PIERCING RANDOM BOXES

LINH V. TRAN

ABSTRACT. Consider a set of n random axis parallel boxes in the unit hypercube in $\mathbf{R}^d$, where d is fixed and n tends to infinity. We show that the minimum number of points one needs to pierce all these boxes is, with high probability, at least $\Omega_d(\sqrt{n}(\log n)^{d/2-1})$ and at most $O_d(\sqrt{n}(\log n)^{d/2-1}\log\log n)$.

1. INTRODUCTION

Let $\mathbf{U}_d$ be the unit hypercube in $\mathbf{R}^d$, where d is a constant: $\mathbf{U}_d := [0,1]^d$. We generate a d-dimensional random box by taking the product of d independent random sub-intervals of $[0,1]$, where each random interval is determined by two random (end-) points, each chosen independently with respect to the uniform measure on $[0,1]$.

Consider a family of n (independently) random boxes. Our investigation is motivated by the following basic questions:

What is the size of the largest sub-family of pairwise disjoint boxes ?

What is the minimum number of points one needs to pierce all the boxes ?

Let us denote these quantities by $\nu_d(n)$ and $\tau_d(n)$, respectively. These quantities, usually referred to as the matching and covering (piercing) numbers, are of fundamental interest and have been studied for a large variety of hypergraphs, deterministic and random alike. It is useful to notice that the piercing number is at least the matching number, by definition.

About ten years ago, Coffman, Lueker, Spencer and Winkler studied $\nu_d(n)$ [1]. They showed

Theorem 1.1. *For $d = 2$, $\nu_2(n) = \Theta(\sqrt{n})$. For $d \geq 3$*

$$\Omega(\sqrt{n}) \leq \nu_d(n) \leq O(\sqrt{n\log^{d-1} n}). \quad (1)$$

The focus of this paper is on the piercing number. We are able to prove a nearly sharp estimate for this quantity

Department of Mathematics, Rutgers University.

Theorem 1.2. *For any fixed $d \geq 2$, we have, almost surely, that*

$$\Omega_d(\sqrt{n}(\log n)^{d/2-1}) = \tau_d(n) = O_d(\sqrt{n}(\log n)^{d/2-1}\log\log n).$$

We conjecture that the lower bound is the truth.

Conjecture 1.3. *For any fixed $d \geq 2$, almost surely,*

$$\tau_d(n) = O_d(\sqrt{n}(\log n)^{d/2-1}).$$

Since $\tau_d(n) \geq \nu_d(n)$, Theorem 1.2 improved (the general case of) Theorem 1.1.

Corollary 1.4. *For $d \geq 3$*

$$\Omega_d(\sqrt{n}) = \nu_d(n) = O_d(\sqrt{n}(\log n)^{d/2-1}\log\log n). \tag{2}$$

These results bring a new perspective to the following question, which has been circulated in the combinatorics community for quite some time

Conjecture 1.5. *Let $\mathcal{F}$ be a finite family of axis parallel boxes in $\mathbf{R}^d$. Then the matching and piercing numbers of $\mathcal{F}$ are the same, up to a constant factor (which may depend on the dimension).*

According to Beck (personal communication), he raised the question some 20 years ago. He also mentioned that the same question was raised by other people as well, around the same time.

The conjecture holds for $d = 1$, but, as far as we know, is still open in all other dimensions.

If Conjecture 1.5 is true, then we should be able to improve the lower bound for $\nu_d(n)$ to $\Omega_d(\sqrt{n}(\log n)^{d/2-1})$, thanks to Theorem 1.2. On the other hand, it is not easy to do so.

One can prove the lower bound in (1) as follows. Partition $\mathbf{U}_d$ into $n^{1/2}$ subcubes of volume $n^{-1/2}$ each. By a direct calculation, one can show that for a fixed subcube $\mathbf{C}$, the probability that $\mathbf{C}$ contains a random box is $\Theta_d(n^{-1})$. After generating n random boxes, each subcube contains at least one box with constant probability. On the other hand, the boxes contained in different subcubes are pairwise disjoint. Thus, one obtains (at least in expectation), $\Omega(n^{1/2})$ disjoint boxes. (The reader is invited to work out a detailed proof.) We find it an exciting problem to improve upon this natural construction. Is there a better way to guarantee that a set of random boxes are disjoint than showing that (piecewise) they are contained in a set of disjoint deterministic boxes ?

The rest of the paper will be organized as follows. In the next section, we discretize the problem by replacing random boxes by random canonical boxes. The easier part of Theorem 1.2, the lower bound, will be proved in Section 3. The rest of the paper is devoted to the proof of the upper bound. We are going to partition the boxes into two groups, according to their volumes. In Section 4, we use a well-known construction from discrepancy theory to show that all big boxes can be pierced by $O_d(\sqrt{n}(\log n)^{d/2-1})$ points (this part is deterministic). The hardest part of the proof, which shows that we can pierce all small random boxes by $O_d(\sqrt{n}(\log n)^{d/2-1}\log\log n)$ points (with high probability), is presented in Sections 5 and 6. This proof relies on an earlier result on nearly perfect matchings in hypergraphs [6] and a careful analysis using Azuma's inequality.

Notation. All logarithms are of base 2. The asymptotic notation is used under the assumption that $n \to \infty$. Notation such as O_d or Ω_d mean that the hidden constants may depend on d. An event holds almost surely if its failure probability goes to zero with n. $\mathbf{P}$ and $\mathbf{E}$ denote probability and expectation.

