1 Points and lines

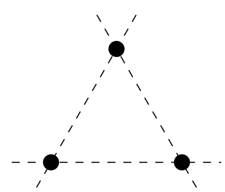
The most basic objects in Euclidean geometry are points and lines. The very first of Euclid's five axioms for plane geometry is about the fundamental relationship between points and lines: any two points define a (unique) line.

In 1893—about 2200 years after Euclid—Sylvester asked the following question, which could easily have appeared in the first book of Euclid's *Elements*.

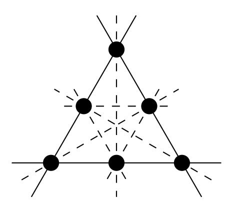
Question. *Is it possible to place finitely many non-collinear points in the plane so that whenever a line passes through two of them, it also passes through a third?*

Note that the non-collinearity is important, since $n \geq 3$ points on a line certainly satisfy this property. Additionally, the requirement that we use finitely many points is necessary, as you'll show on the homework.

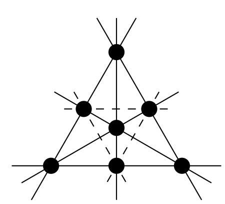
We can just try to place points in the plane so that every line through two of them goes through a third. Since we want them to not all be collinear, we can start with just a triangle of three points:



Of course, this doesn't yet satisfy our desired property, since all three lines only contain two points. We can fix this by adding a point on each of these lines, which I'll now draw as solid to indicate that they now have three points on them.



We fixed three lines, but created six new bad lines, which seems like negative progress. But we can actually fix three of them by placing another point in the center of the triangle!



Though we still have three bad lines... We could of course fix them by placing three more points, but that would create some further bad lines.

If you keep experimenting in this way, you might come to be convinced that the answer to Sylvester's question is no, and that any finite set of non-collinear points in the plane has a line passing through exactly two of them. This is indeed true, and is known as the "Sylvester–Gallai" theorem, even though it was first proved by Melchior in 1941.

Theorem 1.1 ("The Sylvester–Gallai Theorem"). *Any finite set of non-collinear points in the plane define a line containing exactly two of them.*

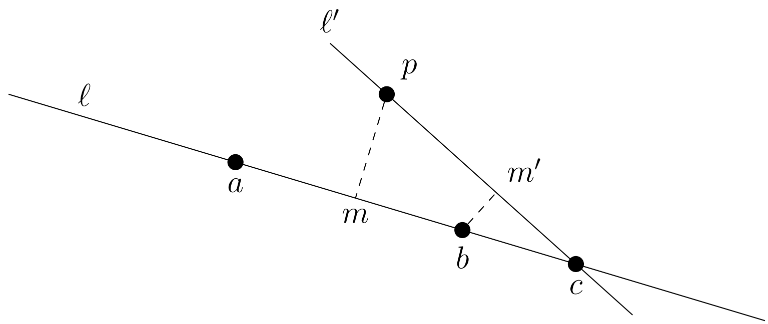
Proof (due to Kelly). Suppose for contradiction that this is false, and let P be the set of points, and let L be the set of lines they define. For a line ℓ and a point p not on ℓ, we let $d(p, \ell)$ denote the distance from p to ℓ, i.e. the length of the segment passing through p and orthogonal to ℓ.

We begin by picking a pair (p, ℓ) of a point and a line not through p whose distance is minimal. Formally, we consider the set of pairs

$$\{(p, \ell) : p \in P, \ell \in L, p \text{ is not on } \ell\}.$$

By the assumption of non-collinearity, this set is non-empty. Moreover, since both P and L are finite, this set is finite. So we can pick p, ℓ with p not on ℓ so that $d(p, \ell)$ is minimized. By assumption, ℓ contains at least three points of P, say a, b, c, and say they appear in this order. Drop a perpendicular from p to ℓ, and say it meets ℓ at the point m.

By the pigeonhole principle, at least two of the points a, b, c must be on the same side of the perpendicular from p. Say that these are b, c. Then the pair of points $p, c \in P$ define another line $\ell' \in L$.



We claim that $d(b, \ell') < d(p, \ell)$, which contradicts the fact that we chose (p, ℓ) to have minimal distance. So to finish the proof, it suffices to prove this claim, which is hopefully pretty intuitive from the picture above.

To prove it rigorously, note that pmc and $bm'c$ are both right triangles. Moreover, these two right triangles are similar, since they have the same angle at c. Finally, note that the length of the segment bc is at most the length of mc, and the length of mc is less than the length of pc since mc is a leg of the right triangle pmc with hypotenuse pc. So the length of bc is less than the length of pc, which shows that the triangle $bm'c$ is smaller than the triangle pmc, where by smaller I mean that the constant of similarity is less than 1. But this implies that the length of bm' is strictly less than the length of pm, as claimed. □

One important consequence of the Sylvester–Gallai theorem is the following theorem, due to de Bruijn and Erdős from 1948.

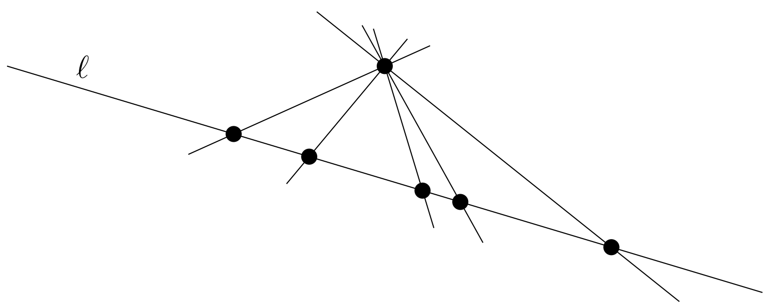
Theorem 1.2. *Any $n \geq 3$ non-collinear points in the plane define at least n lines.*

Proof. We proceed by induction on n. The base case is $n = 3$; in this case, three non-collinear points define a triangle, and thus three lines.

For the inductive case, suppose we have a set P of n points in the plane. By the Sylvester–Gallai theorem, there are two points, say $a, b \in P$, whose line contains no other point of P. We delete a from the configuration to obtain a new set of $n - 1$ points, which we call P'.

If P' is non-collinear, then by the inductive hypothesis it defines at least $n - 1$ lines. Moreover, the line through a, b is not among these, since that line passes through b and no other point of P'. Therefore, P defines at least one more line than P' does, and therefore P defines at least $(n - 1) + 1 = n$ lines.

