Contents

I

Fundamentals

1 Introduction to Problem-Solving

1.1 Introduction

Olympiad mathematics is all about solving problems. But these are different from the problems you'd see elsewhere in school. Most problems here are unique because solving them usually requires you to *think* innovatively and out of the box. The key idea on which the olympiad world is based is that a person with minimal formal math education should be able to understand the solution. Still, even graduates in the subject should struggle to *find* the solution.

Often, the problems you'll see in olympiads (and in this book) will be such that even people around you who aren't really into math can understand them. This makes these problems a more genuine test of your thinking skills and intelligence rather than knowledge. That being said, olympiads *are* trainable to some extent (or else why would I be writing this book?). By "trainable" I mean that experience with a wide set of problems can improve your intuition to a standard where even out-of-the-box problems feel like they were *in your box*.

Through this book, I aim to make your *box* bigger, too, and show you some amazing and beautiful problems that hopefully find a place in your mind forever and help you discover the beauty of this subject.

On this note, some fun problems have been added in the next section. The first 5 require nothing but common sense and very elementary mathematics. The following five are slightly more tricky and will put your logical skills to the test. Gradually, the problems become tricky and require deep thought, but their solutions only need elementary math. Hints have been added for the problems, but I strongly recommend thinking about the problems for some time before using the hints.

1.2 Problems

By the way, the number written after Hint is the hint number you need to refer to in the Appendix at the back of the book if you want a hint on the problem.

Problem 1. A number of bacteria are placed in a glass. One second later, each bacterium divides in two, and the next second, each of the resulting bacteria divides in two again, et cetera. After one minute, the glass is full. When was the glass half-full?

Hints: 5 25

Problem 2. A caterpillar crawls up a pole 75 inches high, starting from the ground. Each day, it crawls up 5 inches, and each night it slides down 4 inches. When will it first reach the top of the pole?

Hints: 247 293

Problem 3. Can one make change of a 25-rupee bill using ten coins, each having a value of 1, 3, or 5 rupees?

Hints: 197 493 241 72

Problem 4 (Oxford MAT). Alice, Bob, and Charlie make the following statements,

Alice: Bob and Charlie are both lying.
Bob: Alice is telling the truth, or Charlie is lying (or both).
Charlie: Alice and Bob are both telling the truth.

Who is telling the truth, and who is lying?

Hints: 570 338

Problem 5 (Oxford MAT). Alice, Bob, and Charlie are well-known expert logicians; they always tell the truth. Each of them is wearing a hat, which is either red or blue in color, and they are sitting in a row so that Alice can see Bob's and Charlie's hats but not her own. Bob can see only Charlie's hat, and Charlie cannot see any hat. All three of them are aware of the arrangement.

An angel says, "At least one of you is wearing a red hat."

Alice begins by saying, "I don't know the color of my hat."

Bob says, "I know the color of my hat!"

What color is Charlie's hat?

Hints: 534 578 459 212

Problem 6 (Oxford MAT). Alice, Bob, Charlie, and Diane are playing when one of them breaks a precious vase. They all know who broke the vase. When questioned, they say the following:

Alice: "It was Bob."
Bob: "It was Diane"
Charlie: "It was not me."
Diane: "What Bob says is wrong."

For each of the following cases, find who broke the vase.

- An angel comes and says that only one of the four is lying
- An angel comes and says that only one of the four is saying the truth

Hints: 484 365

Problem 7 (Oxford MAT). There are n people seated in a circle. Each of them is either a saint who always tells the truth or a liar who always lies. Let's say each of the n people says, "The person on my left is lying, and the person on my right is telling the truth." How many people lied?
Let's say each person says, "Either the people to my left and right are both lying, or both are telling the truth." If at least one person is lying, show that n is a multiple of 3.

Hints: 21 240

Problem 8. Two prisoners are given the following situation: they must go to different rooms with no means of contact with the other person. Both of them toss a coin, after which they must guess the result of the other person. If at least one of them guesses the result of the toss correctly, they are both free. Show that by discussing a strategy beforehand, the prisoners can guarantee freedom.

Hints: 103 404 425

Problem 9. There are ten ants on a rod. Each of them moves at a speed of 1 mm/s, and the length of the rod is 10 cm. They have been given some starting positions and directions, and they begin to move with their given speed. However, here's the catch: every time two ants bump into each other, they reverse their directions. If the ant reaches the end of the rod, it falls off and dies. Find the maximum time for which at least one ant stays alive across all possible starting positions and directions.

Hints: 165 89 92

Problem 10. Three prisoners are now given this situation: each of them is wearing a hat, either red or blue (selected uniformly at random). They don't know whether their hat is red or blue, but they can see the colors of the hats of the other two people. Now, each prisoner has two options: they can either make a guess on the color of their hat or say "I skip". At the end if even one of the prisoners guessed incorrectly or all said "I skip", they are hanged. However, the prisoners are free in all other cases (where all guesses are correct and at least one guess is made). Show that the prisoners can win with 75% probability.

Hints: 183 438 543

Problem 11 (Oxford MAT). A positive integer is written on the forehead of each of Alice and Bob. They can see the number written on the other person's forehead, but they can't see the number on their head. Naturally, neither of them knows the number written on their head. Now imagine an angel arrives and makes the following statement, "The numbers on your foreheads differ by 1." So the number on Alice's forehead is either one more or one less than the number on Bob's forehead.

Imagine the following scenario:

> Alice begins by saying, "I don't know my number."
> Bob replies, "I don't know my number either."
> Alice replies, "I still don't know my number".
> Bob replies, "Aha, I now know my number!"

Find all possible values of Alice and Bob's numbers.

Also, show that in general, given that the numbers are consecutive in some order, Alice and Bob can figure out their numbers by just repeatedly saying "I don't know my number" like above.

Hints: 204 7 596 **Soln:** Page 271, Solution 1

Problem 12. Ten prisoners are given the following situation: they are all given a hat with one of ten colors uniformly at random (it's possible that everyone gets the same color, and it's also possible that all ten prisoners get ten different hat colors). Each prisoner can see the hats of others but not their own. Now, each prisoner must guess what they think is the color of their hat. If even one of them guesses their hat color correctly, they are all free. Show that by discussing a strategy beforehand, the prisoners can guarantee freedom.

Hints: 323 33 109 **Soln:** Page 272, Solution 2

2 Introduction to Counting

A lot of times, we're faced with questions like:

"How many different ways are there to do this?"

One way to figure this out is to explicitly consider all possibilities and count them all individually. In general, however, we seek to avoid that as far as possible (why do so much work when you can do some smart math!), and in this chapter, we'll learn a couple of methods to reduce the computational hard work.

The question to ask here:

Why would I care how many ways are there?

There are various reasons for you to care. Imagine I asked you how many ways you could jumble up a 3 by 3 Rubik's Cube. Sure, there are a lot. But *a lot* is hardly satisfactory to me. Knowing how many ways there are to do a certain thing quantitatively can be really useful in comparing different things. There are many ways of jumbling around ten words in a sentence, too, but is it more than or less than the number of ways to jumble a Rubik's Cube? These methods can tell us how *likely* it is that we get lucky if we start randomly. In particular, imagine we did this with passwords - finding out the number of possibilities for a certain password can give us an idea of how secure it is. A lot of computer science concepts work only based on the fact that you're almost certainly not *lucky* (the same way its really unlikely you jumble a cube randomly and end up at a solved cube). Moreover, it can also help us estimate how long a computer program will take to run (as we'll see in the chapter on algorithms!).

Apart from this, knowing methods of counting forms the additions and subtractions of combinatorics. A wide range of problems require some counting, just like math puzzles you may have done could require you to do some algebra. Besides, counting the total number of possibilities without having to brute-force through all possibilities is pretty satisfying and fun! They open us up to new worlds of sequences with magical patterns and lots of ongoing research.

2.1 Introduction

Let's imagine three balls in a line. In how many ways can we order the three?

The most obvious approach to solving this problem is trying to brute-force through all possibilities. So we have

A B C	B C A
A C B	C A B
B A C	C B A

All in all, there are six ways of doing so. Interesting. The next question: what if we had four balls instead of three? As it turns out, listing out all possibilities will become quite tedious. And if you try listing out possibilities for five balls, it may take you more than an hour or two! So we need to come up with something better.

Let's say our four balls are A, B, C, and D, and we put A at the start of the line.

What remains is mixing up three balls for the remaining three spots. We could begin to calculate that, but wait... we already did that above! Indeed, there are six ways of lining them up. But what if A isn't at the start of the line? Well, some ball has to be at the start of the line, and it's not A, so let's say it's B. Then we need to arrange A, C, and D for the remaining three spots - there are six ways of doing this, too. In particular, you get the same for each ball at the start of the queue. So the total number of ways of doing so should be $6+6+6+6 = 6 \times 4 = 24$.

Exercise. Use this to figure out the answer for five balls.

Hints: 654 651

Problem. Find the answer when there are ten balls.

Hints: 207 93 565

Things are beginning to get messy, and there are a couple of questions:

- Why did we do $6+6+6+6$ and not, say, $6 \cdot 6 \cdot 6 \cdot 6$?
- Do we always need to add things, or are there cases where you need to multiply/divide?

It's important we try and strengthen the basics a little bit before moving forward. As it turns out, all of counting can be reduced to two basic principles as follows:

2.2 Addition and Multiplication Principle

1. **Addition Principle**: This essentially says that if you have to do **either** A **or** B, then the number of ways of doing so is the number of ways of doing A + the number of ways of doing B. One example of the above is the following let's say you have to go from Mumbai to Kolkata. There are two flights and three trains available from Mumbai to Kolkata. Then, the total number of ways of going from Mumbai to Kolkata is $2 + 3 = 5$.

2. **Multiplication Principle**: This says that if you have to do **both** A **and** B then the number of ways of doing so is the number of ways of doing $A \times$ the number of ways of doing B. For this one, consider the following: Once again you have to go from Mumbai to Kolkata. There are two flights from Mumbai to Delhi and three trains available from Delhi to Kolkata. Then the total number of ways of going from Mumbai to Kolkata is $2 \times 3 = 6$. The reasoning is that essentially, you have to first take one of the two flights (F_1, F_2) and then take one of the three trains (T_1, T_2, T_3). So, your six possibilities are

$$\begin{array}{ccc} F_1T_1 & F_1T_2 & F_1T_3 \\ F_2T_1 & F_2T_2 & F_2T_3 \end{array}$$

   In particular, for each of the A options for the first flight, you have B options for the second train, so the total number of ways is $B + B + \cdots + B = B \times A$

Exercise. There are two flights from Mumbai to Kolkata, two flights from Mumbai to Delhi, and two flights from Delhi to Kolkata. In how many ways can I start in Mumbai and end up in Kolkata?

Hints: 228 465 56

Now that we have a basic hang of these two principles, let's try answering the following question:

> **Example 1**
> Find the number of divisors of 3600.

Well, we could try listing them out. There's

$$1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, 18, 20, 24, 25, 30, 36, 40, 45, 50, 60$$

and then $3600 \div$ each of these: so $3600, 1800, \ldots$.

As it turns out, I've missed two of the factors in 1 to 60. Do I really need to go through the entire list again? Urgh.

I think you get the idea. Listing out possibilities is not just a long procedure but also one that is very risk-prone, another reason why we need smarter ways to calculate such things. Let's consider the prime factorization of 3600. We know that

$$3600 = 2^4 \cdot 3^2 \cdot 5^2$$

The idea is that any of its divisors also looks like $2^a \cdot 3^b \cdot 5^c$, with a as any number at most 4, and b and c as any numbers at most 2. In particular, we have five options for a - $\{0,1,2,3,4\}$, and we have 3 options for b and c - $\{0,1,2\}$. We need to pick

(a value for a) AND (a value for b) AND (a value for c)

so there are $5 \cdot 3 \cdot 3 = 45$ factors we can create - and that's our answer.

Example 2

How many 3 digit numbers have all odd digits?

Note that the unit digit can be one from $\{1,3,5,7,9\}$. So there are 5 ways of picking the unit digit (we actually used the addition principle here - we need to pick 1 OR 3 OR 5 OR 7 OR 9).

Similarly, there are 5 ways of picking the tens and hundreds digits too. In all, we need to pick one out of the five for each: so we need to pick

(a unit digit) AND (a ten's digit) AND (a hundred's digit)

so if A is the number of ways of picking a unit digit, B is the number of ways for the ten's digit, and C is the number of ways for hundred's digit, we want to find

$$A \cdot B \cdot C = 5 \times 5 \times 5 = 125$$

Example 3

How many 3 digit numbers with *distinct digits* have all odd digits?

We have now added in this *distinct* constraint. So 113 no longer satisfies the constraint, while 135 continues to satisfy the constraint.

We are, once again, picking a unit's digit AND a ten's digit AND a hundred's digit, so if A is the number of ways of picking the unit digit, B is the number of ways for the ten's digit and C is the number of ways for hundred's digit, we want to find $A \cdot B \cdot C$.

Let's first fix our unit digit as one of the five like the previous time, so $A = 5$. The difference this time is that the ten's digit has only four options now - one of them has been taken away, so $B = 4$. In turn, the hundreds digit has only three options (two are gone), and $C = 3$.

So the answer is $5 \cdot 4 \cdot 3 = 60$

This idea forms the base for the concept of *permutations*, which we tried to understand earlier.

2.3 Permutations

Example 4

Let's say we want to seat 5 people on 5 chairs. How many ways are there to do so?

Just to give an example: let's say the five people are A, B, C, D, E. Then we can seat them in the order $ABCDE$ but also in the order $CEDAB$ or $DAEBC$. We've been tasked with finding the total number of possibilities.

