1. Order of Magnitude notation

In whatever follows, for quantities \( A, B \geq 0 \), the notation
\[
A \lesssim B
\]
means that \( A = O(B) \), i.e. that \( A \leq C B \), for some explicit constant \( C > 0 \) that is independent of \( A \) and \( B \). This is meaningful if the quantities \( A, B \) are allowed to vary: in that case, \( A \lesssim B \) implies that \( A \) is at most \( B \), up to some multiplicative constant - in other words, \( A \) is at most \( B \) when it comes to order of magnitude.

The notation
\[
A \gtrsim B
\]
means that \( B \lesssim A \), i.e. that \( A \geq C' B \), for some constant \( C' > 0 \) that is independent of \( A \) and \( B \).

Finally,
\[
A \sim B
\]
means that \( A \lesssim B \) and \( A \gtrsim B \), i.e. that \( C_1 B \leq A \leq C_2 B \), for constants \( C_1, C_2 > 0 \) that are independent of \( A \) and \( B \). Therefore, \( A \sim B \) means that \( A \) and \( B \) have the same order of magnitude.

Example 1.1. For all \( n \in \mathbb{N} \), it holds that \( \binom{n}{2} \sim n^2 \).

Proof. Since \( \binom{n}{2} = \frac{n(n-1)}{2} \), it holds that
\[
\frac{1}{8} n^2 \leq \binom{n}{2} \leq \frac{1}{2} n^2.
\]

Since the multiplicative constants \( \frac{1}{8} \) and \( \frac{1}{2} \) above are independent of \( n \in \mathbb{N} \), it follows that \( \binom{n}{2} \) has the same order of magnitude as \( n^2 \), i.e. \( \binom{n}{2} \sim n^2 \).

\[\square\]

2. The Crossing Number Inequality.

In this section, for any graph \( G \), \( e \) denotes in principle the number of edges of \( G \) and \( v \) the number of vertices of \( G \).

Definition 2.1. Let \( G \) be a graph. A planar drawing of \( G \) is any drawing of \( G \) on the plane.

In a planar drawing of a graph \( G \), a crossing is any pair of the form \((p, e_1, e_2)\), where \( e_1, e_2 \) are edges of \( G \) that intersect in the drawing at the point \( p \in \mathbb{R}^2 \) that is not a common vertex of \( e_1, e_2 \).

The crossing number of \( G \) is the smallest number of crossings over all planar drawings of \( G \), and is denoted by \( \text{cr}(G) \).
A planar graph (i.e., a graph without crossings) is sparse: it has few edges relative to the number of vertices. More precisely, it has been shown that the following holds.

**Proposition 2.2.** Let $G$ be a simple, planar graph. If $v \geq 3$, then

$$e \leq 3v - 6.$$  

Technically, the above proposition has so far been proved for simple, planar, connected graphs only; however, it trivially generalises to all simple planar graphs, by applying Proposition 2.2 to each connected component.

**Corollary 2.3.** If a simple graph $G$ with at least 3 vertices has more than $3v - 5$ edges, then any planar drawing of $G$ has at least 1 crossing.

However, can we do better than that? The answer is yes, and is given by the crossing number inequality, which quantifies the so far heuristic statement that “many edges induce many crossings”.

**Theorem 2.4. (Crossing number inequality)** Let $G$ be a simple graph, with $e > 7v$. Then,

$$cr(G) \geq \frac{e^3}{29v^2}.$$  

Before proving the crossing number inequality, we perform a reality check.

- **If true, is the crossing number inequality useful/meaningful?**

Yes: the crossing number inequality implies that

$$cr(G) \geq \frac{(7v)^3}{29v^2} \geq \frac{7^3}{29} v = 11.8275862069v;$$

in particular, $cr(G) \geq 12$. This is much better than merely knowing that $cr(G) \geq 1$, which is true by Corollary 2.3 (note that this Corollary can be applied as $G$ has at least 3 vertices, given its large number of edges and the fact that it is a simple, connected graph).

- **Does the crossing number inequality make sense? Could it be that the desired lower bound is too large to be true?**

A trivial upper bound for $cr(G)$ is

$$cr(G) \leq \binom{v}{2};$$

this follows by the fact that, in any planar drawing of $G$ where the edges are drawn as straight line segments, any two edges cross at most once.

So, for the crossing number inequality to have a chance to hold, it should be true that

$$\frac{e^3}{29v^2} \leq \binom{v}{2}.$$  

Observe that it is trivial that $\frac{e^3}{v^2} \leq e^2$, as this is equivalent to $e \lesssim v^2$, something which holds as $e \leq \binom{v}{2}$ (the graph is simple, so it has no multiple edges). So our desired lower bound on $cr(G)$ is indeed at most $\binom{v}{2}$ up to order of magnitude. Thinking along these lines and using the precise bounds for quantities of the form $\binom{n}{2}$ described in Example
1.1 it is easy to prove (1).

Proof of crossing number inequality. The proof is completed in two steps. In Step 1, a simpler inequality is proved for \( \text{cr}(G) \). In Step 2, this simpler inequality is improved to the stronger desired one, via a probabilistic method.

Step 1: The simpler inequality
\[
(2) \quad \text{cr}(G) \geq e - 3v
\]
holds.

Idea: All crossings of \( G \) can be eliminated by deleting \( \leq \text{cr}(G) \) edges of \( G \). The remaining graph is planar, and therefore has few edges. Since we started with many edges, we must have deleted a lot of edges along the way. So, \( \text{cr}(G) \) must be large.

Precisely: Consider a planar drawing of \( G \) with exactly \( \text{cr}(G) \) crossings. Each crossing can be eliminated by deleting one of the two edges responsible for it. Therefore, all \( \text{cr}(G) \) can be eliminated by deleting \( \leq \text{cr}(G) \) edges of \( G \); the remaining subgraph \( G' \) of \( G \) is planar (i.e., has no crossings).

Since \( G' \) is planar, simple and with \( v_{G'} = v \geq 3 \), it follows that \( G' \) has few edges; in particular, by Proposition 2.2,
\[
e_{G'} \leq 3v_{G'} = 3v.
\]
Since \( e_{G'} \geq e - \text{cr}(G) \), it follows that
\[
e - \text{cr}(G) \leq 3v,
\]
which implies that \( \text{cr}(G) \geq e - 3v \), completing the proof of Step 1.

Step 2: Here, the weaker inequality (2) is strengthened to the desired
\[
(3) \quad \text{cr}(G) \geq \frac{e^3}{29v^2}.
\]
This is achieved by using the fact that (2) holds for all simple graphs \( G \) drawn on the plane (rather than merely our fixed one).

Idea: In Step 1, the weaker inequality (2) was established by deleting edges of \( G \) in an appropriate manner. It is therefore conceivable that (3) could be established by the more brutal process of deleting vertices of \( G \). In particular, it would be very promising if a subgraph \( G' \) of \( G \) existed, with
\[
(4) \quad \text{cr}(G') \text{ considerably smaller than } \text{cr}(G),
\]
and yet the number of edges \( e' \) and the number of vertices \( v' \) of \( G' \) satisfying
\[
(5) \quad e' - 3v' \text{ "almost equal to" } e - 3v.
\]
Then, (2) for \( G' \) implies that
\[
\text{cr}(G) \gg \text{cr}(G') \geq e' - 3v' \text{ " } = \text{" } e - 3v;
\]
in particular,
\[
\text{cr}(G) \gg e - 3v,
\]
which is stronger than the weak inequality \( \text{cr}(G) \geq e - 3v \). Therefore, if such a subgraph \( G' \) of \( G \) exists, then it leads to a strengthening of (2). However, constructing such a subgraph can be very challenging. Indeed, while it is easy to get (4) by deleting enough
vertices (and therefore edges) of $G$ (fewer edges lead to fewer crossings), one may think that too many vertices/edges need to be deleted for (5) to simultaneously hold. In other words, control on the quantity $e' - 3v'$ is very limited, as it really depends on the specific graph $G$ we are working with.

To overcome our inability to explicitly construct a subgraph $G'$ that satisfies conditions (4) and (5) above, we resort to the probabilistic method.

The probabilistic method is based on the idea that, if a desired configuration exists (in this case, a good subgraph $G'$), then perhaps many desired configurations exist (i.e., many good subgraphs $G'$), and therefore the average random configuration (i.e., the average random subgraph of $G$) may have the desired properties.

Precisely: For some $p \in (0, 1]$ that will be specified later, keep each vertex of $G$ with probability $p$, and delete it with probability $1 - p$. A random subgraph $G'$ of $G$ is created with this process.

- Since each vertex of $G$ survives with probability $p$, $G'$ has on average $p \cdot v$ vertices. That is, $\mathbb{E}(v') = p \cdot v$.

- An edge of $G$ survives if and only if both its endpoints survive, i.e. with probability $p^2$. Therefore, $G'$ has on average $p^2 \cdot e$ edges. That is, $\mathbb{E}(e') = p^2 \cdot e$.

- Observe that, as long as $G$ is drawn in a way that the minimum number of crossings occurs, any crossing in $G$ is the result of two overlapping edges, where the edges are defined by 4 (rather than 3) distinct vertices. Indeed, if the two overlapping edges were defined by in total 3 vertices, then there would exist a drawing of $G$ on the plane where these edges do not cross, as the following picture demonstrates.
So, a crossing of $G$ survives if and only if all these 4 vertices survive, i.e. with probability $p^4$. Therefore, $G'$ has on average $p^4 \cdot \text{cr}(G)$ crossings. That is, $E(\text{cr}(G')) = p^4 \cdot \text{cr}(G)$.

Note that, as desired, the average random subgraph $G'$ that the above probabilistic process yields has crossing number $p^4 \cdot \text{cr}(G)$, considerably smaller than $\text{cr}(G)$. On the other hand, $e' - 3v'$ is on average $p^2 e - 3pv$, which is “not as far” from $e - 3v$; at least not as far as $p^4(e - 3v)$ is.

Any subgraph $G'$ of $G$ is a simple graph, so

$$\text{cr}(G') \geq e' - 3v'$$

for all such $G'$. In particular, the same holds for the average subgraph $G'$ we get from the probabilistic process above:

$$E(\text{cr}(G')) \geq E(e' - 3v') = E(e') - 3E(v'),$$

i.e.

$$p^4 \text{cr}(G) \geq p^2 e - 3pv;$$

observe that this inequality only depends on $G$, not any particular $G'$.

It has thus been shown that, for all $p \in (0, 1]$,

$$\text{cr}(G) \geq \frac{e}{p^2} - \frac{3v}{p^3}. \tag{6}$$

There exists some $p \in (0, 1]$ for which the lower bound $\frac{e}{p^2} - \frac{3v}{p^3}$ becomes maximal, or at least close to maximal; once found and replaced in (6), (6) becomes (1).

The optimal $p$ will not be pursued here; however, observe that any $p$ such that the quantities $\frac{e}{p^2}$ and $\frac{3v}{p^3}$ are comparable already yields good results. For instance, for $p$ such that $\frac{e}{p^2} = \frac{4v}{p^3}$ (i.e., for $p = \frac{4v}{e}$, which is in $(0, 1]$ iff $e \geq 4v$, rather than the $e > 7v$ condition of the theorem), (6) becomes

$$\text{cr}(G) \geq \frac{v}{p^3} = \frac{1}{64} \frac{e^3}{v^2},$$

a bound almost as good as the one desired; and, in any case, still demonstrating the fact that

$$\text{cr}(G) \gtrsim \frac{e^3}{v^2}.$$

□

3. The Szemerédi–Trotter theorem

This section is dedicated to the Szemerédi–Trotter theorem, a central result in incidence geometry, which controls the number of incidences between points and lines in Euclidean space.

**Theorem 3.1. (Szemerédi–Trotter, version 1)** Let $\mathcal{L}$ be a finite set of $L$ lines in $\mathbb{R}^2$. For any $k \in \mathbb{N}$, let $S_k$ be the set of points in $\mathbb{R}^2$ with the property that each lies in $\sim k$ lines in $\mathcal{L}$. Then, for all $k \geq 2$,

$$\# S_k \preceq \frac{L}{k} + \frac{L^2}{k^3}. \tag{7}$$
The exact meaning of “$\sim k$” in the above statement is irrelevant - the theorem holds independently of the implicit constants hiding behind the $\sim$ notation (of course by adjusting the implicit constants in (22)). This is perhaps not clear yet, so, for the time being, let us focus on 

$$S_k := \{ x \in \mathbb{R}^2 : x \text{ lies in } \geq k \text{ and } < 2k \text{ lines of } \mathcal{L} \}.$$ 

**Observations on the $S$–$T$ theorem:**

- For $k = 1$, the $S$–$T$ bound does not hold. Indeed, one cannot expect any bound on $S_1$ (the set of points with exactly one line in $\mathcal{L}$ through each) that depends only on $L$. For instance, consider $\mathcal{L}$ consisting of a single line $\ell$ ($L = 1$). All points in $\ell$ are in $S_1$, and are infinitely many, so $\#S_1$ is not bounded above.

- According to the value of $k$, either $L_k \lesssim L_2$ or $L_2 k^3$ is dominant on the RHS of (22). More precisely,

$$\frac{L}{k} \lesssim \frac{L^2}{k^3} \text{ iff } k \lesssim L^{1/2}.$$ 

Therefore, $S$–$T$ states that

$$\#S_k \lesssim \begin{cases} \frac{L^2}{k^3}, & \text{if } k \lesssim L^{1/2} \\ \frac{L}{k}, & \text{if } k \gtrsim L^{1/2}. \end{cases}$$

- Observe that each element of $S_k$ is a point in $\mathbb{R}^2$ where at least 2 of the lines in $\mathcal{L}$ meet (since $k \geq 2$). Therefore,

$$\#S_k \lesssim \binom{L}{2} \leq L^2.$$ 

Therefore, for $k \sim 1$, this trivial bound agrees with the one given by $S$–$T$ (at least in terms of order of magnitude). However, for larger $k$ the bound $\frac{L^2}{k^3}$ given by $S$–$T$ becomes tighter. (Note for instance that, for $k \sim L^{1/2}$, $S$–$T$ implies that $\#S_k \lesssim L^{1/2}$, a bound considerably smaller than $L^2$ in terms of order of magnitude.)

- It has been explained that, for $k \gtrsim L^{1/2}$, it holds that $\#S_k \lesssim \frac{L}{k}$; that is,

$$\#S_k \lesssim \frac{L}{k},$$ 

i.e.

$$\sum_{x \in S_k} \#\{\text{lines in } \mathcal{L} \text{ through } x\} \lesssim L.$$

*Morally,* this means that, by adding over all $x \in S_k$ the number of lines through $x$, not too many lines are being double-counted, and therefore the configuration of points and lines looks like a union of essentially disjoint bushes of lines through the points of $S_k$, each bush consisting of $\sim k$ lines.

Of course, there could be other, different situations where (8) holds. It could be, for instance, that too few lines in $\mathcal{L}$ pass through the points of $S_k$, and therefore double-counting lines in (8) is not sufficient to make the LHS as large as $L$. However, it is the “disjoint bushes” scenario that is perhaps the most interesting one, since that is when the bound becomes tighter - and, even so, this “interesting”
scenario is not all that interesting. The significance of S–T lies primarily in the $\frac{L^2}{k}$ bound, which holds for $k \lesssim L^{1/2}$, and which in fact offers a different perspective on the problem.

