ON THE NUMBER OF RICH LINES IN HIGH DIMENSIONAL REAL VECTOR SPACES

MÁRTON HABLICSEK AND ZACHARY SCHERR

ABSTRACT. In this short note we use the Polynomial Ham Sandwich Theorem to strengthen a recent result of Dvir and Gopi about the number of rich lines in high dimensional Euclidean spaces. Our result shows that if there are sufficiently many rich lines incident to a set of points then a large fraction of them must be contained in a hyperplane.

1. INTRODUCTION

Let P be a set of points of size n in $\mathbb{R}^d$, and consider a set of lines L in $\mathbb{R}^d$ so that each line in L contains at least r points of P. We investigate the possible size of L.

We begin our discussion with the case of $d = 2$. The celebrated result of Szemerédi and Trotter (which was generalized to the complex plane by Tóth [8] and Zahl [9]) asserts the following.

Theorem 1.1 ([6])**.** *Given P, a set of points in $\mathbb{R}^2$, and L, a set of lines, the number of incidences $I(L,P)$ between L and P satisfies*

$$I(L,P) = O(|L|^{2/3}|P|^{2/3} + |L| + |P|).$$

In our case, each line contains at least r points of P, therefore, $I(L,P) \geq r|L|$. Rearranging the terms we obtain that

$$|L| = O\left(\frac{n^2}{r^3} + \frac{n}{r}\right).$$

This bound is sharp. In a 2-dimensional square grid of n points, for example, each line parallel to one of the sides of the square contains $O(\sqrt{n})$ points, and there are $O(\sqrt{n}) = O\left(\frac{n^2}{(\sqrt{n})^3}\right)$ such lines.

In the higher dimensional case, the d-dimensional grid of n points contains $O\left(\frac{n^2}{r^{d+1}}\right)$ lines for $r = o(n^{1/d})$ [7]. Similar constructions can be given using low dimensional grids as well. Motivated by these examples, Dvir and Gopi conjectured the following.

Conjecture 1.2 ([1])**.** *Let P be a set of n points in $\mathbb{C}^d$ and let L be a set of lines so that each line contains at least r points of P. There are constants*

Date: February 16, 2021.

K and N, dependent only on d, so that if

$$|L| \geq K\left(\frac{n^2}{r^{d+1}} + \frac{n}{r}\right),$$

then there exists $1 < \ell < d$ and a subset $P' \subseteq P$ of size $N\frac{n}{r^{d-\ell}}$ which is contained in an ℓ-dimensional affine subspace.

In their paper [1], Dvir and Gopi show a weaker version of the conjecture.

Theorem 1.3 ([1])**.** *Let P be a set of n points in $\mathbb{C}^d$ and let L be a set of lines so that each line contains at least r points of P. There are constants K and N, dependent only on d, so that if*

$$|L| \geq K\frac{n^2}{r^d}$$

then there exists a subset of P of size $N\frac{n}{r^{d-2}}$ contained in a $(d-1)$-dimensional hyperplane.

Their proof involves a clever use of design matrices in order to show that almost all the lines lie in a low degree hypersurface (the degree needs to be less than r). In our paper, we prove a stronger version of Theorem 1.3 but over $\mathbb{R}$ rather than over $\mathbb{C}$. The strategy of our proof is similar to that of Dvir and Gopi, except working over $\mathbb{R}$ allows us to use the Polynomial Ham Sandwich Theorem (Theorem 2.3) in place of design matrices.

2. Main results

Our main result shows that if there are too many r-rich lines then most of the lines must lie in a low degree hypersurface.

Theorem 2.1. *Let P be a set of n points in $\mathbb{R}^d$ and let L be a set of lines so that each line contains at least r points of P. There is a constant K, dependent only on d, so that if*

$$|L| \geq K\frac{n^2}{r^{d+1}},$$

then there exists a hypersurface of degree at most $\frac{r}{4}$ containing at least $4\frac{n^2}{r^{d+1}}$ lines of L.

Remark. One can interpret the theorem above as follows. If a set of points is such that there exist a lot of non-generic large subsets, then a large fraction of the points must be non-generic. In our case we know that there are $K\frac{n^2}{r^{d+1}}$ non-generic subsets of size r, and we deduce that a large fraction of points lie in a low degree hypersurface.

As an easy consequence of Theorem 2.1 we obtain a better bound over $\mathbb{R}$ than the bound in Theorem 1.3.