With the same idea as last time, let's first pick a person for the first seat. This can be done in 5 ways. Now, four people are left for the second seat, so there are four options. For the third seat, there are 3 options. For the fourth, there are 2 and for the final seat, there is only one option. We need to pick a person for the first seat AND a person for the second seat AND a person for the third seat, and so on. The multiplication principle allows us to conclude that the total number of ways to do this are $5 \times 4 \times 3 \times 2 \times 1$. Notice that earlier, when we were working with balls we did this recursively, as $6 + 6 + 6 + 6 = 24$, and $24 + 24 + 24 + 24 + 24 = 120$, but this is a more direct way of doing the same calculation.

In general, if we started off with n people and n chairs, we would have ended up with $n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$. This expression is often denoted as $n!$ in short (read out as n factorial). The process of mixing up objects in a certain order is known as "permuting" the objects, and the concept is known as a permutation.

By now, you're probably thinking,

> "Okay, I believe you till now. But do I really care about the number of ways to rearrange people on some chairs?"

You should! This expression can help us solve various combinatorics problems - and is useful in various contexts - not just in mathematics, but even in chemistry, for example. Not convinced? Wait a bit.

2.4 Combinations

Problem. Let's say we have three items and want to take two of them to a party. In how many ways can I do so?

So let's say I have three items: a phone, a charger, a hat.

In practice, I would almost always take my phone and charger (oops!), but for what it's worth, let's see how many possibilities existed.

So I could choose to take my phone and my charger. I could also choose to take my phone and the hat. Finally, I could also take the charger and a hat (I wonder what I would do with a charger without a phone, though). Anyway, all in all, there are three ways.

The natural follow-up questions are: what if I need to pick two items out of four? What if I need to pick two items out of five? Or three items out of five? Let's see!

Problem. You have been given 4 distinct objects. You want to go outside, but you're allowed to take only two of the objects. In how many ways can you do so?

Let's call the objects A, B, C and D this time. I can make the following choices

A B
A C
A D
B C
B D
C D

So there are six ways of doing this

Problem. You have been given 5 distinct objects. You want to go outside, but you're allowed to take only two of the objects. In how many ways can you do so?

Let's call the objects A, B, C, D and E this time. I can make the following choices

A B	A E
A C	B E
A D	C E
B C	D E
B D	
C D	

See what I did there? The possibilities on the left are the same as the ones for Problem 2. And it does make sense - if I need to pick two out of A, B, C, D, E, I can either ignore E and solve the same problem as earlier, or I can pick E - in which case I now need to pick just one other number so there are four ways.

What if I had six objects? Either I ignore the sixth: so there are ten ways for that or I include it and pick one of the remaining five: so there are five ways. All in all, I have to do one of these so the total number of ways is $10 + 5 = 15$.

Problem. Find a general formula for the number of ways of picking 2 objects out of n.

Hints: 80 439

But we don't really need to be stuck on picking just two objects. What if I had to pick 11 objects out of 34?

We now define a new expression.

$$\binom{n}{r}$$

(Read out as n choose r.)

This is the number of ways of picking r people out of n to do a certain job for you, become the organizers of a committee, leaders of the country, etc. We wish to evaluate how many ways there are to do these things.

Let's try and find some smart way of calculating this for, say, $\binom{10}{4}$. So we want to pick four people out of ten. Let's take two cases: either the first person is in the team, or he isn't. Obviously, exactly one of these happens so we can add up the cases at the end. If he was picked, we need to pick 3 out of the remaining 9. If he wasn't, we need to pick 4 out of the remaining 9. So,

$$\binom{10}{4} = \binom{9}{3} + \binom{9}{4}$$

Nice! But this means that to calculate $\binom{10}{4}$, we'll have to calculate $\binom{9}{4}$ and $\binom{9}{3}$, for which we'll have to calculate $\binom{8}{4}$ and many more. Eventually, we should be able to do it, but it'll be quite messy, no? We wish to find a direct method to do it.

And permutations come back to save the day! Consider a permutation of the 10 people - so you have placed them on ten chairs one after the other. I needed to select four - so what if I just select the first four people in this permutation?

Yes! That works. But wait... does it? We calculated that $\binom{4}{2} = 6$. But the number of ways of mixing up four people is 24, not 6. What's the issue?

Let's consider some of the possibilities: $ABCD, ABDC, BACD, BADC$. All four of these correspond to picking A and B. So, we're counting each pair multiple times. If we knew how many times we're counting a single group of people, we could just divide by that, and now we would (hopefully) have the answer with us.

Let's go back to our four of ten examples (it's easier to understand with slightly bigger numbers). So we want to check how many times we're counting the group of A, B, C, D. We'd be counting this group if the first four form a permutation of A, B, C, D, i.e., are something like $CADB$ or $DABC$. There are 4! ways of doing this.

In addition, the other six elements can also be permuted among the last six seats in any way: so $EFGHIJ$ or $GFEJIH$, for example. We want to mix up the first four amongst themselves AND the next six amongst themselves. By the multiplication principle (check this), the total number of times we ended up picking A, B, C, D by this method is $4! \times 6!$. So we've counted each possibility $4! \cdot 6!$ times and the actual number of ways of picking four of ten is

$$\binom{10}{4} = \frac{10!}{4! \cdot 6!}$$

In general (by an identical reasoning),

Theorem

$$\binom{n}{r} = \frac{n!}{r! \cdot (n-r)!}$$

Exercise. Check that $\binom{4}{2} = 6$ as per this formula as we would have liked.

2.5 Inclusion and Exclusion

Let's say you're tasked with finding the answer to the following problem:

Example 5
Find the number of positive integers at most 1000 that are multiples of 2 or 3 (or both).

We can find the multiples of 2 at most 1000 quite easily: they are just the numbers in the set $\{2, 4, 6, \ldots, 998, 1000\}$. In all, there are 500 numbers here. Similarly, the multiples of 3 are $\{3, 6, 9, \ldots, 999\}$, and so there are 333 of them.

So now what? We want A or B, so we just add them up? The issue is the "or both". Consider the number 6. It's getting counted in both sets, but it should be counted only once. This is an issue known as over counting.

The simple way of dealing with this is to subtract the number of values you counted twice - then those numbers will be counted once too.

In particular, every multiple of six is counted twice above. So we subtract the number of positive integers at most 1000 that are multiples of 6 - there are 166 of them. So the final answer becomes

$$500 + 333 - 166 = 667$$

This technique, in general, is known as the principle of "inclusion and exclusion" and you may have encountered it while learning set theory at school.

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

Here A was the set containing multiples of 2, and B was the set containing multiples of 3. Then, the number of elements in the union of the sets is the number of multiples of 2 + the number of multiples of 3 - the number of multiples of six.

Exercise. Find the number of positive integers less than 1000 that are not multiples of 2 or 3.

Hints: 106 158

2.6 Bijection Principle

Let's introduce one final useful trick to help us solve a lot of counting problems: the *Bijection principle*. This principle essentially says that if you can create some conditions on two sets, they must be of the same size.

> **Lemma** (Bijection Principle)
>
> If there is a *bijection* between two finite sets A and B, the sets have the same number of elements so $|A| = |B|$.

First off, let's discuss what a "bijection" means. Imagine the following scenario: you have a set of chairs in a classroom, and a set of students.

If the following two conditions are satisfied, there is a bijection between the sets

- For each chair, there is a student sitting on it.

- For each student, there is a chair they sit on.

It is quite intuitive that if these two conditions hold, there must be an equal number of chairs and students. By the first condition, there can't be any "left-over" chairs, and by the second condition, there couldn't have been a shortage of them either.

The more abstract version of this is to consider sets A and B and a *mapping* from one set to the other.

$$\begin{array}{c}\{1,2,3,4,5\}\\ \begin{array}{ccccc}\downarrow & \downarrow & \downarrow & \downarrow & \downarrow\end{array}\\ \{a,b,c,d,e\}\end{array}$$

In particular, you can think of a mapping like the arrows above.

In general, when you're trying to create a bijection, the simplest way to do so is to ask yourself the following two questions:

> Can I map each element in A to an element in B? And can I reverse the mapping, so given an element in B, can I find an element in A that leads to it?

Now, the first condition means that the size of B is at least as large as A, and the second part means that the size of A is at least as large as B. Together, we conclude that $|A| = |B|$.

Let's now see how we can put this together in an actual problem.

Example 6
You have the set $S = \{1, 2, 3 \ldots, n\}$. Show that S has 2^n subsets.

Let X be the set of subsets. Then for $n = 2$, we have that

$$S = \{1, 2\} \implies X = \{\{\}, \{1\}, \{2\}, \{1, 2\}\}$$

So there are 4 subsets, which is what we wanted to show for $n = 2$. We need to show this for general n, though.

Here comes the magic trick. Consider the set Y of binary strings of length n. As an example

$$Y = \{00, 01, 10, 11\}$$

when $n = 2$. We show that there is, in fact, a bijection between the sets X and Y. We wanted to show that there are 2^n elements in X, so if we have created this bijection, we will just need to show that there are 2^n elements in Y (which turns out to be considerably more direct).

So we need to show that we can create some mapping that takes in subsets of $\{1, 2, \ldots, n\}$ and gives us binary strings of length n and our two conditions are satisfied. This is actually not so hard to achieve. For instance, if for example $n = 5$, map $\{1, 3, 4\}$ to 10110 since the first, third and fourth elements are present. Also, each binary string comes from some subset because you can do the process backwards: 11001 must have come from $\{1, 2, 5\}$. This means that there's a bijection between X and Y, as required!

Problem 7. Show that the number of binary strings of length n are 2^n.

Hints: 162 462

> **Remark.** We could also have done this directly as follows:
>
> > Let's take $n = 3$ as an example, the exact idea generalises. Note that in any such subset: we have the following choices - do we put the first element in the subset? Do we put the second element in the subset? And do we put the third element in the subset? Each of these can be done in two ways.
> >
> > Also, we need to make a choice for the first element AND make a choice for the second element AND make a choice for the third element. So there are $2 \times 2 \times 2$ ways of creating a subset which is 2^3.
>
> Why, then, is the bijection principle useful?
>
> In a lot of scenarios, there may be a method to work directly, but converting the problem into counting something else can make it a lot more intuitive and easy to work with and can also make the proof a little neater.

> **Remark.** Another question to raise is, how do you come up with that set Y? Why would any reasonable person randomly think of taking binary strings? Let's try answering that after the next question!

Example 8

We have an 8×8 chessboard. We start at the bottom left corner and end up at the top right corner. In a move, we can either go upwards one step or go rightwards one step while staying within the board. How many different paths to the top right cell exist?

The idea here is that at each point you want to go either up or right. So maybe consider strings of such moves:

Up, Right, Right, Up, Up

for example. Which strings map to a path from the bottom left to the top right? Well, I must have made seven moves to the right in all to reach the 8th column. Similarly, I must have made seven up moves to reach the 8th row. So I must have moved up and right each a total of seven times.

So these paths are actually equivalent to these *up-right* sequences that have seven up's and seven right's. Does every such path correspond to a unique sequence? Yes, it does. Does every such sequence correspond to a unique path? Indeed. This means that there's a bijection between the two sets. (!!)

So we have that the number of paths is equal to the number of *up-right* sequences that have seven up's and seven right's. So we have a total of 14 moves, seven of which are to the right. In particular, if we fix which seven are up, the remaining seven are right. So all we need to do, is pick 7 out of the 14 moves as right - and that gives us the answer

$$\binom{14}{7}$$

which is the required answer!

> **Remark.** So your set Y is usually quite intuitive - it doesn't just come up out of thin air. In the subsets example, the idea was that we want to decide for each element whether we want to keep it or not. So we were looking at sequences like
>
> Keep, Don't keep, Keep, Keep
>
> And binary strings were just a cool way to rephrase that. Here too, we were just looking at the problem from a slightly different perspective - *move by move* rather than the entire path together, which gave rise to the idea of up's and right's.

2.7 Examples

Example 9

Consider the equation

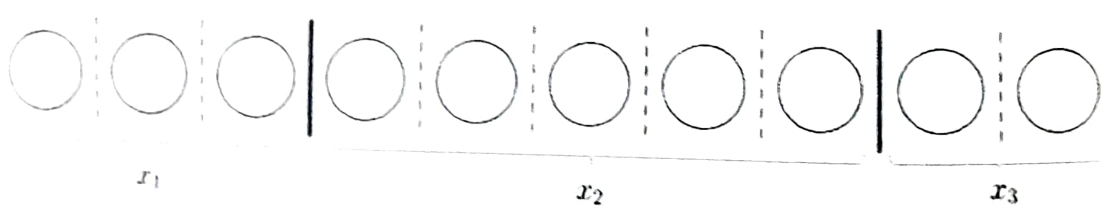
$$x_1 + x_2 + x_3 = 10$$

Here x_1, x_2, x_3 are positive integers. Find the number of positive integer solutions to the equation. (As an example, $(3, 5, 2)$ is a solution to the equation.)

We could obviously go through all the cases but there are two significant disadvantages

- It's a pain.
- You wouldn't be able to do that if I replace 3 and 10 with 500 and 1000 bu I can still give you an expression for the answer.

So let's try interpreting the problem *combinatorially*. Imagine you have 10 chocolates. Now you want to split them across 3 children, such that each child gets a least one chocolate. How do you do this? Well, let's first give some chocolate to the first child. Let's say we gave him 3 chocolates. Now we want to split th remaining 7 among the two. Say we give the second child 5, and then we giv the third child 2. That corresponds to the solution we wrote above. Graphicall we've considered the following splitting of ten chocolates into three parts.



But wait... so are we just breaking ten chocolates into three parts? That's lik creating two lines in between the dots like above. You have 9 spots for the line and you want to pick two of them to create your splitting.

So there are

$$\binom{9}{2}$$

ways of splitting the numbers. We just need to check that there's a bijectio between the splittings and the solutions to the equation. Well, each equatio solution definitely corresponds to a splitting - we can take x_1 in the first group, x

in the second and x_3 in the third. On the other hand, each splitting corresponds to a unique solution too - so there is indeed a bijection!

