Chains, antichains, shadows

Noting that a complete matching of a bipartite graph from X to Y is a vertex disjoint collection of edges such that we cover X.

Theorem 1 (Hall's theorem) *$G = (X \cup Y, E)$ a bipartite graph has a complete matching from X to Y if and only if for all $S \subseteq X$, $|\Gamma(S)| \geq |S|$.*

We then claim that there is a partition of $\mathcal{P}(n)$ into $\binom{n}{\lfloor n/2 \rfloor}$ chains, following as we can find matchings from layer to layer when viewed as Q_n, and there are $\binom{n}{\lfloor n/2 \rfloor}$ elements in the centre allowing us to choose that number of such chains through the layers. Thus we can then prove Sperner's lemma:

Theorem 2 (Sperner's lemma) *An antichain in $\mathcal{P}(n)$ has size at most $\binom{n}{\lfloor n/2 \rfloor}$.*

This is clear as a chain and an antichain meet on at most one element, so partitioning $\mathcal{P}(n)$ into some number of chains means having an antichain with more elements would mean it shares two elements with a chain.

Theorem 3 (LYM Inequality) *Let $\mathcal{F} \subseteq \mathcal{P}(n)$ be an antichain. Then*

$$\sum_{i=0}^{n} \frac{|\mathcal{F} \cap [n]^{(i)}|}{\binom{n}{i}} \leq 1,$$

and we have equality iff $\mathcal{F} = [n]^{(i)}$ for some $i \in [n]$.

Definition 1 *With $\mathcal{F} \subseteq X^{(k)}$ a k-uniform family on X, the lower shadow $\partial\mathcal{F}$ of $\mathcal{F}$ is $\bigcup_{A \in \mathcal{F}} A^{(k-1)}$.*

Essentially, to form a shadow on a k-uniform family $\mathcal{F}$, from each $A \in \mathcal{F}$ we pick $k-1$ elements.

This is proven using a local version with $\mathcal{F} \subseteq [n]^{(r)}$, that

$$\binom{n}{r}|\partial\mathcal{F}| \geq \binom{n}{r-1}|\mathcal{F}|,$$

shown via double counting of edges in Q_n, noting that each element in $\mathcal{F}$ contains k sets of size $k-1$, and each element in $\partial\mathcal{F}$ is contained by $\leq n-(k-1)$ elements in $\mathcal{F}$, so $(n-k+1)|\partial\mathcal{F}| \geq k|\mathcal{F}|$. Equality occurs $\mathcal{F} = [n]^{(k)}$ or $\mathcal{F} = \varnothing$. **Understand proof a bit better.**

As a result, if $\mathcal{F}$ is an antichain in $\mathcal{P}(n)$, $|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}$ with equality for $\mathcal{F} = [n]^{(k)}$, $k \in \{\lfloor n/2 \rfloor, \lceil n/2 \rceil\}$.

Theorem 4 (Dilworth's theorem) *Let $(P, \leq)$ be a finite poset. The minimum number of chains needed to cover P is the maximum size of an antichain.*

One direction of this is clear: the minimum number of chains needed to cover P is certainly not less than the maximum size of an antichain, because then we would have an antichain of maximum size which contains more than one element in one of the covering chains, giving a contradiction.

The other direction is less clear: we need to show that we can construct a covering of P with a chain for each element of a maximum-size antichain. The intuition here is that we should be able to take a maximum size antichain, and generate chains by trawling for elements above and below each element of the antichain. If this doesn't generate a covering, we have an element x which isn't at the top or bottom of any generated chain. x must be comparable with some element of the antichain y, and thus we just select the first element (closest to y) of the chain generated by y with which x is not comparable ...

The above explanation requires some work, but in principle captures how this should operate. We also have a dual to this theorem:

Theorem 5 (Mirsky's theorem) *Let $(P, \leq)$ be a finite poset. The minimum number of antichains needed to cover P is the maximum size of a chain.*

To prove this, we calculate the height of a maximal chain starting at each element $v \in P$, and note that we can get antichains by selecting the elements with a particular height, for each possible height.

