A Size-Sensitive Discrepancy Bound for Set Systems of Bounded Primal Shatter Dimension

Esther Ezra

Abstract

Let \((X, S)\) be a set system on an \(n\)-point set \(X\). The discrepancy of \(S\) is defined as the minimum of the largest deviation from an even split, over all subsets of \(S \in S\) and two-colorings \(\chi\) on \(X\). We consider the scenario where, for any subset \(X' \subseteq X\) of size \(m \leq n\) and for any parameter \(1 \leq k \leq m\), the number of restrictions of the sets of \(S\) to \(X'\) of size at most \(k\) is only \(O(m^{d_1} k^{d-d_1})\), for fixed integers \(d > 0\) and \(1 \leq d_1 \leq d\) (this generalizes the standard notion of bounded primal shatter dimension when \(d_1 = d\)). In this case we show that there exists a coloring \(\chi\) with discrepancy bound \(O^*(|S|^{1/2-d_1/(2d)} n^{(d_1-1)/(2d)})\), for each \(S \in S\), where \(O^*(\cdot)\) hides a polylogarithmic factor in \(n\). This bound is tight up to a polylogarithmic factor [25, 27] and the corresponding coloring \(\chi\) can be computed in expected polynomial time using the very recent machinery of Lovett and Meka for constructive discrepancy minimization [24]. Our bound improves and generalizes the bounds obtained from the machinery of Har-Peled and Sharir [19] (and the follow-up work in [32]) for points and halfspaces in \(d\)-space for \(d \geq 3\). Last but not least, we show that our bound yields improved bounds for the size of relative \((\varepsilon, \delta)\)-approximations for set systems of the above kind.

*Work on this paper has been supported by NSF under grant CCF-12-16689. A preliminary version of this paper has appeared in Proc. Sympos. Discrete Algorithms, 2014, pp. 1378–1388 [16].

†Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA; esther@courant.nyu.edu.
1 Introduction

Let \((X, \mathcal{S})\) be a finite set system with \(n = |X|\). A two-coloring of \(X\) is a mapping \(\chi : X \to \{-1, +1\}\). For a set \(S \in \mathcal{S}\) we define \(\chi(S) := \sum_{x \in S} \chi(x)\). The discrepancy of \(S\) is then defined as

\[
\text{disc}(S) := \min_{\chi} \max_{S \in \mathcal{S}} |\chi(S)|.
\]

In other words, the discrepancy of the set system \((X, \mathcal{S})\) is the minimum over all colorings \(\chi\) of the largest deviation from an even split, over all sets in \(\mathcal{S}\).

Our goal in this paper is to derive discrepancy bounds for \((X, \mathcal{S})\) in the scenario where \((X, \mathcal{S})\) admits a polynomially bounded primal shatter function and has some additional favorable properties. In the bounds that we derive the discrepancy for each \(S \in \mathcal{S}\) is sensitive to its cardinality \(|S|\).

Let us first recall the definition of set systems of this kind (in this definition the ground set \(X\) can be infinite):

**Definition 1.1** (Primal Shatter Function; [22, 27]). The primal shatter function of a set system \((X, \mathcal{S})\) is a function, denoted by \(\pi_{\mathcal{S}}\), whose value at \(m\) is defined by

\[
\pi_{\mathcal{S}}(m) = \max_{Y \subseteq X, |Y| = m} |\mathcal{S}|_{\mathcal{Y}},
\]

where \(\mathcal{S}|_{\mathcal{Y}}\) is the collection of all sets in \(\mathcal{S}\) projected onto (that is, restricted to) \(Y\). In other words, \(\pi_{\mathcal{S}}(m)\) is the maximum possible number of distinct intersections of the sets of \(\mathcal{S}\) with an \(m\)-point subset of \(X\).

From now on we say that a set system \((X, \mathcal{S})\) with \(|X| = n\) (where \(n\) can be assumed to be arbitrarily large) has a primal shatter dimension \(d\) if \(\pi_{\mathcal{S}}(m) \leq Cm^d\), for all \(m \leq n\), where \(d > 1\) and \(C > 0\) are constants.

A typical family of set systems that arise in geometry with bounded primal shatter dimension consists of set systems \((X, \mathcal{S})\) of points in some low-dimensional space \(\mathbb{R}^d\), and \(\mathcal{S}\) is a collection of certain simply-shaped regions, e.g., halfspaces, balls, or simplices (where \(d > 0\) is assumed to be a constant). In such cases, the primal shatter function is \(m^{O(d)}\); see, e.g., [18] for more details. In fact, set systems of this kind are part of a more general family, referred to as set systems of finite VC-dimension [35]; the reader is referred to [18, 22] for the exact definition. Although the formal definition of finite VC-dimension is different, it suffices to have the same requirement as for set systems of polynomially bounded primal shatter function. It is also known that the VC-dimension is finite if and only if the primal shatter dimension is finite, although they do not necessarily have the same value, see, e.g., [18] for more details. From now on we make only the assumption about having a finite primal shatter dimension, in particular, this is the case in our construction and analysis, and the VC-dimension is mentioned here only for the sake of completeness of the presentation.

A major result by Matoušek [25] (see also [27, 29]) is the following:

**Theorem 1.2** (Matoušek [27]). Let \((X, \mathcal{S})\) be a set system as above with \(|X| = n\), \(\pi_{\mathcal{S}}(m) \leq Cm^d\), for all \(m \leq n\), where \(d > 1\) and \(C > 0\) are constants. Then

\[
\text{disc}(\mathcal{S}) = O(n^{1/2-1/(2d)}),
\]

where the constant of proportionality depends on \(d\) and \(C\).

This bound is known to be tight in the worst case (see [27] and our discussion on related work).
Relative \((\varepsilon, \delta)\)-approximations and \(\varepsilon\)-nets. \(\)The motivation to establish size-sensitive discrepancy bounds of the above kind stems from their application in the construction of relative \((\varepsilon, \delta)\)-approximations, introduced by Har-Peled and Sharir [19]\(^1\) based on the work of Li et al. [23] in the context of machine learning theory.

We recall the definition from [19]: For a set system \((X, S)\) (with \(X\) finite), the measure of a set \(S \in S\) is the quantity \(\overline{X}(S) = \frac{|S \cap X|}{|X|}\). Given a set system \((X, S)\) and two parameters, \(0 < \varepsilon < 1\) and \(0 < \delta < 1\), we say that a subset \(Z \subseteq X\) is a relative \((\varepsilon, \delta)\)-approximation if it satisfies, for each set \(S \in S\),

\[
\overline{X}(S)(1 - \delta) \leq Z(S) \leq \overline{X}(S)(1 + \delta), \quad \text{if } \overline{X}(S) \geq \varepsilon, \quad \text{and}
\]

\[
\overline{X}(S) - \delta \varepsilon \leq Z(S) \leq \overline{X}(S) + \delta \varepsilon, \quad \text{otherwise}.
\]

A strongly related notion is the so-called \((\nu, \alpha)\)-sample [18, 20, 23], in which case the subset \(Z \subseteq X\) satisfies, for each set \(S \in S\),

\[
d_{\nu}(\overline{X}(S), Z(S)) := \frac{|Z(S) - \overline{X}(S)|}{Z(S) + \overline{X}(S) + \nu} < \alpha.
\]

As observed by Har-Peled and Sharir [19], relative \((\varepsilon, \delta)\)-approximations and \((\nu, \alpha)\)-samples are equivalent with an appropriate relation between \(\varepsilon, \delta, \nu, \alpha\) (roughly speaking, they are equivalent up to some constant factor). Due to this observation they conclude that the analysis of Li et al. [23] (that shows a bound on the size of \((\nu, \alpha)\)-samples) implies that for set systems of finite VC-dimension \(d\), there exist relative \((\varepsilon, \delta)\)-approximations of size \(\frac{c(d \log(1/\varepsilon) + \log(1/\delta))}{\delta \varepsilon}\), where \(c > 0\) is an absolute constant. In fact, any random sample of these many elements of \(X\) is a relative \((\varepsilon, \delta)\)-approximation with constant probability. More specifically, success with probability at least \(1 - q\) is guaranteed if one samples \(\frac{c(d \log(1/\varepsilon) + \log(1/\delta))}{\delta \varepsilon}\) elements of \(X\).\(^2\)

It was also observed in [19] that \(\varepsilon\)-nets arise as a special case of relative \((\varepsilon, \delta)\)-approximations. Specifically, an \(\varepsilon\)-net is a subset \(N \subseteq X\) with the property that any set \(S \in S\) with \(|S \cap X| \geq \varepsilon |X|\) contains an element of \(N\); in other words, \(N\) is a hitting set for all the “heavy” sets. In this case, if we set \(\delta\) to be some constant fraction, say, \(1/4\), then a relative \((\varepsilon, 1/4)\)-approximation becomes an \(\varepsilon\)-net. Moreover, a random sample of \(X\) of size \(O\left(\frac{\log(d/\varepsilon) + \log(1/\delta)}{\varepsilon}\right)\), with an appropriate choice of the constant of proportionality, is an \(\varepsilon\)-net with probability at least \(1 - q\); see [19] for further details. Our analysis exploits these two structures and the relation between them—see below.