Example 10

There are nine children at the party, three of whom are wearing green shirts, while the remaining six are wearing blue shirts. Find the number of ways of arranging the children in a line such that no two children wearing green shirts are together.

The main condition to keep in mind here is

> "No two children wearing green shirts are together."

We need to somehow keep this condition in mind when arranging the children. Note that this is also equivalent to

> "Between any two children wearing green shirts, there's at least one child wearing a blue shirt."

So the idea here is going to be to randomly arrange the children wearing blue shirts first, and then place the others in the "gaps". In particular, after arranging the six children wearing blue shirts, you have seven gaps as seen below.

$$\square\, 1\, \square\, 2\, \square\, 3\, \square\, 4\, \square\, 5\, \square\, 6\, \square$$

There can be at most one child in each gap (or there'll be two children wearing green shirts together). We need to place three children into these gaps, and there are $\binom{7}{3}$ ways of doing that. Finally, these three children, as well as the six other children can be mixed up amongst themselves, so in all there are

$$\binom{7}{3} \times 6! \times 3!$$

ways of mixing them up.

Example 11

Show that

$$\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = 2^n$$

Imagine you have 3 people.

If you pick zero out of them, this can be done in $\binom{3}{0}$ ways. If you pick one out of them, there are $\binom{3}{1}$ ways. Picking two of them can be done in $\binom{3}{2}$ ways, and if you pick all of them, there are $\binom{3}{3}$ ways.

Now we want them to add up - so we want to get some *ors*. In particular, imagine you wanted to pick some 0 of them, or some 1 of them, or some 2 of them, or some 3 of them. Then the number of ways of doing this is the LHS. What we're basically doing is taking some subset of the people. So the LHS counts the number of distinct subsets of the set of n objects.

But... we did that already! And it comes out as 2^n, and so we're done!

> **Remark.** The problem can also be proved using induction, and you'll be asked to do so in the problems section of the next chapter!

2.8 Problems

Problem 1. Let n, r be positive integers with $0 \leq r \leq n$. Show that

1. $\binom{n}{r} = \binom{n}{n-r}$

   Hints: 260 487

2. $r\binom{n}{r} = n\binom{n-1}{r-1}$

   Hints: 150 200 90

3. $\binom{r}{r} + \binom{r+1}{r} + \cdots + \binom{n}{r} = \binom{n+1}{r+1}$

   Hints: 509 522

Problem 2. Let n be a positive integer. Show:

$$\sum_{r=0}^{n} r\binom{n}{r} = n \cdot 2^{n-1}$$

Hints: 31 499

Problem 3. Show that

$$\binom{100}{0} + \binom{100}{2} + \binom{100}{4} + \cdots + \binom{100}{100} = 2^{99}$$

Find and prove a similar result for 100 replaced by a general n.

Hints: 389 383

Problem 4. Further to Example 9, show that the equation

$$x_1 + x_2 + \cdots + x_r = n$$

has $\binom{n-1}{r-1}$ number of solutions if $x_1, x_2, \ldots, x_r$ are all positive integers.

Hints: 287

Problem 5. You have a deck of 52 cards with four suites: hearts, diamonds, spades, and clubs. Each suite has 13 cards: the Ace, numbers from 2 to 10, Jack, Queen, and King. You pick four cards.

1. In how many ways could you have picked two Aces and two Kings?

   Hints: 416

2. In how many ways could you have picked up three number cards (2-10) and one face card?

   Hints: 455 66

3. In how many ways could you have picked cards from four different suites?

   Hints: 371

Problem 6. There are a men and b women at a party, with $a < b$. In how ma ways can these $a + b$ people be arranged in a row if

1. No two men are together.

   Hints: 593

2. Two particular women, Mrs. X and Mrs. Y, are adjacent.

   Hints: 55

Problem 7. Find (or describe a method to find) the number of numbers less tha 1000 that are multiples of 2, 3, or 5.

Hints: 112 390

Problem 8. There are 10 points marked in a circle, and chords are drawn f each pair of points. If no three chords concur at a point strictly inside the circ find the number of intersections strictly inside the circle (which are the meetin point of two chords).

Hints: 374 210 555

Problem 9 (RMO). Find the number of eight-digit numbers with the sum digits 4.

Hints: 656 346

Problem 10 (2015 RMO/4). Suppose 28 objects are placed along a circle a equal distances. In how many ways can 3 objects be chosen from among them s that no two of the three chosen objects are adjacent or diametrically opposite?

Hints: 392 77

Problem 11 (2005 INMO/4). All six-digit numbers, in each of which the digit occur in non increasing order (from left to right, e.g. 877550), are written as sequence in increasing order. Find the 2005th number in this list.

Hints: 377 98 301

Problem 12. Snow White and the seven dwarves are playing a game, and the need to be split into groups of 2. Show that this can be done in exactly 105 way

Hints: 489 159 292 **Soln:** Page 272, Solution 3

Problem 13. Consider a $2 \times 2 \times 2$ Rubik's cube. Each of its 8 pieces can b oriented in three ways, and each of the 8 pieces can be permuted. However, onl one third of these can be reached with a solved cube (others require a "twist" Find the number of possible jumbles of the cube.

On a $3 \times 3 \times 3$ cube, there are 8 corners and twelve edge pieces. Corners ca be oriented in three ways, edges in two, the corners can be permuted among themselves, and so can the edges. 1/6 of these can be reached with a solved cub Approximate the number of jumbles this time.

Hints: 590 135 387 **Soln:** Page 273, Solution 4

3 Induction and Recurrence Relations

Induction is one of the most useful techniques and finds applications in various fields of mathematics - not just combinatorics but number theory, algebra, and sometimes even geometry! The power of induction often lies in its simplicity.

3.1 Introduction

The idea behind induction is fairly simple. Let's say you want to prove some statement for all natural numbers n. Instead of proving it directly, you do the following:

- Show that the statement is true for $n = 1$.
- Show that if the statement is true for $n = k$, the statement is also true for $n = k + 1$.

To understand why and how induction works, it's useful to think of dominoes. Imagine doing the following two steps:

- Place dominoes one after the other with a small gap between them - so that if one falls then the one after it falls.
- Push down the first domino.

This causes all dominoes to fall down one by one! The idea with induction is similar. We know that the statement is true for $n = 1$ (by the first point). By the second point, we know that the statement is also true for $n = 1 + 1$ (by putting $k = 1$) and so the statement is true for $n = 2$. Repeating this, we get that the statement is true for $n = 2 + 1 = 3$ and so the statement is true for $n = 3 + 1 = 4$. This creates a sort of domino effect, and the statement is true for all positive integers n.

Let's take a very simple example to start off.

Example 1

Prove that

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$$

for all positive integers n.

Following the two steps, we first show that the statement is true when $n = 1$.

We know that

$$1 = \frac{1 \cdot 2}{2} = 1$$

And so the statement is true when $n = 1$. Now to the second part, let's say that the statement is true for $n = k$. We want to prove the statement for $n = k + 1$, so essentially we want to show that

$$1 + 2 + \cdots + k = \frac{k(k+1)}{2} \implies 1 + 2 + \cdots + k + (k+1) = \frac{(k+1)(k+2)}{2}$$

But we know that

$$\begin{aligned} \text{LHS} &= 1 + 2 + \cdots + k + (k+1) \\ &= \frac{k(k+1)}{2} + \frac{2(k+1)}{2} \\ &= \frac{k^2 + k + 2k + 2}{2} \\ &= \frac{(k+1)(k+2)}{2} \\ &= \text{RHS} \end{aligned}$$

And just like that, we've solved the problem!

Some terminology before we move forward:

1. **Induction base case**: The case for $n = 1$ which we need to prove separately. Note however that $n = 1$ need not always be the base case. If the problem asks you to prove, say, "Show that the following statement is true for $n \geq 5$" the case $n = 5$ forms your base case. You can also have multiple base cases if required - we'll see some examples where this is required soon.
2. **Inductive Hypothesis**: The assumption that $n = k$ is true is often called the inductive hypothesis. It is this hypothesis that allows us to complete the inductive step by proving the problem for $n = k + 1$.
3. **Strong induction**: Here the hypothesis is slightly different: we assume that the statement is true for all $n \leq k$, and prove that it's true for $n = k + 1$.

Note that this still works out nicely: the statement is true for all $n \leq 1$ so it's true for $n = 2$. It's now true for $n \in \{1, 2\}$ so it's true for all $n \leq 2$ and thus it's true for $n = 3$. And so on.

3.2 Simple Examples

Example 2

Show that

$$1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

We'll follow the same general method as previous time.

1. Prove that the statement is true for $n = 1$.
2. Show that the statement is true for $n = k + 1$ if it is true for $n = k$.

For the first part, let's evaluate the LHS and RHS.

$$\text{LHS} = 1^2 = 1$$

On the other hand,

$$\text{RHS} = \frac{1 \cdot 2 \cdot 3}{6} = 1$$

So the statement is true for $n = 1$.

Now for the second part, we want to show that

$$1^2 + 2^2 + \cdots + k^2 = \frac{k(k+1)(2k+1)}{6}$$

$$\implies 1^2 + 2^2 + \cdots + k^2 + (k+1)^2 = \frac{(k+1)(k+2)(2k+3)}{6}$$

Following the same pattern,

$$\begin{aligned}
\text{LHS} &= 1^2 + 2^2 + \cdots + k^2 + (k+1)^2 \\
&= \frac{k(k+1)(2k+1)}{6} + \frac{6(k+1)^2}{6} \\
&= \frac{(k+1)(2k^2 + k + 6k + 6)}{6} \\
&= \frac{(k+1)(k+2)(2k+3)}{6} \\
&= \text{RHS}
\end{aligned}$$

Remark. This is the broad textbook way of applying induction. However, as we'll see moving forward, a large number of problems will require a lot more thought, and often proving the induction hypothesis will not be as direct as it was in these examples.

Example 3

The Fibonacci series is defined as follows:

- $F_0 = 0, F_1 = 1.$
- For all non-negative integers $n \geq 0$ we have that

$$F_{n+2} = F_{n+1} + F_n$$

Show that

$$F_0 + F_1 + F_2 + \cdots + F_n = F_{n+2} - 1$$

Induction works very well with a series like the Fibonacci where a term is defined on the basis of previous terms, since that is exactly what induction does - reduce a certain case to the case before it.

In particular, the statement is clearly true when $n = 0$ since we get

$$0 = F_2 - 1 = 1 - 1 = 0$$

as required. Now to the inductive step, we want to show that

$$F_0 + F_1 + \cdots + F_n = F_{n+2} - 1 \implies F_0 + F_1 + \cdots + F_n + F_{n+1} = F_{n+3} - 1$$

$$\begin{aligned}
\text{LHS} &= F_0 + F_1 + F_2 + \cdots + F_n + F_{n+1} \\
&= F_{n+2} - 1 + F_{n+1} \\
&= (F_{n+1} + F_{n+2}) - 1 \\
&= F_{n+3} - 1 \\
&= \text{RHS}
\end{aligned}$$

Remark. The Fibonacci is a really beautiful series, which is known to come up in random, unexpected places - so stay on the lookout for any sequence that begins as $1, 1, 2, 3, 5, 8$.

3.3 Olympiad examples

Example 4

Every road in Sikinia is one-way. Every pair of cities is connected by exactly one direct road. Show that there exists a city which can be reached from every other city either directly or via at most one other city.

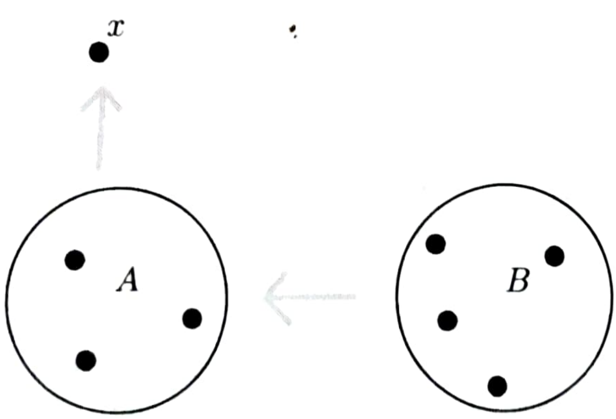
This is the kind of induction based problem that is likely to come up at an olympiad: there is absolutely nothing in the problem that suggests induction. In fact, the problem doesn't have a single math symbol!

Anyway so we have some set of cities, and every pair of cities has a one-way road connecting them.

Exercise. (Not relevant to the problem) How many roads should there be if there are n cities?

Hints: 168 398

We require some city x such that every other city is a part of one of two subsets A and B such that every city in subset A leads to city x and every city in subset B leads to at least one city in A.



The trick is to show this by induction. But... induction on what? We don't even have a variable to induct on! If you think about it, we do. The number of cities is a variable, isn't it? We show that the statement is true when there is only 1 city, and we show that if the statement is true when there are k cities, it remains true when we add in a new city.

When there is only one city, the statement is vacuously true (meaning that there is no condition we need to satisfy) - we need to show that a city exists such that *other* cities lead into it or something of that sort but there is no other city so we're all good.

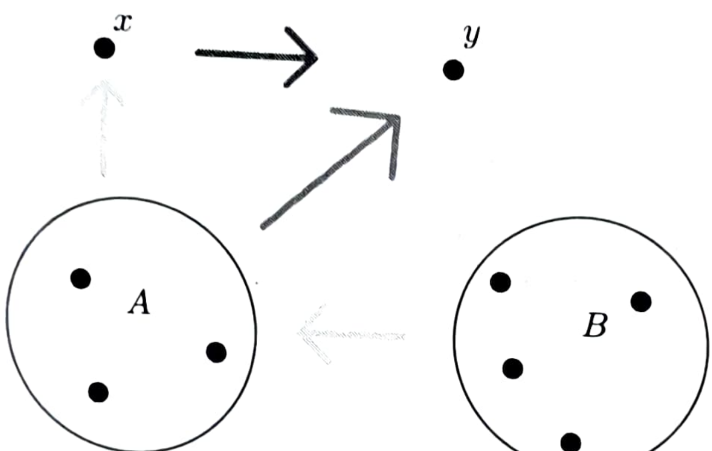
If you're not entirely convinced about this *vacuously* true part, you can try proving the statement for $n = 2$ and $n = 3$ too. They're not that hard, and often doing a couple of base cases also helps with general intuition about a problem.

