SEPARATION WITH RESTRICTED FAMILIES OF SETS

ZSOLT LÁNGI, MÁRTON NASZÓDI, JÁNOS PACH, GÁBOR TARDOS, AND GÉZA TÓTH

Abstract. Given a finite n-element set X, a family of subsets $\mathcal{F} \subset 2^X$ is said to *separate* X if any two elements of X are separated by at least one member of $\mathcal{F}$. It is shown that if $|\mathcal{F}| > 2^{n-1}$, then one can select $\lceil \log n \rceil + 1$ members of $\mathcal{F}$ that separate X. If $|\mathcal{F}| \geq \alpha 2^n$ for some $0 < \alpha < 1/2$, then $\log n + O(\log \frac{1}{\alpha} \log\log \frac{1}{\alpha})$ members of $\mathcal{F}$ are always sufficient to separate all pairs of elements of X that are separated by some member of $\mathcal{F}$. This result is generalized to simultaneous separation in several sets. Analogous questions on separation by families of bounded Vapnik-Chervonenkis dimension and separation of point sets in $\mathbb{R}^d$ by convex sets are also considered.

1. Introduction

For a set X, we say that a subset of X *separates* two elements if it contains one of them and does not contain the other. For a family $\mathcal{F}$ of subsets of X, we say that it *separates* a pair of elements of X if at least one member of $\mathcal{F}$ separates them. Furthermore, $\mathcal{F}$ *separates* X if every pair of distinct elements of X is separated by $\mathcal{F}$.

Suppose your computer is infected by a virus $x \in X$, where X is the set of known computer viruses. You want to perform a number of tests to find out which virus it is. Each test detects a certain set of viruses, which can be associated with the test. Let $\mathcal{F}$ denote the family of subsets of X associated with the tests you can perform. These tests are sufficient to identify the virus if, and only if $\mathcal{F}$ separates X. The number of tests necessary is at least $\log |X|$, where log stands for the base 2 logarithm. On the other hand, there is a family $\mathcal{F} \subset 2^X$ with $|\mathcal{F}| \leq \lceil \log |X| \rceil$ that separates X. This is the starting point of a rich discipline called *combinatorial search theory*; see [AhlW87].

Any fixed pair of distinct elements in X is separated by $2^{|X|-1}$ subsets of X, thus a family $\mathcal{F}$ with $|\mathcal{F}| > 2^{|X|-1}$ separates X. Our first theorem states that in this case, even a small subfamily of $\mathcal{F}$ does the job.

Theorem 1. *Let X be a finite set, $\mathcal{F} \subseteq 2^X$ with $|\mathcal{F}| > 2^{|X|-1}$. Then X can be separated by a subfamily $\mathcal{G} \subseteq \mathcal{F}$ of cardinality at most $\lceil \log |X| \rceil + 1$.*

2010 *Mathematics Subject Classification.* 90B40, 52A37.

Key words and phrases. Search theory, separation, VC-dimension, Erdős–Szekeres theorem.

Zsolt Lángi was supported by the Janos Bolyai Research Scholarship of the Hungarian Academy of Sciences.

Márton Naszódi was supported by the Janos Bolyai Research Scholarship of the Hungarian Academy of Sciences and by the Hungarian National Science Foundation (OTKA) Grant PD-104744.

János Pach was supported by Swiss National Science Foundation Grants 200020-144531 and 200021-137574.

Gábor Tardos was supported by the *Cryptography* "Lendület" project of the Hungarian Academy of Sciences.

Géza Tóth was supported by Hungarian National Science Foundation (OTKA) Grants K-111827 and K-83767.

This statement is almost tight, but not completely. Indeed, for $|X| = 5$, Theorem 1 guarantees the existence of a 4-member separating family, but it is easy to verify that 3 sets suffice. In the following generalization we give the best possible bound.

Theorem 2. *Let $X_1, \dots, X_k$ be pairwise disjoint sets with $|X_i| \le n$ for $i = 1, 2, \dots, k$. Let $X = \bigcup_{i=1}^k X_i$. If $\mathcal{F} \subseteq 2^X$ satisfies $|\mathcal{F}| > 2^{|X|-1}$, then $\mathcal{F}$ has a subfamily of cardinality at most $\lceil \log n \rceil + 1$ that separates X_i for every i ($i = 1, 2, \dots, k$).*

The above bound is tight, that is, the same statement is false for every $n \ge 2$, if we replace $\lceil \log n \rceil + 1$ by $\lceil \log n \rceil$.

We call $|\mathcal{F}|/2^{|X|}$ the *density* of $\mathcal{F}$. If this slips below $1/2$, we cannot guarantee the existence of a small subfamily separating X, as $\mathcal{F}$ itself does not necessarily separate X. But, as claimed by the next theorem, we can still find a small subfamily separating all pairs in X that $\mathcal{F}$ separates. We state this result for simultaneous separation.

Theorem 3. *Let $X_1, \dots, X_k$ be disjoint sets with $|X_i| \le n$ for all $i = 1, 2, \dots, k$. Let $X = \bigcup_{i=1}^k X_i$ and $\mathcal{F} \subseteq 2^X$. Then $\mathcal{F}$ has a subfamily of size at most $\lceil \log n \rceil + C \log \frac{1}{\alpha} \log\log \frac{1}{\alpha}$ separating every pair in each X_i that is separated by $\mathcal{F}$. Here $\alpha = |\mathcal{F}|/2^{|X|}$ is the density of $\mathcal{F}$ and C is a universal constant.*

Definition 1. Let X be a set of n elements. We call a pair (V, W) of disjoint subsets of X a *constraint*, $V \cup W$ is the *support* and $|V \cup W|$ is the *size* of the constraint. A subset A of X *satisfies* the constraint (V, W) if $V \subseteq A$ and $W \cap A = \emptyset$. A family $\mathcal{F}$ of subsets of X *satisfies* a constraint if the constraint is satisfied by some member of $\mathcal{F}$.

