THE CELLULAR METHOD

In this lecture, we introduce the cellular method as an approach to incidence geometry theorems like the Szemerédi-Trotter theorem. The method was introduced in the paper "Combinatorial complexity bounds for the arrangements of curves and spheres" by Clarkson, Edelsbrunner, Guibas, Sharir, and Welzl (Discrete Comput. Geom. (1990) 5, 99-160). In the next lectures, we will combine ideas from the cellular method with polynomial method.

Our goal here is to describe some of the main ideas of the cellular method as context. We will not give complete proofs but instead sketch the proofs and do heuristic calculations. As a model problem, we consider using the cellular method to prove the Szemerédi-Trotter theorem. (It has many other applications, and we will mention some of them later.)

Suppose that we have a set $\mathfrak{L}$ of L lines in $\mathbb{R}^2$ and we fix a number k in the range $2 \leq k \leq L^{1/2}$. We wish to bound the number of k-rich points $|P_k|$.

The cellular method is a divide-and-conquer strategy. We cut the plane into cells. We use elementary estimates to control the number of k-rich points in each cell, and if the points and/or lines are well-divided among the cells, then we get a stronger estimate by dividing into pieces in this way.

1. Good cell decompositions

Suppose we take d lines in $\mathbb{R}^2$ (not necessarily in $\mathfrak{L}$). The complement of the d lines has $\lesssim d^2$ connected components, which we call cells. If the lines are in general position, the number of cells is $\sim d^2$. Each line enters $\leq d + 1$ cells. So each line of $\mathfrak{L}$ enters only a small fraction of all the cells (at most $\lesssim 1/d$).

In each cell, we will employ a simple counting estimate for k-rich points, which we proved in the first lecture on incidence geometry.

Lemma 1.1. *(Counting bound) If $\mathfrak{L}$ is a set of L lines, and if $L \leq k^2/4$, then $|P_k| \leq 2L/k$.*

Our strategy is to emply the counting bound in each cell and add up the results. If all the lines go through a single cell and all the k-rich points lie in that cell, then we haven't gained anything by our cell division. The divide-and-conquer algorithm works well if each cell is roughly equal. We consider two precise conditions.

A Even distribution of points. We suppose that all the k-rich points are in the open cells, and the number of k-rich points in each open cell is $\leq 10|P_k|/d^2$.

B Even distribution of lines. We suppose that all the k-rich points are in the open cells, and the number of lines of $\mathfrak{L}$ that enter each cell is $\leq 10L/d$.

In each case, if we are able to choose d, we will be able to prove the Szemerédi-Trotter bound $|P_k| \lesssim L^2 k^{-3}$. Let us examine how the argument would go in each case. We let O_i denote the open cells. We let L_i denote the number of lines of $\mathfrak{L}$ which intersect O_i. We let N_i be the number of k-rich points in O_i.

Case A.

Lemma 1.2. *If $d \geq 160Lk^{-2}$, and if condition A. holds, then $|P_k| \leq 8Ldk^{-1}$.*

Proof. We have the following bounds for N_i. By the counting bound, if $L_i \leq (1/4)k^2$, then $N_i \leq 2L_i k^{-1}$. Also, by assumption, $N_i \leq 10|P_k|d^{-2}$ for all i. We call a cell 'big' if $L_i \geq (1/4)k^2$. We can bound $|P_k| = \sum N_i$ in terms of the number of big cells as follows:

$$|P_k| = \sum_i N_i \leq \left(\sum_i 2L_i k^{-1}\right) + (\# \text{ big cells}) \cdot 10d^{-2}|P_k|.$$

We also know that $\sum_i L_i \leq L(d+1)$ because each line enters $\leq d+1$ open cells. We can plug this in to the first term of the right-hand side. Also, we see that the number of big cells is at most $\sum L_i/(k^2/4) \leq 8dLk^{-2}$. Therefore we get the following inequality:

$$|P_k| \leq 4Ldk^{-1} + 80Ld^{-1}k^{-2}|P_k|.$$

If the coefficient $80Ld^{-1}k^{-2}$ is ≥ 1, this inequality is vacuous. But as long as $80Ld^{-1}k^{-2} \leq 1/2$, we can shift the term $80Ld^{-1}k^{-2}|P_k|$ to the other side. Let us assume that $80Ld^{-1}k^{-2} \leq 1/2$. This is equivalent to $d \geq 160Lk^{-2}$. Under this assumption, we see that $|P_k| \leq 8Ldk^{-1}$. $\qquad\qquad\square$

Now suppose we were able to arrange d lines in the plane obeying condition A. for any d. We could choose $d = 160Lk^{-2}$, and we would see that $|P_k| \leq 2000L^2k^{-3}$, the Szemerédi-Trotter bound.

