Bounded VC-dimension implies a fractional Helly theorem

Jiří Matoušek*

Department of Applied Mathematics and
Institute of Theoretical Computer Science (ITI)
Charles University
Malostranské nám. 25, 118 00 Praha 1
Czech Republic

Abstract

We prove that every set system of bounded VC-dimension has a fractional Helly property. More precisely, if the dual shatter function of a set system $\mathcal{F}$ is bounded by $o(m^k)$, then $\mathcal{F}$ has fractional Helly number k. This means that for every $\alpha > 0$ there exists a $\beta > 0$ such that if $F_1, F_2, \ldots, F_n \in \mathcal{F}$ are sets with $\bigcap_{i \in I} F_i \neq \emptyset$ for at least $\alpha\binom{n}{k}$ sets $I \subseteq \{1, 2, \ldots, n\}$ of size k, then there exists a point common to at least βn of the F_i. This further implies a (p, k)-theorem: for every $\mathcal{F}$ as above and every $p \geq k$ there exists T such that if $\mathcal{G} \subseteq \mathcal{F}$ is a finite subfamily where among every p sets, some k intersect, then $\mathcal{G}$ has a transversal of size T. The assumption about bounded dual shatter function applies, for example, to families of sets in $\mathbf{R}^d$ definable by a bounded number of polynomial inequalities of bounded degree; in this case, we obtain fractional Helly number $d+1$.

1 Introduction

The well-known theorem of Helly states that if $\mathcal{C}$ is a finite family of convex sets in $\mathbf{R}^d$ such that any $d+1$ or fewer of the sets of $\mathcal{F}$ intersect, then $\bigcap \mathcal{C} \neq \emptyset$; we say that the d-dimensional convex sets have *Helly number* $d+1$. A vast number of Helly-type results are known; see e.g. [6].

*Supported by Project LN00A056 of the Ministry of Education of the Czech Republic. Part of this research was done during a visit to Technion in Haifa.

Here we are going to consider fractional Helly-type theorems. The original fractional Helly theorem for convex sets in $\mathbf{R}^d$, asserts the following (here and in the sequel, we use the notation $[n] = \{1, 2, \ldots, n\}$ and $\binom{X}{k}$ for the system of all k-element subsets of X):

Theorem 1 (Katchalski and Liu [8]) *For every $d \geq 1$ and every $\alpha \in (0, 1]$ there exists a $\beta = \beta(d, \alpha) > 0$ with the following property. Let $C_1, \ldots, C_n$ be convex sets in $\mathbf{R}^d$ such that $\bigcap_{i \in I} C_i \neq \emptyset$ for at least $\alpha\binom{n}{d+1}$ index sets $I \in \binom{[n]}{d+1}$. Then there exists a point contained in at least βn of the C_i.*

Let $\mathcal{F}$ be an arbitrary set system. For sets $F_1, F_2, \ldots, F_n \in \mathcal{F}$ and an index set $I \subseteq [n]$, we write F_I for $\bigcap_{i \in I} F_i$. We say that $\mathcal{F}$ has *fractional Helly number* k if for every $\alpha > 0$ there exists a $\beta > 0$ such that if $F_1, F_2, \ldots, F_n \in \mathcal{F}$ are sets such that $F_I \neq \emptyset$ for at least $\alpha\binom{n}{k}$ sets $I \in \binom{[n]}{k}$, then there exists a point common to at least βn of the F_i. We say that $\mathcal{F}$ has the *fractional Helly property* if it has a finite fractional Helly number.

Although the fractional Helly property appears less intuitive than the Helly property, and its conclusion is weaker, it seems *much better behaved and more robust in general than the Helly property.* Here are some examples.

- There is a fractional Helly theorem for hyperplane transversals of convex sets in $\mathbf{R}^d$ [1] although there is no finite Helly number.
- For convex lattice sets in $\mathbf{Z}^d$ (i.e., intersections of convex sets in $\mathbf{R}^d$ with the d-dimensional integer lattice), the Helly number is 2^d, anomalously large, but the fractional Helly number is only $d+1$ [3].
- If a family $\mathcal{F}$ has fractional Helly number k then the family $\{F_1 \cup F_2 : F_1, F_2 \in \mathcal{F}\}$, too, has fractional Helly number k, as is easily checked; for Helly number this, of course, fails badly.

In this paper, we further support the above thesis by adding a wide class of examples with fractional Helly property: all set systems of bounded VC-dimension.

