Discrepancy and ε-approximations for bounded VC-dimension [◇]

Jiří Matoušek*
Emo Welzl[♯]
Lorenz Wernisch[♯]

B 91–06
April 1991

Abstract

Let $(X,\mathcal{R})$ be a set system on an n-point set X. We investigate colorings $\chi : X \to \{-1,+1\}$, such that the sum of colors in each set $R \in \mathcal{R}$ does not deviate too much from 0. The largest value $|\sum_{x\in R}\chi(x)|$ over all $R \in \mathcal{R}$ is the *discrepancy* of χ on $\mathcal{R}$. We also consider ε-approximations: For $0 < \varepsilon < 1$, a subset A of X is an *ε-approximation* for $(X,\mathcal{R})$, if $|\,|A \cap R|/|A| - |R|/n\,| < \varepsilon$ for all $R \in \mathcal{R}$.

Let d be a fixed integer such that for any $Y \subseteq X$, the number of distinct sets of the form $R \cap Y$, $R \in \mathcal{R}$ is bounded by $O(|Y|^d)$ (i.e., $(X,\mathcal{R})$ is an n-point range space with the *primal shatter function* $\pi_{\mathcal{R}}(m)$ bounded by $O(m^d)$). Then we prove that there is always a coloring with discrepancy bounded by $O(n^{1/2-1/2d}(\log n)^{1+1/2d})$. We show that this implies that, for any r, there exists a $(1/r)$-*approximation* for $(X,\mathcal{R})$ of size $O(r^{2-2/(d+1)}(\log r)^{2-1/(d+1)})$. This improves on a previous bound of $O(r^2 \log r)$ due to Vapnik and Chervonenkis.

If any subcollection of m sets of $\mathcal{R}$ partitions the points into at most $O(m^d)$ classes (i.e., the *dual shatter function* is bounded by $O(m^d)$), then we get a bound of $O(n^{1/2-1/2d}\log n)$ for discrepancy and of $O(r^{2-2/(d+1)}(\log r)^{2-2/(d+1)})$ for $(1/r)$-approximations. These bounds via the dual shatter function can be realized by deterministic polynomial time algorithms.

All the bounds are tight upto polylogarithmic factors in the worst case. Our results allow to generalize several results of Beck bounding the discrepancy in certain geometric settings to the case when the discrepancy is taken relative to an arbitrary measure.

[◇] Work of **J.M.** and **E.W.** was partially supported by the ESPRIT II Basic Research Actions Program of the EC under contract No. 3075 (project ALCOM). **L.W.** acknowledges support from the Deutsche Forschungsgemeinschaft under grant We 1265/1-3, Schwerpunktprogramm "Datenstrukturen und effiziente Algorithmen".

*Department of Applied Mathematics, Charles University, Malostranské nám. 25, 118 00 Praha 1, Czechoslovakia, and School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30 332, USA
[♯] Institut für Informatik, Freie Universität Berlin, Arnimallee 2-6, W 1000 Berlin 33, Germany

1 Introduction and statement of results

In this section we first review basics about discrepancy, ε-approximations, and range spaces of finite VC-dimension, and then we state our results. Section 2 will contain the proofs of our main results. The proofs are based on a combination of tools from discrepancy theory and computational geometry. Finally, in Section 3, we conclude with some implications and algorithmic aspects of our results.

Discrepancy. Let $(X, \mathcal{R})$ be a set system and let $\chi : X \to \{-1, +1\}$ be a mapping; we will call such a mapping a *coloring* of X. For a set $Y \subseteq X$, let $\chi(Y) = \sum_{x \in Y} \chi(x)$. We define the *discrepancy of* χ *on* $\mathcal{R}$ by

$$\operatorname{disc}(\mathcal{R}, \chi) = \max_{R \in \mathcal{R}} |\chi(R)|,$$

and the *discrepancy of* $\mathcal{R}$ by

$$\operatorname{disc}(\mathcal{R}) = \min\{\operatorname{disc}(\mathcal{R}, \chi);\ \chi : X \to \{-1, +1\}\}.$$

The concept of discrepancy originated in the theory of uniform distribution (see e.g. the book [Hla]), and the original problem (how well can a discrete point set in the unit cube approximate the Lebesgue measure on aligned boxes contained in the unit cube) was then extended also to approximating the measure of other geometric figures. The book [BC87] gives a number of results in this direction and further references. We will also say a little more about the geometric discrepancy and its connection to the above defined combinatorial notion in Section 3.