We note that relative approximations appear as a major tool in approximate range counting machinery, where we are given a geometric set system \((X, S)\) (that is, \(X\) is a set points in some low-dimensional space, and \(S\) is a collection of simply-shaped regions, as above), and the goal is to preprocess \(X\) into a data structure that supports efficient queries of the form: Given \(S \in S\) compute a number \(t\) such that,

\[(1 - \delta)\overline{X}(S) \leq t/|X| \leq (1 + \delta)\overline{X}(S),\]

where \(\delta > 0\) is the relative error parameter. This relation has been addressed in [5, 19], see also [1, 4].

Related work. There is a rich body of literature in discrepancy theory, with numerous bounds and results. It is beyond the scope of this paper to mention all results, and we just list those that are most relevant to

\(^1\)In [19] they were introduced, with a slightly different notation, as relative \((p, \varepsilon)\)-approximations.

\(^2\)We note that although in the original analysis for this bound \(d\) is the VC-dimension, this assumption can be replaced by having just a primal shatter dimension \(d\); see, e.g., [18] for the details of the analysis.
our work. We refer the reader to the book of Chazelle [13] for an overview of discrepancy theory and the book of Matoušek [27] for various results in geometric discrepancy. In particular, our work is based on the techniques overviewed in the latter.

We first briefly overview previous results for an abstract set system on an \( n \)-point set \( X \), with \( m = |S| \) sets. The celebrated “six standard deviations” result of Spencer [33], which is an extension to the partial coloring method of Beck [8], implies that for such set systems there exists a coloring \( \chi \) such that \( \chi(S) \leq K\sqrt{n(1 + \log(m/n))} \), for each \( S \in \mathcal{S} \), where \( K > 0 \) is a universal constant. In particular, when \( m = n \) we have \( K = 6 \), in which case the discrepancy bound becomes \( 6\sqrt{n} \). In contrast, one can easily show that using simple probabilistic considerations, a random coloring yields a (suboptimal) discrepancy bound of \( \sqrt{n}\log m \) (or \( \sqrt{n}\log n \) if \( m = O(n) \)). A long-standing open problem was whether the result of Spencer [33] can be made constructive, and this has recently been answered in the positive by Bansal [7] for the case \( m = O(n) \) (for the general case his bound is slightly suboptimal with respect to the bound in [33]). In a follow-up work, Lovett and Meka [24] have shown a new elementary constructive proof of Spencer’s result, resulting in similar asymptotic discrepancy bounds to those in [33], for arbitrary values of \( m \). Very recently, this result has been further generalized by Rothvoss [31].

In geometric set systems, upper bounds were first shown by Beck, where the Lebesgue measure of a class of geometric shapes is approximated by a discrete point set; see, e.g., [9] and the book by Beck and Chen [10] for a bound of \( O(n^{1/4}\sqrt{\log n}) \) for disks as well as rotated rectangles in the plane. For arbitrary points sets, Matoušek et al. [29] have addressed the case of points and halfspaces in \( d \)-space, and showed an almost tight bound of \( O(n^{1/2−1/(2d)}\log^{1/2+1/(2d)} n) \), which has later been improved to \( O(n^{1/2−1/(2d)}) \) [25, 27] (Theorem 1.2). Concerning lower bounds, there is a rich literature where several such bounds are obtained in geometric set systems. We only mention the lower bound \( \Omega(n^{1/2−1/(2d)}) \) of Alexander [2] for set systems of points and halfspaces in \( d \)-space. For further results we refer the reader to [25] and the references therein.

The extension of discrepancy bounds for points and halfspaces in \( d \)-space to be size-sensitive has been addressed in the work of Har-Peled and Sharir [19], who showed a bound of \( O(|S|^{1/4}\log n) \) in the two-dimensional case, using an intricate extension of the technique of Welzl [37] (see also [14]) for constructing spanning trees with low crossing numbers. Nevertheless, their technique cannot be applied in higher dimensions, because already at \( d = 3 \) they showed a counterexample to their construction. The follow-up work of Sharir and Zaban [32] (based on the construction in [19]) addresses these cases, establishing the bound \( O\left(n^{(d−2)/(2d−1)}|S|^{1/(2d(d−1))}\log^{(d+1)/(2d(d−1))} n\right) \). The size-sensitive discrepancy bounds in both studies [19, 32] imply improved bounds for relative \((\varepsilon, \delta)\)-approximations for points and halfspaces in \( d \)-space (roughly speaking, these bounds are better than the bound of Li et al. [23] when \( \delta \) is not too large with respect to \( \varepsilon \)). Specifically, Har-Peled and Sharir [19] showed how to derive such bounds due to a “halving technique”, where in two dimensions the resulting bound is \( O\left(\log^{4/3}\frac{(1/(d\varepsilon))}{\varepsilon^{d/3}}\right) \). Their result in three and higher dimensions is somewhat restricted as they obtained a sequence of \( O(\log(1/\varepsilon)) \) samples, where each halfspace is associated with one such sample that constitutes its relative approximation. The follow-up work in [32] overcomes this difficulty and shows how to construct a single sample (for \( d \geq 3 \)) of size \( O\left(\frac{\log^9 (1/\varepsilon\delta)}{\varepsilon^7\delta^{d/3}}\right) \), where \( \gamma = \frac{2d(d−1)−1}{(d−1)(d−1)} \), \( \mu = \frac{2d}{d+1} \), and \( \eta = \frac{d}{d−1} \).

**Our results.** In this paper we refine the bound in Theorem 1.2 so that it becomes sensitive to the size of the sets \( S \in \mathcal{S} \) in several favorable cases. Specifically, we assume that for any (finite) set system projected onto a ground set of size \( m \leq n \) and for any parameter \( 1 \leq k \leq m \), the number of sets of size at most \( k \) is
only $O(m^{d_1}k^{d-d_1})$, where $d$ is the primal shatter dimension and $1 \leq d_1 \leq d$ is an integer parameter. By assumption, when $k = m$ we obtain $O(m^d)$ sets in total, in accordance with the assumption that the primal shatter dimension is $d$, but the above bound is also sensitive to the size $k$ of the set.

We show that for set systems of this kind there exists a coloring $\chi$ such that

$$\chi(S) = O^*(|S|^{1/2-d_1/(2d)}n^{(d_1-1)/(2d)}),$$

where $O^*(\cdot)$ hides a polylogarithmic factor in $n$. This bound almost matches (up to the polylogarithmic factor) the optimal discrepancy bound in Theorem 1.2 when $|S| = n$, but is more general than this bound as it yields a range of bounds for $1 \leq d_1 \leq d$. Specifically, when $d_1 = d$, the number of sets is just $O(m^d)$ (that is, this bound is not sensitive to their size) in which case we just have a set system of bounded primal shatter dimension, and then, once again, our discrepancy bound almost matches the optimal bound in Theorem 1.2. In the other extreme scenario, when $d_1 = 1$, the set system is what we call “well-behaved”, and our discrepancy bound then becomes $O^*(|S|^{1/2-1/2d})$, which depends only on $|S|$ (up to the polylogarithmic factor). In particular, set systems of points and halfspaces in $d$ dimensions are well-behaved with $d = 2$ and $d = 3$, respectively. In the plane, the resulting bound is slightly suboptimal with respect to the sensitive bound of Har-Peled and Sharir [19] (our hidden polylogarithmic factor is slightly larger than their $\log n$ factor). For points and halfspaces in three dimensions and higher, our bound considerably improves the bound in [32] (which extends the construction in [19]). In particular, the bound in the three-dimensional case in [32] is not purely sensitive in $|S|$ but also contains an additional sublinear term in $n$ (whereas the original technique of Har-Peled and Sharir [19] failed to yield such a bound).