On the other hand, if P' is collinear, then there is some line ℓ containing all the points of P'. The point a must not lie on ℓ, by our assumption that P was non-collinear. For every point p in P', there is a distinct line passing through a and p, which yields $|P'| = n - 1$ lines. Together with the line ℓ, we get n lines in total. □

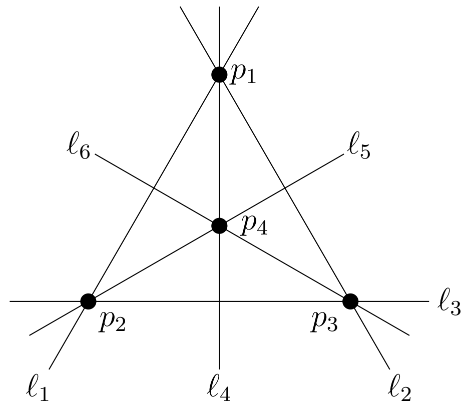


One extremely interesting topic, which you can explore on the homework, is the fact that there are other notions of geometry where statements like the Sylvester–Gallai and de Bruijn–Erdős theorem make sense. In some of these more abstract "geometries", the Sylvester–Gallai theorem is simply false. However, the de Bruijn–Erdős theorem is *always* true; it captures a really fundamental property of geometry. In fact, one can prove the de Bruijn–Erdős theorem simply from the axiom that every two points define a unique line, but the Sylvester–Gallai theorem cannot be proven in this way.

Here is an alternative proof of the de Bruijn–Erdős theorem, which does not use Sylvester–Gallai, or any other special properties of *Euclidean* geometry. This proof uses some linear algebra. On the homework, you can go through the de Bruijn and Erdős's original proof, which does not use linear algebra.

Alternative proof of Theorem 1.2. Let $p_1, \ldots, p_n$ be a set of n points in the plane, and let the lines they define be $\ell_1, \ldots, \ell_m$. We wish to prove that $m \geq n$. For every $1 \leq i \leq n$, we define a vector $v^{(i)} \in \mathbb{R}^m$ as follows. For every $1 \leq k \leq m$, the kth coordinate of the vector $v^{(i)}$ is 1 if p_i lies on the line ℓ_k, and is 0 otherwise.

For example, consider the following configuration of $n = 4$ and $m = 6$ lines.



Then the four vectors $v^{(1)}, v^{(2)}, v^{(3)}, v^{(4)}$ are

$$v^{(1)} = \begin{pmatrix}1 & 1 & 0 & 1 & 0 & 0\end{pmatrix} \qquad v^{(2)} = \begin{pmatrix}1 & 0 & 1 & 0 & 1 & 0\end{pmatrix}$$
$$v^{(3)} = \begin{pmatrix}0 & 1 & 1 & 0 & 0 & 1\end{pmatrix} \qquad v^{(4)} = \begin{pmatrix}0 & 0 & 0 & 1 & 1 & 1\end{pmatrix}$$

For instance, the first coordinate in $v^{(1)}$ is 1 since p_1 lies on ℓ_1, but the first coordinate of $v^{(3)}$ is 0 since p_3 does not lie on ℓ_1.

To prove that $m \geq n$, we will show that the vectors $v^{(1)}, \ldots, v^{(n)}$ are linearly independent. Since they are vectors in $\mathbb{R}^m$, this immediately implies that $n \leq m$, since any collection of linearly independent vectors in $\mathbb{R}^m$ can have at most m elements. If you've never seen linear algebra, the rest of the proof (as well as the previous two sentences) will likely not make a lot of sense, which is OK!

We first claim that every vector $v^{(i)}$ has at least two 1s in it. Indeed, the number of 1s in $v^{(i)}$ is simply the number of lines containing the point p_i. There is at least one 1 in $v^{(i)}$ since p_i lies on at least one line. Moreover, if there is only one 1 in $v^{(i)}$, then p_i lies on a unique line. But this means that there is a single line containing p_i and p_j for every $j \neq i$, meaning that all the points are collinear, a contradiction. So each $v^{(i)}$ has at least two 1s. This, in turn, implies that $v^{(i)} \cdot v^{(i)} \geq 2$ for every i. Indeed, the dot product $v^{(i)} \cdot v^{(i)}$ is simply the sum of the squares of all the coordinates of $v^{(i)}$; since the coordinates of $v^{(i)}$ are just 0 and 1, and there are at least two 1s, this dot product is at least 2.

On the other hand, for every $i \neq j$, we claim that $v^{(i)} \cdot v^{(j)} = 1$. Indeed, when computing $v^{(i)} \cdot v^{(j)}$, we add up the product of the kth coordinate of $v^{(i)}$ and the kth coordinate of $v^{(j)}$, for all $1 \leq k \leq m$. The only way the product of the kth coordinates will be non-zero is if the kth coordinate of both $v^{(i)}$ and $v^{(j)}$ is 1. Therefore, $v^{(i)} \cdot v^{(j)}$ computes the number of lines ℓ_k containing both p_i and p_j. Since every pair of points lies on a unique line, this is exactly 1, so $v^{(i)} \cdot v^{(j)} = 1$ as claimed.

To prove that $v^{(1)}, \ldots, v^{(n)}$ are linearly independent, suppose that there are real numbers $c_1, \ldots, c_n$ such that $c_1 v^{(1)} + \cdots + c_n v^{(n)} = \vec{0}$, where $\vec{0}$ is the zero vector in $\mathbb{R}^m$, and assume for contradiction that not all the c_i equal 0. Taking the dot product of this equation with $v^{(1)}$, we find that

$$0 = c_1(v^{(1)} \cdot v^{(1)}) + c_2(v^{(2)} \cdot v^{(1)}) + \cdots + c_n(v^{(n)} \cdot v^{(1)}) = c_1(v^{(1)} \cdot v^{(1)} - 1) + (c_1 + \cdots + c_n),$$

using the fact that $v^{(1)} \cdot v^{(j)} = 1$ for all $j \geq 2$. Similarly, by taking the dot product with $v^{(i)}$, we find that

$$0 = c_i(v^{(i)} \cdot v^{(i)} - 1) + (c_1 + \cdots + c_n)$$

for all i. Recall that $v^{(i)} \cdot v^{(i)} \geq 2$, so $v^{(i)} \cdot v^{(i)} - 1$ is positive. If $c_1 + \cdots + c_n = 0$, then this shows that $c_i = 0$ for all i, which is a contradiction. If $c_1 + \cdots + c_n > 0$, then we find that $c_i < 0$ for all i, which contradicts the fact that their sum is positive. Similarly, if $c_1 + \cdots + c_n < 0$, then we find that $c_i > 0$ for all i, again a contradiction. □

2 Higher dimensions

While the most well-known geometry in Euclid's *Elements* is planar, Euclid nonetheless devotes a few books to three-dimensional geometry. Are there analogues of the Sylvester–Gallai theorem and the de Bruijn–Erdős theorem in higher dimensions?

