Group Testing with Geometric Ranges

Benjamin Aram Berendsohn
Freie Universität Berlin
beab@zedat.fu-berlin.de

László Kozma
Freie Universität Berlin
laszlo.kozma@fu-berlin.de

***Abstract*—Group testing is a well-studied approach for identifying t defective items in a set X of m items, by testing appropriately chosen subsets of X. In classical group testing *any subset* of X can be tested, and for $t \in \mathcal{O}(1)$ the optimal number of (non-adaptive) tests is known to be $\Theta(\log m)$.**

In this work we consider a novel *geometric setting* for group testing, where the items are points in Euclidean space and the tests are axis-parallel boxes (hyperrectangles), corresponding to the scenario where tests are defined by parameter-ranges (say, according to physical measurements). We present upper and lower bounds on the required number of tests in this setting, observing that in contrast to the unrestricted, combinatorial case, the bounds are *polynomial* in m. For instance, we show that with two parameters, identifying a defective *pair of items* requires $\Omega(m^{3/5})$ tests, and there exist configurations for which $\mathcal{O}(m^{2/3})$ tests are sufficient, whereas to identify a single defective item $\Theta(m^{1/2})$ tests are always necessary and sometimes sufficient. Perhaps most interestingly, our work brings to the study of group testing a set of techniques from extremal combinatorics.

I. Introduction

In the general group testing paradigm (introduced in the 1940s by Dorfman [1]), a set X of m items are given, of which t are defective. The goal is to identify the defectives with as few tests as possible. A test is performed on a selected subset $X' \subseteq X$ and its outcome is *positive* if at least one object in X' is defective, and *negative* otherwise. In *non-adaptive* group testing (our only concern in this work), the tests to be performed are fixed in advance. We assume t is known in advance (see [2, § A] for the case where t is unknown).

Group testing was initially proposed with medical applications in sight (pooling together multiple blood samples when testing for infections), but the approach has seen many other applications (see e.g. the books [3], [4]). Non-adaptive group testing has found connections with error-correcting codes [5], [6], combinatorial designs [7, § 11], compressive sensing [8]–[12], streaming algorithms [13, § 6.1], [14], and inference and learning [15]–[18].

It is well-known that to identify a *constant* number t of defective items, $\Theta(\log m)$ tests are necessary and sufficient (see [3, Thm. 7.2.12 and 7.2.15] for bounds with explicit dependence on t). For this result to hold, however, it is assumed that *arbitrary* subsets of the items can be selected as tests, which could be unrealistic. Group testing under various constraints on the admissible tests (e.g. *sparsity*) has also been studied, see e.g. [4, § 5.8] [19] [2, § 1.3].

In this work we consider items that are described by multivariate numerical data (e.g. size, location, temperature, or any domain-specific parameter), and selection happens by restricting each parameter to some admissible interval. In this way, our items can be viewed as points in d-dimensional Euclidean space and tests are subsets induced by d-dimensional axis-parallel boxes. Such a geometric view is commonly used to model *orthogonal range queries* in databases, but to our knowledge, it has not been considered in group testing.

A standard introductory example of group testing is that of testing lightbulbs connected in series, by probing various sections of the wire. This toy example also fits our geometric model, albeit in its simplest, *one-dimensional* case.

The geometric restriction we impose turns out to increase the number of necessary tests significantly, from *logarithmic* to *polynomial* in m, in stark contrast to the general case. We prove both upper and lower bounds on this quantity. For convenience, we state our results in an equivalent, dual setting, where the number of tests (n) is fixed and the number of items (m) is to be *maximized*. *Upper bounds* on m hold for *arbitrary* configurations of items and tests, and show the inherent hardness of group-testing with orthogonal range queries. *Lower bounds* on m are concrete feasible configurations of items and tests where group testing can be performed efficiently. Finding such configurations turns out to be a non-trivial geometric problem, even for the modest task of improving upon the naïve "one test per item" strategy.

To see that an asymptotic improvement *is* possible, consider the simple example of n^2 points, with a single defective ($t = 1$), in the 2-dimensional plane. Place the points on an $n \times n$ grid, and place $2n$ rectangle-tests, such as to cover each row and column of the grid by a unique rectangle (Figure 2 (a)). Observe that exactly two rectangle-tests will evaluate positive, identifying the defective item by its two coordinates. We argue later (Proposition 1) that this configuration is essentially optimal. Observe however, that the same configuration would fail, in general, to identify *two* defective items.

In general, we view a configuration of *items* (a finite set X) and *tests* ($\mathcal{S} \subseteq 2^X$) as a *set system* $(X, \mathcal{S})$. Let us denote the set of test that contain $x \in X$ as $\mathcal{S}[x] = \{S \subseteq \mathcal{S} \mid x \in S\}$, and the set of tests that contain at least one element of $Y \subseteq X$ as $\mathcal{S}[Y] = \bigcup_{y \in Y} \mathcal{S}[y]$. Let $t \in \mathbb{N}_+$ and assume $|X| > t$. Two properties of set systems are central in non-adaptive group testing: t-*separability* and t-*disjunctness*.

1) $(X, \mathcal{S})$ is t-**separable** if there are no two distinct $Y, Z \subseteq X$ such that $|Y| = |Z| = t$ and $\mathcal{S}[Y] = \mathcal{S}[Z]$;
2) $(X, \mathcal{S})$ is t-**disjunct** if there is no $Y \subseteq X$ and $x \in X \setminus Y$ such that $|Y| = t$ and $\mathcal{S}[x] \subseteq \mathcal{S}[Y]$.