Acknowledgment. I would like to thank V. H. Vu for having called the problem to my attention and for helpful conversations. The case $d = 2$ of Theorem 4.1 was pointed out by B. Sudakov in a conversation with V. H. Vu about 10 years ago. I also thank M. Neiman for reading the first draft and giving useful comments.

2. Change of model: Random canonical boxes

A canonical interval is an interval of length 2^{-i} for some non-negative integer i and has a left end point at some multiple of 2^{-i}. A d-dimensional canonical box is the product of d canonical intervals. Following [1], we show that one can restrict the problem to canonical boxes, sandwiching a random interval by two canonical (random) intervals.

Consider the following model. Let c be a positive constant. For each canonical box B, activate it with probability $\min\{1, cn\mathrm{Vol}(B)^2\}$. Let $\tau_d(c, n)$ be the piercing number of the activated boxes.

Lemma 2.1. *There is a positive constant $c_1 = c_1(d)$ such that the following holds for any positive function $f(d, n)$. If $\tau_d(c_1, n) \geq f(d, n)$ almost surely, then $\tau_d(n) \geq f(d, n)$ almost surely.*

Lemma 2.2. *There is a positive constant $c_2 = c_2(d)$ such that the following holds for any positive function $f(d, n)$. If $\tau_d(c_2, n) \leq f(d, n)$ almost surely, then $\tau_d(n) \leq f(d, n)$ almost surely.*

The proofs of these two lemmas are similar to one used in [1] (with respect to $\nu_d(n)$) and will be sketched in Section 9 for the reader's convenience.

Lemma 2.3. *The number of different shapes of canonical boxes of volume $\epsilon = 2^{-k}$ is*

$$r(\epsilon) = \binom{k+d-1}{d-1}.$$

3. Proof of the lower bound

Consider all random canonical boxes of volume $\epsilon(n) := \frac{(\log n)^{1-d/2}}{\sqrt{n}} = 2^{-k}$. We may assume k is an integer without loss of generality. Choose each box with probability

$$p := c_1 n \text{Vol}(B)^2 = c_1 \log^{2-d} n$$

for some small positive constant c_1.

Claim 3.0.1. *Almost surely, the number of chosen boxes is* $\Omega_{d,c_1}(\sqrt{n}\log^{d/2} n)$.

Proof Let X be the number of chosen boxes, then

$$X = \sum_{\text{Vol}(B)=\epsilon(n)} \mathbf{1}_{\{\text{B is chosen}\}}.$$

The event $\mathbf{1}_{\{\text{B is chosen}\}}$ are independent, so the claim follows from Chernoff bound and the fact that

$$\begin{aligned}\mathbf{E}(X) &= \binom{\log(\epsilon(n)^{-1}) + d - 1}{d-1}\epsilon(n)^{-1}c_1 \log^{2-d} n \\ &= \Omega_{d,c_1}((\log^{d-1} n) \times \sqrt{n}(\log n)^{d/2-1} \times \log^{2-d} n) = \Omega_{d,c_1}(\sqrt{n}\log^{d/2} n).\end{aligned}$$

■

Notice that if we can pierce the chosen boxes by K points, then we can also pierce them by K points whose coordinates are multiples of 2^{-k}. The set P of these points has cardinality at most $\epsilon^{-d} = n^{d/2}(\log n)^{d^2/2-d}$.

For any $v \in P$, v is contained in $r(\epsilon(n)) = O_d(\log^{d-1} n)$ different canonical boxes of volume ϵ (by Lemma 2.3). Let Y_v be the number of boxes chosen among these, then by setting the constant c_1 sufficiently small, we have

$$\mathbf{P}(Y_v \geq d\log n) \leq \binom{r(\epsilon)}{d\log n} p^{d\log n} \leq (\frac{r(\epsilon)ep}{d\log n})^{d\log n} \leq d^{-d\log n} \leq n^{-d}.$$

By the union bound

$$\mathbf{P}(\exists v : Y_v \geq d\log n) \leq \epsilon^{-d} n^{-d} \leq n^{-d/2} = o(1).$$

Thus, almost surely every point in P is contained in at most $d \log n$ chosen boxes. By Claim 3.0.1, the number of points we need to pierce all boxes is, almost surely, $\Omega_{d,c_1}(\sqrt{n}(\log n)^{d/2-1}) = \Omega_d(\sqrt{n}(\log n)^{d/2-1})$, as we can set c_1 to be a small constant depending on d. This, together with Lemma 2.1, completes the proof of the lower bound.

4. Piercing big boxes

We call a (canonical) box *big* if its volume is larger than $\frac{(\log n)^{1-d/2}}{\sqrt{n}}$ and *small* otherwise.

One can pierce big boxes by few points using the following (deterministic) result.

Theorem 4.1. *All canonical boxes of volume at least ϵ can be pierced by $O_d(\epsilon^{-1})$ points.*

One can prove this result using a high dimensional version of Van der Corput sequence of points, introduced by Hammersley [3] and Halton [2], and the Chinese Remainder Theorem. We think that Theorem 4.1 is known, but cannot find any direct reference. However, the proof of Theorem 2.4 of [4] with a few suitable modifications will provide the desired result.

5. Piercing random small boxes

Now we have come to the heart of the proof, the estimate of the number of point needed to pierce all *chosen* small canonical boxes B.

For a canonical box B with volume $\mathrm{Vol}(B)$ at most $\epsilon(n) = \frac{(\log n)^{1-d/2}}{\sqrt{n}}$, set $p_B = c_2 n \mathrm{Vol}(B)^2$, where $c_2 = c_2(d)$ is a large positive constant. Choose each B with probability p_B, independently. From now on, whenever we say random boxes, we mean random canonical boxes generated by this method.