So now let's try to prove the second part: if the statement is true for k cities, it is true when there are $k+1$ cities.

As of now there are k cities, and there is some required city as per the statement. And there are the groups of cities A and B so that every city in A leads into our city x and every city in B leads into at least one city in A (which leads into x).

Now we have added in a new city, y.

- If y leads to x, we just add y to A. So let's say x leads to y.
- If y leads to even one city in A, we can add y to B - so we assume y does not lead to any city in A. In other words, every city in A leads to y



At this point, we're a little stuck. How do we get to x then? If we try to go through a city in B, that'll take 3 roads which isn't allowed, and there's no guarantee if even that's possible - what if even the cities in B lead to y? Well, in that case you're just saying that everything leads to y. Shouldn't you then just make y the main city instead of x? Aha, that works! In fact, in general too - if all the cities in A lead to y and x leads to y, then we know that all the cities in B lead to a city in A, which leads to y - so we can make y the required city and everyone is happy

Example 5 (Spain 1997)

There are n identical cars on a circular track. Among all of them, they have just enough gas for one car to complete a lap. Show that there is a car that can complete a lap by collecting gas from the other cars on its way around. (Assume that there is no limit to the amount of gas the car can carry.)

This is another one of those problems where it's hard to believe that induction could be remotely useful. And yet, as it turns out it makes our job really easy.

We induct on n - so we want to show that the statement is true when there is only one car, and that if it is true for k cars then it is true for $k+1$ cars.

First off, if we have just one car, then we just take a complete lap - the problem says that we have just enough gas for that, and it must have all been in this car. (This was the only car, duh.)

Now let's try to work with the more important part: showing that if it is true for k, it is true for $k+1$. So we know that no matter how you place cars and fuels among them satisfying the constraints, we should be able to complete the lap as long as there are k cars. We want to show that this is true when there are $k+1$ cars.

So let's say we have $k+1$ cars. What now? Well, that's a good question. What *can* we do? The only thing in our hands is where we want to start. So let's say we start with the first car. Now if the fuel of this car runs out before we even reach the second car, we're in trouble. Okay, in that case - let's not pick the first car. But what if the other car you try to pick can also not reach the car after it before running out of fuel?

The idea is that if each car runs out of fuel before reaching the next stop, then the total fuel we had must have been insufficient for the total journey (why?). So there is some car which can at least allow us to reach the next station.

Figure 3.1: A scenario where none of the cars can reach the next stop. Notice that the total fuel must have been insufficient.

The magic trick is to now *merge* these two cars. If we can somehow do this properly, we'll be done because we have reduced $k+1$ cars into just k cars. So how exactly do we merge these two cars? Let's call the first car A, and the second one (which can be reached from A) B. The idea is that when you reach car A, you definitely have enough fuel to reach car B, so we might as well assume that the fuel we get when we reach B is already with us when we are at A. In particular, imagine a situation where all the fuel of car B is already present in A and car B doesn't exist.

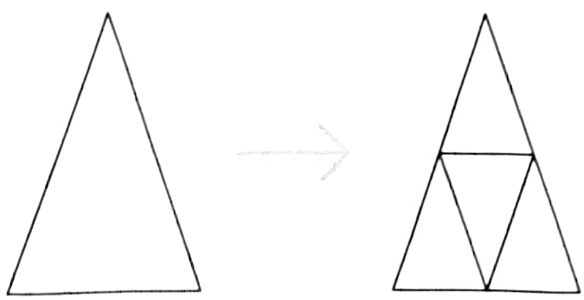
Note that we couldn't have done this for any two consecutive cars - if there isn't sufficient fuel to reach the second one this idea won't work.

Anyway, now we have only k cars, and the total fuel is just enough as required. This can be done by the induction hypothesis, and so we can un-merge these cars (how?) to end up with a valid solution for $k+1$ cars too.

Example 6

For any $n \geq 3$, prove that an equilateral triangle can be divided into n isosceles triangles.

We want to split an equilateral triangle into a lot of isosceles triangles. The idea here to induct is going to be to try and split one of these single isosceles triangles further. In particular, look at the image below:



So a single isosceles triangle can be split into four isosceles triangles by taking the 3 midpoints and joining each pair of them. In particular, if we have a solution for k, we now have a solution for $k+3$. This is slightly different from what we're used to, and there's no real way to go from k to $k+1$ here since you can't necessarily split an isosceles triangle into two such triangles. So how do we work with this weirder inductive idea?

We want to prove the statement for $n \geq 3$. So let's say we prove it for $n = 3$. Then it gets proven for $n = 3+3 = 6$, $n = 6+3 = 9$ and so on. Note that $n = 4$ is still unproven, so let's say we check that by hand as well. Then we're done for $n = 4, 7, \ldots$. Finally, we check $n = 5$ and after this we're done. A slightly more formal approach would be to take 3 cases: $n = 3k$, $n = 3k+1$ and $n = 3k+2$. If we solve the problem for $n = 3, 4, 5$ we're inductively done for each of the 3 cases, and together we're done for all of them.

Okay, so we essentially just want to find a construction for $n = 3$, $n = 4$ and $n = 5$.

How do we do that? Well, that's for you to figure out!

Problem 7. Find constructions for $n = 3, 4, 5$.

Soln: Page 273, Solution 5

3.4 Recurrence Relations

Recurrence relations create a powerful way for us to break complicated problems into parts and help us analyze structures more easily. The idea is best understood with an example, so here we go!

Example 8

You are given a $2 \times n$ grid which needs to be tiled with 2×1 or 1×2 pieces known as dominoes. How many ways are there to complete this tiling?

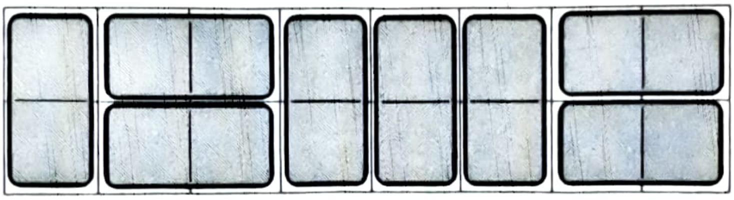


Figure 3.2: An example tiling

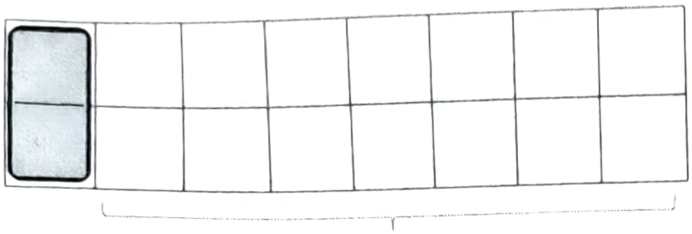
So we are given a $2 \times n$ grid, and we want to tile it with these sorts of dominoes as above. We are asked to find the number of ways of doing so. Let's try and find the value for a couple of small values of n.

Exercise. Evaluate by hand and verify that the answers below are correct:

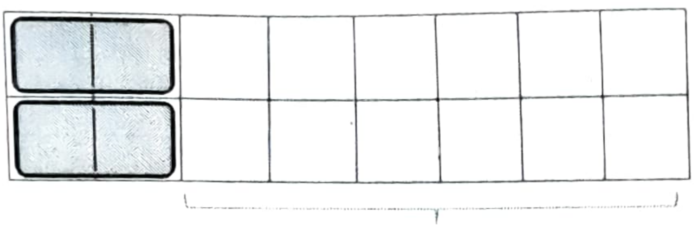
- $n = 1 \to 1$ way
- $n = 2 \to 2$ ways
- $n = 3 \to 3$ ways
- $n = 4 \to 5$ ways

Hmm $1, 2, 3, 5$. Anything that reminds us of? The Fibonacci! We know that the Fibonacci is defined by $F_n = F_{n-1} + F_{n-2}$ so the guess is that we want to somehow relate the number of possibilities for a $2 \times n$ to the number of possibilities for a $2 \times (n-1)$ and a $2 \times (n-2)$. Let's try and figure out what's going on. Let the answer for a $2 \times n$ be A_n.

If we place a domino vertically in the leftmost column, we are now tasked with solving A_{n-1} since the leftmost column is done. On the other hand, if we put a horizontal domino at the top left, we'll have to put another horizontal domino right below (how else do I cover the bottom left square?) and so now I need to solve A_{n-2}.



A_{n-1}



A_{n-2}

Using the technique from the first chapter: we need to do exactly one: the first OR the second, so the number of ways is

$$A_{n-1} + A_{n-2}$$

So in fact this satisfies the same recursion as the Fibonacci, and $A_1 = 1$, $A_2 = 2$. This means that $A_n = F_{n+1}$ ($F_1 = F_2 = 1$) and so the number of possibilities is the $(n+1)$th Fibonacci number.

> **Remark.** If you want an explicit form for the Fibonacci, see the problems section!

Let's look back a little at what we did here. The idea was to look at how things are working at the start/end of the sequence, and reduce the problem to something slightly simpler.

In particular, whenever you see a n somewhere in a question, or a big number that seems too hard to work with - it's useful to try and understand the end of the sequence and try to break it off.

Example 9

An elf has a staircase of n stairs to climb. Each step it takes can cover either one or two stairs. Find a recurrence relation for a_n, the number of different ways for the elf to ascend the n-stair staircase.

This is quite direct, so let's quickly give it a try.

The trick, as with most such recurrence relation problems is going to be - "what happens at the start?"

Well, at the start, the first move either involves the elf moving one stair or the elf moving two stairs. In the first case, it has $n-1$ more stairs to cover, and a_{n-1} ways to do this, and in the second case it has $n-2$ stairs to cover and a_{n-2} ways to do this.

By the addition principle we can add them up, and so

$$a_n = a_{n-1} + a_{n-2}$$

So we have the Fibonacci recurrence, once again.

Let's now try a slightly tricky application of recurrence relations in the next problem.

Example 10

Find the number of derangements of n objects. (A derangement is a permutation of the elements of a set in which no element appears in its original position.)

As an example, 23154 is a derangement, but 21435 isn't because 5 appears at the fifth position.

So the idea is going to be that we let the number of such derangements be D_n, and relate D_n to things like D_{n-1} and D_{n-2} based on some cases.

So to start somewhere, let's say the number k appears at the first position. Now consider the subsequence from 2 to n. The idea is that this is "almost" a derangement as well.

To make this more clear, take the example 23154, and think of the second element to the last one - 3154. If you consider the original set as $\{2,3,4,5\}$, none of the elements appear at their original positions (with the exception of 2, which doesn't appear at all).

So what if we just swap 1 and 2 in the sequence 3154 to get 3254? Aha, now it's perfect - this is a derangement of $\{2,3,4,5\}$! So we can relate D_n to D_{n-1}

since each such valid derangement of $n-1$ numbers creates a derangement for n numbers.

Except... there's still one issue - we know that $3, 4, 5$ won't appear at their original positions, but what if 2 does? As an example, what if we had something like 21534? Now when you consider the elements 1534, and then do the swap 2534, 2 ends up at its original position. The reason this happened was that 1 had gone to 2's position, so when you swapped them 2 ended up at its normal position. But luckily, note that the remaining three elements form a derangement anyway.

So to put what we have together,

- Let's say 2 occurs as the first number.
- If the second number is 2, then the number of solutions is D_{n-2}, since the 3rd, 4th, ..., nth element form a derangement of the set $\{3, 4, \ldots, n\}$ and each such derangement leads to a solution.
- If the second number is not 2, then after swapping 1 and 2, we have that the 2nd, 3rd, ..., nth element form a derangement of $\{2, 3, 4, \ldots, n\}$ and once again each such derangement leads to a valid solution.

In particular, we could have any of the remaining $n-1$ numbers at the first position and the same argument works for each of them, so

$$D_n = (n-1)(D_{n-1} + D_{n-2})$$

Example ~~11~~

Find the number of n digit sequences a_n composed of 0's, 1's and 2's, such that an even number of 0's and an even number of 1's appear.

This problem shows us another very useful idea while working with recurrence relations

> You might need to capture more information than the problem asked you for.

Think of it like this. If the first digit is a 2, then we can just say that we have a_{n-1} cases - since the remaining $n-1$ digits also have an even number of 0's and 1's.

The issue is that the first number might be 0 for example, and in that case you can't easily relate this to a_{n-1} as the remaining $n-1$ numbers must have an odd number of 0's and an even number of 1's.

So our trick is going to be to capture this information as well - consider b_n to be the number of n digit sequences with an *odd* number of 0's and an even number of

1's. To complete the set, we define c_n as sequences with even 0's and odd 1's, and d_n as sequences with an odd number of 0's and an odd number of 1's. Now we try to create as many equations as we can.

- $a_n = a_{n-1} + b_{n-1} + c_{n-1}$ (the first case if 2 is the first digit, second if 0 is, and third if 1 is)
- $b_n = a_{n-1} + b_{n-1} + d_{n-1}$ (the first case if 0 is the first digit, second if 2 is, and third if 1 is)

  In a similar fashion, we get

- $c_n = a_{n-1} + c_{n-1} + d_{n-1}$
- $d_n = b_{n-1} + c_{n-1} + d_{n-1}$

If we know $a_{n-1}, b_{n-1}, c_{n-1}, d_{n-1}$ we can use this to get each of a_n, b_n, c_n, d_n. Knowing the base cases, where $a_1 = b_1 = c_1 = 1$ and $d_1 = 0$, we can build up to get everything we need.