Work supported by DFG grant KO 6140/1-1.

Intuitively, $\mathcal{S}[Y]$ and $\mathcal{S}[Z]$ are the collections of tests that are positive if Y, respectively Z, is the set of defective items. The set system is a valid configuration of tests if and only if it can distinguish the events "Y is the defective set" and "Z is the defective set" for all distinct size-t sets $Y, Z \subseteq X$, as captured by the first definition. Thus, t-separability is *necessary and sufficient* for non-adaptive group testing with t defectives.

Observe that t-separability only guarantees that the defective set *can be inferred* from the test results, but does not imply an efficient algorithm to do so. A naïve approach is to check, for all $\binom{|X|}{t}$ size-t subsets of items, whether they are consistent with the test outcomes. We are not aware of significantly faster algorithms for arbitrary t-separable set systems.

The stronger property of t-disjunctness yields an efficient algorithm. Intuitively, it forces every non-defective item to appear in some test that contains none of the t defectives. We can thus simply discard all items that appear in at least one negative test, so that only the t defectives remain [3, § 7.1].

We observe some simple facts [3] that hold for every set system and $t \geq 1$:

$$(t+1)\text{-separable} \implies t\text{-separable, and}$$
$$(t+1)\text{-disjunct} \implies t\text{-disjunct} \implies t\text{-separable.}$$

We denote the set of all (axis-parallel) *d-rectangles* (also called boxes or hyperrectangles) in $\mathbb{R}^d$ by $\mathcal{R}_d$, and study t-separable and t-disjunct set systems induced by $\mathcal{R}_d$.

Define $\mathrm{S}_t^d(n)$ to be the maximum m for which there exist a set $X \subseteq \mathbb{R}^d$ of size m and a set $\mathcal{S} \subseteq \mathcal{R}_d$ of size n so that $(X, \mathcal{S})$ is t-separable. In other words, $\mathrm{S}_t^d(n)$ is the maximum number of points in $\mathbb{R}^d$ among which we can identify t defectives, using n rectangle-tests. Define $\mathrm{D}_t^d(n)$ accordingly for t-disjunctness. Our goal is to understand the asymptotic behaviour of $\mathrm{S}_t^d(n)$ and $\mathrm{D}_t^d(n)$.

An easy observation is that both $\mathrm{S}_t^d(n)$ and $\mathrm{D}_t^d(n)$ are monotone in both t and d:

$$\mathrm{S}_t^d(n) \geq \mathrm{S}_{t+1}^d(n), \quad \text{and} \quad \mathrm{S}_t^d(n) \leq \mathrm{S}_t^{d+1}(n),$$
$$\mathrm{D}_t^d(n) \geq \mathrm{D}_{t+1}^d(n), \quad \text{and} \quad \mathrm{D}_t^d(n) \leq \mathrm{D}_t^{d+1}(n).$$

Recall that in the unrestricted setting, n tests can handle $2^{\Theta(n)}$ items, assuming that the number of defectives t is constant [3, § 7]. The next upper bound shows that restricting the tests to d-rectangles changes the situation considerably, even when $t = 1$.

Proposition 1. *$\mathrm{S}_1^d(n) \leq (2n-1)^d$ for all $d \in \mathbb{N}_+$.*

Proof. Let $(X, \mathcal{S})$ be a 1-separable set system induced by d-rectangles, where $|\mathcal{S}| = n$. In each dimension, the boundaries of the d-rectangles are defined by at most $2n$ distinct coordinates. These coordinates define a hypergrid with at most $(2n-1)^d$ cells. We may assume $(X, \mathcal{S})$ to be in general position, so that no point is on a cell boundary. Two arbitrary points (items) $x, y \in X$ cannot be in the same cell, as otherwise they would be contained in the same set of d-rectangles (tests), implying $\mathcal{S}[x] = \mathcal{S}[y]$. Thus, $|X| \leq (2n-1)^d$. □

For any d, given a set X of n points in $\mathbb{R}^d$, we can choose n disjoint d-rectangles so that each rectangle contains a single point of X in its interior. This structure is clearly t-disjunct for all t. By our observations and Proposition 1, we have:

$$n \leq \mathrm{D}_t^d(n) \leq \mathrm{S}_t^d(n) \in \mathcal{O}(n^d).$$

Note that in the one-dimensional case ($d = 1$), the trivial upper and lower bounds match. In the case of a single defective item ($t = 1$), we show that the upper bound is tight for all d. For all other cases, i.e. when $t, d \geq 2$, we show that both the trivial lower bound and the trivial upper bound can be improved. For lack of space, we omit or shorten some proofs, referring to our technical report for full details [2].

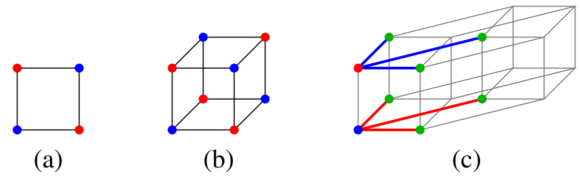

Fig. 1. *(a)* $(P_2, \mathcal{L}_2)$ is not 2-separable; red and blue points indicate sets $A, B \subseteq P_2$, where $\mathcal{L}_2[A] = \mathcal{L}_2[B]$. *(b)* $(P_3, \mathcal{L}_3)$ is not 4-separable. *(c)* $(P_d, \mathcal{L}_d)$ is not $(2d-1)$-separable (example with $d = 4$); red, blue, and green points indicate sets $A \setminus B$, $B \setminus A$, resp. $A \cap B$, where $\mathcal{L}_4[A] = \mathcal{L}_4[B]$.