Theorem 6 (Erdős) *With $\boldsymbol{x} \in \mathbb{R}^n$ s.t. $x_i \geq 1$ for all $i \in [n]$. For every $\alpha \in \mathbb{R}$, there are at most $\binom{n}{\lfloor n/2 \rfloor}$ subsets $I \subseteq [n]$ such that $\sum_{i \in I} x_i \in [\alpha, \alpha + 1)$.*

By using Sperner's lemma, we see that if more such subsets existed, we would have one containing the other, and thus violating the required property.

Indeed we can extend this theorem to the case $|x_i| \geq 1$, noting that replacing $+x_i$ with $-x_i$ just switches around whether we're making an 'include' or 'don't include' decision, translating the collection of possible sums.

We want to generalise this a bit more, and in order to do this we introduce the concept of chain symmetry – we say that a chain in $\mathcal{P}(n)$ is symmetric if each link adds one element, and if the smallest and largest sets in the chain have gross cardinality n. Indeed we can partition $\mathcal{P}(n)$ into symmetric chains by induction. Critically, as every symmetric chain must contain an element in $[n]^{(\lfloor n/2 \rfloor)}$, thus there are $\binom{n}{\lfloor n/2 \rfloor}$ chains in such a decomposition.

Theorem 7 *Let $k, n \geq 1$, suppose that $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n] \in \mathbb{R}^{k \times n}$ s.t. $||\boldsymbol{x}_i||_2 \geq 1$ for all i. With $K \subseteq \mathbb{R}^k$ s.t. $\operatorname{diam}(K) < 1$, there are at most $\binom{n}{\lfloor n/2 \rfloor}$ subsets $I \subseteq [n]$ s.t. $\sum_{i \in I} x_i \in K$.*

It is sufficient to show that we can partition $\mathcal{P}(n)$ into $\binom{n}{\lfloor n/2 \rfloor}$ sets $D_1, \ldots, D_{\binom{n}{\lfloor n/2 \rfloor}}$ which are 'sparse', meaning that for $A, B \in D_i$, $||\sum_{i \in A} \boldsymbol{x}_i - \sum_{j \in B} \boldsymbol{x}_j|| \geq 1$. If such a partition exists then we must only be able to select one valid set from each collection in the partition, and the result is proven.

To prove this, we find a symmetric partition $\mathcal{P}(n)$ into sparse families by induction. For a symmetric partition $\mathcal{F}_1, \ldots, \mathcal{F}_m$, order the elements of $\mathcal{F}_i$ as $A_1, \ldots, A_t$ where $\langle x_{A_1}, x_n \rangle \leq \ldots \langle x_{A_t}, x_n \rangle$, and we hope to then generate new chains for which the sets more aligned with x_n are at the top (and thus $x_n + x_{A_i}$ is kept far from the other sets). We do this and find straightforwardly that with the same procedure to generate a symmetric partition in the first place subject to this particular reordering, we get the desired result.

We would like to get a more precise understanding of the relative cardinalities of r-uniform sets and their shadows. We get an immediate inequality via LYM, but want to understand where we get closer to and further from the equality. To do this, we define lexicographic and colexicographic order:

In *lex*, for $A, B \in [n]^{(r)}$, $A <_{\text{lex}} B$ if $A \neq B$ and $\min(A \triangle B) \in A$. Thus in lex order, the place of a set in an order is primarily determined by its lowest element (and then its next lowest, and on and on).

In *colex*, for $A, B \in [n]^{(r)}$, $A <_{\text{colex}} B$ if $A \neq B$ and $\max(A \triangle B) \in B$. Thus in colex order, the place of a set in an order is primarily determined by its highest element (and then its next highest, and on and on).

Theorem 8 (Kruskal-Katona) *Let $\mathcal{F} \subseteq [n]^{(r)}$, $\mathcal{A}$ be the family of the first $|\mathcal{F}|$ elements of $[n]^{(r)}$ in colex order. Then $|\partial\mathcal{F}| \geq |\partial\mathcal{A}|$.*

Essentially, the claim is that colex order creates the most overlap in the shadow, thus minimising its size. To show this takes some time, but fundamentally we do this by reducing $\mathcal{F}$ to a family which is an initial segment of colex, and demonstrate that at each step $|\partial\mathcal{F}'| \leq |\partial\mathcal{F}|$. Intuitively, our goal is to swap large elements for small elements, to get ourselves towards colex.