Note that the fact that two elements $x, y \in X$ are separated by a member of $\mathcal{F}$ means that at least one of the constraints $(\{x\}, \{y\})$ and $(\{y\}, \{x\})$ is satisfied by $\mathcal{F}$.

Given a family $\mathcal{F} \subseteq 2^X$ and a family of constraints satisfied by $\mathcal{F}$, we are looking for a small subfamily of $\mathcal{F}$ that also satisfies the given constraints. The next theorem establishes what the density of $\mathcal{F}$ has to be (depending on the size of the constraints) for this to be possible.

Theorem 4. *For a positive integer m and $1 - \frac{1}{2^{m-1}} < \alpha < 1$ there exists a constant $0 < c = c(m, \alpha)$ with the following property. If X is a finite set, $\mathcal{F}$ is a family of subsets of X with density above α and $\mathcal{C}$ is a collection of N constraints, each of size m and satisfied by $\mathcal{F}$, then there is a subfamily consisting of at most $c \log N$ sets from $\mathcal{F}$ that satisfies all the constraints in $\mathcal{C}$.*

A similar statement is false for any $m \ge 1$ and $\alpha = 1 - \frac{1}{2^{m-1}}$.

Up to this point, the ground set X and the family $\mathcal{F}$ were not assumed to possess any structure. The first such assumption we make is that of a bounded VC-dimension and prove a linear lower bound on the size of any separating subfamily in this case. Given a family $\mathcal{F}$ of sets, the *Vapnik-Chervonenkis dimension* (in short, VC-dimension) of $\mathcal{F}$ is the largest integer d for which there exists a d-element set A such that for every subset $B \subseteq A$, some member F of $\mathcal{F}$ has $F \cap A = B$.

Let us fix d and assume a set of size n is separated by a family $\mathcal{F}$ of VC-dimension d. In this case $|\mathcal{F}|$ must be at least polynomial in n, namely $|\mathcal{F}| \ge n^{1/(2^{d+1}-1)}$. This follows from the Shatter Function Lemma (or Sauer-Shelah Lemma, cf. Lemma 10.2.5 in [Mat02]) and the fact that the dual of VC-dimension d family has VC-dimension below 2^{d+1}. The size of a separator family of fixed VC-dimension d can, indeed, be a small polynomial of n, namely $|\mathcal{F}| < 2^d n^{1/(2^d-1)}$ can be obtained by considering the dual of the set system $\binom{Y}{2^d-1}$.

We can show even stronger, linear lower bound on the size of a separating family if the base set to be separated can be arbitrarily chosen from an infinite universe with a bounded VC-dimension set system.

Theorem 5. *Let U be an infinite set and $\mathcal{F} \subseteq 2^U$ a family of subsets of U of Vapnik-Chervonenkis dimension d. For every $n > 0$, there is a set $X \subset U$ with $|X| = n$ such that any subfamily of $\mathcal{F}$ which separates X, has at least $\frac{n-1}{d}$ members.*

Note that this theorem is almost tight. Let U be the set of positive integers and let $\mathcal{F}$ consist of the subsets of U containing 1 and separating at most d pairs $(i, i+1)$. The VC-dimension of $\mathcal{F}$ is d and any n element subset of U can be separated by $(n-1)/d + \log d$ members of $\mathcal{F}$.

Now, we turn to separation problems where X and $\mathcal{F}$ have some geometric structure.

Definition 2. A set of points in $\mathbb{R}^d$ is said to be *in general position* if no $d+1$ points lie on a $(d-1)$-dimensional affine subspace.

A prototype of separation questions in the geometric setting was first studied by Gerbner and Tóth [GT13]. They showed that for any set X of n points in general position in the plane, there is a family $\mathcal{F}$ of at most $20n \log\log n/\log n$ convex sets that separates X. On the other hand, there is a set X of n points in the plane in general position for which any separator of X, consisting of convex sets, has cardinality at least $n/(2\log n + 2)$.

It is natural to ask how these results about separating pairs of points could be extended to separation of pairs of k-tuples of points.

Definition 3. Let X be a set and k a positive integer with $k \leq |X| - 1$. We say that a set $F \subset X$ *containment-separates* a pair of k-element subsets of X if F contains one of them and does not contain the other. A family $\mathcal{F}$ of subsets of X is a *containment-separator* of k-subsets of X if for any two k-subsets A and B of X there is at least one member of $\mathcal{F}$ that separates them.

Our goal is to select a small subfamily of $\mathcal{F}$ that containment-separates the k-subsets of X. As a generalization of the question discussed in [GT13], we denote by $c_k^d(n)$ the minimum number c such that, for any set X of n points in general position in d-space, there is a containment-separator of k-subsets of X which consists of c convex sets. This makes sense for $k \leq d+1$ and we have $c_k^d(n) \leq \binom{n}{k}$ because the convex hulls of the k-subsets of X containment-separate the k-subsets. However, two k-subsets with the same convex hull are not containment-separated by any convex set. Thus, $c_k^d(n)$ does not exist if $d+2 \leq k < n$.

The result quoted above from [GT13] can be stated as

$$n/(2\log n + 2) \leq c_1^2(n) \leq 20n \log\log n/\log n.$$

Regarding higher dimensional point sets, it is observed in [GT13] that for any n and d,

$$c\frac{n}{\log^{d-1} n} \leq c_1^d(n) \leq C\frac{n \log\log n}{\log n}.$$

Our goal is to find bounds on $c_k^d(n)$ for $k \geq 2$.

