkis classes. *J. Multivariate Anal.*, 30:155–166, 1989.

[BC87] J. Beck and W. Chen. *Irregularities of distribution*. Cambridge University Press, 1987.

[BC95] P. G. Bradford and V. Capoyleas. Weak ϵ-nets for points on a hypersphere. Report MPI-I-95-1-029, Max-Planck-Institut Inform., Saarbrücken, Germany, 1995.

[BCM93] H. Brönnimann, B. Chazelle, and J. Matoušek. Product range spaces, sensitive sampling, and derandomization. In *Proc. 34th Annu. IEEE Sympos. Found. Comput. Sci.*, pages 400–409, 1993.

[BCP93] H. Brönnimann, B. Chazelle, and J. Pach. How hard is halfspace range searching. *Discrete Comput. Geom.*, 10:143–155, 1993.

[Bec81] J. Beck. Roth's estimate on the discrepancy of integer sequences is nearly sharp. *Combinatorica*, 1(4):319–325, 1981.

[Bec84] J. Beck. Some upper bounds in the theory of irregularities of distribution. *Acta Arith.*, 43:115–130, 1984.

[BEHW89] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. Warmuth. Classifying learnable geometric concepts with the Vapnik-Chervonenkis dimension. *J. ACM*, 36:929–965, 1989.

[BL88] J. Bourgain and J. Lindenstrauss. Distribution of points on the sphere and approximation by zonotopes. *Isr. J. Math.*, 64:25–31, 1988.

[BL93] J. Bourgain and J. Lindenstrauss. Approximating the ball by a Minkowski sum of segments with equal length. *Discr. & Comput. Geom.*, 9:131–144, 1993.

[BLM89] J. Bourgain, J. Lindenstrauss, and V. Milman. Approximation of zonoids by zonotopes. *Acta Math.*, 162:73–141, 1989.

[BPR] S. Basu, R. Pollack, and M.-F. Roy. On the number of cells defined by a family of polynomials on a variety. *Mathematika*. To appear. Preliminary version in *Proc. 35th Annual Symposium on Foundations of Computer Science*, pages 632-641, Santa Fe, 1994.

[BS95] J. Beck and V. Sós. Discrepancy theory. In *Handbook of Combinatorics*, pages 1405–1446. North-Holland, Amsterdam, 1995.

[BVS+92] A. Björner, M. Las Vergnas, B. Sturmfels, N. White, and G. M. Ziegler. *Oriented matroids (Encyclopedia of Mathematics).* Cambridge University Press, 1992.

[CEG+90] K. Clarkson, H. Edelsbrunner, L. Guibas, M. Sharir, and E. Welzl. Combinatorial complexity bounds for arrangements of curves and spheres. *Discrete Comput. Geom.*, 5:99–160, 1990.

[CEG+95] B. Chazelle, H. Edelsbrunner, M. Grigni, L. Guibas, M. Sharir, and E. Welzl. Improved bounds on weak ϵ-nets for convex sets. *Discrete Comput. Geom.*, 13:1–15, 1995.

[CEGS89] B. Chazelle, H. Edelsbrunner, L. J. Guibas, and M. Sharir. Lines in space: combinatorics, algorithms, and applications. In *Proc. 21st Annu. ACM Sympos. Theory Comput.*, pages 382–393, 1989.

[CF90] B. Chazelle and J. Friedman. A deterministic view of random sampling and its use in geometry. *Combinatorica*, 10(3):229–249, 1990.

[Cha89] B. Chazelle. Lower bounds on the complexity of polytope range searching. *J. Amer. Math. Soc.*, 2:637–666, 1989.

[Cha93] B. Chazelle. Cutting hyperplanes for divide-and-conquer. *Discrete Comput. Geom.*, 9(2):145–158, 1993.

[Cha94] B. Chazelle. A spectral approach to lower bounds. In *Proc. 35th Ann. IEEE Symp. on Foundations of Computer Science*, pages 674–682, 1994.

[Cla87] K. L. Clarkson. New applications of random sampling in computational geometry. *Discrete Comput. Geom.*, 2:195–222, 1987.

[Cla93] K. L. Clarkson. Algorithms for polytope covering and approximation. In *Proc. 3rd Workshop Algorithms Data Struct.*, volume 709 of *Lecture Notes in Computer Science*, pages 246–252, 1993.

[CW89] B. Chazelle and E. Welzl. Quasi-optimal range searching in spaces of finite VC-dimension. *Discrete Comput. Geom.*, 4:467–489, 1989.

[DE94] T. K. Dey and H. Edelsbrunner. Counting triangle crossings and halving planes. *Discrete Comput. Geom.*, 12:281–289, 1994.

