Same Type Lemma

Talk by Dmitrii Zakharov

Notes by Sanjana Das

October 6, 2023

§1 The same type lemma

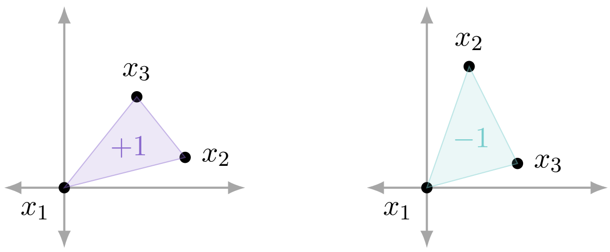
We're working with a set $X \subseteq \mathbb{R}^d$ which is finite and in general position (meaning that there are no $d+1$ points on the same plane). And the information about X that we'll consider is its *order type*, which, informally speaking, keeps track of the orientation of every simplex formed by $(d+1)$ of its points. (For many questions we might be interested in about X, its order type is the only thing we need to know — for example, it tells us whether X is convex.)

To make this more precise, we'll first formalize the definition of orientation.

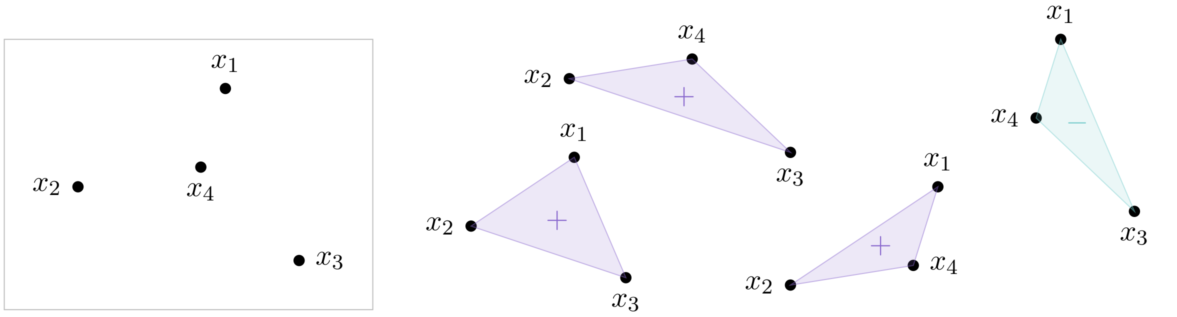
Definition 1.1. The orientation of an ordered $(d+1)$-tuple $(x_1, \ldots, x_{d+1})$ is defined as

$$\operatorname{sgn}\det \begin{bmatrix} | & | & \cdots & | \\ x_2 - x_1 & x_3 - x_1 & \cdots & x_{d+1} - x_1 \\ | & | & \cdots & | \end{bmatrix}.$$

So in words, we're shifting our first point to the origin and looking at the sign of the determinant of the remaining d points. In particular, the orientation is always ± 1.



Definition 1.2. The order type of a point set X, denoted $\mathsf{OT}(X)$, is the 2-coloring of $\binom{X}{d+1}$ with ± 1 where each $(d+1)$-tuple is colored with its orientation.



You might expect $\mathsf{OT}(X)$ to look like a random coloring, but it actually doesn't — it turns out to have a lot of structure. For example, we can consider Ramsey properties — if $d = 2$, then $\mathsf{OT}(X)$ is a 2-coloring of $\binom{X}{3}$ (i.e., the complete 3-uniform hypergraph). And the Erdős–Szekeres theorem on convex subsets of a point set (given n points, we can find roughly $\log n$ of them forming a convex polygon) means this hypergraph contains a monochromatic clique of size $\log |X|$. This is significantly larger than what's guaranteed for an arbitrary 3-uniform hypergraph, which is $\log \log |X|$ (from hypergraph Ramsey).

The same type lemma is a kind of statement of this form (that $\mathsf{OT}(X)$ contains large structured pieces). Specifically, we'll be looking for large *multipartite* structures (rather than cliques).