As a major application, we conclude that our new discrepancy bounds yield improved bounds on the size of relative $(\varepsilon, \delta)$-approximations for the corresponding set systems. Specifically, using a variant of the “halving technique” in [19], we obtain a relative $(\varepsilon, \delta)$-approximation of size $O\left(\frac{\log 2}{\varepsilon \delta^{3/2}}\right)$ for points and halfspaces in three dimensions, which almost matches the bound $O\left(\frac{\log^{3/2}\left(1/(\varepsilon \delta)\right)}{\varepsilon \delta^{3/2}}\right)$ of Har-Peled of Sharir [19] obtained as a sequence of $O\left(\log\left(1/\varepsilon\right)\right)$ samples, whereas we obtain a single sample. In higher dimensions we obtain a bound of $O\left(\frac{\log\left(1/(\varepsilon \delta)\right)}{\varepsilon^{(d+1)/(d+11)}\delta^{2d/(d+1)}}\right)$ (which once again corresponds to a single sample) that is a considerable improvement over the bounds in [19, 32]. In particular, for arbitrarily large values of $d$ the exponent in the term $(1/\varepsilon)$ is close to $3/2$, whereas in the previous bound [32] this exponent is close to 2. See Theorem 3.8 for more details concerning these bounds.

Our construction uses a different machinery than the one in [19, 32] and is a variant of the construction of Matoušek [27] (see also Matoušek [25] for the special case of points and halfspaces in $d$-space), based on Beck’s partial coloring and on the entropy method, and is combined with the property that set systems with bounded primal shatter dimension admit a small packing (see below for details concerning these notions). Our assumption about the set system (that is, the bound on the number of sets of size at most $k \leq m \leq n$ stated above) enables us to refine the analysis in [27] to be size sensitive, which eventually leads to the desired bound.

---

3We ignore the cases where $d_1 = 0$ or $d_1$ takes fractional values, as they do not seem to appear in natural set systems, and, in particular, in the geometric set systems that we consider.
2 Preliminaries

We now briefly review some of the tools mentioned in Section 1, which are applied by our analysis. Then, we present the construction.

Partial coloring and entropy. Let $X$ be a set as above. A partial coloring of $X$ is any mapping $\chi : X \rightarrow \{-1, 0, +1\}$. A point $x \in X$ with $\chi(x) = 1$ or $\chi(x) = -1$ is said to be colored by $\chi$, whereas $\chi(x) = 0$ means that $x$ is uncolored.

A method originated by Beck [8], which was subsequently elaborated by others and is known by now as the entropy method (see, e.g., [3, 28, 33]), shows that under some favorable assumptions there exists a "substantial partial coloring" of $(X, S)$ (that is, at least some constant fraction of $X$ is colored) with good discrepancy bounds. In our analysis we apply the entropy method as a black box, and thus only present the constructive version of Beck’s partial coloring lemma as has recently been formulated by Lovett and Meka [24]; see Section 1 and the references therein for more details. We also refer the reader to [26] for the original (non-constructive) version of the entropy method.

Proposition 2.1 (Lovett and Meka [24]). Let $(X, S)$ be a set system with $|X| = n$, and let $\Delta_S > 0$ be some real number assigned to $S$, for each $S \in S$. Suppose that

$$\sum_{S \in S} \exp\left(-\frac{\Delta_S^2}{16|S|}\right) \leq \frac{n}{16}.$$  

Then there exists a partial coloring $\chi : X \rightarrow \{-1, 0, +1\}$ that colors at least $n/2$ points of $X$, so that $\chi(S) \leq \Delta_S + 1/n^c$, for each $S \in S$, where $c > 0$ is an arbitrarily large constant. Moreover, $\chi$ can be computed in expected polynomial time in $n$ and $|S|$ (where the degree of the polynomial bound depends on $c$).

$\delta$-packing. Let $(X, S)$ be a set system as above, and let $\delta > 0$ be a given integer parameter. A subcollection $P \subseteq S$ is $\delta$-separated if the symmetric difference distance $|S_1 \Delta S_2|$ between any pair of sets $S_1, S_2 \in P$ is strictly greater than $\delta$. A $\delta$-packing for $(X, S)$ is an inclusion-maximal $\delta$-separated subcollection of $S$.

A key property, shown by Haussler [21] (see also [12, 14, 15, 27, 36]), is that set systems of bounded primal shatter dimension admit small $\delta$-packings$^4$. That is:

Theorem 2.2 (Packing Lemma [12, 21]). Let $d > 1$ be a constant, and let $(X, S)$ be a set system on an $n$-point set with primal shatter dimension $d$. Let $1 \leq \delta \leq n$ be an integer parameter, and let $P \subseteq S$ be $\delta$-separated. Then $|P| = O((n/\delta)^d)$, where the constant of proportionality depends on $d$.

3 The Construction

We are now ready to present our construction. Let $(X, S)$ be a set system of bounded primal shatter dimension $d$, with the additional property that the number of sets of size at most $k$ in the projection of $(X, S)$ onto

\footnote{We note that in the original formulation in [21] the assumption is that the set system has a finite VC-dimension. However, a closer inspection at the analysis in [21] shows that this assumption can be replaced with that of having bounded primal shatter dimension.}
any \( m \)-point subset \( X' \subseteq X \) is \( O(m^d k^{d-d'}) \), for any \( m \leq n \). We construct a decomposition of each set \( S \subseteq S \) as a Boolean combination of "canonical sets" obtained from a sequence of packings that we build for \( S \). This decomposition is a variant of the one presented in [27, Chapter 5] (see also [25]), and is also referred to as chaining in the literature (see, e.g., [18]).

**Chaining.** For the sake of completeness, we repeat some of the details in [27, Chapter 5] and use similar notation for ease of presentation. Without loss of generality, we assume that \( \log n \) is an integer, and put \( k := \log n + 1 \). For \( j = 0, \ldots, k \), let \( \mathcal{F}_j \subseteq S \) be an \((n/2^j)\)-packing for \((X, S)\). By definition, this implies that for any pair of sets in \( \mathcal{F}_j \), their distance is larger than \( n/2^j \) and this set is maximal with respect to this property. In particular, we have \( \mathcal{F}_0 = \{ \emptyset \} \), and \( \mathcal{F}_k = S \) by construction.

Observe that for each \( F_j \in \mathcal{F}_j \), there is a set \( F_{j-1} \in \mathcal{F}_{j-1} \) with \( |F_j \triangle F_{j-1}| \leq n/2^{j-1} \). This follows from the inclusion-maximality of \( \mathcal{F}_j \). That is, consider the set \( F_{j-1} \) closest to \( F_j \) in \( \mathcal{F}_{j-1} \). Either \( F_j = F_{j-1} \) and then the claim is trivial, or \( F_j \neq F_{j-1} \), and then the opposite inequality \( |F_j \triangle F_{j-1}| > n/2^{j-1} \) is impossible, for then \( F_j \) could be added to \( \mathcal{F}_{j-1} \), contradicting its maximality. Let us attach to each \( F_j \in \mathcal{F}_j \) its nearest neighbor (closest set) \( F_{j-1} \in \mathcal{F}_{j-1} \), and put \( A(F_j) := F_j \setminus F_{j-1}, B(F_j) := F_{j-1} \setminus F_j \). We now form the set systems \( A_j := \{ A(F_j) \mid F_j \in \mathcal{F}_j \}, B_j := \{ B(F_j) \mid F_j \in \mathcal{F}_j \}, j = 1, \ldots, k \). It has been observed in [27, Chapter 5] that each set \( S \subseteq S \) can be decomposed as

\[
S = (\ldots (((A_1 \cup B_1) \cup A_2) \cup B_2) \cup \cdots \cup A_k) \cup B_k, \tag{1}
\]

where \( \cup \) denotes disjoint union, and \( A_j \subseteq A_j, B_j \subseteq B_j \), for each \( j = 1, \ldots, k \). Indeed, consider the nearest-neighbor “chain” \( S = F_k \to F_{k-1} \to \cdots \to F_0 = \emptyset \) (recall that \( F_0 = \emptyset \) by our assumption on \( \mathcal{F}_0 \), from which it follows that \( A_1 = F_1, B_1 = \emptyset \)). Each set \( F_j \subseteq \mathcal{F}_j \) on this chain can be turned into its nearest neighbor \( F_{j-1} \in \mathcal{F}_{j-1} \) by adding \( B(F_j) \) and subtracting \( A(F_j) \). Moreover, it is easy to verify using induction on \( j \geq 1 \) that \( F_j = (\ldots (((A_1 \cup B_1) \cup A_2) \cup B_2) \cup \cdots \cup A_j) \cup B_j \), and \( S \) is obtained at \( j = k \) (see also [30] for similar considerations).