[DHS94] M. Kearns D. Haussler and R. Schapire. Bounds on the sample complexity of Bayesian learning using information theory and the VC dimension. *Machine Learning*, 14:84–114, 1994.

[DSW94] G.-L. Ding, P. Seymour, and P. Winkler. Bounding the vertex cover number of a hypergraph. *Combinatorica*, 14:23–34, 1994.

[Dud84] R. M. Dudley. *A course on empirical measures.* Lecture Notes in Math. 1097. Springer, 1984.

[Dud85] R. M. Dudley. The structure of some Vapnik-Chervonenkis classes. In *Proc. of Berkeley Conference in honor of Jerzy Neyman and Jack Kiefer, Volume II, (LeCam and Olshen, Eds.)*, pages 495–507, 1985.

[Ede87] H. Edelsbrunner. *Algorithms in Combinatorial Geometry*, volume 10 of *EATCS Monographs on Theoretical Computer Science*. Springer-Verlag, Heidelberg, West Germany, 1987.

[EGS90] H. Edelsbrunner, L. Guibas, and M. Sharir. The complexity of many cells in arrangements of planes and related problems. *Discrete Comput. Geom.*, 5:197–216, 1990.

[Eri96] J. Erickson. New lower bounds for convex hull problems in odd dimensions. In *Proc. 12th Annu. ACM Sympos. Comput. Geom.*, pages 1–9, 1996.

[Fra83] P. Frankl. On the trace of finite sets. *J. Comb. Theory, Ser. A*, 34:41–45, 1983.

[Gia95] A. Giannopoulos. On some vector balancing problems. Manuscript, 1995.

[Glu89] E. D. Gluskin. Extremal properties of orthogonal parallelpipeds and their applications to the geometry of Banach spaces. *Math. USSR Sbornik*, 64(1):85–96, 1989.

[GW94] B. Gärtner and E. Welzl. Vapnik-Chervonenkis dimension and (pseudo-)hyperplane arrangements. *Discr. Comput. Geom.*, 12:399–432, 1994.

[GZ84] E. Giné and J. Zinn. Some limit theorems for empirical processes. *Annals of Probability*, 12:929–989, 1984.

[Hau92] D. Haussler. Decision theoretic generalizations of the PAC model for neural net and other learning applications. *Information and Computation*, 100(1):78–150, 1992.

[Hau95] D. Haussler. Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik-Chervonenkis dimension. *Journal of Combinatorial Theory Ser. A*, 69:217–232, 1995.

[HW87] D. Haussler and E. Welzl. Epsilon-nets and simplex range queries. *Discrete Comput. Geom.*, 2:127–151, 1987.

[KM95a] G. Kalai and J. Matoušek. Guarding galleries where every point sees a large area. KAM Series (tech. Report) 95–293, Charles University, 1995. To appear in Israel J. Math.

[KM95b] M. Karpinski and A. Macintyre. Polynomial bounds for VC dimension of sigmoidal neural networks. In *Proceedings of the 27th Annual ACM Symposium on Theory of Computing*, pages 200–208, 1995.

[KPW92] J. Komlós, J. Pach, and G. Woeginger. Almost tight bounds for ϵ-nets. *Discrete Comput. Geom.*, 7:163–173, 1992.

[KRS96] J. Kollár, L. Rónyai, and T. Szabó. Norm-graphs and bipartite Turán numbers. *Combinatorica*, 1996. To appear. Extended abstract in *Proc. 28th ACM Symposium on Theory of Computing*.

[KSP82] J. Komlós, E. Szemerédi, and J. Pintz. A lower bound for Heilbronn's problem. *J. London Math. Soc.*, 25(2):13–24, 1982.

[KW93] M. Karpinski and T. Werther. VC dimension and uniform learnability of sparse polynomials and rational functions. *SIAM Journal on Computing*, 22(6):1276–1285, 1993.

[Lit88] N. Littlestone. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine Learning*, 2:285–318, 1988.

[Mat92] J. Matoušek. Efficient partition trees. *Discrete Comput. Geom.*, 8:315–334, 1992.

[Mat93] J. Matoušek. Range searching with efficient hierarchical cuttings. *Discrete Comput. Geom.*, 10(2):157–182, 1993.

[Mat94] J. Matoušek. On discrepancy bounds via dual shatter function. KAM Series (Tech. Report) 94–273, Charles University, Prague, 1994.

[Mat95a] J. Matoušek. Approximations and optimal geometric divide-and-conquer. *J. of Computer and System Sciences*, 50:203–208, 1995.

[Mat95b] J. Matoušek. Tight upper bounds for the discrepancy of halfspaces. *Discr. & Comput. Geom.*, 13:593–601, 1995.