II. Lower bounds

We start with a d-dimensional grid-line construction, and fully characterise its t-separability. We then transform it into a combinatorially equivalent set system that is induced by rectangles, yielding the following results:

Theorem 2. *$\mathrm{S}_3^2(n) \in \Omega(n^{3/2})$ and $\mathrm{S}_t^2(n) \in \Omega(n^{1+2/t})$ for all even $t \geq 6$.*

By monotonicity, it follows that $\mathrm{S}_2^2(n) \in \Omega(n^{3/2})$ and $\mathrm{S}_4^2(n) \in \Omega(n^{4/3})$ and bounds for odd t likewise follow.

Let $\ell \geq 2$, let $P_d = [\ell]^d$, and define $\mathcal{L}_d$ to be the set of axis-parallel lines that intersect P_d. In other words, $\mathcal{L}_d$ is the set of grid lines in the d-dimensional $\ell \times \ell \times \cdots \times \ell$ hypergrid. Observe that $|P_d| = \ell^d$ and $|\mathcal{L}_d| = d\ell^{d-1}$. It is not hard to show that $(P_d, \mathcal{L}_d)$ is $(d-1)$-disjunct, but not d-disjunct, for all $d \geq 1$. Determining the t-separability of $(P_d, \mathcal{L}_d)$ is somewhat more difficult.

We start with the two-dimensional case and argue that $(P_2, \mathcal{L}_2)$ is 1-separable. Suppose this is not the case. Then, there exist distinct points $x, y \in P$ such that $\mathcal{L}_2[x] = \mathcal{L}_2[y]$. As x is contained in at least one line of $\mathcal{L}_2$ that does not contain y, this is impossible.

We argue next that $(P_2, \mathcal{L}_2)$ is not 2-separable. Indeed, as shown in Figure 1 (a), we can find two distinct sets of points in P_2, both of size two (containing opposite corners of a grid cell), such that both sets intersect the same four grid lines.

Proposition 3. *$(P_3, \mathcal{L}_3)$ is not 4-separable.*

Proof. Consider a set of 8 points in P_3 forming the corners of a 3-rectangle (i.e. cube). Split these points into two sets A and B, both of size 4, as indicated by the coloring in Figure 1 (b). (The sets correspond to the odd, resp. even layers of the cube.) Then $\mathcal{L}_3[A] = \mathcal{L}_3[B]$, as both A and B hit exactly the lines supporting the edges of the cube. □

The constructions in Figure 1 (a)(b) can easily be extended to higher dimensions, showing that $(P_d, \mathcal{L}_d)$ is not 2^{d-1}-separable. The correct bounds are, however, much lower.

Proposition 4. *For $d \geq 4$, $(P_d, \mathcal{L}_d)$ is not $(2d-1)$-separable.*

Proof. We use the following construction, shown in Figure 1 (c). Let $x, y \in P_d$ be distinct points contained in the same grid line L. Both x and y are contained in $d-1$ lines apart from L, and these $2d-2$ lines are pairwise distinct. Consider an additional point on each of these lines, and call the resulting set of $2d-2$ points Z. Now we have $\mathcal{L}_d[Z \cup \{x\}] = \mathcal{L}_d[Z \cup \{y\}]$ as the two sets of points hit the same set of lines. ☐

With a rather involved case-distinction we can also show:

Proposition 5. *$(P_3, \mathcal{L}_3)$ is 3-separable, and $(P_d, \mathcal{L}_d)$ is $(2d-2)$-separable for all $d \geq 4$.*

Note that, in terms of point inclusion, an axis-parallel line is equivalent to a (sufficiently long and thin) hyperrectangle. Recall that $|P_d| = \ell^d = m$ and $|\mathcal{L}_d| = d\ell^{d-1} = n$, so we already have $\mathrm{S}_3^3(n) \in \Omega(n^{3/2})$ and $\mathrm{S}_{2d-2}^d(n) \in \Omega(n^{1+1/(d-1)})$ for $d \geq 4$ from Proposition 5. We can strengthen these bounds by showing that $(P_d, \mathcal{L}_d)$ is isomorphic to a *two-dimensional* rectangle-induced set system [2], yielding Theorem 2.

We next consider an arrangement of *hyperplanes* on grid points, constructing equivalent *hyperrectangle*-induced set systems, obtaining the t-disjunctness result:

Theorem 6. $\mathrm{D}_t^d(n) \in \Omega\left(n^{1+\left\lfloor \frac{d-1}{t} \right\rfloor}\right)$ *for all $d, t \in \mathbb{N}_+$.*

In the special case $t = 1$, i.e. with a single defective item, the result matches the upper bound of $\mathcal{O}(n^d)$. It follows that $\mathrm{D}_1^d(n), \mathrm{S}_1^d(n) \in \Theta(n^d)$ for all $d \in \mathbb{N}_+$, and our constructions are asymptotically optimal for this case. Theorem 6 can be slightly improved for certain values of t and d:

Theorem 7. *(a)* $\mathrm{D}_t^d(n) \in \Omega\left(n^{1+\frac{d+1-t}{t}}\right)$ *for all $d \geq t \geq 2$.*
(b) $\mathrm{D}_t^d(n) \in \Omega\left(n^{1+\frac{1}{2+t-d}}\right)$ *for all $t \geq d \geq 2$.*

In particular, in two dimensions $\mathrm{D}_t^2(n) \in \Omega(n^{1+1/t})$.