In view of Lemma 2.2, to complete the proof of Theorem 1.2, it suffices to show

Theorem 5.1. *Almost surely, one can pierce the chosen boxes using $O_d(\sqrt{n}(\log n)^{d/2-1} \log\log n)$ points.*

Let us recall the proof of the lower bound $\Omega_d(\sqrt{n}(\log n)^{d/2-1})$. The core of this proof is the fact that any fixed point (with overwhelming probability) can only pierce $O_d(\log n)$ chosen boxes. We proved Theorem 4.1 by showing that we can (with high probability) construct a system of points so that each point pierces as much as $\Omega_d(\log n / \log\log n)$ chosen boxes, and also that there is little overlap (each box is pierced not too many times).

We are going to achieve the same goal for small boxes using a hypergraph construction. Roughly speaking, we construct a hypergraph whose vertices are the chosen boxes and whose edges correspond to points which pierce many boxes. Thus, a small system of piercing points corresponds to an economical covering of the vertices by union of edges. We obtain such an economical covering by showing that the defined hypergraph contains a nearly perfect matching that covers most of the vertices. (The rest of the vertices can be then covered arbitrarily.)

The key ingredient of the proof of Theorem 5.1 is the following result concerning the existence of near perfect matching in a hypergraph.

Large matching in hypergraphs. A *hypergraph* $H(V, E)$ is a pair of a set V of *vertices* and a family E of subsets of V. An element of E is called an edge of H. We say H is $r - uniform$ if every edge has exactly r vertices. The *degree* of a vertex v is the number of edges containing v. H is (D, ϵ)-regular if every degree of H is between $D(1 \pm \epsilon)$.

Given j vertices $v_1, \dots, v_j$, the j-codegree $codeg(v_1, \dots, v_j)$ is the number of edges containing all of them. The j-codegree of H, denoted by $C_j(H)$ is the maximum codegree over all sets of j vertices.

A matching of H is a set of disjoint edges. Let $U(H)$ be the minimum number of vertices left uncovered by a matching. We use the following variant of Theorem 1.2.2 from [6].

Theorem 5.2. *There is an absolute constant C such that the following holds. Let H be a $(r+1)$-uniform, (D, ϵ)-regular hypergraph on N vertices. Suppose that there are numbers $D_1 = D \geq D_2 \geq \dots \geq D_{r+1} = 1$, $\beta > 0$ and $x > 0$ such that*

(1) $\log x \leq (\log D)^{1-\beta}$,
(2) $\epsilon \leq x^{-1} \log^{\beta} D$,
(3) $C_j(H) \leq D_j$ *for* $j = 2, \dots, r$,
(4) $x^3 \leq \frac{D_j}{D_{j+1}}$ *for* $j = 1, \dots, r$.

Then $U(H) \leq Cx^{-1} N \log D$. As a consequence, the vertices of H can be covered by

$$\frac{N}{r+1} + U(H) \leq \frac{N}{r+1} + Cx^{-1} N \log D$$

edges.

6. Proof of Theorem 5.1

In order to prove Theorem 5.1, we will need a somewhat more refined version of the strategy discussed in the previous section. We first need to divide the canonical boxes into small classes by their volumes and shapes, then define a random hypergraph on each class and apply Theorem 5.2 to obtain a upper bound on the covering number of that hypergraph. The desired bound in Theorem 5.1 will be the sum of these numbers. The key lemma is the following.

Lemma 6.1. *Let $b \leq 1$ be a positive constant. Then with probability at least $1 - o(\frac{1}{\log\log n})$ the random canonical boxes of volume $b \times \epsilon(n) = b\frac{(\log n)^{1-d/2}}{\sqrt{n}}$ can be covered by $O_d(b\sqrt{n}(\log n)^{d/2-1}\log\log n)$ points.*

Lemma 6.1 implies Theorem 5.1 by summing over the possible volumes of the canonical boxes (in other words, summing over b). These number b decrease by a factor of 2 and are thus summable to a finite number. Furthermore, we only need to consider boxes with volume at least $n^{-1/2}\log^{-C} n$ for some sufficiently large C, since the number of chosen boxes with smaller volume will be (with high probability) smaller than $n^{1/2}\log^{d/2-1} n$, given that C is sufficiently large. (The proof is left as an exercise.) Consequently, the number of points one needs to pierce these boxes is negligible. This shows that the number of terms in the relevant sum is only $O(\log\log n)$ and the success probability in Lemma 6.1 guarantees that we can use the union bound.

6.2. **Proof of Lemma 6.1.** Consider all canonical boxes of volume $b\epsilon(n) = 2^{-k}$ (we can assume that k is an integer, without loss of generality). We choose each box with probability

$$p = c_2 n(b\epsilon(n))^2 = c_2 b^2 (\log n)^{2-d}$$

with a large constant $c_2 = c_2(d)$.

We divide the boxes into classes according to their shapes. Recall that a shape of canonical box is a d-tuple $(i_1, \dots, i_d)$, where $i_1, \dots, i_d$ are non-negative integers such that $\sum_{j=1}^{d} i_j = k$ (which means the box is the product of canonical intervals of lengths $2^{-i_1}, \dots, 2^{-i_d}$).

For each shape there are 2^k boxes of this shape; these boxes form a partition of the unit cube $\mathbf{U}_d$. First we define the *Big* classes: a Big class consists of all boxes whose shapes have the same last $d-2$ coordinates $(i_3, \dots, i_d)$. Since each i_j can take at most $k+1$ values, there are at most $(k+1)^{d-2}$ Big classes.