There are actually several ways of interpreting this question (depending on what you mean by "analogues"). For the Sylvester–Gallai theorem, the most natural high-dimensional analogue is the following statement, which is indeed true.

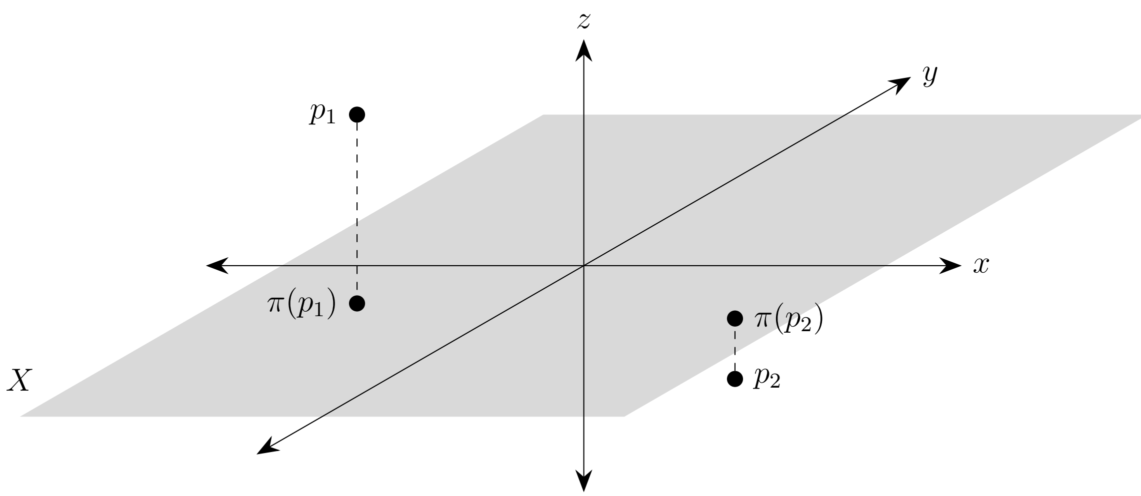
Theorem 2.1 (High-dimensional Sylvester–Gallai). *For every $d \geq 2$ and $n \geq 3$, there is no set of n non-collinear points in $\mathbb{R}^d$ such that every line passing through two of them also passes through a third.*

You'll be asked to prove this on the homework. Once we have Theorem 2.1, we can prove the following high-dimensional de Bruijn–Erdős theorem, using the exact same inductive argument as in our proof of Theorem 1.2.

Theorem 2.2 (High-dimensional de Bruijn–Erdős). *For every $d \geq 2$ and $n \geq 3$, any set of n non-collinear points in $\mathbb{R}^d$ defines at least n lines.*

Although this can be proved from Theorem 2.1 by induction on n, there's actually a totally different proof that deduces the case for general d from the case $d = 2$.

Proof of Theorem 2.2. Consider a set P of n non-collinear points in $\mathbb{R}^d$, and let X denote the two-dimensional xy plane in $\mathbb{R}^d$. Let $\pi : \mathbb{R}^d \to \mathbb{R}^2$ denote the projection onto X. Formally, recall that every point of $\mathbb{R}^d$ is a vector with d coordinates; then π is just the function that takes in a vector and throws away every coordinate except for the first two. Let $\pi(P)$ be the subset of $\mathbb{R}^2$ gotten by applying π to every point in P.



Let's imagine for a moment that we got extremely lucky. Specifically, let's assume that no pair of points in P are mapped to the same point under π. Additionally, let's assume that every triple of non-collinear points in P is also mapped to a triple of non-collinear points by π. If we really got this lucky, then $\pi(P)$ is a set of n non-collinear points in $\mathbb{R}^2$. Additionally, there is a bijection between the lines defined by P and the lines defined by $\pi(P)$. By the two-dimensional de Bruijn–Erdős theorem, Theorem 1.2, $\pi(P)$ defines at least n lines, which implies that P defines at least n lines as well.

However, we obviously might not get so lucky. The trick to get around this is to pick X to be a *random* two-dimensional plane in $\mathbb{R}^d$, rather than the specific xy plane. If we do this, then we will get lucky (in the sense above) with 100% probability. Proving this probabilistic statement is actually not so easy, and requires some background in measure theory, but I hope it's intuitively reasonable. Indeed, there are only finitely many "problems" that can arise: there are $\binom{n}{2}$ pairs of points that might collide under π, and at most $\binom{n}{3}$ non-collinear triples that can be made collinear by π. But we have (uncountably) infinitely many choices for the random plane X, so there's a 0% probability that we'll run into one of the finitely many problems.

By making this argument rigorous, we can in particular find *some* plane X where we get lucky (indeed, there will be infinitely many such planes). If we project onto this plane X, then the argument above works: $\pi(P)$ consists of n non-collinear points in $\mathbb{R}^2$, which define at least n lines by Theorem 1.2, which then implies that P defines at least n lines since there is a bijection between the lines made by P and those made by $\pi(P)$. □

This proof shows that the high-dimensional statement Theorem 2.2 is true, but it actually tells us something more: this is a *fundamentally* two-dimensional statement. Even though we are dealing with points and lines in $\mathbb{R}^d$, all the real mathematical structure comes from what happens when $d = 2$.

Can we come up with some "genuinely d-dimensional" version of the de Bruijn–Erdős theorem? One natural thing to try is to use the new structure that exists in higher dimensions, namely planes, hyperplanes, and so on.

Recall that in $\mathbb{R}^d$, we have *hyperplanes* of every dimension $0 \le i \le d-1$. Zero-dimensonal hyperplanes are just points, one-dimensional hyperplanes are just lines, two-dimensional hyperplanes are our usual notion of planes. We don't have special words for higher-dimensional hyperplanes, simply because our brains can't really visualize what an 11-dimensional hyperplane of $\mathbb{R}^{24}$ looks like. But these are perfectly natural and well-defined mathematical notions.