Theorem 6. *For any $d = 2, 3, \ldots$, and any $k = 2, \ldots, d+1$, there exists $c(k,d) > 0$ such that the following holds for any $n > k+1$.*

$$c_2^1(n) = 2n - 4 \tag{1}$$

$$\left\lfloor \frac{n-2}{2} \right\rfloor \leq c_2^2(n) \leq 2n - 4, \tag{2}$$

$$\frac{n^2 - 2n - 3}{8} \leq c_3^2(n) \leq n^2 - n, \tag{3}$$

$$c(k,d)\frac{n^{\lfloor (k+1)/2 \rfloor}}{(\log n)^{\lfloor (2d-1-k)/2 \rfloor}} \leq c_k^d(n) \leq 2\binom{n}{k-1} \tag{4}$$

It is a challenging problem to narrow the gap between the two bounds in (4). Theorems 2 and 3 are shown in Section 2, Theorem 4 in Section 3, Theorem 5 in Section 4. Theorem 6 is proved in Sections 5. We briefly discuss intersection-separation, a relative of containment-separation, in Section 6.

2. Proofs of Theorems 2 and 3

Proof of Theorem 2. We replaced our original proof of the first part of this theorem by a more elegant argument of András Mészáros [M14], which was submitted as a solution to our problem at the Miklós Schweitzer competition in 2014.

We regard $V = 2^X$ as a vector field over $\mathbb{F}_2$ with respect to the symmetric difference of sets, which we denote by $+$. As $|X_i| \leq n$ for all i we can find a family U of at most $\lceil \log n \rceil$ subsets of X (not necessarily in $\mathcal{F}$) that separates each X_i. Let W denote the linear subspace of V spanned by U. The translates of W partition V, and thus, there is a translate $W + c$, more than half of whose members are in $\mathcal{F}$. Pick a $d \in (W + c) \cap \mathcal{F}$ and consider the set $\{x \in W : \ x + d \in \mathcal{F}\}$. This set has cardinality larger than $|W|/2$, so it spans W, and thus, it contains a basis Z of W. Now, $|Z| = \dim W \leq |U| \leq \lceil \log n \rceil$. We claim that the set $S = \{d\} \cup \{z + d : \ z \in Z\}$ separates each X_i. To see this, observe that for any pair of elements $x, y \in X$, the set of those elements of V that do not separate x and y form a linear subspace of V, so if S did not separate x and y neither did any set generated by S. But this is not possible for $x, y \in X_i$ as S generates each element z of Z through $z = d + (z + d)$ and thus S also generates the subspace W including the sets in U, one of which separates x from y. This finishes the proof of the first part of the theorem.

Now we prove the second part. Let X be the union of $k = 2^{n-1}$, pairwise disjoint n-element sets, let $\mathcal{F}$ consist of the elements of 2^X that intersect at least one class X_i in $\emptyset$ or X_i. Then $2^X \setminus \mathcal{F}$ has $(2^n - 2)^{2^{n-1}}$ elements, and

$$\frac{|2^X \setminus \mathcal{F}|}{|2^X|} = \left(1 - \frac{1}{2^{n-1}}\right)^{2^{n-1}} < \frac{1}{e} < \frac{1}{2}.$$

Hence, $\mathcal{F}$ contains more than half of the elements of X. On the other hand, assume that $\mathcal{G} \subset \mathcal{F}$ separates X, and let $G \in \mathcal{G}$. Let X_i be a class in X which is intersected by G either in $\emptyset$ or in X_i. Clearly, to separate X_i, we need at least $\lceil \log n \rceil$ more elements of $\mathcal{F}$, which implies the assertion. $\square$

Before presenting the details of the proof of Theorem 3 we introduce some notation and give a sketch of the proof.

In the setup of both Theorems 2 and 3 we are given a base set X partitioned into the parts X_i of size at most n and we want to refine this partition by selecting a small subset of a family $\mathcal{F} \subseteq 2^X$. In the proof of the latter theorem we will select this separating subfamily in phases. After selecting a subfamily $\mathcal{F}_1 \subseteq \mathcal{F}$ we measure the progress by the maximal size m of a yet un-separated part of some X_i. Clearly, we need at least $\log(n/m)$ sets to partition a set of size n to sets of size m or smaller. Accordingly, we call the quantity $|\mathcal{F}_1| - \log(n/m)$ the *loss* incurred in decreasing the part size from n to m.

In this terminology we can phrase Theorem 2 as stating that if the density of $\mathcal{F}$ is above $1/2$, then we can decrease the part size all the way to 1 with a loss of less than 2. Similarly, Theorem 3 states that if $\mathcal{F}$ has density α and separates all X_i, then we can decrease the part size to 1 with a loss of $O(\log(1/\alpha)\log\log(1/\alpha))$. Note that a loss of $\log(1/\alpha)$ is unavoidable in certain cases, for example if we have a small set $Y \subset X_1$ and $\mathcal{F}$ consists of all sets disjoint from Y, containing Y, or having size 1. We are not sure if the $\log\log(1/\alpha)$ factor is needed.

We will prove Theorem 3 by constructing the separating family in phases. The first stage will reduce the part size to at most $1/\alpha$ for a loss of less than 2. This stage is a generalization of Theorem 2 and proved very similarly.

In the second phase we further reduce the part size to $O(\log(1/\alpha))$ for a loss of $O(\log(1/\alpha))$. We will select the separating sets in this phase one by one, but we remark that selecting them at once one can decrease the loss incurred in this phase to a constant for the small price of reducing the part size to $O(\log^2(1/\alpha))$ or even $O(\log(1/\alpha)\log\log(1/\alpha))$ (instead of $O(\log(1/\alpha))$ as presented here). Unfortunately, the loss incurred in the third phase is much larger and dominates the losses in the other phases.

In the third phase we make sure that all but $O(\log(1/\alpha))$ elements of X form singleton parts, while in the final fourth phase we separate the remaining few non-singleton parts to "atomic" parts not even separated by $\mathcal{F}$.

Let A and Y be finite sets. We say that *A cuts Y well* if $|Y|/4 \le |Y \cap A| \le 3|Y|/4$. In the second phase the following trivial observation is going to be useful:

Lemma 1. Let Y be a finite set of size $m \ge 2$. The density α of the subsets of Y that do not cut Y well satisfies $\alpha \le 1/2$ and $\alpha \le 2^{-m/10}$.