We define a compression operator to swap in i instead of j and j instead of i wherever this is coherent, and to leave all other sets in $\mathcal{F}$ untouched otherwise. We refer to this as a left compression where $i < j$, and conjecture that left compressions should not increase the shadow.

$$C_{ij}(A) = \begin{cases} (A \setminus j) \cup i & \text{if } i \notin A, j \in A \\ A & \text{otherwise} \end{cases}$$
$$C_{ij}(\mathcal{F}) = \{C_{ij}(A) : A \in \mathcal{F}\} \cup \{A \in \mathcal{F} : C_{ij}(A) \in \mathcal{F}\}.$$

Indeed, the shadow of $C_{ij}(\mathcal{F})$ is no larger than the shadow of $\mathcal{F}$, and there is a family $\mathcal{A}$ which is a fixed point of all left-compressions (which we call left-compressed). Unfortunately, not every left-compressed family is in colex form, as $\{12, 13, 14\}$ is left-compressed but contains 14, which we can't get rid of e.g. via C_{34} because we get it back from 13. Indeed we want to be able to convert this into $\{12, 13, 23\}$, but this is impossible without doing two steps simultaneously.

Thus we want to replace i and j with disjoint sets $U, V \in \mathcal{P}(n)$, $|U| = |V|$, and technical details allow us to get the compression as desired.

Combinatorial Nullstellensatz

Theorem 9 (Combinatorial Nullstellensatz) *Let $\mathbb{F}$ be a field, $f \in \mathbb{F}[x_1, \ldots, x_n]$ a polynomial of degree t, with $t_1, \ldots, t_n$ such that $\sum_{i=1}^n t_i = t$ and $\prod_{i=1}^n x_i^{t_i}$ has non-zero coefficient in f.*

If there are sets $S_1, \ldots, S_n$ in $\mathbb{F}$ each with $|S_i| \geq t_i + 1$, then there is $\boldsymbol{s} \in \prod_{i=1}^n S_i$ such that $f(\boldsymbol{s}) \neq 0$.

The proof is by induction. For $t = 1$, f is linear in an argument, and the statement is immediate. Assuming that there are $S_1, \ldots, S_n$ in $\mathbb{F}$ with $|S_i| \geq t_i + 1$, and taking $s_1 \in S_1$,

$$f(x_1, \ldots, x_n) = (x_1 - s_1)g(x_1, \ldots, x_n) + h(x_2, \ldots, x_n)$$

where g is a polynomial of degree $t-1$ with monomial $x_1^{t_1-1} \prod_{i=2}^n x_i^{t_i}$, and $h \in \mathbb{F}[x_2, \ldots, x_n]$ with degree t. Immediately, for $(s_2, \ldots, s_n) \in \prod_{i=2}^n S_i$, $f(s_1, s_2, \ldots, s_n) = h(s_2, \ldots, s_n)$. Furthermore, by the inductive hypothesis there is some $(s_1', \ldots, s_n') \in (S_1 \setminus \{s_1\}) \times \prod_{i=2}^n S_i$ such that $g(s_1', \ldots, s_n') \neq 0$, so

$$\begin{aligned} f(s_1', \ldots, s_n') &= (s_1' - s_1)g(s_1', \ldots, s_n') + h(s_2', \ldots, s_n') \\ &= (s_1' - s_1)g(s_1', \ldots, s_n') + f(s_1, s_2', \ldots, s_n') \end{aligned}$$

and thus $f(s_1', \ldots, s_n') \neq f(s_1, s_2', \ldots, s_n')$, so at least one of these is non-zero.

This is naturally an incredibly wide-reaching theorem. The key idea is that given a problem, we want to construct a polynomial with informative zeros, and then take a subset of its domain big enough to guarantee a non-zero value, which indicates that we've exited our initial case.

Theorem 10 *With $H_1, \ldots, H_m$ a family of hyperplanes in $\mathbb{R}^n$, and*

$$\left| \{0,1\}^n \cap \bigcup_{i=1}^m H_i \right| = 2^n - 1.$$

Then $m \geq n$.

To prove this, construct a polynomial $P(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{R}^n$, which is zero on $\{0,1\}^n$ if $\boldsymbol{x}$ is contained in a hyperplane and non-zero, or equal to zero. This can be easily constructed with degree $n \vee m$, and if $m < n$, the combinatorial nullstellensatz gives us that there is $\boldsymbol{s} \in \{0,1\}^n$ such that $P(\boldsymbol{s}) \neq 0$, so thus there must be a point not contained in a hyperplane.