Case B.

This case is similar and even a little easier.

Lemma 1.3. *If $d \geq 40Lk^{-2}$, and if condition B. holds, then $|P_k| \leq 4dLk^{-1}$.*

Proof. Since we have assumed condition B, $L_i \leq 10L/d$. If $10L/d \leq (1/4)k^2$, then we can apply the counting bound to deduce that $N_i \leq 2L_i/k$. Then we see that $|P_k| = \sum_i N_i \leq 2k^{-1}\sum_i L_i \leq 2L(d+1)/k \leq 4dLk^{-1}$. $\qquad\qquad\square$

Now suppose we were able to arrange d lines in the plane obeying condition B. for any d. We could choose $d = 40Lk^{-2}$, and we would see that $|P_k| \leq 200L^2k^{-3}$, the Szemerédi-Trotter bound.

This raises the question, "can we actually find d lines in the plane obeying condition A or B?" We explore this in the next section.

2. ARE THERE GOOD CELL DECOMPOSITIONS?

Can we find d lines obeying condition A? Morally the answer is no. Here is a more precise related question.

Question 2.1. *Given a set of N points in the plane, and an integer $d \leq N^{1/2}$, can we find d lines which cut the plane into cells so that each open cell contains $\leq 1000N/d^2$ points?*

The answer to this question is definitely no. Let γ be a closed strictly convex curve, such as a circle. Pick N points on this curve. Pick $d \sim N^{1/2}$. Consider d lines in the plane. Each line contains ≤ 2 of our points, so only a small fraction of the points are in the lines. More importantly, each line intersects γ in at most 2 points. Therefore, the lines cut γ into $\leq 2d$ pieces. One of those pieces must contain $\geq (N - 2d)/(2d) \gtrsim N^{1/2}$ points of our set. This badly violates our goal. We wanted to find $\sim N^{1/2}$ lines that cut $\mathbb{R}^2$ into cells with $\lesssim 1$ point in each cell – but in fact one of the cells must have $\gtrsim N^{1/2}$ points in it.

Next we ask: Can we find d lines obeying condition B? Morally the answer is yes (although I'm not sure if the answer is literally yes). The main idea is to choose a subset of d random lines from $\mathfrak{L}$. If we do this, a typical edge of the cell decomposition will intersect $\lesssim L/d$ lines. To get a rough idea of what's happening, consider a line $l \in \mathfrak{L}$. For simplicity, let's suppose that it intersects all the other lines of $\mathfrak{L}$ at different points. The intersection points are $L - 1$ points along l. Now we randomly pick d of the L lines - so essentially we randomly pick d of the intersection points. Now the line l is cut into the segments between the selected points. The average number of points in each segment is L/d, and the probability that a given segment has $\geq KL/d$ points falls off exponentially in K. So very few edges intersect more than $1000L/d$ lines of $\mathfrak{L}$. Next we consider the cells of our decomposition. If a cell has ≤ 1000 edges, and each edge intersects $\leq 1000L/d$ lines of $\mathfrak{L}$, then the number of lines of $\mathfrak{L}$ which intersect the cell is $\leq 10^6L/d$. The cell decomposition may have a few cells with > 1000 edges, but these are also pretty rare.

Here's another perspective. Suppose we first choose $d/2$ random lines of $\mathfrak{L}$ and look at the resulting cells. Suppose that one of the cells intersects $> KL/d$ lines of $\mathfrak{L}$ for a very large K. When we choose $d/2$ more random lines of $\mathfrak{L}$, we are very likely to choose one of the lines intersecting this popular cell. The probablility that we will not choose any of these lines is $\leq exp(-cK)$. As we keep adding random lines,

popular cells are likely to be cut down to size. If we make this analysis quantitative, I believe we find that the fraction of cells O_i where $L_i \geq KL/d$ is $\lesssim exp(-cK)$. So this random decomposition nearly obeys condition B.