The VC-dimension of a set system $\mathcal{F}$ on a ground set X is the maximum size of a set $A \subseteq X$ that is shattered by $\mathcal{F}$, meaning that $\{A \cap F : F \in \mathcal{F}\} = 2^A$. Examples of set systems with bounded VC-dimension abound in geometry; see, e.g., [9] for a wider background. The *dual shatter function* of $\mathcal{F}$ is a function $\pi^*_{\mathcal{F}}: \mathbf{N} \to \mathbf{N}$, and $\pi^*_{\mathcal{F}}(m)$ is the maximum number of nonempty fields of the Venn diagram of m sets of $\mathcal{F}$. More formally, we call two points $x, y \in X$ *equivalent* w.r.t. sets $F_1, \ldots, F_m$ if $\{i \in [m] : x \in$

$F_i\} = \{i \in [m] : y \in F_i\}$, and $\pi^*_{\mathcal{F}}(m)$ is the maximum possible number of classes of this equivalence over all choices of $F_1, \ldots, F_m \in \mathcal{F}$. The *dual VC-dimension* of $\mathcal{F}$ is the maximum possible number of sets in $\mathcal{F}$ with a complete Venn diagram, i.e., $\max\{k : \pi^*_{\mathcal{F}}(k) = 2^k\}$. It is well known that if the dual VC-dimension is d^* then $\pi^*_{\mathcal{F}}(m) \le \sum_{i=0}^{d^*} \binom{m}{i}$. Moreover, $d^* \le 2^d$, where d is the VC-dimension, and in particular, the VC-dimension is finite iff the dual VC-dimension is.

The dual shatter function seems to be a crucial quantitative parameter of geometric set systems; for example, it is relevant to the performance of range-searching data structures [5], and in many cases it essentially determines the discrepancy of the set system [9]. The following theorem shows a similar phenomenon for the fractional Helly number.

Theorem 2 (Fractional Helly for bounded VC-dimension) *Let $\mathcal{F}$ be a set system whose dual shatter function satisfies $\pi^*_{\mathcal{F}} = o(m^k)$ as $m \to \infty$, where k is a fixed integer (in particular, this holds if the dual VC-dimension of $\mathcal{F}$ is at most $k-1$). Then $\mathcal{F}$ has fractional Helly number k.*

In contrast, bounded VC-dimension does not guarantee any Helly property. A very simple example is the system $\{[n] \setminus \{i\} : i \in [n]\}$, and more complicated examples will be mentioned later.

Let us remark that the original Katchalski–Liu theorem is not a special case of Theorem 2, since convex sets in $\mathbf{R}^d$ have infinite VC-dimension.

A primary example of geometric families of bounded VC-dimension are semialgebraic sets in $\mathbf{R}^d$ of bounded description complexity. We recall that a set $A \subseteq \mathbf{R}^d$ is *semialgebraic* if it can be defined by a Boolean combination of polynomial inequalities; that is, if $A = \{x \in \mathbf{R}^d : \Phi(p_1(x) \ge 0, p_2(x) \ge 0, p_r(x) \ge 0)\}$, where Φ is a Boolean formula and $p_1, \ldots, p_r \in \mathbf{R}[x_1, \ldots, x_d]$ are polynomials. (The definition of a semialgebraic set may also involve quantifiers. But by a well-known result of Tarski, quantifiers can be eliminated, and so each such set has an equivalent quantifier-free definition.) Let us call the number $\max(d, r, D)$, where D is the maximum degree of the p_i, the *description complexity* of A. Standard estimates on the number of sign patterns of real polynomials (due to Oleinik, Petrovskii, Milnor, Thom; see, e.g., [4] for precise results and references) imply that if $\mathcal{F}$ is the family of all semialgebraic sets in $\mathbf{R}^d$ of description complexity at most B, then $\pi^*_{\mathcal{F}}(m) \le Cm^d$ for some $C = C(B)$ and all m. More generally, if $\mathcal{F}$ is as before and $\mathcal{F}' = \{F \cap V : F \in \mathcal{F}\}$, where V is a k-dimensional algebraic variety in $\mathbf{R}^d$, then $\pi^*_{\mathcal{F}'} = O(m^k)$ [4]. We thus have:

Corollary 3 *For every fixed B, the family of all semialgebraic subsets of $\mathbf{R}^d$ of description complexity at most B has fractional Helly number $d+1$. The system of all intersections of sets of this family with a fixed k-dimensional algebraic variety has fractional Helly number $k+1$.*

Here is a nice more concrete example. If $F \subseteq \mathbf{R}^d$ is a semialgebraic set of bounded description complexity, then the set of all j-flats in $\mathbf{R}^d$ intersecting F can be represented a semialgebraic subset of the affine Grassmannian, which is a $(j+1)(d-j)$-dimensional algebraic variety. Consequently, there is a fractional Helly theorem: If $\mathcal{F}$ is the family all semialgebraic subsets of $\mathbf{R}^d$ of description complexity at most B, $F_1, \ldots, F_n \in \mathcal{F}$, and at least $\alpha\binom{n}{k}$ of the k-tuples of the F_i have a j-flat transversal, where $k = (j+1)(d-j)$, then there is a j-flat intersecting at least βn of the F_i. In particular, for line transversals for semialgebraic sets of bounded description complexity in $\mathbf{R}^3$ we obtain fractional Helly number 5.