For the discrepancy of general set systems, various bounds are known (see [Spe87]). For example, for a set system $(X, \mathcal{R})$ with polynomially many ranges, the following bound is of interest:

$$\operatorname{disc}(\mathcal{R}) = O(\sqrt{s \log |\mathcal{R}|}\,), \quad s = \max_{R \in \mathcal{R}} |R|.$$

To see this one considers a random coloring χ. For any fixed set $Y \subseteq X$, we know that

$$\operatorname{Prob}(|\chi(Y)| > \lambda\sqrt{|Y|}\,) < 2e^{-\lambda^2/2}. \tag{1}$$

Hence, if we set $\lambda = \sqrt{2\ln(4|\mathcal{R}|)}$, then the above bound becomes $1/(2|\mathcal{R}|)$, and, with probability at least $\frac{1}{2}$, a random coloring satisfies $|\chi(R)| \le \sqrt{2s\ln(4|\mathcal{R}|)}$ for all $R \in \mathcal{R}$.

If $|\mathcal{R}| = O(n^{O(1)})$, this gives $\operatorname{disc}(\mathcal{R}) = O(\sqrt{n \log n}\,)$. A probabilistic construction shows that, in general, this bound cannot be pushed below $O(\sqrt{n \log n}\,)$. However, as we will see, a substantial improvement is possible if the set system has bounded VC-dimension.

Range spaces of finite Vapnik-Chervonenkis dimension and ε-approximations. A significant part of new results in computational geometry over the last few years use probabilistic methods and algorithms. Different approaches ([Cla87], [HW87]) introduce abstract frameworks for their considerations. Our results will be based on the concept used in [HW87], so-called range spaces of bounded VC-dimension.

A *range space* is a set system, i.e. a pair $\Sigma = (X, \mathcal{R})$, where X is a set and $\mathcal{R}$ is a set of subsets of X. We will usually call the elements of X the *points* of Σ and the elements of $\mathcal{R}$ the *ranges* of Σ. If Y is a subset of X, we denote by $\mathcal{R}|_Y$ the set system $\{R \cap Y;\ R \in \mathcal{R}\}$, and we call $(Y, \mathcal{R}|_Y)$ the *subspace induced by* Y.

Let us say that a subset $Y \subseteq X$ is *shattered* (by $\mathcal{R}$), if every possible subset in the subspace induced by Y is a range, i.e. if $\mathcal{R}|_Y = 2^Y$. We define the *Vapnik-Chervonenkis dimension*, VC-*dimension* for short, of the range space $\Sigma = (X, \mathcal{R})$ as the maximum size of a shattered subset of X (if there are shattered subsets of any size, then we say that the VC-dimension is infinite).

This notion has been introduced by Vapnik and Chervonenkis [VC71]. Range spaces of VC-dimension bounded by a constant occur naturally in geometry, but also e.g. in learning theory (see [BEHW89]). A simple geometric example is the following: X is a finite point set in the plane, and every set defined as the intersection of X with a halfplane is a range. This example can of course be generalized in many ways.

An important notion in applications (and also for our proofs) is that of an ε-net. A subset $S \subseteq X$ is called an *ε-net for Σ* provided that $S \cap R \neq \emptyset$ for every range $R \in \mathcal{R}$ with $|R|/|X| > \varepsilon$.

Bounded VC-dimension guarantees the existence of small ε-nets as stated in the following theorem.

Theorem 1.1 [HW87] *Let d be fixed and let $(X, \mathcal{R})$ be a range space of* VC-*dimension d. Then for every $r > 1$, there exists a $(1/r)$-net for $(X, \mathcal{R})$ of size $O(r \log r)$.* $\square$

The bound on the ε-net size has been improved several times (concerning the dependency of the constant on d), the best result being due to Komlós, Pach and Wöginger [KPW91]. But for a fixed d, the dependency on ε cannot be improved in general; this was shown in [PW90] by a probabilistic construction; a big open problem is whether an improvement is possible in geometric settings.

Another related concept – implicitly contained in [VC71] – is that of an ε-approximation. A subset $A \subseteq X$ is an *ε-approximation* for a range space $(X, \mathcal{R})$, provided that

$$\left| \frac{|A \cap R|}{|A|} - \frac{|R|}{|X|} \right| \leq \varepsilon$$

for every range $R \in \mathcal{R}$. Again, one can show the existence of small ε-approximations:

Theorem 1.2 [VC71] *Let d be fixed and let $(X, \mathcal{R})$ be a range space of VC-dimension d. Then for every $r > 1$, there exists a $(1/r)$-approximation for $(X, \mathcal{R})$ of size $O(r^2 \log r)$.* $\square$

While ε-nets have been applied in computational geometry since their introduction, ε-approximations have lived somehow in their shadow. However, ε-approximations have some nice properties not shared by ε-nets, which were applied for designing an efficient deterministic algorithm for computing ε-approximations for range spaces of bounded VC-dimension, see [Mat91a]. The only known way for efficient deterministic computation of ε-nets is via ε-approximations.

Results. In this paper we prove two bounds on discrepancy of finite range spaces of bounded VC-dimension, and these bounds will imply bounds on the size of ε-approximations.

