Topics in combinatorics

W.T. Gowers

This course will cover a miscellaneous collection of topics in combinatorics and closely related fields. What the topics have in common is that they all involve proofs that at one time surprised experts by their simplicity. Sometimes these were the first proofs of long-standing open problems, and sometimes they were new proofs of results that had previously been established by much longer arguments. Several of these arguments use ideas and techniques that have gone on to be used by many other people.

Another theme of the course is the sheer diversity of methods that are used in combinatorics. We shall see uses of probability, linear algebra, linear analysis, topology, entropy, multivariate polynomials, tensor rank, concentration of measure, and more. (There will also be one or two arguments that are completely elementary.)

Contents

1 Averaging arguments

Let $a_1, \dots, a_n$ be real numbers with average a. Then there exists i such that $a_i \geq a$ and there exists i such that $a_i \leq a$. More generally, we have the following extremely basic fact.

Proposition 1.1. *If X is a random variable, then $\mathbb{P}[X \geq \mathbb{E}X] > 0$ and $\mathbb{P}[X \leq \mathbb{E}X] > 0$.*

Proof. It might seem as though this fact is too obvious even to need a proof. That is almost true, and for the purposes of this course there is no problem if you just assume it, but for completeness here is a proof.

If $\mathbb{P}[X \geq \mathbb{E}X] = 0$, then for each n let E_n be the event that

$$X \in (\mathbb{E}X - (n-1)^{-1}, \mathbb{E}X - n^{-1}],$$

where we interpret $\mathbb{E}X - 0^{-1}$ to be $-\infty$. Then $\sum_n \mathbb{P}[E_n] = 1$. Therefore,

$$\mathbb{E}X \leq \sum_n (\mathbb{E}X - n^{-1})\mathbb{P}[E_n] = \mathbb{E}X - \sum_n n^{-1}\mathbb{P}[E_n],$$

which is strictly less than $\mathbb{E}_X$, since there must exist n with $\mathbb{P}[E_n] > 0$. This is a contradiction. □

Another way of stating the proposition is to say that $\mathbb{E}X \in [\min X, \max X]$. Bizarrely, this fact, which is particularly basic if the random variable X is discrete, the case that will normally concern us, can quite often be used to prove highly non-trivial theorems. This section will discuss three main examples: Sperner's theorem, the Erdős-Ko-Rado theorem, and the Szemerédi-Trotter theorem.

Before I get on to those, let me mention a closely related very basic principle, which underlies so-called double-counting arguments. It can be formalized in various ways, but here's a simple one.

Proposition 1.2. *Let G be a bipartite graph with finite vertex sets X and Y. Suppose that the average degree of a vertex in X is $\delta|Y|$. Then the average degree of a vertex in Y is $\delta|X|$.*

Proof. The number of edges of G is $\delta|X||Y|$ and the result follows instantly. □

This is "double counting" because we are counting the number of edges in two different ways: as $|X|$ times the average degree of a vertex in X, and as $|Y|$ times the average degree of a vertex in Y.

The following proposition, which has essentially the same proof, perhaps captures better how double-counting arguments often work.

Proposition 1.3. *Let G be a bipartite graph with finite vertex sets X and Y. Suppose that the average degree of a vertex in X is at least d_1 and the average degree of a vertex in Y is at most d_2. Then $|Y| \geq d_1|X|/d_2$.*

Proof. The number of edges of G is at least $d_1|X|$ and at most $d_2|Y|$. The result follows. □

Here are two simple examples of double-counting arguments.

Example 1.4. Let G be a planar graph with V vertices, E edges and F faces. Then each face contains at least three edges and each edge is contained in at most two faces. It follows from Proposition 1.3 that $E \geq 3F/2$. (Let X be the set of faces of G and Y be the set of edges of G, and join a face to an edge if and only if the face contains the edge. Then apply the proposition.)

Before I give the next example, let me introduce some standard notation that will be useful throughout the course. I shall write $[n]$ for $\{1, 2, \dots, n\}$, and $[n]^{(r)}$ for the set of subset of $[n]$ of size r.

Example 1.5. Let $0 \leq r < s \leq n$, let $\mathcal{A}$ be a subset of $[n]^{(r)}$, and let $\partial_s\mathcal{A}$ be the *s-upper shadow* of $\mathcal{A}$, which is defined to be $\{B \in [n]^{(s)} : \exists A \in \mathcal{A} \text{ s.t. } A \subset B\}$. Then every set in $\mathcal{A}$ is contained in $\binom{n-r}{n-s}$ sets in $\partial_s\mathcal{A}$, and every set in $\partial_s\mathcal{A}$ contains at most $\binom{s}{r}$ sets in $\mathcal{A}$. It follows that $|\partial_s\mathcal{A}| \geq \binom{n-r}{n-s}|\mathcal{A}|/\binom{s}{r}$, which can also be written as $\frac{(n-r)!r!}{(n-s)!s!}|\mathcal{A}|$, and therefore as $\binom{n}{s}|\mathcal{A}|/\binom{n}{r}$.

1.1 Sperner's theorem

Let $\mathcal{A}$ be a family of subsets of $[n]$. How large can $\mathcal{A}$ be if no member of $\mathcal{A}$ is contained in any other? A simple way to construct such a family is to ensure that all its members have the same size. Sperner showed that that was the best one can do.

Theorem 1.6 (Sperner)**.** *Let $\mathcal{A}$ be a family of subsets of $[n]$ such that no member of $\mathcal{A}$ is contained in any other. Then $|\mathcal{A}| \leq \binom{n}{\lfloor n/2 \rfloor}$.*

Proof. For $0 \leq i \leq n$ let $\mathcal{A}_i$ be the set $\mathcal{A} \cap [n]^{(i)}$ – that is, the collection of sets in $\mathcal{A}$ of size i. Let $x_1, \dots, x_n$ be a random ordering of $[n]$ and for $0 \leq i \leq n$ let E_i be the set $\{x_1, \dots, x_i\}$. Then the number of i such that $E_i \in \mathcal{A}$ is at most 1, since no member of $\mathcal{A}$ is contained in any other. But for each i, E_i is uniformly distributed in $[n]^{(i)}$, so the *expected* number of i such that $E_i \in \mathcal{A}$ is $\sum_{i=0}^n |\mathcal{A}_i|/\binom{n}{i}$. (Here we used linearity of expectation, another very simple principle that surprisingly often has non-trivial consequences.) It follows that $\sum_{i=0}^n |\mathcal{A}_i|/\binom{n}{i} \leq 1$, and therefore, since the largest size of a binomial coefficient is $\binom{n}{\lfloor n/2 \rfloor}$, that $|\mathcal{A}| = \sum_i |\mathcal{A}_i| \leq \binom{n}{\lfloor n/2 \rfloor}$ as claimed. □

Note that the last step in the proof above threw away a lot of interesting information: the above proof shows that, as one might expect, including lots of sets that are either very small or very big makes it harder to avoid containments. The proof itself is not the original proof of Sperner.

Another remark is that the proof used the contrapositive of Proposition 1.1.

1.2 The Erdős-Ko-Rado theorem

Let $\mathcal{A}$ be a family of subsets of $[n]$ of size k. If any two members of $\mathcal{A}$ intersect, then how large can $\mathcal{A}$ be? If $k > n/2$, then the answer is trivially $\binom{n}{k}$. If $k = n/2$, then the answer is not trivial but it is still easy. If $A \subset [n]$ is a set of size k, then we cannot pick both A and A^c, and it is simple to see that if from every pair (A, A^c) we pick exactly one set, then any two members of the resulting family of sets intersect. So when $k = n/2$ the answer is $\frac{1}{2}\binom{n}{k}$. When $k < n/2$, the problem is no longer easy, though the answer is simple and natural: the largest intersecting family is obtained by taking all sets that contain some given element.

The proof we shall give of the next theorem is not the one given by Erdős, Ko and Rado, but rather a later one discovered by Gyula Katona. Like the proof of Sperner's theorem above, it makes use of a random ordering.

Theorem 1.7 (Erdős, Ko, Rado)**.** *Let $k < n/2$ and let $\mathcal{A} \subset [n]^{(k)}$ be an intersecting family. Then $|\mathcal{A}| \leq \binom{n-1}{k-1}$, with equality if and only if $\mathcal{A}$ is of the form $\{A \in [n]^{(k)} : i \in A\}$.*

Proof. Let $x_1, \dots, x_n$ be a random cyclic ordering of $[n]$. (That is, we think of x_1 as the successor of x_n.) For each j, let I_j be the "interval" $\{x_j, x_{j+1}, \dots, x_{j+k-1}\}$, where addition is mod n. We now show that at most k of these intervals can be part of an intersecting family.

To see this, let us assume, without loss of generality, that $I_1 = \{x_1, \dots, x_k\}$ belongs to the family. Also, for each j let us define I_j^- to be the interval $\{x_{j-k}, \dots, x_{j-1}\}$. Then every interval that belongs to the family must intersect I_1, which means that it must be either I_j for some $j \in \{1, 2, \dots, k\}$ or I_j^- for some $j \in \{2, 3, \dots, k\}$. (I did not mention I_{k+1}^- because it is equal to I_1, but this is not an important point.) Also, we cannot include both I_j and I_j^- since they are disjoint (because $k \leq n/2$). It follows, as claimed, that at most k intervals can belong to an intersecting family.

For each j, the probability that I_j belongs to the family is $|\mathcal{A}|/\binom{n}{k}$, since I_j is uniformly distributed. So the expected number of intervals that belong to the family is $n|\mathcal{A}|/\binom{n}{k}$. Therefore, $n|\mathcal{A}|/\binom{n}{k} \leq k$, and therefore $|\mathcal{A}| \leq \frac{k}{n}\binom{n}{k} = \binom{n-1}{k-1}$.

Note that if equality holds, then for every cyclic ordering there must be k consecutive intervals in $\mathcal{A}$. (Here we are using the fact that if $\mathbb{P}[X \leq \mathbb{E}X] = 1$, then $\mathbb{P}[X = \mathbb{E}X] = 1$.) But if exactly k intervals belong to an intersecting family, then we can say slightly more than we said above. For $1 < j < k$, it is not possible for I_j^- and I_{j+1} both to belong to $\mathcal{A}$ (here we use the fact that $n \geq 2k+1$), but for each such $1 < j \leq k$ we have to choose either I_j^- or I_j in order to have enough intervals. Therefore, if ever we choose I_j^-, then we must choose I_{j+1}^-, which forces us to choose I_{j+2}^-, and so on all the way up to I_k^-. It follows that there is some $1 \leq r \leq k$ such that the intervals we choose are $I_1, I_2, \dots, I_r, I_{r+1}^-, \dots, I_k^-$, or in other words all intervals of length k that contain x_r.

This observation enables us to show uniqueness of the extremal example. Let $x_1, \dots, x_{2k-1}$ be such that the intervals $\{x_j, \dots, x_{j+k-1}\}$ with $1 \leq j \leq k$ all belong to $\mathcal{A}$, and let u be an element not equal to any of $x_1, \dots, x_{2k-1}$ (which exists since $2k \leq n$). Now let A be any set of size k that contains x_k, and enumerate its elements as $\{y_1, \dots, y_k\}$ in such a way that $y_1, \dots, y_r$ belong to the set $\{x_1, \dots, x_k\}$, $y_r = x_k$, and $y_{r+1}, \dots, y_k$ do not. Now construct a cyclic order of $[n]$ that begins with u, then continues with an ordering of $x_1, \dots, x_k$ that finishes $y_1, \dots, y_r$, and continues after that with $y_{r+1}, \dots, y_k$ (and then finishes in some arbitrary way). Write this ordering as $z_0, z_1, \dots, z_{n-1}$ with $z_0 = u$. Then the interval $\{z_0, z_1, \dots, z_{k-1}\}$ is equal to the set $\{u, x_1, \dots, x_{k-1}\}$, which does not belong to $\mathcal{A}$ because it does not intersect $\{x_k, \dots, x_{2k-1}\}$. Since there must be k consecutive intervals in this cyclic order that belong to $\mathcal{A}$, and $\{z_1, \dots, z_k\} = \{x_1, \dots, x_k\}$ does, it follows that A, which equals $\{z_{k-r+1}, \dots, z_{2k-r}\}$, belongs to $\mathcal{A}$.

Therefore, $\mathcal{A}$ contains every set of size k that contains the element x_k. Since this is a maximal intersecting family and has size $\binom{n-1}{k-1}$, we are done. □

1.3 The Szemerédi-Trotter theorem

Let $x_1, \dots, x_n$ be points in the plane and let $L_1, \dots, L_m$ be lines. An *incidence* is a pair (x_i, L_j) such that $x_i \in L_j$. If all the points and all the lines are distinct, then how many incidences can there be? This question is answered, up to a multiplicative constant, by a beautiful theorem of Szemerédi and Trotter. Again, the proof we give is not the original one, but a much simpler argument, due to György Elekes, which came as a major surprise and led to significant progress in many questions in combinatorial geometry.

The starting point is another question, which is interesting in its own right. Let G be

a graph. The *crossing number* of G is defined to be the smallest number of crossings (that is, pairs of edges that intersect) in any drawing of G in the plane. So a planar graph is a graph that has crossing number 0, and the larger the crossing number, the further the graph is from being planar.

Let us now investigate the connection between the crossing number of a graph and how many edges it has.

Lemma 1.8. *A planar graph with n vertices has at most $3n-6$ edges.*

Proof. Let e and f be the number of edges and faces of G, respectively. Euler's formula tells us that $n-e+f=2$. We also saw in Example 1.4 that $e \geq 3f/2$. It follows that $2 = n-e+f \leq n-e+2e/3 = n-e/3$. Rearranging gives the lemma. □

Corollary 1.9. *A graph with n vertices and m edges has crossing number at least $m-3n$.*

Proof. Let G be such a graph and suppose we have drawn it in the plane. By Lemma 1.8, there is at least one crossing. Remove one of the two crossing edges and the number of edges is still greater than $3n-6$, so there is another crossing. We can keep doing that $m-3n$ times with the number of edges remaining greater than $3n-6$, so there must be at least $m-3n$ crossings in the original drawing of G. □

Now comes the averaging argument, which for large m will greatly boost the bound just proved.

Corollary 1.10. *Let G be a graph drawn in the plane with n vertices and m edges with $m \geq 6n$. Then there are at least $m^3/72n^2$ crossings.*

Proof. Let p be a probability to be chosen later, and let H be an induced subgraph of G chosen by picking each vertex of G independently at random with probability p.

Suppose that the edges xy and zw cross in G. Then the probability that this crossing belongs to H as well is p^4. Therefore, if t is the number of crossings in G, then the expected number of crossings in H is p^4t.

But the expected number of edges in H is p^2m and the expected number of vertices is pn. Therefore, by Corollary 1.9 the expected number of crossings in H is at least p^2m-3pn. It follows that $p^4t \geq p^2m-3pn$.

It remains merely to make a good choice of p. We choose it so that $p^2m = 6pn$: that is, we take p to be $6n/m$. That gives us that $p^4t \geq 3pn$, so $t \geq 3n/p^3 = m^3/72n^2$. □

It is not immediately obvious what crossing numbers have to do with incidences. However, as we shall see, given a system of points and lines, one can define a graph drawing in a natural way, and the crossing number estimate turns out to give us exactly the information we are looking for.

Theorem 1.11 (Szemerédi, Trotter)**.** *Given n distinct points and m distinct lines in the plane, the number of incidences is at most $8\max\{m, n, m^{2/3}n^{2/3}\}$.*

Proof. Regard the system of points and lines as a drawing of a graph as follows. The points are the vertices. For each line L, enumerate the vertices along that line in order as $x_1, \dots, x_k$, and for each $i < k$ regard the segment of L from x_i to x_{i+1} as an edge.

Let the number of incidences be s. Let the lines be $L_1, \dots, L_m$ and for each i let the number of points on line L_i be s_i. Then $s_1 + \cdots + s_m = s$, and the number of edges of the graph is $(s_1 - 1) + \cdots + (s_m - 1) = s - m$. Therefore, the number of crossings is, by Corollary 1.10, at least $(s-m)^3/72n^2$ if $s - m \geq 6n$.

However, since two lines cross in at most one point, we also know that the number of crossings is at most $\binom{m}{2} \leq m^2/2$. So if $s \geq 6n + m$, then $(s-m)^3/72n^2 \leq m^2/2$, which implies that $(s-m)^3 \leq 36m^2n^2$. From this it follows that either $s \leq 6n + 2m \leq \max\{8m, 8n\}$ or $(s/2)^3 \leq 36m^2n^2$. In the second case, $s \leq 8m^{2/3}n^{2/3}$. The result follows. □

Three examples help to make sense of the bound in the theorem. If there is just one line and it contains all the points, then the number of incidences is n. Similarly, if there is just one point and it is contained in all the lines, then the number of incidences is m. And if we take the grid $\{0, 1, \dots, r-1\} \times \{0, 1, \dots, 2r^2 - 1\}$ and all lines that pass through one of the points $(0, y)$ with $y < r^2$ and have slope belonging to the set $\{1, 2, \dots, r\}$, then there are $2r^3$ points and r^3 lines, and each line passes through r points, which gives r^4 incidences, equalling the Szemerédi-Trotter bound up to a constant. With a bit more thought one can generalize these examples and show that the bound is sharp up to a constant for all values of m and n.

2 A few bounds on binomial coefficients

Combinatorics often involves estimates and those estimates often involve factorials and binomial coefficients, so it is helpful to have a few rough and ready bounds for them.

For factorials one can use Stirling's formula, but for most applications that level of accuracy isn't needed, so I prefer the following bounds, which are much easier to prove.

Lemma 2.1. *For every n, $(n/e)^n \leq n! \leq e(n+1)(n/e)^n$.*

Proof. Suppose you want to obtain crude bounds for $\int_1^n \log x \, dx$. If you decompose the interval $[1, n]$ into intervals of width 1 and look at the upper and lower Riemann sums, you will deduce that

$$\sum_{m=1}^{n-1} \log m \leq \int_1^n \log x \, dx \leq \sum_{m=1}^{n-1} \log(m+1).$$

But the antiderivative of $\log x$ is $x \log x - x$, so the expression in the middle is equal to $n \log n - n + 1$. Taking exp of both sides, we deduce that

$$(n-1)! \leq e^{n \log n - n + 1} \leq n!$$

for every n.

Now $e^{n\log n-n+1} = e(n/e)^n$, from which we obtain the lower bound. Also, $(n+1)^n = (1+1/n)^n n^n \leq en^n$, so

$$e((n+1)/e)^{n+1} = e^{-n}(n+1)^n(n+1) \leq e(n+1)(n/e)^n,$$

which gives us the upper bound. □

This tells us in particular that $(n/e)^n$ is a rough and ready approximation to $n!$ that is useful in contexts where we are not concerned by a stray multiple of n or so.

For some purposes this is not enough. For example, it is often useful to have a good estimate for $2^{-n}\binom{n}{n/2}$ (when n is even) and if we simply plug in the bounds for factorials then we get a trivial upper bound of 1 and a fairly weak lower bound. But we can use a completely elementary argument to obtain a bound that is correct up to a constant multiple.

Lemma 2.2. *For every even integer,* $\frac{1}{\sqrt{2n}} \leq 2^{-n}\binom{n}{n/2} \leq \frac{1}{\sqrt{n}}$.

Proof. Let us begin by writing an exact formula for the binomial coefficient in question. We have

$$2^{-n}\binom{n}{n/2} = \frac{n!}{2^n(n/2)!(n/2)!} = \frac{1.2.3\dots n}{2.4.6\dots n.2.4.6\dots n} = \frac{1.3.5\dots(n-1)}{2.4.6\dots n} = \prod_{m=1}^{n/2}(1-\frac{1}{2m}).$$

To obtain an upper bound for this, let us observe that

$$\left(\frac{1.3.5\dots(n-1)}{2.4.6\dots n}\right)^2 \leq \frac{1.2.3.4\dots n}{2.3.4.5\dots(n+1)} = \frac{1}{n+1}.$$

Similarly, a lower bound is given by

$$\frac{1.1.2.3\dots(n-1)}{2.2.3.4\dots n} = \frac{1}{2n}$$

Taking square roots gives the result. □

Remark 2.3. The quantity $2^{-n}\binom{n}{n/2}$ is the probability that an n-step random walk ends up at the origin. The end point of such a random walk is a sum $X = X_1 + \dots + X_n$ where the X_i are independent random variables that take the values ± 1, each with probability $1/2$. Since variances of independent random variables add, $\mathrm{Var}X = n$, so X has standard deviation $\sqrt{n}$. We expect the values near zero to be taken with roughly equal probability (as long as they are even), so it is no surprise that the probability of landing exactly on zero should be of order $n^{-1/2}$.

Another bound that is often useful when m is a lot smaller than n is the trivial bound

$$\binom{n}{m} = \frac{n(n-1)\dots(n-m+1)}{m!} \leq n^m.$$

Sometimes one wants to be just a little more accurate than this, in which case we can use the fact that $m! \geq (m/e)^m$ to improve this upper bound to $(en/m)^m$. Note that that is true whether or not m is small, and indeed it can be useful even when $m = \alpha n$ for a constant α, as long as that constant is not too close to $1/2$ (because then it does not improve on the trivial bound of 2^n).

In a later section we shall give a more accurate bound for $\binom{n}{m}$ when $m = \alpha n$ for a constant α independent of n.