The method of Alon and Kleitman [2] immediately implies that a family $\mathcal{F}$ as in Theorem 2 satisfies a (p,k)-theorem (for every $p \geq k$):

Theorem 4 ((p,q)-theorem for bounded VC-dimension) *Let $\mathcal{F}$ be a set system with $\pi^*_{\mathcal{F}}(m) = o(m^k)$ for some integer k, and let $p \geq k$. Then there is a constant T such that the following holds for every finite family $\mathcal{G} \subseteq \mathcal{F}$: If $\mathcal{G}$ has the (p,k)-property, meaning that among every p sets of $\mathcal{F}$, some k intersect, then $\tau(\mathcal{G}) \leq T$; that is, there is a T-point set intersecting all sets of $\mathcal{G}$.*

For readers familiar with the Alon–Kleitman proof, we remark that the first step (showing that the fractional packing number of $\mathcal{G}$ is bounded) goes through unchanged based on the fractional Helly property, as well as the second step (LP duality), and the third step (ε-net property, or bounding τ in terms of τ^*) is just the well known theorem of Haussler and Welzl [7] about the existence of ε-nets for systems of bounded VC-dimension.

2 Proof of Theorem 2

Let $\mathcal{F}$ and k be as in Theorem 2, let $\alpha > 0$ be given, and let $F_1, F_2, \ldots, F_n \in \mathcal{F}$ be sets such that $F_I \neq \emptyset$ for at least $\alpha\binom{n}{k}$ k-tuples $I \in \binom{[n]}{k}$. We may assume that n is larger than any given constant, for otherwise, for β sufficiently small, it is enough to have a point in a single F_i.

Using the assumption $\pi^*_{\mathcal{F}} = o(m^k)$, we fix m so that $\pi^*_{\mathcal{F}} < \frac{1}{4}\alpha\binom{m}{k}$, and we set $\beta = \frac{1}{2m}$. Finally, we assume that n is so large that $\beta n \geq m$.

For contradiction, we suppose that no point is common to βn of the F_i. Let us consider an index set $J \in \binom{[n]}{m}$ and a k-tuple $I \in \binom{J}{k}$. Let us call the pair (J, I) *good* if there is a point x with $x \in F_i$ for all $i \in I$ and $x \notin F_j$ for all $j \in J \setminus I$. We bound below the probability that a pair (J, I) chosen uniformly at random is good.

We first choose a random $I \in \binom{[n]}{k}$, and then we choose $m-k$ elements of $J \setminus I$ at random from $[n] \setminus I$. The probability that $F_I \neq \emptyset$ is at least α. If $F_I \neq \emptyset$, we fix one point $x \in F_I$. By the assumption, x is contained in fewer than βn of the F_i, and so the probability that none of the sets F_j with $j \in J \setminus I$ contains x is at least

$$\frac{\binom{\lceil(1-\beta)n\rceil}{m-k}}{\binom{n-k}{m-k}} \geq \prod_{i=0}^{m-k-1} \frac{(1-\beta)n - i}{n-i} \geq \left(\frac{(1-\beta)n - m}{n-m}\right)^m.$$

Since we assumed $m \leq \beta n$ and $\beta = \frac{1}{2m}$, the above expression is at least $(1-2\beta)^m = (1-\frac{1}{m})^m \geq \frac{1}{4}$. Therefore, the probability of a random pair (J, I) being good is at least $\frac{1}{4}\alpha$.

If we choose a random $J \in \binom{[n]}{m}$, the expected number of $I \in \binom{J}{k}$ with (J, I) good is at least $N = \frac{1}{4}\alpha\binom{m}{k}$, and so there exists a J with at least this many I. But this violates the assumption $\pi^*_{\mathcal{F}}(m) < N$, since the sets indexed by J have at least N nonempty fields in their Venn diagram. □

Acknowledgment

I would like to thank Roy Meshulam and Meir Katchalski for inspiring discussions.

References

[1] N. Alon and G. Kalai. Bounding the piercing number. *Discrete Comput. Geom.*, 13:245–256, 1995.

[2] N. Alon and D. Kleitman. Piercing convex sets and the Hadwiger Debrunner (p, q)-problem. *Adv. Math.*, 96(1):103–112, 1992.

[3] I. Bárány and J. Matoušek. A fractional Helly theorem for convex lattice sets. *Adv. Math.*, 2002. In press.

[4] S. Basu, R. Pollack, and M.-F. Roy. On the number of cells defined by a family of polynomials on a variety. *Mathematika*, 43:120–126, 1996.

[5] B. Chazelle and E. Welzl. Quasi-optimal range searching in spaces of finite VC-dimension. *Discrete Comput. Geom.*, 4:467–489, 1989.

[6] J. Eckhoff. Helly, Radon and Carathéodory type theorems. In P. M. Gruber and J. M. Wills, editors, *Handbook of Convex Geometry.* North-Holland, Amsterdam, 1993.

[7] D. Haussler and E. Welzl. Epsilon-nets and simplex range queries. *Discrete Comput. Geom.*, 2:127–151, 1987.

[8] M. Katchalski and A. Liu. A problem of geometry in R^n. *Proc. Amer. Math. Soc.*, 75:284–288, 1979.

[9] J. Matoušek. *Geometric Discrepancy (An Illustrated Guide).* Springer-Verlag, Berlin, 1999.