In the third phase we use the following result of Brace and Daykin [BD71].

Theorem 7 (Brace and Daykin, 1971)**.** *Let $t > 1$ be an integer, Y be a set of size s and $\mathcal{F}'$ be a subset of 2^Y of density exceeding $(t+2)/2^{t+1}$. If $\bigcup \mathcal{F}' = Y$, then there are t elements of $\mathcal{F}'$ whose union is also Y.*

Proof of Theorem 3. In the first phase of the selection of the separating subfamily of $\mathcal{F}$ we decrease the maximal size of a part from n to at most $1/\alpha$. In case $n \le 1/\alpha$ one can simply skip this phase.

We mimic the proof of Theorem 2. We assume without loss of generality, that $n = \max_i |X_i|$ and set $r = \lceil \log n \rceil$. We let $V = \mathbb{F}_2^r$ be the r-dimensional vector space over the two element field with the usual inner product and choose $f : X \to V$ that is injective on each X_i. We set $W = \{O_x \mid x \in V\}$, where O_x consists of the elements a of X with

$f(a)$ not orthogonal to x. We regard 2^X as a group with the symmetric difference operation (denoted by $+$) and note that W is a subgroup. As the density of $\mathcal{F}$ is α we can choose $C \in \mathcal{F}$ such that the density of $\mathcal{F}$ within the coset $C + W$ is at least α, that is the set $S = \{x \in V : C + O_x \in \mathcal{F}\}$ satisfies $|S| \geq \alpha 2^r$.

With $\alpha \leq 1/2$ the set S does not necessarily generate the whole of V. But we can still choose a basis $B \subseteq S$ for the subspace of V generated by S and it is easy to see that the set $\mathcal{F}_1 = \{C + O_x \mid x \in B\} \cup \{C\}$ of size $|B| + 1$ separates each X_i into parts of size at most $2^{r-|B|} \leq 1/\alpha$. Thus, we have decreased the part size to at most $1/\alpha$ for a loss of less than 2. We have $|\mathcal{F}_1| < \log n + 2$.

In the second phase we do similarly as in the first phase but we do not have a linear structure on V any more. Let m be the maximal size of a part in the current partition of $\mathcal{F}$. We take V to be an m element set and select a function $f : X \to V$ that is injective in every part. We set $W = \{f^{-1}(H) \mid H \subseteq V\}$. As before, W is a subgroup of 2^X (considered with the symmetric difference). We select $C \in \mathcal{F}$ such that the density of $\mathcal{F}$ in the coset $C + W$ is at least α, that is, the set $S = \{H \subseteq V \mid C + f^{-1}(H) \in \mathcal{F}\}$ satisfies $|S| \geq \alpha 2^m$. In case S contains a set H that cuts V well, we choose one such set and include the sets C and $C + f^{-1}(H)$ in our separating family. This makes all the parts in the current partition be at most $3m/4$.

By Lemma 1 the selection of the set $H \in S$ cutting V well is possible as long as $\alpha > 2^{-m/10}$, that is $m > 10 \log(1/\alpha)$. In the second phase we repeat the above procedure till all parts of the current partition is of size at most $10 \log(1/\alpha)$ and call $\mathcal{F}_2$ the set of separating sets collected. As the maximal size of a part was at most $1/\alpha$ in the beginning of phase two, and this maximal size decreases by a constant factor in each round when we add two sets to $\mathcal{F}_2$ we have $|\mathcal{F}_2| = O(\log(1/\alpha))$.

In the third phase we use a different strategy. Let $\{X'_i \mid i \in I\}$ be the set of parts in the current partition of X. We call a part X'_i *good* if there is a set $A \in \mathcal{F}$ cutting X_i well and we call X'_i *bad* if it has at least 2 elements, but it is not good. Let $G = \{i \in I \mid X'_i \text{ is good}\}$. For $i \in G$ we select a set $B_i \subset 2^{X'_i}$ of size $|B_i| = 2^{|X'_i|-1}$ such that all sets in B_i cuts X'_i well and we have a set $A \in \mathcal{F}$ with $A \cap X'_i \in B_i$. This is possible as by Lemma 1 at least half the subsets of X'_i cut X'_i well.

We define the function $f : 2^X \to 2^G$ by setting $f(A) = \{i \in G \mid A \cap X'_i \in B_i\}$. Let $\mathcal{F}' = f(\mathcal{F})$. Clearly, f takes all values an equal number of times, thus the density of $\mathcal{F}'$ in 2^G is at least the density α of $\mathcal{F}$ in 2^X. Note that for all $i \in G$ we have $A \in \mathcal{F}$ with $i \in f(A)$, thus we have $\bigcup \mathcal{F}' = G$. We choose $t = O(\log(1/\alpha))$ such that $(t+2)/2^{t+1} < \alpha$ and apply Theorem 7 to find t sets $I_1, \ldots, I_t \in \mathcal{F}'$ with $\bigcup_{i=1}^t I_i = G$. We find sets $A_i \in \mathcal{F}$ with $f(A_i) = I_i$ and include these t sets in our partitioning family. Note that if $x \in X$ is contained in a good part X'_i of size m, then the size of the part containing x after considering these t new separating sets is at most $3m/4$.

We repeat the above procedure for $\lceil \log(10 \log(1/\alpha))/\log(4/3) \rceil = O(\log \log(1/\alpha))$ times and obtain $\mathcal{F}_3$ as the union of all the elements of $\mathcal{F}$ we selected. We clearly have $|\mathcal{F}_3| = O(\log(1/\alpha) \log \log(1/\alpha))$.