The main fact that we will need to know about hyperplanes is that for every $1 \le i \le d-1$, every set of $i+1$ points in $\mathbb{R}^d$ defines a *unique* i-dimensional hyperplane, unless this set of $i+1$ points lies in some $(i-1)$-dimensional hyperplane. This generalizes the well-known fact ($i = 2$) that any three points define a unique plane, unless they are collinear (i.e. contained in some 1-dimensional hyperplane). This also generalizes the fact that any two distinct points define a unique line, which is just the case $i = 1$.

It is natural to expect that there is a higher-dimensional de Bruijn–Erdős theorem, which tells us something about the number of i-dimensional hyperplanes defined by a set of n points in $\mathbb{R}^d$. Let's try to come up with what such a statement might be.

First, we must assume that our set of n points is "genuinely d-dimensional". For instance, if we take n points in $\mathbb{R}^{100}$ that all lie on a two-dimensional plane, then they will not define any three-dimensional hyperplanes, nor any four-dimensional hyperplanes, nor any 99-dimensional hyperplanes. Of course, we already encountered essentially this same issue: in the de Bruijn–Erdős theorem, we assumed that our n points were non-collinear, i.e. "genuinely 2-dimensional".

Definition 2.3. A set P of n points in $\mathbb{R}^d$ is called *genuinely d-dimensional* if P is not contained in any $(d-1)$-dimensional hyperplane.

So from now on, let's assume that we have a set P of n points in $\mathbb{R}^d$, which is genuinely d-dimensional. This assumption automatically implies that P defines at least one line, at least one plane, at least one three-dimensional hyperplane, and so on. Again, this is the natural higher-dimensional analogue of our non-collinearity assumption in Theorem 1.2.

Let $F_i(P)$ denote the number of i-dimensional hyperplanes[1] defined by P. To get intuition for what type of results to expect, let's do some examples.

1. Let P consist of the four vertices of a tetrahedron in $\mathbb{R}^3$. They define 6 lines and four planes (the edges and faces of the tetrahedron, respectively), so
$$F_0(P) = 4 \qquad F_1(P) = 6 \qquad F_2(P) = 4.$$
2. Let P consist of the eight vertices of a cube in $\mathbb{R}^3$, so $F_0(P) = 8$. We again have that every pair of points defines a unique line, so $F_1(P) = \binom{8}{2} = 28$. Counting the planes is a bit trickier: there are six faces of the cube, six additional planes containing four points (namely two opposite edges of the cube), and finally eight more planes containing only three points (there are two such planes orthogonal to any "long diagonal" of the cube). So in total,
$$F_0(P) = 8 \qquad F_1(P) = 28 \qquad F_2(P) = 20.$$

[1]The notation F_i is standard, since some people prefer the word "flats" to "hyperplanes".

3. Suppose we add the center of the cube to the previous set of points. Then we actually generate no new lines, since every line containing the center of the cube and one of its vertices passes through another vertex, and so was already counted. Similarly, one can check that we generate no new planes in this way. So for this example,

$$F_0(P) = 9 \qquad F_1(P) = 28 \qquad F_2(P) = 20.$$

4. The simplest genuinely d-dimensional set of points we can place in $\mathbb{R}^d$ is the vertices of the *simplex*, which is the d-dimensional analogue of a triangle, tetrahedron, etc. It consists of $d+1$ points such that no triple is collinear, no quadruple is coplanar, and in general, no set of $i+1$ points lies on an $(i-1)$-dimensional hyperplane. Concretely, we can take our points to be

$$(0,0,0,\dots,0), \quad (1,0,0,\dots,0), \quad (0,1,0,\dots,0), \quad (0,0,1,\dots,0), \quad (0,0,0,\dots,1).$$

Then $F_0(P) = d+1$, since we have $d+1$ points. Every pair of points defines a unique line, so $F_1(P) = \binom{d+1}{2} = \frac{d^2+d}{2}$. Similarly, every triple of points defines a unique plane, so $F_2(P) = \binom{d+1}{3}$. More generally,

$$F_i(P) = \binom{d+1}{i+1} \qquad \text{for every } 0 \le i \le d-1.$$

Thus, the sequence F_i is simply the $(d+1)$th row of Pascal's triangle, without the 1 at the beginning and the end.

So far, all the examples we've seen do have some common behavior. As guaranteed by the de Bruijn–Erdős theorem, for instance, we always have $F_0(P) \le F_1(P)$. Additionally, in all the three-dimensional examples, we had $F_1(P) \ge F_2(P)$. Additionally, we know that the binomial coefficients $\binom{d+1}{i+1}$ are increasing until $i+1 = \lfloor \frac{d+1}{2} \rfloor$, and then are decreasing starting from $i+1 = \lceil \frac{d+1}{2} \rceil$. This is consistent with our three-dimensional examples, which are increasing from F_0 to F_1, and then decreasing from F_1 to F_2. It is natural to conjecture that this pattern continues, namely that for odd d,

$$F_0(P) \le F_1(P) \le \cdots \le F_{\frac{d-1}{2}}(P) \ge F_{\frac{d+1}{2}}(P) \ge \cdots \ge F_{d-2}(P) \ge F_{d-1}(P),$$ (🤔)

and a similar thing for even d. However, the next example shows that this conjecture is too ambitious.

5. Let P be a set of n points in *general position* in $\mathbb{R}^d$. General position means that no triple of points in P is collinear, no quadruple in P is coplanar, and so on: no set of $i+1$ points in P lie on an $(i-1)$-dimensional hyperplane. It turns out that for every $n \ge d+1$, there is a genuinely d-dimensional set in general position. We explicitly constructed such a set—the simplex—for $n = d+1$, but it turns out that it's not so easy to explicitly construct such sets for arbitrary $n \ge d+1$. Nonetheless, we can again do it randomly! For example, we can let P consist of n randomly chosen points

in the unit ball in $\mathbb{R}^d$. Just as in our proof of Theorem 2.2, there are only finitely many "bad events" (e.g. there are only $\binom{n}{3}$ triples that might be collinear, $\binom{n}{4}$ quadruples that might be coplanar, etc.). Because of this, with 100% probability, we will get lucky and obtain a set in general position in this way.