Remark. Getting an explicit form is tricky here, but can be done - to reduce the work you can try showing that $b_n = c_n$ and $a_n + b_n + c_n + d_n = 3^n$, for starters. After that we have only two variables, and it boils down to some computation.

3.5 Problems

Problem 1. Use induction to show that each of the following equations is true:

1. $1^3 + 2^3 + \cdots + n^3 = \left(\frac{n(n+1)}{2}\right)^2$

   Hints: 295

2. $\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = 2^n$

   Hints: 479 413

Problem 2. Give an inductive proof to show that the sum of the interior angles of a convex polygon with n sides is $(n-2) \times 180°$. (You may assume that the result is true for a triangle.)

Hints: 95

Problem 3. Prove that a convex n-gon has $\frac{n(n-3)}{2}$ diagonals.

Hints: 111

Problem 4. Let F_n denote the nth Fibonacci number, defined by

$$F_n = \begin{cases} 0 & n = 0 \\ 1 & n = 1 \\ F_{n-1} + F_{n-2} & n \geq 2 \end{cases}$$

Prove the following for all $m, n \geq 0$:

$$F_{m+n+1} = F_m F_n + F_{m+1} F_{n+1}$$

Hints: 114

Problem 5. Using the previous problem, conclude that F_{kn} is divisible by F_n for all positive integers n, k.

Hints: 399 19

Problem 6. Let $a = \frac{1+\sqrt{5}}{2}$ and $b = \frac{1-\sqrt{5}}{2}$ be roots of the equation $x^2 - x - 1 = 0$. Prove by induction that that for the Fibonacci series,

$$F_n = \frac{a^n - b^n}{\sqrt{5}}$$

Hints: 303 515

Problem 7. Prove that a square can dissected into n squares for $n \geq 6$.

Hints: 154 182

Idk for n=7
rest ✓✓

Problem 8 (2015 IMOSL C1). In Lineland there are $n \geq 1$ towns, arranged along a road running from left to right. Each town has a left bulldozer (put to the left of the town and facing left) and a right bulldozer (put to the right of the town and facing right). The sizes of the $2n$ bulldozers are distinct. Every time when a left and right bulldozer confront each other, the larger bulldozer pushes the smaller one off the road. On the other hand, bulldozers are quite unprotected at their rears; so, if a bulldozer reaches the rear-end of another one, the first one pushes the second one off the road, regardless of their sizes.

Let A and B be two towns, with B to the right of A. We say that town A can sweep town B away if the right bulldozer of A can move over to B pushing off all bulldozers it meets. Similarly town B can sweep town A away if the left bulldozer of B can move over to A pushing off all bulldozers of all towns on its way.

Prove that there is exactly one town that cannot be swept away by any other one.

Hints: 243 471

Problem 9 (IMO 2000/4). A magician has one hundred cards numbered 1 to 100. He puts them into three boxes, a red one, a white one and a blue one, so that each box contains at least one card. A member of the audience draws two cards from two different boxes and announces the sum of numbers on those cards. Given this information, the magician locates the box from which no card has been drawn. How many ways are there to put the cards in the three boxes so that the trick works?

Hints: 434 524 **Soln:** Page 273, Solution 6

Problem 10 (2015 INMO/4). There are four basketball players A, B, C, D. Initially the ball is with A. The ball is always passed from one person to a different person. In how many ways can the ball come back to A after seven moves? (for example $A \to C \to B \to D \to A \to B \to C \to A$, or $A \to D \to A \to D \to C \to A \to B \to A$).

Hints: 538 195 641

Problem 11. Write a recurrence relation for a_n, the number of n digit binary sequences with at least one instance of consecutive 0s.

Hints: 277 239 592

Problem 12. Find a recurrence relation for a_n, the number of n-digit ternary sequences without any occurrence of the subsequence "012". (A ternary sequence is a sequence composed of 0s, 1s, and 2s.)

Hints: 351

II

Techniques for Problem Solving

4 Invariants and Monovariants

4.1 Introduction

The quote I follow here is:

> "When too much is changing, look for what doesn't change"

This is a pretty useful life quote too, the people who support us even when everything around us is changing are the ones that should really matter to us. But philosophy aside, what does the quote even mean?

Often, in the olympiads, we are faced with math problems that give us a certain starting position, and allow us to do a certain set of operations on that position. (We'll take some nice examples of this in a little bit, this is just for you to get the gist of it.) Let's say you want to show that no matter how you perform those operations, you just cannot reach a certain ending position. Of course, when you're doing those operations, a lot of things are probably changing in the position. But if we can find some quantity that does not change (or changes in a somewhat predictable way) no matter which operation you choose to perform, we can compare the quantity at the starting and ending to (hopefully) derive a contradiction. This *quantity* that does not change is usually turned an invariant, and if it changes in a *predictable way*, i.e. always increases or always reduces - we term it a monovariant.

But what really do we mean by a *quantity*? This isn't physics, where you could be dealing with physical quantities, right? Well, let's see a few examples of what this *quantity* could actually be.

Example 1

The numbers $1, 2, 3, 4, 5$ are written on a blackboard. In an operation, you can choose any two numbers a and b currently on the board, erase them, and write the numbers x and $a + b - x$ on the board for any x of your choice. Show that no matter what you do, you cannot reach a position where the numbers written on the board are

(a) $3, 3, 4, 5$

(b) $2, 3, 4, 5, 6$

Just so that you get the idea, a possible move could be to erase 3 and 5 from the board and replace them with 2 and $3 + 5 - 2 = 6$ so that the board now has the numbers $1, 2, 2, 4, 6$ written on it.

We need to show that no matter what we do, we cannot reach each of the two given positions. So let's see what doesn't change when the two numbers are replaced with these other two numbers. Well, as a pretty basic observation, we notice that we're erasing two numbers, and adding two numbers, so the number of values on the board should always remain constant.

Aha! We have already got an invariant - the number of numbers written on the board always remains constant. Since the number of values is 5 at the start and in the first subpart there are only four numbers, there is no way we could have reached it.

But the second part has five numbers too - so it seems like this invariant is not too useful over there. This can happen often, not every invariant you find will be useful so you must stay on the lookout for these *quantities that don't change.*

Anyway, we're replacing a, b with $x, a + b - x$. The important observation is that the sum of these two pairs is the same: both sum to $a + b$. So no matter what we do, we can't really change the total sum of numbers on the board. And there you go, we have it! A second invariant - the sum of all numbers. Since it is $1 + 2 + 3 + 4 + 5 = 15$ at the start but $2 + 3 + 4 + 5 + 6 = 20$ at the end, there is no way we could have reached that final position.

This is the general idea of how invariants work - you find these quantities that don't change - compare them at the start and end and then either celebrate or return to square one.

But let's take a step back, and consider what I said earlier.

> "This isn't physics, where you could be dealing with physical quantities, right?"

It turns out, you could. Invariants exist in physics too, and across many other divisions of science.

If you've ever been introduced to momentum, someone would've told you that it is the product of the mass and velocity of an object. But what does this quantity *physically* mean? Temperature tells us if something is hot or cold, and velocity tells us if something is fast or slow. What does momentum tell us? There are probably ways you can justify what momentum is by saying it's something like *how easily* you can change its motion or something like that, but this is not satisfying enough. If you've ever learnt the *Law of conservation of momentum*, it states that the momentum of a system is constant if no external forces are acting on the system. Woah! That's quite lucky! Well actually, it isn't. It wasn't just luck that scientists invented a random quantity and it turned out to follow this law. The entire quantity was made *because* they realised that it was an *invariant* under the condition of no external force (!!)

Elsewhere, entropy is a monovariant, while equivalent weights in chemistry are an invariant. The point I'm trying to put across is that the concept of invariant, and in general a lot of concepts in mathematics are widely used in various sciences and other disciplines - because at its heart, math is all about solving various kinds of problems and sciences deal with solving such problems as well.

But enough science. Let's get back to math!

4.2 Introductory Examples

Example 2

A sequence of $+$ and $-$ signs has been written on the board as follows:

$$\{+, +, -, +, -, -, -, +, -\}$$

In a move, we're allowed to choose any two of the signs on the board and erase them. If both signs we removed were the same, we add back a $+$, and if the signs we removed were different, we add back a $-$. Find all possibilities for the final sign.

Just as an example, if we pair up $+$ and $-$ we need to add back a $-$ and if we remove two $-$ signs, we need to add in a $+$. Each time we do this, the number of signs reduces by one, so eventually we would be left with exactly one sign. The question asks us to find which sign this is.

The important observation here is that this is kind of exactly how multiplication works. We all know that $-1 \times -1 = +1$, $+1 \times -1 = -1$ and so on. So if we think of $+$ as $+1$ and $-$ as -1, we're essentially erasing two numbers and replacing them with their product. From here, it's quite clear that the product of all numbers should be invariant. Indeed, the total product was

$$(\text{product of the 2 terms we chose}) \times (\text{product of the remaining } n-2 \text{ terms})$$

before as well as after the operation. In particular note that the final number should then just be the product of the original set of numbers (which is -1 for the given set of numbers) and so no matter what we do, we would have to end up with a $-$ at the end.

Example 3

The set of integers $1, 2, \ldots, 100$ is written on a board. In a move, we can replace a and b with the single value $ab + a + b$. Find all possibilities for the final number.

So we're erasing a and b, but adding back $ab + a + b$. Had we been given just adding back ab, our life would've been perfect - we could've just said that the product of all numbers remains the same. However, we now have to tweak this a little. A very cool identity which comes up a lot in Number theory, as well as combinatorics, is the following (popularly known as Titu's favorite factoring trick).

$$ab + a + b + 1 = (a + 1)(b + 1)$$

Aha, so if we consider each term $+1$, and then take the product, the product went from $(a + 1)(b + 1)(\text{other terms})$ to $((ab + a + b) + 1)(\text{other terms})$. Since these two are the same, this product remains the same! At the start it is

$$(1 + 1)(2 + 1) \ldots (100 + 1) = 101!$$

Thus when only one number N is left, we should have $N + 1 = 101!$ so we have that $N = 101! - 1$, yay!

Example 4

The set of integers $1, 2, \ldots, 100$ is written on a board. In a move, we can replace a and b by the pair of values $0.8a + 0.6b, 0.6a - 0.8b$. Can we reach the set $2, 3, \ldots, 100, 101$ by performing only this operation?

The remarkable thing about this problem is that at first glance 0.8 and 0.6 look perfectly random, but as it turns out, they're really special numbers. Perhaps putting them into fractions could give you the answer: $\frac{4}{5}$ and $\frac{3}{5}$. Hmm.

What is the most famous thing that involves the numbers $3, 4, 5$? Of course. $3^2 + 4^2 = 5^2$, its a Pythagorean triplet! But... how is that relevant? Well, we know that $(\frac{3}{5})^2 + (\frac{4}{5})^2 = 1$, so ...

Problem. Figure out a way to use this neat fact about 0.6 and 0.8 to find the invariant and solve the problem.

Hints: 45 118

4.3 Parity

Lemma 5 (Odd and even numbers)

The sum of two odd numbers is even, the sum of two even numbers is even and the sum of an odd number with an even number is odd.

The somewhat trivial idea of numbers being odd and even is really useful as an invariant - often certain numbers don't change their *parity* and we can take advantage of this.

In what follows, we consider the notation

$$x \equiv 0 \text{ or } 1 \pmod 2$$

The expression $a \equiv b \pmod m$ means that the difference between a and b is a multiple of m. So when m is 2, $a \equiv 0 \pmod 2$ essentially means that a is even (since the difference between a and 0 is divisible by 2) and $a \equiv 1 \pmod 2$ means that a is odd.

Example 6 (RMO 2016)

A box contains answer 4032 scripts out of which exactly half have an odd number of marks. We choose 2 scripts randomly and, if the scores on both of them are odd numbers, we add one mark to one of them, put the script back in the box, and keep the other script outside. If both scripts have even scores, we put back one of the scripts and keep the other outside. If there is one script with even score and the other with an odd score, we put back the script with the odd score and keep the other script outside. After following this procedure a number of times, there are 3 scripts left among which there is at least one script each with odd and even scores. Find, with proof, the number of scripts with odd scores among the three left.

This problem feels quite messy when you read it at first but if you just simplify the information you've been given a little, it turns out that the problem isn't so hard at all.

$$\text{(odd, even)} \to \text{odd}$$
$$\text{(even, odd)} \to \text{odd}$$
$$\text{(even, even)} \to \text{even}$$
$$\text{(odd, odd)} \to \text{even}$$

(For the last case note that adding 1 mark turned it into an even score.)

What happens to the number of odds and evens in each case?

Case 1 : odd: no change	even: -1
Case 2 : odd: no change	even: -1
Case 3 : odd: no change	even: -1
Case 4 : odd: -2	even: +1

For a second I almost believed that the number of odds just doesn't change. But thankfully we can use parity!! Notice that a -2 is really no change, if we look modulo 2. So the number of odds never changes parity. At the start, there were 2016 of them - so at the end, there should be zero or two. The question says that there is at least one script of each type so this forces there to be exactly two sheets with odd scores among the three left!

Remark. Noting down information that a problem gives you in neat ways is key to finding an invariant, as our mind can pick up patterns from short and neat data with ease but it would struggle to do the same if you were trying to find the invariant directly from the problem text.

Example 7 (Classic)

You are given 100 tokens of type A, 102 tokens of type B and 104 tokens of type C. You are allowed to perform the following operation on the tokens:

> Choose 2 chips of different types and replace them with a single chip of the third type.

Can you reach a configuration where only one token remains?

Consider the tuple (a, b, c). The allowed algorithm is basically to subtract 1 from two of the numbers here and add 1 to the third. We want to end up at one of $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$.