We next refine this decomposition as follows. We partition the sets in \( S \) into \( k \) subfamilies \( S_1, \ldots, S_k \) where \( S \subseteq S_i \) if

\[
\frac{n}{2^i} < |S| \leq \frac{n}{2^{i-1}},
\]

for \( i = 1, \ldots, k \) (by definition, \( S_k \) contains all singleton sets). Fix an index \( i = 1, \ldots, k \). For each \( S \subseteq S_i \), we modify (1) by iterating from \( k \) down to \( i \), that is, we eliminate \( A_j, B_j \) from the decomposition for \( j = 1, \ldots, i - 1 \), and replace it by \( F_{i-1} \in \mathcal{F}_{i-1} \). Specifically, we have

\[
S = (\ldots (((F_{i-1} \cup A_i) \cup B_i) \cup A_{i+1}) \cup B_{i+1}) \cup \cdots \cup A_k) \cup B_k. \tag{2}
\]

Indeed, similarly to decomposition (1), it is easy to verify by induction on the index \( j \geq i \) of the sets that \( F_j = (\ldots (((F_{i-1} \cup A_i) \cup B_i) \cup A_{i+1}) \cup B_{i+1}) \cup \cdots \cup A_j) \cup B_j \), and our claim is obtained when \( j = k \). In particular, when \( i = 1 \) we obtain the same decomposition as in (1), as \( F_0 = \emptyset \).

Let us now fix an index \( i \), and construct the subsets \( \mathcal{F}_j^i \) of the packings \( \mathcal{F}_j \), for each \( j = i - 1, \ldots, k \), as follows. For each \( S \subseteq S_i \), we follow its nearest-neighbor chain (where at this time we stop at \( F_{i-1} \)) \( S = F_k \to F_{k-1} \to \cdots \to F_j \to \cdots \to F_{i-1} \), and put in \( \mathcal{F}_j^i \) the corresponding element \( F_j^i \in \mathcal{F}_j \), \( j = i - 1, \ldots, k \). We next show that the size of each of these members, which we now denote as \( F_j^i \in \mathcal{F}_j^i \), is at most \( O(|S|) = O(n/2^{i-1}) \). First we show:
Claim 3.1. For each of the sets $F_j^i \in \mathcal{F}_j^i$, $j = i - 1, \ldots, k$, we have:

$$|S \triangle F_j^i| < \frac{n}{2^{j-1}}.$$  

Proof. By construction we have $|F_j^i \triangle F_j^{i-1}| \leq n/2^{j-1}$, $F_j^i \in \mathcal{F}_j^i$, and $F_j^{i-1} \in \mathcal{F}_j^{i-1}$, for each $j = i, \ldots, k$.

We now apply the triangle inequality on the symmetric difference distance and obtain:

$$|S \triangle F_j^i| \leq |S \triangle F_k^i_{i-1}| + |F_{k-1}^i \triangle F_{k-2}^i| + \cdots + |F_{j+1}^i \triangle F_j^i|$$

$$\leq \frac{n}{2^{k-1}} + \frac{n}{2^{k-2}} + \cdots + \frac{n}{2^{j-1}} < \frac{n}{2^{j-1}},$$

as asserted.

Combining this with the fact that $|S| \leq n/2^{i-1}$ by construction (and the obvious relation $F_j^i \subseteq S \cup (S \triangle F_j^i)$), we obtain:

Corollary 3.2. For each of the sets $F_j^i \in \mathcal{F}_j^i$, $j = i - 1, \ldots, k$, we have:

$$|F_j^i| = O\left(\frac{n}{2^{i-1}}\right).$$

Remark: We note that by construction $|F_j^i \setminus F_{j-1}^i|, |F_{j-1}^i \setminus F_j^i|$ are bounded by $n/2^{j-1}$, and this fact is used later in the entropy method. Nevertheless, the property that the actual sets $F_j^i$ have size $O(n/2^{i-1})$ is stronger, and it enables us to prove a tighter bound on the number of the canonical sets $F_j^i$ (Theorem 3.3), and thus on the number of “pair sets” $F_j^i \setminus F_{j-1}^i, F_{j-1}^i \setminus F_j^i$. This is crucial for our analysis, as it enables us to reduce the bound on the entropy, from which we will derive the desired discrepancy bound—see below.

Bounding the size of the packing. Having a fixed index $i$ as above, we consider all sets $S \in \mathcal{S}_i$ and the corresponding canonical sets $F_j^i \in \mathcal{F}_j^i$ participating in decomposition (2), $j = i - 1, \ldots, k$. For a fixed index $j = i - 1, \ldots, k$, Theorem 2.2 implies that the size of $\mathcal{F}_j^i$ is $O(2^{jd})$, but our goal is to show that the actual bound can be made sensitive to the size of the sets in $\mathcal{F}_j^i$, that is, to $O(n/2^{i-1})$.

Theorem 3.3 (Sensitive Packing Lemma).

$$|\mathcal{F}_j^i| = O\left(\frac{j^{d}2^{jd}}{2^{d(d-1)}(i-1)}\right).$$

Proof. We use a variant of the analysis of Dudley [15] and refine it for our scenario. Put $\delta := n/2^j$. Since the following analysis will restrict sets $S \in \mathcal{S}$ to subsets of $X$, we will refer to $|S|$ more explicitly as $|S \cap X|$, for the sake of presentation.

We form the set system $(X, \mathcal{D})$, where $\mathcal{D} = \{S_1 \triangle S_2 \mid S_1, S_2 \in \mathcal{S}\}$. As observed in [15], this set system admits an $\varepsilon$-net of size $O((1/\varepsilon) \log (1/\varepsilon))$, with a constant of proportionality depending on $d$ (see

\footnote{This follows from the property that for each triple of sets $X, Y, Z$, we have $X \triangle Z \subseteq (X \triangle Y) \cup (Y \triangle Z)$.}
our discussion in Section 1). In fact, a random sample of that size with a sufficiently large constant of proportionality (that depends on \( d \)) is an \( \varepsilon \)-net with probability at least \( 3/4 \), say.

Set \( \varepsilon := \delta/n = 1/2^j \) and let \( N \subseteq X \) be an \( \varepsilon \)-net for \((X, \mathcal{D})\); \(|N| = O(j2^j)\), with a constant of proportionality as above. By the \( \varepsilon \)-net property, whenever the symmetric difference between two sets \( S_1, S_2 \in \mathcal{S} \) contains at least \( \varepsilon n = \delta \) elements, we must have \((S_1 \Delta S_2) \cap N \neq \emptyset\). We thus must have \( S_1 \cap N \neq S_2 \cap N \) for any pair of such sets. But this implies that each set in \( \mathcal{F}_j \) is mapped to a distinct set in the set system \((X, \mathcal{S})\) projected onto \( N \). Let \((N, \mathcal{S}|_N)\) be this set system. Thus the number of sets in \( \mathcal{F}_j \) is bounded by \(|\mathcal{S}|_N|\).

Recall that in view of Corollary 3.2 we are interested only in those sets whose size is \( O(n/2^i-1) \). We now claim that when we project \((X, \mathcal{S})\) onto \( N \), each set \( S \in \mathcal{S} \) with \(|S \cap X| = O(n/2^i-1)\) satisfies \(|S \cap N| = O(j2^j/2^i-1)\), with probability at least \( 3/4 \), and thus we only need to bound the number of sets of this size in the projected set system. Indeed, since \( N \) is a random sample of size \( O(j2^j) \) with a sufficiently large constant of proportionality (that depends on \( d \)), then it is also a relative \((1/2^j, 1/4)\)-approximation for \((X, \mathcal{S})\) with probability at least \( 3/4 \) (see our discussion in Section 1 and [19]). In particular, this means that \( N \) is both a \((1/2^j)\)-net for \((X, \mathcal{D})\) and a relative \((1/2^j, 1/4)\)-approximation for \((X, \mathcal{S})\) with probability at least \( 1/2 \) (and thus we obtain a single sample with a “double role”). The latter property implies that

\[
\frac{|S \cap N|}{|N|} - \frac{|S \cap X|}{|X|} \leq \frac{1}{4} \cdot \frac{|S \cap X|}{|X|},
\]

if \( \frac{|S \cap X|}{|X|} \geq 1/2^j \), and

\[
\frac{|S \cap N|}{|N|} - \frac{|S \cap X|}{|X|} \leq \frac{1}{4} \cdot \frac{1}{2^j},
\]

otherwise. Combining the facts that \(|S \cap X| = O(n/2^i-1), j \geq i - 1\), and the bound on \(|N|\), we obtain that \(|S \cap N| = O(j2^j/2^i-1)\), as asserted. In particular, due to Corollary 3.2, this is also the bound on the cardinality of the sets \( F_j \in \mathcal{F}_j \) restricted to \( N \).