Next, it is surprisingly useful to think about ratios of consecutive binomial coefficients. Directly from the formula one has that

$$\binom{n}{m+1}\Big/\binom{n}{m} = \frac{n-m-1}{m}.$$

In particular, if $m = \alpha n$ then the ratio is roughly $(1-\alpha)/\alpha$. From this it follows if $\alpha < 1/2$ that

$$\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{m} \leq \binom{n}{m}\frac{1}{1-\alpha/(1-\alpha)} = \binom{n}{m}\frac{1-\alpha}{1-2\alpha}.$$

The important point here is not the precise estimate, but the fact that if α is bounded away from $1/2$, then to within a constant the sum of the all the binomial coefficients up to $\binom{n}{m}$ is the same as $\binom{n}{m}$ itself. Often we don't care about constant factors, so then we can say that the sum of the binomial coefficients up to $\binom{n}{m}$ is basically $\binom{n}{m}$.

To give an idea of what one can do with ratios, we now prove a lemma that shows that almost all of the discrete cube is concentrated around the middle layers. This is a simple example of the *concentration of measure phenomenon*, which we shall see more about later in the course.

Lemma 2.4. *Let $m = (1/2-\theta)n$ with $0 < \theta \leq 1/2$. Then $2^{-n}\binom{n}{m} \leq e^{-\theta^2 n/2}$.*

Proof. If $k \leq (1/2-\theta/2)n$, then

$$\binom{n}{k-1}\Big/\binom{n}{k} = \frac{k}{n-k+1} \leq \frac{1/2-\theta/2}{1/2+\theta/2} \leq 1-\theta,$$

It follows that

$$\binom{n}{(1/2-\theta)n} \leq (1-\theta)^{\theta n/2}\binom{n}{(1/2-\theta/2)n} \leq e^{-\theta^2 n/2}.2^n,$$

which proves the result. □

Remark 2.5. I was not careful about whether $(1/2-\theta/2)n$ was an integer, so the proof above is strictly speaking not quite correct. However, it is not worth the bother of correcting it, since we are about to see a different proof that does not run into this kind of difficulty. (But if you do want to correct it, you could note that the proof works in the case that m and $n/2-m$ are both even integers, and we threw quite a lot away, so getting it for the coefficients in between is not too hard.)

Next, we prove a rather more general estimate. It can itself be considerably generalized, but even in this form it is useful.

Lemma 2.6. *Let $X_1, \dots, X_n$ be independent random variables of mean zero taking values in $[-1, 1]$ and let $X = \sum_i X_i$. Then $\mathbb{P}[X \geq \epsilon n] \leq e^{-\epsilon^2 n/4}$.*

Proof. We use a trick that can be used to prove many results of this flavour, which is to look at the *exponential moment* $\mathbb{E}e^{\lambda X}$, apply Markov's inequality, and optimize over λ. We have

$$\begin{aligned}\mathbb{E}e^{\lambda X} &= \mathbb{E}e^{\lambda \sum_i X_i} \\ &= \mathbb{E}\prod_i e^{\lambda X_i} \\ &= \prod_i \mathbb{E}e^{\lambda X_i}\end{aligned}$$

Now since $\mathbb{E}X_i = 0$ for each i,

$$\mathbb{E}e^{\lambda X_i} = 1 + \frac{\lambda^2}{2!}\mathbb{E}X_i^2 + \frac{\lambda^3}{3!}\mathbb{E}X_i^3 + \dots.$$

If $\lambda \leq 1$, then since $|X_i|$ is always at most 1, the right-hand side is at most $1 + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots \leq e^{\lambda^2}$. So $\mathbb{E}e^{\lambda X} \leq e^{\lambda^2 n}$.

By Markov's inequality,

$$\mathbb{P}[X \geq \epsilon n] = \mathbb{P}[e^{\lambda X} \geq e^{\lambda \epsilon n}] \leq e^{\lambda^2 n - \lambda \epsilon n}.$$

Choosing $\lambda = \epsilon/2$ we obtain an upper bound of $e^{-\epsilon^2 n/4}$. □

Remark 2.7. The constant $1/4$ in the bound above can be improved, but at the cost of making the argument a bit more complicated.

Remark 2.8. Replacing each X_i by $-X_i$, we find that $\mathbb{P}[X \leq -\epsilon n] \leq e^{-\epsilon^2 n/4}$ as well.

Corollary 2.9. *Let $\epsilon > 0$, let n be a positive integer, and let $m = (1/2 - \epsilon)n$. Then $2^{-n}\sum_{k=0}^{m}\binom{n}{k} \leq e^{-\epsilon^2 n}$.*

Proof. Let $X_1, \dots, X_n$ be independent random variables that take the values ± 1, each with probability $1/2$ and let $X = \sum_i X_i$. Then the quantity we are trying to bound is $\mathbb{P}[X \leq -2\epsilon n]$. By Remark 2.8, this is at most $e^{-\epsilon^2 n}$. □

3 Well-separated sets and vectors

Variants of the following question come up frequently when one works on combinatorial problems. Let $A_1, \dots, A_N$ be subsets of $[n]$ of size $n/2$ and suppose that $|A_i \cap A_j| \leq \alpha n$ for every $i \neq j$. How large can N be?

The answer turns out to depend quite heavily on α, and in particular on how α compares with $1/4$, the significance of which is that the expected size of the intersection of two random sets of size $n/2$ is $n/4$.

Theorem 3.1. *Let $\alpha > 1/4$. Then there can be exponentially many subsets of $[n]$ of size $n/2$ such that no two intersect in more than αn.*

Proof. Let A be a random set of size $n/2$. We begin by estimating the probability that $|A \cap \{1, 2, \dots, n/2\}| > \alpha n$. Writing $m = \alpha n$, we can give a formula for this probability, namely

$$\binom{n}{n/2}^{-1} \sum_{r=m+1}^{n/2} \binom{n/2}{r}\binom{n/2}{n-r} \leq 2^{n/2}\binom{n}{n/2}^{-1} \sum_{r=0}^{n/2-m-1} \binom{n/2}{r},$$

where the equality follows by replacing r with $n/2 - r$ and using the fact that $\binom{n}{k} = \binom{n}{n-k}$ for any n and k.

Now $n/2-m-1 = \beta n/2$ for a constant β that is less than $1/2$. Therefore, by Lemma 2.9, each binomial coefficient $\binom{n/2}{r}$ in the above sum is exponentially small compared with $2^{n/2}$, which implies that the entire sum is an exponentially small fraction of 2^n, and therefore the entire probability is exponentially small. In principle we could obtain an explicit bound for the exponent, but we content ourselves here with stating that there is a constant c that depends on α (it will be $(\alpha - 1/4)^2$ up to a constant) such that the probability is at most e^{-cn}.

By symmetry, we now know that if we pick sets $A_1, \dots, A_N$ of size $n/2$ independently at random, then for each $i < j$ the probability that $|A_i \cap A_j| > \alpha n$ is at most e^{-cn}. Therefore, the expected number of bad pairs $i < j$ is at most $e^{-cn}\binom{N}{2} \leq e^{-cn}N^2/2$.

It follows that with non-zero probability the number of bad pairs is at most $e^{-cn}N^2/2$. (Strictly speaking, we are using Proposition 1.1 here.) Therefore, there exist sets $A_1, \dots, A_N$ such that the number of bad pairs is at most $e^{-cn}N^2/2$. If we choose N such that this is at most $N/2$, we can remove one set from each bad pair and still be left with $N/2$ sets, and now there are no bad pairs. Taking $N = e^{cn}$ does the job. □

3.1 Associating vectors with sets

We now turn to the cases $\alpha = 1/4$ and $\alpha < 1/4$. For these we shall use an extremely important idea with applications all over combinatorics, which is to treat sets as 01-valued functions and then to view those functions as living in a Euclidean (or occasionally a more general) space. Given a set A, we shall write $\mathbb{1}_A$ for its characteristic function – that is, the function that takes the value 1 in A and 0 outside A. An immediate observation we

can make is that if A and B are subsets of $[n]$, then

$$|A \cap B| = \sum_{i=1}^{n} \mathbb{1}_A(i)\mathbb{1}_B(i) = \langle \mathbb{1}_A, \mathbb{1}_B \rangle.$$

However, it is often convenient to associate with a set not its characteristic function but its *balanced* function, which is defined as follows. If A is a subset of a finite set X and $|A| = \delta|X|$, then for each $x \in X$ we define $f_A(x)$ to be $1-\delta$ if $x \in A$ and $-\delta$ if $x \notin A$. We call this function balanced because it averages zero: indeed $\mathbb{E}_x f(x) = (1-\delta)\delta - \delta(1-\delta) = 0$.

Lemma 3.2. *Let A and B be two subsets of $[n]$ of size $n/2$. Let f_A and f_B be the balanced functions of A and B. Then $\langle f_A, f_B \rangle = |A \cap B| - n/4$*

Proof.

$$\langle f_A, f_B \rangle = \langle \mathbb{1}_A - 1/2, \mathbb{1}_B - 1/2 \rangle = |A \cap B| - |A|/2 - |B|/2 + n/4 = |A \cap B| - n/4.$$

□

In particular, if we have sets A and B of size $n/2$ and their intersection has size at most $(1/4 - \delta)n$, then $\langle f_A, f_B \rangle \leq -\delta n$. This enables us to prove the following rather surprising result, which is that if $\alpha < 1/4$, then the largest possible number of sets of size $n/2$ that intersect in at most αn is bounded independently of n.

Theorem 3.3. *Let $A_1, \dots, A_m$ be subsets of $[n]$ of size $n/2$ and suppose that $|A_i \cap A_j| \leq (1/4 - \delta)n$ for every $i \neq j$. Then $m \leq 1 + 1/4\delta$.*

Proof. Write f_i for the balanced function f_{A_i}. Then, writing $\|.\|$ for the Euclidean norm (that is, $\|f\|^2 = \langle f, f \rangle$),

$$\begin{aligned} 0 &\leq \|\sum_{i=1}^{m} f_i\|^2 \\ &= \sum_{i,j} \langle f_i, f_j \rangle \\ &= \sum_i \|f_i\|^2 + \sum_{i \neq j} \langle f_i, f_j \rangle \\ &\leq mn/4 - m(m-1)\delta n, \end{aligned}$$

where for the last inequality we used Lemma 3.2. The result follows on rearranging. □

Of course, this is just (equivalent to) a special case of a more general theorem, which can be proved in the same way.

Theorem 3.4. *Let $v_1, \dots, v_m$ be unit vectors in a Euclidean space, let $\delta > 0$, and suppose that $\langle v_i, v_j \rangle \leq -\delta$ for every $i \neq j$. Then $m \leq 1 + 1/\delta$.*

Proof.

$$\begin{aligned} 0 &\leq \|\sum_i v_i\|^2 \\ &= m + \sum_{i \neq j} \langle v_i, v_j \rangle \\ &\leq m - \delta m(m-1), \end{aligned}$$

which proves the result. □

If $\delta = 1/k$ for a positive integer k, then it is possible to find $1 + 1/\delta = k+1$ unit vectors that all have inner products at most $-\delta$ (and in fact necessarily equal to $-\delta$ by the proof above). One takes the vertices of a regular simplex of dimension k.

A nice way to construct such an object is to take the standard basis vectors $e_1, \dots, e_{k+1}$ in $\mathbb{R}^{k+1}$ and to project them to the subspace orthogonal to the vector $(1, 1, \dots, 1)$, or in other words to the subspace of vectors whose coordinates add up to zero. This gives us vectors $v_1, \dots, v_{k+1}$ where

$$v_i = \Big(-\frac{1}{k+1}, \dots, -\frac{1}{k+1}, 1 - \frac{1}{k+1}, -\frac{1}{k+1}, \dots, -\frac{1}{k+1} \Big),$$

where the $1 - \frac{1}{k+1}$ is in the ith place. If $i \neq j$, then

$$\langle v_i, v_j \rangle = \frac{k-1}{(k+1)^2} - \frac{2}{k+1} + \frac{2}{(k+1)^2} = -\frac{1}{k+1}.$$

We also have that

$$\|v_i\|^2 = \frac{k}{(k+1)^2} + 1 - \frac{2}{k+1} + \frac{1}{(k+1)^2} = 1 - \frac{1}{k+1} = \frac{k}{k+1}.$$

Therefore, if we normalize the vectors by setting $u_i = ((k+1)/k)^{1/2} v_i$, we have $k+1$ unit vectors and $\langle u_i, u_j \rangle = -1/k$ when $i \neq j$.

Now let us turn to the question of how many unit vectors we can find in $\mathbb{R}^n$ if their inner products are all at most zero. (This corresponds to the case of sets of size $n/2$ with intersections of size at most $n/4$.) A simple example of such a collection of unit vectors is $\pm u_1, \dots, \pm u_n$, where $u_1, \dots, u_n$ is an orthonormal basis. This turns out to be best possible, as we now show. (For this particular question it is of course unimportant that the vectors are unit vectors – we just need them not to be the zero vector.)

Theorem 3.5. *Let $x_1, \dots, x_m$ be non-zero vectors in $\mathbb{R}^n$ such that $\langle x_i, x_j \rangle \leq 0$ for every $i \neq j$. Then $m \leq 2n$, and if $m = 2n$ then there is an orthonormal basis $a_1, \dots, a_n$ such that each x_i is a multiple of some a_j (which implies that amongst the x_i there is exactly one positive multiple and one negative multiple of each a_j).*

Proof. We prove this by induction on n. The case $n = 1$ is trivial. Assume now that $n > 1$ and that $m \geq 2n$. Let $a_1 = x_1/\|x_1\|$ and let $y_2, \dots, y_m$ be the orthogonal projections of $x_2, \dots, x_m$ to the subspace orthogonal to x_1. Then for each $i \geq 2$, $y_i = x_i - \langle x_i, a_1\rangle a_1$. If $i \neq j$, then

$$\langle y_i, y_j\rangle = \langle x_i, x_j\rangle - \langle x_i, a_1\rangle\langle x_j, a_1\rangle.$$

Since neither of $\langle x_i, a_1\rangle$ and $\langle x_j, a_1\rangle$ is positive, this is at most $\langle x_i, x_j\rangle$, and therefore at most 0.

We therefore have $m-1$ vectors $y_2, \dots, y_m$ in an $(n-1)$-dimensional space with $\langle y_i, y_j\rangle \leq 0$ for every $i \neq j$. Since $m - 1 > 2(n - 1)$, it follows from our inductive hypothesis that $y_2, \dots, y_m$ are not all non-zero. Without loss of generality $y_2 = 0$, which implies that x_2 is a multiple of a_1, and therefore a negative multiple of a_1.

But if x_1 is a positive multiple of a_1 and x_2 is a negative multiple of a_1, then the only way another vector can avoid having a positive inner product with one of x_1 and x_2 is if it is orthogonal to a_1. Therefore, $x_3, \dots, x_m$ are at least $2(n - 1)$ vectors living in the space orthogonal to a_1, which is $(n - 1)$-dimensional. By induction we can find an orthonormal basis $a_2, \dots, a_n$ of this space with the properties stated in the theorem, and combining this with a_1 creates an orthonormal basis that proves the theorem. □

Combining this result with Lemma 3.2 and observing that the balanced function of a subset of $[n]$ of size $n/2$ lives in the $(n-1)$-dimensional subspace orthogonal to $(1, 1, \dots, 1)$, we deduce immediately that it is not possible to find more than $2(n - 1)$ sets of size $n/2$ that intersect in at most $n/4$. But can we find that many? To do so, we need to find a large set of orthogonal vectors with ± 1 coordinates. (The balanced functions have coordinates equal to $\pm 1/2$, but that is clearly equivalent.)

3.2 Hadamard matrices

Here are two constructions that give such sets. The first is a very simple inductive one. Let W_0 be the 1×1 matrix (1) and in general define

$$W_m = \begin{pmatrix} W_m & W_m \\ W_m & -W_m \end{pmatrix}.$$

It is straightforward to check the following facts.

- W_m is a $2^m \times 2^m$ matrix with ± 1 entries.
- The first row of W_m consists entirely of 1s.
- The rows of W_m are orthogonal to each other.

The matrices W_m are called *Walsh matrices*. If $n = 2^m$, we have an $n \times n$ ± 1 matrix W with first row consisting of 1s. If we now take A_i to be $\{j : W_{ij} = 1\}$ and $B_i = \{j : W_{ij} = -1\}$ for $i = 2, 3, \dots, n$, we obtain $2(n - 1)$ sets $A_2, \dots, A_n, B_2, \dots, B_n$ of size $n/2$ such that $A_i \cap B_i = \emptyset$ for each i and otherwise any two distinct sets from the collection have

intersection of size $n/4$. (The fact that the sets have size $n/2$ follows is equivalent to the fact that the later rows of H_n are orthogonal to the first row.)

The second construction is algebraic. Recall that if $0 \leq x < p$, then the *Legendre symbol* $\left(\frac{x}{p}\right)$ stands for 1 if x is a quadratic residue, -1 if x is a non-residue, and 0 if $x = 0$. Let p be a prime of the form $4m+3$, let $n = p+1$ and define the *Paley matrix* $P = P_n$ by setting $P_{xy} = \left(\frac{x-y}{p}\right)$ if $0 \leq x, y \leq p-1$ and $x \neq y$, -1 if $x = y$, and 1 if either of x, y is equal to p. In other words, we first create a $p \times p$ matrix by setting P_{xy} to be 1 or -1 according to whether $x - y$ is or is not a quadratic residue, considering 0 to be a non-residue for this purpose, and then we attach to it an extra row and column that consist of 1s.

Here is the Paley matrix P_8. Note that if you ignore the last row and column, then it is constant down diagonals (or even "mod-7 diagonals").

$$\begin{pmatrix} -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \\ -1 & -1 & 1 & 1 & -1 & 1 & -1 & 1 \\ -1 & -1 & -1 & 1 & 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 & -1 & 1 \\ -1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

The proof that the Paley matrix is orthogonal uses some elementary number theory in a very pleasant way.

Lemma 3.6. *For every prime p and every $d \neq 0 \bmod p$ we have that $\sum_x \left(\frac{x}{p}\right)\left(\frac{x+d}{p}\right) = -1$, where the summation is over all $x \bmod p$.*

Proof. We shall make use of the fact that the function $x \mapsto \left(\frac{x}{p}\right)$ is multiplicative.

$$\begin{aligned} \sum_x \left(\frac{x}{p}\right)\left(\frac{x+d}{p}\right) &= \sum_{x \neq 0,-d} \left(\frac{x^2}{p}\right)\left(\frac{1+dx^{-1}}{p}\right) \\ &= \sum_{x \neq 0,-d} \left(\frac{1+dx^{-1}}{p}\right). \end{aligned}$$

Now as x ranges over all values other than 0 or d, $1+dx^{-1}$ ranges over all values other than 1 or 0. But half the non-zero elements mod p are quadratic residues and 1 is a quadratic residue, so this last sum is equal to -1. □

Corollary 3.7. *The Paley matrix has orthogonal rows.*

Proof. The inner product of two distinct rows of the Paley matrix is closely related to the sum calculated in Lemma 3.6. If neither row is the last, then it is of the form

$$\sum_{x\neq 0,-d}\left(\frac{x}{p}\right)\left(\frac{x+d}{p}\right)-\left(\frac{d}{p}\right)-\left(\frac{-d}{p}\right)+1,$$

since when $x=0$ we replace $\left(\frac{x}{p}\right)$ by -1, and the product of the last entries in the two rows is 1. But $p\equiv 3 \bmod 4$, so -1 is not a quadratic residue mod p, from which it follows that $\left(\frac{d}{p}\right)=-\left(\frac{-d}{p}\right)$. Therefore, by Lemma 3.6 (and the fact that $\left(\frac{0}{p}\right)=0$), the inner product is $-1+1=0$.

If one of the rows is the last row, then we obtain the sum

$$-1+\sum_{x\neq 0}\left(\frac{x}{p}\right)+1.$$

Since exactly half the non-zero numbers mod p are quadratic residues, this is also 0. □

A ± 1 matrix with orthogonal rows is called a *Hadamard matrix*. The construction of Walsh matrices relies on the observation that if H is a Hadamard matrix, then so is $\begin{pmatrix} H & H \\ H & -H \end{pmatrix}$. From that and the construction of Paley matrices, we find that if n is of the form $2^m(p+1)$ for p a prime congruent to 3 mod 4, then there is an $n\times n$ Hadamard matrix. The following question is a famous open problem.

Question 3.8. Let n be a multiple of 4. Does there exist an $n\times n$ Hadamard matrix?

The smallest n that is a multiple of 4 for which no Hadamard matrix is known is 668.

4 The sum-product problem

Let A be a set of integers. The *sumset* $A+A$ is defined to be $\{x+y : x,y\in A\}$. There are many very interesting results about the sizes of sumsets and what they tell us about the sets themselves. Here is a simple one.

Theorem 4.1. *Let A be a set of size n. Then $|A+A|\leq 2n-1$, with equality if and only if A is an arithmetic progression.*

Proof. Let $A=\{a_1,\dots,a_n\}$ with $a_1<\dots<a_n$. Then the $2n-1$ numbers

$$a_1+a_1, a_1+a_2,\dots,a_1+a_n,a_2+a_n,\dots,a_n+a_n$$

form a strictly increasing sequence, so they are all distinct. More generally, given any sequence

$$(i_1,j_1),(i_2,j_2),\dots,(i_{2n-1},j_{2n-1})$$

such that $i_1 = j_1 = 1$, $i_{2n-1} = j_{2n-1} = n$, and each term in the sequence is obtained by adding 1 to either the x-coordinate or the y-coordinate of the previous term, then all the resulting sums $a_{i_t} + a_{j_t}$ form a strictly increasing sequence.