Now consider a fixed Big class. For all boxes in this class, we have

$$i_1 + i_2 = h,$$

for some $h = O(k)$.

Let $L = c\log\log n$ where c is a constant to be chosen later, and set

$$s := \frac{h}{L} - 1 = O(\frac{\log n}{\log\log n}). \tag{3}$$

We can assume, without loss of generality, that s is an integer. Now we split the (given) Big class into L small classes, according to $i_1 (\text{mod } L)$. Technically speaking, for $q = 0, \dots, L-1$, the q-th small class consists of all boxes with one of the following shapes:

$$(q, h-q, i_3, \dots, i_d),$$
$$(q+L, h-q-L, i_3, \dots, i_d),$$
$$\dots$$
$$(q+sL, h-q-sL, i_3, \dots, i_d).$$

In order to apply Theorem 5.2, we define an $(s+1)$-uniform hypergraph H_q whose vertices are the chosen boxes in the q-th small class. Thus, for a box B in the class,

$$\mathbf{P}(B \text{ is a vertex of } H_q) = \mathbf{P}(B \text{ is chosen }) = p.$$

The expectation of the number of vertices is thus

$$p(s+1)2^k = \Theta((s+1)b\sqrt{n}(\log n)^{1-d/2}).$$

Using Chernoff bound, one can prove that this number is strongly concentrated around the expectation.

Claim 6.2.1. *With probability at least* $1 - \frac{1}{n}$, *the number of vertices of* H_q *is*

$$N = \Theta((s+1)b\sqrt{n}(\log n)^{1-d/2}).$$

Now comes an important definition. An edge of H_q consists of $s+1$ chosen boxes of *different shapes* whose intersection is a canonical box of *positive* volume. It is easy to see that the intersection box, which we call an *atom box*, should have the shape

$$(q+sL, h-q, i_3, \dots, i_d)$$

and volume $2^{-(k+sL)}$.

We list all atom boxes. For each atom box B, there is a unique set of $(s+1)$ boxes of different shapes in the q-th class whose intersection is B. We say that B defines an edge of H_q if all these $s+1$ boxes are chosen. It is clear that

$$\mathbf{P}(\text{ An atom box defines an edge of } H_q) = p^{s+1}.$$

Consider a vertex v in H_q. We are interested in its degree D. By definition, v corresponds to a box B_v of volume 2^{-k} which contains 2^{sL} atom boxes. The probability that an atom box defines an edge is p^s (as we already fix v). Thus,

$$\mathbf{E}(D) = p^s 2^{sL}.$$

We now show that with high probability, H_q satisfies the conditions of Theorem 5.2. Let

$$x = \frac{(\log n)^2}{\log\log n}$$

thus we can choose $\beta = 0.9$ and

$$\epsilon = x^{-1} \log^{0.9}(\mathbf{E}(D)) = O(\frac{\log\log n}{\log^{1.1} n}).$$

Then

Claim 6.2.2. *If c is sufficiently large, then H_q is $(\mathbf{E}(D), \epsilon)$-regular with probability at least $1 - \frac{1}{\log^d n}$.*

Let $\mathbf{C}_j$ be the j-codegree of H_q, $j = 2\dots, s$.

Claim 6.2.3. *For $j = 2, \dots, s$, put*

$$D_j = 2^{(s+1-j)L} p^{s+1-j}(1+\epsilon)$$

then with probability at least $1 - \frac{1}{\log^d n}$,

$$\mathbf{C}_j \le D_j$$

for all j, given that c is sufficiently large.

Choose c sufficiently large, Claims 6.2.2 and 6.2.3 take care of conditions (1) and (2) of Theorem 5.2. For condition (3), we have, again by setting c large, that

$$\frac{\mathbf{E}(D)}{D_2} = \frac{p2^L}{1+\epsilon} = \Theta((\log n)^{c+2-d}) > x^3,$$
$$\frac{D_j}{D_{j+1}} = p2^L > x^3.$$

Theorem 5.2 gives the following upper bound on the number of vertices left uncovered by the largest matching of H_q

$$\begin{aligned} U(H_q) &= O(x^{-1} N \log \mathbf{E}(D)) \\ &= O((s+1)b\sqrt{n}(\log n)^{1-d/2}\frac{\log\log n}{\log n}) \\ &\le O(b\sqrt{n}(\log n)^{1-d/2}) \end{aligned}$$

Thus, with probability at least $1 - \frac{1}{\log^d n}$, we can pierce all chosen boxes in the q-th small class by

$$\frac{N}{s+1} + U(H_q) \le O(b\sqrt{n}(\log n)^{1-d/2})$$

points: one point for each edge in the matching and one point for each box not covered by the matching.

Sum up over all small classes and using the union bound (noticing that the number of small classes is $O((k+1)^{d-2} \times \log\log n) = o(\log^{d-1} n))$, we conclude that with probability at least $1 - \frac{1}{\log n} = 1 - o(\frac{1}{\log\log n})$, the number of points needed to pierce all random boxes of volume 2^{-k} is at most

$$\begin{aligned} &\le k^{d-2} L \times b\sqrt{n}(\log n)^{1-d/2} \\ &\le O_d\big((\log n)^{d-2} \log\log n)(b\sqrt{n}(\log n)^{1-d/2})\big) \\ &= O_d(b\sqrt{n}(\log n)^{d/2-1} \log\log n) \end{aligned}$$

proving Lemma 6.1.

7. Proof of Claim 6.2.2

We shall use the bounded martingale method, relying on Azuma's inequality.