Recall that \( \mathcal{S}|_N \) is the family \( \mathcal{S} \) projected onto \( N \), and, due to the double role of \( N \), each \( F_j \in \mathcal{F}_j \) is mapped to a distinct set of \( \mathcal{S}|_N \) of size \( O(j2^j/2^i-1) \). By assumption, the number of sets of \((N, \mathcal{S}|_N)\) of size \( 1 \leq k \leq |N| \) is \( O(|N|^{d_1}k^d-d_1) \), and hence the number of its sets of size \( O(j2^j/2^i-1) \) is at most

\[
O \left( |N|^{d_1} (j2^j/2^i-1)^{d-d_1} \right) = O \left( \frac{jd^2j^d}{2(d-d_1)(i-1)} \right),
\]

from which we obtain the bound on \(|\mathcal{F}_j|\).

\(\square\)

**Applying the entropy method.** Let us fix an index \( i \) for the family \( \mathcal{S}_i \) under consideration. Returning to decomposition (2), we denote, with a slight abuse of notation, the canonical sets \( A_j, B_j \) by \( A^i_j, B^i_j \), respectively. By construction, \( A^i_j = F_j^i \setminus F_{j-1}^i, B^i_j = F_{j-1}^i \setminus F_j^i \). Let \( \mathcal{M}_j^i \) be the collection of these sets \( A^i_j, B^i_j \) (or \( F_j^i \) if \( j = i - 1 \)). We now set a parameter \( \Delta_j^i \) for the discrepancy bound (with respect to partial coloring) for each of the canonical sets in \( \mathcal{M}_j^i \), where all sets in this subcollection are assigned the same discrepancy bound \( \Delta_j^i \). We then use the entropy method on this new set system in order to obtain a partial coloring \( \chi \) achieving the pre-assigned discrepancy bounds. Having these bounds at hand, we can
then conclude that the discrepancy of any \( S \in S_i \) with respect to \( \chi \) is at most \( 2 \sum_{j=i-1}^k \Delta_j \) (using standard arguments, see, e.g., [27]), and this will yield the desired size-sensitive bound for \( \chi(S) \).

In order to apply the entropy method as presented in Proposition 2.1 we need to have, for each \( j = i - 1, \ldots, k \), (i) a bound on \( |\mathcal{M}_j^i| \), (ii) a bound on the size of the canonical sets in \( \mathcal{M}_j^i \), and (iii) an appropriate choice for the parameter \( \Delta_j \), uniformly assigned to all these sets. We set each of the bounds in (i)–(iii) for a fixed index \( i \), and then sum them up over all \( i = 1, \ldots, k \).

For the bound in (i), we observe that each canonical set \( A_j \) (resp., \( B_j \)) corresponds to a pair \((F_j^i, F_{j-1}^i)\), but each of these pairs can uniquely be charged to \( F_j^i \), as \( F_{j-1}^i \) is the corresponding nearest neighbor in the packing \( \mathcal{F}_{j-1}^i \). Thus the number of these canonical sets is \( |\mathcal{F}_j^i| \), for \( j = i, \ldots, k \). For \( j = i - 1 \), the bound is trivially \( |\mathcal{F}_j^i| \). Thus by applying the Sensitive Packing Lemma (Theorem 3.3) we conclude that the overall number of canonical sets is at most \( C \cdot \frac{j^d2^jd}{2^{(d-d_1)(i-1)}} \), for \( j = i - 1, \ldots, k \), where \( C > 0 \) is an appropriate constant whose choice depends on \( d \). By construction, the size of the sets \( A_j, B_j \) is at most \( s_j = n/2^{j-1} \) (recall that \( |F_j^i\Delta F_{j-1}^i| \leq n/2^{j-1} \) and \( |F_{j-1}^i| = O(n/2^{j-1}) \) by Corollary 3.2, whence we get the bound for (ii). For the choice in (iii) we set:

\[
\Delta_j := A \cdot \frac{1}{(1 + |j - j_0|)^2} \left( \frac{n^{1/2-1/(2d)}}{2^{(i-1)(1/2 - d_1/(2d))}i} \right) \log^{1/2+1/2d} n,
\]

where \( j_0 := (1/d) \log n + (1 - d_1/d)(i - 1) - (1 + 1/d) \log \log n - B \), for an appropriate constant \( B > 5 + \log C \), and for a sufficiently large constant of proportionality \( A > 0 \), whose choice depends on \( B \), and will be determined shortly (note that all the three constants \( A, B, \) and \( C \) depend on \( d \)).

We first provide a brief justification for our choice of \( j_0 \) and the appearance of the coefficient \( \frac{1}{(1 + |j - j_0|)^2} \). For the first, our goal is to bound the entropy function, bounded by the sum (4) appearing in Proposition 3.4, and at \( j = j_0 \) the corresponding summands achieve their maximum value (which is some linear function of \( n/\log n \) with an appropriate constant of proportionality). Then when \( j \geq j_0 \) the exponents “take over” this summation, in which case it decreases superexponentially, and when \( j < j_0 \) the terms \( C \cdot \frac{j^d2^jd}{2^{(d-d_1)(i-1)}} \) representing the packings take over this summation and decrease geometrically. This eventually leads to the bound stated in Proposition 3.4. The coefficient \( \frac{1}{(1 + |j - j_0|)^2} \) in (3) guarantees that the sum \( \sum_{j=i-1}^k \Delta_j \) (corresponding to the asymptotic bound on the discrepancy of any \( S \in S_i \)) converges to \( O \left( \frac{n^{1/2-1/2d}}{2^{(i-1)(1/2 - d_1/(2d))}i} \log^{1/2+1/2d} n \right) \). In particular, it does not contain an extra logarithmic factor over the initial bound of \( \Delta_j \)—see below. Similar ideas have been used in [27]. We defer the remaining technical details to Appendix A, where we present the proof of Proposition 3.4.

**Proposition 3.4.** The choice in (3), for \( A > 0 \) sufficiently large (whose choice depends on \( C \) and thus on \( d \)), satisfies

\[
\sum_{i=1}^k \sum_{j=i-1}^k C \cdot \frac{j^d2^jd}{2^{(d-d_1)(i-1)}} \exp \left( - \frac{(\Delta_j)^2}{16s_j} \right) \leq \frac{n}{16},
\]

**Remark:** Currently, our analysis is slightly suboptimal, as our bound in (3) contains an extra fractional logarithmic power, which results from the following reasons: (i) We may overcount in our bound on the
entropy function the same set $A_j^i$ (or $B_j^i$) when we iterate over $i = 1, \ldots, k$; recall that for each $S \in S_i$ we follow its nearest-neighbor chain and put in each $F_j^i$ the corresponding element from $F_j$, thus an element from $F_j$ may appear in multiple layers $i$. (ii) In the Sensitive Packing Lemma (Theorem 3.3) the size of the random sample $N$ is $O((n/\delta) \log (n/\delta))$, whereas in the analysis of the original Packing Lemma (Theorem 2.2), it is sufficient to have a sample of $(n/\delta)$ elements, resulting in the bound $O((n/\delta)^d)$. Nevertheless, due to the fact that our sample needs also to be a relative approximation (to exploit the property that our packing consists of sets of size at most $O(n/2^{i-1})$), we had to use a slightly larger sample. It is an interesting open problem whether the extra logarithmic factor can be removed from the bound on the size of the sensitive packing. If so, this will imply that the actual bound in Theorem 3.3 is $O\left(\frac{2^d}{2^d-d_1(n/\delta)}\right)$.