- The bound $\#S_k \lesssim L^2$ is trivial, as at least two lines meet at any point in $S_k$. However, the tighter bound

  \[ \#S_k \lesssim \frac{L^2}{k^2} \]

is also trivial. The reason is that $\#S_k k^2$ counts the number of pairs of distinct lines in $L$ that pass through points in $S_k$, while $L^2$ counts all pairs of distinct lines in $L$. More precisely,

\[
\begin{align*}
\#S_k k^2 &\sim \sum_{x \in S_k} \#\{\text{lines in } L \text{ through } x\}^2 \\
&\sim \sum_{x \in S_k} \left( \#\{\text{lines in } L \text{ through } x\} \right)^2 \\
&\sim \sum_{x \in S_k} \#\{\text{pairs of distinct lines in } L \text{ through } x\}.
\end{align*}
\]

Since any two distinct lines can meet at at most one point, the sum above counts different pairs for different $x$. Thus, by the addition principle,

\[ \#S_k k^2 \lesssim \#\{\text{all pairs of distinct lines in } L\} \sim \left( \frac{L}{k} \right)^2 \sim L^2. \]

It is this last observation that shows that the Szemerédi–Trotter theorem is highly non-trivial. The fact that, for $k \lesssim L^{1/2}$, $\#S_k k^2$ is not equal to the number of pairs of all lines in $L$, but in fact is at most a $\frac{1}{k}$ portion of all these pairs, means that a lot of pairs of lines meet at points in $S_\lambda$, for $\lambda \neq k$. In other words, for sufficiently large $k \lesssim L^{1/2}$, the lines in $L$ do not only meet on $S_k$; they are responsible for points of different multiplicity than $k$. The picture is richer that what one would perhaps initially guess.
Proof of the Szemerédi–Trotter theorem (version 1) – by Székely, 1993. This is a very fast proof of the S–T theorem, using the crossing number inequality.

Fix $\mathcal{L}$ and $k$ as in the statement of the S–T theorem. Consider the graph $G$ on the plane, with set of vertices $S_k$, and an edge connecting two vertices $u,v \in S_k$ iff $u$ and $v$ are consecutive points in $S_k$ on a line in $\mathcal{L}$.

Idea: If the edges are many, then the crossing number inequality $\text{cr}(G) \gtrsim \frac{e^3}{v^2}$ holds, and hopefully each of the three quantities in it can be expressed in terms of $\# S_k$, $k$ and $L$, the three quantities whose relationship we wish to understand. Given that the crossing number inequality is non-trivial, perhaps it will lead to a non-trivial estimate for $\# S_k$ relative to $L$.

Precisely: For the graph $G$, it holds that

- $\text{cr}(G) \leq \left( \frac{L}{2} \right)^2 \leq L^2$. It is this larger quantity $L^2$ that will be used in the crossing number inequality in place of $\text{cr}(G)$. Note that, in principle, not much is lost by this replacement. Indeed, if the S–T theorem is believed to be true, then, for $k \lesssim L^{1/2}$, most pairs of lines in $\mathcal{L}$ do not meet on $S_k$, and therefore contribute to $\text{cr}(G)$; in other words, we should expect that $\text{cr}(G) \sim L^2$, at least when no two lines in $\mathcal{L}$ are parallel to each other.

- $v = \# S_k$.

- $e \geq k\# S_k - L$. Indeed, since $G$ is a simple graph, it holds that
  
  \[ 2e = \sum_{x \in S_k} \text{deg}(x). \]

Now, for each $x \in S_k$, $\text{deg}(x)$ is equal to the number of neighbours of $x$ in $G$. If $x$ has a neighbour on either side of it on each line in $\mathcal{L}$ through $x$, then $\text{deg}(x) = 2\# \{ \text{lines in } \mathcal{L} \text{ through } x \} \sim 2k$. However, it could be that $x$ is a single point of $S_k$ on certain lines $\ell \in \mathcal{L}$ through it, in which case $x$ has 0 neighbours on each such $\ell$, or $x$ is an “endpoint” on certain lines $\ell \in \mathcal{L}$ through it, but not a single point of $S_k$ on $\ell$, in which case $x$ has 1 neighbour on each $\ell$. In other words, counting 2 neighbours of $x$ on each line in $\mathcal{L}$ through $x$ is too much; the correct formula for $\text{deg}(x)$ is

\[ \text{deg}(x) = 2\# \{ \ell \in \mathcal{L} \text{ through } x \} - 2\# \{ \ell \in \mathcal{L} \text{ through } x \text{ with } x \text{ a single point of } S_k \text{ on } \ell \} - \# \{ \ell \in \mathcal{L} \text{ through } x \text{ with } x \text{ an “endpoint” of } S_k \text{ on } \ell, \text{ but } x \text{ not a single point of } S_k \text{ on } \ell \}. \]

Adding over all $x \in S_k$, one obtains

\[ 2e = \sum_{x \in S_k} 2\# \{ \ell \in \mathcal{L} \text{ through } x \} - 2\# \{ \ell \in \mathcal{L} : \ell \text{ contains a single point of } S_k \} - 2\# \{ \ell \in \mathcal{L} : \ell \text{ contains 2 “endpoints” in } S_k \} \]
\[ = \sum_{x \in S_k} \# \{ \ell \in \mathcal{L} \text{ through } x \} - 2L \]
\[ \geq 2k\# S_k - 2L. \]

Therefore, $e \geq k\# S_k - L$, as desired.
– Case 1: It holds that \( k \# S_k - L > 7 \# S_k \). In this case, \( e > 7v \), so the crossing number inequality

\[
\text{cr}(G) \gtrsim \frac{e^3}{v^2}
\]

holds. Therefore, by the above estimates on \( \text{cr}(G) \), \( v \) and \( e \), one obtains

\[
L^2 \gtrsim \frac{(k \# S_k - L)^3}{\# S_k^2}.
\]

1i) If \( L \leq \frac{k \# S_k}{2} \) (in which case \( \# S_k \geq \frac{L}{2} \), thus the “uninteresting” S–T bound does not hold), then \( k \# S_k \) is dominant in the numerator above, leading to

\[
L^2 \gtrsim \frac{(\frac{k \# S_k}{2})^3}{\# S_k^2} \sim k^3 \# S_k.
\]

Rearranging, one obtains

\[
\# S_k \lesssim \frac{L^2}{k^3};
\]

the non-trivial bound of S–T holds, in this case where the uninteresting one does not.

1ii) If \( L \geq \frac{k \# S_k}{2} \), then \( \# S_k \lesssim \frac{L}{k} \), thus the uninteresting S–T bound holds.

– Case 2: It holds that \( k \# S_k - L \leq 7 \# S_k \).

2i) If \( k > 7 \), then \( k - 7 \sim k \). Thus, the above inequality \((k - 7) \# S_k \leq L\) becomes \( k \# S_k \lesssim L \), i.e.

\[
\# S_k \lesssim \frac{L}{k}.
\]

2ii) If \( k \leq 7 \), then \( k \sim 1 \), so \( L^2 \sim \frac{L^2}{k^2} \). So, the trivial bound \( \# S_k \lesssim \left( \frac{L}{2} \right) \leq L^2 \) becomes

\[
\# S_k \lesssim \frac{L^2}{k^3}.
\]

Observe that all possible cases have been exhausted, and each case led to either \( \# S_k \lesssim \frac{L^2}{k^2} \) or \( \# S_k \lesssim \frac{L}{k} \). Therefore, in all cases,

\[
\# S_k \lesssim \frac{L}{k} + \frac{L^2}{k^3}.
\]

Observe now that, due to our definition of \( S_k \), the following (seemingly stronger) version of the Szemerédi–Trotter theorem is easily deduced.

**Theorem 3.2. (Szemerédi–Trotter, version 2)** Let \( \mathcal{L} \) be a finite set of \( L \) lines in \( \mathbb{R}^2 \). For any \( k \in \mathbb{N} \), let \( S_{\geq k} \) be the set of points in \( \mathbb{R}^2 \) with the property that each point in \( S_{\geq k} \) lies in at least \( k \) lines in \( \mathcal{L} \). Then, for all \( k \geq 2 \),

\[
\# S_{\geq k} \lesssim \frac{L}{k} + \frac{L^2}{k^3}.
\]

**Proof.** It holds that

\[
S_{\geq k} = S_k \cup S_{2k} \cup S_{4k} \cup S_{8k} \cup S_{16k} \cup \ldots
\]
Therefore, by the addition principle,
\[ \#S_{\geq k} = \#S_k + \#S_{2k} + \#S_{4k} + \#S_{8k} + \#S_{16k} + \ldots \]
Applying the Szemerédi–Trotter theorem (version 1) to each of the sets above, one obtains
\[
\#S_{\geq k} \lesssim \left( \frac{L}{k} + \frac{L}{2k} + \frac{L}{4k} + \frac{L}{8k} + \frac{L}{16k} + \ldots \right) \\
+ \left( \frac{L^2}{k^3} + \frac{L^2}{(2k)^3} + \frac{L^2}{(4k)^3} + \frac{L^2}{(8k)^3} + \frac{L^2}{(16k)^3} + \ldots \right) \\
\sim \frac{L}{k} \left( 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \ldots \right) + \frac{L^2}{k^3} \left( 1 + \frac{1}{2^3} + \frac{1}{4^3} + \frac{1}{8^3} + \frac{1}{16^3} + \ldots \right) \\
\sim \frac{L}{k} + \frac{L^2}{k^3},
\]
because the above series converge to positive constants independent of \( L \) and \( k \).

□

**Observation.** If \( S_k \) had been defined as the set of points in \( \mathbb{R}^2 \) where exactly \( k \) lines in \( \mathcal{L} \) meet, then the above proof would not have worked, because the geometric series would have given way to another, much larger sum. Indeed, one would merely be able to use that
\[
S_{\geq k} = S_k \sqcup S_{k+1} \sqcup S_{k+2} \sqcup S_{k+3} \sqcup \ldots \sqcup S_L
\]
(clearly, no more than \( L \) lines in \( \mathcal{L} \) can pass through a given point in \( \mathbb{R}^2 \)), which would lead to
\[
\#S_{\geq k} \lesssim \left( \frac{L}{k} + \frac{L}{k+1} + \frac{L}{k+2} + \frac{L}{k+3} + \ldots + \frac{L}{L} \right) \\
+ \left( \frac{L^2}{k^3} + \frac{L^2}{(k+1)^3} + \frac{L^2}{(k+2)^3} + \frac{L^2}{(k+3)^3} + \ldots + \frac{L^2}{L^3} \right) \\
\sim L \left( \frac{1}{k} + \frac{1}{k+1} + \frac{1}{k+2} + \frac{1}{k+3} + \ldots + \frac{1}{L} \right) + \\
+ L^2 \left( \frac{1}{k^3} + \frac{1}{(k+1)^3} + \frac{1}{(k+2)^3} + \frac{1}{(k+3)^3} + \ldots + \frac{1}{L^3} \right) \\
\sim L \cdot \int_k^L \frac{1}{x} + L^2 \int_k^L \frac{1}{x^3} \\
\sim L \cdot (\log L - \log k) + L^2 \left( \frac{1}{k^2} - \frac{1}{L^2} \right),
\]
a bound nowhere near the desired one.

4. Point-line incidences

**Definition 4.1.** Let \( \mathcal{L} \) be a set of lines and \( \mathcal{P} \) a set of points in some vector space \( V \). We say that the pair \((p, \ell)\), for \( p \in \mathcal{P} \) and \( \ell \in \mathcal{L} \), is an incidence between \( \mathcal{P} \) and \( \mathcal{L} \) if \( p \in \ell \).

Denote by \( I(\mathcal{P}, \mathcal{L}) \) the number of all incidences between \( \mathcal{P} \) and \( \mathcal{L} \).

By its definition,
\[
I(\mathcal{P}, \mathcal{L}) = \# \{ (p, \ell) \in \mathcal{P} \times \mathcal{L} : p \in \ell \}. 
\]
Therefore,
\[ I(P, L) = \sum_{p \in P} \#\{\text{lines in } L \text{ through } p\} \]
while also
\[ I(P, L) = \sum_{\ell \in L} \#\{\text{points in } P \text{ on } \ell\} \]
It is now clear that the Szemerédi–Trotter theorem counts incidences between points and lines. Indeed, it asserts that
\[ \#S_k \lesssim L + \frac{L^2}{k^2}, \]
i.e. that
\[ I(S_k, L) \lesssim L + \frac{L^2}{k^2}, \]
since
\[ \#S_k \sim \sum_{x \in S_k} \#\{\text{lines in } L \text{ through } x\} = I(S_k, L). \]
In other words, S–T ensures that, under conditions, there are few incidences between points and lines.

Being able to control incidences between points and lines is crucial in a variety of problems. An important example is the Erdős distinct distance on the plane, conjecturing that \( n \) points on the plane define many distinct distances, only solved in 2010; this problem can be transformed into a point-line incidences problem in \( \mathbb{R}^3 \).

The next section is dedicated to a powerful new tool for counting point-line incidences.

5. The Polynomial Method

The idea behind the polynomial method in incidence geometry is the following:

Suppose we have a set \( L \) of lines in \( \mathbb{F}^n \) and a set \( P \) of points where certain of our lines intersect with particular properties. Suppose we want to count the points in \( P \). If we find a polynomial \( p \in \mathbb{F}[x_1, \ldots, x_n] \) whose zero set \( Z_p \) passes through all points in \( P \), then we know that each line \( \ell \) that is not fully in \( Z_p \) intersects \( Z_p \) at most \( \deg p \) times, as \( p_\ell \) is a polynomial in one variable of degree \( \leq \deg p \), and therefore has at most \( \deg p \) roots. Since \( P \) lies in \( Z_p \), it follows that each line \( \ell \) that is not fully in \( Z_p \) contains \( \leq \deg p \) points of \( P \). If \( \deg p \) is small enough, this gives us good control on the number of incidences between \( P \) and the lines in \( L \) that do not fully lie in \( Z_p \). Thus, what remains is to control the incidences between \( P \) and the lines in \( L \) that are inside \( Z_p \). This provides much more structural information that we would have if we were ignoring the existence of a \( Z_p \) with the above properties.

Let us now explain the above.

**Definition 5.1.** Let \( \mathbb{F} \) be a field and \( n \in \mathbb{Z}_{\geq 1} \).

Denote by \( \mathbb{F}[x_1, \ldots, x_n] \) the set of polynomials in \( n \) variables with coefficients in \( \mathbb{F} \).

Let \( p \in \mathbb{F}[x_1, \ldots, x_n] \). We define the zero set of \( p \) in \( \mathbb{F}^n \) to be the set
\[ Z_p := \{ x \in \mathbb{F}^n : p(x) = 0 \}. \]
We will fix our attention on counting point-line incidences in \( \mathbb{R}^n \), therefore we will be interested in polynomials in \( p \in \mathbb{R}[x_1, \ldots, x_n] \). For such \( p \), one may morally think of \( Z_p \) as an \((n-1)\)-dimensional surface in \( \mathbb{R}^n \) (at least when \( p \) is not the zero polynomial, whose zero set is the whole of \( \mathbb{R}^n \)). In general the situation is a bit more complicated; here are some examples.

**Examples.**

- For any \( p \in \mathbb{R}[x] \) polynomial of one variable, \( Z_p \) is the set of roots of \( p \) in \( \mathbb{R} \), therefore \( Z_p \) is a finite set of \( \leq \deg p \) points.

- The line \( \ell = \{(x, y) \in \mathbb{R}^2 : y = ax + b\} \) in \( \mathbb{R}^2 \) is the zero set of the polynomial \( p(x, y) = y - ax - b \in \mathbb{R}[x, y] \).

- The parabola \( P = \{(x, y) \in \mathbb{R}^2 : y = x^2\} \) is the zero set of the polynomial \( p(x, y) = y - x^2 \in \mathbb{R}[x, y] \).

- The hyperplane \( H = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : a_1x_1 + \ldots + a_nx_n = b\} \) is the zero set of the polynomial \( p(x_1, \ldots, x_n) = a_1x_1 + \ldots + a_nx_n - b \in \mathbb{R}[x_1, \ldots, x_n] \).