Theorem 7.1 (Azuma's inequality)**.** *Let X_k, $k = 0, \dots, N$ be a martingale. Suppose there are constants $C_k > 0$ so that $|X_k - X_{k-1}| \le C_k$. Then for all $t > 0$*

$$Pr(|X_N - X_0| > t) \le 2\exp(-\frac{t^2}{2\sum_{k=1}^{N} C_k^2}).$$

Before starting the proof, we would like to point out that in our application of Azuma's inequality, the order in which we expose the atom variables will play a significant role.

We want to prove a concentration bound for the degree D of one vertex of the random hypergraph H_q. In fact, we will do slightly more. Fix a canonical box B which belongs to a q-th small class. We will show that the number of *potential* degree of this box (vertex) is close to its mean (regardless if this box will be chosen or not).

Recall that each box (in a mall class) is chosen with probability $p = c_2 b^2 \log^{2-d} n$. For a box S, let t_S be the indicator variable of the event "S is chosen", i.e. $t_S = 1$ with probability p and $t_S = 0$ otherwise. Using the definition of an edge, we can write the formula for the potential degree of B

$$D = \sum_{A \text{ atom}, A \subset B} \prod_{B' \supset A, B' \neq B} t_{B'}.$$

Suppose B is of the shape $(q + mL, h - q - mL, i_3, \dots, i_d)$, with $m \le s/2$ (which we can assume by symmetry). We consider all boxes B' in the same small class which have nonempty intersection with B.

For $j = 1, 2, \dots, m$, let $\mathcal{B}_{2j-1}$ be the set of boxes of the shape

$$(q + (m-j)L, h - q - (m-j)L, i_3, \dots, i_d),$$

and $\mathcal{B}_{2j}$ be the set of boxes of the shape

$$(q+(m+j)L, h-q-(m+j)L, i_3, \dots, i_d).$$

For $j = m+1, \dots, s-m$, let $\mathcal{B}_{m+j}$ be the set of boxes of the shape

$$(q+(m+j)L, h-q-(m+j)L, i_3, \dots, i_d).$$

It is easy to see that every box should belong to some set $\mathcal{B}$ and also boxes in the same set $\mathcal{B}$ are disjoint. Furthermore, $|\mathcal{B}_{2j-1}| = |\mathcal{B}_{2j}| = 2^{jL}$ and $|\mathcal{B}_{m+j}| = 2^{jL}$ for all relevant j.

We order the sets $\mathcal{B}_l$ increasingly in l. In each set, order the boxes arbitrarily. Overall, we obtain an ordering over all boxes that intersect B. We are going to expose the indicator random variables in this order.

Let $t_1, t_2, \dots, t_M$ be the indicator variables in question. Define a martingale as follow:

$$\mathbf{D}_0 := \mathbf{E}(D)$$

$$\mathbf{D}_i := \mathbf{E}(D|t_1, \dots, t_i), \quad i = 1, \dots, M.$$

Notice that $D = \mathbf{D}_M$.

Now let us take a closer look at the martingale differences Z_i, where

$$Z_i := \mathbf{D}_i - \mathbf{D}_{i-1} = \mathbf{E}(D|t_1, \dots, t_{i-1}, t_i) - \mathbf{E}(D|t_1, \dots, t_{i-1}).$$

If t_i is the indicator variable of (a box) B_i, then in order to bound Z_i, we only need to take into account the edges defined by the atom boxes in the intersection of B_i and B. We next consider three cases, depending on where B_i belongs.

Case 1. $B_i \in \mathcal{B}_{2j-1}$ for some $1 \le j \le m$. In this case the intersection of B and B_i has volume $2^{-(k+jL)}$, so thus contains $2^{(s-j)L}$ atom boxes. Each atom box defines an edge of H_q if all s boxes containing it are chosen, each box from a distinct set $\mathcal{B}_l$. At the time t_i is exposed, only $s-(2j-1) = s-2j+1$ among these boxes remain random. Therefore,

$$|Z_i| \le 2^{(s-j)L} p^{s-2j+1} < 2^{(s-j)L} p^{s-2j}.$$

Case 2. $B_i \in \mathcal{B}_{2j}$ for some $1 \le j \le m$. The same argument gives

$$|Z_i| \le 2^{(s-j)L} p^{s-2j}.$$

Case 3. $B_i \in \mathcal{B}_{m+j}$ for some $m+1 \le j \le s-m$. In this case the intersection $B_i \cap B$ contains $2^{(s-j)L}$ atom boxes, and at the time t_i is exposed, $s-m-j$ boxes remain open. So

$$|Z_i| \le 2^{(s-j)L} p^{s-m-j} \le 2^{(s-j)L} p^{s-2j}.$$

Thus we the parameters C_i's in Azuma's inequality to be $2^{(s-j)L}p^{s-2j}$ in all cases.