We call $x \in X$ fully separated if it forms a singleton part in the current partition after the third phase. Clearly, if x is not fully separated, it must have been in a bad part at some time. Let us choose the earliest bad part containing x and consider all the distinct sets $Y_1, \dots, Y_j$ obtained this way from not fully separated elements. Clearly, these sets are pairwise disjoint. Let $Y = \bigcup_{i=1}^{j} Y_i$. By Lemma 1 the ratio of subsets of X not cutting Y_i well is at most $2^{-|Y_i|/10}$. A random subset $A \subseteq X$ intersects the sets Y_i independently, thus the probability that it cuts none of the Y_i well is at most $\prod_{i=1}^{j} 2^{-|Y_i|/10} = 2^{-|Y|/10}$. By the definition of bad parts, no set $A \in \mathcal{F}$ cuts any of the sets Y_i well, so we have $\alpha \le 2^{-|Y|/10}$. Therefore, all non-singleton parts in the current partition after the third phase is contained in the set Y of size at most $10 \log(1/\alpha)$.

Finally in the last phase we select any set in $\mathcal{F}$ that refines our current partition and repeat this process until no further refinement is possible. Clearly, the set $\mathcal{F}_4$ selected in this phase satisfies $|\mathcal{F}_4| < |Y| = O(\log(1/\alpha))$.

The family $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3 \cup \mathcal{F}_4$ separates any pair of elements in the same part X_i that is separated by $\mathcal{F}$ and the size of this set is $\log n + O(\log(1/\alpha) \log\log(1/\alpha))$. This finishes the proof of Theorem 3. □

3. Proof of Theorem 4

For $0 < \varepsilon < 1$, we call a constraint *ε-good* if at least $\varepsilon|\mathcal{F}|$ members of $\mathcal{F}$ satisfy it, and *ε-bad* otherwise.

The proof of the first statement of Theorem 4 consists of two steps. First, a standard application of the probabilistic method shows that for any $\varepsilon > 0$, all ε-good constraints can be satisfied by $(\log N)/\varepsilon$ randomly chosen members of $\mathcal{F}$. Second, we show that if ε is set sufficiently small as a function of m and α but independent of $|X|$ and N, then the number of ε-bad constraints is bounded by another value Z depending on m and α, and independent of $|X|$ and N. Since all constraints are satisfiable, it means that adding Z (well chosen) members of $\mathcal{F}$ to the $(\log N)/\varepsilon$ random members satisfying the ε-good constraints we obtain a collection satisfying all the N constraints.

To prove the first step, let C be a ε-good constraint, and choose $(\log N)/c$ members of $\mathcal{F}$ randomly, uniformly.

$$\mathbb{P}(C \text{ is not satisfied by any of the chosen sets}) \le (1-c)^{\frac{\log N}{c}} < \frac{1}{N}.$$

Thus, with non-zero probability, all the at most N ε-good constraints are satisfied by the randomly chosen members of $\mathcal{F}$.

For the second step let a be a large enough number depending on α and m to be set shortly. Assume that the supports of $a2^m$ distinct ε-bad constraints form a "sunflower", that is any two of them intersect in the same set C. Let $b = |C|$ and note that $0 \le b \le m-1$. Clearly, we can find a of these constraints, say (V_i, W_i) for $1 \le i \le a$ that contain the elements of C "on the same side", that is C is the support of a constraint (V_0, W_0) with $V_0 \subseteq V_i$ and $W_0 \subseteq W_i$ for $i = 1, \dots, a$. Now consider a uniform random subset $A \subseteq X$. This set satisfies the constraint (V_0, W_0) with probability 2^{-b}. If it satisfies (V_0, W_0), then the conditional probability that it also satisfies (V_i, W_i) is 2^{b-m} for any $i \ge 1$. From the sunflower property we see that assuming A satisfies (V_0, W_0) the events that A satisfy (V_i, W_i) are mutually independent, so the overall probability P that A satisfies at least one of them is exactly

$$P = 2^{-b}(1 - (1 - 2^{b-m})^a).$$

On the other hand, A has the chance at least α to be in $\mathcal{F}$, and assuming it is in $\mathcal{F}$, it has a chance of less than ε to satisfy any of those constraints. Therefore we have

$$P < (1-\alpha) + a\alpha\varepsilon.$$

Let us select $\varepsilon > 0$ small enough and a large enough (depending on m and α) such that

$$2^{-b}(1-(1-2^{b-m})^a) \geq (1-\alpha) + a\alpha\varepsilon$$

holds for any $0 \leq b \leq m-1$. This is possible as increasing a the left hand side approaches $2^{-b} \geq 2^{1-m}$ and for fixed a, and ε approaching 0, the right hand side approaches $1-\alpha < 2^{1-m}$.

With this choice of ε and a the imminent contradiction in the last three displayed equations shows that the supports of no $a2^m$ ε-bad constraints form a sunflower. By the sunflower lemma of Erdős and Rado [ER60] we find that the total number of ε-bad constraints is at most $Z = m!2^m(a2^m)^m$. (The extra 2^m factor is coming from the possibility that many ε-bad constraints may have the same support.) This bound is independent of N and $|X|$ as claimed and finishes the proof of the first statement of Theorem 4.

For the second part, for any m and N we construct a base set X, a family $\mathcal{F}$ of density strictly above $1-2^{1-m}$ and N one sided constraints $(V_i, \emptyset)$ of size m, all satisfied by exactly one member of the family $\mathcal{F}$ and such that different constraints are satisfied by different members of $\mathcal{F}$. Thus, all constraints are satisfied by $\mathcal{F}$ but no subset of cardinality less than N satisfies them all.

For this we set $|X| = N+m-1$ and identify a subset $Y \subset X$ of size $|Y| = m-1$. Let $\mathcal{F}$ consist of all the subsets of X not containing Y plus all the m element subsets. We select the constraints $(V_i, \emptyset)$ with all possible m-subsets V_i of X containing Y.