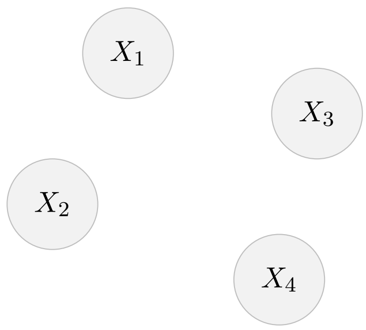
Theorem 1.3 (Same type lemma)

Let $X \subseteq \mathbb{R}^d$ be finite and in general position, and let $m \geq d$. Then there exist $X_1, \ldots, X_m \subseteq X$ with the following properties:

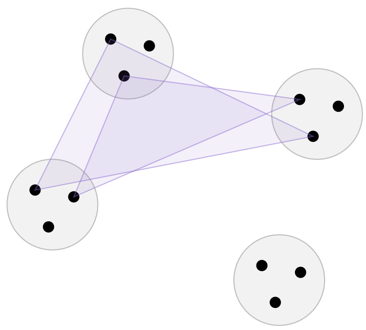
- We have $|X_i| \geq c(m, d)\,|X|$ for all i, where $c(m, d)$ is a constant depending only on m and d.
- The sets X_i are in 'same type position' — for any $d + 1$ indices $i_1, \ldots, i_{d+1}$, the orientations of the $(d + 1)$-tuples $(x_{i_1}, \ldots, x_{i_{d+1}})$ are the same over all choices of $x_{i_1} \in X_{i_1}, \ldots, x_{i_{d+1}} \in X_{i_{d+1}}$.

So in words, we've got a bunch of *linear*-sized blobs such that the orientation of any $(d + 1)$-tuple only depends on which blobs its points come from. (In contrast, if $\mathsf{OT}(X)$ were replaced by an arbitrary coloring, then the Kővári–Sós–Turán theorem would only guarantee *logarithmic*-sized sets X_i.)

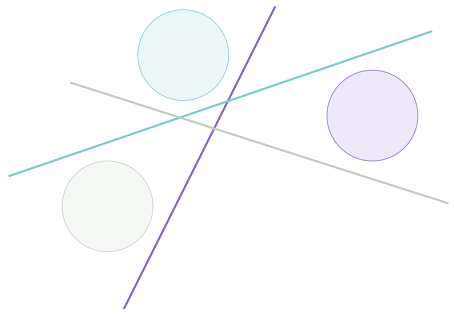
As an example of what such sets $X_1, \ldots, X_m$ could look like in two dimensions, suppose that we take a bunch of small circles at various points in X (which are spaced out reasonably well).



Then if we fix any three of these circles, any three points we choose from them will have the same orientation.



To be more precise, to guarantee that this intuition actually holds, the condition we want on our circles is that we can separate them by lines (i.e., that for any three circles, we can find a line separating any one from the other two).



This ensures that if we move around our points inside their respective blobs, they'll never be on the same line (since if we can separate our sets in this way, then no line intersects all three sets), which means this movement can't flip the orientation.

And so in the same type lemma, our goal is essentially to get a bunch of sets $X_1, \ldots, X_m$ that look like this and all have a lot of points.

§2 A first proof

We'll now give a proof of the same type lemma. Our main focus for today will be the behavior of $c(m, d)$ (the constant in the theorem statement), but for now we won't worry about it — we'll just prove the statement for *some* constant $c(m, d)$.

§2.1 A continuous formulation

First, we're going to replace the statement we're trying to prove (Theorem 1.3) with a *continuous* version. Instead of a set of points $X \subseteq \mathbb{R}^d$, we'll imagine we have a *measure* μ on $\mathbb{R}^d$. We'll suppose that μ has certain reasonable properties — it's a probability measure (i.e., $\mu(\mathbb{R}^d) = 1$), it's absolutely continuous, and it's compactly supported (i.e., it's only nonzero on some finite ball). Now instead of finding large sets X_1, $\ldots$, X_m in same type position, we want to find large chunks of our measure in same type position. So we'll prove the following statement.

Lemma 2.1

For all d and m, given any absolutely continuous, compactly supported probability measure μ on $\mathbb{R}^d$, there exist convex sets $K_1, \ldots, K_m$ such that $\mu(K_i) \geq c(m, d)$ for all i (where $c(m, d)$ is a constant only depending on m and d), and such that for all $(d+1)$-tuples of distinct indices $i_1, \ldots, i_{d+1}$, there is no hyperplane that intersects all of $K_{i_1}, \ldots, K_{i_{d+1}}$.