Hence

$$\sum_{i=1}^{M} C_i^2 \le \sum_{j=1}^{s-m} 2^{jL}(2^{(s-j)L}p^{s-2j})^2$$
$$= (\mathbf{E}(D))^2 \sum_{j=1}^{s-m} \big(\frac{1}{2^L p^4}\big)^j.$$

Since $2^L p^4 = c_2^4 b^8 (\log n)^{c+8-4d}$ then for $c > 4d-8$ we have

$$\sum_{i=1}^{M} C_i^2 = O((\mathbf{E}(D))^2 (\log n)^{-(c+8-4d)}).$$

Put

$$\lambda = \epsilon \mathbf{E}(D)$$

and applying Azuma's Inequality we obtain

$$\mathbf{P}(|D - \mathbf{E}(D)| > \epsilon\mathbf{E}(D)) \le \exp(-\frac{(\frac{\log\log n}{\log^{1.1} n})^2 (\mathbf{E}(D))^2}{O((\mathbf{E}(D))^2 (\log n)^{-(c+8-4d)})})$$
$$= O(e^{-(\log n)^{c-4d+5.8}(\log\log n)^2})$$
$$= o(1/n)$$

if c is sufficiently large. Since there are $(s+1)2^k = O(\sqrt{n}(\log n)^{d/2})$ boxes, by union bound we have

$$\mathbf{P}(H_q \text{ is not } (\mathbf{E}(D), \epsilon)\text{-regular}) = o(\frac{1}{n}\sqrt{n}(\log n)^{d/2}) = o(\frac{1}{(\log n)^d})$$

This completes the proof. ■

8. Proof of Claim 6.2.3

This proof has the same spirit as the one of Claim 6.2.2. We want to estimate the maximal j-codegree of H_q. More precisely, for j canonical boxes in the q-th class, we show that, with high probability, the potential j-codegree of them is less than an upper bound. Recall that an edge of H_q which contains all theses boxes is determined by an atom box lying in all of them. We only consider boxes of different shapes since the ones of same shape are disjoint. Fix j canonical boxes in H_q of shapes

$$(q + m_1 L, h - q - m_1 L, i_3, \dots, i_d),$$
$$(q + m_2 L, h - q - m_2 L, i_3, \dots, i_d),$$
$$\dots$$
$$(q + m_j L, h - q - m_j L, i_3, \dots, i_d),$$

where $0 \le m_1 < m_2 < \cdots < m_j \le s$, and by symmetry we can assume that $m_1 + m_j \le s$. Let I be the intersection. Then I has shape $(q + m_j L, h - q - m_1 L, i_3, \dots, i_d)$.

We are going to distribute all boxes B that belong to the q-th class and have non-empty intersection with I into sets $\mathcal{B}_l$ for $l = 1, 2, \dots$. Put $a = m_j - m_1 - (j-1) \ge 0$. For $l = 1, \dots, a$, we arbitrarily assign the boxes B's of shape $(q + mL, h - q - mL, i_3, \dots, i_d)$ where $m \in [m_1, m_j] \setminus \{m_1, m_2, \dots, m_j\}$ to the sets B_l so that each set contains only one box.

For $l = a + 1, \dots, a + m_1$, let $\mathcal{B}_{2l-a-1}$ be the set of boxes of the shape

$$(q + (m_1 + a - l)L, h - q - (m_1 + a - l)L, i_3, \dots, i_d),$$

and $\mathcal{B}_{2l-a}$ be the set of boxes of the shape

$$(q + (m_j - a + l)L, h - q - (m_j - a + l)L, i_3, \dots, i_d).$$

The sizes of the sets are

$$|\mathcal{B}_{2l-a-1}| = |\mathcal{B}_{2l-a}| = 2^{(l-a)L}.$$

For $l = m_1 + a + 1, \dots, s - m_j + a$, let $\mathcal{B}_{m_1+l}$ be the set of boxes of the shape

$$(q + (m_j - a + l)L, h - q - (m_j - a + l)L, i_3, \dots, i_d).$$

The sizes of the sets are

$$|\mathcal{B}_{m_1+l}| = 2^{(l-a)L}.$$

We order the sets $\mathcal{B}_l$ increasingly in l. In the set $\mathcal{B}_1$, order the boxes arbitrarily as $B_1, \dots, B_{|\mathcal{B}_1|}$. In the set $\mathcal{B}_2$, order the boxes arbitrarily as $B_{|\mathcal{B}_1|+1}, \dots, B_{|\mathcal{B}_1|+|\mathcal{B}_2|}$, and so on. Let M be the sum of sizes of all sets, and let t_i be the indicator variable of B_i, $i = 1, \dots, M$. Using the definition of an edge, we can write the formula for the potential codegree of the fixed j boxes

$$\mathbf{C}_j = \sum_{A \text{ atom}, A \subset I} \prod_{B' \supset A} t_{B'}.$$

Observe that

$$\mathbf{E}(\mathbf{C}_j) = 2^{(s-(m_j - m_1))L} p^{s+1-j} \le \frac{D_j}{1+\epsilon}.$$

We define the martingale as

$$X_0 = \mathbf{E}(\mathbf{C}_j),\ X_i = \mathbf{E}(\mathbf{C}_j | t_1, \dots, t_i),\ i = 1, \dots, M,$$

and the martingale difference

$$Z_i = X_i - X_{i-1}.$$

We bound Z_i by using similar argument as in the proof of Claim 6.2.2.

Case 1. $B_i \in \mathcal{B}_l$ for some $1 \le l \le a$. In this case the intersection $B_i \cap I$ contains $2^{(s-(m_j-m_1))L}$ atom boxes, and at the time t_i is exposed, $s+1-j-l$ boxes remain random. So

$$|Z_i| \le 2^{(s-(m_j-m_1))L} p^{s+1-j-l}.$$

Case 2. $B_i \in \mathcal{B}_{2l-a-1}$ for some $a+1 \le l \le a+m_1$. In this case the intersection of I and B_i has volume $2^{-(k+(m_j-m_1+l-a)L)}$, and thus contains $2^{(s-(m_j-m_1)-l+a)L}$ atom boxes. Each atom box defines an edge of H_q if all s boxes containing it are chosen, each box from a distinct set $\mathcal{B}_l$. At the time t_i is exposed, only $(s+1-j)-(2l-a-1) = s+1-j-2l+a+1$ among these boxes remain random. Therefore,

$$|Z_i| \le 2^{(s-(m_j-m_1)-l+a)L} p^{s+1-j-2l+a+1} < 2^{(s-(m_j-m_1)-l+a)L} p^{s+1-j-2l}.$$