- The \((n-1)\)-dimensional sphere in \( \mathbb{R}^n \) centered at \( y = (y_1, \ldots, y_n) \) with radius \( r \) is the zero set of the polynomial \( p(x_1, \ldots, x_n) = (x_1 - y_1)^2 + \ldots + (x_n - y_n)^2 - r^2 \in \mathbb{R}[x_1, \ldots, x_n] \). Observe that zero sets of polynomials are not necessarily graphs of functions \( f : \mathbb{R}^{n-1} \rightarrow \mathbb{R} \).

- The union of zero sets of finitely many polynomials is a zero set of a polynomial. Indeed, for any \( p_1, \ldots, p_k \in \mathbb{R}[x_1, \ldots, x_n] \), it holds that

  \[ Z_{p_1} \cup \ldots \cup Z_{p_k} = Z_{p_1 \cdot p_2 \cdots p_k}. \]

  In particular, this implies that zero sets of polynomials are not necessarily smooth manifolds. For instance, the zero set of the polynomial \( (x - y)(x + y) \in \mathbb{R}[x, y] \) is the union of the bisectors of the first and third quadrant in \( \mathbb{R}^2 \), which is clearly not a smooth manifold.

- The intersection of zero sets of finitely many polynomials in \( \mathbb{R}[x_1, \ldots, x_n] \) is the zero set of a polynomial in \( \mathbb{R}[x_1, \ldots, x_n] \). Indeed, for any \( p_1, \ldots, p_k \in \mathbb{R}[x_1, \ldots, x_n] \), it holds that

  \[ Z_{p_1} \cap \ldots \cap Z_{p_k} = Z_{p_1^2 + \ldots + p_k^2}. \]

  In particular, this implies that zero sets of polynomials in \( \mathbb{R}[x_1, \ldots, x_n] \) are not necessarily \((n-1)\)-dimensional objects. For instance, the zero set of any \( p \in \mathbb{R}[x_1, \ldots, x_k] \) in \( \mathbb{R}^k \), where \( k < n \), is also the zero set, in \( \mathbb{R}^n \), of the polynomial \( \tilde{p}(x_1, \ldots, x_k)^2 + x_{k+1}^2 + x_{k+2}^2 + \ldots + x_n^2 \). The original zero set is at most \( k \)-dimensional, and yet it is the zero set, in \( \mathbb{R}^n \), of a polynomial in \( n \) variables.

- Combining the two bullet points above, it follows that zero sets of polynomials can be unions of objects of different dimensions.

**Definition 5.2.** Let \( \mathbb{F} \), \( n \in \mathbb{Z}_{\geq 1} \). Any polynomial \( p \in \mathbb{F}[x_1, \ldots, x_n] \) of degree at most \( d \), for some \( d \in \mathbb{Z}_{\geq 1} \), is of the form

\[
p(x_1, \ldots, x_n) = \sum_{(a_1, \ldots, a_n) \in \mathbb{Z}_{\geq 0}^n : a_1 + \ldots + a_n \leq d} c_{a_1, \ldots, a_n} x_1^{a_1} x_2^{a_2} \ldots x_n^{a_n},
\]

for coefficients \( c_{a_1, \ldots, a_n} \in \mathbb{F} \).
Any line $\ell$ in $\mathbb{F}^n$ is of the form

$$\ell := \{ \vec{b} + t\vec{v} : t \in \mathbb{F} \},$$

where $\vec{b} = (b_1, \ldots, b_n) \in \mathbb{F}^n$ and $\vec{v} = (v_1, \ldots, v_n) \in \mathbb{F}^n$; $\vec{v}$ is called the direction of $\ell$, while $\vec{b}$ is an (any) element of $\mathbb{F}^n$ that $\ell$ passes through.

For any polynomial $p$ and line $\ell$ as above, define $p_{|\ell}$ to be the polynomial in $\mathbb{F}[t]$ (of one variable) with

$$p_{|\ell}(t) := \sum_{(a_1, \ldots, a_n) \in \mathbb{Z}_n^+ : a_1 + \cdots + a_n \leq d} c_{a_1, \ldots, a_n} (b_1 + tv_1)^{a_1} (b_2 + tv_2)^{a_2} \cdots (b_n + tv_n)^{a_n}.$$ 

Observe that $\deg p_{|\ell} \leq \deg p$.

**Definition 5.3.** Let $\mathbb{F}$, $n \in \mathbb{Z}_{\geq 1}$. We say that $p \in \mathbb{F}[x_1, \ldots, x_n]$ is identically zero if all of its coefficients are zero.

**Observation.** It holds that $p \in \mathbb{R}[x_1, \ldots, x_n]$ is identically zero iff $p(x) = 0$ for all $x \in \mathbb{R}^n$. However, in fields of finite characteristic this is not necessarily true. For instance, the polynomial $p \in \mathbb{Z}_p[x]$ with $p(x) = x^p - x$ is such that $p(x) = 0$ for all $x \in \mathbb{Z}_p$, but $p$ is not identically zero, since it has a non-zero coefficient.

We will primarily focus on $\mathbb{R}^n$ from now on, therefore the above distinction will never be an issue.

One of the main reasons why the polynomial method is so useful when it comes to counting point-line incidences is that lines interact very well with zero sets of polynomials: zero sets of polynomials "infect" lines. More precisely, the following holds.

**Proposition 5.4.** Let $\ell$ be a line in $\mathbb{F}^n$, and $p \in \mathbb{F}[x_1, \ldots, x_n]$. Then, either

$$p_{|\ell}(t) \equiv 0 \ (\text{in which case } \ell \subseteq \mathbb{Z}_p),$$

or

$\ell$ meets $\mathbb{Z}_p$ at most $\deg p$ times (i.e., $p$ vanishes at at most $\deg p$ points of $\ell$).

**Proof.** It holds that

$$\ell := \{ b + tv : t \in \mathbb{F} \}$$

for some fixed $b, v \in \mathbb{F}^n$. Suppose that $\ell$ meets $\mathbb{Z}_p$ more than $\deg p$ times. This means that there exist (pairwise distinct) $y_1, \ldots, y_m \in \ell$, with $m > \deg p$, such that

$$p(y_i) = 0 \text{ for all } i = 1, \ldots, m.$$

Now, for each $i = 1, \ldots, m$, it holds that $y_i = b + t_iv$, for some $t_i \in \mathbb{F}$; all these $t_i$ are pairwise distinct. Therefore, $p(b + t_iv) = 0$ for all $i = 1, \ldots, m$; by the fact that $p_{|\ell}$ was defined so that $p_{|\ell}(t) = p(b + tv)$ for all $t \in \mathbb{F}$, the above means exactly that

$$p_{|\ell}(t_i) = 0 \text{ for all } i = 1, \ldots, m.$$

Therefore, the polynomial $p_{|\ell} \in \mathbb{F}[t]$ (of one variable) has $> \deg p$ roots in $\mathbb{F}$. Since its degree is $\geq \deg p$, $p_{|\ell}$ is identically 0. 

$\square$
In particular, in the case of fields $\mathbb{F}$ of infinite size, such as $\mathbb{R}$ or $\mathbb{C}$, it holds that, for any line $\ell$ in $\mathbb{F}^n$, $Z_p$ either contains finitely many points of the line $\ell$, or fully engulfs the line $\ell$ (which has infinitely many points) - one of these extremal situations holds, no middle ground exists.

For instance, line segments are not zero sets of polynomials in $\mathbb{R}^n$; if a zero set of a polynomial contains a line segment, then it has to contain the whole line this segment lives in. This demonstrates how rigid zero sets of polynomials are; it is hard for a manifold to be the zero set of a polynomial.

By the above, it indeed makes sense, when counting incidences between a set $L$ of lines and a set $P$ of points, to try and find some low-degree polynomial $p$ that vanishes on $P$ (i.e., whose zero set $Z_p$ passes through the points in $P$). This will at least ensure that every $\ell \in L$ that does not fully lie in $Z_p$ contains $\leq \deg p$ points of $P$, and thus contributes few incidences. In fact, the smaller $\deg p$ is, the fewer incidences these lines contribute.

However, we cannot hope that, no matter what set $P$ of points we start with, we can find a polynomial $p$ of degree as low as we wish with $Z_p$ containing $P$. For instance, we cannot in general hope that we can find such $p$ of degree 1: this would mean that $p$ is linear, therefore $Z_p$ is a hyperplane, and it is quite unlikely that the points in $P$ all lie on a hyperplane.

So, in principle, we need to extend our search to polynomials $p$ of larger degree to achieve that $Z_p$ contains $P$. Thankfully, the following theorem asserts that, even so, we can achieve what we need using $p$ of not too large degree.

[This idea first appeared in number theory in work of Thue (1917), and was since used in number theory and computer science. It was in 2008 that it first appeared in incidence geometry, when Dvir solved the Kakeya problem in finite fields, inducing a revolution in the area.]

**Theorem 5.5.** Let $n \geq 1$ and $\mathbb{F}$ be a field. Then, for any finite set $P$ of points in $\mathbb{F}^n$, there exists a non-zero $p \in \mathbb{F}[x_1, \ldots, x_n]$, with $\deg p \lesssim \left(\#P\right)^{1/n}$, that vanishes on $P$.

In the above, $\deg p \lesssim \left(\#P\right)^{1/n}$ means that $\deg p \leq c_n(\#P)^{1/n}$ for some constant $c_n > 0$ that depends only on the dimension $n$; not on the set $P$ of points or the field $\mathbb{F}$.

Before proving the theorem, observe that it is not absurd that smaller degree should be required as the dimension increases. Indeed:

- For $n = 1$, any polynomial $p$ that vanishes on $P$ (a subset of $\mathbb{F}$ in this 1-dimensional case) has degree at least $\#P$. Indeed, if $P = \{y_1, y_2, \ldots, y_m\}$, then $p(x)$ must have $x - y_1$, $x - y_2, \ldots, x - y_m$ as factors (when seen as a polynomial with coefficients in the algebraic closure of $\mathbb{F}$), so $\deg p \geq m = \#P$. Since the polynomial $p(x) = (x - y_1) \cdots (x - y_m)$ has degree $\#P$ and vanishes on $P$, it is the lowest degree polynomial in $\mathbb{F}[x]$ with this property.

- For $n = 2$, unlike in the 1-dim case, lines are zero sets of non-zero polynomials in $\mathbb{F}^2$. So:
  
  - Consider $P$ line on a line $\ell$ inside $\mathbb{F}^2$ (the picture is essentially 1-dimensional). In this case, unlike for $n = 1$, the exists $p \in \mathbb{F}[x_1, x_2]$ (crucially, 2 variables are allowed now) of degree only 1 that vanishes on $P$: a linear polynomial whose zero set is $\ell$. 


• Consider $m$ lines $\ell_1, \ell_2, \ldots, \ell_m$ in $\mathbb{P}^2$ in general position, i.e. each two intersecting at a different point in $\mathbb{P}^2$. The union of these lines is the zero set of the polynomial $p := p_1 p_2 \cdots p_m$, where each $p_\ell$ is a linear polynomial that vanishes on $\ell_i$; $\deg p = m$. However, the set $\mathcal{P}$ of pairwise intersections of the above lines has size $\binom{m}{2} \approx m^2$, and $p$ vanishes on $\mathcal{P}$. That is, $p$ has degree $\approx (\#\mathcal{P})^{1/2}$, and yet vanishes on $\mathcal{P}$.

In other words, in the example above the degree of $p$ can be lowered from $\#\mathcal{P}$ to $(\#\mathcal{P})^{1/2}$ because only $(\#\mathcal{P})^{1/2}$ lines are needed to define $\mathcal{P}$: few lines can intersect at many points in $\mathbb{P}^2$. And, as mentioned above, unlike in the 1-dim case, lines are zero sets of non-zero polynomials in $\mathbb{P}^2$.

As the dimension $n$ gets higher, there is more room in $\mathbb{F}^n$ for more interesting objects to be zero sets of non-zero polynomials, and for them to intersect each other more.

**Proof of Theorem 5.5.** Fix $n$, $\mathbb{F}$ and $\mathcal{P} = \{y_1, \ldots, y_m\}$. We are looking for $d \geq 1$ and for coefficients $c_{a_1, \ldots, a_n} \in \mathbb{F}$, not all simultaneously 0, so that the polynomial

$$p(x_1, \ldots, x_n) = \sum_{(a_1, \ldots, a_n) \in \mathbb{Z}_{\geq 0}^n : a_1 + \cdots + a_n \leq d} c_{a_1, \ldots, a_n} x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$$

vanishes on $\mathcal{P}$, i.e. so that

$$p(y_1) = 0, \ldots, p(y_m) = 0.$$  

(11)

Observe that, for each $i = 1, \ldots, m$, the equation $p(y_i) = 0$ is a linear equation with unknowns the coefficients $c_{a_1, \ldots, a_n}$, i.e. with $\binom{d+n}{n}$ unknowns. Therefore, (11) is a system of $m$ linear equations with unknowns the $\binom{d+n}{n}$ coefficients. If the unknowns are more than the equations, i.e. if $d \geq 1$ is such that

$$\binom{d+n}{n} > m,$$

then (11) has a non-trivial solution; i.e., there exist coefficients not all simultaneously 0 (and therefore a non-zero $p \in \mathbb{F}[x_1, \ldots, x_n]$ of degree at most $d$) for which (11) holds.

Now, $\binom{d+n}{n} \geq \frac{d^n}{n!}$ (in order of magnitude notation, $\binom{d+n}{n} \gtrsim d^n$). So, for any $d \geq 1$ such that $\frac{d^n}{n!} > m$ (equivalently, for any $d > (n!)^{1/n} m^{1/n}$ or $d \gtrsim m^{1/n}$ in o.o.m. notation), there exists non-zero $p \in \mathbb{F}[x_1, \ldots, x_n]$ of degree at most $d$, that vanishes on $\mathcal{P}$.

In particular, the above holds for the smallest $d \geq 1$ such that $\frac{d^n}{n!} > m$, i.e. for $d = \lceil (n!)^{1/n} m^{1/n} \rceil + 1$ (for $d \sim m^{1/n}$ in o.o.m. notation). In other words, for $d = \lceil (n!)^{1/n} m^{1/n} \rceil + 1$, there exists a non-zero $p \in \mathbb{F}[x_1, \ldots, x_n]$, of degree at most

$$d \leq (n!)^{1/n} m^{1/n} + 1 \leq 2(n!)^{1/n} m^{1/n} \lesssim m^{1/n} = \#\mathcal{P}^{1/n}$$

that vanishes on $\mathcal{P}$.

\[ \square \]

6. Better understanding of zero sets of polynomials - $\nabla p$

**Definition 6.1.** For any $p \in \mathbb{R}[x_1, \ldots, x_n]$, define

$$\nabla p := \left( \frac{\partial p}{\partial x_1}, \ldots, \frac{\partial p}{\partial x_n} \right).$$
Simply understanding whether $\nabla p$ vanishes or not at a point $x_0 \in Z_p$ provides a lot of structural information on $Z_p$. In particular, the following holds for points of $Z_p$ where $\nabla p$ does not vanish.

**Proposition 6.2.** Let $n \geq 2$ and $p \in \mathbb{R}[x_1, \ldots, x_n]$. If $x_0 \in Z_p$ is such that
$$\nabla p(x_0) \neq 0,$$
then $Z_p$ is a smooth $(n-1)$-dimensional manifold locally around $x_0$. In particular, the tangent space $T_{Z_p}(x_0)$ to $Z_p$ at $x_0$ exists, and is the hyperplane through $x_0$ that is normal (i.e. perpendicular) to the vector $\nabla p(x_0)$.