Let us sketch the first construction. For $k \in \mathbb{N}_+$, consider the vectors $c_i = (1, (i-1), (i-1)^2, \ldots, (i-1)^{k-1})$ in $\mathbb{R}^k$, for $i \in [\ell]$, where $\ell = (k-1)t + 1$. Observe that every set of k distinct vectors c_i is linearly independent. Define the hyperplanes $H_{i,j} = \{x \in \mathbb{R}^k \mid c_i \cdot x = j\}$ for all $i \in [\ell]$ and $j \in \mathbb{N}_+$. Observe that $H_{i,j}$, $H_{i',j'}$ are distinct, unless $i = i'$ and $j = j'$, and that $H_{i,j}$, $H_{i,j'}$ are parallel for all i, j, j'.

Lemma 8. *Let $m \in \mathbb{N}_+$ with $m > k$, and let $\mathcal{S} = \{H_{i,j} \mid i \in [\ell], j \in [\ell^k m]\}$. The set system $([m]^k, \mathcal{S})$ is t-disjunct.*

Lemma 8 implies that for each fixed $t \geq 1$, there is a set $P \subseteq \mathbb{R}^d$ of points and a set $\mathcal{S}$ of hyperplanes in $\mathbb{R}^d$ so that $(P, \mathcal{S})$ is t-disjunct and $|P| \in \Omega(|\mathcal{S}|^d)$. (A somewhat similar construction was also used in [6, Lemma 3.1].)

We now derive the lower bound on $\mathrm{D}_t^d(n)$. Observe that in Lemma 8 we construct ℓ $(\ell^k m)$-partitions of the grid $[m]^k$, each consisting of parallel hyperplanes. The following lemma shows that with d-dimensional hyperrectangles, we can construct a set of d arbitrary partitions. The idea is simply to use values of the i-th coordinate to encode the i-th partition. Each part of a partition corresponds to a d-rectangle.

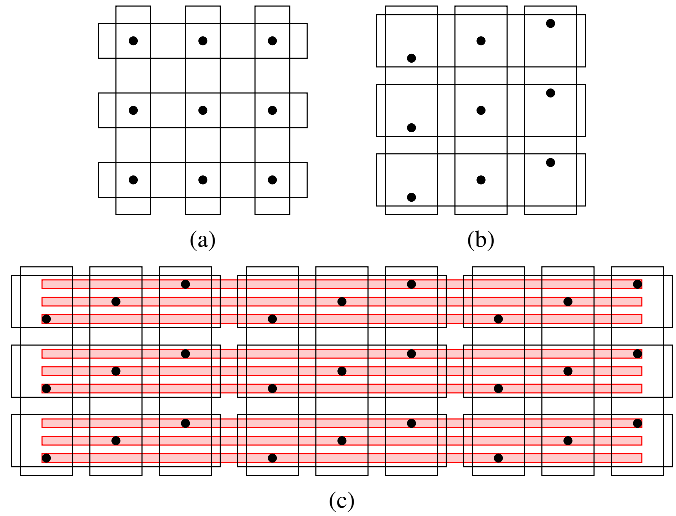

Fig. 2. The first step of the long rectangle construction in two dimensions. *(a)* Base case for $k = 3$, with $2k$ rectangles and k^2 points (simplified from what Lemma 8 yields). *(b)* Perturbation so that all points have distinct y-coordinates. *(c)* Three copies of the perturbed configuration, arranged along x-coordinate with added long rectangles in red.

Lemma 9. *Let $\Pi_1, \Pi_2, \ldots, \Pi_d$ a set of q-partitions of some set X, and let $\mathcal{S} = \Pi_1 \cup \Pi_2 \cup \cdots \cup \Pi_d \subseteq 2^X$. Then there are mappings $r \colon \mathcal{S} \to \mathcal{R}_d$ and $p \colon X \to \mathbb{R}^d$ such that for all $x \in X$ and $S \in \mathcal{S}$, we have $x \in S \Leftrightarrow p(x) \in r(S)$.*

From Lemma 8 (setting $\ell = d$) and Lemma 9 it follows that for each k, t, m and $d = (k-1)t + 1$ there is a set $X \in \mathbb{R}^d$ of m^k points and a set $\mathcal{S}$ of $d^{k+1}m$ hyperrectangles, such that $(X, \mathcal{S})$ is t-disjunct. This implies $\mathrm{D}_t^d(n) \geq (n/d^{k+1})^k$. After solving for k, we obtain Theorem 6.

Theorem 7 (a) uses a generalization of the $(P_d, \mathcal{L}_d)$ example, replacing grid lines with axis-parallel affine subspaces.

Theorem 7 (b) is proven using an easier iterative construction that we sketch next. The main idea is to transform a t-disjunct arrangement into a $(t+1)$-disjunct arrangement, so that the *ratio* of rectangles to points increases only moderately. Repeating this step yields superlinear lower bounds for $\mathrm{D}_t^d(n)$ for all $t \geq d$.

An example of a single step in two dimensions is shown in Figure 2 (c). Start with an arbitrary t-disjunct arrangement of m points and n rectangles, and perturb the points without changing the combinatorial structure. Then make k copies of the arrangement (k is to be optimized later) and place them side by side, horizontally. Finally, for each original point p, add one long and thin rectangle that covers all copies of p. This yields a $(t+1)$-disjunct arrangement.