It follows that if $|A + A| = 2n - 1$, then these sequences are all the same. From that it follows that $a_i + a_j$ depends only on $i + j$. From this it follows that the numbers $a_1, \dots, a_n$ form an arithmetic progression, since $a_{i-1} + a_{i+1} = 2a_i$ for every i strictly between 1 and n. □

We also define the *product set* $A.A$ to be the set $\{xy : x, y \in A\}$. If A consists of positive numbers, then by taking the logarithm of those numbers and applying what we have just proved for sumsets (after checking that the proof goes through unchanged if $a_1, \dots, a_n$ are real numbers, which it does), we see that if $|A| = n$, then $|A.A| \geq 2n - 1$, with equality if and only if A is a geometric progression.

So far, so simple, but we are a short step away from a famous unsolved problem of Erdős and Szemerédi.

Conjecture 4.2. *For every $c > 0$ there exists n_0 such that for every $n \geq n_0$ and every set A of positive integers of size n, either $A + A$ or $A.A$ has size at least n^{2-c}.*

Note that the conjecture is saying that either the sumset or the product set is almost as big as it can be, since trivially the maximum size is at most $n(n + 1)/2$ (and simple examples show that this maximum can be attained – for example to obtain a large sumset, take any sequence $a_1, a_2, \dots, a_n$ of positive integers such that $a_i \geq 2a_{i-1}$ for every i).

One reason the conjecture is quite plausible is that sets with small sumsets tend to be related to arithmetic progressions, which have very large product sets, whereas sets with small product sets tend to be related to geometric progressions, which have very large sumsets. However, this does not rule out some kind of strange "intermediate" set with, say, a sumset and a product set of size at most $n^{1.99}$, and indeed, almost nothing is known about the structure of sets with sumsets that are of this sort of size.

For some years, the best known bound for the problem, due to Jozsef Solymosi, was that, up to logarithmic factors, either the sumset or the product set was of size at least $n^{4/3}$. The exponent 4/3 has subsequently been improved a few times, each time by just a tiny amount, and the current record, due to Misha Rudnev and Sophie Stevens, stands at 4/3+2/1167. But these improved bounds require complicated arguments, whereas the proof of Solymosi, which we give now, is beautifully simple.

We start with a useful definition and an almost trivial, but for the proof very important, lemma.

Definition 4.3. Let A be a finite subset of $\mathbb{R}_+$. For each $x \in \mathbb{R}_+$, write $\rho_A^+(x)$ for the number of pairs $(a, b) \in A^2$ such that $a + b = x$, $\rho_A^\times(x)$ for the number of pairs $(a, b) \in A^2$ such that $ab = x$, and $\rho_A^\div(x)$ for the number of pairs $(a, b) \in A^2$ such that $a/b = x$. The *additive energy* of A is $\sum_{x \in A+A} \rho_A^+(x)^2$, and the *multiplicative energy* is $\sum_{x \in A.A} \rho_A^\times(x)^2$. The *quotient set* A/A of A is the set $\{x/y : x, y \in A\}$.

Lemma 4.4. *The multiplicative energy is also equal to $\sum_{x \in A/A} \rho_A^\div(x)^2$.*

Proof. The multiplicative energy $\sum_{x\in A.A}\rho_A^\times(x)^2$ is equal to the number of quadruples $(a,b,c,d)\in A^4$ such that $ab=cd$, which is the number of quadruples such that $a/c=d/b$, which is equal to $\sum_{x\in A/A}\rho_A^\div(x)^2$. □

Solymosi deduces his sum-product theorem from a more precise statement, which is that for every set A of positive real numbers, $|A.A||A+A|^2\geq |A|^4/4\lceil\log|A|\rceil$. This he proves by obtaining both an upper bound and a lower bound for the multiplicative energy of A. The lower bound is easier, so we give it first. We shall use the following inequality, which is immensely useful throughout combinatorics. Because it is so important, we give two proofs.

Lemma 4.5. *Let $a_1,\dots,a_n$ be real numbers. Then $\sum_i|a_i|^2\geq n^{-1}(\sum_i|a_i|)^2$.*

Proof 1. By the Cauchy-Schwarz inequality,

$$\sum_i|a_i|=\sum_i 1\cdot|a_i|\leq n^{1/2}\Big(\sum_i|a_i|^2\Big)^{1/2}.$$

The result follows on rearranging. □

Proof 2. Let X be a random variable. Then its variance is

$$\mathbb{E}[X-\mathbb{E}X]^2=\mathbb{E}X^2-(\mathbb{E}X)^2.$$

The left-hand side is clearly non-negative, so $(\mathbb{E}X)^2\leq\mathbb{E}X^2$. Now apply this fact to the random variable that takes the value $|a_i|$ with probability n^{-1} to deduce that $n^{-2}(\sum_i|a_i|)^2\leq n^{-1}\sum_i|a_i|^2$. □

Lemma 4.6. *Let A be a set of positive real numbers. Then the multiplicative energy of A is at least $|A|^4/|A.A|$.*

Proof. The multiplicative energy is $\sum_{x\in A.A}\rho_A^\times(x)^2$. Now $\sum\rho_A^\times(x)=|A|^2$, since each pair $(a,b)\in A^2$ contributes 1 to exactly one $\rho_A^\times(x)$. Also, the number of x for which $\rho_A^\times(x)\neq 0$ is $|A.A|$. The result now follows from Lemma 4.5. □

The above proof, though simple, is worth digesting fully, since this kind of use of the Cauchy-Schwarz inequality has a huge number of applications.

We now turn to the upper bound, and for this we use Lemma 4.4, which tells us that we can give an upper bound for $\sum_{m\in A/A}\rho_A^\div(m)^2$ instead, which for convenience we shall rewrite as $\sum_{m\in A/A}\rho_A^\div(m^{-1})^2$, which we can do because $\rho_A^\div(m)=\rho_A^\div(m^{-1})$ for every $m\in A/A$.

Lemma 4.7. *Let A be a set of positive real numbers. Then*

$$\sum_{m\in A/A}\rho_A^\div(m^{-1})^2\leq 2|A+A|^2\lceil\log|A|\rceil.$$

Proof. We begin by applying another trick that is extremely useful in a number of contexts. It applies when we have a function defined on a finite set and have difficulty handling it because it is not sufficiently close to being constant. In such a situation, if the values are not too spread out and we are prepared to sacrifice a logarithmic factor, we can simply partition the domain according to the values taken up to the nearest power of 2. This process is called *dyadic decomposition*.

Here every value of $\rho_A^{\div}(m^{-1})$ is between 1 and $|A|$, so if we divide the set $\{1, 2, \dots, |A|\}$ up into intervals of the form $\{x : 2^{k-1} \leq x < 2^k\}$, we need at most $\lceil \log_2 |A| \rceil$ of those intervals. It follows by averaging that there exists one of the intervals, call it I, such that

$$\sum_{\rho_A^{\div}(m^{-1}) \in I} \rho_A^{\div}(m^{-1})^2 \geq \frac{1}{\lceil \log_2 |A| \rceil} \sum_{m \in A/A} \rho_A^{\div}(m^{-1})^2.$$

We shall now concentrate our efforts on finding an upper bound for the left-hand side.

Let us think a bit about what $\rho_A^{\div}(m^{-1})^2$ means. The number of $(a, b) \in A \times A$ such that $a/b = m^{-1}$ is the number of points in the intersection of $A \times A$ with the line $y = mx$. (Note that I am writing $A \times A$ for the Cartesian product of A with itself – it does *not* stand for the product set.) Therefore, the square of this number is the number of pairs of such points. That is not very interesting as it stands, but now that we know that the lines intersect $A \times A$ in sets of roughly the same size, we can make it more interesting by observing that up to a factor of 2 it is equal to the number of pairs of points that can be obtained by taking one point from one line and one point from another. And if the points in one line are $p_1, \dots, p_r$ and the points in another line are $q_1, \dots, q_s$, then because the lines have different gradients, the points $p_i + q_j$ are distinct and they all belong to $(A \times A) + (A \times A) = (A + A) \times (A + A)$. (Note that this $\times$ is a Cartesian product: $A \times A$ is not to be confused with $A.A$).

Let $m_1, \dots, m_t$ be the elements of A/A, in increasing order, for which $\rho_A^{\div}(m^{-1}) \in I$, and for each i let L_i be the line $y = m_i x$ and let $B_i = L_i \cap (A \times A)$. Then the sets $B_i + B_{i+1}$ are disjoint (because $B_i + B_{i+1}$ lies between L_i and L_{i+1} and the gradients of the L_i are in increasing order) and contained in $(A+A) \times (A+A)$. Also, for each i, $|B_i + B_{i+1}| \geq |B_i|^2/2$, since $|B_i + B_{i+1}| = |B_i||B_{i+1}|$ and all the B_i have the same size up to a factor of 2.

It follows that $\sum_{i=1}^{t-1} |B_i|^2 \leq 2|A + A|^2$. This is not quite what we want, because we would prefer to include $|B_t|^2$ on the left-hand side. There is a small trick that deals with this problem, which is to consider the vertical line that passes through the leftmost point of B_t, which is a point $(a, m_t a)$ for some $a \in A$. The intersection of this line with $A \times A$ contains all points $(a, m_t b)$ such that $(b, m_t b) \in B_t$. Setting B_{t+1} to be this set of points, we then have that $|B_{t+1}| = |B_t|$, that $|B_{t+1} + B_t| = |B_t|^2$, and that $B_{t+1} + B_t$ is disjoint from the other sets $B_i + B_{i+1}$. We may therefore conclude that

$$\sum_{\rho_A^{\div}(m^{-1}) \in I} \rho_A^{\div}(m^{-1})^2 = \sum_{i=1}^{t} |B_i|^2 \leq 2|A + A|^2,$$

which implies the result, by our choice of I. □

We have more or less finished the argument.

Theorem 4.8 (Solymosi)**.** *Let A be a set of positive real numbers. Then*

$$|A.A||A+A|^2 \geq \frac{|A|^4}{2\lceil \log |A| \rceil},$$

and therefore either $|A+A|$ or $|A.A|$ is at least $|A|^{4/3}/(2\lceil \log |A| \rceil)^{1/3}$.

Proof. By our upper and lower bounds for the multiplicative energy, we find that

$$2|A+A|^2\lceil \log |A| \rceil \geq \frac{|A|^4}{|A.A|}.$$

The result follows. □

If $A = \{1, 2, \dots, n\}$, then $|A+A| \leq 2n$ and $|A.A| \leq n^2$, so $|A.A||A+A|^2 \leq 4n^4$. Thus, the lower bound for $|A.A||A+A|^2$ given above is sharp up to the log factor. Of course, this does not imply that the lower bound for $\max\{|A+A|, |A.A|\}$ is sharp (and indeed, as mentioned earlier, it is known not to be).

5 The chromatic number of the Kneser graph

The Kneser graph $G_{n,k}$ is defined as follows. Its vertex set is $[2n+k]^{(n)}$ – the set of all subsets of $\{1, 2, \dots, 2n+k\}$ of size n – and two vertices are joined by an edge if and only if the corresponding subsets are disjoint. We normally think of k as fixed and n tending to infinity, so two sets of size n that are joined are almost complementary. However, this is not an assumption in the arguments that follow.

This section concerns the chromatic number of $G_{n,k}$, the determination of which was an open problem for over 20 years until it was solved by Lázsló Lovász in 1978. Kneser observed that the chromatic number is at most $k+2$. Indeed, we can define a colouring as follows. For $i = 1, 2, \dots, k+1$ we let C_i consist of all sets of size n with minimal element i. Then all the remaining sets to into a colour class C_{k+2}. It is trivial that two sets in C_i intersect if $i \leq k+1$. If they both belong to C_{k+2}, then they are n-element subsets of the set $\{k+2, k+3, \dots, 2n+k\}$, which has size $2n-1$, and therefore we see again that they intersect. By the definition of the Kneser graph, that implies that $C_1, \dots, C_{k+2}$ is indeed a proper colouring.

Lovász's proof is regarded as something of a milestone because it introduced the idea of using topological methods to solve combinatorial problems, which at the time was a big surprise. With the benefit of hindsight one can reduce this surprise by introducing topological methods in a simpler context that bears some resemblance to the problem we wish to solve. The main tool he used was the Borsuk-Ulam theorem, so let us begin by looking at that.

5.1 The Borsuk-Ulam theorem and some variants

The Borsuk-Ulam theorem is the following statement.

Theorem 5.1. *Let $f : S^n \to \mathbb{R}^n$ be a continuous function. Then there exists $x \in S^n$ such that $f(x) = f(-x)$.*

It is equivalent to the following result, which is itself sometimes referred to as the Borsuk-Ulam theorem.

Theorem 5.2. *Let $A_1, \dots, A_{n+1}$ be closed subsets of S^n with union equal to S^n. Then there exist $x \in S^n$ and $i \in \{1, 2, \dots, n+1\}$ such that $x \in A_i$ and $-x \in A_i$.*

Proof of equivalence. Assume Theorem 5.1, and let $A_1, \dots, A_{n+1}$ are closed subsets of S^n with union S^n. Define a function $f : S^n \to \mathbb{R}^{n+1}$ by $f(x) = (d(x, A_1), \dots, d(x, A_n))$. Note that the image of f lies in the subset of $\mathbb{R}^{n+1}$ that consists of all vectors with non-negative coordinates such that at least one coordinate is zero, the second property following from the fact that the sets A_i cover S^n. This set is homeomorphic to $\mathbb{R}^n$, as can be seen by taking the orthogonal projection from it to the subspace $\{y : \sum_i y_i = 0\}$.

By Theorem 5.1, there therefore exists x such that $f(x) = f(-x)$. But since the union of the A_i is S^n, there exists i such that $x \in A_i$, and therefore $d(x, A_i) = 0$, and therefore $d(-x, A_i) = 0$. Since A_i is closed, this implies that $-x \in A_i$.

Conversely, suppose we can find a continuous function $f : S^n \to \mathbb{R}^n$ for which $f(x)$ is never equal to $f(-x)$. Define a continuous function $g : S^n \to S^{n-1}$ by

$$g(x) = \frac{f(x) - f(-x)}{\|f(x) - f(-x)\|_2}$$

and note that $g(-x) = -g(x)$ for every x. Now there exist n closed sets $B_1, \dots, B_{n+1}$ that cover S^{n-1} without any of them containing a pair of antipodal points. For example, take a regular n-dimensional simplex and project its faces on to the sphere in the obvious way. Since each face is part of an affine hyperplane that does not contain 0, its projection lies strictly inside a hemisphere and therefore does not contain a pair of antipodal points. But then the $n+1$ sets $g^{-1}(B_i)$ are closed and cover S^n, and do not contain a pair of antipodal points (because $g(-x) = -g(x)$), contradicting Theorem 5.2. □

This in turn implies (and as we shall soon see, is equivalent to) the same statement but about open sets.

Theorem 5.3 (Open-sets version)**.** *Let $A_1, \dots, A_{n+1}$ be open subsets of S^n with union equal to S^n. Then there exist $x \in S^n$ and $i \in \{1, 2, \dots, n+1\}$ such that $x \in A_i$ and $-x \in A_i$.*

Theorem 5.2 implies Theorem 5.3. For each i let $f_i(x) = d(x, A_i^c)$ and let $f(x) = \max_i f_i(x)$. Then each f_i is continuous, and therefore so is f. Also, since every x belongs to some A_i and every A_i is open, $f(x) > 0$ for every x. By compactness, it follows that there is some $\epsilon > 0$ such that $f(x) \geq \epsilon$ for every x.

Now for each i define B_i to be $\{x : d(x, A_i^c) \geq \epsilon\}$. Then each B_i is a closed subset of A_i and the union of the B_i is S^n. Therefore, by Theorem 5.2 we can find i and x such that $x \in B_i$ and $-x \in B_i$, which proves the result. □

We now show that the open-sets version implies the following "mixed" version.

Theorem 5.4 (Open or closed version)**.** *Let $A_1, \ldots, A_{n+1}$ be subsets of S^n, each of which is open or closed, with union equal to S^n. Then there exist $x \in S^n$ and $i \in \{1, 2, \ldots, n+1\}$ such that $x \in A_i$ and $-x \in A_i$.*

Theorem 5.3 implies Theorem 5.4. We prove this by induction. Suppose that $A_1, \ldots, A_t$ are closed and $A_{t+1}, \ldots, A_{n+1}$ are open. If $t = 0$ then the result is Theorem 5.3 so we are done by hypothesis. Otherwise, suppose that A_t does not contain a pair of antipodal points. Define a function f by $f(x) = \max\{d(x, A_t), d(-x, A_t)\}$. Then f is continuous and never takes the value zero, so is bounded below by some $\epsilon > 0$. Write $(A_t)_\epsilon$ for the set $\{x \in S^n : d(x, A_t) < \epsilon\}$. Then $(A_t)_\epsilon$ is open, contains A_t and does not contain a pair of antipodal points. But then by our inductive hypothesis at least one of the sets $A_1, \ldots, A_{t-1}, A_{t+1}, \ldots, A_{n+1}$ contains a pair of antipodal points and we are done. □

Note that this last result includes the case where all the sets are closed, so all the results we have proved so far are equivalent. It may look as though the last theorem is of interest principally as an inductive step useful for getting from the all-open version to the all-closed version, but actually we shall be using one of the intermediate cases.

To get a more intuitive feel for why the mixed result is true, it is helpful to consider the case $d = 1$ and assume that the sets C_1 and C_2 are both intervals, or rather circular arcs. If they could be half-open intervals, then it would be easy to ensure that neither contained a pair of antipodal points – for example, if we identify the unit circle with the interval $[0, 2\pi)$ in the usual way, then C_1 could be the interval $[0, \pi)$ and C_2 the interval $[\pi, 2\pi)$. But if one of the intervals is required to contain both end points, then we can no longer do this, or anything like it.

5.2 What has the Borsuk-Ulam theorem got to do with combinatorics?

You may possibly have encountered the beautiful problem of proving that there exists a triangle-free graph with chromatic number at least 2020 (or whatever the year happens to be). There are several different approaches to this problem, of which one uses so-called *sphere graphs*. These are graphs where the vertex set is the unit sphere in $\mathbb{R}^d$ for some d, and two vertices are joined if and only if their inner product satisfies some given inequality (or equivalently the angle between them satisfies some given inequality).

One simple proof of the result uses the open-sets version of the Borsuk-Ulam theorem. Let us state it slightly differently to bring out its relevance to colouring: if we colour the sphere S^d with d open colours (allowing points to have more than one colour), then there

will necessarily be a pair of antipodal points that have the same colour (or more precisely have a colour in common).

By slightly modifying this statement, we can remove any topological requirements on the colours.

Theorem 5.5. *Let $\delta > 0$ and define a graph on S^d by joining vectors u and v if $\langle u, v\rangle < -1 + \delta$. Then this graph has chromatic number at least $d + 2$.*

Proof. If u and v are unit vectors, then $\|u+v\|^2 = 2+2\langle x, y\rangle$. It follows that $\langle u, v\rangle < -1+\delta$ if and only if $\|u + v\|^2 < 2\delta$ (which is saying that the distance from u to $-v$ is less than $\sqrt{2\delta}$).

Now let $C_1, \dots, C_{d+1}$ be sets that partition S^d, let $\epsilon = \sqrt{\delta/2}$, and for each i define U_i to be the set of all points x such that $d(x, C_i) < \epsilon$. Then U_i is open for each i, so by Theorem 5.3 there exists i such that U_i contains a pair $x, -x$ of antipodal points. By the definition of U_i, C_i contains points u, v with $\|u - x\| < \epsilon$ and $\|v + x\| < \epsilon$, which implies by the triangle inequality that $\|u + v\| < 2\epsilon = \sqrt{2\delta}$, which as we saw above implies that u and v are joined by an edge. Therefore, there is no proper colouring with fewer than $d + 2$ colours. □

Remark 5.6. If δ is small enough, then the bound above is sharp. We have basically seen a construction that demonstrates this in the previous subsection: we can take $d + 2$ unit vectors $u_1, \dots, u_{d+2}$ in $\mathbb{R}^{d+1}$ that form the vertices of a regular simplex, and let C_i be the set of x such that $\langle x, u_i\rangle > \theta$. Provided θ is at most about d^{-1}, these sets cover all of S^d, and the largest distance between two vectors in C_i is slightly less than 2. I'll leave it as an exercise to work out the details more precisely.

Remark 5.7. If δ is small, then two points that are joined to each other must be almost antipodal. This clearly implies that any odd cycle in the graph must be quite long. (I'll leave it as another exercise to prove a precise statement along these lines.) In particular, this construction gives us a big supply of graphs with no short odd cycles and high chromatic number.