The above follows by the implicit function theorem, as well as the fact that polynomials are smooth functions; a full proof is not provided here.

Knowing that the tangent space $T_{Z_p}(x_0)$ exists at a point $x_0 \in Z_p$ gives structural information about the lines through $x_0$ that happen to lie in $Z_p$.

**Lemma 6.3.** Let $n \geq 2$ and $p \in \mathbb{R}[x_1, \ldots, x_n]$. Let $x_0 \in Z_p$ be such that the tangent space $T_{Z_p}(x_0)$ exists. Then, all lines through $x$ that are contained in $Z_p$ lie in $T_{Z_p}(x_0)$.

**Proof.** Let $\ell$ be a line through $x$ that is contained in $Z_p$. Then, since $T_{Z_p}(x_0)$ is tangent to $Z_p$ at $x_0$, it must also be tangent to the line $\ell$ at $x_0$. This is only possible if $\ell$ is fully contained in $T_{Z_p}(x_0)$.

\[\Box\]

6.1. **Remark on Proposition 6.2.** While this is not exam material, let us get some intuition on why Proposition 6.2 holds by explaining at least why $\nabla p(x_0)$ is normal to $T_{Z_p}(x_0)$ when this tangent space exists.

**Proposition 6.4.** Let $p \in \mathbb{R}[x_1, \ldots, x_n]$. If $x_0 \in Z_p$ is such that $T_{Z_p}(x_0)$ exists, then the vector $\nabla p(x_0)$ is normal to $T_{Z_p}(x_0)$, i.e.
$$\nabla p(x_0) \cdot v = 0 \text{ for all } v \in T_{Z_p}(x_0)$$
(where $\cdot$ denotes the inner product in $\mathbb{R}^n$).

**Proof.** For any $v \in T_{Z_p}(x_0)$, there exists a curve $\gamma$ in $Z_p$, passing through $x_0$, with tangent vector at $x_0$ equal to $v$. Therefore, it suffices to show the following:

For any curve $\gamma : I \to Z_p$, for some interval $I \subset \mathbb{R}$, with the property that there exists $t_0 \in \mathbb{R}$ such that
$$\gamma(t_0) = x_0$$
(i.e., such that $\gamma$ passes through $x_0$), and that
$$\gamma'(t_0) \text{ exists}$$
(i.e., such that $\gamma$ has a tangent vector at $x_0$), it holds that
$$(13) \quad \nabla p(x_0) \cdot \gamma'(t_0) = 0$$
(i.e., $\nabla p(x_0)$ is perpendicular to $\gamma'(t_0)$).

Indeed, (13) trivially holds: There exist $\gamma_t : I \to \mathbb{R}$ with
$$\gamma(t) = (\gamma_1(t), \ldots, \gamma_n(t)), \text{ for all } t \in I.$$
Since $\gamma \subset Z_p$, the function
$$p_{|\gamma} : I \to \mathbb{R}$$
with \[ p_{|\gamma}(t) := p(\gamma_1(t), \ldots, \gamma_n(t)) \]
is identically 0, therefore
\[ (p_{|\gamma})' = 0; \]
in particular,
\[ (p_{|\gamma})'(t_0) = 0. \]
Since by the chain rule
\[ (p_{|\gamma})'(t_0) = \frac{d}{dt} \left( p(\gamma_1(t), \ldots, \gamma_n(t)) \right) \bigg|_{t=t_0} \]
\[ = \left( \frac{\partial p}{\partial x_1}(x) \right)_{x=\gamma(t_0)} \gamma_1'(t_0) + \ldots + \left( \frac{\partial p}{\partial x_n}(x) \right)_{x=\gamma(t_0)} \gamma_n'(t_0) = \nabla p(x_0) \cdot \gamma'(t_0), \]
it follows that
\[ \nabla p(x_0) \cdot \gamma'(t_0) = 0, \]
as required.

\[ \square \]

7. The Joints Problem

\textbf{Definition 7.1.} Let \( \mathcal{L} \) be a family of lines in \( \mathbb{R}^n \), for \( n \in \mathbb{Z}_{\geq 2} \). We say that a point \( x \in \mathbb{R}^n \) is a joint formed by \( \mathcal{L} \) if there exist at least \( n \) lines in \( \mathcal{L} \) through \( x \) with the property that their directions span \( \mathbb{R}^n \).

Denote by \( J(\mathcal{L}) \) the set of joints formed by \( \mathcal{L} \).

Observe that:

- Any two distinct lines in \( \mathbb{R}^2 \) meeting at a point form a joint there (since their two directions are linearly independent, and therefore span \( \mathbb{R}^2 \)). Therefore, in \( \mathbb{R}^2 \) joints are not special; \( J(\mathcal{L}) \) is merely the set of pairwise intersections of the lines in \( \mathcal{L} \), and therefore \#\( J(\mathcal{L}) \) can be as large as \( \#\mathcal{L}^2 \sim L^2 \).

- For three lines meeting at a point \( x \in \mathbb{R}^3 \) to form a joint at \( x \), not all three should lie in the same plane. Therefore, in \( \mathbb{R}^3 \) joints are harder to encounter than in \( \mathbb{R}^2 \). It therefore makes some sense that \#\( J(\mathcal{L}) \) should never be as large as \( \#\mathcal{L}^2 \) in \( \mathbb{R}^3 \).

A further indication for the above is the following: Suppose that all lines in \( \mathcal{L} \) are lying on a 2-dimensional plane \( P \) in \( \mathbb{R}^3 \). Let \( x \) be a point in the plane \( P \), lying in the intersection of at least three lines in \( \mathcal{L} \). For the lines through \( x \) to form a joint at \( x \), at least one of them (call it \( \ell \)) has to be rotated around \( x \), so that it intersects the plane only at \( x \). So, while before \( \ell \) was intersecting potentially all the other \#\( \mathcal{L} - 1 \) lines in \( \mathcal{L} \), after its rotation all these points of intersection are lost. In other words, trying to impose “3-dimensionality” in the picture has cost us a lot of point-line incidences; few points should survive this process, therefore it is not inconceivable that joints in 3 dimensions should be few.
Conjecture 7.2. (1993) Let \( n \in \mathbb{Z}_{\geq 2} \). For any finite family \( \mathcal{L} \) of \( L \) lines in \( \mathbb{R}^n \), it holds that
\[
\# J(\mathcal{L}) \lesssim L^{\frac{n}{n-1}}
\]
(i.e., that \( \# J(\mathcal{L}) \leq c_n \ L^{\frac{n}{n-1}} \), for a constant \( c_n > 0 \) depending only on the dimension \( n \)).

Observe that in \( \mathbb{R}^2 \) the conjecture holds: it states that \( \# J(\mathcal{L}) \lesssim L^2 \), which is trivially true (since \( \# J(\mathcal{L}) \leq \binom{L}{2} \)).

Already in \( \mathbb{R}^3 \) however the conjecture is far from trivial; it states that \( \# J(\mathcal{L}) \lesssim L^{3/2} \). Not only is \( L^{3/2} \) much smaller than the trivial \( L^2 \) bound, it also is a quantity to which one cannot easily assign a combinatorial meaning (\( L^2 \) can be seen as the size of the set of pairs of lines in \( \mathcal{L} \), \( L^3 \) can be seen as the size of the set of triples of lines in \( \mathcal{L} \), however there is no obvious combinatorial meaning for \( L \) raised to a fractional power.)

And, as the dimension \( n \) gets higher, the power \( \frac{n}{n-1} \) of \( L \) in the joints conjecture gets smaller and closer to 1, making the problem harder.

It turns out that the joints conjecture is true; this was proved by Quilodr´an and independently by Kaplan, Sharir and Shustin in 2009. The proof provided here is by Quilodr´an (done while he was a graduate student at UC Berkeley!).

Theorem 7.3. Let \( n \in \mathbb{Z}_{\geq 2} \). For any finite family \( \mathcal{L} \) of \( L \) lines in \( \mathbb{R}^n \), it holds that
\[
\# J(\mathcal{L}) \lesssim L^{\frac{n}{n-1}}.
\]

Proof. Crucial in the proof of the joints theorem is the following fact, which clearly demonstrates the benefits of knowing that \( J(\mathcal{L}) \) lies in the zero set of some low-degree polynomial.

Claim 7.4. If a non-zero \( p \in \mathbb{R}[x_1, \ldots, x_n] \) vanishes on \( J(\mathcal{L}) \), then
\[
\# J(\mathcal{L}) \leq \deg p \cdot L.
\]

Once the Claim is proved, the proof is easily completed due to the fact that there exists non-zero \( p \in \mathbb{R}[x_1, \ldots, x_n] \) that vanishes on \( J(\mathcal{L}) \), with \( \deg p \lesssim (\# J(\mathcal{L}))^{1/n} \). More precisely, the existence of such \( p \) implies by Claim 7.4 that
\[
\# J(\mathcal{L}) \lesssim (\# J(\mathcal{L}))^{1/n} \cdot L,
\]
i.e. that
\[
\# J(\mathcal{L}) \lesssim L^{\frac{n}{n-1}}.
\]

\[ \square \]

Proof of Claim 7.4. Observe that we need to show that the joints are few. This is proved in two steps.

- In Step 1, it is shown that, if a polynomial \( p \) vanishes on the set of joints formed by some finite set of lines, then one of the lines contains few joints.

- In Step 2, we remove from our set of lines a line that contains few joints, together with the joints it contains. We continue iteratively: one of the remaining lines contains few joints, so we remove that line and the joints it contains from our collection, and so on. We continue this process until all the joints are removed. We thus observe that the original joints cannot have been too many, since we eventually removed all joints by removing only few at each step.
More precisely, the proof unfolds as follows.

Step 1: This step ensures the validity of the following Claim, which states that, under the assumption of Claim 7.4, one of the lines contains few joints.

**Claim 7.5.** If a non-zero \( p \in \mathbb{R}[x_1, \ldots, x_n] \) vanishes on \( J(L) \), then there exists \( \ell \in L \) containing \( \leq \deg p \) joints of \( J(L) \).

To prove Claim 7.5, one may assume without loss of generality that, amongst all non-zero polynomials in \( \mathbb{R}[x_1, \ldots, x_n] \) that vanish on \( J(L) \), \( p \) has lowest degree. Indeed, if \( p \) does not have this property, then \( p \) can be replaced with a non-zero \( g \in \mathbb{R}[x_1, \ldots, x_n] \) of lower degree that vanishes on \( J(L) \), and deduce that there exists \( \ell \in L \) containing \( \leq \deg g \) (and therefore \( \leq \deg p \)) joints in \( J(L) \).

Suppose for contradiction that each \( \ell \in L \) contains \( > \deg p \) joints in \( J(L) \). Since \( J(L) \subseteq \mathbb{Z}_p \), it follows that each \( \ell \in L \) meets \( \mathbb{Z}_p > \deg p \) times, therefore

\[ \ell \subseteq \mathbb{Z}_p \text{ for all } \ell \in L. \]

Crucially, this implies that

\[ \nabla p(x) = 0 \text{ for all } x \in J(L). \]  \( (14) \)

Indeed, let \( x \in J(L) \). There exist (at least) \( n \) lines in \( L \) through \( x \), whose directions span \( \mathbb{R}^n \), and all of which fully lie in \( \mathbb{Z}_p \). Therefore, there does not exist a tangent space \( T_{\mathbb{Z}_p}(x) \) to \( \mathbb{Z}_p \) at \( x \): if \( T_{\mathbb{Z}_p}(x) \) existed, then, by Lemma 6.3, it would contain all lines in \( L \) through \( x \) (since all of these lie in \( \mathbb{Z}_p \), and should thus be tangent to \( T_{\mathbb{Z}_p}(x) \)), therefore these lines would span an at most \( (n-1) \)-dimensional space, rather than the whole of \( \mathbb{R}^n \). Thus, by Proposition 6.2, \( \nabla p(x) = 0 \).

By \( (14) \), the partial derivatives of \( p \) vanish on \( J(L) \); and these are all polynomials of lower degree than the degree of \( p \). Since by our initial assumption \( p \) has lowest degree amongst all non-zero polynomials in \( \mathbb{R}[x_1, \ldots, x_n] \) that vanish on \( J(L) \), it follows that the partial derivatives of \( p \) are all the zero polynomial; i.e.,

\[ \frac{\partial p}{\partial x_i} \equiv 0 \text{ for all } i = 1, \ldots, n. \]

This implies that \( p \) is a constant. However, \( p \) vanishes somewhere (since it vanishes on \( J(L) \)). So, \( p \equiv 0 \), a contradiction.

Therefore, the assumption that each \( \ell \in L \) contains \( > \deg p \) joints in \( J(L) \) is false. Hence, there exists \( \ell \in L \) containing \( \leq \deg p \) joints in \( J(L) \).

**Step 2:** Claim 7.4 is now proved by appropriately iterating Step 1.

Let \( p \in \mathbb{R}[x_1, \ldots, x_n] \) be a non-zero polynomial that vanishes on \( J(L) \).

By Step 1 (Claim 7.5), there exists a line \( l_1 \in L \) containing at most \( \deg p \) elements of \( J(L) \). We remove \( l_1 \) and the elements of \( l_1 \cap J(L) \) from our collection of lines and joints, respectively (thus removing \( \leq \deg p \) joints from \( J(L) \)).

The remaining joints in \( J(L) \), i.e. the elements of \( J_1 := J(L) \setminus (l_1 \cap J(L)) \), are joints formed by \( L_1 := L \setminus \{l_1\} \). Now, \( p \) is a non-zero polynomial in \( \mathbb{R}[x_1, \ldots, x_n] \) that vanishes on \( J(L_1) \) (since \( J(L_1) \subseteq J(L) \)). Again by Step 1 (Claim 6.5), there exists \( l_2 \in L_1 \), containing at most \( \deg p \) elements of \( J(L_1) \), and thus of \( J_1 \). We remove \( l_2 \) and the elements of \( l_2 \cap J_1 \) from our collection of lines and joints, respectively.
We continue as above, until we have removed enough lines of \( L \) (and the corresponding elements of \( J(L) \) on each) so that no elements of \( J(L) \) are remaining. This is achieved in at most \( L \) steps, and \( J(L) \) is the union of its subsets that were removed in each step, each of which has size at most \( \deg p \). Therefore,

\[
\# J(L) \leq \deg p \cdot L.
\]

8. Polynomial partitioning

8.1. What’s the point? This is a preliminary discussion, which, while not necessary to understand polynomial partitioning, may shed some light on what it is about; and, in general, on why the polynomial method has proved so useful (other than merely providing nice computational bounds).

Recall that Theorem 5.5 ensures that, for any finite set \( P \) of points, there exists a polynomial \( p \) of relatively low degree that vanishes on \( P \):

**Theorem 5.5**. Let \( n \geq 1 \) and \( \mathbb{F} \) be a field. Then, for any finite set \( P \) of points in \( \mathbb{F}^n \), there exists a non-zero \( p \in \mathbb{F}[x_1, \ldots, x_n] \), with \( \deg p \lesssim \left( \# P \right)^{1/n} \), that vanishes on \( P \).

It is clear by now that, the lower \( \deg p \) is (for \( p \) vanishing on \( P \)), the better bounds we get (see for instance Proposition 5.4 and Claim 7.4). In general, however, one cannot expect to find a polynomial vanishing on \( P \) of degree considerably smaller than \( \left( \# P \right)^{1/n} \) (see the discussion after the statement of Theorem 5.5 in Section 5). Still though, one would hope that, under extra assumptions on \( P \), this might be possible - that, after working hard enough, one could reduce to a situation where \( P \) (or at least many points in \( P \)) indeed lie in the zero set of a polynomial of degree much lower than \( \left( \# P \right)^{1/n} \).