Lemma 10. *Let $X \in \mathbb{R}^d$ be a set of m points and let $\mathcal{S}$ be a set of n d-rectangles such that $(X, \mathcal{S})$ is t-disjunct, and let $k \in \mathbb{N}_+$. Then there is a set $\mathcal{S}'$ of d-rectangles and a set $X' \in \mathbb{R}^d$ of points such that $(X', \mathcal{S}')$ is $(t+1)$-disjunct, $|\mathcal{S}'| = k \cdot n + m$ and $|X'| = k \cdot m$.*

Proof. We may assume that X is in general position (no two points agree in a coordinate) and that the coordinates of points in X and corners of hyperrectangles in $\mathcal{S}$ are integers in $[m]$.

For each $i \in [k]$, let X_i be a copy of X and let $\mathcal{S}_i$ be a copy of $\mathcal{S}$, both shifted by $(i-1)m$ in the first coordinate. For each point $p \in X$, consider the hyperrectangle $R_p = \{q \mid 1 \leq q_1 \leq km, \forall i \in [d] \setminus \{1\} : q_i = p_i\}$ that contains all copies of p. Now let $X' = \bigcup_{i=1}^{k} X_i$ and $\mathcal{S}' = \{R_p \mid p \in X\} \cup \bigcup_{i=1}^{k} \mathcal{S}_i$. Clearly, $|X'| = km$, and $|\mathcal{S}'| = kn + m$. It remains to show that $(X', \mathcal{S}')$ is $(t+1)$-disjunct.

Let $Y \subset X'$ with $|Y| = t+1$, and $p \in X' \setminus Y$. We claim that there is some rectangle $R \in \mathcal{S}'$ such that $p \in R$ and $Y \cap R = \emptyset$. This implies that $\mathcal{S}[p] \not\subseteq \mathcal{S}[Y]$.

Recall that there is some $q \in X$ such that p is a copy of q, i.e. $p \in R_q \cap X'$. First assume that $R_q \cap Y = \emptyset$. Then the claim is true simply for $R = R_q$. Otherwise, let $s \in R_q \cap Y$. Let $i \in [k]$ such that $p \in X_i$. By construction, $s \in R_q \setminus \{p\}$ implies that $s \notin X_i$. Let $Y' = Y \cap X_i \subseteq Y \setminus \{s\}$, and observe that $|Y'| \leq t$. By assumption, $(X_i, \mathcal{S}_i)$ is t-disjunct, which means that there is some rectangle $R \in \mathcal{S}_i$ for which $p \in R$ and $Y' \cap R = \emptyset$. Moreover, $R \cap X \subseteq X_i$ by construction, so $Y \cap R = \emptyset$. This proves the claim. □

Corollary 11. *Let $c \geq 1$ and let $\mathrm{D}_t^d(n) \in \Omega(n^c)$. Then $\mathrm{D}_{t+1}^d(n) \in \Omega(n^{2-1/c})$.*

Theorem 7 (b) follows by using Theorem 6 with $t = d-1$ as induction base and Corollary 11 as induction step.

III. Upper bounds

Next, we develop a technique based on avoided patterns (variants and extensions of the Kővári-Sós-Turán theorem [20] in extremal combinatorics), and improve the trivial $\mathcal{O}(n^d)$ upper bound for t-separability, for all $t, d \geq 2$.

Theorem 12. $\mathrm{S}_2^d(n) \in \mathcal{O}\left(n^{d-1/3}\right)$ *for all $d \geq 2$, and* $\mathrm{S}_t^d(n) \in \mathcal{O}(n^{d-1/2})$ *for all $d \geq 2$, $t \geq 3$.*

Thus, in two-dimensions $\mathrm{S}_2^2(n) \in \mathcal{O}(n^{5/3})$, and together with Theorem 7 (b), we have the tight result $\mathrm{S}_3^2(n) \in \Theta(n^{3/2})$.

We remark that it can be assumed without loss of generality that in a hyperrectangle-induced set system $(P, \mathcal{S})$ all points and rectangle corners have integral coordinates in $[4n]$, where $n = |\mathcal{S}|$. We start with a lemma that allows generalizing bounds to higher dimensions.

Lemma 13. *For all constants $d, t, k \in \mathbb{N}_+$, we have*

$$\mathrm{S}_t^{d+k}(n) \leq (4n)^k \cdot \mathrm{S}_t^d(n), \text{ and}$$
$$\mathrm{D}_t^{d+k}(n) \leq (4n)^k \cdot \mathrm{D}_t^d(n).$$

Proof. We only prove the lemma for $k = 1$, as the case $k > 1$ follows by induction.

Let $(P, \mathcal{S})$ be a $(d+1)$-rectangle-induced set system, and let $n = |\mathcal{S}|$ and assume $P \subseteq [4n]^{d+1}$. Let H be an axis-parallel hyperplane in $[4n]^{d+1}$, obtained by fixing the first coordinate to a value in $[4n]$. Let $P' = P \cap H$ and $\mathcal{S}' = \{R \cap H \mid R \in \mathcal{S}\}$.