Note that the condition that no hyperplane intersects all of $K_{i_1}, \ldots, K_{i_{d+1}}$ implies that the sets are in the same type position in the sense of Theorem 1.3, meaning that all choices of $(x_{i_1}, \ldots, x_{i_{d+1}})$ with $x_{i_1} \in K_{i_1}$, $\ldots$, $x_{i_{d+1}} \in K_{i_{d+1}}$ have the same orientation, for the same reason as in the two-dimensional case above — if we imagine wiggling around our points x_{i_j} in their respective sets, they'll never be in the same hyperplane, so they can't flip orientation (more precisely, the corresponding determinant will never be nonzero, so it can't flip sign).

This continuous statement is equivalent to the discrete one — to deduce the discrete statement from the continuous one, we imagine starting with a point set X, and we replace each point in X with a tiny ball to obtain our measure μ.

Then we can use Lemma 2.1 to find convex sets $K_1, \ldots, K_m$ for μ, and we define $X_i = X \cap K_i$ for each i. (There are a few details here — you might need to worry about the case where you have two sets K_i and K_j both chipping away at different parts of the measure corresponding to the same point in X, but this doesn't happen too often.)

For the reverse direction (to deduce the continuous statement from the discrete one), we can imagine starting with μ and sampling a large set of points X according to μ; then we can use the discrete statement to find $X_1, \ldots, X_m$, and we set $K_i = \operatorname{conv}(X_i)$ for each i (so we're taking the convex hulls of our point sets).

Remark 2.2. The fact that the discrete same type condition as in Theorem 1.3 guarantees the desired condition on these sets K_i — that no hyperplane intersects $d+1$ of them — is not obvious. But it can be shown (using facts from convex geometry) that the discrete same type condition on $X_{i_1}, \ldots, X_{i_{d+1}}$ implies that for any subset of these $d+1$ sets, we can separate their convex hulls from those of the others using a hyperplane, and this does mean no hyperplane can intersect all $d+1$ convex hulls.

§2.2 Proof of the continuous version

Now we'll prove the same type lemma in its continuous formulation (Lemma 2.1). We're going to make use of the ham sandwich theorem in this proof.

Theorem 2.3 (Ham sandwich)

For any d sets $S_1, \ldots, S_d$ in $\mathbb{R}^d$, there is a hyperplane H that splits all of them in half (meaning that $\mu(S_i \cap H^+) = \mu(S_i \cap H^-) = \frac{1}{2}\mu(S_i)$ for each i, where H^+ and H^- are the two sides of H).

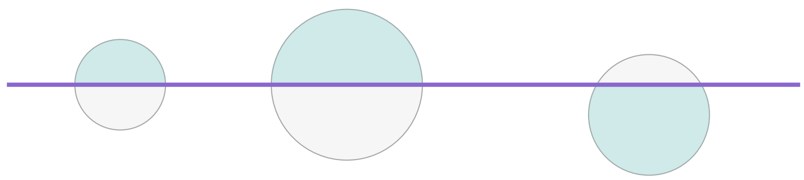
Now the idea is that we'll begin by taking $K_1 = \cdots = K_m = \mathbb{R}^d$, and then iteratively cutting them up to eventually ensure the desired condition (that no hyperplane intersects $d+1$ of them).

We'll iterate over all possible $(d+1)$-tuples of indices $i_1, \ldots, i_{d+1}$, where at each step, we consider the corresponding sets $K_{i_1}, \ldots, K_{i_{d+1}}$ (at the current point in the process), and we want to cut them up to force them to satisfy the condition (that no hyperplane intersects all of them).