5.3 Proof of Kneser's conjecture

Now that you have seen that argument, it should come as less of a surprise that the Borsuk-Ulam theorem will play a central role in the solution of Kneser's conjecture. Indeed, we can think of the Kneser graph as rather like a discrete version of the sphere graph just discussed, since the condition that two sets are joined by an edge is quite similar to the condition that two points should be almost antipodal. However, because of the discreteness of the Kneser graph, is it is not completely obvious how to turn that thought into a rigorous proof. Lovász was the first to do it, and soon afterwards Imre Bárány surprised everybody by finding a much shorter proof (still based on the Borsuk-Ulam theorem). It was thought for a long time that Bárány's proof was the "right" argument, so it came as yet another surprise when almost a quarter of a century later Joshua Greene, who was a student at

the time, came up with the following even shorter proof. He began by observing Theorem 5.4, the "mixed" version of the Borsuk-Ulam theorem we proved in the previous section.

And then came the crazily short proof of Lovász's theorem.

Theorem 5.8. *The chromatic number of the Kneser graph $G_{n,k}$ is $k+2$.*

Proof. We are free to choose any set of size $2n+k$ as the ground set, so let us choose $2n+k$ points in general position in the sphere S^k, which we regard as a subset of $\mathbb{R}^{k+1}$, where the key property we require is that no $k+2$ points are contained in any proper subspace of $\mathbb{R}^{k+1}$. Let E be the set of points chosen.

Let $C_1, \dots, C_{k+1}$ be sets that partition $E^{(n)}$, and let $\phi : [2n+k] \to S^k$ be a map whose image is in general position. In particular, the property we need is that no $k+2$ points of the image should lie in a proper subspace of $\mathbb{R}^k$.

For each i, let $X_i \subset S^k$ be the set of all x such that the open hemisphere $H(x) = \{y : \langle x, y\rangle > 0\}$ contains a set A that belongs to C_i. Also, let $X_0 = S^k \setminus (X_1 \cup \dots \cup X_{k+1})$. The sets X_i with $1 \le i \le k+1$ are open, so X_0 is closed. Therefore, by Theorem 5.4 some X_i contains a pair of antipodal points.

If this happens for $i \ge 1$, then since $H(x)$ and $H(-x)$ are disjoint, we find that C_i contains two disjoint sets and we are done. If it happens for $i = 0$, then the hemispheres $H(x)$ and $H(-x)$ do not contain any sets of size n, so there must be at least $k+2$ points in the subspace $\{y : \langle x, y\rangle = 0\}$, which contradicts the fact that the points are in general position. □

6 Marcus and Tardos's solution to the Stanley-Wilf conjecture

Let $\pi \in S_k$ be a permutation, let $n \ge k$, and let $\sigma \in S_n$ be another permutation. Given any k numbers $x_1 < \dots < x_k \in [n]$, we can obtain a new sequence $\sigma(x_1), \dots, \sigma(x_k)$ and let the ordering of $\sigma(x_1), \dots, \sigma(x_k)$ define a permutation τ of $[k]$ in a natural way: for each $r \le k$ we define $\tau(r)$ to be s if $\sigma(x_r)$ is the sth smallest element of the set $\{\sigma(x_1), \dots, \sigma(x_k)\}$. If we can find $x_1, \dots, x_k$ such that $\tau = \pi$, then we say that σ *contains* π.

An equivalent and more economical way of defining containment is to say that σ contains π if there exist $x_1 < \dots < x_k$ such that $\sigma(x_i) < \sigma(x_j)$ if and only if $\pi(i) < \pi(j)$.

For example, let π be the permutation 2413 in S_4. (Here the notation simply means that 1 maps to 2, 2 maps to 4, 3 maps to 1 and 4 maps to 3: it is not supposed to represent the 4-cycle (2413).) Then the permutation 251983467 contains π, as we see from the subsequences 2513, 2514, 5836, 5837, 5936, and 5937, but not, for example, from the subsequence 2536.

If σ is a permutation that does not contain π, we say that σ is π-*avoiding*. An example of a permutation that avoids 2413 is 152637489. More generally, suppose that $\sigma \in S_n$, $0 < r < n$, and there is a partition of $[n]$ into a set A of size r and a set B of size $n-r$ such that σ maps A to $\{1, 2, \dots, r\}$ in the unique order-preserving way and maps B to

$\{r+1,\dots,n\}$ in the unique order-preserving way. Then σ cannot contain 2413, since in order to find $x_2 < x_3$ such that $\sigma(x_2) > \sigma(x_3)$ it is necessary to take $x_2 \in B$ and $x_3 \in A$. But then x_1 cannot be in A since we would have $\sigma(x_1) < \sigma(x_3)$, and x_4 cannot be in B for similar reasons. But that implies that $\sigma(x_1) > \sigma(x_4)$.

The number of permutations we have just constructed is of size 2^n, since the set A determines the permutation uniquely. However, these are not the only ones: they were enumerated exactly in a paper of Miklós Bóna in 1997. Nevertheless, it turns out that there are only exponentially many. Since $n!$ is around $(n/e)^n$, so superexponential, this shows that avoiding the permutation 2413 is a very strong condition.

The Stanley-Wilf conjecture was that for *every* permutation π, there exists a constant $C = C(\pi)$ such that the number of permutations in S_n that avoid π is at most C^n for every n. It was posed in the late 1980s, received a lot of attention, and was finally solved by Adam Marcus and Gábor Tardos in 2004 with a remarkably simple argument.

Actually, Marcus and Tárdos solved a different problem that was known to imply the Stanley-Wilf conjecture, so let us first discuss that problem and its solution and then see how the implication works.

The problem solved by Marcus and Tárdos concerns a closely related notion of containment that concerns 01-matrices. We say that an $n \times n$ matrix A contains a $k \times k$ matrix B if we can find $x_1 < \cdots < x_k$ and $y_1 < \cdots < y_k$ such that $A_{x_i y_j} = 1$ whenever $B_{ij} = 1$. Two important things to note about this definition are that we do not insist that $A_{x_i y_j} = 0$ if $B_{ij} = 0$, and that the order matters. The second point is particularly important, so let us look at it a different way. We can regard A and B as adjacency matrices of bipartite graphs $G(A)$ and $G(B)$ (so for example $G(A)$ has two vertex sets X, Y that are copies of $[n]$, with $i \in X$ joined to $j \in Y$ if and only if $A_{ij} = 1$). Then the condition that A contains B is saying more than that $G(A)$ contains $G(B)$ as a subgraph: it is saying that there is an embedding of $G(B)$ into $G(A)$ in such a way that the orderings of the two sets of vertices are preserved.

As an an example, the matrix $\begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$ contains the matrix $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$, as is shown by the entries that are highlighted in red.

As with permutations, we say that a matrix A *avoids* a matrix B if it does not contain B.

Of particular interest is the case where B is a permutation matrix. We can ask how many $n \times n$ matrices avoid B, but actually the question Marcus and Tárdos looked at was how many non-zero entries such a matrix can have. It had been conjectured by Füredi and Hajnal that for any given B the number was at most linear in n. This was the conjecture that Marcus and Tárdos proved.

Theorem 6.1 (Marcus, Tárdos)**.** *Let P be a permutation matrix. Then there exists a constant $C = C(P)$ such that every $n \times n$ 01-matrix that avoids P has at most Cn non-*

zero entries.

Before we move on to the proof, let us briefly look in a slightly different way at what it means to contain a permutation matrix. Define an *ordered partition* of $[n]$ to be a partition into sets $X_1, \dots, X_m$ such that if $i < j$, then every element of X_i is less than every element of X_j.

A matrix A contains a permutation matrix $P \in S_k$ if and only if there exist ordered partitions $R_1, \dots, R_k$ and $C_1, \dots, C_k$ of the rows and columns of A such that if $P_{ij} = 1$ then there is a non-zero entry in the submatrix of A given by the rows in R_i and the columns in C_j. To illustrate this with a diagram,

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \text{ is contained in } \left(\begin{array}{cc|cc|c} 1 & 1 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ \hline 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right)$$

because each of the red blocks contains at least one 1. (Note that other blockings would have worked just as well.) Thus, we can think of matrix containment as saying that the small matrix is a "subset of a quotient" of the big matrix. We shall see that blockings like this play an important part in the proof of Marcus and Tárdos.

Let us make this observation more precise. Let A be an $n \times n$ 01-matrix, let $R_1, \dots, R_k$ and $C_1, \dots, C_k$ be as above, and let B be the $k \times k$ matrix obtained by setting $B_{ij} = 1$ if and only if there is a 1 in the block $R_i \times C_j$ – that is, a 1 whose row belongs to R_i and whose column belongs to C_j. We shall call such a matrix B an *ordered quotient* of A.

Lemma 6.2. *Let P be a permutation matrix and let A be an 01-matrix that avoids P. Then every ordered quotient of A avoids P.*

Proof. Let $\sigma \in S_k$, let P be the corresponding permutation matrix (that is, $P_{ij} = 1$ if $j = \sigma(i)$ and 0 otherwise), and suppose that A has an ordered quotient B that contains P. Let $R_1, \dots, R_t$ and $C_1 \dots, C_t$ be the ordered partitions that determine the quotient. Let $x_1 < \cdots < x_k$ and $y_1 < \cdots < y_k$ be such that $B_{x_i y_{\sigma(i)}} = 1$ for $i = 1, 2, \dots, k$, and for each i pick an entry $A_{r_i s_i} = 1$ such that $r_i \in R_{x_i}$ and $s_i \in C_{y_{\sigma(i)}}$. Since $x_1 < \cdots < x_k$, $y_1 < \cdots < y_k$ and the partitions are ordered, it follows that $r_1 < \cdots < r_k$ and $s_1 < \cdots < s_k$. Therefore, A contains P. □

Remark 6.3. If that proof did not seem trivial to you, then I recommend drawing a diagram, or looking at the diagram of the block matrix given above.

Now let us start the proof in earnest. Let us fix a $k \times k$ permutation matrix P and for each positive integer n let $f(n)$ be the largest possible number of non-zero entries in an $n \times n$ 01-matrix that does not contain P. For convenience, let us assume for the time being that n is divisible by k^2. This allows us to take ordered partitions $R_1, \dots, R_{n/k^2}$ and $C_1, \dots, C_{n/k^2}$ of the rows and columns, with each R_i and C_j of size k^2.

Given an $n \times n$ 01-matrix A with $k^2|n$, let us write A_{ij} for the submatrix obtained by restricting to the rows in R_i and columns in C_j. Call these $k^2 \times k^2$ submatrices *blocks*. The key to the proof is the following definition. We shall call a block *wide* if at least k different columns of the block contain a 1, and *tall* if at least k different rows contain a 1. The structure of the proof will be to show that if A does not contain P, then there cannot be too many wide blocks, or too many tall blocks, or too many non-empty blocks that are neither wide nor tall.

Lemma 6.4. *If A does not contain a given $k \times k$ permutation matrix P, then for each j the number of wide blocks with columns in C_j is at most $(k-1)\binom{k^2}{k}$ and for each i the number of tall blocks with rows in R_i is at most $(k-1)\binom{k^2}{k}$.*

Proof. If there are more than $(k-1)\binom{k^2}{k}$ wide blocks with columns in C_j, then by the pigeonhole principle we must be able to find a set K of k columns in C_j and a set B of k blocks such that every block in B contains a 1 in every column in K. But then the submatrix of A that is formed out of the blocks in B (placed one on top of another) has a $k \times k$ ordered quotient with 1s everywhere: we split the rows according to which block of B they belong to, and we split the columns in such a way that each of the chosen k columns belongs to a different cell of the partition. Therefore, A contains the $k \times k$ all-1s matrix, which trivially implies that it contains P.

The same argument, interchanging the roles of rows and columns, proves the statement for tall blocks. □

Lemma 6.4 allows us to prove a recursion for the number of entries that A can have.

Corollary 6.5. *Let P be a $k \times k$ permutation matrix, let n be divisible by k^2, and let $f(n)$ be the largest number of entries of any $n \times n$ 01-matrix that does not contain P. Then*

$$f(n) \leq 2k^2 n(k-1)\binom{k^2}{k} + (k-1)^2 f(n/k^2).$$

Proof. The number of wide blocks is at most $\frac{n}{k^2}(k-1)\binom{k^2}{k}$, by Lemma 6.4, and each such block contains at most k^4 1s. The same is true of tall blocks. So the number of 1s contained in wide or tall blocks is at most the first term in the expression above.

Let B be the ordered quotient of A defined by $R_1, \ldots, R_{n/k^2}$ and $C_1, \ldots, C_{n/k^2}$. Then by Lemma 6.2, B does not contain P. It follows that B has at most $f(n/k^2)$ non-zero entries. So the number of non-zero blocks of A is at most $f(n/k^2)$. We have already counted the entries that belong to wide or tall blocks. Any other block has at most $(k-1)^2$ non-zero entries, so the number of entries not so far counted is at most $(k-1)^2 f(n/k^2)$. This proves the result. □

The interesting part of the proof is over: it remains to obtain an upper bound for $f(n)$ from this recursion.

Proof of Theorem 6.1. If P is a $k \times k$ permutation matrix, we shall prove by induction on n that $f(n) \leq 2k^4 n\binom{k^2}{k}$.

If $n \leq k^2$, then trivially $f(n) \leq k^4 \leq k^4 n$, so we are done.

Otherwise, let m be the largest multiple of k^2 that is less than n. Then, again trivially, $f(n) \leq f(m) + 2k^2 n$. And by our inductive hypothesis and Corollary 6.5,

$$f(m) \leq 2k^2(k-1)m\binom{k^2}{k} + (k-1)^2.2k^4\frac{m}{k^2}\binom{k^2}{k}.$$

But

$$2k^2(k-1) + (k-1)^2.2k^2 = 2k^3(k-1),$$

so putting this together we find that

$$f(n) \leq 2(k^4 - k^3)m\binom{k^2}{k} + 2k^2 n \leq 2k^4 n\binom{k^2}{k}.$$

□

Remark 6.6. It is not immediately obvious how to find a matrix with more than $(k-1)^2 n$ non-zero entries that does not contain every $k \times k$ permutation matrix, and for a while it was conjectured that this should be the right bound. However Jacob Fox, in a very nice paper, proved that in fact the constant has to be exponential in k for almost all $k \times k$ permutation matrices. So the use of the pigeonhole principle above, which looks quite inefficient, is in fact not too wasteful.

Let us now turn to the Stanley-Wilf conjecture.

Corollary 6.7. *For every permutation $\pi \in S_k$ there is a constant C such that for every n there are at most C^n permutations $\sigma \in S_n$ that avoid π.*

Proof. Let P be the matrix of π and let α be such that every $n \times n$ 01-matrix with at least αn non-zero entries contains P. For each n let $T(n)$ be the number of $n \times n$ 01 matrices that avoid P. We claim that $T(2n) \leq 15^{\alpha n} T(n)$.

To see this, observe first that if we take ordered partitions into sets of size 2 of the rows and columns of a $2n \times 2n$ matrix that avoids P, then the number of distinct quotient matrices we can obtain is at most $T(n)$, by Lemma 6.2. But each such quotient matrix has at most αn non-zero entries, so the number of matrices with quotient equal to any given $n \times n$ matrix is at most $15^{\alpha n}$ (since there are four entries to choose in each non-empty block and they cannot all be zero). The recursion follows.

It follows immediately by induction that if n is a power of 2, then $T(n) \leq 15^{\alpha n}$. It is also easy to see that the number of permutations in S_n that avoid π is a non-decreasing function of n, so for general n we may deduce that $T(n) \leq 15^{2\alpha n}$. □

7 Entropy arguments

This section is slightly different from previous ones in that we shall need to develop a little bit of theory, and only after that will the arguments be very short. However, the theory consists of a few basic statements, and what really matters is not the proofs of those statements, but how to use them. To put it another way, I recommend treating those statements more like axioms than lemmas. To encourage this, I shall do so myself, but just to give you something to hold on to, which makes some of the axioms more intuitive, you should think of the entropy $H[X]$ of a discrete random variable X as a real number that measures the "information content" of X. Roughly speaking, this is how many bits of information you gain, on average, if you find out the value of X, or equivalently, the expected number of bits needed to specify X.

7.1 The Khinchin axioms for entropy and some simple consequences

Entropy has the following properties, which are called the Khinchin (or Shannon-Khinchin) axioms. (I got some of these from a set of lecture notes by Cosma Shalizi, which I recommend for further discussion of the pros and cons of an axiomatic approach.)

0. *Normalization.* If X takes the values 0 and 1, each with probability $1/2$, then $H[X] = 1$.

1. *Invariance.* $H[X]$ depends only on the probability distribution of X. That is, if $Y = f(X)$ for a function f that is bijective on the values taken by X, then $H[Y] = H[X]$.

2. *Maximality.* If X takes at most k distinct values, then $H[X]$ is maximized when X takes each value with equal probability $1/k$.

3. *Extensibility.* If X is a random variable that takes values in a finite set A and Y is a random variable that takes values in a set B with $A \subset B$, and if $\mathbb{P}[X = a] = \mathbb{P}[Y = a]$ for every $a \in A$ (and hence $\mathbb{P}[Y = b] = 0$ for every $b \in B \setminus A$), then $H[Y] = H[X]$.

4. *Additivity.* For any two random variables X and Y, $H[X, Y] = H[X] + H[Y|X]$, where
$$H[Y|X] = \sum_x \mathbb{P}[X = x] H[Y|X = x].$$

5. *Continuity.* $H[X]$ depends continuously on the probabilities $\mathbb{P}[X = x]$.

Axiom 0 is not really one of Khinchin's axioms, but the remaining axioms determine H only up to a multiplicative constant so it is there to fix that constant to a convenient value. Axioms 1-3 and 5 are rather basic properties of a kind that one might expect, but axiom 4 needs more comment. The quantity $H[X, Y]$ is simply the entropy (whatever that will turn out to mean) of the joint random variable (X, Y). The quantity $H[Y|X]$ is called the

conditional entropy of Y given X: it is the average entropy of Y given the value of X. Note that from its definition and the fact that H takes non-negative values (which implies that $H[Y|X = x]$ is non-negative for each x), it follows that $H[Y|X]$ is non-negative.

We now prove a sequence of lemmas, most of them very simple.

Lemma 7.1. *If X and Y are independent, then $H[Y|X] = H[Y]$ and $H[X, Y] = H[X] + H[Y]$.*

Proof. For each x the distribution of Y given that $X = x$ is the same as the distribution of Y, so $H[Y|X = x] = H[Y]$ for every x, by the invariance axiom. (The reason that axiom is needed is that strictly speaking the random variable $Y|X = x$ takes values of the form (x, y), where y is a value taken by Y.) It follows that

$$H[Y|X] = \sum_x \mathbb{P}[X = x]H[Y|X = x] = \sum_x \mathbb{P}[X = x]H[Y] = H[Y].$$

The second statement then follows from the additivity axiom. □

Lemma 7.2. *If X takes just one value, then $H[X] = 0$.*

Proof. $H[X, X] = H[X]$, by the invariance axiom. But X and (X, X) are independent, so $H[X, X] = 2H[X]$, by Lemma 7.1. □

And another.

Lemma 7.3. *Let $A \subset B$, let X be uniformly distributed on A, and let Y be uniformly distributed on B. Then $H[X] \leq H[Y]$, with equality if and only if $A = B$.*

Proof. By extensibility, $H[X]$ is not affected if we regard it as taking values in B. The inequality then follows from the maximality axiom.

Suppose now that $|A| = r$, $|B| = s$, and $r < s$. If $r = 1$, then the result follows from Lemma 7.2, the normalization axiom, and what we have just proved.

Otherwise, denote by X^n the A^n-valued random variable given by n independent copies of X, and similarly for Y. Then for any n, Lemma 7.1 and induction imply that $H[X^n] = nH[X]$ and $H[Y^n] = nH[Y]$.

Now choose n such that $r^n \leq s^{n-1}$. Then $|A^n| \leq |B^{n-1}|$, so

$$nH[X] = H[X^n] \leq H[Y^{n-1}] = (n - 1)H[Y],$$

where the inequality follows from what we have just proved (together with the invariance axiom). Since $H[X] \geq 1$ (again by what we proved above), it follows that $H[X] < H[Y]$. □

And another.

Lemma 7.4. *Let X be a random variable and let $Y = f(X)$ for some function f. Then $H[Y] \leq H[X]$.*

Proof. By the invariance axiom, $H[X] = H[X, Y]$, since there is a bijection between values x taken by X and values $(x, f(x))$ taken by (X, Y). Therefore, by the additivity axiom, $H[X] = H[Y] + H[X|Y]$. □

In an earlier version of these notes, I assumed that H was non-negative, having failed to see a proof of non-negativity from the axioms. However, Sean Eberhard (a postdoc at Cambridge) pointed out to me the following argument.

Lemma 7.5. *$H[X] \geq 0$ for every discrete random variable X that takes values in a finite set A.*

Proof. First let us suppose that there exists n such that $p_a = \mathbb{P}[X = a]$ is a multiple of n^{-1} for every $a \in A$. Let Y be uniformly distributed on $[n]$, let $(E_a : a \in A)$ be a partition of $[n]$ such that $|E_a| = p_a n$ for each $a \in A$, and let Z be the random variable where $Z = a$ if $Y \in E_a$. Then Z and X are identically distributed, so $H[Z] = H[X]$, by invariance.

Now $H[Y, Z] = H[Z] + H[Y|Z]$, by additivity. Also, $H[Y, Z] = H[Y]$ by Lemma 7.4, since (Y, Z) depends only on Y. And finally, for each $a \in A$, $H[Y|Z = a]$ is uniformly distributed on a set of size at most n, so by Lemma 7.3 it follows that $H[Y|Z = a] \leq H[Y]$. This implies that $H[Y|Z] \leq H[Y, Z]$, and therefore that $H[X] = H[Z] \geq 0$.