Wrapping up. Incorporating Propositions 2.1 and 3.4, we obtain that there exists a partial coloring $\chi$, computable in expected polynomial time, which colors at least $n/2$ points of $X$, such that $\chi(M_j^i) \leq \Delta_j + 1/n^c$, for each $M_j^i \in M_j^i$, where $c > 0$ is an arbitrarily large constant. But then our choice in (3) and our earlier discussion imply that, for each $S \in S_i$,

$$\chi(S) \leq 2 \sum_{j=1}^{k} \Delta_j = O\left(\frac{n^{1/2-1/(2d)}}{2(1-1/2-\ldots-1/(2d))} \log^{1/2+1/2d} n\right),$$

since the sum $\sum_{j=1}^{k} \frac{1}{(1+1/j-3a)^2}$ converges. Due to the fact that $n/2^i \leq |S| < n/2^{i-1}$ (by definition), the latter term is:

$$O\left(|S|^{1/2-d_1/(2d)} n^{(d_1-1)/(2d)} \log^{1/2+1/(2d)} n\right).$$

Applying the partial coloring procedure (Proposition 2.1) for at most $\log n$ iterations (which yields a full coloring for $X$), we obtain that for each $S \in S_i$, $\chi(S) = O\left(|S|^{1/2-d_1/(2d)} n^{(d_1-1)/(2d)} \log^{1/2+1/2d} n\right)$, if $d_1 > 1$, and $\chi(S) = O\left(|S|^{1/2-1/(2d)} \log^{3/2+1/2d} n\right)$, if $d_1 = 1$ (recall that we assume $1 \leq d_1 \leq d$ is an integer parameter). This is because, for $d_1 > 1$, the term $n^{(d_1-1)/(2d)}$ creates a decreasing geometric sequence over the at most $\log n$ iterations. We have just shown:

**Theorem 3.5.** Let $(X, S)$ be a (finite) set system of primal shatter dimension $d$ with the additional property that in any set system restricted to $X' \subseteq X$, the number of sets of size $k \leq |X'|$ is $O(|X'|^{d_1} k^{d-d_1})$, where $d_1$ is an integer parameter between 1 and $d$. Then

$$\text{disc}(S) = \begin{cases} O\left(|S|^{1/2-d_1/(2d)} n^{(d_1-1)/(2d)} \log^{1/2+1/2d} n\right), & \text{if } d_1 > 1, \\ O\left(|S|^{1/2-1/(2d)} \log^{3/2+1/2d} n\right), & \text{if } d_1 = 1, \end{cases}$$

where the constant of proportionality depends on $d$.

**Remark:** We note that although the number of uncolored points in $X$ decreases by at least a half after applying Proposition 2.1, it does not necessarily guarantee that the size of a set $S \in S_i$ decreases by the same factor, and so we can bound it from above only by $n/2^i$ at each round. Moreover, it may happen that a set $S \in S_i$ from the previous round now lies in a different class at the current partition. Thus at each round we need to resume the process from scratch, due to which we obtain an extra logarithmic factor for $d_1 = 1$, as shown in the bound above.
Algorithmic aspects. In order to apply the randomized algorithm of Lovett and Meka [24], we first need to construct, for each $i = 1, \ldots, k$, the canonical sets in $\mathcal{F}_j$.

In order to do so for a fixed $i$, we need to construct a $\delta$-packing, for $\delta = n / 2^j$, $j = i - 1, \ldots, k$, such that the size of each set in the packing does not exceed $K n / 2^{i-1}$, for an appropriate constant $K > 0$. We thus iterate over $j = i - 1, \ldots, k$, and form a $\delta$-packing $\mathcal{F}_j$ as above in a brute force manner by initially picking an arbitrary set $F \in \mathcal{S}$, whose size is at most $K n / 2^{i-1}$, to be included into $\mathcal{F}_j$, and then keep collecting sets $F' \in \mathcal{S}$ into $\mathcal{F}_j$ if (i) $|F'| \leq K n / 2^{i-1}$ and (ii) the symmetric difference distance between $F'$ and each of the elements currently in $\mathcal{F}_j$ is greater than $\delta$. We stop as soon as there are no leftover sets $F'$ of the above kind. The set just created is inclusion-maximal and thus according to the Sensitive Packing Lemma (Theorem 3.3) its size is only $O \left( \frac{j^{d2/d} d}{2^{(d-d_1)(i-1)}} \right)$. It is easy to verify that the construction of each $\delta$-packing $\mathcal{F}_j$ can be performed in polynomial time due to the fact that the number of sets in $\mathcal{S}$ is only $O(n^d)$. Omitting any further details we conclude:

Corollary 3.6. A coloring $\chi$ achieving the discrepancy bound in Theorem 3.5 can be computed in expected polynomial time.

The case of points and halfspaces in $d$ dimensions. When $(X, \mathcal{S})$ is a set system of points and halfspaces in $d$-space, it is known that the number of halfspaces containing at most $k$ points of $\mathcal{S}$ is $O(n^{\lfloor d/2 \rfloor} k^{\lceil d/2 \rceil})$ (see, e.g., [17]). Thus from Theorem 3.5 and Corollary 3.6 we conclude:

Theorem 3.7. Let $(X, \mathcal{S})$ be a set system of points and halfspaces in $d$-space. Then

$$\text{disc}(\mathcal{S}) = O \left( |\mathcal{S}|^{1/4} n^{1/4-1/(2d)} \log^{1/2+1/2d} n \right),$$

for $d \geq 4$ even, and

$$\text{disc}(\mathcal{S}) = O \left( |\mathcal{S}|^{1/4+1/(4d)} n^{1/4-3/(4d)} \log^{1/2+1/2d} n \right),$$

for $d \geq 5$ odd, where the constant of proportionality in these bounds depends on $d$. In particular, when $d = 3$, the bound is $O \left( |\mathcal{S}|^{1/3} \log^{5/3} n \right)$. In each of the above cases the corresponding coloring $\chi$ can be computed in expected polynomial time.

Remark: When $d = 2$ the resulting bound is $O(|\mathcal{S}|^{1/4} \log^{7/4} n)$. However, this is slightly suboptimal with respect to the bound $O(|\mathcal{S}|^{1/4} \log n)$ shown in [19].

3.1 An Application to Relative $(\varepsilon, \delta)$-Approximations.

We next show how to obtain improved bounds on the size of relative $(\varepsilon, \delta)$-approximations, using a variant of the halving technique presented in [19] (which was originally formulated in [29], and is overviewed in detail in [13]). A major technical difficulty in our analysis is the fact that, unlike the coloring properties exploited in [19, 32], our coloring is not necessarily balanced, that is, the red-blue split on $X$ is not necessarily even. This situation is somewhat different and requires extra care when applying the halving technique in [19]. We defer the analysis and the technical details to Appendix B and state our main result below.
Theorem 3.8. Let \((X, S)\) be a (finite) set system of primal shatter dimension \(d\) with the additional property that in any set system restricted to \(X' \subseteq X\), the number of sets of size \(k \leq |X'|\) is \(O(|X'|^{d_1 k^{d-d_1}})\), where \(d_1\) is an integer between 1 and \(d\). Then \((X, S)\) admits a relative \((\varepsilon, \delta)\)-approximation of size

\[
\max \left\{ O \left( \frac{\log \frac{1}{\varepsilon}}{\frac{d_1}{\varepsilon} \frac{2d}{\delta \frac{n}{\varepsilon}}} \right), \right.
\]

for \(d_1 > 1\), and

\[
\max \left\{ O \left( \frac{\log \frac{1}{\varepsilon}}{\frac{d_1}{\varepsilon} \frac{2d}{\delta \frac{n}{\varepsilon}}} \right), \right.
\]

for \(d_1 = 1\), where the constant of proportionality depends on \(d\). Moreover, such a sample can be constructed in expected polynomial time.

When \((X, S)\) is a set system of points and halfspaces in \(d\)-space we thus conclude:

Corollary 3.9. Let \((X, S)\) be a set system of points and halfspaces in \(d\)-space. Then \((X, S)\) admits a relative \((\varepsilon, \delta)\)-approximation of size

\[
\max \left\{ O \left( \frac{\log \frac{1}{\varepsilon}}{\frac{d_1}{\varepsilon} \frac{2d}{\delta \frac{n}{\varepsilon}}} \right), \right.
\]

for \(d \geq 4\), where the constant of proportionality depends on \(d\). When \(d = 3\) this bound is

\[
\max \left\{ O \left( \frac{\log \frac{1}{\varepsilon}}{\frac{d_1}{\varepsilon} \frac{2d}{\delta \frac{n}{\varepsilon}}} \right), \right.
\]

Concluding remarks and further research. We note that whereas our construction is a variant of that of Matoušek [25], a key ingredient in our analysis is the Sensitive Packing Lemma (Theorem 3.3), where we restrict each set under consideration to have a bounded size. As our analysis shows, when these sets are relatively small, the bound on the size of the packing is considerably smaller than the bound \(O((n/\delta)^d)\) derived in the original Packing Lemma (Theorem 2.2). This bound is eventually integrated into the entropy method applied in Proposition 3.4 (in addition to our decomposition (2)), from which we eventually obtain the discrepancy bound in (3).