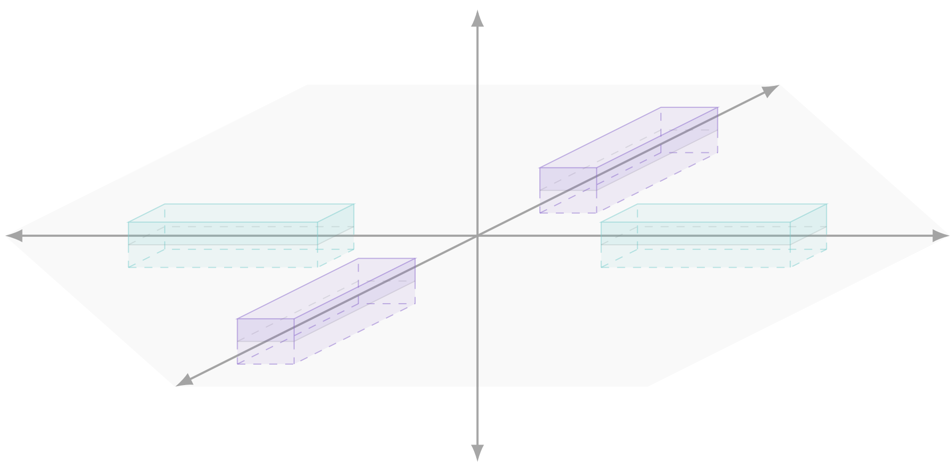
To do so, first, by Ham Sandwich we can find a hyperplane H that halves the first d of these sets — so $\mu(K_{i_j} \cap H^+) = \mu(K_{i_j} \cap H^-) = \frac{1}{2}\mu(K_{i_j})$ for all $1 \le j \le d$. We can't say anything about how H cuts up the last set $K_{i_{d+1}}$, but we *do* know one of the two resulting pieces is at least half of the original — i.e., either $\mu(K_{i_{d+1}} \cap H^+)$ or $\mu(K_{i_{d+1}} \cap H^-)$ is at least $\frac{1}{2}\mu(K_{i_{d+1}})$. And then we replace $K_{i_{d+1}}$ with its bigger half (when cut by H) and each of the other d sets with their halves on the opposite side — so if $\mu(K_{i_{d+1}} \cap H^+) \ge \frac{1}{2}\mu(K_{i_{d+1}})$ then we replace

$$K_{i_{d+1}} \mapsto K_{i_{d+1}} \cap H^+ \quad \text{and} \quad K_{i_j} \mapsto K_{i_j} \cap H^- \text{for all } 1 \le j \le d$$

(and in the opposite case we swap the roles of H^+ and H^-). This at worst halves the sizes of all these sets, and now there's a hyperplane (namely, H) that separates $K_{i_{d+1}}$ from the other d sets.



As a first try, we can imagine repeating this $d+1$ times, once for every cyclic shift of the $d+1$ sets; this ensures that in the end, we have a hyperplane separating each one from the other d. If $d = 2$, this is enough to ensure that no line intersects all three sets. But for larger values of d, this isn't enough — for example, for $d = 3$ (where we have four sets), even if we ensure that each *one* set can be separated from the other three, it isn't necessarily the case that any two sets can be separated from the other two, which is what we need to ensure that no hyperplane intersects all four. For example, consider the following scenario.



Here each of the four sets can be separated from the other three (by a hyperplane parallel to the xz-plane or yz-plane), but the two blue sets can't be separated from the two purple sets. And there does exist a hyperplane that intersects all four (e.g., the xy-plane).

So for $d > 2$, we need to work a bit harder. When we're working with the indices $i_1, \ldots, i_{d+1}$, we also iterate over all possible subsets $J \subseteq \{i_1, \ldots, i_{d+1}\}$, and apply the ham sandwich theorem in the same way to separate the sets with indices in J from the ones with indices not in J. So we find a hyperplane H that halves all but one of the sets K_{i_j}, and then we either choose the upper half-space for all sets in J and the lower half-space for all sets not in J, meaning that we replace

$$K_i \mapsto K_i \cap H^+ \text{ for all } i \in J \quad \text{and} \quad K_i \mapsto K_i \cap H^- \text{ for all } i \notin J,$$

or vice versa (whichever keeps at least half the measure of the last set that H doesn't halve).

So we iterate over all indices $i_1, \ldots, i_{d+1}$ and perform this process for each (where we iterate over all subsets $J \subseteq \{i_1, \ldots, i_{d+1}\}$ and slice up $K_{i_1}, \ldots, K_{i_{d+1}}$ to ensure that the sets with indices in J are separated from those with indices not in J by some hyperplane). And this does ensure that no hyperplane intersects all of $K_{i_1}, \ldots, K_{i_{d+1}}$. (If some hyperplane did intersect all $d+1$ sets, then we'd get $d+1$ points in $\mathbb{R}^{d-1}$ by looking at that hyperplane, and it's a fact that for any such points we can partition them into two groups whose convex hulls intersect. And we're ruling this out by ensuring there's a hyperplane separating those two groups.) So this proves the same type lemma.

§3 A proof with a better constant

In the rest of the talk, we'll give a different proof (from last month) that obtains a better value of the constant $c(m, d)$. We'll redefine $c(m, d)$ as the optimal constant for which the same type lemma holds. Then the previous proof shows that

$$c(m, d) \gtrsim 2^{-\binom{m}{d+1}2^d}$$

(we're iterating over $\binom{m}{d+1}$ $(d+1)$-tuples, and for each we perform 2^d procedures that halve the sets).