The results will not be stated in terms of the VC-dimension of the range space, because the exact value of the VC-dimension is often difficult to determine even for natural geometric examples (e.g., the reader can try to determine the VC-dimension of the space with a point set in the plane and with ranges determined by all triangles). Other related parameters of a range space, which are also easier to estimate, turn out to be essential for the discrepancy bound: the primal shatter function and the dual shatter function.

The *primal shatter function* $\pi_{\mathcal{R}}$ of a range space $(X, \mathcal{R})$ is defined by

$$\pi_{\mathcal{R}}(m) = \max_{A \subseteq X,\ |A| \leq m} |\{R \cap A;\ R \in \mathcal{R}\}|.$$

The *dual shatter function* $\pi^*_{\mathcal{R}}$ is the primal shatter function of the dual range space arising by exchanging the role of points and ranges; thus $\pi^*_{\mathcal{R}}(m)$ is the maximum number of equivalence classes into which the points of X can be partitioned by a collection $\mathcal{A}$ of m ranges in $\mathcal{R}$, where $x, y \in X$ are *equivalent relative to* $\mathcal{A}$ if $\{R \in \mathcal{A};\ x \in R\} = \{R \in \mathcal{A};\ y \in R\}$). For example, consider $(P, \mathcal{B})$, where P is a finite set of points in $\mathbb{E}^d$, and $\mathcal{B}$ is the set of intersections of P with balls. Then the primal shatter function is of order $O(m^{d+1})$, while the dual shatter function is of the order $O(m^d)$; the VC-dimension is $d+1$.

A result obtained independently by several authors ([VC71], [Sau72]) says that the primal shatter function $\pi_{\mathcal{R}}(n)$ of a range space of VC-dimension d is bounded by $\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d} = \Theta(n^d)$, and the bound is tight in the worst case – take all subsets of X with at most d elements as ranges. But in geometric examples, the bound on the primal shatter function is usually better than implied by their VC-dimension. One bound for discrepancy will be expressed in terms of the primal shatter function, and the other one in terms of the dual shatter function. Both bounds come out almost identical, but their area of application differs; a range space can have a much larger primal shatter function than the dual shatter function or vice versa.

Our results are:

Theorem 1.3 *Let $(X, \mathcal{R})$ be an n-point range space and d, C constants, such that $\pi_{\mathcal{R}}(m) \leq Cm^d$ for all $m \leq n$. Then the discrepancy* $\operatorname{disc}(\mathcal{R})$ *of $\mathcal{R}$ is bounded by*

$$O(n^{\frac{1}{2}-\frac{1}{2d}}(\log n)^{1+\frac{1}{2d}}),\ \textit{if } d > 1,\ \textit{and } O(\log^{\frac{5}{2}} n),\ \textit{if } d = 1,$$

and for every $r \leq n$, there exists a $(1/r)$-approximation for $(X, \mathcal{R})$ of size

$$O(r^{2-\frac{2}{d+1}}(\log r)^{2-\frac{1}{d+1}}),\ \textit{if } d > 1,\ \textit{and } O(r \log^{\frac{5}{2}} r),\ \textit{if } d = 1.$$

Theorem 1.4 *Let $(X, \mathcal{R})$ be an n-point range space and d, C constants, such that $\pi^*_{\mathcal{R}}(m) \leq Cm^d$ for all $m \leq n$. Then the discrepancy* $\operatorname{disc}(\mathcal{R})$ *of $\mathcal{R}$ is bounded by*

$$O(n^{\frac{1}{2}-\frac{1}{2d}} \log n),\ \textit{if } d > 1,\ \textit{and } O(\log^{\frac{3}{2}} n),\ \textit{if } d = 1,$$

and for every $r \leq n$, there exists a $(1/r)$-approximation for $(X, \mathcal{R})$ of size

$$O(r^{2-\frac{2}{d+1}}(\log r)^{2-\frac{2}{d+1}}),\ \textit{if } d > 1,\ \textit{and } O(r \log^{\frac{3}{2}} r),\ \textit{if } d = 1.$$

Section 3 will contain some remarks to what extent these bounds are best possible, and we discuss the implications for concrete examples like balls in $\mathbb{E}^d$.

2 Proofs

We first show how the result for discrepancy implies the bounds for ε-approximations. Then we prove the results for primal and dual shatter functions.

Discrepancy versus ε-approximations. We begin with a simple lemma.

Lemma 2.1 *Let $(X, \mathcal{R})$ be a set system on an n-point set with $X \in \mathcal{R}$ and let $\delta = \mathrm{disc}(\mathcal{R})$. Then there exists a $(2\delta/n)$-approximation A for $(X, \mathcal{R})$ with $|A| = \lceil \frac{n}{2} \rceil$.*

Proof. For a coloring χ of X with discrepancy δ let A' be the larger one of the two sets $\{x \in X; \chi(x) = -1\}$ and $\{x \in X; \chi(x) = +1\}$; so $|A'| \geq \lceil \frac{n}{2} \rceil$. From $X \in \mathcal{R}$ and $|\chi(X)| \leq \delta$ we get $|A'| - |X \setminus A'| \leq \delta$, or $|A'| - \frac{n}{2} \leq \frac{\delta}{2}$. Remove $|A'| - \lceil \frac{n}{2} \rceil$ arbitrary elements from A' to obtain the approximation A.

