A Sauer-Shelah-Perles Lemma for Sumsets

Zeev Dvir[*]

Department of Computer Science and Department of Mathematics
Princeton University
Princeton, U.S.A.

zeev.dvir@gmail.com

Shay Moran[†]

School of Mathematics
Institute for Advanced Study
Princeton, U.S.A.

shaymoran1@gmail.com

Submitted: Jun 20, 2018; Accepted: Oct 23, 2018; Published: Nov 16, 2018

Abstract

We show that any family of subsets $A \subseteq 2^{[n]}$ satisfies $|A| \leqslant O\big(n^{\lceil d/2 \rceil}\big)$, where d is the VC dimension of $\{S \triangle T \,|\, S, T \in A\}$, and $\triangle$ is the symmetric difference operator. We also observe that replacing $\triangle$ by either $\cup$ or $\cap$ fails to satisfy an analogous statement. Our proof is based on the polynomial method; specifically, on an argument due to [Croot, Lev, Pach '17].

Mathematics Subject Classifications: 05D05, 05E99

1 Introduction

Let $A \subset 2^{[n]}$ be a family of subsets of an n element set ($[n]$ w.l.o.g). The VC dimension of A, denoted by $\mathsf{VC\text{-}dim}(A)$, is the size of the largest $Y \subseteq [n]$ such that $\{S \cap Y \,|\, S \in A\} = 2^Y$. One of the most useful facts about the VC dimension is given by the Sauer-Shelah-Perles Lemma.

Theorem 1 (Sauer-Shelah-Perles Lemma [12, 13]). *Let $d \leqslant n \in \mathbb{N}$. Suppose $A \subset 2^{[n]}$ satisfies $\mathsf{VC\text{-}dim}(A) \leqslant d$. Then $|A| \leqslant \binom{n}{\leqslant d}$.*

[*] Research supported by NSF CAREER award DMS-1451191 and NSF grant CCF-1523816.

[†] Research supported by the National Science Foundation under agreement No. CCF-1412958 and by the Simons Foundations.

The Sauer-Shelah-Perles Lemma has numerous applications ranging from model theory, probability theory, geometry, combinatorics, and various fields in computer science. A simple-yet-useful corollary of this lemma is that if $\mathsf{VC\text{-}dim}(A) \leqslant d$, and $\star$ is any binary set-operation (e.g. $\star \in \{\cap, \cup, \triangle\}$) then

$$\bigl|\{S \star T \mid S, T \in A\}\bigr| \leqslant \binom{n}{\leqslant d} \cdot \binom{n}{\leqslant d} = O(n^{2d}).$$

This corollary is used, for example, by [2] to derive closure properties for *PAC learnability.* Let $A \circledast A$ denote the family $\{S \star T \mid S, T \in A\}$. In this work we explore the converse direction: Does an upper bound on the VC-dimension $\mathsf{VC\text{-}dim}(A \circledast A)$ imply an upper bound on $|A|$? It is not hard to see that $\mathsf{VC\text{-}dim}(A) \leqslant \mathsf{VC\text{-}dim}(A \circledast A)$ for $\star \in \{\cup, \cap, \triangle\}$, and therefore, by Theorem 1: $\mathsf{VC\text{-}dim}(A \circledast A) < d \implies |A| \leqslant O(n^d)$.

Our main result quadratically improves this naive bound when $\star$ is symmetric difference:

Theorem 2. *Let $d \leqslant n \in \mathbb{N}$. Suppose $A \subset 2^{[n]}$ satisfies $\mathsf{VC\text{-}dim}(A \bigcirc\!\!\!\!\!\triangle A) \leqslant d$. Then*

$$|A| \leqslant 2\binom{n}{\leqslant \lfloor d/2 \rfloor}.$$

We note that Theorem 2 does not hold when $\star \in \{\cup, \cap\}$: pick $d \geqslant 2$, and set

$$A = \{S \subseteq [n] \mid |S| \leqslant d\}.$$

Note that $A = A \bigcirc\!\!\!\!\!\cap A$ and therefore $d = \mathsf{VC\text{-}dim}(A) = \mathsf{VC\text{-}dim}(A \bigcirc\!\!\!\!\!\cap A)$. However $|A| = \binom{n}{\leqslant d} = \Theta(n^d)$, which is not upper bounded by $O(n^{\lceil d/2 \rceil})$. Picking $A = \{S \subseteq [n] \mid |S| \geqslant n - d\}$ shows that $\cup$ behaves similarly to $\cap$ in this context.