But it turns out that $c(m, d)$ actually has *polynomial*, rather than exponential, dependence on m (we think of d as a constant).

Theorem 3.1 (Bukh–Vassilevski 2023)
We have $m^{-d} \gtrsim_d c(m, d) \gtrsim_d m^{-d^2}$.

Remark 3.2. It's easy to see that polynomial dependence is required — we certainly have $c(m, d) \lesssim m^{-1}$, since we need our m sets $X_1, \ldots, X_m$ to be disjoint. But (in addition to proving that polynomial dependence is *sufficient*, by showing $c(m, d) \gtrsim_d m^{-d^2}$) the authors also improved this upper bound to m^{-d}. They did so by taking μ to be the uniform measure on $[0, 1]^d$.

Interestingly, even for the uniform measure on $[0, 1]^d$, we don't know what the optimal partition is — the best bounds we know are the ones in this theorem. To intuitively see where the upper bound of m^{-d} comes from, imagine that we've got m balls of volume m^{-d}, and therefore radius m^{-1} (where both are multiplied by large constants depending on d). Then if we take a random hyperplane, it'll be expected to intersect more than d of these balls; so there must exist *some* hyperplane that intersects $d+1$.

Meanwhile, to get a construction (for the uniform measure on $[0, 1]^d$), we can randomly pick balls of a certain radius; and this gives a bound of m^{-d^2}.

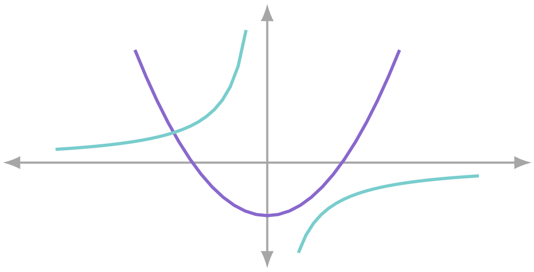
In the rest of this talk, we'll prove the lower bound in Theorem 3.1 (i.e., that $c(m, d) \gtrsim_d m^{-d^2}$).

§3.1 Polynomial partitioning

In the earlier proof, we used the ham sandwich theorem to chop up our sets using hyperplanes. But now we're going to instead use *polynomial partitioning*. The idea is that it's not very efficient to cut up our sets just using hyperplanes, and we can actually do better by cutting them up using the zero sets of polynomials.

Definition 3.3. We define the zero set of a polynomial $f \in \mathbb{R}[x_1, \ldots, x_d]$ as

$$\mathcal{Z}(f) = \{x \in \mathbb{R}^d \mid f(x) = 0\} \subseteq \mathbb{R}^d.$$



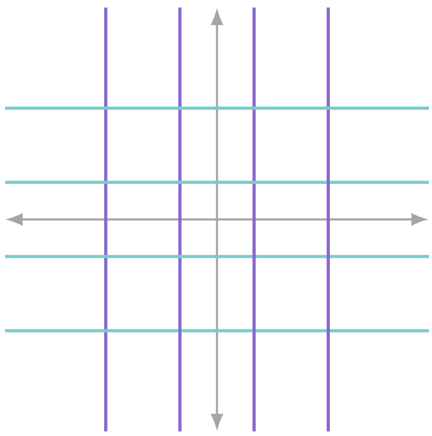
$f(x_1, x_2) = (2x_2 - x_1^2 - 2)(x_1x_2 + 1)$

Then we can think of $\mathcal{Z}(f)$ as cutting up $\mathbb{R}^d$ into several pieces, the connected components of $\mathbb{R}^d \setminus \mathcal{Z}(f)$. And it's a general fact from algebraic geometry that there can't be too many of these connected components.

Lemma 3.4

For any $f \in \mathbb{R}[x_1, \ldots, x_d]$ of degree r, the number of connected components of $\mathbb{R}^d \setminus \mathcal{Z}(f)$ is at most r^d.

As an example, if f is the product of roughly r/d hyperplanes in the horizontal direction, r/d hyperplanes in the vertical direction, and so on (with r/d hyperplanes perpendicular to each of the d axes), then $\mathbb{R}^d \setminus \mathcal{Z}(f)$ will have roughly $(r/d)^d$ connected components. And Lemma 3.4 says that in general, it can't have many more connected components than in this example.