Theorem 2.2. *Let P be a set of n points in $\mathbb{R}^d$ and let L be a set of lines so that each line contains at least r points of P. There are constants K and N, dependent only on d, so that if*

$$|L| \geq K\frac{n^2}{r^{d+1}},$$

then there exists a hyperplane containing $N\frac{n}{r^{d-1}}$ points of P.

The main technique in our proof is the Polynomial Ham Sandwich Theorem which we state below.

Theorem 2.3 (Polynomial Ham Sandwich)**.** *Let S be a finite set of points in $\mathbb{R}^d$, and let $m \geq 1$. Then there exists a non-trivial polynomial f of degree m and a decomposition of $\{x \in \mathbb{R}^d : f(x) \neq 0\}$ into at most $O(m^d)$ cells each of which is an open set with boundary in $\{x \in \mathbb{R}^d : f(x) = 0\}$, and each of which contains at most $O\left(\frac{|S|}{m^d}\right)$ points of S.*

This poweful tool was invented by Guth and Katz in [2] to give a nearly complete solution to the Erdős distinct distance problem and has been applied, for instance, to give a new proof of the Szemerédi-Trotter theorem, the Pach-Sharir theorem [5] (see [3] for more details) and some variants of the joints problem [4].

We remark that the Polynomial Ham Sandwich Theorem relies on the topology of $\mathbb{R}$, and thus our proof only works over $\mathbb{R}$. On the other hand, we believe that Theorem 2.1 holds over any prime field $\mathbb{F}_p$ and over $\mathbb{C}$ as well, and hence it would be nice to see a proof of Theorem 2.1 which does not use the Polynomial Ham Sandwich Theorem.

Question 2.4. Let P be a set of n points in k^d, where k is either a prime field $\mathbb{F}_p$ or the field of complex numbers. Let L be a set of lines so that each line contains at least r points of P. Is there a constant K, dependent only on d, so that if

$$|L| \geq K\frac{n^2}{r^{d+1}},$$

then there exists a hypersurface of degree at most $\frac{r}{4}$ containing at least $4\frac{n^2}{r^{d+1}}$ lines of L?

We remark that recently in [10], Zahl proved a slightly weaker version of Theorem 2.2 over $\mathbb{C}$ using a version of the Polynomial Ham Sandwich Theorem over $\mathbb{C}$ (see [8] or [9]).

3. Proof of the main theorems

In this section we prove Theorems 2.1 and 2.2. We begin with the proof of Theorem 2.1 which we restate below.

Theorem 3.1. *Let P be a set of n points in $\mathbb{R}^d$, and let L be a set of lines so that each line contains at least r points of P. There is a constant K, dependent only on d, so that if*

$$|L| \geq K\frac{n^2}{r^{d+1}},$$

then there exists a hypersurface of degree at most $\frac{r}{4}$ containing at least $4\frac{n^2}{r^{d+1}}$ lines of L.

Proof. Assume that $|L| = K\frac{n^2}{r^{d+1}}$ for a large constant K (which will be chosen in the end of the proof) and fix a positive integer m in the range $\frac{r}{8} < m < \frac{r}{4}$ (the interesting case of the theorem is when r is large). Using the Polynomial Ham Sandwich Theorem (Theorem 2.3), we can find a polynomial f of degree m partitioning $\mathbb{R}^d$ into the zero locus of f as well as $M = O(m^d)$ open cells

$$\mathbb{R}^d = \{x : f(x) = 0\} \cup C_1 \cup C_2 \cup \ldots \cup C_M$$

so that each cell contains at most $O\left(\frac{n}{m^d}\right)$ points of P and has boundary in the zero set of f. We denote $P_i := C_i \cap P$.

Let

$$L_{cell} = \{\ell \in L : \exists i \text{ with } |\ell \cap P_i| \geq 2\}.$$

Since the zero locus of f forms the boundary of the union of the cells, Bézout's theorem guarantees that every line in $\mathbb{R}^d$ intersects at most m cells. If $\ell \in L \backslash L_{cell}$, then $|\ell \cap P_i| \leq 1$ for each i, so in particular

$$\left|\bigcup_{i=1}^{M} \ell \cap P_i\right| = \sum_{i=1}^{M} |\ell \cap P_i| \leq m < \frac{r}{2}. \tag{3.2}$$

By assumption, every line in L contains r points of P so lines in $L \backslash L_{cell}$ must contain at least $\frac{r}{2} > m$ in the zero locus of f. We can again invoke Bézout to conclude that such a line is necessarily contained in the zero locus of f. Since what we are after is a lower bound on the number of lines in L which are contained in the zero locus of f, this discussion shows that it suffices to give an upper bound on the size of L_{cell}.