The above examples rule out the analog of Theorem 2 for exactly one of $\cup, \cap$. This suggests the following open question:

Question 3. Let $d \leqslant n \in \mathbb{N}$. Suppose $A \subset 2^{[n]}$ satisfies $\mathsf{VC\text{-}dim}(A \bigcirc\!\!\!\!\!\cap A) \leqslant d$ and $\mathsf{VC\text{-}dim}(A \bigcirc\!\!\!\!\!\cup A) \leqslant d$. Is it necessarily the case that $|A| \leqslant n^{d/2+O(1)}$?

Another natural question is whether this phenomenon extends to several applications of the symmetric difference operator, for example:

Question 4. Does there exist an $\epsilon < 1/2$ such that for every $d \leqslant n$ and every $A \subset 2^{[n]}$:

$$\mathsf{VC\text{-}dim}\bigl(A \bigcirc\!\!\!\!\!\triangle A \bigcirc\!\!\!\!\!\triangle A\bigr) \leqslant d \implies |A| \leqslant n^{\epsilon \cdot d + O(1)}?$$

In Section 3 we derive a related statement when $\triangle$ is replaced by addition modulo p for a prime p, and the VC dimension is replaced by the interpolation degree (which is defined in the next section).

1.1 Interpolation degree

Since our proof method is algebraic, it is convenient to view $A \subset 2^{[n]}$ as a subset of the n-dimensional vector space $\mathbb{F}_2^n$ over the field of two elements. In this setting $A \mathbin{\bigcirc\mkern-15mu\triangle} A$ is the *sumset* of A, denoted $A + A$.

Theorem 2 will follow from a stronger statement involving a quantity referred to in some places as the *regularity* (as a special case of Castelnuovo-Mumford regularity from algebraic geometry) [11] and in others as the *interpolation-degree* [9]. We will use the more descriptive interpolation-degree for the rest of this paper. We begin with some preliminary notations and definitions.

Let $A \subset \mathbb{F}_2^n$. It is a basic fact that for each function $f : \mathbb{F}_2^n \mapsto \mathbb{F}_2$ there exists a unique multilinear polynomial $P_f \in \mathbb{F}_2[x_1, \ldots, x_n]$ such that $f(a) = P_f(a)$ for all $a \in \mathbb{F}_2^n$ (existence is via simple interpolation and uniqueness follows from dimension counting). For a partial function $f : A \mapsto \mathbb{F}_2$ there are many (precisely $2^{2^n - |A|}$) multilinear polynomials whose restriction to A computes f. Let $\deg_A(f)$ denote the minimal degree of any polynomial whose restriction to A computes f. We define the *interpolation-degree* of A, denoted $\mathsf{int\text{-}deg}(A)$ to be the maximum of $\deg_A(f)$ taken over all functions $f : A \mapsto \mathbb{F}_2$. In other words, $\mathsf{int\text{-}deg}(A)$ is the smallest d such that any function from A to $\mathbb{F}_2$ can be realized by a polynomial of degree at most d. Clearly, $\mathsf{int\text{-}deg}(A)$ is an integer between 0 and n. It is also not hard to see that, if A is a proper subset of $\mathbb{F}_2^n$ then $\mathsf{int\text{-}deg}(A) < n$. Our interest in $\mathsf{int\text{-}deg}(A)$ comes from the following connection to VC-dimension.

Lemma 5 ([1, 8, 14, 9])**.** *For $A \subset \mathbb{F}_2^n$ we have $\mathsf{int\text{-}deg}(A) \leqslant \mathsf{VC\text{-}dim}(A)$.*

This Lemma, under various formulations, was proved in several works. The formulation that appears here can be found in [9]. For completeness, we next sketch the proof: since the set of all multilinear monomials (also those of degree larger than $\mathsf{VC\text{-}dim}(A)$) span the set of functions $f : A \to \mathbb{F}_2$, it suffices to show that any monomial (when seen as an $A \to \mathbb{F}_2$ function) can be represented a polynomial of degree at most $d = \mathsf{VC\text{-}dim}(A)$. The crucial observation is that if $x_S = \pi_{i \in S} x_i$ is a monomial of degree larger than d, then S is not shattered by A. This means that there is a pattern $v : S \to \{0, 1\}$ that does not appear in any of the vectors in A and therefore

$$\Pi_{i \in S}(x_i + v_i + 1) =_A 0,$$

where "$=_A$" means equality as functions over A. Now, expanding this product and rearranging the equation yields a representation of x_S as sum of monomials $x_{S'}$, where $S' \subset S$, which by induction can also be represented by polynomials of degree at most d.