Our key observation is that if $(P, \mathcal{S})$ is t-separable (t-disjunct), then $(P', \mathcal{S}')$ is t-separable (t-disjunct). This implies that $|P'| \leq \mathrm{S}_t^d(|\mathcal{S}'|) \leq \mathrm{S}_t^d(n)$ (respectively $|P'| \leq \mathrm{D}_t^d(n)$). Finally, we can cover $[4n]^{d+1}$ with $4n$ such hyperplanes H, which implies the desired bounds. □

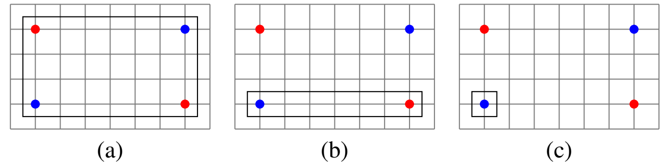

Fig. 3. The three ways a rectangle R can intersect an induced rectangle I, up to rotation. One intersection of type (c) is necessary, as otherwise each rectangle that contains a blue point also contains a red point, and vice versa.

Lemma 13 implies that a bound of $o(n^2)$ in two dimensions translates to bounds $o(n^d)$ in all dimensions $d \geq 2$.

Before proceeding with the technical proof, we give a high-level description of the upper bound in two dimensions. Let $(P, \mathcal{S})$ be a 2-separable rectangle-induced set system, and let $n = |\mathcal{S}|$. Assume that the coordinates of P are positive integers in $[4n]$. We want to show that $|P| \in \mathcal{O}(n^{5/3})$ (which implies that $\mathcal{S}_t^d(n) \in \mathcal{O}(n^{d-1/3})$ for all $t, d \geq 2$).

An *induced rectangle* in P is a set of four points in P that form the corners of an axis-parallel rectangle (which is not necessarily in $\mathcal{S}$). It is known that if P contains no induced rectangle, then $|P| \in \mathcal{O}(n^{3/2})$. (This is a special case of the well-known *Zarankiewicz problem* [20].) We can thus assume that an induced rectangle $I \subseteq P$ exists. Let $P_1, P_2 \subset P$ be the two sets of opposing corners of I. As $(P, \mathcal{S})$ is 2-separable, there must be some rectangle $R \in \mathcal{S}$ with $R \cap P_1 \neq \emptyset$ and $R \cap P_2 = \emptyset$, or vice versa. This means that the rectangle formed by I contains a corner of R (see Figure 3).

Let V be the set of corners of the rectangles in $\mathcal{S}$. Our observations imply that each induced rectangle in P must contain a point of V; we say that V *stabs* P. It is now sufficient to prove that if V stabs P, then $|P| \in \mathcal{O}(|V|^{5/3})$ (observe that $|V| \leq 4n$).

Towards this claim, we cover the square $[4n]^2$ (which contains both P and V) with rectangles $\mathcal{Q}$ such that no rectangle $Q \in \mathcal{Q}$ contains a point from V in its interior. This means that the points of P that fall within Q cannot induce a rectangle. Again invoking the Zarankiewicz problem, we can bound the number of points in $P \cap Q$. Choosing a good rectangle covering (in a non-trivial way) finally yields the desired bound.

A *rectangle covering* of the grid $G = [n]^2$ is a set $\mathcal{Q}$ of rectangles (with integral corners) such that $G \subseteq \bigcup \mathcal{Q}$. We say that $\mathcal{Q}$ is *valid* with respect to some set $V \subseteq G$ of points if $V \cap \mathrm{interior}(Q) = \emptyset$ for all $Q \in \mathcal{Q}$. Given a set of $\mathcal{O}(n)$ points V and a *weight function* $w \colon \mathcal{R}_2 \to \mathbb{R}$, we ask for the minimum total weight $W(\mathcal{Q}) = \sum_{Q \in \mathcal{Q}} w(Q)$ that a valid rectangle covering $\mathcal{Q}$ can have.

In the following, we define a suitable weight function for the case $t = 2$ (for other values of t, different weight functions are needed). For a rectangle Q, the weight $z(Q)$ is the maximum number of integral points that can fit in Q without inducing an axis-parallel rectangle. The total weight of a rectangle covering $\mathcal{Q}$ is denoted $Z(\mathcal{Q})$.

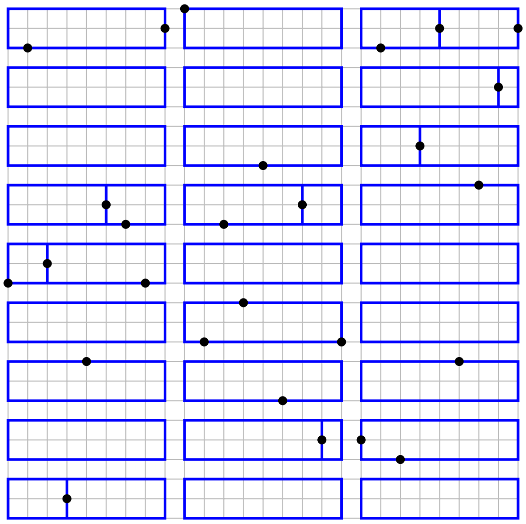

Fig. 4. A two-dimensional rectangle covering as in Lemma 15 with $k = 3$, where rectangles are always divided with vertical lines. Blue lines indicate rectangle borders.

We say that a rectangle (with integral corners) is a *size* $p \times q$ rectangle if it covers a $p \times q$ grid, i.e. has side lengths $p-1$ and $q-1$. The Kővári-Sós-Turán theorem implies the following:

Lemma 14 ([20], [21])**.** *A size $p \times q$ rectangle Q where $p \geq q$ has weight $z(Q) \leq p^{1/2}q + p$.*

Note that among rectangles with the same area, a $k^2 \times k$ rectangle is the "lightest". We use this observation in the following upper bound.