Quite amazingly, the latter has been the case for many point-line incidence problems tackled so far. While the whole picture is unclear yet, the moral is the following.

When lines cluster on zero sets of polynomials of low degree, then they intersect each other too much, creating many point-line incidences (imagine for instance lines lying in a 2-dimensional plane, the zero set of a polynomial of degree 1; any two intersect each other in principle; similarly, we get a lot of incidences between lines clustering on a hyperboloid of one sheet). The amazing part is that, to an extent, a sort of “converse” holds: roughly speaking, when it comes to point-line incidence problems, it seems that the worst case scenarios, or at least the hard-to-tackle situations where we have too many point-line incidences, tend to be those where the lines cluster on zero sets of polynomials of low degree.

Therefore:
If we believe the above, then: when it comes to counting point-line incidences, any method that allows us to reduce to a situation where our points lie in the zero set of a polynomial of low degree could be valuable. The obvious reason is that, this way, we have all the nice bounds that follow from the fact that a polynomial of low degree vanishes on our points. The deeper reason is that, this way, we are hopefully reducing our problem to a worst case scenario (which, for better or worse, we would have to deal with eventually), without having to prove that it is a worst case scenario.

Now, Theorem 5.5 indeed provides a way to reduce to a situation where our points lie in the zero set of a polynomial of (sort of) low degree; without any work in fact. And this is actually why Theorem 5.5 is not always useful: it immediately asserts the existence of such a zero set, without taking into account the special structure of the set of points we are considering. This is where polynomial partitioning comes in, a method developed by Guth and Katz in 2010 to solve the Erdős distinct distances problem on the plane.

Polynomial partitioning is a method that has the potential to allow us to reduce our point-line incidence problem to a case where our points lie in the zero set of a polynomial of very low degree (lower than the one offered by Theorem 5.5). The reason is that polynomial partitioning takes advantage of the structure of our sets of points and lines.

8.2. Polynomial partitioning - the statement. At first glance, it will perhaps not be clear how polynomial partitioning could allow us to reduce to a situation where a lot of our points lie in the zero set of a polynomial of low degree. This will be explained later. For now, let us see the statement - it asserts that, no matter what finite set \( P \) of points we start with, and no matter what degree \( d \) we want to use, we can find a polynomial of degree at most \( d \) that splits our set of points in smaller parts, in a nice, controlled way.

**Theorem 8.1. (Polynomial partitioning)** For any finite set \( P \) of points in \( \mathbb{R}^n \), and any \( d \geq 1 \), there exists a non-zero polynomial \( p \in \mathbb{R}[x_1, \ldots, x_n] \), with \( \deg p \leq d \), whose zero set \( Z_p \) splits \( \mathbb{R}^n \) in \( \sim d^n \) cells, each containing \( \lesssim \frac{\#P}{d^n} \) points of \( P \).

The “cells” above are the connected components of \( \mathbb{R}^n \setminus Z_p \). The polynomial partitioning theorem asserts control on the number of connected components of \( \mathbb{R}^n \setminus Z_p \), and on the number of points of \( P \) contained in each of those connected components.

So, polynomial partitioning allows us to divide and conquer: split our points in smaller parts, and work with each part. Then we can focus on the points that lie on the zero set \( Z_p \); importantly, this \( p \) can now have degree much lower than \( (\#P)^{1/n} \).
8.3. Polynomial partitioning - proof. While not exam material, we present here the proof of the polynomial partitioning - just because it is so beautiful.

The proof is based on the Borsuk–Ulam theorem of topology. This theorem uses the topology of $\mathbb{R}^n$, which is why polynomial partitioning is a theorem about euclidean space only. It asserts that, when the unit sphere in $\mathbb{R}^n$ is compressed into a hyperplane in a continuous and odd manner, then some point of it is mapped to 0.

**Theorem 8.2. (Borsuk–Ulam)** Let $n \geq 2$. Let $f : S^{n-1} \to \mathbb{R}^{n-1}$ be continuous and odd. Then, there exists $x \in S^{n-1}$, with $f(x) = 0$.

The first step towards the proof of the polynomial partitioning Theorem 8.1 is establishing a version of Theorem 5.5 for finitely many sets of finite, positive volume (rather than finitely many points). This is achieved using the Borsuk–Ulam theorem.

**Theorem 8.3. (Polynomial ham sandwich theorem; Stone–Tukey, 1942)** Let $S_1, S_2, \ldots, S_M$ be sets of finite, positive volume in $\mathbb{R}^n$. Then, there exists a non-zero polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$, with $\deg p \lesssim M^{1/n}$, whose zero set $Z_p$ bisects each $S_i$, i.e. such that

$$\text{vol}(S_i \cap \{ p > 0 \}) = \text{vol}(S_i \cap \{ p < 0 \}) \text{ for all } i = 1, \ldots, M.$$

Observe that the numerology in the polynomial ham sandwich theorem is the same as the one in Theorem 5.5. Indeed:

- Theorem 5.5 asserts that, for every finite set of points in $\mathbb{R}^n$, there exists a non-zero $p \in \mathbb{R}[x_1, \ldots, x_n]$, of degree $\lesssim (\# \text{points})^{1/n}$, whose zero set passes through all the points.
- The polynomial ham sandwich theorem asserts that, for every finite set of (appropriate) sets in $\mathbb{R}^n$, there exists a non-zero $p \in \mathbb{R}[x_1, \ldots, x_n]$, of degree $\lesssim (\# \text{sets})^{1/n}$, whose zero set bisects all the sets.

**Proof of the polynomial ham sandwich theorem.** Identify each polynomial in $\mathbb{R}[x_1, \ldots, x_n]$ of degree at most $d$ with its sequence of $\binom{d+n}{n}$ coefficients, which is an element of $\mathbb{R}^{\binom{d+n}{n}}$. This way, the set of polynomials in $\mathbb{R}[x_1, \ldots, x_n]$ of degree at most $d$ can be seen as the space $\mathbb{R}^{\binom{d+n}{n}}$ with the usual metric.

Define the map

$$f : S^{\binom{d+n}{n}-1} \to \mathbb{R}^M,$$

with

$$f(p) := \left( \text{vol}(S_1 \cap \{ p > 0 \}) - \text{vol}(S_1 \cap \{ p < 0 \}), \ldots, \text{vol}(S_M \cap \{ p > 0 \}) - \text{vol}(S_M \cap \{ p < 0 \}) \right)$$

for all $p \in S^{\binom{d+n}{n}-1}$ (i.e., for all $p \in \mathbb{R}[x_1, \ldots, x_n]$ of degree at most $d$, whose vector of coefficients has euclidean length 1.)

It is easy to see that $f$ is continuous (since perturbing slightly the coefficients of $p$ should only slightly perturbe the sets $\{ p > 0 \}$ and $\{ p < 0 \}$) and odd (since $-p > 0$ iff $p < 0$).

Therefore, as long as $\mathbb{R}^M \subseteq S^{\binom{d+n}{n}-1}$, the Borsuk–Ulam theorem holds for $f$, implying that there exists a polynomial $p \in S^{\binom{d+n}{n}-1}$ with $f(p) = 0$, i.e. such that

$$\text{vol}(S_i \cap \{ p > 0 \}) = \text{vol}(S_i \cap \{ p < 0 \}) \text{ for all } i = 1, \ldots, M.$$
That is, for any \( d \geq 1 \) such that \( M \leq \binom{d+n}{n} - 1 \), i.e. such that
\[
M < \binom{d+n}{n},
\]
there exists non-zero \( p \in \mathbb{R}[x_1, \ldots, x_n] \) (in fact, with vector of coefficients having euclidean length 1) of degree at most \( d \), such that \( Z_p \) bisects each \( S_i \).

Since \( \binom{d+n}{n} \geq \frac{1}{n!} d^n \), it follows that, for any \( d \geq 1 \) such that \( \frac{1}{n!} d^n > M \), (15) holds as well. Therefore:

For any \( d \geq 1 \) with \( \frac{1}{n!} d^n > M \) \( \iff d > (n!)^{1/n} M^{1/n} \), there exists non-zero \( p \in \mathbb{R}[x_1, \ldots, x_n] \) of degree at most \( d \), such that \( Z_p \) bisects each \( S_i \).

The smallest \( d \) that satisfies the above conditions is
\[
d = \left\lfloor (n!)^{1/n} M^{1/n} \right\rfloor + 1,
\]
which has the property that it is \( \leq 2(n!)^{1/n} M^{1/n} \lesssim M^{1/n} \). For this \( d \) there exists polynomial as above; i.e., there exists non-zero \( p \in \mathbb{R}[x_1, \ldots, x_n] \) of degree at most \( d \), and therefore \( \lesssim M^{1/n} \), whose zero set bisects each \( S_i \).

The 
second step

The second step towards the proof of the polynomial partitioning Theorem 8.1 is establishing a version of the polynomial ham sandwich Theorem 8.3 for finite sets of points (rather than sets of finite, positive volume). This is achieved by slightly thickening each point, so that we get sets of finite, positive volume, and then applying the polynomial ham sandwich theorem.

Theorem 8.4. (Guth–Katz, 2010) Let \( S_1, S_2, \ldots, S_M \) be finite sets of points in \( \mathbb{R}^n \). Then, there exists a non-zero polynomial \( p \in \mathbb{R}[x_1, \ldots, x_n] \), with \( \deg p \lesssim M^{1/n} \), whose zero set \( Z_p \) bisects each \( S_i \), i.e. such that
\[
\#(S_i \cap \{ p > 0 \}) \leq \frac{\#S_i}{2}
\]
and
\[
\#(S_i \cap \{ p < 0 \}) \leq \frac{\#S_i}{2}
\]
for all \( i = 1, \ldots, M \).

Note that “bisection” in this case is not a too literal term; it means that, for each \( i = 1, \ldots, M \), at most half of the points of \( S_i \) lie on the set where \( p \) is positive, and at most half of the points of \( S_i \) lie on the set where \( p \) is negative. It could even be that all points of some \( S_i \) are contained in \( Z_p \), in which case 0 points of \( S_i \) lie in the sets \( \{ p > 0 \} \) and \( \{ p < 0 \} \); that is still a legitimate bisection of \( S_i \) by \( Z_p \).

The numerology in Theorem 8.4 above is exactly the same as the one for the polynomial ham sandwich theorem. Observe, that, while the upper bound \( M^{1/n} on the degree of the bisecting polynomial depends on the number of sets of points we are bisecting, it does not depend on the number of points in total; each \( S_i \) can contain a very large number of points, yet this will not affect the degree of the polynomial. This is why Theorem 8.4 brings us closer to the proof of the polynomial partitioning Theorem 8.1 (whose statement includes a degree independent of the size of the set of points we start with).

Proof of Theorem 8.4. For each \( \delta > 0 \), let \( U_{i,\delta} \) be the union of the \( \delta \)-balls centred at the points of \( S_i \), \( i = 1, \ldots, M \) (in other words, \( U_{i,\delta} \) is just the set we get if we replace each point in \( S_i \) with a \( \delta \)-ball centered at that point).
By the polynomial ham sandwich Theorem \[8.3\] for each \(\delta > 0\) there exists a non-zero \(p_\delta \in \mathbb{R}[x_1, ..., x_n]\), of degree \(\leq M^{1/n}\), whose zero set \(U_{i,\delta}\) for all \(i = 1, ..., M\). Recall, in fact, that we may assume that \(p_\delta \in S^{(\frac{d+n}{n})-1}\), when \(p_\delta\) is identified with its vector of \(\binom{d+n}{n}\) coefficients.

Recall that \(S^{(\frac{d+n}{n})-1}\) is a compact (and thus sequentially compact) subset of \(\mathbb{R}^{(\frac{d+n}{n})}\) with respect to the euclidian metric. Therefore, there exists sequence \((\delta_n)_{n \in \mathbb{N}}\) of positive radii that converges to 0, and \(p \in S^{(\frac{d+n}{n})-1}\), such that

\[
p_{\delta_n} \to p
\]

in \(S^{(\frac{d+n}{n})-1}\) w.r.t. the euclidean distance. This merely means that we have convergence coordinate-wise (recall that the polynomials above are now identified with their vectors of coefficients). In other words, the coefficients of \(p_{\delta_n}\) converge to the coefficients of \(p\) as \(n \to \infty\). This implies that \(p_{\delta_n}\) converges to \(p\) uniformly on bounded subsets of \(\mathbb{R}^n\) (when these polynomials are seen as functions on \(\mathbb{R}^n\)).

We now show that \(Z_p\) bisects \(S_i\), for all \(i = 1, ..., M\).

Indeed, let us assume that \(Z_p\) does not bisect \(S_i\), for some \(i \in \{1, ..., M\}\). Then, either \(p > 0\) on more than half of the points of \(S_i\), or \(p < 0\) on more than half of the points of \(S_i\). Suppose that \(p > 0\) on more than half of the points of \(S_i\). Let \(S^+_i\) be the subset of \(S_i\) on which \(p\) is positive. Since \(p\) is a continuous function on \(\mathbb{R}^n\) and \(S^+_i\) is a finite set of points, there exists some \(\epsilon > 0\) such that \(p > 0\) on the union of the \(\epsilon\)-balls centred at \(S^+_i\). Now, since \(p_{\delta_n}\) converges to \(p\) uniformly on compact sets, there exists \(n \in \mathbb{N}\), such that \(p_{\delta_n} > 0\) on the union of the \(\delta_n\)-balls centred at \(S^+_i\) and the \(\delta_n\)-balls centred at \(S^-_i\) are disjoint. However, \(S^+_i\) contains more than half of the points of \(S_i\), and thus the zero set of \(p_{\delta_n}\) does not bisect the union of the \(\delta_n\)-balls centred at \(S_i\), i.e. the set \(U_{i,\delta_n}\), which is a contradiction. Similarly, we are led to a contradiction if we assume that \(p < 0\) on more than half of the points of \(S_i\). Therefore, \(Z_p\) bisects \(S_i\), for all \(i = 1, ..., M\).

The third step towards the proof of the polynomial partitioning Theorem \[8.1\] is an iterative application of Theorem \[8.4\] on \(\mathcal{P}\): first bisect \(\mathcal{P}\) with the zero set of a low degree polynomial, then bisect each of the two sets we get by the zero set of a low degree polynomial, then bisect each of the four sets we get with the zero set of a low degree polynomial, and so on. Stop this process when the original set \(\mathcal{P}\) has been split in the number of parts we wanted.

**Proof of polynomial partitioning Theorem \[8.1\].** Let \(\mathcal{P}\) be a finite set of points in \(\mathbb{R}^n\), and \(d \geq 1\).

By Theorem \[8.4\] applied to the finite set of points \(\mathcal{P}\), there exists a non-zero \(p_1 \in \mathbb{R}[x_1, ..., x_n]\), of degree \(\leq 1/n\), whose zero set \(Z_{p_1}\) bisects \(\mathcal{P}\). I.e., \(\mathbb{R}^n \setminus Z_{p_1}\) consists of \(2^1\) disjoint cells (the cell \(\{p_1 > 0\}\) and the cell \(\{p_1 < 0\}\)), each of which contains \(\leq \frac{\#\mathcal{P}}{2}\) points of \(\mathcal{P}\).

