On the number of rich lines in high dimensional real vector spaces

Márton Hablicsek and Zachary Scherr

December 23, 2014

In this paper we strengthen a recent result of Dvir and Gopi about the number of rich lines in high dimensional spaces. 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 asserts the following.

**Theorem 1** ([3]). Given a set of lines, $L$ and a set of points, $P$ in $\mathbb{R}^2$, 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

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

The bound is sharp, consider a 2-dimensional square grid of $n$ points. 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 $r = o(n^{1/d})$ ([4]). Similar constructions can be given using low dimensional grids as well. Motivated by these examples, Dvir and Gopi conjectured the following.

**Conjecture 2** ([1]). Let $L$ be a set of lines and $P$ be a set of points in $\mathbb{C}^d$, so that each line in $L$ contains at least $r$ points of $P$. Let us denote the number of points by $n$. 

1
Then, there exist constants $K$ and $M$ depending only on $d$ so that if

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

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

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

**Theorem 3 (1).** Let $L$ be a set of lines and $P$ be a set of points in $\mathbb{C}^d$, so that each line in $L$ contains at least $r$ points of $P$. Let us denote the number of points by $n$. There exist constants $K$ and $M$ depending only on $d$ so that if the number of $r$-rich lines satisfies

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

then there exists a subset of size $M \frac{n}{r^{d-1}}$ contained in a $d - 1$-dimensional hyperplane.

Their proof involves a clever use of design matrices in order to show that all the lines lie in a low degree hypersurface (the degree needs to be less than $r$). We follow the same strategy in our paper. We show that the Polynomial Ham Sandwich theorem ([2]) ensures that - not all, but - almost all the lines lie in a low degree hypersurface, and the rest follows from the work of Dvir and Gopi. Our result is the following.

**Theorem 4.** Let $L$ be a set of lines and $P$ be a set of points in $\mathbb{R}^d$, so that each line in $L$ contains at least $r$ points of $P$. Let us denote the number of points by $n$. There exist constants $K$ and $M$ so that if

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

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

We remark that our proof only works over the real numbers as opposed to Theorem 3. We believe that the same statement holds over $\mathbb{C}$ as well.

We are ready to prove Theorem 4.

**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), we will show that there exists a hyperplane in $\mathbb{R}^d$ containing $\frac{n}{r^{d-1}}$ points of $P$.

The following lemma is unchanged from the paper of Dvir and Gopi.
Lemma 5 (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 sets \( A' \subset A \) 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 \( \frac{|E|}{4|B'|} \),
- \( |E'| \geq |E|/2 \).

This lemma implies that without loss of generality we can assume that each point of \( P \) is incident to \( \frac{n}{rd} \) lines, since the number of incidences between the points satisfies

\[
I(L, P) \geq r|L| \geq rK \frac{n^2}{rd+1}.
\]

Using the Polynomial Ham Sandwich theorem (see [2]), we can find a polynomial \( f \) of degree \( m < \frac{r}{4} < \frac{r}{2} \) partitioning \( \mathbb{R}^d \) into the zero locus of \( f \) as well as \( M = O(m^d) \) cells so that each cell contains at most \( O \left( \frac{m}{m^d} \right) \) points of \( P \). We order the cells, and we denote by \( P_i \) the subset of points of \( P \) contained in cell \( i \).

Consider those lines in \( L \) which contain at least \( \frac{r}{2m} > 2 \) points of \( P \) in one of the cells. We denote the set of these lines by \( L_{cell} \). The lines which are not in \( L_{cell} \) contain fewer than \( \frac{r}{2m} \) points in each cell. Each line can intersect \( m \) cells so these lines contain fewer than \( \frac{r}{2} \) points in the union of the cells. Therefore they contain at least \( \frac{r}{2} \) points of the zero locus of \( f \). By Bezout’s theorem, we obtain that the lines are contained in the zero locus of \( f \). Furthermore, any line which does not lie in the zero locus of \( f \), contains at least \( \frac{r}{2m} \) points in one of the cells.

Next we estimate the size of \( L_{cell} \). Each pair of points defines a line. The points in the cells thus define

\[
\sum_{i=1}^{M} \binom{|P_i|}{2}
\]

lines with multiplicities. Each line \( l \in L \) is calculated with

\[
\sum_{i=1}^{M} \binom{|l \cap P_i|}{2}
\]

multiplicity. Each line in \( L_{cell} \) intersects at most \( m \) cells, therefore by Cauchy-Schwarz

\[
\sum_{i=1}^{M} \binom{|l \cap P_i|}{2} = \sum_{i: |l \cap P_i| \neq 0} \frac{|l \cap P_i|^2 - |l \cap P_i|}{2} \geq \left( \sum_{i: |l \cap P_i| \neq 0} \frac{|l \cap P_i|}{2m} \right)^2 - \sum_{i: |l \cap P_i| \neq 0} \frac{|l \cap P_i|}{2}.
\]
Since
\[
\sum_{i:|l \cap P_i| \neq 0} |l \cap P_i| \geq \frac{r}{2},
\]
and \(m < \frac{r}{4}\), we have
\[
\sum_{i=1}^{M} \binom{|l \cap P_i|}{2} \geq \frac{r^2}{16m}.
\]

Summarizing the above discussion, since each line in \(L_{cell}\) contains at least two points in some cell, we obtain
\[
\sum_{i=1}^{M} \binom{|P_i|}{2} \geq |L_{cell}| \frac{r^2}{16m}.
\]

The number of cells is bounded above by \(O(m^d)\) and each \(|P_i|\) by \(O\left(\frac{n^2}{m^d}\right)\), therefore the left-hand side is bounded above by
\[
O \left( m^d \frac{n^2}{m^{2d}} \right) = O \left( \frac{n^2}{m^d} \right).
\]

If, \(m\) is on the order of \(r\), we obtain that \(|L_{cell}| = O \left( \frac{n^2}{r^{d+1}} \right)\). As a consequence choosing \(K\) to be large enough we can guarantee that the number of lines inside the zero locus is at least \(4\frac{n^2}{r^{d+1}}\). 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 by Lemma 5 we can assume that each point of \(P_Z\) is incident to at least \(n^d\) lines. 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 \(r\).

There are two cases. First, all the points of \(P_Z\) are ‘joints’, meaning that the directions of the lines of \(L_Z\) incident to each point of \(P_Z\) span \(\mathbb{R}^d\). In this case the gradient of \(g\) vanishes on every point of \(P_Z\). Pick a component of the gradient which is non-zero on the vanishing locus of \(g\). That component vanishes on all the points, it is of degree less than \(r\), therefore, by Bezout’s theorem, the component vanishes on all the lines as well. But, the component is of smaller degree than of \(g\) which is a contradiction.

Second, there is a point \(p\) of \(P_Z\) which is not a joint. Thus, all the lines of \(L_Z\) going through \(p\) lie in the same hyperplane. There are \(\frac{n^2}{r^{d+1}}\) lines going through \(p\), on each such line there are \(r-1\) other points which implies that there are \(O(\frac{n^2}{r^{d+1}})\) points in one hyperplane.

This concludes the proof. \(\square\)
References

[1] Dvir, Z., Gopi, S., *On the number of rich lines in truly high dimensional sets*, preprint, 2014

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

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

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