Theorem 11 (Cauchy-Davenport) *If p is prime and $A, B \subseteq \mathbb{Z}_p$, then*

$$|A + B| \geq \min\{p, |A| + |B| - 1\}.$$

The claim is almost immediate for $|A| + |B| > p$, as we just need to show that $A + B = \mathbb{Z}_p$. In the case $|A| + |B| \leq p$, we want to claim a contradiction if $|A + B| \leq |A| + |B| - 2$, by constructing a polynomial of degree $|A| + |B| - 2$, zero on $A \times B$. Yet combinatorial nullstellensatz should give an element on which the polynomial is nonzero, and this gives us a contradiction.

Intersections and traces

We say that $\mathcal{A} \subseteq \mathcal{P}(n)$ is intersecting if $A \cap B \neq 0$ for $A, B \in \mathcal{A}$. It's immediately clear that $|\mathcal{A}| \leq 2^{n-1}$.

Theorem 12 (Erdös-Ko-Rado) *For $r \leq n/2$, $\mathcal{A} \subseteq [n]^{(r)}$ intersecting implies that $|\mathcal{A}| \leq \binom{n-1}{r-1}$.*

There are two separate proofs for this: the first involves Katona's circle method, and the second using the Kruskal-Katona theorem.

Theorem 13 (Liggett's theorem) *With $\boldsymbol{Y} \overset{\text{i.i.d.}}{\sim} \operatorname{Ber}(p)^n$, $\boldsymbol{\alpha} \geq 0$, $\sum \alpha_i = 1$,*

$$\mathbb{P}\left(\sum_i \alpha_i Y_i \geq 1/2\right) \geq p.$$

The proof of this theorem follows by applying EKR to $\{A \subseteq [n]^{(k)} : \sum_{i \in A} \alpha_i > 1/2\}$.

Theorem 14 (Two Families Theorem) *Let $A_1, \ldots, A_k, B_1, \ldots, B_k$ be finite sets such that each (A_i, B_i) pair is disjoint, but no other (A_i, B_j) pair is disjoint. Then*

$$\sum_{i=1}^{k} \binom{|A_i| + |B_i|}{|A_i|}^{-1} \leq 1.$$

Observe immediately that each summand is equal to the probability that we choose the (A_i, B_i) partition of $A_i \cup B_i$.

VC-dimension

Definition 2 (Trace) *For $\mathcal{F} \subseteq \mathcal{P}(X)$, $S \subseteq X$, the trace of $\mathcal{F}$ on S is*

$$\mathcal{F}\big|_S := \{F \cap S : F \in \mathcal{F}\}$$

and $\operatorname{tr}_{\mathcal{F}}(S) = \left|\mathcal{F}\big|_S\right|$.

We say that S is shattered by $\mathcal{F}$ if $\mathcal{F}\big|_S = \mathcal{P}(S)$, and define the VC-dimension of $\mathcal{F}$ as

$$\max\{|S| : S \subseteq X \text{is shattered by } \mathcal{F}\}.$$

To give some intuition: as a result of this, we're mainly concerned with being able to divide up a finite set S of elements by the various sets in $\mathcal{F}$. For example, the VC-dimension of the set of half-spaces in $\mathbb{R}^2$ is 3, because it's not possible to draw 4 points on a plane without there being two lines between distinct pairs which intersect, and thus any half plane containing two of these opposite points must also contain their line, and thus the intersection, including a third point.

Theorem 15 (Shauer-Shelah) *If $\mathcal{A} \subseteq \mathcal{P}(n)$ has VC-dimension at most d,*

$$|\mathcal{A}| \leq \left|[n]^{(\leq d)}\right|$$

The intuition is that to shatter a set in $[n]^{(d+1)}$, it's sufficient to just have $> \left|[n]^{(\leq d)}\right|$ sets in your system, because at a certain point we have to capture at least one $[n]^{(d+1)}$ set fully.