In the general case, since we can approximate the probabilities p_a arbitrarily closely by multiples of n^{-1} for a suitably large n, we can apply the continuity axiom to obtain the same conclusion. □

Here is a slightly less simple lemma.

Lemma 7.6. *Let X be a random variable that takes at least two values with non-zero probability. Then $H[X] > 0$.*

Proof. Let A be the set of values taken by X, let $\alpha = \max_{a \in A} \mathbb{P}[X = a]$, and for each n denote by X^n the A^n-valued random variable that is given by n independent copies of X. Then the maximum probability of any value taken by X^n is α^n. Since $\alpha < 1$, for any $\epsilon > 0$ there exists n such that $\alpha^n < \epsilon$. It follows that we can partition A^n into two sets E and F, each of which has probability between $1/2 - \epsilon$ and $1/2 + \epsilon$. Now let Y be a random variable that takes the value 0 if $X^n \in E$ and 1 if $X^n \in F$. Then $H[X^n] = nH[X]$, by Lemma 7.1 (and induction), and also $H[X^n] = H[Y] + H[X^n|Y] \geq H[Y]$. But $H[Y] > 0$ for sufficiently small ϵ, by the normalization axiom and continuity. It follows that $H[X^n] > 0$ and therefore that $H[X] > 0$. □

We end this sequence of lemmas with a result that is often useful. It is sometimes known as the *chain rule* for entropy.

Lemma 7.7. *Let $X_1, \dots, X_k$ be random variables taking values in a set A. Then*

$$H[X_1, \dots, X_k] = H[X_1] + H[X_2|X_1] + H[X_3|X_1, X_2] + \cdots + H[X_k|X_1, \dots, X_{k-1}].$$

Proof. By additivity,

$$H[X_1, \dots, X_k] = H[X_1, \dots, X_{k-1}] + H[X_k|X_1, \dots, X_{k-1}].$$

The result therefore follows by induction, with the additivity axiom as the base case. □

7.2 The number of paths of length 3 in a bipartite graph

Let us now consider the following problem. Suppose that G is a bipartite graph with finite vertex sets A and B and density α. (The density is defined to be the number of edges divided by $|A||B|$.) A *labelled* $P3$ is a quadruple (x_1, y_1, x_2, y_2) such that $x_1, x_2 \in A$, $y_1, y_2 \in B$, and x_1y_1, y_1x_2, and x_2y_2 are all edges of G. In other words, it is a path of length 3 in the graph, but we allow degeneracies such as $x_1 = x_2$.

How many labelled $P3$s must a bipartite graph with density α contain? If G is bi-regular, meaning that every vertex in A has degree $\alpha|B|$ and every vertex in B has degree $\alpha|A|$, then there are $\alpha^3|A|^2|B|^2$, since we can choose x_1 in $|A|$ ways, then y_1 in $\alpha|B|$ ways, then x_2 in $\alpha|A|$ ways, and finally y_2 in $\alpha|B|$ ways. We shall now show that this is the smallest number of labelled $P3$s that there can be. The proof will assume that entropy exists – that is, that there is some H that satisfies the Khinchin axioms. Later we shall see that this assumption is valid, and that will put in the final piece of the jigsaw.

Before you read the proof, I strongly recommend you try to prove the result for yourself by elementary means, since it looks as though it ought to be possible (given that the result is true), but turns out to be surprisingly tricky, and if you haven't experienced the difficulty, then you won't appreciate the power of the entropy approach.

How does one use entropy to prove results in combinatorics? Part of the answer lies in axiom 2. Suppose that X is uniformly distributed on a set of size k, and Y is a random variable with $H[Y] \geq H[X]$. then axiom 2 implies that Y takes at least k different values. Therefore, if we want to prove that a set A has size at least k, one way of doing it is to find a random variable that takes values in A and has entropy at least $H[X]$.

At first this may seem a very strange approach, since by axiom 2 we know that if there is such a random variable, then a random variable that is uniformly distributed on A will also work. If we define $f(n)$ to be the entropy of a random variable that is uniformly distributed on a set of size n, then all we seem to be doing is replacing the cardinality of a set by f of that cardinality, which doesn't look as though it will achieve anything.

However, there is a flaw in that criticism, which is that it might in principle be easier to obtain a lower bound for the entropy of a carefully chosen distribution on a set A (given certain assumptions about A) than it is to find a lower bound on the cardinality of A. And indeed, this turns out to be the case in many interesting situations, including the one at hand.

We wish to obtain a lower bound for the number of labelled $P3$s in a bipartite graph G of density α, and to do so we shall obtain a lower bound for the entropy of the following distribution on the set of labelled $P3$s, which is *not* uniform (except when the graph is regular). We choose an edge x_1y_1 (with $x \in A$ and $y \in B$ uniformly at random, then a vertex x_2 uniformly from the neighbours of y_1, and then a vertex y_2 uniformly from the neighbours of x_2.

Let X_1, Y_1, X_2, and Y_2 be the distributions of x_1, y_1, x_2, and y_2, respectively. We now wish to say something about the entropy $H[X_1, Y_1, X_2, Y_2]$. The chain rule (Lemma 7.7) tells us that it is equal to

$$H[X_1, Y_1] + H[X_2|X_1, Y_1] + H[Y_2|X_1, Y_1, X_2].$$

Now

$$H[X_2|X_1,Y_1] = \sum_{a\in A}\sum_{b\in B}\mathbb{P}[X_1=a,Y_1=b]H[X_2|X_1=a,Y_1=b].$$

But for each fixed b, the distributions of X_1 and X_2 given that $Y_1 = b$ are independent: the way we choose x_2 once we have chosen y_1 depends entirely on y_1 and not on how y_1 was obtained. Therefore, this simplifies to

$$\begin{aligned}\sum_{a\in A}\sum_{b\in B}\mathbb{P}[X_1=a,Y_1=b]H[X_2|Y_1=b] &= \sum_{b\in B}\mathbb{P}[Y_1=b]H[X_2|Y_1=b]\\ &= H[X_2|Y_1].\end{aligned}$$

In a similar way, if we know the value of X_2, then the distribution of Y_2 is independent of the values of X_1 and Y_1, so

$$H[Y_2|X_1,Y_1,X_2] = H[Y_2|X_2].$$

So we are interested in finding a lower bound for

$$H[X_1,Y_1] + H[X_2|Y_1] + H[Y_2|X_2].$$

By the additivity axiom, we can write this as

$$H[X_1,Y_1] + H[Y_1,X_2] + H[X_2,Y_2] - H[Y_1] - H[X_2].$$

Notice that the first three terms are the entropies of the distributions of the three edges of the random labelled $P3$. We now make an important observation.

Lemma 7.8. *Given a random labelled $P3$ from the distribution defined above, the three edges are all uniformly distributed over all edges.*

Proof. The first edge is uniformly distributed by the definition of the distribution. Now the number of edges is $\alpha|A||B|$ and the number of edges x_1y_1 with $y_1 = b$ is $d(b)$ (the degree of b), so the probability that $Y_1 = b$ is $d(b)/\alpha|A||B|$, and the probability that $X_2 = a$ given that $Y_1 = b$ is 0 if ab is not an edge and $d(b)^{-1}$ if ab is an edge. So the probability that $(X_2,Y_1) = (a,b)$ is $1/\alpha|A||B|$ whenever ab is an edge, which is another way of saying that X_2Y_1 is uniformly distributed over all edges. And once we know that, then the same proof shows that X_2Y_2 is uniformly distributed. □

We also know that $H[Y_1]$ and $H[X_2]$ are at most as big as they would be if Y_1 and X_2 were uniformly distributed. So if we let X be uniformly distributed over A, Y be uniformly distributed over B, and E be uniformly distributed over all edges, then a lower bound for the entropy of (X_1,Y_1,X_2,Y_2) is

$$3H[E] - H[X] - H[Y].$$

Consider now the random variable $(X_1, Y_1, X_2, Y_2, X, Y)$, where (X_1, Y_1, X_2, Y_2) is as before, X is a random element of A, and Y is a random element of B, with X and Y independent of each other and of (X_1, Y_1, X_2, Y_2). By the above bound and Lemma 7.1 this random variable has entropy at least $3H[E]$, which is the entropy of the uniform distribution over all triples of edges (by Lemma 7.1). From this and Lemma 7.3, it follows that $|A||B|$ times the number of labelled $P3$s is at least the cube of the number of edges, which is $\alpha^3|A|^3|B|^3$, and from this we get that the number of labelled $P3$s is at least $\alpha^3|A|^2|B|^2$, as required.

The statement just proved is a special case of the following famous conjecture of Sidorenko.

Conjecture 7.9. *Let G be a bipartite graph with finite vertex sets X and Y and density α. Let H be another bipartite graph with vertex sets A and B and let ϕ be a random function that takes A to X and B to Y. Then the probability that $\phi(a)\phi(b)$ is an edge of G for every edge ab of H is at least $\alpha^{|E(H)|}$.*

In rough terms, this can be thought of as saying that if you want to minimize the number of copies of H in a bipartite graph of density α, then you cannot do better than to pick the edges of the bipartite graph independently at random with probability α. The conjecture has been proved for several classes of bipartite graphs, some by entropy methods, but in general it remains stubbornly open.

7.3 The formula for entropy

I once wrote a blog post about the above proof, in which I did things differently. There I defined entropy in a more usual way – by writing down a formula for it – and then I *calculated* entropies, or gave bounds in the form of specific numbers. When I started this section, I decided to try to do it axiomatically, but I wasn't sure how successful I would be. Having completed the exercise, I am now completely convinced that it is the right thing to do, as it makes it much clearer that the proof is comparing the entropy of one distribution with the entropy of another, rather than merely obtaining a numerical lower bound that gives the desired answer. Also, the proof in the blog post used Jensen's inequality, whereas this proof used the simpler maximality axiom.

So when I now give the formula, I recommend that you resist the temptation to latch on to it and use it the whole time. It should be a last resort – your proofs will be clearer if you can avoid it. It's a little like helping a much younger mathematician to get out of the habit of replacing $\sqrt{2}$ by 1.414.... At a certain age, one feels the need to experience the square root of 2 as a number with a decimal expansion, but with experience one comes to realize that what really matters is the "axioms for $\sqrt{2}$" which are

1. $\sqrt{2} > 0$.
2. $(\sqrt{2})^2 = 2$.

Of course, sometimes we use facts such as that $\sqrt{2} > 1$, but those can be deduced from the above properties and the ordered-field axioms that the reals satisfy.

With those remarks out of the way, here's the formula. If X is a discrete random variable taking values in a set A, then, writing p_a for $\mathbb{P}[X = a]$, we have

$$H[X] = \sum_{a \in A} p_a \log(1/p_a),$$

where the logarithm is to base 2. People often write this as $-\sum_{a \in A} p_a \log(p_a)$, a practice I don't like because it has a kind of clever-clever "You thought I was negative but I'm not!" aspect to it.

A quick example to help with orientation: if X is uniformly distributed over a set A of size n, then the formula tells us that $H[X] = \sum_{a \in A} n^{-1} \log n = \log n$. In particular, if $n = 2^k$, then the entropy is k, which reflects the intuitive idea that we need k bits of information to specify an element of A.

It is not hard to prove that this function satisfies the axioms given earlier. Normalization, invariance, extensibility and continuity are obvious. Maximality is a simple consequence of Jensen's inequality: the function $\log x$ is concave, so if A is any finite set, then for any random variable X taking values in A, if we write p_a for $\mathbb{P}[X = a]$, then the p_a are non-negative and sum to 1, so we have

$$\sum_{a \in A} p_a \log(1/p_a) \leq \log(\sum_{a \in A} p_a/p_a) = \log(|A|),$$

which is the entropy of a uniformly distributed random variable taking values in A.

As for the additivity axiom, let X take values in A and Y take values in B and let p_a, p_b and p_{ab} have obvious meanings (in particular, $p_{ab} = \mathbb{P}[X = a, Y = b]$). Then

$$H[X, Y] = \sum_{a \in A} \sum_{b \in B} p_{ab} \log(1/p_{ab}).$$

Now $p_{ab} = p_a \mathbb{P}[Y = b | X = a]$, so the right-hand side equals

$$\sum_{a \in A} \sum_{b \in B} \Big(p_{ab} (\log(1/p_a) + \log(1/\mathbb{P}[Y = b | X = a]) \Big).$$

But

$$\sum_{a \in A} \sum_{b \in B} p_{ab} \log(1/p_a) = \sum_{a \in A} p_a \log(1/p_a) = H[X],$$

and

$$\begin{aligned}
\sum_{a \in A} \sum_{b \in B} p_{ab} \log(1/\mathbb{P}[Y = b | X = a]) &= \sum_{a \in A} p_a \sum_{b \in B} \mathbb{P}[Y = b | X = a] \log(1/\mathbb{P}[Y = b | X = a]) \\
&= \sum_{a \in A} p_a H[Y | X = a] \\
&= \mathbb{E}_{a \in A} H[Y | X = a] \\
&= H[Y | X].
\end{aligned}$$

Thus, $H[X,Y] = H[X] + H[Y|X]$.

This proves that there is a function that satisfies the entropy axioms, and that completes the proof of the lower bound for the number of labelled $P3$s.

7.3.1 The axioms uniquely determine the formula.

It turns out that the formula we have given for entropy is the only one that satisfies the entropy axioms. To show this, we begin by working out $H[X]$ when X is uniformly distributed. (Here H is any function that satisfies the axioms for entropy. Logarithms are to base 2 throughout.)

Lemma 7.10. *If X is uniformly distributed on a set of size 2^k, then $H[X] = k$.*

Proof. Let Y be uniformly distributed on a set of size 2. Then $H[Y] = 1$ by the normalization axiom, which implies that $H[Y^k] = k$ by Lemma 7.1 and induction. Since Y^k is uniformly distributed on a set of size 2^k, $H[X] = k$ as well, by invariance. □

Lemma 7.11. *If X is uniformly distributed on a set of size n, then $H[X] = \log n$.*

Proof. For each r, X^r is uniformly distributed on a set of size n^r, and $H[X^r] = rH[X]$. Therefore, if $2^k \le n^r \le 2^{k+1}$, then $k \le rH[X] \le k+1$, by Lemma 7.3. It follows that $k/r \le H[X] \le (k+1)/r$ whenever $k/r \le \log n \le (k+1)/r$. This implies that $H[X] = \log n$ as claimed. □

Corollary 7.12. *Let X take values in a finite set A, with $p_a = \mathbb{P}[X = a]$. Then $H[X] = \sum_a p_a \log\left(\frac{1}{p_a}\right)$.*

Proof. First let us assume that there exists n such that p_a is a multiple of n^{-1} for every $a \in A$. As in the proof of Lemma 7.5 let Y be uniformly distributed on $[n]$, let Y be partitioned into sets E_a, one for each $a \in A$, with $|E_a| = p_a n$, and let us assume that $X = a$ if and only if $Y \in E_a$ (which by the invariance axiom loses no generality).

Then by additivity, $H[Y] = H[X] + H[Y|X]$. But by Lemma 7.11 $H[Y] = \log n$, and

$$H[Y|X] = \sum_a p_a H[Y|X=a] = \sum_a p_a H[Y|Y \in E_a] = \sum_a p_a \log(p_a n),$$

where for the last equality we again applied Lemma 7.11. It follows that

$$H[X] = \log n - \sum_a p_a(\log p_a + \log n) = \sum_a p_a \log\left(\frac{1}{p_a}\right).$$

As in the proof of of Lemma 7.5 we obtain the general case by applying the continuity axiom. □

7.4 Brégman's theorem

We shall now prove another result using entropy. For this result the final answer is a number, so the proof is not quite as purely based on the entropy axioms. However, the only numerical result needed is Lemma 7.11, the relatively simple result that the entropy of the uniform distribution on a set of size n is $\log n$.

We begin with a definition of an important combinatorial concept.

Definition 7.13. Let A be an $n \times n$ matrix. The *permanent* of A is the sum

$$\operatorname{per}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} A_{i\sigma(i)},$$

where S_n is the symmetric group on $\{1, 2, \dots, n\}$.

The permanent of a matrix can be thought of as "the determinant if you accidentally forget the signs". Unlike the determinant, which has many equivalent definitions and nice algebraic properties that allow one to calculate it efficiently, the permanent is extremely hard to calculate (a theorem of Leslie Valiant states that the problem is #P hard, which makes it at least as hard as any problem in NP, and possibly harder), so it may seem a little surprising that it is ever studied. But one hint that it is a natural concept, even if less fundamental than the determinant, is that if A is the bipartite adjacency matrix of a bipartite graph with both vertex sets X and Y labelled with the integers $\{1, 2, \dots, n\}$, then $\operatorname{per}(A)$ is the number of *perfect matchings* in that graph: that is, the number of bijections $\sigma : X \to Y$ such that $x\sigma(x)$ is an edge for every $x \in X$.

Given that one cannot easily calculate the permanent of a matrix, attention turns naturally to finding bounds. Brégman's theorem is a general upper bound for 01-matrices.

Theorem 7.14. *Let A be an $n \times n$ 01-matrix and for each i let r_i be the number of 1s in row i. Then*

$$\operatorname{per}(A) \leq \prod_{i=1}^{n} (r_i!)^{1/r_i}.$$

Note a simple special case: if A consists entirely of 1s, then the right-hand side gives the nth power of $(n!)^{1/n}$, which equals $n!$, which is the number of perfect matchings in the complete bipartite graph $K_{n,n}$. So the bound is sharp in this case. Note also that if A is a matrix that splits into blocks $\left(\begin{array}{c|c} A_1 & 0 \\ \hline 0 & A_2 \end{array}\right)$ then $\operatorname{per}(A) = \operatorname{per}(A_1)\operatorname{per}(A_2)$ and if we write $f(A)$ for the upper bound in Brégman's theorem, then we also have that $f(A) = f(A_1)f(A_2)$. So this gives us a reasonably wide class of matrices for which the bound is sharp.

We now give a proof, due to Jaikumar Radhakrishnan, that uses entropy. The theorem has a number of proofs, some quite short, so we do not claim that the use of entropy is essential: nevertheless it provides an interesting illustration of the technique.

Entropy proof of the theorem. Let G be the bipartite graph with bipartite adjacency matrix A: that is, the vertex sets of G are two copies X, Y of $[n]$ and xy is an edge of G if and only if $A_{xy} = 1$. Let σ be a perfect matching chosen uniformly at random from all perfect matchings in G. We shall obtain an upper bound for the entropy $H[\sigma]$. (To be clear, σ is a random variable with values in S_n and $H[\sigma]$ is the entropy of that random variable.) Brégman's theorem is equivalent to the assertion that the number of perfect matchings is at most $\prod_{x\in X}(d(x)!)^{1/d(x)}$, where $d(x)$ is the degree of x, which in turn is equivalent to the inequality

$$H(\sigma) \leq \sum_{x\in X} \frac{1}{d(x)} \log(d(x)!).$$

Let $x_1, x_2, \ldots, x_n$ be an enumeration of the vertices of X. (We can't just take the obvious enumeration here, since later we shall average over all enumerations – this is a key idea of the proof.) Then the chain rule (Lemma 7.7) tells us that

$$H[\sigma] = H[\sigma(x_1)] + H[\sigma(x_2)|\sigma(x_1)] + \cdots + H[\sigma(x_n)|\sigma(x_1), \ldots, \sigma(x_{n-1})].$$

We now bound the quantities on the right-hand side, starting with $H[\sigma(x_1)]$. We know that $\sigma(x_1)$ must be one of the $d(x_1)$ neighbours of x_1 in G. The probability that any given neighbour is chosen depends on the matrix A in a horribly complicated way, but since $\sigma(x_1)$ can take only $d(x_1)$ different values, we can at least bound $H[\sigma(x_1)]$ above by the entropy of the uniform distribution on a set of size $d(x_1)$, which by Lemma 7.11 is $\log(d(x_1))$.

Now let us turn to $H[\sigma(x_2)|\sigma(x_1)]$. If we are given $\sigma(x_1)$, then we know two things about $\sigma(x_2)$: that it is a neighbour of x_2, and that it is not equal to $\sigma(x_1)$. For each x and σ, let us define $d_1^\sigma(x)$ to be the number of neighbours of x not equal to $\sigma(x_1)$. Then an upper bound for $H[\sigma(x_2)|\sigma(x_1)]$ is $\mathbb{E}_\sigma \log(d_1^\sigma(x_2))$.

More generally, for each k, if we are given $\sigma(x_1), \ldots, \sigma(x_{k-1})$, then for each x, let $d_{k-1}^\sigma(x)$ be the number of neighbours of x not equal to any of $\sigma(x_1), \ldots, \sigma(x_k)$. Then an upper bound for $H[\sigma(x_k)|\sigma(x_1), \ldots, \sigma(x_{k-1})]$ is $\mathbb{E}_\sigma \log(d_{k-1}^\sigma(x_k))$, so

$$H[\sigma] \leq \mathbb{E}_\sigma \sum_{k=1}^{n} \log(d_{k-1}^\sigma(x_k)).$$

At this point we seem to be facing a serious difficulty, which is that it is hard to say anything about the distribution of d_{k-1}^σ, and hence hard to bound the expectations we need to bound. However, we still have a card up our sleeve, mentioned above, which is that we can average over all orderings $x_1, \ldots, x_n$ of $[n]$.