Lemma 15. *Let $n = k^3$ for some $k \in \mathbb{N}_+$, and let $V \subseteq [n]^2$ such that $|V| \leq n$. Then there is a rectangle covering $\mathcal{Q}$ with $Z(\mathcal{Q}) \leq 4n^{5/3}$ that is valid with respect to V.*

Proof. We construct $\mathcal{Q}$ as follows. Start with the regular decomposition into n rectangles of size $k^2 \times k$, i.e. $\mathcal{Q}_0 =$

$$\{[(i-1)k^2+1, ik^2] \times [(j-1)k+1, jk] : i \in [k], j \in [k^2]\}.$$

Observe that while these rectangles do not cover the *continuous* square $[1,n]^2$, they do cover all integral points in $[n]^2$. Moreover, each rectangle in $\mathcal{Q}_0$ has weight at most $2k^2$, so $Z(\mathcal{Q}_0) = n \cdot 2k^2 = 2n^{5/3}$.

The obtained $\mathcal{Q}_0$ is not yet a valid rectangle covering, as its rectangles may contain points of V in their interiors. To fix this, for each point $v \in V$ and rectangle $Q \in \mathcal{Q}_0$ where $v \in \text{interior}(Q)$, split Q with an axis-parallel line through v (see Figure 4, where we always split the rectangles with vertical lines). Observe that these splits increase the number of rectangles by at most $|V|$. Let $\mathcal{Q}$ be the set of rectangles thus obtained. As the weight of a rectangle obtained by splitting can only be smaller than the weight of the original rectangle, we have $z(Q) \leq 2k^2$ for all $Q \in \mathcal{Q}$, and thus,

$$Z(\mathcal{Q}) \leq Z(\mathcal{Q}_0) + |V| \cdot 2k^2 \leq 4n^{5/3}. \qquad \square$$

Let now $V, P \subseteq [n]^2$ be point sets. We say that P *induces* a rectangle R if all four corners of R are in P, and we say that V *stabs* P if every rectangle induced by P contains a point from V on its interior.

Lemma 16. *Let $n = k^2$ for some $k \in \mathbb{N}$, and let $V, P \subseteq [n]^2$ such that $|V| \leq n$ and V stabs P. Then $|P| \leq 4n^{5/3}$.*

Proof. Lemma 15 implies that there is a rectangle covering $\mathcal{Q}$ with $Z(\mathcal{Q}) \leq 4n^{5/3}$ that is valid with respect to V. Now assume that $|P| > Z(\mathcal{Q})$. By the pigeonhole principle, there is a rectangle $Q \in \mathcal{Q}$ that contains more than $z(Q)$ points. By the definition of z, inside Q there must be 4 points of P that induce a rectangle. The interior of this induced rectangle contains no point from V, contradicting the assumption that V stabs P. Thus, $|P| \leq Z(\mathcal{Q})$. $\square$

Note that while Lemma 15 is tight, we do not know whether Lemma 16 is also tight. We are now ready to prove Theorem 12; we only show the case $d = 2$, the other cases follow by Lemma 13.

Let $(P, \mathcal{S})$ be a 2-separable rectangle-induced set system. Let $n = |\mathcal{S}|$ and let V be the set of $n' = 4n$ corners of rectangles in $\mathcal{S}$. We assume that n' is of the form k^3, that $P, V \subseteq [n']$, and that points $p \in P$ and $v \in V$ differ in all coordinates.

Let $P' \subseteq P$ be a set of 4 points that induce a rectangle R. Let P_1 and P_2 be the pairs of opposite corners of R. We claim that there must be a rectangle $S \in \mathcal{S}$ that contains only one point of P. Suppose not, then each rectangle $S \in \mathcal{S}$ either contains no point of P or at least one "edge" of R, and thus both a point from P_1 and a point from P_2 (see Figure 3). This implies $\mathcal{S}[P_1] = \mathcal{S}[P_2]$, contradicting that $(P, \mathcal{S})$ is 2-separable. As such, R contains a point of V, which is in the interior of R by assumption. As a consequence, V must stab P, and Lemma 16 implies $|P| \leq 4(n')^{5/3} \in \mathcal{O}(n^{5/3})$.

This concludes the proof.

For disjunctness we can use a similar technique, but with a different weight function and different avoided patterns (*stars* instead of *induced rectangles*) [2]. We obtain:

Theorem 17. $\mathrm{D}_t^d(n) \in \mathcal{O}\left(n^{d-1+\frac{1}{\min(d,t)}}\right)$ *for all $d, t \geq 2$; in particular,* $\mathrm{D}_2^2(n) \in \Theta(n^{3/2})$.

IV. Conclusion

We have shown several bounds on the possible sizes of configurations for group testing with axis-parallel boxes. Our results indicate that non-trivial improvements over the naïve configurations are possible, yet these improvements are limited by elusive geometric and combinatorial constraints.

In contrast to classical (unrestricted) group testing, a precise description of optimal configurations seems out of reach. Gaps remain between upper and lower bounds, even in the planar, 2-separable case, where we have $c \cdot n^{3/2} \leq \mathrm{S}_2^2(n) \leq c' \cdot n^{5/3}$, for some $c, c' > 0$. (The known bounds are summarized in tabular form in [2, § B]).