Theorem 5 reduces Theorem 2 to the following stronger statement that is proved in the next section.

Theorem 6. *Let $d \leqslant n \in \mathbb{N}$, and let $A \subset \mathbb{F}_2^n$ satisfy $|A| > 2\binom{n}{\leqslant \lfloor d/2 \rfloor}$. Then $\mathsf{int\text{-}deg}(A + A) > d$.*

2 Proof of Theorem 6

The main technical tool will be a lemma of Croot-Lev-Pach [3] that was the main ingredient in the recent solution of the cap-set problem [5] and has found many other applications since then (e.g., [7, 15, 4, 6] to name a few).

Lemma 7 (CLP lemma [3]). *Let $P \in \mathbb{F}_q[x_1, \ldots, x_n]$ be a polynomial of degree at most d over any finite field $\mathbb{F}_q$, and let M denote the $q^n \times q^n$ matrix with entries $M_{x,y} = P(x+y)$ for $x, y \in \mathbb{F}_q^n$. Then $\mathsf{rank}(M) \leqslant 2 \cdot m_{\lfloor d/2 \rfloor}(q,n)$, where $m_k(q,n)$ denotes the number of monomials in n variables $x_1, \ldots, x_n$ such that each variable appears with individual degree at most $q-1$ and the total degree of the monomial is at most k.*

Specializing to our setting of $\mathbb{F}_2$ multilinear polynomials, we see that $m_k(2,n) = \binom{n}{\leqslant k}$ and so we conclude:

Corollary 8. *Let $P \in \mathbb{F}_2[x_1, \ldots, x_n]$ be a polynomial of degree at most d and let M be as in Lemma 7. Then $\mathsf{rank}(M) \leqslant 2\binom{n}{\leqslant \lfloor d/2 \rfloor}$.*

We are now ready to prove Theorem 6.

Proof of Theorem 6. Suppose $A \subset \mathbb{F}_2^n$ is such that $|A| \geqslant 2\binom{n}{\leqslant \lfloor d/2 \rfloor}$. Let $f : A + A \mapsto \mathbb{F}_2$ be such that $f(\bar{0}) = 1$, where $\bar{0}$ is the all zero vector in $\mathbb{F}_2^n$, and $f(a) = 0$ for all non-zero $a \in A + A$. It suffices to show that $deg_{A+A}(f) \geqslant \lfloor d/2 \rfloor$ (notice that since $A \neq \emptyset$ it follows that $\bar{0} \in A + A$ and so f is not constantly 0 on $A + A$). Let M be the $2^n \times 2^n$ matrix whose rows and columns are indexed by $\mathbb{F}_2^n$ and with entries $M_{x,y} = f(x+y)$. By our definition of f we have that the sub-matrix of M whose rows and columns are indexed by A is just the $|A| \times |A|$ identity matrix. This implies

$$\mathsf{rank}(M) \geqslant |A|.$$

Let $d_f = \deg_{A+A}(f)$ denote the smallest degree of a polynomial whose restriction to $A+A$ computes f. Applying Corollary 8 we get that

$$\mathsf{rank}(M) \leqslant 2\binom{n}{\leqslant \lfloor d_f/2 \rfloor}.$$

Combining the two inequalities on $\mathsf{rank}(M)$ and using the bound on the size of A we get that

$$2\binom{n}{\leqslant \lfloor d/2 \rfloor} < |A| \leqslant \mathsf{rank}(M) \leqslant 2\binom{n}{\leqslant \lfloor d_f/2 \rfloor},$$

which implies $\lfloor d/2 \rfloor < \lfloor d_f/2 \rfloor$. This means that $d_f > d$ and so $\mathsf{int\text{-}deg}(A + A) > d$. □