Let σ be fixed, and let $x_1, \ldots, x_n$ be a random ordering. Let x be a vertex in X and let us think about the contribution x makes to the sum

$$\sum_{k=1}^{n} \log(d_{k-1}^\sigma(x_k)),$$

which is what we wish to estimate.

Let the neighbours of x be $y_1, \dots, y_d$. Then the ordering of the vertices $\sigma^{-1}(y_1), \dots, \sigma^{-1}(y_d)$, which all belong to X, is distributed uniformly amongst all $d!$ possible orderings, and the position of x (which is $\sigma^{-1}(y_j)$ for some j) in this ordering is therefore uniformly distributed amongst the d possibilities. If x comes in position s in the ordering, and k (which depends on the ordering) is such that $x_k = x$, then $d^\sigma_{k-1}(x_k) = d - s + 1$. Thus, for any fixed σ, $d^\sigma_{k-1}(x_k)$ is uniformly distributed in the set $\{1, 2, \dots, d(x)\}$. It follows that

$$\mathbb{E} \sum_{k=1}^{n} \log(d^\sigma_{k-1}(x_k)) = \sum_{x \in X} \frac{1}{d(x)} \sum_{s=1}^{d(x)} \log(d(x) + s - 1) = \sum_{x \in X} \frac{1}{d(x)} \log(d(x)!),$$

where the expectation is over all orderings.

The above computation was for a fixed σ. But since the answer is independent of σ, we may conclude that

$$\mathbb{E}\mathbb{E}_\sigma \sum_{k=1}^{n} \log(d^\sigma_{k-1}(x_k)) = \sum_{x \in X} \frac{1}{d(x)} \log(d(x)!),$$

where again the first expectation is over all orderings.

Since $\mathbb{E}_\sigma \sum_{k=1}^{n} \log(d^\sigma_{k-1}(x_k))$ was an upper bound for $H[\sigma]$ for every ordering $x_1, \dots, x_n$, the result follows. □

7.5 Shearer's entropy lemma

We begin this section by proving a further useful fact about entropy. When I first wrote this section of the notes, I didn't manage to find an axiomatic proof so resorted to the formula. Subsequently, Chris West, a wonderful double bass player who read mathematics at Cambridge and has been following this course on YouTube, notified me that he had found an axiomatic proof. So here are two proofs, the first a boring one that uses the formula, and the second Chris West's much nicer axiomatic proof (which is almost certainly known, but I haven't found it anywhere online).

Lemma 7.15. *Let X and Y be discrete random variables. Then $H[X, Y] \leq H[X] + H[Y]$.*

Proof using the formula. Since $H[X, Y] = H[Y] + H[X|Y]$, by the additivity axiom, this is equivalent to the assertion that $H[X|Y] \leq H[X]$, which makes intuitive sense, since knowing the value of Y shouldn't *increase* the amount of information we obtain from X. To prove it, let A be the set of values taken by X, let B be the set of values taken by Y,

and write p_a for $\mathbb{P}[X = a]$, q_b for $\mathbb{P}[Y = b]$, and $p_{a,b}$ for $\mathbb{P}[X = a, Y = b]$. Then

$$\begin{aligned} H[X|Y] &= \sum_b q_b H[X|Y = b] \\ &= \sum_b q_b \sum_a \frac{p_{a,b}}{q_b} \log\left(\frac{q_b}{p_{a,b}}\right) \\ &= \sum_a p_a \sum_b \frac{p_{a,b}}{p_a} \log\left(\frac{q_b}{p_{a,b}}\right). \end{aligned}$$

Now $\sum_b \frac{p_{a,b}}{p_a} = 1$, so by Jensen's inequality this last expression is at most

$$\sum_a p_a \log\left(\sum_b \frac{p_{a,b}}{p_a}.\frac{q_b}{p_{a,b}}\right) = \sum_a p_a \log\left(\frac{1}{p_a}\right) = H[X].$$

□

Axiomatic proof. First consider the case where Y is uniformly distributed. Let notation be as in the previous proof. Then

$$H[Y|X] = \sum_{a \in A} p_a H[Y|X = a]$$

by the definition of conditional entropy. For each a, the random variable $[Y|X = a]$ takes values in B, so by maximality has entropy at most $H[Y]$, from which it follows that $H[Y|X] \leq H[Y]$, which is equivalent to the assertion that $H[X, Y] \leq H[X] + H[Y]$, and also to the assertion that $H[X|Y] \leq H[X]$.

Now suppose that q_b is rational for every $b \in B$. Then we can find a positive integer n such that every q_b is of the form m_b/n for a non-negative integer m_b. Now define a random variable Z as follows. For each b, let E_b be a set of size m_b, and let these sets be disjoint. If $Y = b$, let Z be chosen uniformly from E_b, and let this be done in such a way that $[Z|Y = b]$ and $[X|Y = b]$ are independent.

Then by this independence and by what we have already proved,

$$H[X|Y] = H[X|Y, Z] = H[X, Y, Z] - H[Y, Z] = H[X, Z] - H[Z] = H[X|Z] \leq H[X],$$

which implies that $H[X, Y] \leq H[X]+H[Y]$. The general case now follows by continuity. □

Although we shall not use it in this course, it is worth knowing a definition that arises naturally from the above fact. The *mutual information* $I[X; Y]$ is defined to be $H[X] + H[Y] - H[X, Y]$. Lemma 7.15 implies that the mutual information is non-negative. Other simple observations are that if $X = Y$ (or more generally X determines Y and vice versa) then $I[X; Y] = H[X]$, that if $Y = f(X)$, then $I[X; Y] = H[Y]$, and that if X and Y are

independent, then $I[X;Y] = 0$. One can think of the mutual information as being the amount of information that is "shared" by X and Y. The formula

$$H[X,Y] = H[X] + H[Y] - I[X;Y],$$

which follows from the definition, is highly reminiscent of the first case

$$|A \cup B| = |A| + |B| - |A \cap B|$$

of the inclusion-exclusion formula.

Starting with Lemma 7.15 and applying induction in the obvious way, we deduce that entropy has the following more general subadditivity property: if $X_1, \dots, X_n$ are random variables, then

$$H[X_1, \dots, X_n] \leq H[X_1] + \cdots + H[X_n].$$

Shearer's lemma, which has a number of applications, is a slightly less obvious generalization of this fact.

Theorem 7.16 (Shearer's lemma)**.** *Let S be a random subset of $[n]$, chosen according to some probability distribution, and suppose that for every i, $\mathbb{P}[i \in S] \geq \mu$. Then for any discrete random variables $X_1, \dots, X_n$,*

$$\mu H[X_1, \dots, X_n] \leq \mathbb{E}_S H[X_S],$$

where if $S = \{i_1, \dots, i_k\}$, then X_S is short for the random variable $(X_{i_1}, \dots, X_{i_k})$.

Note that if S is a random singleton, chosen uniformly, then μ can be taken to be n^{-1}, and we conclude that $n^{-1}H[X_1, \dots, X_n] \leq \mathbb{E}_i H[X_i]$, which is just a restatement of the subadditivity just mentioned.

Proof of Theorem 7.16. As in the statement of the theorem, let us write $S = \{i_1, \dots, i_k\}$, and let this be done in such a way that $i_1 < \cdots < i_k$. (Note that S is a random set, so the elements i_j are not fixed elements of $[n]$.)

By the chain rule for entropy (Lemma 7.7),

$$H[X_S] = H[X_{i_1}] + H[X_{i_2}|X_{i_1}] + \cdots + H[X_{i_k}|X_{i_1}, \dots, X_{i_{k-1}}].$$

Write $H[X_i|X_{<i}]$ for $H[X_i|X_1, \dots, X_{i-1}]$. Since conditioning on more variables reduces entropy, the above equality implies that

$$H[X_S] \geq \sum_{i \in S} H[X_i|X_{<i}].$$

Taking expectations, we deduce that

$$\begin{aligned}\mathbb{E}_S H[X_S] &\geq \mathbb{E}_S \sum_{i\in S} H[X_i|X_{<i}] \\ &= \sum_{i=1}^n \mathbb{P}[i \in S] H[X_i|X_{<i}] \\ &\geq \mu \sum_{i=1}^n H[X_i|X_{<i}] \\ &= \mu H[X_1, \dots, X_n],\end{aligned}$$

where the second inequality is by the main hypothesis of the theorem and the last equality is the chain rule again. □

We now give a representative application of Shearer's lemma. It gives a bound for the following question: if a graph has m edges, how many triangles can it have? It is natural to guess that the best construction will be a complete graph on n vertices, at least if we can find n such that $\binom{n}{2} = m$. In that case we have $\binom{n}{2}$ edges and $\binom{n}{3}$ triangles. These are roughly $n^2/2$ and $n^3/6$, which would suggest that the number of triangles is at most something along the lines of $(2m)^{3/2}/6$. This is in fact exactly the bound we shall obtain. But although the result is not unexpected, it is notable that for the proof we do not have to get our hands in the slightest bit dirty: there are no arguments, for instance, that "compress" the graph.

Theorem 7.17. *Let G be a graph with m edges and t triangles. Then $t \leq (2m)^{3/2}/6$.*

Proof. Let T be a triangle chosen uniformly at random from G and let $X = (X_1, X_2, X_3)$ be some ordering of its vertices. Then X is uniformly distributed on a set of size $6t$, so $H[X] = \log(6t)$.

We now let S be a random pair of elements of $\{1, 2, 3\}$ (with each pair chosen with probability $1/3$. Then we can set $\mu = 2/3$ in Shearer's lemma, from which we deduce that

$$\frac{2}{3} H[X] \leq \mathbb{E}_S H[X_S].$$

It follows (by averaging again) that there is some pair S such that $\frac{2}{3}H[X] \leq H[X_S]$. But the values taken by X_S are ordered pairs of vertices that form end-points of edges, so it is supported on a set of size $2m$. It follows that $H[X_S] \leq \log(2m)$. We conclude that

$$\frac{2}{3} \log(6t) \leq \log(2m),$$

which is equivalent to the assertion of the theorem. □

Now for a somewhat more subtle use of Shearer's lemma. We shall use it to obtain a bound on the following natural problem. Let $\mathcal{G}$ be a family of graphs with vertex set $[n]$. We say that $\mathcal{G}$ is Δ-intersecting if for any two graphs $G_1, G_2 \in \mathcal{G}$, the intersection $G_1 \cap G_2$ contains a triangle. And now we ask the obvious question: how large can a Δ-intersecting family be?

As with intersecting families, there is also an obvious construction: just take all graphs that contain some fixed triangle. This gives us a family of size $2^{\binom{n}{2}}/8$. In the other direction, if we just use the fact that any two graphs in a Δ-intersecting family have a non-empty intersection, we get an upper bound of $2^{\binom{n}{2}}/2$. (You might ask whether we can do better by using the fact that the intersection of any two graphs in the family has size at least 3, but the set of graphs with at least $\frac{1}{2}\binom{n}{2}+2$ edges has that property and its size is not that much smaller than $2^{\binom{n}{2}}/2$.)

It turns out that the bound $2^{\binom{n}{2}}/8$ *is* the best possible – this is a relatively recent result of David Ellis, Yuval Filmus and Ehud Friedgut – but for a long time the best known bound was $2^{\binom{n}{2}}/4$, which we shall prove now, using Shearer's lemma. This result is due to Chung, Frankl, Graham, and Shearer himself.

Theorem 7.18. *A Δ-intersecting family $\mathcal{G}$ of graphs with vertex set $[n]$ has size at most $2^{\binom{n}{2}}/4$.*

Proof. As usual, we let X be an element of $\mathcal{G}$ chosen uniformly at random, and our aim will be to obtain an upper bound for $H[X]$, which will translate directly into an upper bound for $|\mathcal{G}|$. In the argument that follows, it may help to think of each element of $\mathcal{G}$ as a function from $[n]^{(2)}$ to $\{0,1\}$ – the characteristic function of the graph – so we are thinking of X as a random variable $(X_e)_{e\in[n]^{(2)}}$, where $X_e = 1$ if e belongs to the graph and 0 otherwise.

To apply Shearer's lemma, we shall choose S as follows. We first pick a set $R \subset [n]$ uniformly at random, and we let G_R be the graph that consists of a clique with vertex set R and a clique with vertex set $[n] \setminus R$. Note that every edge e has a probability $1/2$ of being an edge of R. Therefore, if we define X_R to be the intersection $X \cap G_R$, which corresponds to the random variable $(X_e)_{e\in R}$, then Shearer's lemma gives us that

$$\frac{1}{2}H[X] \leq \mathbb{E}_R H[X_R].$$

Now we use the fact that $\mathcal{G}$ is Δ-intersecting. Since any two graphs in $\mathcal{G}$ intersect in at least a triangle, and since every triangle intersects every graph G_R, we find that the restriction of $\mathcal{G}$ to any G_R is an intersecting family, and therefore has size at most $2^{|G_R|-1}$. Therefore, $H[X_R] \leq |G_R| - 1$ for every R. This gives us that

$$H[X] \leq 2\mathbb{E}_R|G_R| - 2$$

Since each element of $[n]^{(2)}$ has a probability $1/2$ of belonging to G_R, the right-hand side is equal to $\binom{n}{2} - 2$, and this proves the result. □

It's worth reflecting a little on what Shearer's lemma did for us in the above argument. It was a little bit similar in spirit to the averaging argument that proved the Erdős-Ko-Rado theorem, in that we showed that a suitably chosen "random part" of our Δ-intersecting family was not too large, and then we exploited that to get a bound on the size of the whole family. However, unlike in Katona's argument, the "random part" was not the intersection of the family with a random set (such as the set of intervals in a cyclic order) but rather the *projection* of the family on to a random set of coordinates, where each coordinate had an equal probability of belonging to the set. So Shearer's lemma tells us that if we can find a nicely balanced set of projections of a set system and can give good upper bounds for their sizes, then we can get a good upper bound for the size of the whole family. This general theme of bounding the size of a set in terms of the size of its projections is a significant one. (See for example the Loomis-Whitney inequality and its generalizations.)

8 The polynomial method

It is not very easy to say exactly what the polynomial method is, but in broad terms it is exploiting facts about zeros of polynomials to prove results that do not appear to have anything to do with polynomials. Typically, one tries to prove that a small combinatorial structure cannot exist by showing that if it did, then it would allow us to define a non-zero polynomial of low degree that vanishes on a large set, sometimes with some extra structure, which we can then contradict by proving results that show that non-zero polynomials of low degree cannot vanish on such sets.

8.1 Dvir's solution to the Kakeya problem for finite fields

In this section we shall illustrate the method with three results. The first is a celebrated result of Dvir, who solved the so-called Kakeya problem for finite fields. This is the following question: suppose that A is a subset of $\mathbb{F}_p^n$ that contains a translate of every line. Must A have size $p^{n-o(1)}$? Here, we are taking n to be fixed and p to be tending to infinity. Dvir's solution gave the following theorem.

Theorem 8.1. *Let A be a subset of $\mathbb{F}_p^n$ that contains a translate of every line. Then $|A| \geq c(n)p^n$.*

The proof needs a couple of simple (but surprisingly powerful) lemmas.

Lemma 8.2. *Let $A \subset \mathbb{F}_p^n$ be a set of size less than $\binom{n+d}{d}$. Then there exists a non-zero polynomial $P(x_1, \dots, x_n)$ of degree d that vanishes on A.*

Proof. A polynomial of degree d in the variables $x_1, \dots, x_n$ is a linear combination of monomials of degree at most d. The number of monomials of degree k is the number of ways of writing a number less than or equal to d in the form $a_1 + \cdots + a_n$ with each a_i non-negative. By a standard holes-and-pegs argument, this is $\binom{n+d}{d}$. (Given $n+d$ holes with d pegs, then a_i is the number of pegs that follow the ith hole.) Therefore, for a

polynomial of degree d to vanish at m given points, a certain set of m linear combinations of the coefficients must all be zero. By dimension considerations, this is possible if m is less than the number of coefficients. The result follows. □

When $d = p - 1$, this gives us that for every set A of size less than $\binom{n+p-1}{p-1} = \binom{n+p-1}{n}$ we can find a non-zero polynomial of degree d that vanishes on A. Since $\binom{n+p-1}{p-1} > p^n/n!$, this is in particular true when $|A| \leq p^n/n!$.

The next lemma is the main idea behind the proof of Dvir's theorem.

Lemma 8.3. *Suppose that $A \subset \mathbb{F}_p^n$ contains a line in every direction, that $d < p$, and that there exists a non-zero polynomial f of degree at most d that vanishes on A. Then there is a non-zero degree-d polynomial that vanishes everywhere on $\mathbb{F}_p^n$.*

Proof. Without loss of generality the degree of f is exactly d. Let $a, z \in \mathbb{F}_p^n$ with $z \neq 0$ and let L be the line consisting of all points $a + tz$. The restriction of f to L is a polynomial of degree d in t, and its leading coefficient is $f_d(z)$, where f_d is the degree-d part of f. To see this, observe that for any monomial $\prod_{i=1}^n x_i^{r_i}$ its value at $a + tz$ is $\prod_{i=1}^n (a_i + tz_i)^{r_i}$, so if the monomial has degree d, then to obtain a term in t^d we must choose tz_i from each bracket, which gives $t^d \prod_{i=1}^n z_i^{r_i}$.

Now if f vanishes everywhere on L, then since its dependence on t is given by a polynomial of degree less than p, all the coefficients of that polynomial must be zero. It follows that $f_d(z) = 0$. But z was an arbitrary non-zero element of $\mathbb{F}_p^n$, and f_d vanishes at zero as well, so it vanishes everywhere. □

To finish the proof, we need the second of our simple but powerful lemmas. Note that there is an important distinction between the zero polynomial, which is the polynomial with all coefficients of all monomials equal to zero, and a polynomial that takes the value zero everywhere on $\mathbb{F}_p^n$. For instance, the polynomial $x^p - x$ is not the zero polynomial but takes the value zero everywhere on $\mathbb{F}_p$.

Lemma 8.4. *Let f be a non-zero polynomial on $\mathbb{F}_p^n$ of degree less than p. Then f is not identically zero.*

Proof. We shall prove this by induction on n. If $n = 1$, then a non-zero polynomial that vanishes everywhere has p roots, so must be of degree p. Essentially the same argument works for general n. If f vanishes everywhere, then for each a it vanishes on the set $x_1 = a$. But the restriction of f to that set is a polynomial of degree less than p in the variables $x_2, \dots, x_n$, so by induction it is the zero polynomial.

In other words, if we substitute a for x_1 in the definition of f, we obtain the zero polynomial. If we think of f as a polynomial in x_1 with coefficients in $\mathbb{F}_p[x_2, \dots, x_n]$, then the polynomial division algorithm tells us that we can write $f(x)$ in the form $P(x_1, \dots, x_n)(x_1 - a) + Q(x_2, \dots, x_n)$, so if this polynomial is the zero polynomial when we substitute $x_1 = a$ we obtain that Q is the zero polynomial and $(x_1 - a)$ divides f. But if this is true for every a, then again we find that f has to have degree at least p, contradicting our assumption. □

Note that this result is sharp, as can be seen by considering the polynomial $p(x) = x_1^p - x_1$.

Proof of Theorem 8.1 Combine Lemmas 8.2, 8.3 and 8.4 and the result follows straight away, with $c(n) = 1/n!$. □

Since we are most of the way to a proof, we briefly discuss a result called the Schwartz-Zippel lemma, which tells us that a polynomial of degree d in n variables cannot have too many roots. Some sort of intuition for how many roots we might expect comes from thinking about linear polynomials: there we do not get more than p^{n-1} roots. A very useful and general principle in applications of arithmetic geometry is that polynomials behave in a rather similar way to linear functions. For example, a linear function from $\mathbb{R}$ to $\mathbb{R}$ has at most one root, while a polynomial of degree d has at most d roots, which is the same up to a constant.

Lemma 8.5. *Let f be a non-zero polynomial of degree at most d on $\mathbb{F}_p^n$. Then f has at most dp^{n-1} roots.*

Proof. Without loss of generality the degree is exactly d. As we have already seen, if we restrict f to the line consisting of points $a + tz$, then we obtain a polynomial of degree d in the single variable t with leading coefficient $f_d(z)$, where f_d is the degree-d part of f. By Lemma 8.4 we may choose z such that $f_d(z) \neq 0$. But that means that on every line L in direction z the restriction of f to L is given by a polynomial in one variable of degree d. So f has at most d roots in any line, and therefore at most dp^{n-1} roots in total. □

The bound here is sharp as well, as can be seen by considering polynomials that depend on x_1 only. Note also that the Schwarz-Zippel lemma implies Lemma 8.4.

One use of the Schwarz-Zippel lemma appears in theoretical computer science in connection with the general problem of *polynomial identity testing*. Here one is given two polynomials P and Q in n variables over a finite field $\mathbb{F}_p$, and the task is to decide whether they are in fact the same polynomial. If the polynomials are presented as sums of products of other polynomials, then it may be impractical to determine this by expanding everything out as a sum of monomials, of which there will be n^d if the polynomials have degree d, and in general the problem appears to be very hard.

However, the Schwarz-Zippel lemma provides a probabilistic algorithm. One selects r choices for $x_1, \dots, x_n$ independently at random and checks whether $P(x_1, \dots, x_n) = Q(x_1, \dots, x_n)$. If ever they differ, we know that the polynomials are not identical. Conversely, if P and Q have degree d and are different, then by the Schwarz-Zippel lemma the probability that we will not detect this with one of our random choices of $x_1, \dots, x_n$ is at most $(d/p)^r$, since for each choice we have to pick a root of $P - Q$ and the probability of choosing a root is at most d/p. So by making r large (but not ridiculously large) we can achieve an extremely high degree of confidence (falling short of total certainty) that either the two polynomials are the same or we will detect that they are different.