One proof of Lemma 3.4 is to first perturb f a bit so that it has no weird singularities; then each connected component has either a local maximum or minimum point of f. And then we can consider the d partial derivatives of f, each of which is a polynomial of degree at most r — they must all vanish at each of these maxima and minima. But if we have d degree r polynomials over $\mathbb{R}^d$, their number of common intersections is at most r^d by Bezout's theorem — so there are at most r^d critical points, and therefore at most r^d cells.

And the tool we'll need in order to use polynomials to cut up our sets is the *polynomial partitioning lemma*.

Theorem 3.5 (Polynomial partitioning lemma, Guth–Katz)

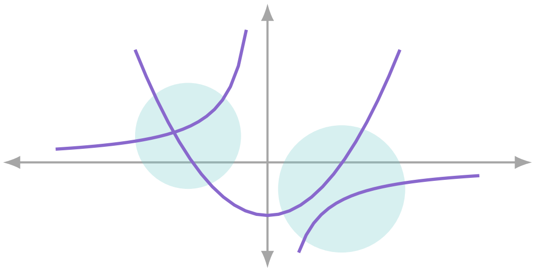
Let μ be a measure on $\mathbb{R}^d$ (which is a probability measure, is absolutely continuous, and compactly supported), and let $r \geq 1$. Then there exists a polynomial f of degree at most r such that for every connected component C of $\mathbb{R}^d \setminus \mathcal{Z}(f)$, we have $\mu(C) \lesssim_d r^{-d}$.

This means we can find a polynomial (of decently low degree) whose connected components split our measure fairly 'evenly' — by Lemma 3.4, $\mathbb{R}^d \setminus \mathcal{Z}(f)$ will have at most r^d connected components, so on average each component will have measure r^{-d}. And Theorem 3.5 says we can guarantee that each piece has measure only a constant multiple of this average. (We'll refer to these connected components as *cells*.)

To prove this, we'll iterate the *polynomial* ham sandwich theorem, which is a version of the ham sandwich theorem (Theorem 2.3) where instead of a hyperplane splitting d measures in half, we find a *polynomial* splitting many more measures in half.

Theorem 3.6 (Polynomial ham sandwich)
For any $t \le \binom{r+d}{d} - 1$ sets $S_1, \ldots, S_t$ in $\mathbb{R}^d$, there is a polynomial $f \in \mathbb{R}[x_1, \ldots, x_d]$ of degree at most r such that for each cell C of $\mathbb{R}^d \setminus \mathcal{Z}(f)$, we have $\mu(S_i \cap C) \le \frac{1}{2}\mu(S_i)$ for each i.

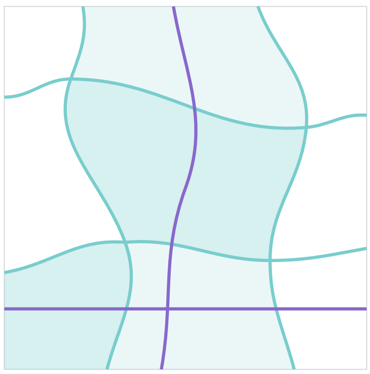
(Here we ask only for each cell to contain *at most* half of each set S_i, rather than *exactly* half, because there may be much more than two cells.)



The $r = 1$ case corresponds to the usual ham sandwich theorem. For large r we have $\binom{r+d}{d} \gtrsim_d r^d$, so this lets us split roughly r^d sets with a polynomial of degree r.

Proof sketch of Theorem 3.5. We're going to iterate the polynomial ham sandwich theorem — we'll start with $\mathbb{R}^d$ and then repeatedly split up our cells until they're not too big.

Suppose that at some point in the process, we have an intermediate polynomial f of degree s, which partitions $\mathbb{R}^d$ into at most s^d cells. Some of these cells might be too large, but there can't be too many of them; so we'll use the polynomial ham sandwich theorem to find a low-degree polynomial g which halves all of them, and replace f with $f \cdot g$.



(Here $\mathcal{Z}(f)$ is drawn in blue, the large cells are shaded in blue, and $\mathcal{Z}(g)$ is drawn in purple.)

If we do this carefully, then the last step of this iteration will be the most costly, and we'll end up with degree r and cell sizes r^{-d} (up to a constant). □

Polynomial partitioning is now a standard tool in incidence geometry. But it turns out to be very fitting for this kind of convex geometry problem as well, as we'll now see.