3 Generalization to sums modulo p

In this section we observe that our proof can be generalized to give stronger bounds in the case when we take p-fold sums of boolean vectors over $\mathbb{F}_p$. The case proved in the last section corresponds to (two fold) sums modulo 2. For a subset $A \subset \mathbb{F}_p^n$ and a positive integer k, we denote by

$$k \cdot A = \{a_1 + \ldots + a_k \,|\, a_i \in A\}$$

the k-fold sumset of A. To formally define the interpolation degree over $\mathbb{F}_p$ we need to consider, instead of multilinear polynomials, polynomials in which each variable has degree at most $p-1$. We call such polynomials *p-reduced* polynomials. The space of all p-reduced polynomials has dimension p^n and can uniquely represent any function $f : \mathbb{F}_p^n \mapsto \mathbb{F}_p$. The degree of such a function is defined to be the total degree of the unique p-reduced polynomial representing it and can range between 0 and $(p-1)n$. The interpolation degree of a set $A \subset \mathbb{F}_q^n$ is the minimum d such that any function $f : A \mapsto \mathbb{F}_p$ can be represented by a p-reduced polynomial of degree at most d. To avoid confusion we will denote the interpolation degree over $\mathbb{F}_p^n$ as $\mathsf{int\text{-}deg}_p(A)$.

We denote by $\mathcal{M}_d(p,n)$ the set of monomials in n variables $x_1, \ldots, x_n$ in which each variable has degree at most $p-1$ and the total degree is at most d. When $p = 2$ we have the closed formula $|\mathcal{M}_d(2,n)| = \binom{n}{\leqslant d}$. When $p > 2$ the quantity $|\mathcal{M}_d(p,n)|$ is a bit more tricky to compute but is known to satisfy certain asymptotic inequalities (e.g., large deviations [10] showing that $\mathcal{M}_{\delta n}(p,n) \leqslant 2^{\epsilon n}$ with $\epsilon(\delta)$ going to zero with δ).

The following theorem generalizes Theorem 2 when $p > 2$.

Theorem 9. *Let p be any prime number and let $A \subset \{0,1\}^n \subset \mathbb{F}_p^n$ be such that $|A| > p \cdot |\mathcal{M}_{\lfloor d/p \rfloor}(p,n)|$. Then $\mathsf{int\text{-}deg}_p(p \cdot A) > d$.*

The proof of the theorem requires the notion of *slice-rank* of a tensor which was introduced by Tao in his symmetric interpretation of the proof of the cap-set conjecture [16]. By a k-fold tensor of dimension D over a field $\mathbb{F}$ we mean a function T mapping ordered tuples $(j_1, \ldots, j_k) \in [D]^n$ to $\mathbb{F}$. The *slice-rank* of a k-fold tensor T is a the smallest integer R such that T can be written as a sum $T = \sum_{i=1}^R T_i$ such that, for every $i \in [R]$ there is some $j_i \in [k]$ so that $T_i(j_1, \ldots, j_k) = A(j_i)B(j_1, \ldots, j_{i-1}, j_{i+1}, \ldots, j_k)$. In other words, we define the 'rank one' tensors to be those in which the dependence on one of the variables is multiplicative (by a function $A(j_i)$) and the rank of a tensor is the smallest number of rank one tensors needed to describe it. For 2-fold tensors (or matrices) this notion coincides with the usual definition of matrix rank.

The proof of Theorem 9 will follow from a combination of two lemmas regarding slice rank. The first lemma generalizes the Croot-Lev-Pach lemma (and proved in an a similar way).

Lemma 10. *Let $f : \mathbb{F}_p^n \mapsto \mathbb{F}_p$ be of degree d. Then the p-fold p^n dimensional tensor $T : (\mathbb{F}_p^n)^k \mapsto \mathbb{F}_p$ defined by $T(X^1, \ldots, X^p) = f(X^1 + \ldots + X^p)$ has slice rank at most $p \cdot \mathcal{M}_{\lfloor d/p \rfloor}(p,n)$.*

Proof. Consider T as a polynomial in p groups of variables $X^i = (x^i_1, \ldots, x^i_n)$ with $i = 1, 2, \ldots, p$. Since the degree of f is d, the degree of T as a polynomial will also be at most d. This means that, in each monomial of $T(X^1, \ldots, X^p) = f(X^1 + \ldots + X^p)$, the degree of at least one group of variables will be at most $\lfloor d/p \rfloor$. Grouping together monomials according to which group has low degree (if there is more than one group take the one with lowest index) we can represent T as a sum of p tensors, each having rank at most $\mathcal{M}_{\lfloor d/p \rfloor}(p, n)$. This completes the proof. □

The second lemma needed to prove Theorem 9 is due to Tao and shows that the 'diagonal' tensor has full rank.