It remains open whether $\mathrm{S}_{t+1}^d(n)$ and $\mathrm{S}_t^d(n)$ are asymptotically separated for *all* t (even in the planar case). Whether $\mathrm{D}_t^d(n)$ and $\mathrm{S}_t^d(n)$ are asymptotically separated for *any* $t, d \geq 2$ is also left as an intriguing open question.

References

[1] R. Dorfman, "The detection of defective members of large populations," *Ann. Math. Statist.*, vol. 14, no. 4, pp. 436–440, 12 1943. [Online]. Available: https://doi.org/10.1214/aoms/1177731363

[2] B. A. Berendsohn and L. Kozma, "Geometric group testing," *CoRR*, 2020. [Online]. Available: https://arxiv.org/abs/2004.14632

[3] D. Du and F. K. Hwang, *Combinatorial group testing and its applications*. Singapore: World Scientific, 1993.

[4] M. Aldridge, O. Johnson, and J. Scarlett, "Group testing: An information theory perspective," *Foundations and Trends in Communications and Information Theory*, vol. 15, no. 3-4, pp. 196–392, 2019. [Online]. Available: http://dx.doi.org/10.1561/0100000099

[5] P. Indyk, H. Q. Ngo, and A. Rudra, "Efficiently decodable non-adaptive group testing," in *Proc. of Symposium on Discrete Algorithms, SODA*, 2010, pp. 1126–1142. [Online]. Available: https://doi.org/10.1137/1.9781611973075.91

[6] E. Porat and A. Rothschild, "Explicit non-adaptive combinatorial group testing schemes," in *Automata, Languages and Programming, 35th International Colloquium, ICALP 2008, Proc. Part I: Tack A: Algorithms, Automata, Complexity, and Games*, 2008, pp. 748–759. [Online]. Available: https://doi.org/10.1007/978-3-540-70575-8_61

[7] D. Stinson, *Combinatorial Designs: Constructions and Analysis*. Springer, 2007. [Online]. Available: https://books.google.de/books?id=MJUMBwAAQBAJ

[8] A. C. Gilbert, M. A. Iwen, and M. J. Strauss, "Group testing and sparse signal recovery," in *2008 42nd Asilomar Conference on Signals, Systems and Computers*. IEEE, 2008, pp. 1059–1063.

[9] A. C. Gilbert and P. Indyk, "Sparse recovery using sparse matrices," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 937–947, 2010. [Online]. Available: https://doi.org/10.1109/JPROC.2010.2045092

[10] A. C. Gilbert, B. Hemenway, A. Rudra, M. J. Strauss, and M. Wootters, "Recovering simple signals," in *2012 Information Theory and Applications Workshop, ITA 2012*, 2012, pp. 382–391. [Online]. Available: https://doi.org/10.1109/ITA.2012.6181772

[11] H. Q. Ngo, E. Porat, and A. Rudra, "Efficiently decodable compressed sensing by list-recoverable codes and recursion," in *29th International Symposium on Theoretical Aspects of Computer Science, STACS 2012*, 2012, pp. 230–241. [Online]. Available: https://doi.org/10.4230/LIPIcs.STACS.2012.230

[12] A. Emad and O. Milenkovic, "Poisson group testing: A probabilistic model for nonadaptive streaming boolean compressed sensing," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 3335–3339.

[13] S. Muthukrishnan *et al.*, "Data streams: Algorithms and applications," *Foundations and Trends in Theoretical Computer Science*, vol. 1, no. 2, pp. 117–236, 2005.

[14] G. Cormode and S. Muthukrishnan, "What's hot and what's not: tracking most frequent items dynamically," *ACM Trans. Database Syst.*, vol. 30, no. 1, pp. 249–278, 2005. [Online]. Available: https://doi.org/10.1145/1061318.1061325

[15] S. Ubaru and A. Mazumdar, "Multilabel classification with group testing and codes," in *Proceedings of the 34th International Conference on Machine Learning, ICML 2017*, 2017, pp. 3492–3501. [Online]. Available: http://proceedings.mlr.press/v70/ubaru17a.html

[16] Y. Zhou, U. Porwal, C. Zhang, H. Q. Ngo, L. Nguyen, C. Ré, and V. Govindaraju, "Parallel feature selection inspired by group testing," in *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014*, 2014, pp. 3554–3562. [Online]. Available: http://papers.nips.cc/paper/5296-parallel-feature-selection-inspired-by-group-testing

[17] D. M. Malioutov and K. R. Varshney, "Exact rule learning via boolean compressed sensing," in *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ser. ICML13, 2013, p. III765III773.

[18] M. Shi, T. Furon, and H. Jégou, "A group testing framework for similarity search in high-dimensional spaces," in *Proceedings of the 22nd ACM International Conference on Multimedia*, ser. MM 14, 2014, p. 407416. [Online]. Available: https://doi.org/10.1145/2647868.2654895

[19] H. A. Inan, P. Kairouz, and A. Özgür, "Sparse combinatorial group testing," *IEEE Trans. Inf. Theory*, vol. 66, no. 5, pp. 2729–2742, 2020. [Online]. Available: https://doi.org/10.1109/TIT.2019.2951703

[20] T. Kővári, V. Sós, and P. Turán, "On a problem of K. Zarankiewicz," *Colloquium Mathematicae*, vol. 3, no. 1, pp. 50–57, 1954. [Online]. Available: http://eudml.org/doc/210011

[21] C. Hyltén-Cavallius, "On a combinatorical problem," in *Colloquium Mathematicae*, vol. 6, no. 1, 1958, pp. 61–65.