If we have a set P in general position, then $F_i(P) = \binom{n}{i+1}$ for every $0 \le i \le d-1$. Indeed, every set of $i+1$ points will define an i-dimensional hyperplane, and the general position assumption implies that this hyperplane will not contain any other point in P. Because of this, we will get $\binom{n}{i+1}$ distinct hyperplanes.

In particular, if $n \ge 2d$, then the sequence $\binom{n}{i+1}$ will simply be increasing for all i, i.e. we'll have

$$F_0(P) < F_1(P) < F_2(P) < \cdots < F_{d-1}(P).$$

This demonstrates that the earlier conjecture can't be true in general. Similarly, if $d+1 < n < 2d$, then the sequence will eventually start decreasing, but it won't happen at the midpoint.

Nonetheless, there is a reasonable conjecture we can salvage out of this data. Namely, it seems that our (false) conjecture (🤔) can only fail "upwards": if (🤔) is false, then it's because the sequence $F_i(P)$ doesn't start decreasing "when it's supposed to", at the halfway mark. This behavior is sometimes called *top-heavy* behavior, since there's "more stuff at the top than at the bottom" of the sequence.

Many people over the years observed that the naive conjecture is false, but that it seems to only be false in one way, which caused them to formulate a number of more refined conjectures. Maybe the most natural is the following, known as *Rota's unimodality conjecture.*

Conjecture 2.4 (Rota, 1971)**.** *Let P be a genuinely d-dimensional set of n points in $\mathbb{R}^d$. Then there exists some $0 \le m \le d-1$ such that*

$$F_0(P) \le F_1(P) \le \cdots \le F_{m-1}(P) \le F_m(P) \ge F_{m+1}(P) \ge \cdots \ge F_{d-1}(P).$$

In other words, Rota's conjecture says that the sequence $F_i(P)$ goes up for a while, and then starts coming down. Note that if $m = d-1$, then it will actually never start coming down, which is the behavior we saw in our last example.

Despite 50 years of effort, and despite it being such a simple and basic problem, Rota's conjecture remains open. Nonetheless, the past decade has seen a flurry of activity on a number of related problems; essentially, results of the same form have been proved in a number of closely related contexts, though we still don't know how to prove such a result for point sets in $\mathbb{R}^d$.

Note that Rota's conjecture doesn't say anything about the "upward failure" of (🤔) that I alluded to above. That behavior was observed by Dowling and Wilson, who formulated the following *top-heavy conjecture.*

Conjecture 2.5 (Dowling–Wilson, 1974)**.** *Let P be a genuinely d-dimensional set of n points in $\mathbb{R}^d$. Then for every $i < j$ with $i + j \le d - 1$, we have that $F_i(P) \le F_j(P)$. In particular, we have that*

$$F_0(P) \le F_1(P) \le \cdots \le F_{\lfloor \frac{d-1}{2} \rfloor}(P)$$

and

$$F_i(P) \le F_{d-1-i}(P)$$

for every $0 \le i \le \left\lfloor \frac{d-1}{2} \right\rfloor$.

Note that for $d = 2$, the de Bruijn–Erdős theorem proves both Rota's conjecture and the Dowling–Wilson conjecture. Thus, both conjectures are natural generalizations of the de Bruijn–Erdős theorem to higher dimensions.

Unlike Rota's conjecture, which is still open, the Dowling–Wilson conjecture was proved very recently, by Huh and Wang. Even more recently, Braden, Huh, Matherne, Proudfoot, and Wang gave a new proof of a more general theorem, which I won't state; but roughly speaking, it proves the Dowling–Wilson conjecture in the most general possible setting (of so-called *matroids*), which form a vast generalization of point sets in $\mathbb{R}^d$.

Theorem 2.6 (Huh–Wang 2017, Braden–Huh–Matherne–Proudfoot–Wang 2020)**.** *Conjecture 2.5 is true (as are many generalizations of it).*

One astonishing thing about these proofs is their complexity. The conjecture is about the most basic objects in Euclidean geometry: points, lines, planes, etc. Moreover, the $d = 2$ case of this conjecture, namely the de Bruijn–Erdős theorem, is pretty straightforward to prove. However, the original proof of Huh and Wang used some extraordinarily complicated mathematics, namely the theory of intersection cohomology for ℓ-adic perverse sheaves. These are algebraic structures that don't obviously have much to do with point sets in $\mathbb{R}^d$, so it's already remarkable that Huh and Wang were able to find a connection between these disparate areas of math.

The more recent and more general proof, of Braden, Huh, Matherne, Proudfoot, and Wang, is in a certain sense simpler. Namely, they don't need to use any of this intersection cohomology as a black box, and their paper is basically self-contained. However, it's almost 100 pages long, and for good reason: they construct a "combinatorial Hodge theory", which is some analogue to this intersection cohomology, and which is well-suited to deal with their problem directly. However, to use it, they need to prove that it satisfies certain complicated algebraic relations (the so-called *Kähler package*), and these proofs are in some sense inspired by the corresponding proofs for ℓ-adic perverse sheaves.

I don't really understand either of the proofs of the Dowling–Wilson conjecture, and even if I did, it'd be impossible to cover a complicated 100-page paper in a three-day course. But at an immensely high level, one can view their proof(s) as a generalization of the linear-algebraic proof above of the de Bruijn–Erdős theorem. Recall that there, we wished to prove that $n \le m$, where n is the number of poitns in the plane and m is the number of lines they define. To do so, we defined vectors $v^{(1)}, \ldots, v^{(n)} \in \mathbb{R}^m$, and showed that these vectors were linearly independent, which implied that $n \le m$. The vectors were defined in a very natural way, namely the kth coordinate of $v^{(j)}$ was 1 if the jth point was on the kth line, and 0 otherwise.

Let's say we wish to prove that $F_i(P) \leq F_{d-1-i}(P)$ for some genuinely d-dimensional point set P in $\mathbb{R}^d$. Let $N = F_i(P)$ and $M = F_{d-1-i}(P)$. Braden–Huh–Matherne–Proudfoot–Wang do effectively define vectors $v^{(1)}, \ldots, v^{(N)} \in \mathbb{R}^M$, and then prove that these vectors are linearly independent, implying that $N \leq M$ as claimed. Moreover, the kth coordinate of the vector $v^{(j)}$ is non-zero if the jth i-dimensional hyperplane is a subset of the kth $(d-1-i)$-dimensional hyperplane, and zero otherwise. However, it turns out that if we let the vectors $v^{(j)}$ have only 0 and 1 as their coordinates, then they will *not* be linearly independent in general. Instead, in order to prove the Dowling–Wilson conjecture, it is necessary to put some other non-zero numbers in the vectors $v^{(1)}, \ldots, v^{(N)}$. And it really seems that the only way to come up with *which* numbers to put in the vectors is to build up some complicated algebraic machinery, as these authors do.