By Theorem \[8.4\] applied to the finite sets of points \(\mathcal{P} \cap \{p_1 > 0\}\), \(\mathcal{P} \cap \{p_1 < 0\}\), there exists a non-zero \(p_2 \in \mathbb{R}[x_1, ..., x_n]\), of degree \(\leq 2^{\frac{1}{n}}\), whose zero set \(Z_{p_2}\) bisects \(\mathcal{P} \cap \{p_1 > 0\}\) and \(\mathcal{P} \cap \{p_1 < 0\}\). I.e., \(\mathbb{R}^n \setminus (Z_{p_1} \cup Z_{p_2})\) consists of \(2^2\) disjoint cells (the cells \(\{p_1 > 0\} \cap \{p_2 > 0\}\), \(\{p_1 > 0\} \cap \{p_2 < 0\}\), \(\{p_1 < 0\} \cap \{p_2 > 0\}\) and \(\{p_1 < 0\} \cap \{p_2 < 0\}\)), each of which contains \(\leq \frac{\#\mathcal{P}}{2^2}\) points of \(\mathcal{P}\).
We continue in a similar way; by the end of the $j$-th step, we have produced non-zero $p_1, \ldots, p_j \in \mathbb{R}[x_1, \ldots, x_n]$, of degrees $\lesssim 2^{(1-1)/n}, \ldots, \lesssim 2^{(J-1)/n}$, respectively, such that $\mathbb{R}^n \setminus (Z_{p_1} \cup \ldots \cup Z_{p_j})$ consists of $2^j$ disjoint cells, each of which contains $\lesssim \frac{\#P}{2^j}$ points of $\mathcal{P}$.

We stop this procedure at the $J$-th step, where $J$ is such that the polynomial $p := p_1 \cdots p_J$ has degree $\leq d$ and the number of cells in which $\mathbb{R}^n \setminus (Z_{p_1} \cup \ldots \cup Z_{p_j}) = \mathbb{R}^n \setminus Z_p$ is decomposed is $\sim d^n$. (In other words, we stop when $2^{(1-1)/n} + 2^{(2-1)/n} + \ldots + 2^{(J-1)/n} \lesssim d$ and $2^J \sim d^n$, for appropriate constants hiding behind the $\lesssim$ and $\sim$ symbols; this can be achieved, as $2^{(1-1)/n} + 2^{(2-1)/n} + \ldots + 2^{(J-1)/n} \sim 2^{J/n}$.) The polynomial $p$ has the desired properties.

\[ \square \]

**Remark.** It is clearer from the proof of the polynomial partitioning theorem that the “cells” in its statement are simply the $2^J$ subsets of $\mathbb{R}^n$, each of which is fully determined by a choice of signs for the polynomials $p_1, \ldots, p_J$ that are used during the $J$ steps of the method. So, technically some of the cells may be empty (for instance, the cell $\{p_1 < 0\} \cap \{p_2 < 0\} \cap \ldots \cap \{p_J < 0\}$ is empty if the polynomials $p_1, \ldots, p_J$ never take negative values; if, for instance, each of these polynomials is the square of another polynomial).

So, the connected components of $\mathbb{R}^n \setminus Z_p$ are actually the *non-empty* cells. Thus, while the cells are $\sim d^n$, the connected components of $\mathbb{R}^n \setminus Z_p$ may be fewer than $\sim d^n$.

9. **How to use polynomial partitioning.**

Recall the statement of the polynomial partitioning theorem:

**Theorem 8.1 (Polynomial partitioning)** For any finite set $\mathcal{P}$ of points in $\mathbb{R}^n$, and any $d \geq 1$, there exists a non-zero polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$, with $\deg p \leq d$, whose zero set $Z_p$ splits $\mathbb{R}^n$ in $\sim d^n$ cells, each containing $\lesssim \frac{\#P}{d^n}$ points of $\mathcal{P}$.

It has been mentioned above that the reason polynomial partitioning is useful is that it has the potential to allow us to reduce the problem of controlling $\#\mathcal{P}$ to a case where $\mathcal{P}$ lies in the zero set of a polynomial of degree considerably lower than $(\#\mathcal{P})^{1/n}$.

It is here described how such a reduction, if possible, can be achieved.

Let $\mathcal{P}$ be a finite set of points in $\mathbb{R}^n$ that we want to count. Let $d \geq 1$. Fix a non-zero $p \in \mathbb{R}[x_1, \ldots, x_n]$, with $\deg p \leq d$, whose zero set partitions $\mathbb{R}^n$ as above. Then, no matter what $d$ is, there are two possible cases:

**The cellular case:** $\gtrsim \#\mathcal{P}$ points of $\mathcal{P}$ (say, more than 1% of the points in $\mathcal{P}$) lie outside $Z_p$ (i.e., in the union of the cells).

**The algebraic case:** $\gtrsim \#\mathcal{P}$ points of $\mathcal{P}$ (say, more than 90% of the points in $\mathcal{P}$) lie on $Z_p$.

If for some $d \ll (\#\mathcal{P})^{1/n}$ we can somehow deal with the cellular case, either by contradicting it or by solving our problem in that case, then we can assume that we are in the algebraic case. That is, the problem is reduced to a case where most of our points cluster on the zero set of a polynomial of degree considerably smaller than $(\#\mathcal{P})^{1/n}$. 
9.1. **Dealing with the cellular case.** Suppose that we want to control $\#P$ relative to $\#L$, for some family $L$ of lines containing $P$.

Note that, a priori, there is no reason at all why one should be able to deal (at least in a standard manner) with the cellular case for small $d$. Indeed, if there was a standard way to reduce to the algebraic case no matter what problem we started with, it would mean that all problems have an underlying algebraic nature, which does not have to be true.

There exists, however, a standard way to attempt to deal with the cellular case: working with a cell with average behaviour, which, in the cellular case, means a cell that contains the average number of points but is crossed (entered) by few lines in $L$.

More precisely, the existence of such a cell is ensured by the following Lemma.

**Lemma 9.1.** Suppose that we are in the cellular case in $\mathbb{R}^n$. Then, for any finite family $L$ of $L$ lines in $\mathbb{R}^n$, there exists a cell $C$ such that

- $C$ contains $\sim \frac{\#P}{d^n}$ points of $P$
- $C$ is crossed by $\lesssim \frac{L}{d^n-1}$ lines in $L$.

It will soon become clearer why such a cell exhibits average (rather than exceptional) behaviour. In any case:

In a cell $C$ of average behaviour (as in Lemma 9.1) we have a microscopic picture of what we started with. In particular, we have fewer points (but not too few) lying in fewer lines than those we started from. Therefore, induction on the number of lines in $C$, or applying a known estimate inside $C$, may resolve the cellular case.

Precise examples will be seen later. For now, Lemma 9.1 is proved. This is achieved by ensuring that

- a (perhaps small but) definite proportion (say, 1%) of the cells are such that each contains $\sim \frac{\#P}{d^n}$ points of $P$, and
- a large proportion of the cells are such that each is crossed by $\lesssim \frac{L}{d^n-1}$ lines in $L$. By large proportion here we mean large enough so that, when added to the proportion above, we get more than 100% (say, 99.5%).

By the above, a definite proportion of the cells (at least 0.5% with the numbers used above) will have the properties described in Lemma 9.1.

**Remark.** Observe that, by the above, a lot of the cells satisfy the requirements of Lemma 9.1, not just a single one; this will not be useful to our applications later though. The reason is that indeed these cells demonstrate average behaviour, therefore using many of them, rather than just one, is not advantageous.

The two bullet points above are now explained.

- A **definite proportion of the cells have the property that each contains $\sim \frac{\#P}{d^n}$ points of $P$.**
Recall that 
\[ \#\text{cells} \sim d^3, \]
each cell contains \[ \leq \frac{\#P}{d^n} \] points of \( P \),
and, since we are in the cellular case,
\[ \geq \#P \] points of \( P \) are in the cells.

This implies that there exist constants \( c_1, c_2 > 0 \), depending only on the dimension \( n \), such that
\[ \text{each cell contains} \leq c_1 \frac{\#P}{\#\text{cells}} \text{ points of } P \]
and
\[ \geq c_2 \#P \] points of \( P \) are in the cells.

For these particular constants \( c_1 \) and \( c_2 \), the following holds.

**Lemma 9.2.** In the cellular case, \( \geq \frac{c_2}{c_1} \% \) of the cells have the property that each contains \( \sim \frac{\#P}{d^n} \) points of \( P \).

**Proof.** It suffices to show that \( \geq \frac{c_2}{c_1} \% \) of the cells have the property that each contains \( > 0.0001c_2 \frac{\#P}{\#\text{cells}} \) points of \( P \).

[Indeed, since \( \#\text{cells} \sim d^n \), the above implies that each of the above cells contains \( \geq \frac{\#P}{d^n} \) points of \( P \). Since by the polynomial partitioning Theorem 8.1 each cell contains \( \leq \frac{\#P}{d^n} \) points of \( P \), the proof is complete.]

The proof is a simple counting argument. Define
\[ C_{\text{good}} := \{ \text{cells each containing} > 0.0001c_2 \frac{\#P}{\#\text{cells}} \text{ points of } P \} \]
and
\[ C_{\text{bad}} := \{ \text{cells each containing} \leq 0.0001c_2 \frac{\#P}{\#\text{cells}} \text{ points of } P \}. \]

The idea is the following: Each cell is either good or a bad. There are a lot of points in the cells (\( \geq c_2 \#P \)). Since each of the bad cells contains very few points, most of the points must lie in the good cells. However, like all cells, each good cell contains at most “its fair share” of points of \( P \) (that is, \( \leq c_1 \frac{\#P}{\#\text{cells}} \) points of \( P \)). So, the good cells cannot contribute a lot of points in total, unless they are many.

More precisely,
\[ \sum_{\text{all cells}} \#\{ \text{points of } P \text{ in cell} \} \geq c_2 \#P. \]

Since each cell is either good or bad,

\[ \sum_{\text{cells in } C_{\text{good}}} \#\{ \text{points of } P \text{ in cell} \} + \sum_{\text{cells in } C_{\text{bad}}} \#\{ \text{points of } P \text{ in cell} \} \geq c_2 \#P. \]

It holds that
\[ \sum_{\text{cells in } C_{\text{bad}}} \#\{ \text{points of } P \text{ in cell} \} \leq \#C_{\text{bad}} \cdot 0.0001c_2 \frac{\#P}{\#\text{cells}}. \]
(17) \[ \leq \#\text{cells} \cdot 0.0001c_2 \frac{\#P}{\#\text{cells}} \leq 0.0001c_2 \#P \]

(the contribution of the bad cells is very small).

Now, just like all cells, also every good cell contains \( \leq c_1 \frac{\#P}{\#\text{cells}} \) points of \( P \). Therefore, assuming for contradiction that \( \#\text{C}_{\text{good}} \leq \frac{c_2}{c_1} \% \#\text{cells} \), the contribution of the good cells is also small:

\[ \sum_{\text{cells in } C_{\text{good}}} \#\{ \text{points of } P \text{ in cell} \} \leq \#C_{\text{good}} \cdot c_1 \frac{\#P}{\#\text{cells}} \]

(18) \[ \leq \frac{c_2}{c_1} \#\text{cells} \cdot c_1 \frac{\#P}{\#\text{cells}} \leq 0.01c_2 \#P. \]

By (17) and (18), it follows that

\[ \sum_{\text{cells in } C_{\text{good}}} \#\{ \text{points of } P \text{ in cell} \} + \sum_{\text{cells in } C_{\text{bad}}} \#\{ \text{points of } P \text{ in cell} \} < 0.1c_2 \#P, \]

contradicting (16). Therefore, the assumption that \( \#C_{\text{good}} \leq \frac{c_2}{c_1} \% \#\text{cells} \) is false. \( \square \)

- A large proportion of the cells have the property that each is crossed by \( \lesssim \frac{L}{d^n} \) lines in \( L \). This proportion can be made as large as we want, as long as we appropriately enlarge the constant hiding behind the \( \lesssim \) symbol.

[Observe that, the larger \( C \) is, the easier it is for a cell to be crossed by \( \leq C \frac{L}{d^n} \) lines in \( L \). So, it is not absurd that, by enlarging \( C \), we can make the proportion above as large as we wish.]

The precise statement is the following (and is true whether we are in the cellular case or not).

**Lemma 9.3.** Let \( L \) be a finite family of lines in \( \mathbb{R}^n \). For any \( a \in (0, 100) \), \( > a\% \) of the cells have the property that each is crossed by \( \lesssim \frac{L}{d^n} \) lines in \( L \).

More precisely, by \( \lesssim \frac{L}{d^n} \) above we mean \( \leq A \frac{L}{d^n} \), for \( A := \frac{10^{10}}{(100-a)\%c_0} \), where \( c_0 > 0 \) is a constant depending only on the dimension for which

\[ \#\text{cells} \geq c_0 \, d^n \]

(such \( c_0 \) exists, since \( \#\text{cells} \sim d^n \)).

The proof of Lemma 9.3 is a simple counting argument. However, this is achieved thanks to the following crucial lemma, which is in fact one of the main reasons why polynomial partitioning (i.e., partitioning with the zero set of a polynomial, rather than another object) is so useful.

**Lemma 9.4.** Any line in \( \mathbb{R}^n \) enters at most \( d + 1 \) cells.

**Proof.** Let \( \ell \) be a line in \( \mathbb{R}^n \). Suppose that \( \ell \) enters \( \geq d + 2 \) cells. As \( \ell \) passes from one cell to the next, it hits the zero set \( Z_p \).

[Indeed, recall that the polynomial \( p \) achieving the partitioning equals \( p_1 \cdots p_J \), for certain polynomials \( p_1, \ldots, p_J \) used in the partitioning process. Each cell is fully determined by a]
choice of signs for the \(p_i\)'s. Therefore, for any two cells \(C_1, C_2\), one of the \(p_i\)'s is positive on \(C_1\) and negative on \(C_2\). So, the polynomial \((p_i)_\ell\) (of one variable) is positive on \(\ell \cap C_1\) and negative on \(\ell \cap C_2\). Therefore, by the intermediate value theorem, \((p_i)_\ell\) vanishes in the transition between the cells.

So, \(\ell\) hits \(Z_p\) at least \(d + 1 \geq \deg p + 1\) times. Thus, \(\ell \subseteq Z_p\), so \(\ell\) enters no cells, a contradiction under the assumption that it enters \(\geq d + 2\) cells.

So, \(\ell\) enters \(\leq d + 1\) cells.

\[
\boxed{}
\]

Proof of Lemma 9.3. Let \(\mathcal{L}\) be a family of \(L\) lines in \(\mathbb{R}^n\). By the Lemma above,

\[
\sum_{\ell \in \mathcal{L}} \#\{\text{cells crossed by } \ell\} \lesssim L d.
\]

On the other hand,

\[
\sum_{\ell \in \mathcal{L}} \#\{\text{cells crossed by } \ell\} = \sum_{\text{cells}} \#\{\text{lines in } \mathcal{L} \text{ through the cell}\}.
\]

Therefore,

\[
\sum_{\text{cells}} \#\{\text{lines in } \mathcal{L} \text{ crossing the cell}\} \leq 2Ld.
\]

[Observe that this means that each cell is crossed on average by

\[
\frac{\sum_{\text{cells}} \#\{\text{lines in } \mathcal{L} \text{ crossing the cell}\}}{\#\text{cells}} \leq \frac{2Ld}{\#\text{cells}} \sim \frac{Ld}{d^n} \sim \frac{L}{d^{n-1}}
\]

lines in \(\mathcal{L}\). So, at least one cell should have this property. Lemma 9.3 however asserts that this average behaviour is exhibited by many cells, not simply one.]