We have $|\,2|A \cap R| - |R|\,| \leq |\,2|A' \cap R| - |R|\,| + 2(|A'| - |A|) \leq 2\delta$. Thus, for n even,

$$\left| \frac{|A \cap R|}{|A|} - \frac{|R|}{|X|} \right| = \frac{1}{n} |\,2|A \cap R| - |R|\,| \leq \frac{2\delta}{n};$$

similarly, one derives the bound for n odd, where one uses that actually $|\,2|A \cap R| - |R|\,| \leq 2\delta - 1$ in this case. $\Box$

Now we show how to obtain small ε-approximations from colorings with small discrepancy.

Let $(X, \mathcal{R})$ be a set system with $X \in \mathcal{R}$, and let $\delta(m)$ be a function bounding the discrepancy of any m-point subspace of our set system. We construct sets $A_0, A_1, \ldots, A_i, \ldots$ as follows: $A_0 = X$, and, for $i \geq 0$, write a_i short for $|A_i|$ and let A_{i+1} be a $(2\delta(a_i)/a_i)$-approximation for $(A_i, \mathcal{R}|_{A_i})$ with $|A_{i+1}| = \lceil a_i/2 \rceil$. We conclude that $|A_k| = \lceil n/2^k \rceil$ and A_k is an ε-approximation for $(X, \mathcal{R})$ where

$$\varepsilon \leq 2 \left(\frac{\delta(a_0)}{a_0} + \frac{\delta(a_1)}{a_1} + \cdots + \frac{\delta(a_{k-1})}{a_{k-1}} \right), \quad \text{with } a_i = \lceil \frac{n}{2^i} \rceil.$$

Lemma 2.2 *Let $(X, \mathcal{R})$ be a set system on an n-point set with $X \in \mathcal{R}$, and let δ be a function with $\mathrm{disc}(\mathcal{R}|_Y) \leq \delta(|Y|)$ for all $Y \subseteq X$. Then, for every k, there exists an ε-approximation A for $(X, \mathcal{R})$ with $|A| = a = \lceil \frac{n}{2^k} \rceil$ and*

$$\varepsilon \leq \frac{2}{n} \left(\delta(n) + 2\delta(\lceil \frac{n}{2} \rceil) + \ldots + 2^{k-1}\delta(\lceil \frac{n}{2^{k-1}} \rceil) \right).$$

In particular, if there exists a constant $c > 1$ such that $\delta(m) \geq \frac{c}{2}\delta(2m)$ for $m \geq a$, then $\varepsilon = O(\frac{\delta(a)}{a})$. $\Box$

Since adding X as a range to $\mathcal{R}$ increases the primal shatter function by 1 at most, and leaves the dual shatter function unchanged, the bounds for approximations in Theorems 1.3 and 1.4 readily follow from those for discrepancy.

Discrepancy bounds via primal shatter functions. Our proof uses the following lemma, which follows from the proof of Lemma 8.10 in [BC87], due to Beck. For the reader's convenience, we recall the proof.

Lemma 2.3 *Let $\mathcal{L}$, $\mathcal{S}$ be set systems on an n-point set X, $|\mathcal{S}| > 1$, such that $|S| \leq s$ for every $S \in \mathcal{S}$ and*

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

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

Proof. Let $\mathcal{C}_0$ be the set of all colorings $\chi : X \to \{-1, +1\}$, and let $\mathcal{C}_1$ be the subcollection of mappings χ with $|\chi(S)| \leq \sqrt{2s \ln(4|\mathcal{S}|)}$ for all $S \in \mathcal{S}$. We have seen in Section 1 that $|\mathcal{C}_1| \geq \frac{1}{2}|\mathcal{C}_0| = 2^{n-1}$.

Now let us define a mapping $v : \mathcal{C}_1 \to \mathbb{Z}^{|\mathcal{L}|}$, assigning to a coloring χ the $|\mathcal{L}|$-component integer vector $v(\chi) = (\chi(L);\ L \in \mathcal{L})$. Since $|\chi(L)| \leq |L|$ and $\chi(L) - |L|$ is even for every L, the image of v contains at most

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

integer vectors. Hence there is a vector $v_0 = v(\chi_0)$ such that v maps at least $2^{4(n-1)/5}$ elements of $\mathcal{C}_1$ to v_0. Let $\mathcal{C}_2$ be the collection of all $\chi \in \mathcal{C}_1$ with $v(\chi) = v_0$. Let us pick one $\chi_0 \in \mathcal{C}_2$ and for every $\chi \in \mathcal{C}_2$, we define a new mapping $\chi' : X \to \{-1, 0, 1\}$ by $\chi'(x) = (\chi(x) - \chi_0(x))/2$. Then $\chi'(L) = 0$ for every $L \in \mathcal{L}$, and also $\chi'(S) \leq \sqrt{2s \ln(4|\mathcal{S}|)}$ for every $S \in \mathcal{S}$. Let $\mathcal{C}_2'$ be the collection of χ' for all $\chi \in \mathcal{C}_2$.