Case 3. $B_i \in \mathcal{B}_{2l-a}$ for some $a+1 \le l \le a+m_1$. The same argument gives

$$|Z_i| \le 2^{(s-(m_j-m_1)-l+a)L} p^{s+1-j-2l}.$$

Case 4. $B_i \in \mathcal{B}_{m_1+l}$ for some $m_1+a+1 \le l \le s-m_j+a$. In this case the intersection $B_i \cap I$ contains $2^{(s-(m_j-m_1)-l+a)L}$ atom boxes, and at the time t_i is exposed, $(s+1-j)-(m_1+l)$ boxes remain random. So

$$|Z_i| \le 2^{(s-(m_j-m_1)-l+a)L} p^{s+1-j-m_1-l} < 2^{(s-(m_j-m_1)-l+a)L} p^{s+1-j-2l}.$$

Hence

$$\sum_{i=1}^{M} C_i^2 \le \sum_{l=1}^{a} (2^{(s-(m_j-m_1))L} p^{s+1-j-l})^2 + \sum_{l=a+1}^{s-m_j+a} 2^{(l-a)L} (2^{(s-(m_j-m_1)-l+a)L} p^{s+1-j-2l})^2$$

If $a = 0$ then the first sum is 0, otherwise

$$\begin{aligned}
\sum_{l=1}^{a} (2^{(s-(m_j-m_1))L} p^{s+1-j-l})^2 &= (2^{(s+1-j)L} p^{s+1-j})^2 2^{2(j-1-(m_j-m_1))L} \sum_{l=1}^{a} p^{-2l} \\
&= (\frac{D_j}{1+\epsilon})^2 (\frac{1}{2^L p})^{2a} (1+o(1)) \\
&= O((\frac{D_j}{1+\epsilon})^2 (\log n)^{-(c+2-d)2a}).
\end{aligned}$$

The second sum is

$$\sum_{l=a+1}^{s-m_j+a} 2^{(l-a)L}(2^{(s-(m_j-m_1)-l+a)L}p^{s+1-j-2l})^2$$

$$= (2^{(s+1-j)L}p^{s+1-j})^2 2^{-La} \sum_{l=a+1}^{s-m_j+a} (\frac{1}{2^L p^4})^l$$

$$= O((\frac{D_j}{1+\epsilon})^2 (\log n)^{-(c(2a+1)+(8-4d)(a+1))})$$

if $c > 4d-8$.. Take c large enough so that

$$\min\{c(2a+1)+(8-4d)(a+1), 2a(c+2-d)\} > 5,$$

we have

$$\sum_{i=1}^{M} C_i^2 = O((\frac{D_j}{1+\epsilon})^2 (\log n)^{-5}).$$

Apply the Azuma-Hoeffding Inequality we obtain

$$\mathbf{P}(\mathbf{C}_j > D_j) < \mathbf{P}(C_j > \mathbf{E}(C_j) + \epsilon\frac{D_j}{1+\epsilon})$$

$$\leq \exp\big(-\frac{(\frac{\log\log n}{\log^{1.1} n})^2(\frac{D_j}{1+\epsilon})^2}{2O((\frac{D_j}{1+\epsilon})^2(\log n)^{-5})}\big)$$

$$= O(\exp(-(\log n)^{2.8}(\log\log n)^2))$$

$$= o(1/n^2).$$

Now by union bound we have to sum up over all possible collections of j boxes of different shapes. Note that is we fix j different shapes then a collection of j boxes of those shapes is uniquely determined by the intersection of the boxes. Recall that with the notation used before, the intersection box has volume $2^{-(k+(m_j-m_1)L)}$, thus there are $2^{k+(m_j-m_1)L} < 2^{2k}$ such intersection boxes. Number of such collections is at most

$$\sum_{j=2}^{s}\binom{s}{j}2^{2k} < 2^s n(\log n)^{d-2} = n^{1+1/\log\log n}(\log n)^{d-2}$$

thus

$$\mathbf{P}(\mathbf{C}_j > D_j,\ \forall j=2,\dots,s) = o(\frac{1}{(\log n)^d})$$

and this concludes the proof. ■

9. Appendix

9.1. **Proof of Lemma 2.1.** We shall first poissonize the problem without affecting the result. Poissonization will help us keep the independence of random variables, which can be seen later. Let $\Pi(\lambda)$ be a random variable with Poisson distribution of mean λ. The number of random boxes will now be a random variable $\Pi(n)$ instead of a deterministic n. If the asymptotic lower and upper bounds for $\tau_d(\Pi(n))$ - the piercing number of the poissonized version - can be obtained, they're also the bounds for $\tau_d(n)$, modulo a constant factor. Indeed, suppose almost surely $\tau_d(\Pi(n))$ has lower bound L, i.e.

$$\mathbf{P}(\tau_d(\Pi(n)) < L) = o(1).$$

It's well-known that as n tends to infinity

$$P(|\Pi(n) - n| > \sqrt{n}\log n) = o(1).$$

Since $\tau_d(n)$ is non-decreasing as n increases, then

$$\begin{aligned}\mathbf{P}(\tau_d(\Pi(n)) < L) &\geq \mathbf{P}(\Pi(n) \geq n - \sqrt{n}\log n)\mathbf{P}(\tau_d(n - \sqrt{n}\log n) < L) \\ &\geq \frac{1}{2}\mathbf{P}(\tau_d(n) < L)\end{aligned}$$

which implies

$$\mathbf{P}(\tau_d(n) < L) = o(1)$$

so L is also the lower bound for $\tau_d(n)$ almost surely. A similar argument applies for the upper bound.