To do so, we take advantage of the fact that every line $\ell \in L_{cell}$ has the property that $|\ell \cap P_i| \geq 2$ for some i. The total number of lines, counted with multiplicity, in $\mathbb{R}^d$ which intersect some P_i in at least two points is

$$\sum_{i=1}^{M} \binom{|P_i|}{2} \tag{3.3}$$

where each such line ℓ is counted with multiplicity

$$k_\ell := \sum_{i=1}^{M} \binom{|\ell \cap P_i|}{2} = \frac{1}{2}\left(\sum_{i=1}^{M} |\ell \cap P_i|^2 - \sum_{i=1}^{M} |\ell \cap P_i|\right).$$

We have already observed that a line not contained in the zero locus of f can only intersect at most m cells. If

$$a_i = \begin{cases} 0, & \ell \cap P_i = \emptyset \\ 1, & \text{otherwise} \end{cases}$$

then this observation, combined with the the Cauchy-Schwarz inequality, gives

$$\begin{aligned} \left(\sum_{i=1}^{M} |\ell \cap P_i| \right)^2 &= \left(\sum_{i=1}^{M} a_i \, |\ell \cap P_i| \right)^2 \\ &\leq \sum_{i=1}^{M} a_i^2 \cdot \sum_{i=1}^{M} |\ell \cap P_i|^2 \\ &\leq m \sum_{i=1}^{M} |\ell \cap P_i|^2 . \end{aligned}$$

Therefore we get a lower bound

$$k_\ell \geq \frac{1}{2} \left(\frac{\left(\sum_{i=1}^{M} |\ell \cap P_i| \right)^2}{m} - \sum_{i=1}^{M} |\ell \cap P_i| \right). \tag{3.4}$$

If $\ell \in L_{cell}$, then (3.2) guarantees that

$$\sum_{i=1}^{M} |\ell \cap P_i| \geq \frac{r}{2}.$$

For such ℓ, (3.4) becomes

$$k_\ell \geq \frac{r}{4} \left(\frac{r}{2m} - 1 \right) = \frac{r^2 - 2mr}{8m}.$$

Since $m < \frac{r}{4}$, it follows that

$$k_\ell \geq \frac{r^2 - r^2/2}{8m} \geq \frac{r^2}{16m}$$

when r is large enough.

Every $\ell \in L_{cell}$ is counted with multiplicity k_ℓ in (3.3). Thus

$$\sum_{i=1}^{M} \binom{|P_i|}{2} \geq \sum_{\ell \in L_{cell}} k_\ell \geq |L_{cell}| \frac{r^2}{16m}. \tag{3.5}$$

We know that $M = O(m^d)$ and $|P_i| = O\left(\frac{n}{m^d}\right)$, so we can rewrite (3.5) as

$$|L_{cell}| = \frac{16m}{r^2} O\left(m^d \frac{n^2}{m^{2d}} \right) = \frac{1}{r^2} O\left(\frac{n^2}{m^{d-1}} \right).$$

Since $\frac{r}{8} < m$, this last equation becomes

$$|L_{cell}| = O\left(\frac{n^2}{r^{d+1}}\right).$$

The set of lines in L which are contained in the zero locus of f has size

$$|L| - |L_{cell}| \geq K\frac{n^2}{r^{d+1}} - |L_{cell}|,$$

and so we can choose K large enough so as to ensure that this last quantity is bounded below by $4\frac{n^2}{r^{d+1}}$. □

As an easy corollary we prove Theorem 2.2. In order to do so, we use the following standard graph theoretic lemma which can also be found in the paper of Dvir and Gopi.

Lemma 3.6 (Lemma 2.8, [1])**.** *Let $G = (A \sqcup B, E)$ be a bipartite graph with a non-empty edge set $E \subset A \times B$. Then there exist non-empty subsets $A' \subset A$ and $B' \subset B$ such that if we consider the induced subgraph $G' = (A' \sqcup B', E')$, then*

- *The minimum degree in A' is at least $\frac{|E|}{4|A|}$,*
- *The minimum degree in B' is at least $\frac{|E|}{4|B|}$,*
- *$|E'| \geq |E|/2$.*

We are ready to prove the theorem.