To prove the lemma, it remains to show that there is a mapping $\chi' \in \mathcal{C}_2'$ whose value is nonzero in at least $n/10$ points of X. The number of mappings $X \to \{-1, 0, +1\}$ with at most $n/10$ nonzero elements is bounded by

$$\sum_{q=0}^{\lfloor n/10 \rfloor} \binom{n}{q} 2^q,$$

and standard estimates show that this number is smaller than $2^{4(n-1)/5} \leq |\mathcal{C}_2'|$ (see [BC87]). Hence there exists a mapping $\chi' \in \mathcal{C}_2'$ with at least $n/10$ nonzero values. □

For the **proof of Theorem 1.3** we first describe a construction of a partial coloring for a range space using the previous lemma, which will then be applied iteratively.

Let $(X, \mathcal{R})$ be a range space with $\pi_{\mathcal{R}}(m) = O(m^d)$. Let us define another range space $(X, \mathcal{R}')$ by $\mathcal{R}' = \{R_1 \setminus R_2;\ R_1, R_2 \in \mathcal{R}\}$. The range space $(X, \mathcal{R})$ has bounded VC-dimension, and hence also $(X, \mathcal{R}')$ has a bounded VC-dimension (see [HW87]). Let $N \subseteq X$ be a $(1/r)$-net for $(X, \mathcal{R}')$ of size $O(r \log r)$, where r is a parameter to be fixed later (the existence of such N is guaranteed by Theorem 1.1).

Let us call two ranges $R_1, R_2 \in \mathcal{R}$ equivalent if $R_1 \cap N = R_2 \cap N$. Since the ranges of $\mathcal{R}$ have at most $O((r \log r)^d)$ distinct intersections with N, this equivalence has at most $O((r \log r)^d)$ classes. Let a collection $\mathcal{L}$ contain exactly one range of each equivalence class. For a range $R \in \mathcal{R}$, let L_R be the member of $\mathcal{L}$ equivalent to R.

Let us put

$$\mathcal{S} = \{R \setminus L_R;\ R \in \mathcal{R}\} \cup \{L_R \setminus R;\ R \in \mathcal{R}\}.$$

For every R, $L_R \setminus R$ and $R \setminus L_R$ contain no points of N, and thus by the $(1/r)$-net property of N, the cardinality of any set of $\mathcal{S}$ is at most n/r. Also we have $|\mathcal{S}| \leq 2|\mathcal{R}| \cdot |\mathcal{L}| = O(n^{2d})$.

We want to apply Lemma 2.3 on the set systems $\mathcal{L}$ and $\mathcal{S}$, so we need an estimate on $\prod_{L\in\mathcal{L}}(|L|+1)$. This is bounded by $(n+1)^{|\mathcal{L}|} \leq (n+1)^{(Kr\log r)^d}$ for some constant K. Thus if we set $r = cn^{1/d}/(\log n)^{1+1/d}$ for a small enough positive constant c, we get that the product is bounded by $2^{(n-1)/5}$ as required. Then the size of sets of $\mathcal{S}$ is bounded by $s = n/r = O(n^{1-1/d}(\log n)^{1+1/d})$, and Lemma 2.3 guarantees the existence of a mapping $\chi : X \to \{-1, 0, 1\}$, such that $\chi(L) = 0$ for all $L \in \mathcal{L}$,

$$|\chi(S)| \leq \sqrt{2s\ln(4|\mathcal{S}|)} = O(n^{1/2-1/2d}(\log n)^{1+1/2d}),$$

and the set $Y_1 = \{x \in X;\ \chi(x) \neq 0\}$ has at least $n/10$ elements.

If R is any range, we can write

$$R = (L_R \cup S_1) \setminus S_2,$$

where $S_1 = R \setminus L_R \in \mathcal{S}$, $S_2 = L_R \setminus R \in \mathcal{S}$, S_1 and L_R are disjoint and S_2 is contained in L_R. Hence

$$|\chi(R)| = |\chi(R \cap Y_1)| \leq |\chi(L_R)| + |\chi(S_1)| + |\chi(S_2)| = O(n^{1/2-1/2d}(\log n)^{1+1/2d}).$$

To prove Theorem 1.3, we apply the construction described above inductively. We set $X_1 = X$, and we obtain a partial coloring χ_1 nonzero on a set Y_1 as above. We set $X_2 = X_1 \setminus Y_1$, and we obtain a partial coloring χ_2 of X_2 by applying the above construction on the range space $(X_2, \mathcal{R}|_{X_2})$, etc. We repeat this construction until the size of the set X_k becomes trivially small (e.g., smaller than a suitable constant). Then we define $Y_k = X_k$ and we let χ_k be the constant mapping with value 1 on Y_k.