I mentioned earlier that many results related to Rota's conjecture have been proved, although Rota's conjecture itself remains open. Most of these related results have come in the past decade, by several of the same authors I mentioned above, and also using techniques related to this sort of combinatorial Hodge theory. I personally wouldn't be too surprised if Rota's conjecture were proved in the next few years, using many of these same techniques.

3 Almost orthogonal lines

Let us say that a collection of lines passing through a common point is *pairwise orthogonal* if every pair of them is orthogonal. It's easy to show that three lines in the plane cannot all be pairwise orthogonal; moreover, this is basically an immediate consequence of the fourth of Euclid's five axioms for plane geometry.

In three dimensions, of course, we can have three pairwise orthogonal lines, but four lines cannot be pairwise orthogonal. The generalization of this, unsurprisingly, is the following.

Theorem 3.1. *The maximum number of pairwise orthogonal lines in $\mathbb{R}^n$ is n.*

Proof. One can prove this using linear algebra, but let's prove it by induction on n. The base case $n = 1$ is simple: $\mathbb{R}^1$ consists of a single line, so the maximum number of *distinct* lines in $\mathbb{R}^1$ (let alone pairwise orthogonal) is 1.

For the inductive step, suppose the statement is true for $n-1$. Say we are given any collection L of pairwise orthogonal lines in $\mathbb{R}^n$. Fix any line ℓ in the collection, and let H be the $(n-1)$-dimensional hyperplane orthogonal to ℓ. Let $L' = L \setminus \{\ell\}$ be the collection of lines gotten by deleting ℓ. Since every line in L' is orthogonal to ℓ, we see that L' must lie in the hyperplane H. But since H is identical to $\mathbb{R}^{n-1}$, we have that $|L'| \leq n-1$ by the inductive hypothesis. Therefore, $|L| = |L'| + 1 \leq (n-1) + 1 = n$. □

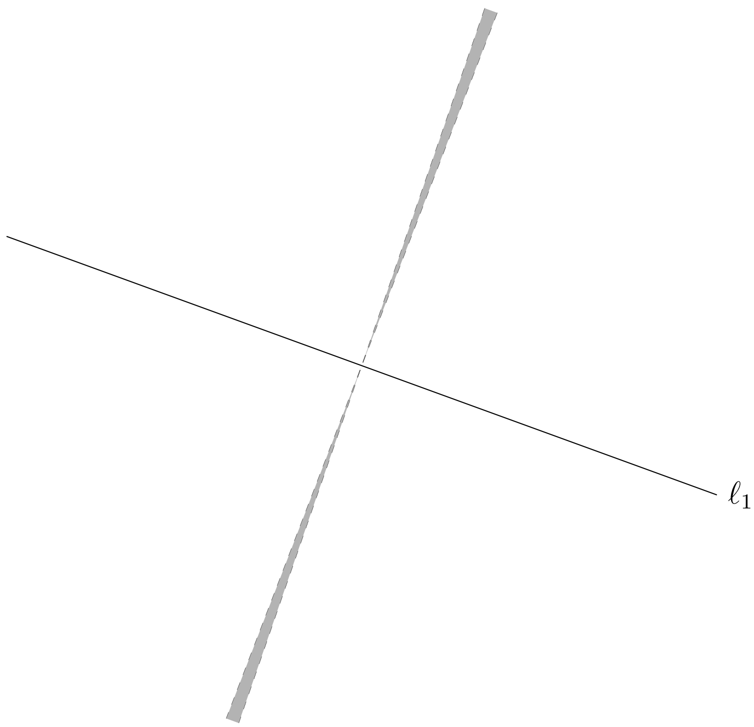
Although Euclid didn't know about dimensions higher than 3, this proof was basically known to him, and is essentially given in Propositions XI.4 and XI.5 in the *Elements*.

So far, we haven't done anything too interesting. But we can get a new notion by slightly weakening our notion of pairwise orthogonality.

Definition 3.2. Let us say that a collection of lines through a common point is *almost orthogonal* if for every pair, the angle between them is between 89° and 91°.

Let $f(n)$ denote the maximum number of almost orthogonal lines in $\mathbb{R}^n$. Note that since a pairwise orthogonal collection is, in particular, almost orthogonal, we definitely have $f(n) \geq n$. In other words, a collection of n pairwise orthogonal lines in $\mathbb{R}^n$ is certainly almost orthogonal.

Can $f(n)$ be larger than n? We can try some small examples for intuition; for instance, I claim that $f(2) = 2$. Indeed, consider any line ℓ_1 in $\mathbb{R}^2$. Any line which is almost orthogonal to ℓ_1 must lie in the tiny shaded region in the following picture.



We can put any line ℓ_2 we want inside the shaded region, but once we pick such a line, we can't find another line ℓ_3 that is nearly orthogonal to both ℓ_1 and ℓ_2.

Similarly, one can prove that $f(3) = 3$, in much the same way. However, it's pretty easy to believe that as the number of dimensions increases, the amount of wiggle room we have increases. My intuition suggests that once n is large enough, we can actually use this wiggle room and squeeze in one more line, so that $f(n_1) = n_1 + 1$ for some sufficiently large n_1. It is then reasonable to imagine that for many more dimensions, $n + 1$ is the best we can do, but eventually we accumulate enough wiggle room so that $f(n_2) = n_2 + 2$ for some much larger n_2. I would expect the pattern to continue in this fashion, so that as n tends to infinity, we have that $f(n)$ equals n plus some very slowly growing function of n. In particular, I'd expect that $f(n) < 2n$ for all n.

As it turns out, this intuition is *wildly* wrong, and high dimensions behave totally differently. For very large n, $f(n)$ will be waaaaaay larger than n.

Theorem 3.3. $f(n) \geq \lfloor 1.0007^n \rfloor$.

In particular, in $100,000$*-dimensional space, there are more than* 10^{30} *almost orthogonal lines. In million-dimensional space, there are more than a googol cubed almost orthogonal lines.*

To prove Theorem 3.3, we need to set up some background first. Recall that for two vectors $v, w \in \mathbb{R}^n$, the dot product $v \cdot w$ is defined $v \cdot w = \sum_{i=1}^n v_i w_i$, where $v_1, \ldots, v_n$ are the n coordinates of v, and similarly for w. The *length* of a vector v is defined by $\|v\| = \sqrt{v \cdot v}$.