4. Proof of Theorem 5

Assume without loss of generality that U is the set of positive integers. Since the VC-dimension of $\mathcal{F}$ is d, for any $(d+1)$-element subset $A = \{a_1, a_2, \ldots, a_{d+1}\} \subset U$ with $a_1 < a_2 < \ldots < a_{d+1}$, there is a set $T(A) \subseteq \{1, 2, \ldots, d+1\}$ such that no $F \in \mathcal{F}$ satisfies $F \cap A = \{a_i | i \in T(A)\}$. If there is more than one such set, we fix $T(A)$ arbitrarily, and we call it the *type* of A.

By Ramsey's Theorem, there is a set $X = \{y_1, y_2, \ldots y_n\} \subset U$ with $y_1 < \ldots < y_n$, such that all $(d+1)$-element subsets of X have the same type T.

Let us call the index $1 \leq i \leq d$ *regular* if T separates i from $i+1$ and *singular* otherwise. We let k stand for the number of singular indices.

Consider any $F \in \mathcal{F}$. We claim that there is a set F^* of at most k elements of X and a partition of X into at most $d-k+1$ intervals such that the symmetric difference $F+F^*$ does not separate any two elements in the same interval. To see this consider the greedy process looking for indices $i_1 < i_2 < \ldots < i_{d+1}$ such that for the set $H = \{y_{i_1}, \ldots, y_{i_{d+1}}\}$ we have $H \cap F = \{y_{i_j} \mid j \in T\}$. As such a set could not have type T, we cannot find all these indices, nevertheless, some of the indices can be found by a greedy process (for example, if $1 \in T$, then we start with $i_1 = \min\{j \mid y_j \in F\}$). We can satisfy the claim by making F^* consist of the elements y_{i_j}, where j is singular and starting a new interval in the partition of X at every element y_{i_j} with j regular.

Now assume the family $\{F_1, \dots, F_m\} \subseteq \mathcal{F}$ separates X. Let F_i^* be the corresponding sets of size at most k and consider the set $V = X \setminus \bigcup_{i=1}^m F_i^*$. We have $|V| \geq n - mk$ determining at least $n - mk - 1$ neighboring pairs of elements. By the above claim each set F_i separates at most $d - k$ of these neighboring elements, thus we must have $n - mk - 1 \leq m(d - k)$, that is, $m \geq (n-1)/k$. This completes the proof of Theorem 5.

5. Containment-separation

In this section we prove Theorem 6.

We start with the proof of (1). Let $H = \{x_1, \dots, x_n\}$ be an arbitrary n element subset of the real line and assume $x_1 < \dots < x_n$. Note that the family $\{[x_1, x_i] \mid 1 < i < n\} \cup \{[x_i, x_n] \mid 1 < i < n\}$ of $2n - 4$ convex sets containment-separates all the pairs in H. This proves $c_2^1(n) \leq 2n - 4$. On the other hand realize that no convex set containment-separates more than one of the $2n - 4$ pairs $(\{x_1, x_i\}, \{x_1, x_{i+1}\})$ for $1 < i < n$ and $(\{x_i, x_n\}, \{x_{i-1}, x_n\})$ for $1 < i < n$. This proves $c_2^1(n) \geq 2n - 4$.

We will use the following monotonicity property:

$$c_k^d(n) \leq c_k^{d-1}(n),$$

for all choices of $k \leq d$ and n. To see this consider a set X of n points in general position in d-space and find a projection π to $(d-1)$-space such that $\pi(X)$ is again n points in general position. A generic projection π satisfies this. Now find $c_k^{d-1}(n)$ convex sets to containment separate all the k-sets of $\pi(X)$ and consider the inverse images of these sets for the projection π. Clearly, these sets are convex and they containment-separate the k-subsets of X.

A similar monotonicity also holds in k if $n > k + 1$:

$$c_k^d(n) \leq c_{k+1}^d(n).$$

This is because any collection of sets containment-separating the $(k+1)$-subsets of an n-set H also containment-separate the k-subsets. Indeed, if A and B are k-subsets, then any set containment separating $A \cup \{x\}$ from $B \cup \{x\}$ also containment-separates A from B. This trick works if we can choose $x \in H$ outside $A \cup B$. In case $A \cup B = H$ we pick $x \in A \setminus B$ and $y \in B \setminus A$ and use that any set containment-separating $A \cup \{y\}$ from $B \cup \{x\}$ also containment separates A from B. This latter trick fails if $|A \cap B| = k - 1$, but then $|A \cup B| = k + 1 < n$.

Next we prove (2). The upper bound follows from (1) via the monotonicity mentioned above.

For the lower bound in (2) let $p = (1, 1)$ and consider the set X consisting of p and $n - 1$ points on the unit circle around the origin, all in the first quadrant. This set is in general position. Let us denote the points of X on the circle $x_1, \dots, x_{n-1}$ in order of increasing x-coordinate. The property of the arrangement we use is that the the convex hull of $\{x_i, x_j, p\}$ contains all x_k with $i < k < j$. This implies that any convex set can containment-separate at most two of the pairs of two element sets $(\{x_i, p\}, \{x_{i+1}, p\})$ for $1 \leq i \leq n - 1$ and the stated lower bound on $c_2^2(n)$ follows.

Next we turn to the upper bound in (4). By the monotonicity mentioned above it is enough for us to prove $c_k^{k-1}(n) \leq 2\binom{n}{k-1}$. Let us consider a set X of n points in general position in the $(k-1)$-dimensional space. Each $(k-1)$-subset of X determines a hyperplane. Consider all the closed half-spaces bounded by one of these hyperplanes. This is a collection of $2\binom{n}{k-1}$ convex sets and it containment-separates all k-subsets of X. Indeed, if A and B are distinct k-subsets of X, then either the convex hull of A does not contain B, or vice versa. In the former case a supporting half-space of the convex hull of A containment-separates the sets.

We turn to the proof of (3). The upper bound is a special case of the upper bound in (4). For the lower bound we give a construction.