Let $R \in \mathcal{R}$ be a range. We have

$$|\chi(R)| \leq |\chi_1(R \cap Y_1)| + |\chi_2(R \cap Y_2)| + \cdots + |\chi_k(R \cap Y_k)|. \tag{3}$$

For every i, $|\chi_i(R)|$ is bounded by $O(n_i^{1/2-1/2d}(\log n_i)^{1+1/2d})$, where $n_i = |X_i| \leq (9/10)^{i-1}n$. Thus, for $d > 1$ the summands on the right hand side of (3) decrease geometrically, and we obtain $\mathrm{disc}(\mathcal{R}) = O(n^{1/2-1/2d}(\log n)^{1+1/2d})$ as claimed. For $d = 1$, we get $\mathrm{disc}(\mathcal{R}) = O(\log^{5/2} n)$. □

Discrepancy bounds via dual shatter functions. The proof of Theorem 1.4 uses results on spanning paths with a low crossing number by Chazelle and Welzl [Wel88], [CW89]. Let us give the necessary definitions.

Let $(X, \mathcal{R})$ be a range space. If $\{x, y\}$ is a two-point subset of X and R a range, we say that R *crosses* $\{x, y\}$ if $|\{x, y\} \cap R| = 1$. A *spanning path* P on X is a linear ordering $(x_1, \ldots, x_n)$ of the points of X; its *edges* are $\{x_1, x_2\}$, $\{x_2, x_3\}$, ..., $\{x_{n-1}, x_n\}$. The *crossing number* of P is the maximum number of edges of P crossed by a single range of $\mathcal{R}$, over all ranges $R \in \mathcal{R}$. A spanning path with a low crossing number will help us to establish Theorem 1.4.

Theorem 2.4 [CW89] *Let $(X, \mathcal{R})$ be a range space whose dual shatter function $\pi^*_{\mathcal{R}}(m)$ is bounded by Cm^d for some constants C, d. Then there exists a spanning path on X with crossing number $O(n^{1-1/d}\log n)$, if $d > 1$ and $O(\log^2 n)$, if $d = 1$.* □

Hence for a **proof of Theorem 1.4**, it is enough to show the following:

Lemma 2.5 *Let* $(X, \mathcal{R})$ *be a range space with a spanning path with crossing number* κ. *Then* $\operatorname{disc}(\mathcal{R}) = O(\sqrt{\kappa \log |\mathcal{R}|})$.

Proof. What we actually need is a matching on X with a small crossing number. Let us suppose that the number of points of X is even (if not, we may ignore one point temporarily; the discrepancy grows at most by one by adding it back). Let $P = (x_1, \ldots, x_n)$ be a spanning path with crossing number κ, and consider the set

$$M = \{\, \{x_1, x_2\}, \{x_3, x_4\}, \ldots, \{x_{n-1}, x_n\}\}$$

of $n/2$ edges of P. We let $\mathcal{C}$ be the set of all colorings $\chi : X \to \{-1, +1\}$ with $\chi(\{x, y\}) = 0$ for any pair $\{x, y\} \in M$. We show that a random element of $\mathcal{C}$ satisfies $\operatorname{disc}(\mathcal{R}, \chi) \leq \sqrt{2\kappa \ln(4|\mathcal{R}|)}$. Indeed, let $R \in \mathcal{R}$ be a range, and let M_R be the union over the set of edges of M crossed by R; we know that $|M_R \cap R| \leq \kappa$ and $\chi(R) = \chi(M_R \cap R)$ for every $\chi \in \mathcal{C}$. For a random $\chi \in \mathcal{C}$, we thus have

$$\operatorname{Prob}(|\chi(R)| > \lambda\sqrt{\kappa}\,) \;\leq\; \operatorname{Prob}(|\chi(M_R \cap R)| > \lambda\sqrt{|M_R \cap R|}\,) \;<\; 2e^{-\lambda^2/2} \;=\; \frac{1}{2|\mathcal{R}|}$$

for $\lambda = \sqrt{2\ln(4|\mathcal{R}|)}$ (according to (1)), and hence some mapping $\chi \in \mathcal{C}$ gives the claimed discrepancy. $\Box$

3 Discussion

We discuss some implications of our results and their proofs.