[Mat95c] J. Matoušek. Geometric range searching. *ACM Comput. Surveys*, 26:421–461, 1995.

[Mat96a] J. Matoušek. Derandomization in computational geometry. *J. Algorithms*, 20:545–580, 1996.

[Mat96b] J. Matoušek. Improved upper bounds for approximation by zonotopes. KAM Series (Tech. Report) 96–??, Charles University, Prague, 1996.

[Mil64] J. Milnor. On the betti numbers of real algebraic varieties. *Proc. Amer. Math. Soc.*, 15:275–280, 1964.

[MR95] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

[MS96] J. Matoušek and J. Spencer. Discrepancy in arithmetic progressions. *J. Amer. Math. Soc.*, 9:195–204, 1996.

[Mul94] K. Mulmuley. *Computational Geometry: An Introduction Through Randomized Algorithms*. Prentice Hall, Englewood Cliffs, NJ, 1994.

[MWW93] J. Matoušek, E. Welzl, and L. Wernisch. Discrepancy and ε-approximations for bounded VC-dimension. *Combinatorica*, 13:455–466, 1993.

[Nie92] H. Niederreiter. *Random number generation and quasi-Monte Carlo methods*. CBMS-NSF Regional Conference Series in Applied Mathematics. 63. Philadelphia, PA: SIAM, 1992.

[Ole51] O.A. Oleinik. Estimates of the Betti numbers of real algebraic hypersurfaces (in Russian). *Mat. Sbornik (N. S.)*, 28(70):635–640, 1951.

[PA95] J. Pach and P. K. Agarwal. *Combinatorial Geometry*. John Wiley & Sons, New York, NY, 1995.

[Pac91] J. Pach. Geometric graphs. In J.E. Goodman, R. Pollack, and W. Steiger, editors, *Computational Geometry: papers from the DIMACS special year*. Amer. Math. Soc., 1991.

[Pol90] D. Pollard. *Empirical processes: Theory and applications*. NSF-CBMS Regional Conference Series in Probability and Statistics vol. 2. Institute of Math. Stat. and Am. Stat. Assoc., 1990.

[PS85] F. P. Preparata and M. I. Shamos. *Computational Geometry: An Introduction*. Springer-Verlag, New York, NY, 1985.

[PSS92] J. Pach, W. Steiger, and E. Szemerédi. An upper bound on the number of planar k-sets. *Discrete Comput. Geom.*, 7:109–123, 1992.

[Rot64] K. F. Roth. Remark concerning integer sequences. *Acta Arithmetica*, 9:257–260, 1964.

[Rot76] K. Roth. Developments in the triangle problem. *Adv. Math.*, 22:364–385, 1976.

[RV] K. Romanik and J. Vitter. Using Vapnik-Chervonenkis dimension to analyze the testing complexity of program segments. *Information and Computation*. To appear.

[Sau72] N. Sauer. On the density of families of sets. *Journal of Combinatorial Theory Ser. A*, 13:145–147, 1972.

[She72] S. Shelah. A combinatorial problem, stability and order for models and theories in infinitary languages. *Pacific J. Math.*, 41:247–261, 1972.

[Spe85] J. Spencer. Six standard deviations suffice. *Trans. Amer. Math. Soc.*, 289:679–706, 1985.

[Spe87] J. Spencer. *Ten lectures on the probabilistic method*. CBMS-NSF. SIAM, 1987.

[ST83] E. Szemerédi and W. T. Trotter, Jr. Extremal problems in discrete geometry. *Combinatorica*, 3:381–392, 1983.

[Szé95] L. Székely. Crossing numbers and hard Erdős problems in discrete geometry. Manuscript, 1995.

[Tho65] R. Thom. Sur l'homologie des variétés algébriques reélles. In *Differential and Combinatorial Topology (S.S. Cairns ed.)*. Princeton Univ. Press, 1965.

[Val95] P. Valtr. Guarding galleries where no point sees a small area. KAM Series (tech. Report) 95–305, Charles University, 1995.

[Vap82] V. N. Vapnik. *Estimation of dependences based on empirical data*. Springer-Verlag, Berlin, 1982.

[VC71] 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.

[Wel88] E. Welzl. Partition trees for triangle counting and other range searching problems. In *Proc. 4th Annu. ACM Sympos. Comput. Geom.*, pages 23–33, 1988.

[WF94] M. Watanabe and P. Frankl. Some best possible bounds concerning the traces of finite sets. *Graphs Comb.*, 10:283–292, 1994.

[Wil82] D. E. Willard. Polygon retrieval. *SIAM J. Comput.*, 11:149–165, 1982.