Let us assume $n \geq 8$. Let X_{ex} be the vertex set of a regular k-gon around the origin for $k = 2\lfloor n/4 \rfloor$. We call the opposite pairs of points in X_{ex} a *diameter*. Let us find a point x distinct from the origin but so close to it that it is contained in the interior of the convex hull of any two diameters. We further assume that $X_{\text{ex}} \cup \{x\}$ is in general position. Let $l = n - k$ and $X_{\text{in}} = \{x_i \mid 1 \leq i \leq l\}$, where $x_i = \frac{i}{l}x$.

Consider the pairs of 3 element sets $\{\{p, -p, x_i\}, \{p, -p, x_{i+1}\}\}$, where $p \in X_{\text{ex}}$ (so $\{p, -p\}$ is a diagonal) and $1 \leq i \leq l - 1$. No set containing no diagonal can containment-separate any of these pairs, and convex sets containing more than one diagonal contain all x_i in their interior, so they do not separate these pairs either. A convex set S containing a single diagonal $\{p, -p\}$ separates at most one of these pairs, since, if $x_i \in S$, then we also have $x_j \in S$ for all $j < i$.

This shows that $X_{\text{ex}} \cup X_{\text{in}}$ is a good choice for a hard to separate set, but it is not in general position. Fortunately, the above arguments are robust against small perturbations. Let us obtain X as the union of X_{ex} and a set $\{x'_i \mid 1 \leq i \leq n\}$ where x'_i is ε-close to x_i but perturbed in such a way that X is in general position.

It is easy to see that if $\varepsilon > 0$ is small enough, then each convex set can containment-separate at most one of the pairs $(\{p, -p, x'_i\}, \{p, -p, x'_{i+1}\}$. To containment-separate all these pairs, one needs at least $(l - 1)k/2$ convex sets, proving the lower bound in (3).

Finally, we prove the lower bound in (4). For this we need the following result. For any $n > d \geq 2$, Károlyi and Valtr [KV03] constructed a set of n points in the d-dimensional space which contains at most $c_d \log^{d-1} n$ points in convex position, where c_d depends only on d. We call such a point set a *Károlyi-Valtr construction*.

Let us assume that k is odd as the case of even k comes from the monotonicity of $c_k^d(n)$ in k. We need to construct a set X of n points in general position in d-space, whose k-subsets is hard to containment-separate with convex sets. Let $m = \frac{2d+1-k}{2}$. Take a set A of size $\frac{n}{4}$ and a set A' of size m such that all the $\frac{n}{4} + m$ points in $A \cup A'$ are at unit distance from the origin and the unit vectors corresponding to any d of them are linearly independent. Let $X_{\text{ex}} = -A \cup A$, where $-A = \{-p \mid p \in A\}$. We call the points p and $-p$ in X_{ex} an *opposite pair*. Let F be an m-flat passing through the origin and the points of A'. Consider a ball B centered at the origin that is so small that the convex hull of any d opposite pairs of X_{ex} contains B in its interior. Let X_{in} be an m-dimensional Károlyi-Valtr construction of $\frac{n}{4}$ points in $F \cap B$ and assume (without loss of generality) that X_{in} as an m-dimensional set is in general position. (Note, however, that neither X_{ex} nor X_{in} is in general position in d-space.)

We claim that if a convex set S contains $\frac{k-1}{2}$ opposite pairs, then the points of X_{in} on the boundary of S are in convex position. Indeed, otherwise there would be a set H of $m + 1$ points in X_{in}, all on the boundary of S and forming a simplex in F such that the simplex contains yet another point $x \in X_{\text{in}}$ on the boundary of S. The contradiction comes from the fact that in this case the convex hull of the union of H and the opposite pairs in S has x in its interior.

As a consequence, we see that if a convex set contains $\frac{k-1}{2}$ opposite pairs, then its boundary contains at most $c_m \log^{m-1} n$ points of X_{in}.

Set $\varepsilon > 0$ very small and let X be a point set consisting of a point ε-close to each point in X_{ex} and two distinct points ε-close to each point in X_{in}. We choose the points in X to be in general position and let $f : X_{\text{ex}} \to X$, $f_1 : X_{\text{in}} \to X$ and $f_2 : X_{\text{in}} \to X$ be the functions showing our choices.

Consider a set H of $\frac{k-1}{2}$ opposite pairs, and a point $x \in X_{\text{in}}$ and let $P(H, x)$ be the pair $(f(H) \cup \{f_1(x)\}, f(H) \cup \{f_2(x)\})$. We claim that if ε is small enough, then any convex set can containment-separate at most $c_m \log^{m-1} n$ pairs $P(H, x)$ with a fix H. Indeed, to separate the pair $P(H, x)$ the convex set must contain $f(H)$ and has to have a boundary point on the (short) interval $f_1(x)f_2(x)$. In the limit for $\varepsilon \to 0$ we find a convex set containing H and having x on its boundary. As we saw above this is possible for at most $c_m \log^{m-1} n$ points $x \in X_{\text{in}}$.

If a convex set S contains d opposite pairs, then it contains X_{in} in its interior. Therefore, for small enough ε, a convex set containing $f(H)$ for a collection H of d opposite pairs containment-separates no pair $P(H, x)$.

To containment-separate $P(H, x)$ a (convex) set must contain $f(H)$, so the above bounds mean that (again, for small enough ε) no convex set containment-separates more than $c_m \binom{d-1}{\frac{k-1}{2}} \log^{m-1} n$ of the pairs $P(H, x)$. Comparing this with the total number of $\frac{n}{4}\binom{\frac{n}{4}}{\frac{k-1}{2}}$ of the pairs $P(H, x)$ shows that we need many convex sets to containment-separate the k-subsets of X. This finishes the proof of the lower bound in (4) and with that the proof of Theorem 6.