§3.2 Proof of Theorem 3.1

Now we'll give a second proof of the same type lemma, achieving the better bound stated in Theorem 3.1 (of $c(m,d) \gtrsim m^{-d^2}$ — we're going to write $\gtrsim$ rather than $\gtrsim_d$ from now for convenience, but all implicit constants are allowed to depend on d). We'll again work with the continuous version, where we've got a measure μ on $\mathbb{R}^d$ and we want to find m convex sets $K_1, \ldots, K_m$ in same type position with $\mu(K_i) \gtrsim m^{-d^2}$ for all i.

First, we'll let r be a parameter representing the degree of our polynomial; we'll take r to be a large constant times m^d. We then apply polynomial partitioning to find some polynomial f of degree at most r such that each cell C of $\mathbb{R}^d \setminus \mathcal{Z}(f)$ has small mass, specifically $\mu(C) \lesssim r^{-d}$.

First, since there are at most r^d cells (by Lemma 3.4), this means each has mass at most a constant times their average; and so by an averaging argument, lots of these cells must have mass reasonably close to the average — more precisely, there is some small constant c (depending on d) such that $\mu(C) \geq cr^{-d}$ for $n \gtrsim r^d$ cells C, which we'll call $C_1, \ldots, C_n$.

Now the idea is that we're going to pick m random cells among $C_1, \ldots, C_n$ and take $K_1, \ldots, K_m$ to be their convex hulls.

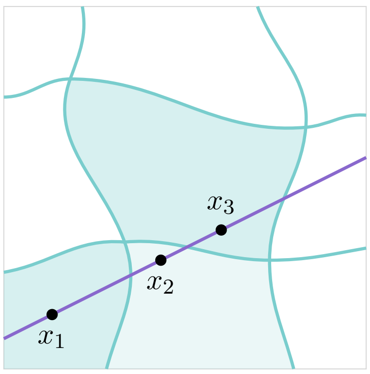
To show that this works, we'll need to estimate the probability that a given $(d+1)$-tuple $(C_{i_1}, \ldots, C_{i_d})$ is 'bad' (meaning that all these cells can be simultaneously intersected by a hyperplane, which is what we're trying to avoid); the next lemma is designed to do that.

> **Lemma 3.7**
>
> The number of tuples $(i_1, \ldots, i_{d+1})$ such that $C_{i_1}, \ldots, C_{i_{d+1}}$ can all be intersected by one hyperplane is at most n^{d+1}/r (up to a constant).

This lemma is the crux of the argument. If we can show this, then it makes sense that the probabilistic argument should work — there's roughly n^{d+1} possible $(d+1)$-tuples, so this means a *random* $(d+1)$-tuple has probability at most $1/r$ of being bad. We're choosing m cells (we'll actually choose a constant factor extra and delete some), so we'll have m^{d+1} possible $(d+1)$-tuples, and the *expected* number of bad $(d+1)$-tuples will be roughly m^{d+1}/r (which is a small constant multiple of m, since r is a large constant multiple of m^d); so we can delete a cell from each and win.

Proof. The idea is that we're going to reduce to Lemma 3.4 — we're going to show that bad $(d+1)$-tuples correspond to cells of a certain polynomial in a higher dimension, and use Lemma 3.4 to say this polynomial can't have too many cells.

If $(C_{i_1}, \ldots, C_{i_{d+1}})$ is bad (meaning that there exists a hyperplane intersecting all $d+1$ cells), then we can choose points $x_1 \in C_{i_1}, \ldots, x_{d+1} \in C_{i_{d+1}}$ which lie on the same hyperplane.



And this means there is an linear dependence between $x_1, \ldots, x_{d+1}$ — more precisely, we can write x_{d+1} as an affine combination of the first d points, meaning that

$$x_{d+1} = a_1x_1 + \cdots + a_dx_d$$

for some reals $a_1, \ldots, a_d$ with $a_1 + \cdots + a_d = 1$. So we can think of $(C_{i_1}, \ldots, C_{i_{d+1}})$ as being 'certified' as bad by the $(2d-1)$-tuple $(x_1, \ldots, x_d, a_1, \ldots, a_{d-1})$ (we don't need to specify a_d because it's determined by $a_1, \ldots, a_{d-1}$ — specifically, $a_d = 1 - a_1 - \cdots - a_{d-1}$).