Theorem 3.7. *Let P be a set of n points in $\mathbb{R}^d$, and let L be a set of lines so that each line contains at least r points of P. There are constants K and N, dependent only on d, so that if*

$$|L| \geq K\frac{n^2}{r^{d+1}},$$

then there exists a hyperplane containing $N\frac{n}{r^{d-1}}$ points of P.

Proof. We may use the previous theorem to conclude that if K is large enough then there exists at least $4\frac{n^2}{r^{d+1}}$ lines contained in a degree $m < \frac{r}{4}$ hypersurface. Let us denote the set of these lines by L_Z and the set of points of P on the lines of L_Z by P_Z. Each line of L_Z is still r-riched, thus the total number of incidences between L_Z and P_Z satisfies

$$I(L_Z, P_Z) \geq r\,|L_Z| = 4\frac{n^2}{r^d}.$$

By Lemma 3.6 we may, after removing lines and points, therefore assume without loss of generality that each point of P_Z is incident to at least $\frac{n}{r^d}$ lines in L_Z.

Let g be a non-zero polynomial of minimum degree vanishing on L_Z. We know that f vanishes on L_Z, therefore the degree of g is less than r.

Now, we call a point $p \in P_Z$ a *joint* if the directions of the lines in L_Z incident to p span $\mathbb{R}^d$. If every $p \in P_Z$ is a joint, then surely the gradient of g must vanish on all of P_Z. Pick a component of the gradient which is non-zero on the vanishing locus of g. This component vanishes on all the points in P_Z and is of degree less than r. Therefore, by Bézout's theorem, this component vanishes on all the lines in L_Z as well, but the component is of smaller degree than of g which is a contradiction.

Thus there must be a point $p \in P_Z$ which is not a joint, whence all the lines of L_Z going through p lie in the same hyperplane. We know that there are $\frac{n}{r^d}$ lines going through p, and on each such line there are $r-1$ other points, implying that there are at least

$$(r-1)\frac{n}{r^d} + 1 = \Omega\left(\frac{n}{r^{d-1}}\right)$$

points in one hyperplane. □

References

[1] Dvir, Z., Gopi, S., *On the number of rich lines in truly high dimensional sets*, 31st International Symposium on Computational Geometry (SoCG 2015), Leibniz International Proceedings in Informatics (LIPIcs), 34, 584–598, 2015

[2] Guth, L., Katz, N., *On the Erdős distinct distances problem in the plane*, Annals of Math., Volume 181, 155–190, 2015

[3] Kaplan, H., Matoušek, J., Sharir, M., *Simple Proofs of Classical Theorems in Discrete Geometry via the Guth–Katz Polynomial Partitioning Technique*, Discrete & Computational Geometry, (3), 499–517, 2012

[4] Iliopoulou, M., *Incidence bounds on multijoints and generic joints*, Discrete Comput. Geom. 54 (2), 481–512, 2015

[5] Pach, J., Sharir, M., *On the number of incidences between points and curves*, Combinat. Probab. Comput., (7), 121–127, 1998

[6] Szemerédi, E., Trotter, W.T., *Extremal problems in discrete geometry*, Combinatorica, (3), 381–392, 1983

[7] Solymosi, J., Vu, VH., *Distinct distances in high dimensional homogeneous sets*, Cont. Math., (342) 259–268, 2004

[8] Toth, Cs., *The Szemerédi-Trotter theorem in the complex plane*, preprint, 2003

[9] Zahl, J., *A Szemerédi-Trotter type theorem in* $\mathbb{R}^4$, Discrete. Comput. Geom. 54 (3), 513–572, 2012

[10] Zahl, J., *A note on rich lines in truly high dimensional sets*, FoM, Sigma. 4 (e2), 1–13, 2016

Department of Mathematics, Leiden University, Snellius Building, Niels Bohrweg 1, 2333 CA Leiden, Netherlands

Email address: `hablicsekhm@math.leidenuniv.nl`

Department of Mathematics and Computer Science, Fisher Hall Susquehanna University, 514 University Ave., Selinsgrove, Pa. 17870 USA

Email address: `scherr@susqu.edu`