If the heuristics above are correct, we can arrange that every cell obeys $L_i \leq C(\log L)L/d$, and this implies the S-T estimate up to logarithmic losses. In order to prove the real Szemerédi-Trotter theorem with the cellular method, one has to subdivide the popular cells by adding some line segments. This requires some care, and we don't discuss the details here.

3. GOOD CELL DECOMPOSITIONS IN THREE DIMENSIONS

Having warmed up in two dimensions, now we consider a set $\mathfrak{L}$ of L lines in $\mathbb{R}^3$. We consider d planes in $\mathbb{R}^3$ which typically divide $\mathbb{R}^3$ into $\sim d^3$ cells. Each line can only enter $\leq d+1$ open cells, so each line enters only a small fraction of the cells. If the lines and/or points are evenly distributed among the cells, then we get a good bound for the number of k-rich points. We again consider two precise conditions.

A Even distribution of points. We suppose that all the k-rich points are in the open cells, and the number of k-rich points in each open cell is $\leq 10|P_k|/d^3$.

B Even distribution of lines. We suppose that all the k-rich points are in the open cells, and the number of lines of $\mathfrak{L}$ that enter each cell is $\leq 10L/d^2$.

In either case, we get good bounds for P_k especially if we can choose d.

Lemma 3.1. *If $d \geq 13L^{1/2}k^{-1}$, and if condition A. holds, then $|P_k| \leq 8Ldk^{-1}$.*

Proof. We have the following bounds for N_i. By the counting bound, if $L_i \leq (1/4)k^2$, then $N_i \leq 2L_ik^{-1}$. Also, by assumption, $N_i \leq 10|P_k|d^{-3}$ for all i. We call a cell 'big' if $L_i \geq (1/4)k^2$. We can bound $|P_k| = \sum N_i$ in terms of the number of big cells as follows:

$$|P_k| = \sum_i N_i \leq \left(\sum_i 2L_ik^{-1}\right) + (\# \text{ big cells}) \cdot 10d^{-3}|P_k|.$$

We also know that $\sum_i L_i \leq L(d+1)$ because each line enters $\leq d+1$ open cells. We can plug this in to the first term of the right-hand side. Also, we see that the number of big cells is at most $\sum L_i/(k^2/4) \leq 8dLk^{-2}$. Therefore we get the following inequality:

$$|P_k| \leq 4Ldk^{-1} + 80Ld^{-2}k^{-2}|P_k|.$$

If the coefficient $80Ld^{-2}k^{-2}$ is ≥ 1, this inequality is vacuous. But as long as $80Ld^{-2}k^{-2} \leq 1/2$, we can shift the term $80Ld^{-2}k^{-2}|P_k|$ to the other side. Let us assume that $80Ld^{-2}k^{-2} \leq 1/2$. This is implied by $d \geq 13L^{1/2}k^{-1}$. Under this assumption, we see that $|P_k| \leq 8Ldk^{-1}$. $\qquad\square$

Now suppose that for any $d \geq 1$ we could choose d hyperplanes so that A holds. We would choose $d \sim L^{1/2}k^{-1}$, and then we would get the bound $|P_k| \lesssim L^{3/2}k^{-2}$.

Condition B is similar. If we could find $d = 20L^{1/2}k^{-1}$ planes obeying condition B, then it would again follow that $|P_k| \lesssim L^{3/2}k^{-2}$.

This looks like a promising route towards our target theorem:

Theorem 3.2. *If $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$ with $\leq L^{1/2}$ lines in any plane and $3 \leq k \leq L^{1/2}$, then $|P_k| \lesssim L^{3/2}k^{-2}$.*

4. ARE THERE GOOD CELL DECOMPOSITIONS IN THREE DIMENSIONS?

So we are led to the following question.

Question 4.1. *If $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$ with $\leq L^{1/2}$ lines in a plane, and if $d \leq L^{1/2}$, can we find d planes obeying condition A or condition B?*

We haven't used or mentioned the restriction $\leq L^{1/2}$ lines in any plane so far. We take a moment to see how it may be relevant. Suppose we consider L lines which all lie in a plane π. If we include the plane π among the d planes, then the planes will include all the k-rich points which completely violates condition A or B. Each other plane intersects π in a line (or not at all). So we are now effectively partitioning the point of $P_k \subset \pi$ with d lines. But these d lines only cut π into d^2 pieces, and so an average piece must have $\gtrsim |P_k|d^{-2}$ k-rich points and must meet $\gtrsim Ld^{-1}$ lines, violating either condition.