This study raises several open problems. A main problem, concerning the remarks following Proposition 3.4 and Theorem 3.5, is whether the logarithmic factor in our discrepancy bound can be reduced or even removed completely, in which case it becomes optimal. A major step towards this goal is to remove the polylogarithmic factor on the bound in the Sensitive Packing Lemma (Theorem 3.3). The current analysis is based on the suboptimal bound of Dudley [15], who showed a bound of \(O((n/\delta)^d \log^d (n/\delta))\) on the largest \(\delta\)-separated set in a set system of primal shatter dimension \(d\). Thus the question at hand is whether the analysis of Haussler [21], who showed the bound \(O((n/\delta)^d)\), as stated in Theorem 2.2, which is optimal up to a constant factor, can be adapted to the scenario of our problem.

Another related problem concerns the discrepancy bounds in geometric set systems of “low degree”. Specifically, in abstract set systems \((X, S)\), this is the so-called Beck-Fiala problem, where each point of \(X\) appears in at most \(t \leq n\) sets of \(S\) (where \(n = |X|\)). In this case, the discrepancy bound is conjectured to be \(O(\sqrt{t})\) [11]. In fact, using the entropy method, one can obtain a constructive bound of \(O(\sqrt{t} \log n)\) [7, 24].
(see also [34]), where the best currently known bound is by Banaszczyk [6], who showed a non-constructive bound of $O(\sqrt{t \log n})$. The question at hand is what should be the corresponding bounds in geometric set systems. Specifically, assume we have a set system of $n$ points and halfspaces in $d$-space, is it possible to obtain a bound, which is roughly $o(\sqrt{t})$? In an on-going work, the author has shown that such bounds exist in two and three dimensions. We note that such bounds do not follow directly from the work on this paper, as in such settings $S$ may contain a few large sets, whereas the bound on the discrepancy should depend only on the degree $t$ (say, up to a logarithmic factor in $n$).

Last but not least, for further research, we suggest to integrate our bound on relative $(\varepsilon, \delta)$-approximations for points and halfspaces in $d$ dimensions (Corollary 3.9) with the existing approximate range counting machinery in [5], for $d \geq 4$, and [19], for $d = 3$. We hope this will improve the current performance bounds.

Acknowledgments. The author wishes to thank Boris Aronov and Micha Sharir for helpful discussions and for their help in writing this paper. The author also wishes to thank Jiri Matoušek for several useful comments about this paper and Shachar Lovett for several discussions concerning discrepancy theory.

References


A Proof of Proposition 3.4:

We first note that at \( j_0 \) the above exponent becomes a constant, whereas the size of the packing becomes roughly \( n/ \log n \) (for a fixed index \( i \)). Indeed, applying our choice in (3), we have

\[
\exp \left( -\frac{(\Delta_j^i)^2}{16s_j} \right) = \exp \left( -\frac{A^2 \cdot 2^{j-1} \log^{1+1/d} n}{16(1 + |j - j_0|)4^{n/d}2^{(1-d_1/d)(i-1)}} \right),
\]

which is \( \exp \left( -\frac{A^2}{16 \cdot 2^{d+1}} \right) \) at \( j = j_0 = (1/d) \log n + (1 - d_1/d)(i-1) - (1 + 1/d) \log \log n - B \). Concerning the bound on the packing size, \( C \cdot \frac{n2^{(d-d_1)(i-1)}}{2^{d-d_1}(i-1)2^{dB} \log^{d+1} n} \), since \( j \) can always be bounded by \( k = \log n \), at \( j = j_0 \) we obtain:

\[
C \cdot \frac{n2^{(d-d_1)(i-1)}}{2^{d-d_1}(i-1)2^{dB} \log^{d+1} n} = C \cdot \frac{n}{2^{dB} \log n}.
\]

We now fix an index \( i \), split the summation into the two parts \( j \geq j_0 \) and \( i-1 \leq j < j_0 \), and then bound each part in turn. In the first part, the exponent will “take over” the summation in the sense that it decreases superexponentially, making the other factors (with \( j > j_0 \)) insignificant, and in the second part, the packing size will decrease geometrically. Thus the “peak” of this summation is obtained at \( j = j_0 \), and is decreasing as we go beyond or below \( j \).

For the first part, put \( j := j_0 + l \), for an integer \( l \geq 0 \), and then

\[
\sum_{i=1}^{k} \sum_{j=j_0}^{k} C \cdot \frac{j^{d}2^{j}d}{2^{d-d_1}(i-1)} \exp \left( -\frac{(\Delta_j^i)^2}{16s_j} \right)
\]
\[
\leq \sum_{i=1}^{k} \sum_{l=0}^{k-j_0} C \cdot \frac{n2^{ld}}{2^{dB} \log n} \exp \left( -\frac{A^2}{16} \cdot \frac{2^{l-(B+1)}}{(1+l)^4} \right)
\]

\[
\leq C \cdot n2^{-dB} \sum_{l=0}^{k-j_0} 2^{ld} \exp \left( -\frac{A^2}{16} \cdot \frac{2^{l-(B+1)}}{(1+l)^4} \right),
\]

where the logarithmic factor in the packing size is now eliminated due to the summation over \(i\). The exponents in the above sum decrease superexponentially. Choosing \(A\) sufficiently large (say, \(A > 2^{6+(B+1)+\log d}\)) and having \(B > 5 + \log C\) as above, we can guarantee that the latter sum is strictly smaller than \(n/32\).

When \(j < j_0\), put \(j := j_0 - l\), \(l > 0\) as above. We now obtain, by just bounding the exponent from above by 1, and using similar considerations as above:

\[
\sum_{i=1}^{k} \sum_{j=1-i}^{j_0-1} C \cdot \frac{j^{d2^ld}}{2^{d(d-d_1)(i-1)}} \exp \left( -\frac{A_1^2}{16s_j} \right) \leq \sum_{l=1}^{j_0-(i-1)} C \cdot \frac{n}{2^{d(l+B)}}.
\]

Once again, our choice for \(B\) guarantees that the above (geometrically decreasing) sum is strictly smaller than \(n/32\). Thus the entire summation is bounded by \(n/16\), as asserted.

\[\text{B Proof of Theorem 3.8}\]

Following the arguments in [19], it is sufficient to construct a \((\nu, \alpha)\)-sample for \((X, S)\) as such a sample is equivalent to a relative \((\varepsilon, \delta)\)-approximation (where \(\nu, \varepsilon\) and \(\alpha, \delta\) are within some constant factor from each other—see Section 1).

Our construction proceeds over iterations, where we repeatedly “halve” \(X\) until we obtain a subset of an appropriate size, which we argue to comprise the resulting \((\nu, \alpha)\)-sample. Put \(X_0 := X\). Then, at each iteration \(i \geq 1\), we let \(X_i, X'_i\) be the two corresponding portions of \(X_{i-1}\), where the points in \(X_i\) are, say, colored +1 and the points in \(X'_i\) are colored −1. Assume, w.l.o.g., \(|X_i| \geq |X'_i|\). We now keep \(X_i\), remove \(X'_i\) and continue in this “halving” process, and thus the desired \((\nu, \alpha)\)-sample corresponds to some set \(X_i\) obtained in an appropriate iteration. Put \(n_i := |X_i|\). On an even split we have \(n_i = n/2^i\) (recall that we assume \(n\) is an integer power of 2), nevertheless, \(n_i\) may be slightly larger when applying the coloring \(\chi\) corresponding to the discrepancy bound in Theorem 3.5. We bound its size as follows.

We first assume, w.l.o.g., \(X = X_0\) is part of \(S\), and thus, at each iteration \(i\), \(X_{i-1}\) is part of the collection \(S\) projected onto \(X_{i-1}, i \geq 1\). Applying Theorem 3.5 at iteration \(i\) we obtain6:

\[
||S \cap X_i| - |S \cap X'_i|| \leq K \cdot |S \cap X_{i-1}|^{1/2-d_1/(2d)} |X_{i-1}|^{(d_1-1)/(2d)} \log^{1/2+1/(2d)} |X_{i-1}|, \tag{5}
\]

for an appropriate constant \(K > 0\) which depends on \(d\). Letting \(S = X_{i-1}\), and using the fact that \(|X_i| + |X'_i| = |X_{i-1}|\), we obtain:

\[
||X_i| - (|X_{i-1}| - |X_i|)| \leq K \cdot |X_{i-1}|^{1/2-d_1/(2d)} |X_{i-1}|^{(d_1-1)/(2d)} \log^{1/2+1/(2d)} |X_{i-1}|
\]

\[
= K \cdot |X_{i-1}|^{1/2-1/(2d)} \log^{1/2+1/(2d)} |X_{i-1}|,
\]

6For the time being, we assume \(d_1 > 1\) and use the corresponding discrepancy bound in Theorem 3.5. The case \(d_1 = 1\) is handled later on, and its analysis follows almost verbatim—see below.
From the bound on $\delta$

By adding and subtracting $i$ process at iteration $n$ and then integrate the resulting bound on $X$ from which we obtain

$|X| \leq \frac{|X_{i-1}|}{2} \left(1 + \frac{K \log^{1/2+1/(2d)} |X_{i-1}|}{|X_{i-1}|^{1/2+1/(2d)}}\right)$.