Fix \(a \in (0, 100)\), and let \(A := \frac{10^{10}}{(100 - a)c_0}\), where \(c_0 > 0\) is a constant depending only on the dimension such that

\[
\#\text{cells} \geq c_0 d^n.
\]

Now, define

\[
C_{\text{good}} := \left\{\text{cells each crossed by } \leq A \frac{L}{d^{n-1}} \text{ lines of } \mathcal{L}\right\}
\]

and

\[
C_{\text{bad}} := \{\text{all cells}\} \setminus C_{\text{good}} = \left\{\text{cells each crossed by } > A \frac{L}{d^{n-1}} \text{ lines of } \mathcal{L}\right\}.
\]

It is now shown that \#\(C_{\text{good}}\) > \(a\%\) \#\(\text{cells}\). Indeed, assume for contradiction that \#\(C_{\text{good}}\) \(\leq a\%\) \#\(\text{cells}\). Since each cell is either good or bad, it holds that

\[
\sum_{\text{cells in } C_{\text{good}}} \#\{\text{lines in } \mathcal{L} \text{ crossing the cell}\} = \sum_{\text{cells in } C_{\text{good}}} \#\{\text{lines in } \mathcal{L} \text{ crossing the cell}\} + \sum_{\text{cells in } C_{\text{bad}}} \#\{\text{lines in } \mathcal{L} \text{ crossing the cell}\}.
\]

Thus,

\[
\sum_{\text{cells in } C_{\text{good}}} \#\{\text{lines in } \mathcal{L} \text{ crossing the cell}\} + \sum_{\text{cells in } C_{\text{bad}}} \#\{\text{lines in } \mathcal{L} \text{ crossing the cell}\} \leq 2Ld.
\]

The assumption that \(C_{\text{good}}\) is small implies that \(C_{\text{bad}}\) is large: \#\(C_{\text{bad}}\) > \((100 - a)\%\) \#\(\text{cells}\). Observe however that this is impossible; each bad cell is crossed by a lot of lines, so the
existence of this many bad cells would mean that the second term above is much larger than $2Ld$. Indeed,

$$\sum_{\text{cells in } C_{\text{bad}}} \#\{\text{lines in } L \text{ crossing the cell}\} > \#C_{\text{bad}} \cdot A \frac{L}{d^{n-1}}$$

$$> (100 - a)^\% \#\text{cells} \cdot \frac{10^{10}}{(100 - a)^\%c_0} \frac{L}{d^{n-1}}$$

$$\geq (100 - a)^\% c_0 \cdot \frac{10^{10}}{(100 - a)^\%c_0} \frac{L}{d^{n-1}}$$

$$\geq 10^{10} Ld,$$

a contradiction (as it is at most $2Ld$).

□

Proof of Lemma 9.1. Apply Lemma 9.3 for $a$ such that $a^\% + \frac{a^\%}{c_1} > 100\%$. This implies that the set of $C_1$ of cells that each contain a lot of points and the set $C_2$ of cells that are each crossed by few lines are not disjoint. Therefore, there exists at least one cell with both properties (in fact, there exists a definite proportion of the cells satisfying both properties).

□

The final two sections are dedicated to reproving two theorems (the joints theorem and the Szemerédi–Trotter theorem) using polynomial partitioning. In the joints theorem proof, the cellular case will be dealt with by using induction in an average cell. In the Szemerédi–Trotter theorem proof, the cellular case will be dealt with by using a known formula in an average cell. As stated earlier, these are the two possible approaches when dealing with the cellular case.

10. Counting joints using polynomial partitioning

The primary aim of this section is to demonstrate how one can deal with the cellular case using an inductive hypothesis in an average cell. For this to work, however, the numerology of the problem has to be agreeable. And, unfortunately, the numerology in the desired estimate

$$\#J(L) \lesssim (\#L)^{\frac{n}{n-1}}$$

is border-line disagreeable; that is, any larger exponent on the right-hand side would have been agreeable, but $\frac{n}{n-1}$ is not. This will be clearer through the proof. In any case, if one’s mind is set on using induction in an average cell (rather than another known formula), then this is the best that can be proved.\footnote{Using polynomial partitioning, but applying the Szemerédi–Trotter theorem in an average cell, rather than doing induction, gives much better estimates for $\#J(L)$. In fact, one can prove that $\#J(L) \lesssim (\#L)^{\frac{n}{n-1}}$, but also prove refined estimates for $\#J(L)$, according to how many lines of $L$ pass through each joint.}

Theorem 10.1. (Joints theorem with a loss) Let $n \geq 2$ and $\varepsilon > 0$. Then, there exists $C_\varepsilon > 0$, depending only on $\varepsilon$ and the dimension $n$, such that

$$\#J(L) \leq C_\varepsilon (\#L)^{\frac{n}{n-1} + \varepsilon},$$

for any finite family $L$ of lines in $\mathbb{R}^n$.\footnote{Using polynomial partitioning, but applying the Szemerédi–Trotter theorem in an average cell, rather than doing induction, gives much better estimates for $\#J(L)$. In fact, one can prove that $\#J(L) \lesssim (\#L)^{\frac{n}{n-1}}$, but also prove refined estimates for $\#J(L)$, according to how many lines of $L$ pass through each joint.}
Proof. The proof is achieved by induction on \( \#\mathcal{L} \). More precisely, let \( C_\epsilon > 0 \) be a constant that will be specified later.

- The estimate \( \#J(\mathcal{L}) \leq C_\epsilon (\#\mathcal{L})^{\frac{n}{n-1}+\epsilon} \) is trivially true for any family \( \mathcal{L} \) of lines with \( \#\mathcal{L} = 1 \) (in fact, no matter what \( C_\epsilon \) is). Indeed, for any such \( \mathcal{L} \), it holds that \( \#J(\mathcal{L}) = 0 \), as at least \( n \) lines are needed to form a joint in \( \mathbb{R}^n \).

- Assume that the estimate \( \#J(\mathcal{L}) \leq C_\epsilon (\#\mathcal{L})^{\frac{n}{n-1}+\epsilon} \) holds for any family \( \mathcal{L} \) of lines in \( \mathbb{R}^n \), with \( \#\mathcal{L} \leq L - 1 \), for some \( L \geq 2 \).

- It is now proved that the estimate \( \#J(\mathcal{L}) \leq C_\epsilon (\#\mathcal{L})^{\frac{n}{n-1}+\epsilon} \) holds for any family \( \mathcal{L} \) of lines in \( \mathbb{R}^n \) with \( \#\mathcal{L} = L \).

Let \( \mathcal{L} \) be a family of lines in \( \mathbb{R}^n \) with \( \#\mathcal{L} = L \).

For some \( d \geq 1 \) to be specified later, apply polynomial partitioning for \( J(\mathcal{L}) \), with a polynomial of degree at most \( d \). (When the time comes, \( d \) will be specified so that \( \#J(\mathcal{L}) \leq C_\epsilon (\#\mathcal{L})^{\frac{n}{n-1}+\epsilon} \) will hold in the cellular case, allowing us to move on to the algebraic case.)

[Observe here: The point of polynomial partitioning is to reduce to the algebraic case: the case where most joints are in \( Z_p \), for \( \deg p \) considerably lower than expected, i.e. than \( (\#J(\mathcal{L}))/n \). So, for polynomial partitioning to make sense as an approach, we’d better use \( d \ll (\#J(\mathcal{L}))/n \). Otherwise, we have just made our life harder, to get to a worse conclusion than that of Theorem 5.5. Getting back to the proof:]

There exists non-zero \( p \in \mathbb{R}[x_1, \ldots, x_n] \), with \( \deg p \leq d \), such that \( Z_p \) splits \( \mathbb{R}^n \) in \( \sim d^n \) cells, each containing \( \lesssim \#J(\mathcal{L})/d^n \) joints in \( J(\mathcal{L}) \). There exist two possible cases:

**The cellular case:** \( \gtrsim \#J(\mathcal{L}) \) joints in \( J(\mathcal{L}) \) (say, > 1% of the joints in \( J(\mathcal{L}) \)) lie in the union of the cells. Then, by Lemma 9.1, there exists a cell demonstrating agreeable average behaviour; that is,

\[
\#(J(\mathcal{L}) \cap \text{cell}) \sim \frac{\#J(\mathcal{L})}{d^n}
\]

and

\[
\#L_{\text{cell}} \lesssim \frac{L}{d^{n-1}}
\]

where

\[
L_{\text{cell}} := \{ \text{lines in } \mathcal{L} \text{ that cross the cell} \}.
\]

The hope is to use the inductive hypothesis in this cell, in order to derive the desired estimate for the same constant \( C_\epsilon \) as the one used in the inductive hypothesis. For that reason, the implicit constants above cannot be ignored, as they could mess up the inductive constant. Thus, let \( c_1, c_2 > 0 \) be constants, depending only on the dimension, for which

\[
(19) \quad \#(J(\mathcal{L}) \cap \text{cell}) \geq c_1 \frac{\#J(\mathcal{L})}{d^n}
\]

and

\[
(20) \quad \#L_{\text{cell}} \leq c_2 \frac{L}{d^{n-1}}
\]
hold. Now, observe that the joints in $J(L)$ contained in this cell are joints formed by the lines of $L$ that cross the cell; i.e.,

$$J(L) \cap \text{cell} \subseteq J(L_{\text{cell}}).$$

Importantly, assuming that $d > c_2 \frac{n-1}{c_1}$, the bound (20) on $#L_{\text{cell}}$ implies that $#L_{\text{cell}} \leq L$, and therefore the inductive hypothesis

$$#J(L_{\text{cell}}) \leq C \varepsilon (#L_{\text{cell}})^{\frac{n}{n-1}+\varepsilon}$$

holds. By the bounds (19) and (20) above, it follows that

$$c_1 \frac{#J(L)}{d^{n-1}} \sim \frac{#(J(L) \cap \text{cell})}{#J(L_{\text{cell}})} \leq #J(L_{\text{cell}}) \leq C \varepsilon (#L_{\text{cell}})^{\frac{n}{n-1}+\varepsilon} \leq C \varepsilon \left( c_2 \frac{L}{d^n-1} \right)^{\frac{n}{n-1}+\varepsilon}.$$

Observe that the left-hand side and the right-hand side of this inequality are now independent of the cell we were working with; they only depend on $#J(L)$ and $#L$ (the quantities of interest), and the still unspecified $d$. Rearranging, it follows that

$$#J(L) \leq C \varepsilon \cdot c_2 \frac{n^{n-1}+\varepsilon - 1}{c_1} \cdot \frac{d^{n-1}}{L^{n-1}+\varepsilon}.$$

In other words, whenever we partition $J(L)$ with the zero set of a polynomial of degree at most $d$, then (21) holds in the cellular case. Observe that, the larger $d$ is, the better (21) is. And, in fact, if

$$\frac{c_2}{c_1} \frac{n^{n-1}+\varepsilon - 1}{d^{n-1}+\varepsilon} < 1,$$

i.e. if

$$d > \left( c_2 \frac{n^{n-1}+\varepsilon - 1}{c_1} \right)^{\frac{1}{n-1}+\varepsilon},$$

the desired estimate

$$#J(L) \leq C \varepsilon (#L)^{\frac{n}{n-1}+\varepsilon}$$

holds in the cellular case. Therefore, we simply work with the smallest possible $d$ that satisfies the three conditions that have so far been imposed on $d$. That is, we work with

$$d \sim \max \left\{ 1, c_2 \frac{1}{c_1}, \left( c_2 \frac{n^{n-1}+\varepsilon - 1}{c_1} \right)^{\frac{1}{n-1}+\varepsilon} \right\}$$

(for a multiplicative constant larger than 1, than ensures that $d$ is an integer larger than this maximum). Applying polynomial partitioning for this $d$, the desired estimate is proved in the cellular case.

---

2 This is the first restriction imposed on $d$ (apart from the trivial restriction $d \geq 1$).

3 This is exactly how the cellular case is dealt with here; the inductive hypothesis is applied in a single cell of good average behaviour.

4 Observe here that there would be no denominator on the right-hand side if not for the existence of the extra $\varepsilon$ in the exponent. So, the only reason the cellular case can be dealt with here is that the exponent on the right-hand side is strictly larger than $\frac{n}{n-1}$. That is, $\frac{n}{n-1}$ is the magic threshold beyond which the cellular case has a chance of being dealt with using induction. This is true of any problem; observe that we never use in the cellular case that we are dealing with joints (apart from when we actually apply the inductive hypothesis, which of course is expected to hold because we are dealing with joints).

5 This is the last condition imposed on $d$; this is the vital condition so that the joints estimate proved in the cellular case has multiplicative constant equal to the constant in the inductive hypothesis.
The algebraic case: \( \gtrsim \#J(L) \) joints in \( J(L) \) joints in \( J(L) \) (say, > 99% of the joints in \( J(L) \)) lie in \( Z_p \). Importantly, \( \text{deg } p \leq d \), where
\[
 d \sim \max \left\{ 1, c_2^{\frac{1}{n-1}}, \left( c_2^{\frac{n}{n-1}} + \epsilon \right)^{\frac{1}{(n-1)}} \right\},
\]
a constant depending only on \( \epsilon \) and the dimension \( n \).

Just like in the original proof, the above information ensures that the joints on \( Z_p \) (and therefore all the joints) are fewer than \( \text{deg } p \cdot \#L \), which is considerably smaller than \( C \epsilon \left( \#L \right)^{\frac{n}{n-1} + \epsilon} \), completing the proof.

In more detail, the following analogue of Claim 7.4 holds, controlling the size of subsets of \( J(L) \) on which a polynomial vanishes.

**Claim 10.2.** If a non-zero \( p \in \mathbb{R}[x_1, \ldots, x_n] \) vanishes on \( J \subseteq J(L) \), then
\[
 \#J \leq \text{deg } p \cdot L.
\]

The proof of the Claim is not presented here, as it is the same as the proof of Claim 7.4.

Denoting by \( J_{Z_p} \) the set of joints in \( J(L) \) that lie in \( Z_p \), Claim 10.2 immediately implies that
\[
 \#J_{Z_p} \leq \text{deg } p \cdot L \leq d \cdot L,
\]
whence
\[
 \#J(L) \leq c \cdot d \cdot L,
\]
for some \( c \) depending only on the dimension \( n \). Aiming for \( C \epsilon := c \cdot d \) from the start (this is allowed as \( d \) simply depends on the dimension and \( \epsilon \), and therefore so does \( C \epsilon \)), it follows that
\[
 \#J(L) \leq C \epsilon \cdot L \leq C \epsilon (\#L)^{\frac{n}{n-1} + \epsilon}
\]
in the algebraic case as well.

Recall once again that the fact that we were counting joints was not used at all in the cellular case, apart from when the inductive hypothesis was employed (an estimate which, in retrospect, is true for joints, but not necessarily for other sets of points lying in the lines in \( L \)). Therefore:

---

6Observe in how better a situation we are now than in the original joints proof. Then, the best we knew was that the joints were in the zero set of a polynomial of degree \( \lesssim (\#J(L))^{1/n} \). Now, we have reduced to a situation where most of the joints are in the zero set of a polynomial of constant degree, i.e. as low degree as possible.
11. Proving the Szemerédi–Trotter theorem using polynomial partitioning

The primary aim of this section is to demonstrate how one can deal with the cellular case using a known formula in an average cell. When following this approach, the numerology of the problem is not as restrictive (as when trying to deal with the cellular case using induction). However, one needs to be much more imaginative. What formula is the best to use? What degree should one apply polynomial partitioning with? There is much more freedom now.

Recall the Szemerédi–Trotter theorem.

**Theorem 3.1 (Szemerédi–Trotter)** Let \( L \) be a finite set of \( L \) lines in \( \mathbb{R}^2 \). For any \( k \in \mathbb{N} \), let \( S_k \) be the set of points in \( \mathbb{R}^2 \) with the property that each lies in \( \sim k \) lines in \( L \). Then, for all \( k \geq 2 \),

\[
\#S_k \lesssim \frac{L}{k} + \frac{L^2}{k^3}.
\]

Recall that Székely’s proof of the Szemerédi–Trotter theorem was based on the crossing number inequality, which, in turn, was proved in two steps: a simple bound on the crossings was proved in the first step, and it was strengthened in the second step, using a probabilistic argument.