So we'll now define a polynomial based on this information — we define F as the polynomial

$$F(x_1, \ldots, x_d, a_1, \ldots, a_{d-1}) = f(x_1)\cdots f(x_d)f(a_1x_1 + \cdots + a_{d-1}x_{d-1} + (1 - a_1 - \cdots - a_{d-1})x_d).$$

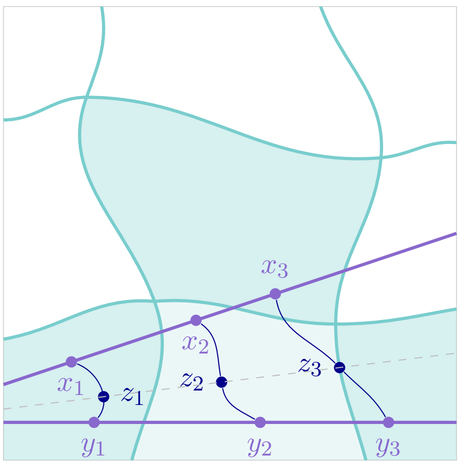
In particular, if we plug in $(x_1, \ldots, x_d, a_1, \ldots, a_{d-1})$ as described above, then this polynomial takes the value $f(x_1)\cdots f(x_d)f(x_{d+1})$, since the last term corresponds to computing x_{d+1} and plugging it in.

Note that F is a polynomial in $d^2 + d - 1$ variables (each x_i is an element of $\mathbb{R}^d$ so has d components, while each a_i is an element of $\mathbb{R}$), and $\deg(F) \lesssim_d r$ (there's only a constant number of factors of f).

Then the key point is the following.

Claim 3.8 — For F defined in this way, the number of bad $(d+1)$-tuples $(C_{i_1}, \ldots, C_{i_{d+1}})$ is at most the number of connected components of $\mathbb{R}^{d^2+d-1} \setminus \mathcal{Z}(F)$.

Proof. Each bad $(d+1)$-tuple $(C_{i_1}, \ldots, C_{i_{d+1}})$ corresponds to some $(x_1, \ldots, x_{d+1})$ as above, and therefore to a point in some cell of $\mathbb{R}^{d^2+d-1} \setminus \mathcal{Z}(F)$. So we just need to show that different bad $(d+1)$-tuples correspond to points in different cells — equivalently, that if we have $(x_1, \ldots, x_{d+1})$ and $(y_1, \ldots, y_{d+1})$ corresponding to two different bad $(d+1)$-tuples, and we connect them by a path, then this path has to run into some point $(z_1, \ldots, z_{d+1})$ corresponding to a zero of F. And the reason this is true is that as we move $(x_1, \ldots, x_{d+1})$ around in a higher-dimensional space, we can imagine each of the points $x_1, \ldots, x_{d+1}$ moving around in $\mathbb{R}^d$ (to $y_1, \ldots, y_{d+1}$, respectively). And since $(x_1, \ldots, x_{d+1})$ and $(y_1, \ldots, y_{d+1})$ correspond to different $(d+1)$-tuples of cells, one of these points z_j is going to have to cross a cell boundary at some point; and then at that point we'll have $f(z_j) = 0$, and therefore F (which is the product of all $f(z_j)$) will be 0 as well.



Explicitly, if the two $(d+1)$-tuples differ in the jth cell (meaning that $i_j \neq i'_j$, where the two $(d+1)$-tuples are $(C_{i_1}, \ldots, C_{i_{d+1}})$ and $(C_{i'_1}, \ldots, C_{i'_{d+1}})$), then we must have $f(z_j) = 0$ for some $(z_1, \ldots, z_{d+1})$ along the path (because as we walk along the path, z_j starts in C_{i_j} and ends in $C_{i_{j'}}$). ☐

This means we can relate the number of bad $(d+1)$-tuples to the number of connected components corresponding to this new polynomial F. And by Lemma 3.4, this number is not too large — we have $\deg(F) \lesssim_d r$, so the number of connected components of $\mathbb{R}^{d^2+d-1} \setminus \mathcal{Z}(F)$ is at most (up to constants) $r^{d^2+d-1} \lesssim n^{d+1}/r$. □

And this finishes — Lemma 3.7 means there's very few bad $(d+1)$-tuples, so we can choose our m sets by sampling cells and deleting these bad tuples.

> **Remark 3.9.** Technically, the way we've written this guarantees that there's no hyperplane intersecting any $d+1$ of our chosen *cells* C_i, but the condition we really want is that no hyperplane intersects their *convex hulls* $K_i = \operatorname{conv}(C_i)$. But it doesn't really make a difference.