6. A remark: Intersection-separation

In the geometric setting, we discussed containment-separation. We can extend the notion of separation of points to k-tuples in another way as well. We say that a set F *intersection-separates* a pair of k-element subsets of X, if F intersects one of them and is disjoint from the other. A family $\mathcal{F}$ of subsets of X *intersection-separates* the k-element subsets of X if, for any pair of k-element subsets of X, there is a member of $\mathcal{F}$ that intersection-separates that pair. And thus, we can define the intersection-separation numbers as

$$i_k^d(n) = \max_{\substack{X \subset \mathbb{R}^d, |X| = n \\ X \text{ in general pos.}}} \min \left\{ |\mathcal{G}| : \mathcal{G} \subset \mathcal{C}^d \; \begin{array}{c} \text{an intersection-separator of} \\ k\text{-subsets of } X \end{array} \right\},$$

where $\mathcal{C}^d$ denotes the family of convex subsets of $\mathbb{R}^d$.

The number $i_k^d(n)$ is always defined and at most $n-1$ as the singleton subsets intersection-separate, even if we omit one of them.

The following monotonicity properties can be verified exactly as for containment-separation.

$$i_k^d(n) \le i_k^{d-1}(n), \quad \text{and} \quad i_k^d(n) \le i_{k+1}^d(n) \tag{5}$$

Proposition 1. *For any $d = 2, 3, \ldots$ there is a constant $c_d > 0$ such that for any $n > k \geq 2$ we have*

$$\left\lfloor \frac{n+3}{6} \right\rfloor \leq \quad i_k^2(n) \quad \leq n-1 \tag{6}$$

$$c_d \frac{n}{\log^{d-1} n} \leq \quad i_k^d(n) \quad \leq n-1 \tag{7}$$

We have already mentioned the upper bounds. The lower bound in (7) can be proved by replacing each point of a Károlyi–Valtr construction (see Section 5) by a pair of twins (two very close points). We provide a construction to show the lower bound in (6).

By (5) it is sufficient to consider the case $k = 2$. Suppose without loss of generality that $n = 3m$. We give the points in polar coordinates (r, ϕ). Let $\varepsilon > 0$ be very small. For $0 \leq i \leq m-1$, let p_i be the point $(1-\varepsilon, 2i\pi/m)$, q_i the point $(1, 2i\pi/m + \varepsilon^2)$, and r_i the point $(1, 2i\pi/m - \varepsilon^2)$. Let X be the set of these $n = 3m$ points. For any i, $0 \leq i \leq m-1$, consider the following pair of pairs: $(\{p_i, q_i\}, \{p_i, r_i\})$.

To finish the proof, we claim that no convex set intersection-separates more than two of these m pairs of pairs. Indeed, suppose that a convex set K intersection-separates the pairs corresponding to the indices i, j and k. We may assume that the greatest angle of the triangle $p_i p_j p_k$ is at p_j. Now, K contains one of q_i and r_i, one of q_j and r_j, and one of q_k and r_k. It is easy to see that, if ε is small enough, K contains p_j, a contradiction.

References

[AhlW87] R. Ahlswede and I. Wegener, *Search problems. Wiley-Interscience Series in Discrete Mathematics and Optimization*, John Wiley & Sons, Chichester, 1987.

[B72] J. A. Bondy, Induced subsets, *Journal of Combinatorial Theory, Series B* **12** (1972), 201–202.

[BU95] R. P. Boland, J. Urrutia, Separating collections of points in Euclidean spaces, *Information Processing Letters* **53** (1995), 177–183.

[BD71] A. Brace and D. E. Daykin, A finite set covering theorem. I, II. *Bull. Austral. Math. Soc.*, 5:197–202; ibid. 6 (1972), 19–24, 1971.

[DH94] D. Z. Du and F. K. Hwang, *Combinatorial Group Testing and Its Applications*, World Scientific, 2nd ed., 2000.

[ER60] P. Erdős and R. Rado, Intersection theorems for systems of sets, *J. London Math. Soc.*, **35** (1960), 85–90.

[ES35] P. Erdős and G. Szekeres, A combinatorial problem in geometry, *Compositio Mathematica* **2** (1935), 463–470.

[ES60] P. Erdős and G. Szekeres, On some extremum problems in elementary geometry, *Ann. Universitatis Scientiarum Budapestinensis, Eötvös, Sectio Mathematica* **3/4** (1960–61), 53–62.

[GT13] D. Gerbner and G. Tóth, Separating families of convex sets, *Computational Geometry: Theory and Applications* **46** (2013), 1056–1058.

[KV03] Gy. Károlyi and P. Valtr, Point configurations in d-Space without large subsets in convex position, *Discrete and Computational Geometry* **30** (2003), 277–286.

[Mat02] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.

[M14] A. Mészáros. *Personal communication:* a solution to a problem at the Miklós Schweitzer competition in 2014.

[TV05] G. Tóth and P. Valtr, The Erdős-Szekeres theorem, upper bounds and generalizations, in: *Discrete and Computational Geometry–Papers from the MSRI Special Program* (J. E. Goodman et al., eds.), MSRI Publications **52**, Cambridge University Press, Cambridge (2005), 557–568.

[W09] G. Wiener, Rounds in combinatorial search, in: *Search Methodologies* (eds. R. Ahlswede et al., eds.), Dagstuhl Seminar Proceedings, 2009., Paper 5., 1–7.

(Zs. Lángi) Budapest University of Technology, Budapest
E-mail address: zlangi@math.bme.hu

(M. Naszódi) EPFL, Lausanne and Eötvös University, Budapest
E-mail address: marton.naszodi@math.elte.hu

(J. Pach) EPFL, Lausanne and Rényi Institute, Budapest
E-mail address: pach@cims.nyu.edu

(G. Tardos) Rényi Institute, Budapest
E-mail address: tardos.gabor@renyi.mta.hu

(G. Tóth) Rényi Institute, Budapest and Budapest University of Technology
E-mail address: toth.geza@renyi.mta.hu