Lemma 11 ([16])**.** *Let $\delta(j_1, \ldots, j_k) : [D]^k \mapsto \mathbb{F}$ be defined as $\delta(j, j, \ldots, j) = 1$ for all j and is zero otherwise. Then the slice rank of δ is equal to D.*

Proof of Theorem 9. To prove the bound on $\mathsf{int\text{-}deg}_p(p \cdot A)$ we describe a function $f : p \cdot A \mapsto \mathbb{F}_p$ that cannot be represented by a low degree polynomial. We take f to be equal to 1 on the zero vector and zero otherwise. We now consider the tensor $T(X^1, \ldots, X^p) = f(X^1 + \ldots + X^p)$ defined on A^p. Notice that, since $A \subset \{0, 1\}^n$, the sum of p of them is equal to zero iff all p summands are identical. This implies that T is the diagonal tensor δ of Lemma 11 and hence has rank equal to $|A|$. On the other hand, if the degree of f (over $p \cdot A$) is at most d then, by Lemma 10, the tensor T has rank at most $p \cdot \mathcal{M}_{\lfloor d/p \rfloor}(p, n)$. Since we assume that $|A| > p \cdot \mathcal{M}_{\lfloor d/p \rfloor}(p, n)$ this cannot happen and so $\mathsf{int\text{-}deg}_p(pA) > d$. □

Acknowledgements

We thank Stijn Cambie for helpful comments on an earlier draft of this manuscript.

References

[1] L. Babai and P. Frankl. *Linear Algebra Methods in Combinatorics.* University of Chicago, 1992.

[2] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *J. Assoc. Comput. Mach.*, 36(4):929–965, 1989. doi:10.1145/76359.76371.

[3] E. Croot, V. F. Lev, and P. P. Pach. Progression-free sets in $\mathbb{Z}_4^n$ are exponentially small. *Ann. of Math. (2)*, 185(1):331–337, 2017. doi:10.4007/annals.2017.185.1.7.

[4] Z. Dvir and B. Edelman. Matrix rigidity and the Croot-Lev-Pach lemma. Manuscript, 08 2017.

[5] J. S. Ellenberg and D. Gijswijt. On large subsets of $\mathbb{F}_q^n$ with no three-term arithmetic progression. *Ann. of Math. (2)*, 185(1):339–343, 2017. doi:10.4007/annals.2017.185.1.8.

[6] J. Fox and L. M. Lovász. A tight bound for green's arithmetic triangle removal lemma in vector spaces. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '17, pages 1612–1617, Philadelphia, PA, USA, 2017. Society for Industrial and Applied Mathematics. URL http://dl.acm.org/citation.cfm?id=3039686.3039792.

[7] B. Green. Sarkozy's theorem in function fields. *Quarterly Journal of Mathematics*, 68, 2016.

[8] L. Gurvits. Linear algebraic proofs of VC-dimension based inequalities. In *Computational Learning Theory*, pages 238–250. Springer, 1997.

[9] S. Moran and C. Rashtchian. Shattered sets and the hilbert function. In *MFCS*, volume 58 of *LIPIcs*, pages 70:1–70:14. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2016.

[10] F. Rassoul-Agha and T. Seppäläinen. *A course on large deviations with an introduction to Gibbs measures.* American Mathematical Society, 05 2015. ISBN 978-0-8218-7578-0.

[11] Z. Remscrim. The hilbert function, algebraic extractors, and recursive fourier sampling. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 197–208, Oct 2016. doi:10.1109/FOCS.2016.29.

[12] N. Sauer. On the density of families of sets. *J. Comb. Theory, Ser. A*, 13:145–147, 1972. doi:10.1016/0097-3165(72)90019-2.

[13] S. Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific J. Math.*, 41(1):247–261, 1972.

[14] R. Smolensky. Well-known bound for the VC-dimension made easy. *Computational Complexity*, 6(4):299–300, 1997. doi:10.1007/BF01270383. URL http://dx.doi.org/10.1007/BF01270383.

[15] J. Solymosi. The sum of nonsingular matrices is often nonsingular. *Linear Algebra and its Applications*, 552, 01 2018.

[16] T. Tao. A symmetric formulation of the Croot - Lev - Pach - Ellenberg - Gijswijt capset bound. https://terrytao.wordpress.com/2016/05/18/a-symmetric-formulation-of-the-croot-lev-pach-ellenberg, 2016.