The new proof of the Szemerédi–Trotter theorem presented here (due to Kaplan, Matousek and Sharir in 2011) follows a similar scheme: simple bounds are proved in the first step, and are strengthened in the second step, using polynomial partitioning. (In fact, the cellular case will be dealt with using a simple bound in an average cell).

The simple bounds on point-line incidences required for the proof are the following.

**Lemma 11.1.** For any set \( P \) of points and any family \( L \) of lines, it holds that

\[
I(P, L) \lesssim (\#P)^2 + \#L \quad \text{(23)}
\]

and

\[
I(P, L) \lesssim (\#L)^2 + \#P. \quad \text{(24)}
\]

**Proof.** By definition,

\[
I(P, L) = \#\{(p, l) \in P \times L : p \in l\}.
\]

Therefore,

\[
I(P, L) = \sum_{\ell \in L} \#\{\text{points in } P \text{ on } \ell\} \quad \text{(25)}
\]

and

\[
I(P, L) = \sum_{p \in P} \#\{\text{lines in } L \text{ through } p\}. \quad \text{(26)}
\]
By (25),
\[
I(\mathcal{P}, \mathcal{L}) = \sum_{\ell \in \mathcal{L}} \# \{ \text{points in } \mathcal{P} \text{ on } \ell \}
\]
\[
= \sum_{\ell \in \mathcal{L} \text{ each containing 1 point of } \mathcal{P}} \# \{ \text{points in } \mathcal{P} \text{ on } \ell \}
+ \sum_{\ell \in \mathcal{L} \text{ each containing } \geq 2 \text{ points of } \mathcal{P}} \# \{ \text{points in } \mathcal{P} \text{ on } \ell \}
\leq \# \mathcal{L} + \# \{ \text{pairs of distinct points in } \mathcal{P} \text{ on } \ell \}
\leq \# \mathcal{L} + \left( \# \mathcal{P} \right)^2,
\]
where the second to last inequality is due to the fact that any two distinct points in \( \mathcal{P} \) cannot lie on more than one line, therefore any pair of distinct points in \( \mathcal{P} \) cannot be counted for more than one line in the sum above. Thus, (23) has been proved.

Similarly, by (26),
\[
I(\mathcal{P}, \mathcal{L}) = \sum_{p \in \mathcal{P}} \# \{ \text{lines in } \mathcal{L} \text{ through } p \}
\]
\[
= \sum_{p \in \mathcal{P} \text{ each lying in 1 line of } \mathcal{L}} \# \{ \text{lines in } \mathcal{L} \text{ through } p \}
+ \sum_{p \in \mathcal{P} \text{ each lying in } \geq 2 \text{ lines of } \mathcal{L}} \# \{ \text{lines in } \mathcal{L} \text{ through } p \}
\leq \# \mathcal{P} + \# \{ \text{pairs of distinct lines in } \mathcal{L} \text{ through } p \}
\leq \# \mathcal{P} + \left( \# \mathcal{L} \right)^2,
\]
where the second to last inequality is due to the fact that any two distinct lines in \( \mathcal{L} \) cannot intersect at more than one point, therefore any pair of distinct lines in \( \mathcal{L} \) cannot be counted for more than one point in the sum above. Thus, (24) has been proved. \( \square \)

The Szemerédi–Trotter theorem (which, observe, provides a bound on \( I(S_k, \mathcal{L}) \)) will now be used using polynomial partitioning. As stated earlier, this time the cellular case will be dealt with by using one of the simple bounds above.

**Proof of Theorem 3.7 (S-T) using polynomial partitioning.** Let \( \mathcal{L} \) be a family of \( L \) lines in \( \mathbb{R}^2 \). Observe that

\[
\# S_k \cdot k \sim I(S_k, \mathcal{L}),
\]
since each point in \( S_k \) has \( \sim k \) lines of \( \mathcal{L} \) through it. Therefore, the Szemerédi–Trotter theorem provides a bound on \( I(S_k, \mathcal{L}) \). So, by (23),

\[
\# S_k \cdot k \sim I(S_k, \mathcal{L}) \lesssim (\# S_k)^2 + L.
\]

Thus, if \( (\# S_k)^2 \lesssim L \), it follows that \( \# S_k \cdot k \lesssim L \), i.e.

\[
S_k \lesssim \frac{L}{k},
\]
i.e. the Szemerédi–Trotter theorem is proved.

**Side comment.** Observe how, by (23), the estimate \( \# S_k \lesssim L^{1/2} \) directly implies that \( \# S_k \cdot k \lesssim L \), a much better estimate for \( k \gtrsim L^{1/2} \). The fact that \( \# S_k \lesssim L^{1/2} \) is self-improving actually makes a lot of sense, because, at least morally, it implies that each
line in $\mathcal{L}$ contains $\sim 1$ point of $S_k$. So, essentially the picture is a set of disjoint bushes of lines, one through each point of $S_k$. Therefore, $S_k \cdot k \lesssim L$, i.e. $S_k \lesssim \frac{L}{k}$.

Let us explain this. Try to evaluate $I(S_k, \mathcal{L})$ like this instead (a new proof of (23) actually): we need to add, for every $p \in S_k$, the number of lines through $p$. Let us pretend that each line in $\mathcal{L}$ through each $p$ contains at least $2$ elements of $S_k$. Fix $p \in S_k$, label the lines through $p$; say, $\ell_1, \ldots, \ell_n$. For each $\ell_i$, fix a point $p_i \in S_k$ on that line, different from $p$ (it exists). Then,

$$\# \{\text{lines in } \mathcal{L} \text{ through } p\} = \# \{(p, p_i) : i = 1, \ldots, n\} \leq \# \{\text{pairs of } p \text{ with any point in } S_k \text{ lying in the lines through } p\}.$$ 

Observe that the right-hand side above counts each line through $p$ as many times as the number of elements of $S_k$ on it, while the left-hand side counts it only once. However, adding the right-hand side over all $p$ we are getting at most $(#S_k)^2$, which is $\lesssim L$ by our assumption. Therefore, not much double counting of lines has actually occurred. I.e., it principle each line was counted only $\sim 1$ times by the right-hand side above, which means that each line contains $\sim 1$ points of $S_k$.

By the above analysis, it may be assumed that

$$(#S_k)^2 \gtrsim L.$$ 

This means that the ratio $\frac{(#S_k)^2}{L}$, and all its powers, are numbers $\gtrsim 1$, therefore they could play the role of a degree of a polynomial. Indeed, we apply polynomial partitioning on $S_k$, with a polynomial of degree at most

$$d := A \frac{(S_k)^{2/3}}{L^{1/3}},$$

for a constant $A > 0$ independent of $S_k$ and $\mathcal{L}$, large enough for $d$ to be at least $1$ (such constant exists by the above).

More precisely, for this $d$, there exists non-zero $p \in \mathbb{R}[x_1, \ldots, x_n]$, with $\deg p \leq d$, whose zero set $Z_p$ splits $\mathbb{R}^2$ in $\sim d^2$ cells, each containing $\lesssim \frac{#S_k}{d^2}$ points of $S_k$.

Now, there are two possible cases.

**The cellular case.** $\gtrsim #S_k$ points of $S_k$ lie in the union of the cells. As usual, we simply work with a cell that exhibits good average behaviour. More precisely, by Lemma [9.1] there exists a cell with

$$\#(S_k \cap \text{cell}) \sim \frac{#S_k}{d^2}$$

and

$$\# \mathcal{L}_{\text{cell}} \lesssim \frac{L}{d},$$

where $\mathcal{L}_{\text{cell}}$ is the set of lines in $\mathcal{L}$ that cross the cell.

---

7 This is quite a clever choice of degree that happens to work; some intuition for this choice will be given during the treatment of the cellular case.

8 Since this time we do not work by induction, the precise constants hiding in the $\sim$ symbols can be suppressed.
Each point in $S_k \cap \text{cell}$ has $\sim k$ lines in $\mathcal{L}_{\text{cell}}$ through it (since the lines in $\mathcal{L}$ through the points in $S_k \cap \text{cell}$ are all in $\mathcal{L}_{\text{cell}}$; they have to cross the cell to pass through points in it). Thus,

$$I(S_k \cap \text{cell}, \mathcal{L}_{\text{cell}}) \sim \#(S_k \cap \text{cell}) \cdot k \sim \frac{\#S_k}{d^2} \cdot k.$$ 

Importantly, $I(S_k \cap \text{cell}, \mathcal{L}_{\text{cell}})$ can be controlled by a known formula: the trivial bound [23]. That is,

$$I(S_k \cap \text{cell}, \mathcal{L}_{\text{cell}}) \leq \left(\#(S_k \cap \text{cell})\right)^2 + \#\mathcal{L}_{\text{cell}}.$$ 

By the earlier bounds on the three quantities above, one obtains

$$\frac{\#S_k}{d^2} \cdot k \lesssim \left(\frac{\#S_k}{d^2}\right)^2 + \frac{L}{d}.$$ 

Observe that the two quantities on the right-hand side above are essentially equal for the chosen $d$; indeed\(^9\)

$$\left(\frac{\#S_k}{d^2}\right)^2 \sim \frac{L}{d} \iff d \sim \frac{(\#S_k)^{2/3}}{L^{1/3}}.$$ 

Therefore, [27] becomes

$$\frac{\#S_k}{d^2} \cdot k \lesssim \frac{L}{d}.$$ 

i.e.

$$\#S_k \lesssim d \cdot \frac{L}{k} \sim \frac{(\#S_k)^{2/3}}{L^{1/3}} \cdot \frac{L}{k}$$

which, after rearranging, becomes

$$\left(\#S_k\right)^{1/3} \lesssim \frac{L^{2/3}}{k},$$

i.e.

$$\#S_k \lesssim \frac{L^2}{k^3},$$

the non-trivial Szemerédi–Trotter bound.

**The algebraic case.** \(\gtrsim \#S_k\) points of $S_k$ lie in $\mathbb{Z}_p$. I.e., for the set

$$S_{Z_p} := S_k \cap \mathbb{Z}_p,$$

it holds that $\#S_{Z_p} \sim \#S_k$. Therefore,

$$I(S_{Z_p}, \mathcal{L}) \sim \#S_{Z_p} \cdot k \sim \#S_k \cdot k.$$ 

Therefore, to control $\#S_k \cdot k$, it suffices to control $I(S_{Z_p}, \mathcal{L})$. Now,

$$I(S_{Z_p}, \mathcal{L}) = I(S_{Z_p}, \mathcal{L}_{\mathbb{Z}_p}) + I(S_{Z_p}, \mathcal{L}_{\subset \mathbb{Z}_p}),$$

where

$$\mathcal{L}_{\mathbb{Z}_p} := \{\text{lines in } \mathcal{L} \text{ that do not fully lie in } \mathbb{Z}_p\}$$

and

$$\mathcal{L}_{\subset \mathbb{Z}_p} := \mathcal{L} \setminus \mathcal{L}_{\mathbb{Z}_p} = \{\text{lines in } \mathcal{L} \text{ that fully lie in } \mathbb{Z}_p\}.$$ 

\(^9\)This provides some intuition as to why one would choose this $d$; at this early stage of the proof, there is no reason to believe that one of the two quantities should be favourable relative to the other, therefore one may try assuming that they are essentially the same.
Observe that \( I(S_{Z_p}, \mathcal{L}_{Z_p}) \) can be easily controlled, as each line that is not fully in \( Z_p \) can intersect \( Z_p \) at most \( \deg p \leq d \) times, therefore it can contain at most \( d \) points of \( S_{Z_p} \) (since those points are in \( Z_p \)). That is,
\[
I(S_{Z_p}, \mathcal{L}_{Z_p}) = \sum_{\ell \in \mathcal{L}_{Z_p}} \# \{ \text{points in } S_{Z_p} \text{ on } \ell \} \leq \# \mathcal{L}_{Z_p} \cdot d \leq Ld. (29)
\]
On the other hand, \( I(S_{Z_p}, \mathcal{L}_{\subseteq Z_p}) \) can be controlled due to the fact that \( \mathcal{L}_{\subseteq Z_p} \) is small. Indeed, \( p \) is a polynomial in two variables, therefore the number of lines in \( \mathbb{R}^2 \) on which it identically vanishes is \( \leq \deg p \). In other words, \( Z_p \) contains at most \( \deg p \leq d \) lines. Therefore,
\[
\# \mathcal{L}_{\subseteq Z_p} \leq d. 
\]
So, by the trivial bound \( 10 \) \( (30) \)
\[
I(S_{Z_p}, \mathcal{L}_{\subseteq Z_p}) \leq (\mathcal{L}_{\subseteq Z_p})^2 + \# S_{Z_p} \leq d^2 + \# S_k.
\]
Now, observe that \( d^2 \lesssim \# S_k \). Indeed, recalling the formula for \( d \), it is easy to see that
\[
d^2 \lesssim \# S_k \quad \text{iff} \quad \# S_k \lesssim L^2,
\]
which holds \( (k \geq 2, \text{ so each point in } S_k \text{ lies in at least } 2 \text{ lines in } \mathcal{L}, \text{ hence } S_k \text{ is at most as large as the set of all pairwise intersections of lines in } \mathcal{L}). \]

**Side comment.** Observe that the fact that \( d^2 \lesssim \# S_k \), i.e. that \( d \lesssim (\# S_k)^{1/2} \), means that the degree of the polynomial we used for partitioning is smaller than the degree of the polynomial coming from Theorem 5.5 which vanishes to begin with at all points of \( S_k \). In fact, if we believe the Szemerédi–Trotter theorem, then \( \# S_k \) is considerably smaller than \( L^2 \), so \( d \ll (\# S_k)^{1/2} \). Observe how, once again, polynomial partitioning has allowed us to reduce to a situation where almost all the points we want to count lie in the zero set of a polynomial of degree considerably smaller than expected.

By the above, \( 10 \) \( (30) \) simply becomes
\[
I(S_{Z_p}, \mathcal{L}_{\subseteq Z_p}) \lesssim \# S_k. \quad (31)
\]
Combining \( 10 \) \( (29) \) and \( 10 \) \( (31) \), \( 10 \) \( (28) \) implies that
\[
\# S_k \cdot k \lesssim Ld + \# S_k. \quad (28)
\]
According to which of the two terms on the right-hand side above is dominant, a bound follows for \( \# S_k \cdot k \). More precisely:

— If \( Ld \gtrsim \# S_k \), then \( \# S_k \cdot k \lesssim Ld \), i.e.
\[
\# S_k \cdot k \lesssim L \cdot \frac{(\# S_k)^{2/3}}{L^{1/3}},
\]
which, by rearranging, becomes
\[
(\# S_k)^{1/3} \lesssim \frac{L^{2/3}}{k},
\]
i.e.
\[
\# S_k \lesssim \frac{L^2}{k^3}. \quad (32)
\]

\(^{10}\)It makes sense here to use \( 10 \) \( (24) \) rather than \( 10 \) \( (23) \), as \( 10 \) \( (24) \) features \( (\# \mathcal{L})^2 \), which we know is small in our case; while \( 10 \) \( (23) \) features \( (\# P)^2 \), which is large in our case: it is \( (\# S_k)^2 \).
— — If \( \#S_k \gtrsim Ld \), then \( \#S_k \cdot k \lesssim \#S_k \), so \( k \sim 1 \). Therefore, the trivial upper bound \( L^2 \) for \( \#S_k \) is equal to \( \frac{L^2}{k^3} \) (up to order of magnitude). In other words,

\[
\#S_k \lesssim L^2 \sim \frac{L^2}{k^3}.
\]

The proof of the Theorem is complete.