The three steps we're allowed to perform are

$$(a, b, c) \to (a + 1, b - 1, c - 1)$$
$$(a, b, c) \to (a - 1, b + 1, c - 1)$$
$$(a, b, c) \to (a - 1, b - 1, c + 1)$$

The key observation is that $+1$ and -1 can actually become the same thing - if we consider the process modulo 2, and notice that after a single move,

$$(a, b, c) \to (a + 1, b + 1, c + 1)$$

(as $c - 1 \equiv c + 1 \pmod 2$)) Aha! so the only possible ending states are (a, b, c) and $(a+1, b+1, c+1) \pmod 2$ (why?).

Note that we started at $(100, 102, 104)$ which is $(0, 0, 0) \pmod 2$. So we can only finish at a position that is $(0, 0, 0)$ or $(1, 1, 1)$ mod 2, so ending at one of $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$ is impossible.

> **Remark.** I like to think of the above as the "$\pm$ trick". Whenever you see a mix of pluses and minuses, but you want to view them the same way, it is very useful to think of modulo 2 since $+a \equiv -a \pmod 2$

Example 8

We begin with the numbers $1, 2, \ldots 2023$ written on the blackboard. At each step, we can erase any two of the numbers a and b and then write down the number $|a - b|$. We continue until one number remains. Determine whether this final number could be equal to 1.

Let's imagine the question instead said that we need to write down $a + b$ instead of a and b. Well, now the question is pretty easy - the sum of the numbers remains invariant, and so we can easily determine the final answer as $1 + 2 + \cdots + 2023$. The issue is that we don't have $a + b$ here, we have $|a - b|$. But is that really an issue? $\pm$ trick to the rescue! If we consider everything $\pmod 2$, we have that (check this)

$$|a - b| \equiv a + b \pmod 2$$

and so the sum of numbers mod 2 remains invariant.

Exercise. Show that as a result of this invariant, our final number couldn't have been 1.

Soln: Page 274, Solution 7

> **Remark.** A key thing to notice here is that if the question had asked us "Can we reach 2?", our method would be a complete disaster. In fact, invariants like these only tell us when a certain thing is unreachable. In some cases, we can in fact reach the final number required. This is something we deal with in the chapter of algorithms.

4.4 Working modulo m

Till now, we've been working mod 2 everywhere. As it turns out, there are a lot of problems where using mod 2 isn't very useful at all, but a different mod can help us solve the problem. But how do we figure out which m is going to be useful to us? Let's learn through a couple of examples.

Example 9

A room is initially empty. Every minute, either one person enters or three people leave. After exactly 2023 minutes, could the room contain exactly 201 people?

So after each minute, we have a -3 or $+1$. The idea is that we want these to be the same modulo m, i.e.

$$1 \equiv -3 \pmod{m}$$

and so $m = 4$ makes the most sense.

Exercise. Finish from here by showing that the number of people after 2023 minutes should be $3 \pmod 4$)

Soln: Page 275, Solution 8

Example 10

You are given 100 tokens of type A, 101 tokens of type B and 102 tokens of type C. You are allowed to perform the following operation on the tokens:

> Choose 2 chips of different types and replace them with **two** chips, both of the third type.

Can you reach a configuration where all tokens are of the same type?

The three steps we're allowed to perform are

$$(a,b,c) \to (a+2, b-1, c-1)$$
$$(a,b,c) \to (a-1, b+2, c-1)$$
$$(a,b,c) \to (a-1, b-1, c+2)$$

We want to reach $(303, 0, 0)$ or one of its permutations. Last time, we motivated mod 2 by noticing that we wanted $+1 \equiv -1 \pmod{m}$. This time, we need $+2 \equiv -1 \pmod{m}$. This motivates setting $m = 3$. So each time, we go from $(a, b, c) \to (a-1, b-1, c-1) \pmod 3$.

Exercise. Show that the desired final position can never be reached from the given initial one.

Hints: 290 83

4.5 Monovariants

Let's say you want to show that after a certain sequence of operations, you can *never* return to the original position. Well, you can try all you want to find an invariant but it's just not going to be too useful because the starting and ending positions you would have to compare after finding the invariant are literally the same. On the other hand, what if you can show that a certain quantity always increases or reduces (strictly)? Well, then you can never increase the quantity again and again and end up at the same value (which would have happened had you reached the same position). Such a quantity that always increases or always reduces is known as a monovariant.

Another place where monovariants are seriously useful is if you want to show that eventually, you won't be able to perform any more operations. In this case, you can argue that the quantity you found as your monovariant actually has a maximum value or minimum value. So if it increases by at least 1 each time (usually your monovariant will only take positive integer values) and the maximum possible value of the quantity is some B then you can perform the operations at most B times and so the process must end. This is an idea we will re-encounter when studying greedy algorithms later, where we need to show that the operation we came up with can be applied only finitely many times.

An idea in invariants/monovariants that comes up a lot is interpreting numbers in terms of some sort of binary representation. If you can show that the number decreases/increases on every step, this would mean that the process has to end - similar to the monovariants idea we saw earlier.

Let's start with a simple example.

Example 11

The numbers $1, 2, 3, 4, 5$ are written on a blackboard. In an operation, you can choose any two non-zero numbers a and b currently on the board, erase them, and write the numbers x and $a + b - x - 1$ on the board for any x of your choice under the condition that the two numbers you put should be non-negative.

(a) Is it possible to perform this operation one or more times and end up at $1, 2, 3, 4, 5$ again?

(b) Can the process go on forever?

The good part about this chapter is that you already know the answers to the questions - *No, it's not possible.* This is a liberty you lose during the later chapters

- and it's important to realise how significant being able to guess the answer can be while attempting a problem. Imagine looking for an invariant or monovariant for hours only to be told that there was a way to make the process go on forever.

Anyway, this is almost identical to the original problem we had - and there we had the sum of all numbers constant. This time the sum of numbers goes from $a + b +$ (other stuff) to $a + b - 1 +$ (other stuff), i.e. it reduces by one. And there we have it - a monovariant! After k moves, the total sum would've reduced by k, and so we can never reach back the original position after $k \geq 1$ moves.

The problem statement says that the numbers we must create should be non-negative, so the sum of numbers remains non-negative, and so we can decrease it by 1 at most $1 + 2 + 3 + 4 + 5 = 15$ times. At this point, all the numbers become zero, and so we cannot pick two non-zero numbers to operate on!

4.6 Weighting

Example 12

We place n stones at certain integer positions on a number line. In a move, if a certain position has at least two stones, we move one stone to the left and one stone to the right. Is it possible to perform a non-zero number of moves end up at the same position as the one we started off with?

Well, the first observation is that the number of stones is invariant - but that isn't too useful, is it? As we've seen - an invariant can never tell us whether we can end up at the same position - we're going to need a monovariant.

The key idea here is to assign *weights*. In particular, we label the number line from $-\infty$ to ∞ so that the coins are placed at some positions. Now think of it like this - if your coin is at position 100, we value it a lot more than a coin at position 2 (no idea why, but we just do). So we assign a weight w_i to the coins at position i, and take the weighted sum of the coins

$$\sum_i w_i x_i$$

where x_i is the number of coins currently at position i. What happens when we make a move? Well, 2 coins are removed from i so the sum reduces by $w_i + w_i$. But it also increases by $w_{i+1} + w_{i-1}$.

The cool trick is that if we can create weights so that the expression $w_{i+1} + w_{i-1} - 2w_i$ is positive for each i, then we're only increasing that weighted summation - and so it should be impossible to come back to the same position. Showing that such weights exist is easy, just set $w_{i+1} = 2w_i - w_{i-1} + 1$ or something recursively, and then the weights all work out. If you want explicit constructions, note that $w_i = i^2$ and $w_i = 2^i$ work.

Example 13 (IMOSL 2009 C1)

Consider 2009 cards, each having one gold side and one black side, lying on parallel on a long table. Initially, all cards show their gold sides. Two players, standing by the same long side of the table, play a game with alternating moves. Each move consists of choosing a block of 50 consecutive cards, the leftmost of which is showing gold, and turning them all over, so those which showed gold now show black and vice versa. The last player who can make a legal move wins.

(a) Does the game necessarily end?

(b) Does there exist a winning strategy for the starting player?

For now we'll concern ourselves with only part (a). Part (b) is also a very interesting problem and is quite reasonable so I encourage you to give it a try!

Once again, we want to prove that some process cannot go on forever. Hmm. Here comes an idea that is really useful whenever you see cards/coins/markers having two sides: think *binary*. The idea is that you can think of gold as a 0 and black as a 1. So initially you have a long binary string that reads 00000...00. After each move, you're flipping 50 of the bits, with the constraint that the leftmost bit you flip is a 0.

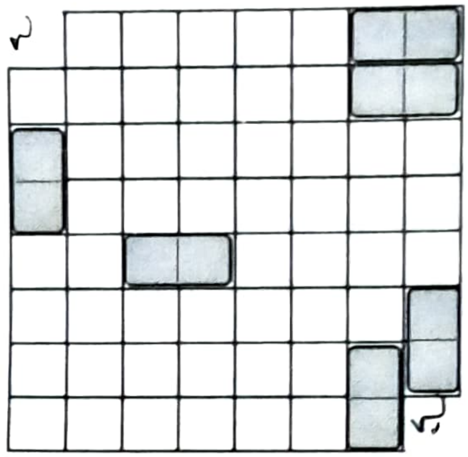
As it turns out, binary is really special. The idea is that the value of each bit is more than the value of all previous bits combined. Mathematically, I mean $32 > 16 + 8 + 4 + 2 + 1$, which means that the value of flipping the 32 bit can be more than the value of flipping everything to the right of it however you wish. In particular, here the leftmost bit we flip is a 0, and it has become a 1. This creates a $+2^n$. Even if everything else to its right goes from 1 to 0, we'd have a total subtraction of $2^{n-1} + 2^{n-2} + \cdots + 2 + 1$, but 2^n is still bigger than this, so the number represented by the binary string must have increased! It's easy to finish from here - as the length of the binary string is fixed at 2009, it can peak at a maximum value of $2^{2009} + 2^{2008} + \cdots + 2 + 1$ and so at that point, the process must end.

We could also motivate this entire method using the same weights method as last time, but I felt that binary is something that comes up a lot so it's useful to think of binary straight away, rather than end up at it by chance.

4.7 Colourings

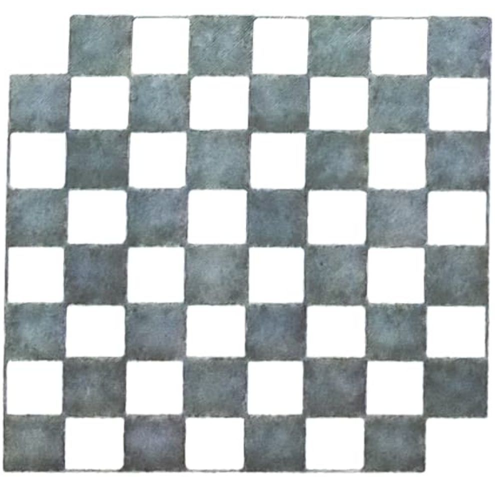
Example 14

Two opposite corners are removed from a standard chessboard. Is it possible to tile the resulting board with dominoes?



We'll show that it is impossible to tile this chessboard. To do so, we use this very powerful trick: *colouring the board*! In particular, we color the cells of the board with a certain number of colors, so that no matter how you place a domino, you occupy one vertex of each color.

In fact, achieving such a coloring is hardly difficult. Take a chessboard coloring - so color alternate cells in each row white and black just like a chessboard. (In particular you can also think of this as coloring the cells (x, y) such that $x + y \equiv 0 \pmod 2$ black and leaving the rest as white.)



But... what was the point? The point is that any domino you place uses one cell of each color. So if the entire board is tiled, we must have used 31 dominoes, and in particular, there should be 31 cells of each color. But as it turns out, there are 32 of one color and 30 of the other because both the cells deleted are of the same color (check this!).

So in general, when working with colorings, we're going to be looking at trying to color the grid in a way that there is some sort of *invariance* regardless of how you place your domino/object on the board.

Example 15

Determine whether it is possible to tile a 10×10 square floor using 1×4 rectangular tiles.

The first thought is to color it the same way, because each 1×4 tile will use two of each color. The issue is that there are in fact 50 squares of each color, so we've hardly got a contradiction. Instead, let's look at how we had defined our coloring:

> Color cell (x, y) black if $x + y \equiv 0 \pmod 2$ and white if $x + y \equiv 1 \pmod 2$.

See a way to extend this for when we have a 1×4 tile?

Problem. Find a new coloring, and show that there aren't an equal number of cells of each color (sorry but you'll just have to bash this out!).

Hints: 256 **Soln:** Page 275, Solution 9

1 2 3 0 1 2 3

0 1 2 3

2 3 0

4.8 Problems

Problem 1. Find the minimum number of breaks required to break an $m \times n$ bar of chocolate into 1×1 squares.

Hints: 244

Problem 2. A zombie plague is infecting a town. Each night, the townspeople will meet randomly in groups of three. If at least one of the members of the group is a zombie, all others will become zombies at the end of the night. Given that the infection starts in one random individual, and that the town has a population of 1000, can there be exactly 100 affected individuals after 10 nights?

Hints: 300 453

Problem 3 (Classic). Initially, 4 chips are placed at the point (0, 0). At each step we can remove one chip from some point (a, b) and replace it with 2 chips, one at the point $(a + 1, b)$ and the other at $(a, b + 1)$. Show that, after finitely many steps, there will always be some point with at least two chips sitting on it.

Hints: 217 550 507 432

Problem 4 (2014 ISL C2). We have 2m sheets of paper, with the number 1 written on each of them. We perform the following operation. In every step we choose two distinct sheets; if the numbers on the two sheets are a and b, then we erase these numbers and write the number $a + b$ on both sheets. Prove that after $m2^{m-1}$ steps, the sum of the numbers on all the sheets is at least 4^m.