8.2 Alon's combinatorial Nullstellensatz

The next result, due to Noga Alon, concerns polynomials, but has a number of applications to problems that do not appear to have anything to do with polynomials.

Lemma 8.6. *Let f be a non-zero polynomial in n variables over $\mathbb{F}_p$ of degree $k_1 + \cdots + k_n$, where the k_i are non-negative integers and the coefficient of $x_1^{k_1} \dots x_n^{k_n}$ is non-zero. Let $S_1, \dots, S_n$ be subsets of $\mathbb{F}_p$ such that $|S_i| > k_i$ for each i. Then f does not vanish on $S_1 \times \cdots \times S_n$.*

Proof. We prove this by induction on the degree of f. The result is easy when the degree is zero.

If the degree is greater than zero, then without loss of generality $k_1 > 0$. Let $a \in S_1$ and use polynomial division to write $f(x)$ in the form $(x_1 - a)P(x) + Q(x)$, where Q does not depend on x_1. Since the term in $x_1^{k_1} \dots x_n^{k_n}$ has a non-zero coefficient in f, and the degree of P is $k_1 + \cdots + k_n - 1$, the term in $x_1^{k_1-1} x_2^{k_2} \dots x_n^{k_n}$ has non-zero coefficient in P.

Suppose that the conclusion is false. Then f vanishes on $\{a\} \times S_2 \times \cdots \times S_n$, from which it follows that Q vanishes on this set too. Since Q does not depend on x_1, it vanishes on all of $S_1 \times \cdots \times S_n$. Therefore, $(x_1 - a)P$ vanishes on $S_1 \times \cdots \times S_n$, which implies that P vanishes on $(S_1 \setminus \{a\}) \times S_2 \times \cdots \times S_n$. By induction it follows that P is also the zero polynomial, and we are done. □

As our first application of the combinatorial Nullstellensatz, we give a short proof of the Cauchy-Davenport theorem, which is the following result (discovered independently by Cauchy in 1813 and Davenport in 1935 – apparently it was not until 1947 that Davenport found out that Cauchy had beaten him to it by over a century). Neither Cauchy nor Davenport used the method below.

Theorem 8.7. *Let p be a prime and let A and B be subsets of $\mathbb{F}_p$. Then $|A + B| \geq \min\{p, |A| + |B| - 1\}$.*

Proof. Once we have the clue that this can be proved using the combinatorial Nullstellensatz, we need to find a suitable sequence of sets $S_1, \dots, S_n$ and a polynomial that vanishes on $S_1 \times \cdots \times S_n$ and that has small degree unless $A + B$ is large.

This gives enough of a clue to complete the proof. The obvious sequence of sets to take, since the only sets we have around are A and B, is (A, B). We want the degree of our polynomial to depend on the size of $A + B$, and we also want it to vanish on $A \times B$. The most economical way of getting it to vanish at a point (a, b) is to ensure that the polynomial has a factor $x + y - (a + b)$, which leads to the idea of considering the polynomial

$$f(x, y) = \prod_{c \in A+B} (x + y - c).$$

This vanishes on $A \times B$ and has degree equal to $|A + B|$, so it looks promising.

Suppose now that $|A| + |B| \leq p$ and $|A + B| \leq |A| + |B| - 2$. We want to contradict the combinatorial Nullstellensatz, so we need a monomial $x^r y^s$ with non-zero coefficient with

$r < |A|$, $s < |B|$ and $r + s = |A + B|$. But if we pick any r and s that satisfy the last three conditions, which we clearly can if $|A + B| \leq |A| + |B| - 2$, then the coefficient of $x^r y^s$ in the polynomial is $\binom{r+s}{r}$, and this is non-zero because p is prime and $r + s < p$.

If $|A| + |B| > p$, then for any x the sets A and $x - B$ intersect, so $|A + B| = p$. □

Note that this result is sharp if A and B are arithmetic progressions in $\mathbb{F}_p$ with the same common difference. Note too that the result is false in $\mathbb{Z}_n$ if n is composite, since then $\mathbb{Z}_n$ has proper subgroups

Now we shall use the combinatorial Nullstellensatz to prove a variant of the Cauchy-Davenport theorem. The result is due to da Silva and Hamidoune, who found a somewhat involved combinatorial proof. This short argument was discovered by Alon, Nathanson and Ruzsa. Let us write $A\dot{+}B$ for the set of sums $a + b$ such that $a \in A$, $b \in B$ and $a \neq b$.

Theorem 8.8. *Let p be a prime and let A and B be subsets of $\mathbb{Z}_p$. Then $|A\dot{+}B| \geq \min\{p, |A| + |B| - 3\}$.*

Proof. First we show that if $|A| + |B| \geq p + 2$, then $A\dot{+}B = \mathbb{Z}_p$. If $p = 2$ the result is trivial. To see it for $p > 2$, let $x \in \mathbb{Z}_p$. Then $A \cap (x - B)$ contains at least two elements, so at least one of them, a does not satisfy the equation $2a = x$. But then $a \in A$ and $b = x - a \in B$ are distinct elements that add up to x.

We would now like to apply the combinatorial Nullstellensatz, so we need to show that if $A\dot{+}B$ is too small, then some polynomial of low degree vanishes everywhere on a product set. The natural product set to try to take is $A \times B$. What low-degree polynomial would vanish on $A \times B$? Well, we know that $A\dot{+}B$ is small, so a first approximation would be to take the polynomial $\prod_{c \in A\dot{+}B}(x+y-c)$. The trouble with that is that it does not necessarily vanish when $x = y \in A \cap B$. But we can take care of that by simply multiplying by the polynomial $x - y$.

So now we have a polynomial of degree $|A\dot{+}B|+1$ that vanishes on $A \times B$. For technical reasons it will be convenient to modify it slightly. Let us assume that the result is false and let C be a set that contains $A\dot{+}B$ and has cardinality exactly $|A| + |B| - 4$. Let P be the polynomial $(x - y)\prod_{c \in C}(x + y - c)$. This polynomial vanishes on $A \times B$ and has degree $|A| + |B| - 3$.

Let us look at the terms in $x^{|A|-1}y^{|B|-2}$ and $x^{|A|-2}y^{|B|-1}$, since if either of these has a non-zero coefficient then we will have contradicted the combinatorial Nullstellensatz. Let us write $r = |A|$, $s = |B|$, $t = r + s - 3$. Then the coefficient of $x^{r-1}y^{s-2}$ is $\binom{t-1}{r-2} - \binom{t-1}{r-1}$ and the coefficient of $x^{r-2}y^{s-1}$ is $\binom{t-1}{r-3} - \binom{t-1}{r-2}$.

Now it is not possible for three consecutive binomial coefficients to be congruent mod p. This can be checked by hand if $p = 2$. Otherwise, $\binom{a}{b-1}/\binom{a}{b} = \frac{b}{a-b+1}$ and $\binom{a}{b}/\binom{a}{b+1} = \frac{b+1}{a-b}$, so we would require $b - 1 \equiv a - b \equiv b + 1$. It follows that the two coefficients above are not both zero, so the result is proved. □

Note that if $A = B = \{0, 1, 2, \ldots, r - 1\}$, then $A\dot{+}B = \{1, 2, \ldots, 2r - 3\}$, so the result is sharp when the sets have equal size. It is not sharp for general sizes, since for example if B is a singleton, then $|A| + |B| \geq |A| - 1 = |A| + |B| - 2$.

The above two results can be proved by other means, but there are results that have been proved using the combinatorial Nullstellensatz for which no other proof is known – just as no proof of Dvir's theorem is known that does not go via polynomials.

It is interesting to note that there is an important difference between Dvir's proof and proofs that use the combinatorial Nullstellensatz. The former uses the fact that a non-zero polynomial of low degree cannot have too many roots. That is, the zero set of a low-degree non-zero polynomial cannot be too large. The combinatorial Nullstellensatz places a restriction (under a slightly different hypothesis) on not just the size but also the *structure* of the zero set: it cannot be a Cartesian product of sets that are too large.

8.3 The solution to the cap-set problem

The following result is a well-known theorem in additive combinatorics.

Theorem 8.9. *For every $\delta > 0$ there exists n such that every subset $A \subset \mathbb{F}_3^n$ of density at least δ contains three distinct vectors x, y, z with $x + y + z = 0$.*

Note that the condition $x + y + z = 0$ is the same as the condition $x + z = 2y$, so we can think of such triples as arithmetic progressions in $\mathbb{F}_3^n$. We can also think of them as one-dimensional affine subspaces.

The theorem is due to Roy Meshulam, though the proof is a straightforward adaptation of a proof due to Roth of the following closely related result.

Theorem 8.10. *For every $\delta > 0$ there exists n such that every subset $A \subset \{1, 2, \dots, n\}$ of density at least δ contains an arithmetic progression of length 3.*

It is natural to ask for more quantitative versions of these results. Meshulam's proof tells us that there is an absolute constant C such that the conclusion of the theorem holds if $n \geq C\delta^{-1}$. Turning that round, this says that if $A \subset \mathbb{F}_3^n$ is a set of density at least Cn^{-1}, then the conclusion holds. However, the densest known sets that did not contain triples x, y, z with $x + y + z = 0$ were far sparser than this – they had exponentially small density (as a function of n).

Bridging this huge gap became a notorious open problem, known as the cap-set problem, until a remarkable development in 2016. First Ernie Croot, Seva Lev, and Péter Pál Pach proved that a subset of $\mathbb{Z}_n^4$ with no progression of length 3 has to be of exponentially small density, and then within a couple of weeks Jordan Ellenberg and Dion Gijswijt independently showed that the Croot-Lev-Pach technique could be made to solve the cap-set problem itself. Not long after that, Terence Tao came up with a related argument that, unlike the arguments of Ellenberg and Gijswijt, treated the three variables x, y, z symmetrically. We give Tao's argument here.

The argument begins with a notion of rank for 3-tensors (that is, 3-dimensional generalizations of matrices). First, note that if X, Y are finite sets, $\mathbb{F}$ is a field, and $f : X \times Y \to \mathbb{F}$ is a function of two variables, then the rank of f, if we regard it as a matrix, is the minimum

k such that we can find functions $u_1, \dots, u_k : X \to \mathbb{F}$ and $v_1, \dots, v_k : Y \to \mathbb{F}$ such that

$$f(x,y) = \sum_{i=1}^{k} u_i(x)v_i(y)$$

for every $(x,y) \in X \times Y$.

If we now look at functions of three variables, there are a number of natural generalizations of rank. Let $f : X \times Y \times Z \to \mathbb{F}$ be a function. Then the most obvious way that we might try to decompose f into simpler parts is by writing it in the form

$$f(x,y,z) = \sum_{i=1}^{k} u_i(x)v_i(y)w_i(z).$$

Another, slightly less obvious but very natural when one stops to think about it, is to write it in the form

$$f(x,y,z) = \sum_{i=1}^{k} u_i(x,y)v_i(y,z)w_i(x,z).$$

Both the above definitions have their uses, but Tao introduced an "intermediate" definition, where instead of using products of one-variable functions or products of two-variable functions, one takes products of a one-variable function and a two-variable function.

Definition 8.11. Let X, Y and Z be finite sets, let $\mathbb{F}$ be a field, and let $f : X \times Y \times Z \to \mathbb{F}$. The *slice rank* of f is the smallest r such that there exist positive integers $0 \leq r_1 \leq r_2 \leq r$ and functions $u_1, \dots, u_r$ and $v_1, \dots, v_r$ such that

$$f(x,y,z) = \sum_{i=1}^{r_1} u_i(x)v_i(y,z) + \sum_{i=r_1+1}^{r_2} u_i(y)v_i(x,z) + \sum_{i=r_2+1}^{r} u_i(z)v_i(x,y),$$

for every $(x,y,z) \in X \times Y \times Z$.

In the above definition, the functions u_i and v_i do not all have the same domains – for instance, if $r_1 < i \leq r_2$ then $u_i : Y \to \mathbb{F}$ and $v_i : X \times Z \to \mathbb{F}$, as is suggested by the use of the variables. Note also that we use the convention that if $r_1 = 0$ then there are no terms in the first sum, if $r_1 = r_2$ then there are no terms in the second sum, and if $r_2 = r$ then there are no terms in the third sum.

To put the above definition in a more wordy way, we could define a function from $X \times Y \times Z$ to be "basic" if it is a product of a one-variable function and a two-variable function. Then the slice rank of f is the smallest r such that f is a sum of r basic functions.

Now let us prove a lemma that generalizes to slice ranks the fact that the rank of a diagonal matrix is the number of non-zero entries of that matrix.

Lemma 8.12. *Let X be a finite set, let $A \subset X$, let $\mathbb{F}$ be a field, and let $f : X^3 \to \mathbb{F}$ be a function such that $f(x,y,z) \neq 0$ if and only if $x = y = z$ and $x \in A$. Then the slice rank of f is $|A|$.*

Proof. Let $\delta_a : X \to \mathbb{F}$ be defined by $\delta_a(x) = 1$ if $x = a$ and 0 otherwise. Then by hypothesis we have that

$$f(x,y,z) = \sum_{a \in A} f(a,a,a)\delta_a(x)\delta_a(y)\delta_a(z)$$

for every $(x,y,z) \in X \times Y \times Z$, which implies that f has slice rank at most $|A|$ (since a product of two one-variable functions is a two-variable function).

Now let us suppose that we have a decomposition

$$f(x,y,z) = \sum_{i=1}^{r_1} u_i(x)v_i(y,z) + \sum_{i=r_1+1}^{r_2} u_i(y)v_i(x,z) + \sum_{i=r_2+1}^{r} u_i(z)v_i(x,y).$$

Without loss of generality $r_1 > 0$, so that the first term above is not empty.

We now claim that there is a function $h : X \to \mathbb{F}$ such that $\sum_{x \in X} h(x)u_i(x) = 0$ for $i = 1, 2, \dots, r_1$ and such that h is non-zero outside a set of size r_1. To see this, form an $r_1 \times |X|$ matrix M with the functions u_i as its rows (that is, $M(i,x) = u_i(x)$). Then we are trying to find a solution to the equation $Mh = 0$ such that h does not have many zeros. We can put M into reduced row-echelon form, and then it has $s \leq r_1$ non-zero rows, and a set of s columns such that those rows and columns form a copy of the $s \times s$ identity matrix. Let $S \subset X$ be the subset corresponding to the set of columns. Then we may choose arbitrary values $h(x)$ for every $x \notin S$, and then the values of h inside S are uniquely determined by the equations. The claim follows.

Now consider the function $g : Y \times Z \to \mathbb{F}$ given by the formula

$$g(y,z) = \sum_x h(x)f(x,y,z).$$

Then $g(y,z) = 0$ if $y \neq z$ or if $y = z \notin A$. If $y = z \in A$, then it takes the value $h(a)f(a,a,a)$. Since h is non-zero outside a set of size at most r_1, $h(a)f(a,a,a)$ is non-zero on a set of size at least $|A| - r_1$. From this it follows that g has rank at least $|A| - r_1$.

However, we also know that $g(y,z)$ is given by the formula

$$\sum_{i=r_1+1}^{r_2} u_i(y) \sum_x h(x)v_i(x,z) + \sum_{i=r_2+1}^{r} u_i(z) \sum_{x \in X} h(x)v_i(x,y),$$

which is a sum of $r - r_1$ products of two one-variable functions. This proves that g has rank at most $r - r_1$.

It follows that $|A| \leq r$, so the slice rank is at least $|A|$. □

The connection between slice rank and the cap-set problem comes with the following key observation. Suppose that A is a subset of $\mathbb{F}_3^n$ that does not contain distinct elements x, y, z with $x + y + z = 0$. Then if x, y, z belong to A and are not all the same, we must

have that $x+y+z \neq 0$ and hence that there exists i such that $x_i+y_i+z_i \neq 0$. And this last assertion can be written in a polynomial form as

$$1-(x_i+y_i+z_i)^2=0.$$

Now let $f: A^3 \to \mathbb{F}_3$ be defined by setting $f(x,y,z)=1$ if $x=y=z \in A$ and 0 otherwise. Then

$$f(x,y,z)=\prod_{i=1}^{n}(1-(x_i+y_i+z_i)^2)$$

for every $(x,y,z) \in |A|^3$. By the above lemma, the slice rank of f is $|A|$, so any upper bound we can obtain for the slice rank will translate directly into an upper bound for the size of A.

Lemma 8.13. *The slice rank of the polynomial $P(x,y,z)=\prod_{i=1}^{n}(1-(x_i+y_i+z_i)^2)$ is at most $3M$, where M is the number of 012-sequences of length n that sum to at most $2n/3$.*

Proof. The polynomial P is a polynomial of degree $2n$ in the $3n$ variables $x_1,\dots,x_n$, $y_1,\dots,y_n$, and $z_1,\dots,z_n$, and it is of degree at most 2 in each variable. It can be written as a linear combination of monic monomials, that is, polynomials of the form

$$x_1^{a_1}\dots x_n^{a_n}y_1^{b_1}\dots y_n^{b_n}z_1^{c_1}\dots z_n^{c_n}.$$

Let us now partition the monomials according to which of $a_1+\dots+a_n$, $b_1+\dots+b_n$ and $c_1+\dots+c_n$ is the smallest (breaking ties arbitrarily when they occur). This expresses P as a sum $P_1+P_2+P_3$, where P_1 is of degree at most $2n/3$ in the x_i, P_2 is of degree at most $2n/3$ in the y_i, and P_3 is of degree at most $2n/3$ in the z_i.

Now P_1 is a sum of the form $\sum_j Q_j(x)R_j(y)S_j(z)$, where Q_j, R_j and S_j are monomials and Q_j is monic. Let us now collect these terms according to what Q_j is. That enables us to write P_1 as a sum $\sum_h T_h(x)U_h(y,z)$, where the T_h are *distinct* monic monomials of degree at most $2n/3$. (Each T_h is equal to a Q_j and each $U_h(y,z)$ is the sum of all the $R_j(y)S_j(z)$ such that $Q_j=T_h$.) Note that the number of possibilities for T_h is M, so there are at most M terms in the sum.

We can do the same with P_2 and P_3, and this proves the result. □

It remains only to obtain an upper bound for M. For this we can use Lemma 2.6. Let $X_1,\dots,X_n$ be independent random variables, each of which is uniformly distributed in the set $\{-1,0,1\}$, and let $X=X_1+\dots+X_n$. Then it is not hard to see (by considering $Y_i=1-X_i$) that $M=3^n\mathbb{P}[X \geq n/3]$, which by Lemma 2.6 is at most $3^n e^{-n/36}$, and in particular is exponentially smaller than 3^n. It follows that $|A|$ is also exponentially smaller than 3^n, and we have proved Theorem 8.9 with an exponential bound for the density required.

One can of course be more careful at the end here and obtain a precise estimate for the number of 012-sequences that add up to at most $2n/3$.

9 Huang's solution to the sensitivity conjecture

Just over a year ago (at the time of writing in late 2020), a young mathematician called Hao Huang astonished theoretical computer scientists by not merely solving a famous open problem, known as the sensitivity conjecture, but by doing so with an extraordinarily short proof. As Scott Aaronson put it on his blog, it is a good illustration of the difference between the complexity classes P and NP, because the proof is extremely short and easy to understand, but it took a lot of very good mathematicians a few decades to find it.

I shall not say what the sensitivity conjecture is here, because it had been shown to follow from a very simply stated problem about the discrete cube, so it is the latter problem and its solution that I shall discuss. Write Q^n for the *Hamming cube*, that is, the graph with vertex set $\{0,1\}^n$ where we join x to y if they differ in exactly one coordinate. The theorem that Huang proved is the following.

Theorem 9.1. *Let G be an induced subgraph of Q^n with $2^{n-1}+1$ vertices. Then the maximum degree of G is at least $\sqrt{n}$.*

Note that the Q^n is bipartite, since if two vertices are joined, then the numbers of 1s in the corresponding sequences have different parity. So the theorem says that if we go from 2^{n-1} vertices to $2^{n-1}+1$ vertices, then the smallest possible maximum degree jumps from zero to $\sqrt{n}$.

To begin the proof, we define a sequence of orthogonal matrices, the inductive construction of which is reminiscent of that of the Walsh matrices we saw earlier in the course. We start with $A_0=(0)$, and then we define

$$A_n=\begin{pmatrix} A_{n-1} & I \\ I & -A_{n-1}\end{pmatrix}.$$

Here, I is the $2^{n-1}\times 2^{n-1}$ identity matrix.

It is easy to see (by induction) that A_n is symmetric. We also have that its rows are orthogonal. To see this, note that it is trivial by induction if the two rows are either both in the top half or both in the bottom half. Otherwise, if we take the ith row from the top half and the jth row from the bottom half, we obtain the inner product $(A_{n-1})_{ij}-(A_{n-1})_{ji}$, which is zero because A_{n-1} is symmetric.