Discrepancy of balls in $\mathbb{E}^d$. We consider the case when X is a set of points in $\mathbb{E}^d$, and the ranges are those subsets which can be obtained as an intersection of X with a ball. It was shown in [CW89], that every set of n points in $\mathbb{E}^d$ allows a spanning path with crossing number $\kappa = O(n^{1-1/d})$ for balls (which is better by a log-factor compared to the general bound in Theorem 2.4, using the fact that the dual shatter function is of the order $O(m^d)$). We get

Corollary 3.1 *Let* P *be a set of* n *points in* $\mathbb{E}^d$, *and let* $\mathcal{B} = \{P \cap B; B \text{ a ball in} \mathbb{E}^d\}$. *Then* $\operatorname{disc}(\mathcal{B}) = O(n^{1/2-1/2d}\sqrt{\log n})$. $\Box$

As we mentioned in the introduction, the notion of discrepancy originated in geometric settings and many results in this direction are known. One type of a geometric discrepancy discussed in [BC87] is defined as follows:

Let μ be the probabilistic measure in $\mathbb{E}^d$. Let $\mathcal{F}$ be a family of μ-measurable subsets of $\mathbb{E}^d$ (usually of simple geometric objects, as e.g. balls or boxes); e.g. take μ as the Lebesgue measure in the unit cube. For an n-point set $P \subseteq \mathbb{E}^d$, one defines the μ-discrepancy of $\mathcal{F}$ on P by

$$D_\mu(P, \mathcal{F}) = \sup_{F \in \mathcal{F}} |n\mu(F) - |P \cap F|\,|,$$

and the μ-discrepancy function of $\mathcal{F}$ is then

$$D_\mu(n, \mathcal{F}) = \inf_{P \subset \mathbb{E}^d,\ |P|=n} D_\mu(P, \mathcal{F}).$$

(also other variations of discrepancy are treated in [BC87], as e.g. the toroidal discrepancy of a point set, but we will not go into details here).

The book [BC87] contains many upper and lower bounds on the discrepancy functions for various families, as e.g. aligned boxes (with sides parallel to coordinate axes), boxes, balls, convex sets. General bounds are given for the case when A is a fixed compact convex body and $\mathcal{F}$ is the family of all its copies under rotations, translations and contracting homotheties (or only translations and homotheties).

It is not difficult to show that the investigation of discrepancy with respect to the Lebesgue measure (restricted to the unit cube) can be reduced to the investigation of discrepancy with respect to a measure concentrated on the points of a sufficiently fine grid. In fact, some of the Beck's upper bounds were gained via theorems about discrepancy of set systems, and a detailed discussion of the transformation from discrete to continuous setting and back can be found in [BC87].

The general results in this paper allow to re-derive many of the upper bounds, and all the bounds gained in this way hold for arbitrary probabilistic measures μ.

Corollary 3.2 *Let $\mathcal{B}$ be the set of balls in $\mathbb{E}^d$. Then $D_\mu(n, \mathcal{B}) = O(n^{1/2-1/2d}\sqrt{\log n})$ for every probabilistic measure μ, with the constant depending on d only (and not on μ).*

Proof. For a given n, we construct a set P of n points with $|n\mu(B) - |P \cap B||\, = O(n^{1/2-1/2d}\sqrt{\log n})$ for all $B \in \mathcal{B}$ as follows.

Let $\varepsilon_1 = n^{1/2-1/2d}\sqrt{\log n}/n$. We first take a finite set Q of N points in $\mathbb{E}^d$, such that $|\mu(B) - |B \cap Q|/N\,| < \varepsilon_1$ for all balls $B \in \mathcal{B}$; (it is shown in [VC71] that a random – according to μ – set of $N = O((d/\varepsilon_1)^2 \log(d/\varepsilon_1))$ points will have the desired property; however, the size of Q is not crucial in our argument, as long as it is finite). Next we choose an ε_2-approximation P of Q with $|P| = n$ and $\varepsilon_2 = O(\varepsilon_1)$. The existence of P follows from Lemma 2.2. We obtain

$$|\mu(B) - \frac{|B \cap P|}{n}\,| \le |\mu(B) - \frac{|B \cap Q|}{N}\,| + |\,\frac{|B \cap P|}{n} - \frac{|B \cap Q|}{N}\,| < \varepsilon_1 + \varepsilon_2,$$

and the claimed bound follows. □

The corollary improves on results in [Bec84], where the bounds hold only under certain assumptions on the measure, and the constant in the asymptotic bound depends also on this measure.

The bound can be readily generalized to the case where $\mathcal{F}$ is the set of k-fold boolean combination of balls, since the crossing number of a spanning path increases by a factor at most k with respect to such a family (compared to balls alone).

For other families, we can easily guarantee the bound $O(m^d)$ for the dual shatter function. This is for instance if the sets of $\mathcal{F}$ are bounded by algebraic surfaces of a fixed degree: m such surfaces give rise to $O(m^d)$ d-wise intersections (ignoring degeneracies), and the number of cells in the arrangement of m surfaces is not greater than the number of vertices. For these cases we obtain a slightly worse discrepancy bound $O(n^{1/2-1/2d}\log n)$. This result translated to the Lebesgue measure case is not directly implied by Beck's results, since he considers convex bodies only.