Hints: 214 26

> **Remark** (Small hint). You might need the AM-GM inequality to solve this.
>
> $$\frac{x_1 + x_2 + \cdots + x_n}{n} \geq (x_1 x_2 \ldots x_n)^{1/n}$$

Problem 5 (Classic). The numbers $1, 2, 3, \ldots, n$ are written in a row. It is permitted to swap any two numbers. If 2023 such operations are performed, is it possible that the final arrangement of numbers coincides with the original?

Hints: 368 286 102 442

Problem 6 (2012 ISL C1). Several positive integers are written in a row. Iteratively, Alice chooses two adjacent numbers x and y such that $x > y$ and x is to the left of y, and replaces the pair (x, y) by either $(y + 1, x)$ or $(x - 1, x)$. Prove that she can perform only finitely many such iterations.

Hints: 375 289

Problem 7 (Russia 1998/9.8). Let the distinct positive integers a, b be written on a board. Every minute, the smaller number is erased and replaced with

$$\frac{ab}{|a-b|}$$

Prove that, eventually the two numbers are equal (and so the process must end, otherwise we would have to divide by 0 next). What is the final pair of numbers?

Hints: 40 82 331

Problem 8. An 8×8 chessboard is colored in the usual way, but that's boring, so you decide to fix this. You can take any row, column, or 2×2 square, and reverse the colors inside it, switching black to white and white to black. Prove that it's impossible to end up with 63 white squares and 1 black square.

Hints: 587

Problem 9 (IMO 1993/3, simpler version). The following game is played on an infinite chessboard. Initially, each cell of an $3n \times 3n$ grid is occupied by a single chip. A *move* consists of a jump over a chip in the horizontal or vertical direction on to a cell directly behind it. The chip jumped over is removed. Show that we can never reach a position where exactly one chip remains.

Hints: 505 457 236

Problem 10. Initially, $n-1$ of the n^2 squares in a $n \times n$ grid are infected. During each unit time interval, each square which has 2 or more infected neighbours (a neighbour being a square which shares an edge) also becomes infected. Determine whether it is possible that all n^2 squares will eventually become infected.

Hints: 585 386 466

Problem 11 (MOP 1998). On an infinite (in both directions) strip of squares, indexed by the integers, are placed several stones (more than one may be placed on a single square). We perform a sequence of moves, each move being one of the following types:

- Remove one stone from each of the squares $n-1$ and n and place one stone on square $n+1$. -1
- Remove two stones from square n and place one stone on each of the squares $n+1, n-2$. Same

Prove that any sequence of such moves will lead to a position in which no further moves can be made.

Hints: 401 537 255 199 649 146

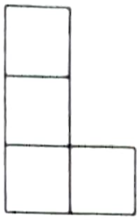
Problem 12 (RMM SL 2016). Start with any finite list of distinct positive integers. We may replace any pair $n, n+1$ (not necessarily adjacent in the list) by the single integer $n-2$, now allowing negatives and repeats in the list. We may also replace any pair $n, n+4$ by $n-1$. We may repeat these operations as many times as we wish. What is the most negative integer which can appear in a list?

Hints: 428 32 119

Problem 13. The numbers $1, 2, \ldots, 2008$ are written on a blackboard. Every second, Jimmy erases four numbers of the form $a, b, c, a+b+c$, and replaces them with the numbers $a+b, b+c, c+a$. Prove that this can continue for at most 10 minutes.

Hints: 227 164 269 22

Problem 14 (Pablo Soberon). Is it possible to cover a 10×10 board with the L shaped as below without them overlapping? (The pieces can be flipped and turned.)



Hints: 427 581 470

Problem 15 (INMO 2020/6). A stromino is a 3×1 rectangle. Show that a 5×5 board divided into twenty-five 1×1 squares cannot be covered by 16 strominos such that each stromino covers exactly three squares of the board, and every square is covered by one or two strominos. (A stromino can be placed either horizontally or vertically on the board.)

Hints: 443 304 218 631 409

5 Combinatorial Games

5.1 Introduction

In general, many combinatorial game problems in olympiads look like the following:

> **Example**
>
> Two players, Alice and Bob (in favor of being able to call them A and B), play the following game. They are given ... and:
>
> - Alice, on her move, is allowed to perform the following operations: ...
> - Bob, on his move, is allowed to perform the following operations: ...
>
> The players alternate moves with Alice going first, and whoever cannot make a move loses. Find a winning strategy for one of the players.

Most problems set the operations each player is allowed to perform to be the same, but there are some which have different allowed operations for each player.

Usually, the games in olympiads are going to have the following properties:

- There is usually perfect information. You can think of this as being able to see the cards in the other player's hand (which you can't see in a lot of card games generally).
- There's no element of luck and chance involved - so there won't be any probability analysis going around here.
- There is no cheating allowed (unfortunately).

The key question to answer here is - what exactly *is* a winning strategy? It turns out that there is the following theorem (which we'll prove towards the end of this chapter but isn't really too relevant to the discussions in the chapter):

Theorem 1 (Zermelo's theorem)

In any finite-stage two-player game with alternating moves, either

1. player 1 has a winning strategy.
2. player 2 has a winning strategy.
3. Both have a strategy to force a draw.

This basically means that in any game with finite number of possible positions - say tic-tac-toe or even chess (which can last a very large number of moves but is still finite since there are finitely many positions and threefold repetition would mean a draw) - there can either be a strategy by which no matter what the second player does, player one can force a win, or vice versa, or both can force a draw. A majority of the problems we'll do here won't have the concept of a draw, so in fact we can say that one of the two players must have a winning strategy.

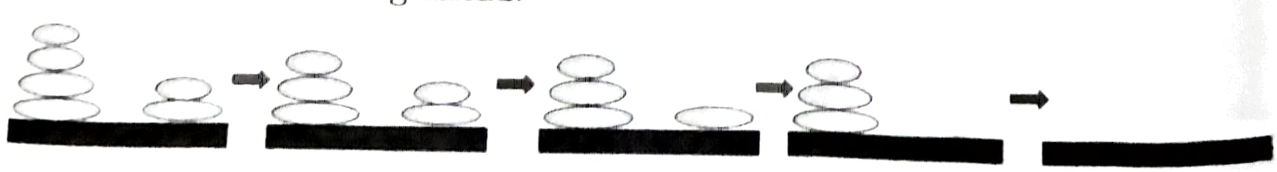
5.2 A first example

Example 2

Alice and Bob play the following game: There are two piles of stones - one of them has 20 stones and the other has 23. In a move, a player can select one of the two piles and take out some positive number of stones from it. Players play alternatively and the player who cannot make a move loses (so you lose if on your turn both piles end up being empty). Find who has a winning strategy and describe it.

This is a finite-stage game: there are at best 20×23 possible reachable positions and in particular the game ends in at most $20 + 23 = 43$ moves. There's no concept of a draw so as the theorem says, one of the players must have a winning strategy. 20 and 23 are pretty arbitrary (I just put them in randomly) and so maybe its a good idea to generalise this game to piles having a and b stones.

Figure 5.1: *A simplified version of the game with the piles having 4 and 2 stones instead. Try figuring out who wins here if both players play optimally before reading ahead.*



First guess: okay, what if I take away all stones from one pile on my first turn. That seems reasonable, but then the other player just takes all the stones from the second pile and wins - so that isn't too good. Okay, what if I take all stones except one? Turns out, that is usually not a great move either, and as to why so, we'll see in just a bit.

An idea that is really useful in combinatorial games is to consider base cases and analyse what sort of results you're getting. So in what follows we analyse different the winner for various pairs (a, b), i.e. who wins if one pile has a stones and one pile has b stones.

For instance, if one of the piles has zero stones (and the other has some positive number of stones), Alice wins because she can just take all the stones from the other pile and now Bob is left with no move to make. So (here n is a positive integer):

$$(n, 0) \rightarrow \text{Alice wins}$$

What about $(1, 1)$? Alice has to convert this to either $(1, 0)$ or $(0, 1)$ so either way Bob just takes the final coin and wins.

$$(1, 1) \rightarrow \text{Bob wins}$$

It is important to realise that using each other's strategy is something that comes up a lot in such games. We knew that $(1, 0)$ and $(0, 1)$ are winning to the player who plays first - so as soon as we reached such a position with Bob playing first - we can be sure that Bob would win in such a case. In particular, this means that induction works really well with combinatorial games - as you can reduce to a smaller situation and use the claimed winning strategy in that case.

Anyways, coming back to the problem - lets consider say $(2, 1)$. Alice can convert this to $(1, 1)$, $(0, 1)$ and $(2, 0)$. The second and third as we've seen are winning for the first player (which is now Bob) and so we want to avoid those. The first case though is exactly what we need - the player who plays second wins there (which is now Alice since Bob is the one who needs to make a move now). So putting stuff more formally, on the first move Alice needs to look for an outcome starting from where the second player would've won. If it exists, Alice has a winning strategy but if it doesn't, Bob can win no matter what Alice does and so Bob has a winning strategy.

What about $(10, 1)$. There are lots of possibilities but note that Alice just needs one case where the second player wins - and we know $(1, 1)$ works for that. So what we have so far is (where n is a positive integer):

$$(n, 0) \to \text{Alice wins}$$
$$(1, 1) \to \text{Bob wins}$$
$$(1, 1+n) \to \text{Alice wins}$$

How about, $(2, 2)$? Alice needs a position where the second player wins on the first move but so far the only ones we know to be of that type are $(1, 1)$ and $(0, 0)$ and neither of those can be reached from here (and you can check that all of the 4 possible outcomes are winning for the first player) so we get that Bob wins here. By now, you're probably getting a feel of what the answer is going to end up as - it seems to be as if Bob wins if and only if the two piles have equal number of stones. The question is - how are we going to *prove* it.

In general, there are usually multiple ways to go about *proving* who has a winning strategy. One of them is to use strong induction - and so claim that the only way Bob wins is if the piles have equal number of stones. This is clearly true for the base case. Now we assume this holds whenever the minimum of the two piles is at most n. We now show that $(n+1, n+1)$ is winning for Bob, which $(n+1, n+1+x)$ is winning for Alice. The second one is quite easy once you're done with the first - Alice just removes x and plays as Bob. As for the first one, notice that any move Alice makes leaves the piles with unequal number of stones, and the minimum of the two piles is now at most n so we can use the induction hypothesis.

However, as it turns out - there is a much neater way of framing this argument without using induction, which goes as follows:

Let's say the two piles have an equal number of stones. If Alice takes out k stones from the first pile, Bob takes out k stones from the other one - and the piles now have equal stones as well. The idea is that if Alice can take out c stones from a pile now, so can Bob from the other one because the piles end up being symmetric. So eventually Alice needs to finish a pile, at which point Bob finishes the other one and wins. On the other hand if the two piles don't have equal stones, Alice spends her first move equalising the two piles - and then plays as Bob in an equal stone game to win.

In particular, the answer for $(20, 23)$ is that Alice wins: by removing 3 from the second pile and then copying whatever Bob does onto the other pile.

5.3 Copying and symmetry

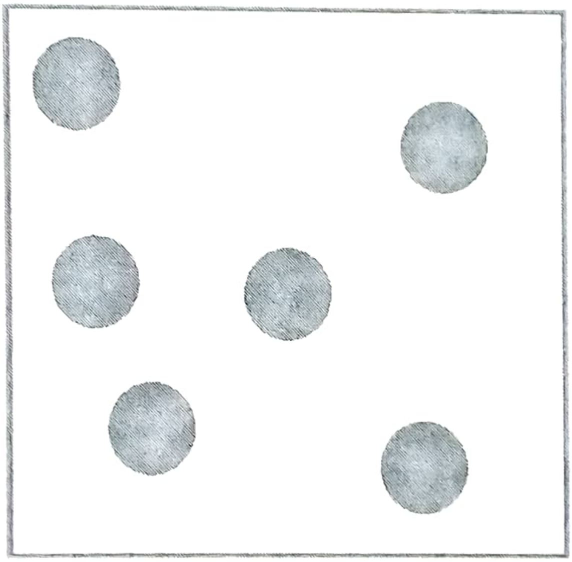
One of the ideas we used above was that of *copying* the other player. There's a funny quote on this (inspired by Pranav Sriram's book)

"They said no cheating, but we can still copy."

The idea is that many games require us to find a way to *survive* the longest - whoever can't make a move loses. So if we can create a strategy such that whenever the opponent can make a move, we can too - there's no way we could lose. Since the game is finite this would necessarily imply that the other player must lose. While this might seem slightly vague right now, an example would definitely clear this up.

Example 3

Two players, Alice and Bob place circular coins of 1 cm diameter on a square table of area one square meter. The catch is that no coin being placed can touch or overlap with a coin that has already been placed. The player who cannot make a move loses. Who has the winning strategy?



So you have a big table and people are placing coins on it. Based on what we discussed above, our strategy as the second player should be to somehow ensure that if the other player could make a move, we can make a move. So if he plays at a corner, we can play at the opposite corner. If he plays at a third corner, we play at a fourth. In general, we could *reflect* his move about the centre and place our coin at this new point. But - what if the first player played at the centre itself? Of course we can't place a coin above it, and there doesn't seem to be any *nice* way to make things work. However if there was already a coin at the centre - we would be in good shape. So in fact the key realisation is that Alice can first play in the centre and then copy whatever Bob does.

Exercise. Convince yourself that this strategy works.

> **Remark.** Unfortunately, life in combinatorial games isn't so fair. It was Bob that did all the hard-work of finding the strategy and the idea of reflection but it was Alice who used it.

Example 4 (ELMO 2019 SL)

Bella and James are playing a game. Initially, $n \geq 3$ points are given on a circle. On a player's turn, that player must draw a triangle using three unused points as vertices, without creating any crossing edges. The first player who cannot move loses. If Bella goes first and players alternate turns, who wins?

Let us change the question for a while, and solve a slightly different one. Imagine there are 2 circles, and not one. The game goes on the same way - in each step you choose one of the circles, and then draw a triangle using three unused vertices of that circle it following the conditions as given. The only difference is that choice in the start. Who wins now?