The main fact we will need is essentially the same as the *law of cosines*. It says that for two vectors v, w, the angle θ between them satisfies

$$\cos\theta = \frac{v \cdot w}{\|v\|\|w\|}.$$

One can check (e.g. on Wolfram Alpha) that an angle θ is between $89°$ and $91°$ if and only if $|\cos\theta| \leq \beta$, for some real number $\beta \approx 0.017452$. So in order to prove Theorem 3.3, we will actually find a set of $M \geq \lfloor 1.0007^n \rfloor$ vectors $v^{(1)}, \ldots, v^{(M)}$ in $\mathbb{R}^n$ such that

$$|v^{(i)} \cdot v^{(j)}| \leq \beta \|v^{(i)}\|\|v^{(j)}\| \qquad \text{for all distinct } 1 \leq i, j \leq M. \tag{1}$$

It is actually not so easy to explicitly construct such a set of vectors. To get around this, we will use an immensely powerful technique called the *probabilistic method*. Basically, it turns out that if we pick these vectors at random, then property (1) holds with positive probability. In particular, since the probability that (1) holds is positive when the vectors are chosen randomly, there must exist *some* set of vectors for which it holds!

Before we do the proof, we'll collect a few probabilistic tools that we will need.

3.1 Probabilistic tools

Let's say we make some random choices, and let A be an *event*, i.e. a thing that can happen after the random choices. For instance, if our randomness is rolling a die, then the event A might be that die comes up 1, or that it comes up a prime number, or that it comes out even. If the randomness is that we deal a random card to everyone in this class, then the event A could be that we all receive red cards, or that I get the ace of spades, or that exactly two of us got queens. We denote by $\Pr(A)$ the probability of A, which is a number between 0 and 1.

The first basic result we'll need about probability is the following, known as the *union bound*.

Proposition 3.4 (Union bound). *For any two events* A, B,

$$\Pr(A \textit{ happens or } B \textit{ happens (or both)}) \leq \Pr(A) + \Pr(B).$$

More generally, for any events $A_1, \ldots, A_t$,

$$\Pr(\textit{at least one of } A_1, \ldots, A_t \textit{ happens}) \leq \sum_{i=1}^t \Pr(A_i).$$

Rigorously proving the union bound requires setting up the actual formal mathematical theory of probability, which we won't do. But hopefully it is intuitively clear. For instance, if I pick a random card from a deck, the probability that it is either red *or* a king is at most the probability that it is red, plus the probability that it is a king.

We say that two events A and B are *independent* if $\Pr(A \text{ and } B) = \Pr(A)\Pr(B)$. Intuitively, what independence "means" is that whether or not A happens has no influence on the probability that B happens. For example, if I flip a fair coin twice, then the event that the first flip is heads is independent of the event that the second flip is tails. If we have events $A_1, \ldots, A_t$, then we say that they are *mutually independent* if for every set $I \subseteq \{1, \ldots, t\}$,

$$\Pr(A_i \text{ happens for all } i \in I) = \prod_{i \in I} \Pr(A_i).$$

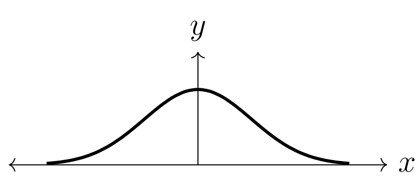
The technical condition is perhaps a bit tricky to wrap your head around, but again the intuition is that the events are mutually independent if they all have no influence on each other.

A *random variable* is just any number whose value is determined by the randomness. For instance, if we roll a die, then the outcome of the die roll is a random variable. If we denote this random variable by X, then we have

$$\Pr(X = 1) = \Pr(X = 2) = \Pr(X = 3) = \Pr(X = 4) = \Pr(X = 5) = \Pr(X = 6) = \frac{1}{6}.$$

This information is called the *distribution* of X: formally, the distribution is the list of probabilities $\Pr(X = x)$ for every value x that X can take. We say that two random variables X, Y are *identically distributed* if they have the same distribution: the probability that X takes on the value 5 equals the probability that Y takes on the value 5, and this holds for every possible choice of 5. Additionally, we say that random variables X, Y are independent if the event $X = x$ is independent of the event $Y = y$ for all x, y. We can analogously define what it means for a number of random variables $X_1, \ldots, X_t$ to be mutually independent.

Perhaps the most important result in all of probability theory is the *central limit theorem*, which was first observed by Gauss and Laplace. The central limit theorem explains the shocking prevalence of the *bell curve* in a huge number of natural phenomena. The bell curve (or *Gaussian distribution*) is the curve in the plane given by the equation $y = e^{-x^2}$.



I won't state it rigorously or prove it, but the central limit theorem roughly says that if $X_1, X_2, \ldots$ are independent random variables that are identically distributed, then $X_1 + \cdots + X_t$ converges to a Gaussian distribution as $t \to \infty$. In real life, you could imagine that we are doing some scientific experiment, trying to find the value of some experimentally determined number (e.g. the speed of light, the mass of an electron, etc.). Since our equipment is not perfect, we imagine that every experiment returns a random outcome, which is probably close to the truth, but with some random noise. We can also assume that in different experiments, the random noise is independent. This implies that when we add up (or average) the results of the experiments, we are adding up independent, identically distributed random variables, which should then yield a bell curve by the central limit theorem. This is the reason why the bell curve shows up in many scientific experiments, as well as in many other contexts in real life.

Instead of the central limit theorem, we will need the following result which can be seen as an "approximate" version of the central limit theorem. Results of this type are variously known as the Chernoff bound, Hoeffding's inequality, and/or Azuma's inequality. Rather than stating it for arbitrary random variables, we'll only state it for variables that are $+1$ or -1 with probability $1/2$.

Proposition 3.5 (Chernoff bound)**.** *Let $Z_1, \ldots, Z_n$ be mutually independent random variables, with $\Pr(Z_i = 1) = \Pr(Z_i = -1) = \frac{1}{2}$ for all $1 \le i \le n$. For any $a > 0$, we have*

$$\Pr(Z_1 + \cdots + Z_n > a) \le e^{-a^2/(2n)}$$

and

$$\Pr(Z_1 + \cdots + Z_n < -a) \le e^{-a^2/(2n)}.$$

Therefore,

$$\Pr(|Z_1 + \cdots + Z_n| > a) \le 2e^{-a^2/(2n)}.$$

The reason that I say that this is an approximate form of the central limit theorem is that we see that quadratic behavior in the exponent, matching what is given by the bell curve $y = e^{-x^2}$.