The primal shatter function bound does not seem to be so useful in geometric settings, since it depends on the complexity of the figures determining ranges rather than the space dimension (e.g., for disks in the plane, the shatter function is $\Theta(n^3)$, while the dual shatter function only $O(n^2)$).

Lower bounds. The results of Beck [BC87] also imply almost matching lower bounds (upto logarithmic factors) for both discrepancy bounds in this paper. First of all, one of Beck's results gives a lower bound of $\Omega(n^{1/2-1/2d})$ for the discrepancy function of a family $\mathcal{F}$, where each member of $\mathcal{F}$ is a union of at most 2^d balls in $\mathbb{E}^d$ (this follows from a bound on toroidal discrepancy for balls). Since it is easy to see that the dual shatter function $\pi^*_{\mathcal{F}}(m)$ is of order $O(m^d)$, we get that the discrepancy bound in Theorem 1.4 cannot be improved below $O(n^{1/2-1/2d})$.

For the primal shatter function, the lower bound question is more delicate. Let $\mathcal{F}$ be the family of all halfspaces in $\mathbb{E}^d$. Then it is easy to see that $\pi_{\mathcal{F}}(m) = O(m^d)$. In dimension $d = 2$, Beck proves a lower bound $\Omega(n^{1/4}(\log n)^{-7/2})$ for the discrepancy function of $\mathcal{F}$ (where the measure μ is the Lebesgue measure restricted to the unit disk instead of the unit square). His proof works in any dimension, giving a lower bound of the form $\Omega(n^{1/2-1/2d}(\log n)^{c_d})$ (c_d a constant) for discrepancy under the assumptions of Theorem 1.3.

Other geometric lower bound results were given by Alexander [Ale90] using different methods than Beck. However, it would be nice to find a more direct lower bound proof for our combinatorial setting.

Algorithmic aspects. The proof of Theorem 1.4 can be easily turned into a polynomial algorithm for computing good colorings or ε-approximations with the claimed bounds. Note that such an algorithm will go through two stages. First, it has to compute a spanning path, or actually a matching with small crossing number. Second, it has to decide which of every two matched points gets sign '+' and '−'.

The first stage can be solved in polynomial time by the "iterative reweighting algorithm" in [Wel88], [CW89]. The second phase can either be solved by choosing the signs randomly (as suggested by the proof), or by the "hyperbolic cosine algorithm" in [Spe87], if one prefers a determinisitic algorithm with guaranteed performance.

Corollary 3.3 *Given a range space $(X, \mathcal{R})$ with dual shatter function $\pi^*(m) = O(m^d)$, a coloring with discrepancy $O(n^{1/2-1/2d}\log n)$ can be constructed in time polynomial in $|X|$ and $|\mathcal{R}|$.* □

It would be interesting to give more specific bounds for deterministic computing of good colorings and ε-approximations (particularly in specific geometric settings). When $1/\varepsilon$ is much smaller than n, one can use a method of [Mat91a] for a faster computation of an ε-approximation. For computing a spanning path with low crossing number, the methods of [Mat91b] can be applied in some geometric settings to get an efficient algorithm. However, the actual complexity of such algorithms is a matter of further research.

Unfortunately, the proof for the bound via the primal shatter function uses the pigeon hole principle, and thus it is not clear how it can be turned into an efficient algorithm.

References

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

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

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

[BEHW89] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. Warmuth. Classifying learnable geometric concepts with the Vapnik-Chervonenkis dimension. *Journal of the ACM*, 36(4):929–965, 1989.

[Cla87] K. L. Clarkson. Applications of random sampling in computational geometry. *Discrete & Computational Geometry*, 2:195–222, 1987.

[Cla88] K. L. Clarkson. Applications of random sampling in computational geometry II. In *Proc. 4. ACM Symposium on Computational Geometry*, pages 1–11, 1988.

[CW89] B. Chazelle and E. Welzl. Quasi-optimal range searching in spaces of finite VC-dimension. *Discrete & Computational Geometry*, 4:467–490, 1989.

[Hla] E. Hlawka. *The theory of uniform distribution.* A B Academic, 1984.

[HW87] D. Haussler and E. Welzl. ε-nets and simplex range queries. *Discrete & Computational Geometry*, 2:127–151, 1987.

[KPW91] J. Komlós, J. Pach and G. Wöginger. Almost tight bounds for epsilon-nets. Discrete & Computational Geometry 1991. To appear.

[Mat91a] J. Matoušek. Approximations and optimal geometric divide-and-conquer. In *Proc. 23. ACM Symposium on Theory of Computing*, 1991. To appear.

[Mat91b] J. Matoušek. Efficient patrition trees. In *Proc. 7th ACM Symposium on Computational Geometry*, 1991. To appear.

[PW90] J. Pach and G. Wöginger. Some new bounds for epsilon-nets. In *Proc. 6. ACM Symposium on Computational Geometry*, pages 10–15, 1990.

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

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

[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. 4. ACM Symposium on Computational Geometry*, pages 23–33, 1988.