Now we return to our set of lines $\mathfrak{L}$ with $\leq L^{1/2}$ lines in any plane. Is it possible that with this restriction, we can find a good cell decomposition obeying one of the conditions. The answer is still no. We have the same problems as before with condition A. If P is a set of points on a convex surface like a sphere, then for any cell decomposition with d planes, one of the cells will have $\gtrsim |P|d^{-2}$ points. Also, P could be a set of points on a curve γ. There are many curves γ that intersect every plane in ≤ 10 points - say a typical trefoil knot. If P is a set of points on such a curve, then any cell decomposition with d planes has a cell with $\gtrsim |P|d^{-1}$ points on it.

We also have a problem with condition B. Suppose that P is a set of points on a convex curve γ in $\mathbb{R}^2$, and $\mathfrak{L}$ is $P \times \mathbb{R} \subset \mathbb{R}^3$. Suppose that one plane is transverse to the x_3-axis. This plane intersects the lines in $|P|$ points lying on a convex curve. Each other plane intersects the first plane in a line. All together, the other planes cut the first plane into $\lesssim d^2$ faces. But they cut the convex curve into $\lesssim d$ segments. Therefore, we get a 2-dimensional face of our decomposition which transversely intersects $\gtrsim Ld^{-1}$ lines. Any open cell bordering this face must intersect $\gtrsim Ld^{-1}$ lines of $\mathfrak{L}$. We could also try using planes that are all parallel to the x_3 axis, but there are similar problems, and one of the cells still contains $\gtrsim Ld^{-1}$ lines.

The cellular method works well for incidences of codimension 1 objects, such as planes or spheres in $\mathbb{R}^3$. In this case, we can build an interesting cell decomposition by taking a random subset of the planes or spheres. For objects of codimension > 1, such as lines in $\mathbb{R}^3$, it has been difficult to apply the cellular method (at least directly).

Returning to our question, there are many examples where we cannot cut space into evenly matched cells. It's not clear if these examples share a useful structure or property that we could take advantage of. In our counterexamples, it seems that the points or lines fit onto a nice 2-dimensional surface or 1-dimensional curve. Does that always happen, or is it just wishful thinking?

5. Polynomial cell decompositions

A union of d planes is a special case of an algebraic surface of degree d. The main idea in this chapter is to cut space into pieces with a degree d algebraic surface. Allowing an arbitrary degree d surface instead of just d planes greatly increases our flexibility. (When we pick d planes, we have $3d$ parameters to play with, but when we pick a degree d surface we have $\sim (1/6)d^3$ parameters to play with!) With all this extra flexibility, we can do a much better job of decomposing space into evenly matched cells. On the other hand, if Z is a degree d surface, then a line either lies in Z or intersects Z in $\leq d$ points. Therefore, each line intersects $\leq d+1$ components of the complement of Z – exactly the same bound as if Z was a union of d planes.

Theorem 5.1. *If X is any finite subset of $\mathbb{R}^n$ and d is any degree, then there is a non-zero degree d polynomial P so that each component of $\mathbb{R}^n \setminus Z(P)$ contains $\leq C(n)|X|d^{-n}$ points of X.*

We will prove this theorem next time. The proof is a cousin of finding a degree d polynomial that vanishes at $\sim d^n$ prescribed points, but it uses topology instead of linear algebra.

We should also give a caveat. The theorem does NOT guarantee that the points of X lie in the complement of $Z(P)$. In fact it is possible that $X \subset Z(P)$. There are two extreme cases. If all the points of X lie in the complement of $Z(P)$, then we get optimal equidistribution, and we have a good tool to do a divide-and-conquer argument. If all the points of X lie in $Z(P)$, then we see that $deg(X) \leq d$, and we get a good degree bound on X. Generally, X will have some points in $Z(P)$ and some points in the complement. On one part of X we get a degree bound and on the other part of X we get good equidistribution.