You'll prove the Chernoff bound in the homework.

3.2 Back to almost orthogonal lines

As we said earlier, in order to prove the existence of a set of vectors satisfying (1), we will pick the vectors $v^{(i)}$ randomly. Specifically, each coordinate of each vector will be $+1$ or -1 with probability $\frac{1}{2}$, and all these choices will be made independently at random.

The following two lemmas will allow us to prove (1).

Lemma 3.6. *Let $X_1, \ldots, X_n$ be mutually independent, identically distributed random variables with $\Pr(X_i = 1) = \Pr(X_i = -1) = \frac{1}{2}$ for all i. Let $v = (X_1, \ldots, X_n)$ be the random vector they define. Then $\|v\| = \sqrt{n}$.*

Proof. By definition,

$$\|v\|^2 = v \cdot v = \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} 1 = n.$$ □

Lemma 3.7. *Let $X_1, \dots, X_n, Y_1, \dots, Y_n$ be mutually independent, identically distributed random variables, each equal to $+1$ or -1 with probability $\frac{1}{2}$. Let $v = (X_1, \dots, X_n)$ and $w = (Y_1, \dots, Y_n)$. Then*

$$\Pr(|v \cdot w| > \beta n) \leq 2e^{-\beta^2 n/2}.$$

Proof. Let $Z_i = X_i Y_i$ for all $1 \leq i \leq n$. Then by the definition of the dot product,

$$v \cdot w = Z_1 + \cdots + Z_n.$$

We first observe that $Z_1, \dots, Z_n$ are mutually independent, since all the X_i, Y_i were mutually independent, and therefore the randomness in Z_i has no effect on the randomness in Z_j for any $j \neq i$. Additionally, we claim that $\Pr(Z_i = 1) = \Pr(Z_i = -1) = \frac{1}{2}$. Indeed, Z_i will be 1 if X_i and Y_i are both either equal to 1 (which happens with probability $\frac{1}{4}$) or both equal to -1 (which also happens with probability $\frac{1}{4}$), so $\Pr(Z_i = 1) = \frac{1}{2}$. By essentially the same argument, we see that $\Pr(Z_i = -1) = \frac{1}{2}$.

Therefore, we are in a position to apply Proposition 3.5, which tells us that for any $a > 0$,

$$\Pr(|v \cdot w| > a) = \Pr(|Z_1 + \cdots + Z_n| > a) \leq 2e^{-a^2/(2n)}.$$

Plugging in $a = \beta n$, we get the desired result. □

We are now ready to prove Theorem 3.3.

Proof of Theorem 3.3. Let $M = \lfloor 1.0007^n \rfloor$, and let $v^{(1)}, \dots, v^{(M)}$ be random vectors in $\mathbb{R}^n$, each of which has all its coordinates equal to ± 1 with probability $\frac{1}{2}$, with all these choices made independently. By Lemma 3.6, we see that $\|v^{(i)}\| = n$ for all $1 \leq i \leq M$.

For all $1 \leq i < j \leq M$, let A_{ij} be the event that $|v^{(i)} \cdot v^{(j)}| > \beta n$. By Lemma 3.7, we know that

$$\Pr(A_{ij}) \leq 2e^{-\beta^2 n/2}$$

for all i, j. Therefore, by the union bound,

$$\Pr(A_{ij} \text{ happens for at least one pair } i,j) \leq \sum_{1 \leq i < j \leq M} \Pr(A_{ij}) \leq \binom{M}{2} \cdot 2e^{-\beta^2 n/2} < M^2 e^{-\beta^2 n/2},$$

where the final inequality uses the fact that $\binom{M}{2} = \frac{M^2 - M}{2} < \frac{M^2}{2}$.

To conclude, we note that by our choice of M, we have that $M \leq 1.0007^n$, and thus $M^2 \leq 1.0015^n$. Moreover, if we recall that $\beta = \cos(89°) \approx 0.017452$, then we see that $e^{-\beta^2/2} < 0.99$. Therefore,

$$\Pr(A_{ij} \text{ happens for at least one pair } i,j) < (1.0015 \cdot 0.99)^n < 1,$$

since $1.0015 \cdot 0.99 \approx 0.9915 < 1$.

Because of this, we find that with positive probability, *none* of the events A_{ij} happen. Since this probability is positive, there must exist some vectors $v^{(1)}, \dots, v^{(M)}$ in $\mathbb{R}^n$ such that

$|v^{(i)} \cdot v^{(j)}| \leq \beta n$ and $\|v^{(i)}\| = n$ for all distinct i, j. Consider the set of M lines going through these vectors. By the law of cosines, the angle θ_{ij} between any two of them satisfies

$$\cos\theta_{ij} = \frac{v^{(i)} \cdot v^{(j)}}{\|v^{(i)}\|\|v^{(j)}\|} \leq \frac{\beta n}{n} = \beta.$$

Thus, these M lines all have angles between 89° and 91°, and thus they are almost orthogonal, as claimed. □

The final thing to remark is that the result about almost orthogonal lines in $\mathbb{R}^n$ is one instantiation of a very general phenomenon, often called the "curse of dimensionality". Basically, a bunch of weird things happen once one moves to very high dimensions, even though the basic rules of geometry are the same as in the dimensions we're used to. Here are just a few other examples.

- Volumes and distances get all weird in high dimensions. For instance, if you take the unit ball in $\mathbb{R}^n$ and inscribe it in a hypercube, then the volume of the ball is *exponentially small* in the volume of the cube. In other words, the vast majority of points with coordinates in $[-1, 1]$ are *really far* from the origin.
- It turns out that almost all the points on the unit sphere in $\mathbb{R}^n$ are extremely close to the equator: if you pick a random point on the unit sphere, then the probability that its latitude is greater than 1° or less than −1° is exponentially small. But this is extremely weird, since the sphere is symmetric, so this holds for all equators! So the sphere is somehow very squashed like a pancake, since almost all of it is very close to the equator. But it's squashed in *every* direction, so it's kind of like a spiky ball with spikes going out in every direction. It's really hard to visualize!