Theorem 16 (Kleitman's theorem) *Let $\mathcal{A}, \mathcal{B} \subseteq \mathcal{P}(n)$ be downsets. Then*

$$|\mathcal{A} \cap \mathcal{B}| \geq 2^{-n}|\mathcal{A}||\mathcal{B}|$$

The intuition for this should be that $|\mathcal{A} \cap \mathcal{B}|/2^n \geq |\mathcal{A}|/2^n \cdot |\mathcal{B}|/2^n$, and thus that downsets must intersect considerably (in particular, that if we include each element in $[n]$ with probability $1/2$, $\mu(A \cap B) \geq \mu(A)\mu(B)$). This has an important application in problem sheet 3, and more generally has applications wherever we want to bound the cardinality of families which are likely to agree considerably due to intersection properties.

Definition 3 *$\mathcal{A} \subseteq \mathcal{P}(n)$ is t-intersecting if $|A \cap B| \geq t$ for all $A, B \in \mathcal{A}$.*

Note that a system is 1-intersecting iff it is intersecting, and if $s \leq t$, t-intersecting implies s-intersecting.

Theorem 17 (Generalised Erdös-Ko-Rado) *With $1 < t \leq k$ integers, there is an integer n_0 such that for $n > n_0$, $\mathcal{A} \subseteq [n]^{(k)}$ is t-intersecting implies*

$$|\mathcal{A}| \leq \binom{n-t}{k-t}$$

with equality iff $\mathcal{A} = \{A \in [n]^{(k)} : T \subseteq A\}$ for some $T \in [n]^{(t)}$.

This is a generalisation of the Erdös-Ko-Rado theorem for large n. The point of this is to give us even stronger bounds on sets which we know are 'very intersecting'. Note that for any $(t+1)$-intersecting system, we can replace one element within any set with another, and this gives us a set which is t intersecting with all others, thus we can assume the existence of $A, B \in \mathcal{A}$ with $|A \cap B| = t$ for $\mathcal{A}$ maximal. We can then argue that for n large enough, $A \cap B \subseteq C$ for all $C \in \mathcal{A}$.

It turns out we get a distinctly tight bound when the sense of t-intersection is strict:

Theorem 18 (Fisher's inequality) *Let $k \geq 1$. If $\mathcal{A} \subseteq \mathcal{P}(n)$ satisfies $|A \cap B| = k$ for all distinct $A, B \in \mathcal{A}$, then $|\mathcal{A}| \leq n$.*

This follows from a new type of argument which we continue to use in this section: to identify sets with their characteristic function (a vector in $\mathbb{F}_2^n$), and then use properties of linear independent sets.

Theorem 19 (Oddtown theorem) *Let $\mathcal{A} \subseteq \mathcal{P}(n)$ be a family such that $|A|$ is odd for $A \in \mathcal{A}$ and $|A \cap B|$ is even for all distinct $A, B \in \mathcal{A}$. Then $|\mathcal{A}| \leq n$.*

Working over $\mathbb{F}_2^n$, $\langle \chi_A, \chi_B \rangle = |A \cap B| = \chi_{A=B}$ (as we're working modulo 2), and thus we can claim that the set $\{\chi_A : A \in \mathcal{A}\}$ is linearly independent, as $\langle \sum_{A \in \mathcal{A}} \lambda_A \chi_A, \chi_B \rangle = \lambda_B = 0$ for all B if the initial sum is zero. Thus $|\mathcal{A}| \leq n$.

Theorem 20 (Modular Frankl-Wilson theorem) *Let p be prime, $S \subseteq \{0, 1, \ldots, p-1\}$, and $\mathcal{A} \subseteq \mathcal{P}(n)$ satisfies*

- *$|A| \notin S \bmod p$ for all $A \in \mathcal{A}$, and*
- *$|A \cap B| \in S \bmod p$ for all $A, B \in \mathcal{A}$ distinct.*

Then

$$|\mathcal{A}| \leq \sum_{i=0}^{|S|} \binom{n}{i}.$$

The proof of this involves once again working in $\mathbb{F}_p^n$. For $A \in \mathcal{A}$, define

$$f_A(x) = \prod_{s \in S} \left(\sum_{i \in A} x_i - s \right)$$

Now on the basis that we only want to evaluate our polynomials on characteristic vectors, we can 'multilinearise' them, replacing each power x^k by x in order to squeeze down the dimension of the space within which we're working. We then get a set of linearly independent multilinear polynomials of degree at most $|S|$, and as this is spanned by the monomials $\{\prod_{i \in A} x_i : |A| \leq |S|\}$, the result follows by the same principle just by calculating the dimension.