For any interval $I' = (a, b)$ there is a unique canonical interval $I = [k2^{-i}, (k+1)2^{-i})$ so that the two end points of I' are in the two different halves of I. Indeed, suppose $a \in [k2^{-i}, (2k+1)2^{-i-1})$ and $b \in [(2k+1)2^{-i-1}, (k+1)2^{-i})$, then (a, b) can not be contained in any canonical interval of length 2^{-i-1} since $[k2^{-i}, (2k+1)2^{-i-1})$ and $[(2k+1)2^{-i-1}, (k+1)2^{-i})$ are the only two canonical intervals of length 2^{-i-1} that have nonempty intersections with (a, b). On the other hand, since no canonical interval has non-empty intersections with both halves of a longer canonical interval, then if I is contained in any longer canonical interval, it is contained in one half of that interval, and so is (a, b). We say I is the *canonical outer interval* of I', and the *canonical outer box* is product of the canonical outer intervals. Thus we can easily see that, if I' is a random subinterval of $[0, 1]$ then

$$\mathbf{P}(I \text{ is the canonical outer interval of } I') = \frac{1}{2}2^{-2i}.$$

In high dimensional case, if B' is a d-dimensional random box, and B is a product of d canonical intervals of length $2^{-i_1}, 2^{-i_2}, \ldots, 2^{-i_d}$. Then

$$\begin{aligned}\mathbf{P}(B \text{ is the canonical outer box of } B') &= \big(\frac{1}{2}\big)^d 2^{-2(i_1+i_2+\cdots+i_d)} \\ &= \big(\frac{1}{2}\big)^d (Vol(B))^2 =: p_B.\end{aligned}$$

Consider a new model as follows: Take $\Pi(n)$ random boxes and replace each box by its canonical outer box. Then for a fixed canonical box B, the number of copies of B that appears will be independent of the number of appearances of other canonical boxes and has Poisson distribution with mean

$$\lambda_B = np_B = n\big(\frac{1}{2}\big)^d (Vol(B))^2.$$

To see this, recall that if we color $\Pi(\lambda)$ balls with some colors c, each color has probability to be used p_c, then the number of balls colored by c will be independent to those of other colors and has Poisson distribution $\Pi(\lambda p_c)$. Here we have an analogue with random boxes being the balls, canonical boxes being the colors, $\lambda = n$, $p_c = p_B$. Note that without Poissonization we cannot have the independence

between different canonical boxes.
We say that B is chosen (activated) if B is the canonical outer box of at least one random box. Then

$$\mathbf{P}(B \text{ is chosen }) = 1 - e^{-n(1/2)^d (Vol(B))^2} \approx \left(\frac{1}{2}\right)^d n(Vol(B))^2 = c_1 n(Vol(B))^2.$$

A lower bound on size of the point covering of all chosen canonical boxes is also a lower bound for size of a point covering of all random boxes. The Lemma is proved. ■

9.2. **Proof of Lemma 2.2.** Let I' be a random subinterval of $[0,1]$. I' contains one or two longest canonical intervals. *The canonical inner interval of* I' is its longest canonical interval if only one such exists, or is uniformly and randomly chosen if two exist. Let I be a canonical interval of length 2^{-i}. By similar argument as in the proof of Lemma 2.1, one can show that [1]

$$\mathbf{P}(I \text{ is the canonical inner interval of } I') \leq \frac{3}{2} 2^{-2i}.$$

and in high dimension

$$p_B = \mathbf{P}(B \text{ is the canonical inner box of } B') \leq \left(\frac{3}{2}\right)^d (Vol(B))^2.$$

We now take $\Pi(n)$ random boxes and replace them by their canonical inner boxes. Then for a fixed canonical box B, the number of copies of B that appears will be independent of the number of appearances of other canonical boxes and has Poisson distribution with mean

$$\lambda_B = np_B \leq n\left(\frac{3}{2}\right)^d (Vol(B))^2.$$

Add a number of additional copies of B with distribution $\Pi(n(3/2)^d (Vol(B))^2 - \lambda_B)$ (independently for each B) then the number of copies of B will have distribution $\Pi(n(3/2)^d (Vol(B))^2)$. Since the number of boxes increases, the upper bound for piercing number of the new system is also the upper bound for the original one. The rest of proof is similar to the proof of Lemma 2.1. ■

References

[1] E. G. Coffman, G.S. Lueker, J. Spencer and P.M. Winkler, *Packing random rectangles.* Probab. Theory Relat. Fields 120, 585-599 (2001).
[2] J. H. Halton, *On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals.* Numer. Math. 2 1960 84–90.
[3] J. M. Hammersley, *Monte-Carlo methods for solving multivariable problems*, Ann. New York Acad. Sci 86:844-874, 1960.
[4] J. Matousek, *Geometric Discrepancy*, Springer, 1999.
[5] J.H. Kim, V. H. Vu, *Concentration of multi-variate polynomials and its applications*, Combinatorica, 20 (3) (2000), 417-434.
[6] V. H. Vu, *New bounds on nearly perfect matchings in hypergraphs: higher codegrees do help*, Random Structures and Algorithms, 17 (1)(2000), 29-63.

Department of Mathematics, Rutgers, Piscataway, NJ 08854

E-mail address: linhtran@math.rutgers.edu