Another feature of the matrix A_n is that its entries are all $-1, 0$ or 1 and that there are $n+1$ non-zero entries in each row (and hence also in each column). In fact, we can say more. If we index the rows and columns not with elements of the set $\{1,2,\dots,2^n\}$ but instead with 01-sequences of length n, where the first half of the indices for A_n are obtained from those of A_{n-1} by adding a 0 on the end, and the second half are obtained by adding a 1 on the end, then $(A_n)_{xy}$ is non-zero only if x and y differ in exactly one coordinate. Indeed, if x and y share their last coordinate, then this is true by induction, and if they do not share it, then $(A_n)_{xy}$ is an entry of one of the two identity-matrix parts of A_n, which implies that x and y are equal in all the other coordinates.

Thus, A_n is obtained from the adjacency matrix of Q^n by attaching signs to the edges. We can describe these signs inductively by saying that when $x_n=0$ we use the signs from

A_{n-1}, when $x_n = 1$, we multiply all those signs by -1, and for the edges that go across from $x_n = 0$ to $x_n = 1$ we have 1s everywhere. For a non-inductive description (which we will not need), if x and y agree except at the ith coordinate, then A_{xy} is $(-1)^{x_{i+1}+\cdots+x_n}$.

An immediate consequence of these observations is that $A_n^2 = nI$. (The orthogonality implies that $A_n A_n^T = nI$ and the symmetry implies that $A_n = A_n^T$.) From this it follows that every eigenvalue of A_n is either $\sqrt{n}$ or $-\sqrt{n}$. Since $\operatorname{tr} A_n = 0$, the two eigenvalues each occur with multiplicity 2^{n-1}.

We are now ready to prove the theorem.

Proof of Theorem 9.1. Let V be a subset of $\{0,1\}^n$ of size $2^{n-1}+1$. Let us define a linear map $\alpha : \mathbb{R}^V \to \mathbb{R}^V$ by taking a function $f : V \to \mathbb{R}$ to the function $g : V \to \mathbb{R}$ defined by

$$g(v) = \sum_{w \in V} (A_n)_{vw} f(w).$$

In other words, we multiply f (considered as a column vector) by the matrix A_n and restrict it to V.

If Δ is the maximum degree of the subgraph of Q^n induced by V, then

$$\|\alpha f\|_1 \leq \sum_{v \in V} \sum_{w \in V} |(A_n)_{vw}||f(w)| = \sum_{w \in V} |f(w)| \sum_{v \in V} |(A_n)_{vw}| \leq \Delta \|f\|_1,$$

since $|(A_n)_{vw}| = 1$ if vw is an edge and 0 otherwise.

Now the set of all functions from $\{0,1\}^n$ to $\mathbb{R}$ that are supported in V is a space of dimension $2^{n-1}+1$, and therefore, since the two eigenspaces of A_n have dimension 2^{n-1}, there is an eigenvector f with eigenvalue $\sqrt{n}$ supported in V. For this f we have that $\alpha f = \sqrt{n} f$, and therefore that $\|\alpha f\|_1 = \sqrt{n}\|f\|_1$. It follows that $\Delta \geq \sqrt{n}$. □

How should one react to an argument like this, which at first sight appears to work by magic? I recommend thinking about it until it starts to seem more natural and less magical. Note, for instance, that it is another example of a very useful general technique, which is to bound a cardinality in terms of something else. An earlier example has been our use of entropy to provide bounds for cardinalities. Here one can imagine a thought process that begins with the well known observation that the largest eigenvalue of the adjacency matrix of a graph is at most the maximum degree of the graph, then spots that the same is true for any matrix with entries that are bounded in magnitude by those of the adjacency matrix. So it is sufficient to find a matrix with that property and an eigenvector with eigenvalue $\sqrt{n}$. But we don't know all that much about the graph, since its vertex set is an arbitrary subset of $\{0,1\}^n$ of size $2^{n-1}+1$. That does tell us that we'd be done if we could find a matrix with an eigenvector $\sqrt{n}$ of multiplicity at least 2^{n-1}. And although it's not trivial that such a matrix exists, once one starts to think about the properties it ought to have, one arrives reasonably quickly at the wish that its non-zero entries should all be ± 1 and that the rows ought to be orthogonal, and actually finding it becomes a not too difficult exercise at that point.

Actually, I know for a fact that that is not quite Huang's thought process, since he used a result called Cauchy's interlace theorem, though the basic ideas were the same. The even simpler proof presented above was spotted by Shalev Ben-David and shared in a comment on Scott Aaronson's blog post.

10 Dimension arguments

The topic of this last section wraps up the course quite nicely, as it will require us to touch on a number of earlier topics that we have covered, including averaging arguments, set systems with restrictions on intersections, combinatorial geometry, the Borsuk-Ulam theorem, and use of polynomials.

The basic idea behind a dimension argument is that one can sometimes obtain an upper bound for the size of a combinatorial object by constructing an injection from that object to a low-dimensional vector space such that the image is a linearly independent set. Then the dimension of the vector space gives the desired lower bound. This technique is particularly useful for extremal problems where the maximum is far from unique, since it transforms those problems into other problems (of the form, "How large an independent set can we find in this vector space?") that have the same property. We shall use dimension arguments to prove a number of results about set systems, and then we shall discuss a famous solution by Jeff Kahn and Gil Kalai to a problem in combinatorial geometry.

10.1 Subsets of $\mathbb{R}^n$ that give rise to at most two distances

Let X be a subset of $\mathbb{R}^n$ and suppose that there is only one distinct distance between any two distinct points. How large can X be?

This is easy to answer. The condition tells us that X is a regular simplex, and the largest regular simplex that fits into $\mathbb{R}^n$ is the n-dimensional simplex, which has $n+1$ vertices. (To make this argument rigorous, let $X = \{x_1, \dots, x_m\}$, let $x = (x_1 + \cdots + x_m)/m$, and let $y_i = (x_i - x)/\|x_i - x\|$ for every i. Then one can check that $\langle y_i, y_j \rangle = -1/(m-1)$ for every $i \neq j$, and then results about well-separated vectors from earlier in the course yield that $m \leq n+1$.)

What if we allow at most *two* distances? An example that gives $\binom{n}{2}$ points is the set of vectors $e_i + e_j$, where $e_1, \dots, e_n$ are the standard basis vectors of $\mathbb{R}^n$. It is not known whether this is best possible, but a comparable upper bound can be obtained with a very nice linear algebra argument.

Theorem 10.1. *Let $a_1, \dots, a_m$ be points in $\mathbb{R}^n$ such that the number of distinct distances $d(a_i, a_j)$ with $i \neq j$ is at most 2. Then $m \leq (n+1)(n+4)/2$.*

Proof. Let the two distances be a and b, and define a polynomial in $2n$ variables by

$$f(x, y) = (d(x, y)^2 - a^2)(d(x, y)^2 - b^2).$$

Note that $d(x, y)^2 = \sum_{i=1}^n (x_i - y_i)^2$, so this is a polynomial of degree 4.

For each i, let f_i be the n-variable polynomial given by $f_i(x) = f(x, a_i)$. Then these polynomials are linearly independent, since $f_i(a_i) = a^2b^2$, and $f_i(a_j) = 0$ for every $j \neq i$.

To squeeze as much information as we can out of this observation, we would now like to prove that the polynomials f_i all lie in a low-dimensional subspace of the space of n-variable polynomials. Since they are all quartics, we can obtain an upper bound of order n^4 with ease, but with slightly more thought we can do substantially better. Note that

$$d(x, a_i)^2 = \sum_j x_j^2 - 2\sum_j x_j a_{ij} + \sum_j a_{ij}^2$$

, where we write a_{ij} for the jth coordinate of a_i. It follows that $f_i(x)$ is a linear combination of $(\sum_j x_j^2)^2$, the n polynomials $x_h \sum_j x_j^2$, the $n(n+1)/2$ polynomials $x_h x_j$ with $h \leq j$, the n polynomials x_j, and 1. So the f_i live in a space of dimension $1+n+n(n+1)/2+n+1 = (n+1)(n+4)/2$. □

The above bound is not the best known: using a strengthening of the same method, Blokhuis obtained an upper bound of $(n+1)(n+2)/2$. But it is still an open problem to determine the exact bound.

10.2 Set systems with intersections of restricted parity

Let $\mathcal{A}$ be a family of subsets of $[n]$ such that every set in $\mathcal{A}$ has even cardinality and any two distinct sets in A have an intersection of even cardinality. How large can $\mathcal{A}$ be?

A simple example shows that $\mathcal{A}$ can be of size $2^{\lfloor n/2 \rfloor}$: we just take all sets that are unions of some of the sets $\{1,2\}, \{3,4\}, \ldots$. It also seems to be hard to do better than this.

How about if we ask for the intersections to have odd size? Now it seems to be much more challenging to find a large example. If n is odd, the system $\{1,2\}, \{1,3\}, \ldots, \{1,n\}$, $\{2,3,\ldots,n\}$ has size n and satisfies the conditions, but it cannot be extended, since any set of even size that contains 1 will intersect at least one $\{1,i\}$ in a set of size 2, and any set of even size that does not contain 1 intersects $\{2,3,\ldots,n\}$ in a set of even size.

Rather surprisingly (if you haven't seen it before), these radically different bounds are both sharp, and one can prove it quite simply using linear algebra. Let us start with the second.

Theorem 10.2. *Let $\mathcal{A}$ be a family of subsets of $[n]$ of even size such that any two distinct sets in the family have an intersection of odd size. Then $|\mathcal{A}| \leq n$ if n is odd, and $|\mathcal{A}| \leq n-1$ if n is even.*

Proof. We begin by proving the same statement but with "even" and "odd" reversed: that is, we assume that the sets have odd size and that their intersections have even size.

Associate with each set $A \in \mathcal{A}$ its characteristic vector χ_A and regard that vector as an element of $\mathbb{F}_2^n$. We can define an $\mathbb{F}_2$-valued inner product in the obvious way, namely $\langle f, g \rangle = \sum_x f(x)g(x)$, where the sum is in $\mathbb{F}_2$. Note that with respect to this inner product,

we have that $\langle \chi_A, \chi_A \rangle = 1$ and $\langle \chi_A, \chi_B \rangle = 0$ for every $A, B \in \mathcal{A}$. In other words, the characteristic vectors form an "orthonormal basis" over $\mathbb{F}_2$.

The usual proof that orthogonal vectors are independent works here too: if $\sum_A \lambda_A \chi_A = 0$, then for every B, $0 = \langle \sum_A \lambda_A \chi_A, \chi_B \rangle = \lambda_B$. Therefore, $|\mathcal{A}| \leq n$.

Note that the sets $\{1\}, \{2\}, \ldots, \{n\}$ show that this bound is best possible.

Now let us assume that the sets have even size and odd intersections.

If n is odd, we can replace each A by $[n] \setminus A$, which yields a family of sets of odd size with even intersections. (This last assertion follows easily from the fact that if A and B are sets with cardinalities of the same parity, then $|A \setminus B|$ and $|B \setminus A|$ have the same parity.) Thus $|\mathcal{A}| \leq n$.

As noted above, the sets $\{1, 2\}, \ldots, \{1, n\}, \{2, 3, \ldots, n\}$ show that this bound is best possible.

If n is even, then replace each $A \in \mathcal{A}$ that contains n by $[n] \setminus A$ and leave the remaining sets as they are. If A and B are any two sets and $|A|$ is even, then $|A \cap B|$ and $|A \cap B^c|$ have the same parity, and since n is even, B and B^c have the same parity as well. Also, $\mathcal{A}$ cannot contain a complementary pair since all intersections have odd size. Therefore, we obtain a new set system with the same property and the same number of sets, but now with all sets contained in $[n-1]$. Since $n-1$ is odd, we find that $|\mathcal{A}| \leq n-1$.

The sets $\{1, 2\}, \ldots, \{1, n\}$ show that this bound is best possible. □

Now let us consider what happens if the sets and their intersections have even sizes.

Theorem 10.3. *Let $\mathcal{A}$ be a collection of subsets of $[n]$ and suppose that $|A|$ and $|A \cap B|$ are even for every $A, B \in \mathcal{A}$. Then $|\mathcal{A}| \leq 2^{\lfloor n/2 \rfloor}$.*

Proof. Bounding $\mathcal{A}$ is equivalent to bounding the size of a subset $\mathcal{F} \subset \mathbb{F}_2^n$ with the property that $\langle f, g \rangle = 0$ for every $f, g \in \mathcal{F}$.

Let k be maximal such that $\mathcal{F}$ contains a linearly independent set of size k. Then $\mathcal{F}$ lies in the orthogonal complement of that set, which has size 2^{n-k}, so $|\mathcal{F}| \leq 2^{n-k}$. But $\mathcal{F}$ also lies in the linear span of that set, so $|\mathcal{F}| \leq 2^k$. It follows that $|\mathcal{F}| \leq \max_k \min\{2^k, 2^{n-k}\} = 2^{\lfloor n/2 \rfloor}$. □

10.3 More general restricted intersections

Earlier in the course, we looked at set systems where the sizes of intersections were required to be large. Here we shall look at set systems where we ask for intersections not to be of certain sizes, and we shall see that even ruling out one quite large size can force a set system to have cardinality exponentially smaller than 2^n. The next result is just a sample of what can be done.

Theorem 10.4. *Let $\mathcal{A}$ be a family of subsets of $[n]$ such that the size of every $A \in \mathcal{A}$ is a multiple of p, but the size of no intersection of two distinct sets in $\mathcal{A}$ is a multiple of p. Then*

$$|\mathcal{A}| \leq \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{p-1}.$$

Proof. Given a set $A \in \mathcal{A}$, define a polynomial P_A over $\mathbb{F}_p$ in n variables by

$$P_A(x) = 1 - (\sum_{i \in A} x_i)^{p-1}.$$

Note that $P_A(x) = 1$ if $\sum_{i \in A} x_i = 0$ and $P_A(x) = 0$ otherwise (by Fermat's little theorem). Note also that P_A has degree $p - 1$.

If x is the characteristic function of a set $B \in A$, then $P_A(x) = 1$ if $A = B$ and 0 otherwise, since in the second case our hypothesis implies that $P_A(x)$ is not zero mod p. This proves that the polynomials P_A are linearly independent.

However, this is not enough to give us the bound stated. For that, we need a further trick, which is to observe that if P is any polynomial in n variables and we replace each occurrence of x_i^r by x_i (so for example the monomial $x_1^3 x_4^6 x_5^{11}$ would become the monomial $x_1 x_4 x_5$), then we obtain a new polynomial that takes the same value as P on any x that has all its coordinates equal to 0 or 1, since for such an x we have that $x_i^r = x_i$ for every $r \geq 1$.

So now let us replace each P_A by a polynomial Q_A that agrees with P_A on all characteristic functions of sets and has degree at most 1 in each variable x_i. Then $Q_A(\chi_B) = P_A(\chi_B) = \delta_{AB}$ for every $A, B \in \mathcal{A}$, so the Q_A are also independent.

The number of monomials of degree at most $p - 1$ and of degree at most 1 in each variable x_i is

$$\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{p-1},$$

so the Q_A live in a space of this dimension, which proves the result. □

Note that in the case $p = 2$ this gives a bound of $n + 1$, which is pretty close to the exact bound we obtained in the previous subsection.

An interesting special case of the above result is where $n = 4p$ and the sets all have size $2p$. Then the result tells us that we have a collection $\mathcal{A}$ of subsets of $[4p]$ of size $2p$, no two of which are disjoint and no two of which have an intersection of size exactly p. With this constraint, we deduce that $|\mathcal{A}| \leq \sum_{m<p} \binom{4p}{m}$, which is exponentially smaller than 2^{4p}. From this, we can obtain the following geometrical consequence.

Corollary 10.5. *Let p be an odd prime and let $n = 4p$. Then the largest measure of a set X of unit vectors in $\mathbb{R}^{4p}$ that contains no pair of orthogonal vectors is exponentially small.*

Proof. Let Y be the collection of unit vectors all of whose coordinates are either $1/2\sqrt{p}$ or $-1/2\sqrt{p}$ and that sum to zero. If $A \subset [4p]$ is a set of size $2p$, let v_A be the vector in Y that takes the value $1/2\sqrt{p}$ on A and $-1/2\sqrt{p}$ on $[4p] \setminus A$. Observe that $\langle v_A, v_B \rangle = 0$ if and only if $|A \cap B| = p$. It therefore follows from Theorem 10.4 that $|X \cap Y| \leq \sum_{m=0}^{p-1} \binom{n}{m}$, which is exponentially smaller than 2^n since $p = n/4$.

Since the property of not containing a pair of orthogonal vectors is invariant under rotation, it also follows that if ρ is any rotation of $\mathbb{R}^n$, the intersection $\rho X \cap Y$ also has size at most $\sum_{m=0}^{p-1} \binom{n}{m}$.

Now we do an averaging argument very similar to the averaging step in Katona's proof of the Erdős-Ko-Rado theorem. Let ρ be a random rotation of $\mathbb{R}^n$. Then for each $y \in Y$, the probability that $y \in \rho X$ is μX (where μ is the rotation-invariant probability measure on the sphere), so the expected size of $\rho X \cap Y$ is $\mu|Y|$. But it is also at most $\sum_{m=0}^{p-1} \binom{n}{m}$, from which we deduce that $\mu \leq \binom{n}{2p}^{-1} \sum_{m=0}^{p-1} \binom{n}{m}$. The estimates in Section 2 imply easily that this is exponentially small as a function of n. □

To justify the last sentence of the above proof slightly more, $\binom{n}{2p} = \binom{n}{n/2}$, which is $n^{-1/2}2^n$ up to a constant factor, and we also know that if $\alpha < 1/2$, then the sum $\sum_{m \leq \alpha n} \binom{n}{m}$ is exponentially smaller than 2^n, by Lemma 2.6.

10.4 Kahn and Kalai's disproof of Borsuk's conjecture

The following conjecture was made by Borsuk in 1933.

Conjecture 10.6. *Let X be a bounded subset of $\mathbb{R}^n$. Then there exist sets $Y_1, \dots, Y_{n+1}$ of diameter less than that of X such that $X \subset \bigcup_i Y_i$.*

To see why this is a reasonable conjecture, consider the case where X is the unit ball of ℓ_2^n – that is, the set of all vectors of norm at most 1. As we saw in the section on using the Borsuk-Ulam theorem, if we cover the boundary of X with n closed sets, then one of those sets will contain an antipodal pair, and therefore will have diameter at least that of X. Therefore, the bound $n+1$, if correct, is best possible. (On the third examples sheet I ask you to show that Borsuk's conjecture is true for this X.)

It is not hard to see that we may assume that X and $Y_1, \dots, Y_n$ are closed and convex, since taking the closure of the convex hull of a set does not change its diameter. The conjecture came to seem even more reasonable when it was proved to be true when X is smooth. (I won't say here precisely what this means, but roughly speaking it means that the boundary of X has no sharp angles.)

It was therefore a huge surprise (yet another) when Jeff Kahn and Gil Kalai not only disproved it, but obtained an exponentially large (in $\sqrt{n}$) lower bound for the number of sets of smaller diameter needed to cover X, and even more of a surprise (yet another) that the proof was very short and easy to understand. I was a graduate student at the time, and was lucky enough to see Jeff Kahn presenting the proof in Cambridge just after it had been announced in 1993.

I'll give an argument that's not quite identical to theirs, but it's very similar. (It will give a very slightly worse bound, but avoid a tiny technical improvement to the argument in the previous section.) The next result is another corollary to Theorem 10.4.

Corollary 10.7. *Let $n = 4p$ for a prime p. Then $\mathbb{R}^{n^2}$ contains a set X of size $\binom{n}{2p}$ such that every subset of X of smaller diameter has size at most $\sum_{m=0}^{p-1} \binom{n}{m}$.*

Proof. Let Y be the set of unit vectors defined in the proof of Corollary 10.5 and, again as in that proof, for each $A \in [4p]^{(2p)}$ let $v_A \in Y$ be the vector with all its positive coordinates in A. As noted in that proof, v_A and v_B are orthogonal if and only if $|A \cap B| = p$.

Now let $X \subset \mathbb{R}^{n^2}$ be the set of all points $v \otimes v$ such that $v \in Y$. (We saw this notation in the section on the cap-set problem: $v \otimes v$ denotes the matrix with ijth entry $v_i v_j$, which we can think of as a point in $\mathbb{R}^{n^2}$.) For any $v, w \in \mathbb{R}^n$,

$$\langle v \otimes v, w \otimes w = \sum_{ij} v_i v_j w_i w_j = \langle v, w \rangle^2 \geq 0,$$

with equality if and only if v and w are orthogonal. This calculation also shows that if v and w are unit vectors, then so are $v \otimes v$ and $w \otimes w$, which implies that

$$\begin{aligned} \|v \otimes v - w \otimes w\|^2 &= \|v \otimes v\|^2 + \|w \otimes w\|^2 - 2\langle v \otimes v, w \otimes w \rangle \\ &= 2 - 2\langle v, w \rangle^2 \\ &\leq 2, \end{aligned}$$

with equality if and only if $\langle v, w \rangle = 0$. It follows that X has diameter $\sqrt{2}$, and that no subset of X of smaller diameter contains a pair of vectors $v_A \otimes v_A$ and $v_B \otimes v_B$ such that v_A and v_B are orthogonal, or equivalently such that $|A \cap B| = p$. The result now follows from Theorem 10.4. □

It follows immediately that X cannot be covered by fewer than $\binom{n}{2p} / \sum_{m=0}^{p-1} \binom{n}{m}$ subsets of smaller diameter. This is exponentially large in n, and so exponentially large in the square root of the dimension of the space that contains X.