Theorem 21 (Frankl-Wilson theorem) *Let $\mathcal{A} \subseteq \mathcal{P}(n)$ be S-intersecting (for $A, B \in \mathcal{A}$ distinct, $|A \cap B| \in S$). Then*

$$|\mathcal{A}| \leq \sum_{i=0}^{|S|} \binom{n}{i}$$

A reduced form of this with the assumption $|A| \notin S$ for $A \in \mathcal{A}$ is immediate from the above, and we don't prove the more general version.

Theorem 22 (Ray-Chaudhuri-Wilson) *Let $\mathcal{A} \subseteq [n]^{(k)}$ be S-intersecting with $S \subseteq \{0, \ldots, k-1\}$. Then*

$$|\mathcal{A}| \leq \binom{n}{|S|}.$$

In this sense, the above theorems can be ordered roughly in strength for demonstrating that an intersecting family must be small.

$$\text{RCW} \sim \text{Fisher} \succ \text{M-FW} \succ \text{G-EKR} \sim \text{FW}.$$

Note that the oddtown theorem is merely a specific case of Modular Frankl-Wilson.

We now consider a key application of the above theorems, which is the counterexample to Borsuk's conjecture.

Theorem 23 (Kahn-Kalai) *Let $k(d)$ be the smallest integer such that every subset of $\mathbb{R}^d$ of diameter 1 can be partitioned into $k(d)$ sets of smaller diameter. Then there is $c > 1$ such that*

$$k(d) \geq c^{\sqrt{d}}$$

for infinitely many d.

Our overarching goal within this proof is to construct, with d an increasing function of p, a set $K \subseteq \mathbb{R}^d$ corresponding to a set of graphs, such that if $L \subseteq K$ has a smaller diameter, by the nature of the identification the graph of L takes up a bounded fraction of the graph of K (which we can then use to bound below the number of such Ls we need to divide K up).

To do this, take $d = \binom{4p}{2}$, and with the natural identification of $\{0,1\}^d$ with subsets of $E(K_{4p})$, write for $A \in [4p]^{(2p)}$, E_A as the complete bipartite graph connecting A and A^c, $\mathcal{F} := \{E_A : A \in [4p]^{(2p)}\}$. We then define our $K \subseteq \mathbb{R}^d$ as the natural identification of $\mathcal{F}$, and note that for $A, B \in [4p]^{(2p)}$, $||v_{E_A} - v_{E_B}||^2 = |E_A| + |E_B| - 2|E_A \cap E_B|$, so thus any $L \subseteq K$ with smaller diameter, for A, B distinct in $\mathcal{A}$ the set of subgraphs corresponding to L, $|E_A \cap E_B|$ cannot be minimal (occurring iff $|A \cap B| = p$), and thus we get our result from the lemma, that

$$|\mathcal{F}|/|\mathcal{A}| \geq \frac{1}{6}\binom{4p}{2p} \Big/ \binom{4p}{p-1}$$

and thus covering K requires this number of sets, which one can see is asymptotically exponential in p, which is itself asymptotically equivalent to $\sqrt{d}$.

Lemma 24 *Let p be a prime and $\mathcal{A} \subseteq [4p]^{(2p)}$. If there is no distinct pair $A, B \in \mathcal{A}$ with $|A \cap B| = p$, then $|\mathcal{A}| \leq 4\binom{4p}{p-1}$.*

Take $x \in [4p]$ which occurs in at least half the sets in $\mathcal{A}$ (see existence by a probabilistic argument). Note that for distinct $A, B \in \mathcal{A}$ both containing x, $|A \cap B| \in \{1, \ldots, p-1\} \bmod p$, and $|A| = 2p \notin \{1, \ldots, p-1\} \bmod p$, so applying M-FW gives

$$\frac{1}{2}|\mathcal{A}| \leq |\{A \in \mathcal{A} : x \in A\}| \leq \sum_{i=0}^{p-1} \binom{4p}{i} \leq 2\binom{4p}{p-1}$$

where the last equality is from rearrangement with the inequality $\binom{4p}{i-1} \leq \binom{4p}{i}/2$, and thus $|\mathcal{A}| \leq 4\binom{4p}{p-1}$ (indeed we can also get $|\mathcal{A}| \leq 3\binom{4p}{p-1}$ by the same method, but this is unnecessary).