We thus write the bound on $|X|$ as

$|X| = \frac{|X_{i-1}|}{2} (1 + \delta_{i-1})$,  \hspace{1cm} (6)

where $0 \leq \delta_{i-1} \leq \frac{K \log^{1/2+1/(2d)} |X_{i-1}|}{|X_{i-1}|^{1/2+1/(2d)}}$. Applying (6) recursively on $i$, we obtain:

$|X| = \frac{|X_0|}{2^i} \prod_{j=0}^{i-1} (1 + \delta_j) \leq \frac{|X_0|}{2^i} \exp \left\{ \sum_{j=0}^{i-1} \delta_j \right\}$.

From the bound on $\delta_j$ we have (recall that $n_i = |X_i|$ and $|X_i| \geq |X_{i-1}|/2$ by assumption):

$n_i \leq \frac{n}{2^i} \exp \left\{ K (\log n)^{1/2+1/(2d)} \sum_{j=0}^{i-1} \left( \frac{2^j}{n} \right)^{1/2+1/(2d)} \right\}$,  \hspace{1cm} (7)

and the exponent in the latter term is $O(1)$ when $i \leq i^* = \log n - \log \log n - \log^2 \frac{2d}{d+1} K$. We thus stop the process at iteration $i^*$ (or earlier—see below), from which we obtain a lower bound of $\Omega(\log n)$ on $n_{i-1}$ (with a constant of proportionality that depends on $K$).

We next proceed with the presentation of the “halving” process in order to obtain a relative error of $\alpha$, and then integrate the resulting bound on $n_{i-1}$ with the one above. We use a variant of the considerations in [19] and proceed as follows. Our goal is to bound, for each $S \in \mathcal{S}$, the difference $\frac{|S \cap X_{i-1} / |X_{i-1}| - |S \cap X_i / |X_i| |}{|X_{i-1}|}$, which we also denote by $|X_{i-1}(S) - X_i(S)|$ (recall our notation for a measure of a set from Section 1).

Since $|X_i| = \frac{|X_{i-1}(1+\delta_{i-1})}{2}$ and $X_{i-1} = X_i \cup X_i'$, this difference is

$$\left| \frac{|S \cap X_i| + |S \cap X_i'|}{|X_{i-1}|} - \frac{2|S \cap X_i|}{|X_{i-1}|(1 + \delta_{i-1})} \right| = \left| \frac{|S \cap X_i'|}{|X_{i-1}|} - \frac{|S \cap X_i|}{|X_{i-1}|(1 + \delta_{i-1})} \right|.$$

By adding and subtracting $\frac{|S \cap X_i'|}{|X_{i-1}|}$, we obtain that the latter term is:

$$\left| \frac{|S \cap X_i'|}{|X_{i-1}|} - \frac{|S \cap X_i|}{|X_{i-1}|(1 + \delta_{i-1})} \right| = \left| \frac{|S \cap X_i'|}{|X_{i-1}|} - \frac{|S \cap X_i|}{|X_{i-1}|} \right| + \frac{2\delta_{i-1}}{1 + \delta_{i-1}} \cdot \frac{|S \cap X_i|}{|X_{i-1}|(1 + \delta_{i-1})}.$$

since $|X_{i-1}| = \frac{2|X_i|}{1 + \delta_{i-1}}$. We have thus shown:

$$|X_{i-1}(S) - X_i(S)| \leq \left| \frac{|S \cap X_i'|}{|X_{i-1}|} - \frac{|S \cap X_i|}{|X_{i-1}|} \right| + \delta_{i-1} \cdot X_i(S).$$  \hspace{1cm} (8)
Using (5), we obtain:
\[
|X_{i-1}(S) - X_i(S)| \leq K \cdot |S \cap X_{i-1}|^{1/2-d_i/(2d)} |X_{i-1}||(d_1-1)/(2d) \log^{1/2+1/2d} |X_{i-1}|
+ \delta_{i-1} \cdot X_i(S).
\]

The latter term in the above sum is obviously bounded by \(\delta_{i-1} \cdot (X_i(S) + X_{i-1}(S) + \nu)\). Concerning the first term, we rewrite it as
\[
K \cdot |X_{i-1}(S)|^{1/2-d_i/(2d)} |X_{i-1}|^{1/2+1/2d} \log^{1/2+1/2d} |X_{i-1}|
= K \cdot |X_{i-1}(S)|^{1/2-d_i/(2d)} \log^{1/2+1/2d} |X_{i-1}|,
\]
and use the observation that \(x^p \leq (x+y)^{1-p}\), for \(x \geq 0, y > 0\) and \(0 \leq p \leq 1\) (stated in [19]) in order to bound the latter term by (we now replace \(|X_{i-1}|\) with \(n_{i-1}\), and set \(p := 1/2 - d_1/(2d)\)):
\[
K \cdot \log^{1/2+1/2d} n_{i-1} \cdot \frac{X_{i-1}(S) + \nu}{\nu^{1/2+d_i/(2d)}}
\leq K \cdot \log^{1/2+1/2d} n_{i-1} \cdot \frac{X_i(S) + X_{i-1}(S) + \nu}{\nu^{1/2+d_i/(2d)}}.
\]

This implies that
\[
d_\nu(X_{i-1}(S), X_i(S)) = \frac{|X_{i-1}(S) - X_i(S)|}{X_i(S) + X_{i-1}(S) + \nu}
\leq \frac{K \cdot \log^{1/2+1/2d} n_{i-1}}{\nu^{1/2+d_i/(2d)} n_{i-1}^{1/2+1/2d}} + \delta_{i-1}
\leq \frac{K \cdot \log^{1/2+1/2d} n_{i-1}}{(n_{i-1})^{1/2+1/2d}} \left(\frac{1}{\nu^{1/2+d_i/(2d)}} + 1\right),
\]
due to the bound on \(\delta_{i-1}\).

Since \(d_\nu(\cdot, \cdot)\) satisfies the triangle inequality (see [23]), we obtain:
\[
d_\nu(X_0(S), X_i(S)) \leq \sum_{j=1}^i d_\nu(X_{j-1}(S), X_j(S))
\leq K \cdot \left(\frac{1}{\nu^{1/2+d_i/(2d)} + 1}\right) \sum_{j=1}^i \frac{\log^{1/2+1/2d} n_{j-1}}{n_{j-1}^{1/2+1/2d}}.
\]
By (7) \( n_j = O(n/2^j) \), for each \( j < i^* \), and since we also have \( n_j \geq n/2^j \) by assumption, we obtain that the latter expression is bounded by

\[
O \left( \frac{\log^{1/2 + 1/(2d)} n_{i-1}}{\log^{1/2 + d_1/(2d)} n_{i-1}} \right).
\]

We next bound \( d_\nu(X_0(S), X_i(S)) \) by the relative error \( \alpha \), in order to conclude that this is valid as long as

\[
n_{i-1} = \Omega \left( \frac{\log^{1/2 + 1/(2d)} n_{i-1}}{\log^{1/2 + d_1/(2d)} n_{i-1}} \right).
\]

We thus stop at that iteration \( i \) for which the set \( X_{i-1} \) is the smallest that still satisfies this lower bound.

Combining these considerations with the fact that we stop the process no later than iteration \( i^* \), we obtain

\[
n_{i-1} = \max \left\{ \Omega \left( \log n \right), \Omega \left( \frac{\log \frac{1}{\nu \alpha}}{\log^{1/2 + d_1/(2d)} n_{i-1}} \right) \right\},
\]

from which we obtain the asserted bound on the size of the \((\nu, \alpha)\)-sample.

In case \( d_1 = 1 \), the analysis proceeds almost verbatim, with the slight difference that now \( 0 \leq \delta_{i-1} \leq K \frac{\log^{3/2 + 1/(2d)} |X_{i-1}|}{|X_{i-1}|^{1/2 + 1/(2d)}} \). Following similar considerations, this leads to the lower bound

\[
\max \left\{ \Omega \left( \log^{3d+1 \delta_{i-1}} n \right), \Omega \left( \frac{\log^{3d+1 \delta_{i-1}} \frac{1}{\nu \alpha}}{\log^{2d/(d+1)} n \alpha} \right) \right\},
\]

on \( n_{i-1} \).