If you think about it, it'll be quite clear that the second player must win. Why? Well, if Bella makes some triangle on the first circle, James can make it on the second one. If Bella makes a triangle on the second, James can make the same triangle on the first. This is exactly the symmetry and copying idea we've been seeing so far and it all works out perfectly.

The question, though, is what do we do now that there's only one circle? You want to copy - but you don't have some sort of identical sub-problem. Well, what if you can convert this position into one with two identical sub-problems? Aha! That would be perfect. If Bella in the first move can split the game into two identical non interacting sub-games, she can copy whatever James does onto the other sub-game and win. How on earth are we going to do this though?

At this point, it's time for you to take out your pencils, and figure out a way to break the polygon into 2 smaller polygons of roughly half the size, so that every new triangle must be a part of exactly one of those polygons.

Problem. Find this starting move.

Hints: 322 259 **Soln:** Page 275, Solution 10

5.4 Winning and Losing (P and N) Positions

A lot of combinatorial games boil down to an analysis of what positions are winning for the player who currently must make the move.

This is once again similar to what we saw in the first problem - we considered lots of base cases and figured out for which positions the player who moves now wins. However, there we only interpreted it as places where Alice won and places where Bob won.

In general, it is slightly more convenient to just think of whether its the player who just moved that wins or the player that is going to move that wins. The positions in the game from where the player who is just about to play (equivalent to Alice with the game starting at that position) has a winning strategy are **P positions** and the others are **N positions**. It's slightly weird terminology - but P here stands for player and N stands for next (player).

Example 5

Alice and Bob play the following game (often played during truth and dare games to decide the next person who has to face the truth/dare). On the first turn, Alice says one of $1, 2, 3$. Players play alternately and on each following turn, the player moving can add one of $1, 2, 3$ to the current number. The player who says a number ≥ 21 loses, who has a winning strategy?

The main idea here is that instead of starting at 0, we consider what happens when we start at various numbers, and figure out whether they are P positions or N positions.

- So the question says that any number ≥ 21 is a P position (this might be a little confusing, but essentially if its your turn with the number currently being 21 or more, it means that the previous number said that number and so you won).
- What about 20? Any move I make will make the number ≥ 21 so 20 is an N position.
- For 19, I can just say 20, and I'm now the second player of an N position and so I win. So 19 is a P position. Similarly, 18 and 17 are P positions too (Check this). The general idea is that if we can make a move that leads to an N position, the current position is a P position else it is an N position.
- Now if we start at 16, we can move to $17, 18, 19$, all of which are P positions so 16 is an N position. For $13, 14, 15$ we can convert this to 16 and since we know this is an N position, each of $13, 14, 15$ are P positions.

Getting the pattern yet? The observation is that all numbers divisible by 4 seem to be N positions.

Exercise. Prove this pattern using induction or by giving an explicit strategy.

Problem. Solve this problem for 3 and 21 replaced with k and n.

Hints: 113 436 563

Example 6 (Codechef)

Two positive integers: N and k are fixed. A blackboard initially has N written on it. Alice and Bob play the following game: in a move, they must subtract k^x for some non negative integer x from the number on the blackboard, with the added criteria that the number on the board should remain non negative. A player who cannot make a move loses. Who has a winning strategy (in terms of N and k)?

In the previous problem, we started at a big number and went downwards. Another methodology that is quite useful is to start at 0 and see who wins for larger and larger numbers.

But before we get into that, lets try and get a feel of the problem. You're given two numbers - lets say 100 and 3, and initially the board has 100 on it. You're allowed to subtract a number from the set $\{1, 3, 9, 27, 81, 243, \dots\}$, from this but the remaining number should be non negative. If you can get the number to zero, you win. In other words, 0 is an N position, irrespective of N and k.

Okay, lets try and keep things simple - what if simply $k = 1$?

Exercise. Show that Alice wins if and only if N is odd if $k = 1$

Hints: 335

Okay that was almost too easy since the only number we were allowed to subtract was 1. Lets try $k = 3$ instead (we'll come to $k = 2$ a little later).

- $N = 1$ is a P position since you can subtract 1
- $N = 2$ is a N position since the only number you can go to is 1 which is not an N position.
- $N = 3$ is a P position since you can subtract 3 and get to zero
- $N = 4$ is a N position since you can only reach 1 and 3, neither of which is an N position.
- $N = 5$ is a P position since you can subtract 1 and get to an N position.
- $N = 6$ is a N position since both 5 and 3 are P positions.

All in all, we have

1	P	3	P	5	P
2	N	4	N	6	N

See the pattern yet? Seems like Alice wins if and only if N is odd, once again.

Exercise. Show this

Hints: 477 411

The key reason why all of this was working is that we were only able to select odd numbers - so each time we did this, the parity changed. If $k = 2$ however, we're allowed to subtract 1 which is odd but also a set of evens, which makes it a lot trickier.

So lets see what $k = 2$ has in store for us. I strongly suggest that you try finding out the answers for the first 10 values on your own first, and then read further

- For $N = 1$, you can subtract one and so its a P position.
- For $N = 2$, you can subtract two, and so its a P position as well. Our guess of odd and even has failed already!
- For $N = 3$, both the allowed moves take you to P positions, so its an N position.
- For $N = 4$, you can subtract 4 and win so its a P position
- For $N = 5$, you can subtract two to go to 3 which we know to be an N position do $N = 5$ is a P-position
- For $N = 6$, you can go to $5, 4, 2$ but all of them are P positions so 6 is an N position.
- For $N = 7$, you can subtract one and go to an N position so its a P position.
- For $N = 8$ we can subtract two and go to 6 which was an N position (or well we could subtract 8 directly) so its a P position.
- For $N = 9$ we can go to $8, 7, 5, 1$ but all of them were P positions so 9 is an N position.

So essentially we have that

1	P	2	P	3	N
4	P	5	P	6	N
7	P	8	P	9	N

So either someone is fooling with us or the pattern here is that the numbers divisible by 3 are N positions.

Lets see if we can try proving this by induction. Clearly if the pattern holds till $3n$, we can go from $3n+1$ and $3n+2$ to $3n$ which was an N position, so $3n+1$ and $3n+2$ are P positions. From $3n+3$, any number we go to cannot be a multiple of 3 (we're subtracting a power of 2 which obviously isn't divisible by 3) so the resulting position is going to be a P position (by the induction hypothesis the only N positions were the ones divisible by 3) no matter what we do - so $3n+3$ is an N position and the induction is complete!

But... what about general k?

Problem. Try the problem for $k = 4$, find the pattern and prove it in general

Soln: Page 275, Solution 11

5.5 Proof by contradiction

This is a technique I'm sure most of you have heard of somewhere or the other, but we're going to see how it can be used to prove that a certain player has a winning strategy in a certain position. The idea is that we assume for the sake of contradiction that the other player has the winning strategy and then we find a magical way to use this strategy against him to win the game.

This is something that can be understood well only with an example so lets take one below.

Example 7 (MBL 2023 Application)
Let n be a positive integer. Initially, the numbers

$$1, 2, 3, \ldots, n-1, n$$

are written on a blackboard. Two players, Alice and Bob take turns making moves, with Alice going first. On their turn, a player erases a number as well as all of its remaining divisors from the blackboard. The player who cannot make a move loses. Show that Alice has a winning strategy for every n.

So according to our ideology, lets say for the sake of contradiction that Bob has the winning strategy. Then no matter Alice does in her first move, Bob has some move that helps him win.

The key observation is the following: what happens if Alice removes 1 in the first turn? Now Bob apparently has a winning strategy by removing some number (say x) from $2, 3, \ldots, n$. But notice that if Alice had removed x in her first move itself, 1 would have been erased along with it and we could then use Bob's strategy against him. Boom.

Example 8

Alice and Bob are playing a game on an $m \times n$ board - all the cells of which are initially blue. In a move you're allowed to choose any cell (a, b) that is currently blue and change the color to all cells in the rectangle with corners at $(1, 1)$ and (a, b) to red (cells that were already red remain red). If at the end of your turn, no cell is blue - you lose. Show that if $mn > 1$, A has a winning strategy.

The idea in this problem is actually remarkably similar to the previous problem - to an extent that they really feel like they're the "same problem".

There are lot of ways you can try and approach this, and try to create symmetry and cases - but I am not aware of any solution that actually characterizes this winning strategy for general m, n - all we can prove (easily) is that Alice wins if both players play optimally.

So let's think like we did in the previous problem. The key observation there was to delete 1 - because it's almost like "not making a move" because it would end up going on the next move anyway. Along the same lines, the trick here is to pick $a = b = 1$ on the first turn.

Problem. Find the finish

Soln: Page 276, Solution 12

> **Remark.** Apparently, if the cell $(1, 1)$ is colored red before the game begins, this game is *unsolved* for general (m, n).

Let's now take a slightly trickier example which is based on the same idea.

Example 9 (Russia 2011)

There are $N > n^2$ stones on the table. Peter and Vasya play a game, Peter starts. Each turn, a player can take any prime number p less than n and remove p stones, or any multiple of n stones, or 1 stone. Prove that Peter always can take the last stone (regardless of Vasya's strategy).

Once again, lets assume that no matter what Peter does, Vasya has some *winning move*. So if Peter removes $4n$ stones there is some move that Vasya can do to win. First notice that if this *winning move* is subtracting another $5n$ stones, we could've directly subtracted $4n + 5n = 9n$ stones and used Vasya's strategy. In general, if we used our first move to subtract kn stones, Vasya's winning move canto have been removing ln stones (why?). So her move must have been subtracting some prime.

As an example, let's say if we subtract n stones from the starting N her winning move is to subtract 5. If we subtract $2n$, her winning move is to subtract 3, and so on. In particular, if we subtract kn stones she has a winning strategy by subtracting $f(k)$ stones where f is some function which maps to primes less than n.

So after the first two moves, we're left with $N - kn - f(k)$ stones and apparently no matter what Peter does here he loses. Well, what if we subtract some ln stones now. There are $N - (k+l)n - f(k)$ stones left. Notice that if $f(k) = f(k+l)$ we would have a contradiction (why??). So it remains to show that this function isn't injective, i.e $f(a) = f(b)$ for some $a \neq b$.

But that is true since the number of primes less than n is, well, less than n (probably the worst bound on the number of primes you can come up with) and since $N > n^2$ k can take any value from 1 to n.

Problem. Figure out why $f(k+l) = f(k)$ results in a contradiction, and how you can use this to construct a contradiction given some $a < b$ with $f(a) = f(b)$

Hints: 345 355

5.6 Nim-game

As the first example, we encountered the problem below:

"Alice and Bob play the following game: There are two piles of stones..."

The question we raise here is, what is so special about two? Well, a lot of things - it's the only even prime for instance. But as far as we were concerned, only that it's small. The general nim game deals with solving the same problem but with an arbitrary number of piles instead.

Example 10

Alice and Bob play the following game: There are n piles of stones - having $a_1, a_2, \ldots, a_n$ stones respectively. In a move, a player can select one of the n piles and take out some positive number of stones from it. Players play alternatively, and the player who cannot make a move loses (so you lose if, on your turn, the final pile has just become empty). Who has the winning strategy?

Let's try solving the problem for three piles first. We'll try and find as many N positions as we can. The idea is that (a, b, c) and (a, b, d) can't both be losing as if $c > d$, we can convert (a, b, c) into (a, b, d) with one move. So essentially if we fix a and b they should lead to a (hopefully) unique c (hopefully since there may be no c at all). So for instance if $a = b = 1$, we can see that $c = 0$ is a losing position. Let's in fact define a function $f(a, b) = c$ if (a, b, c) is an N position and -1 if none of the positions are N.

Clearly $f(a, a) = 0$. Now let's fine $f(1, 2)$. $(1, 2, 0)$ is P, $(1, 2, 1)$ and $1, 2, 2$ are also P (just get rid of the unequal pile). Note that $(1, 2, 3)$ is in fact N (check this!). So $f(1, 2) = 3$. Similarly, we get that $f(1, 4) = 5$ and $f(1, 6) = 7$.

Seeing this we may guess that it has something to do with one number being the sum of the other two. Indeed, $(2, 4, 6)$, $(2, 5, 7)$ are N positions too. The scare comes in however, when you see that $(3, 5, 6)$ is in fact an N position! Looking further, we also get that $(3, 9, 10)$ $(5, 9, 12)$ and $(5, 11, 14)$ are N positions.

Exercise. Check that the following table is correct.

f	0	1	2	3	4	5	6	7
0	0	1	2	3	4	5	6	7
1	1	0	3	2	5	4	7	6
2	2	3	0	1	6	7	4	5
3	3	2	1	0	7	6	5	4
4	4	5	6	7	0	1	2	3
5	5	4	7	6	2	3	0	1
6	6	7	4	5	2	3	0	1
7	7	6	5	4	3	2	1	0

The more you look at the table, the more it becomes evident that powers of 2 are involved. the second row seems to be working in patterns of 2 consecutive numbers, while the third and fourth in patterns of four consecutive numbers. All of this suggests one thing: *binary*. So let's consider some of the *exceptions* we saw earlier in binary.

3 :	011	3 :	0011	5 :	0101	5 :	0101
5 :	101	9 :	1001	9 :	1001	11 :	1011
6 :	110	10 :	1010	12 :	1100	14 :	1110

See the pattern yet? The guess is that each binary digit seems to have an even number of ones! In fact, this is the answer. The position is N if and only if there are an even number of 1s for each binary digit. Why? Well, that's for you to figure out!

5.7 Problems

Problem 1 (Lithuania 2010). In an $m \times n$ rectangular chessboard,there is a stone in the lower leftmost square. Two persons A,B move the stone alternately. In each step one can move the stone upward or rightward any number of squares. The one who moves it into the upper rightmost square wins. Find all (m, n) such that the first person has a winning strategy.

Hints: 560 482

Problem 2 (European Mathematical Cup 2016). For two positive integers a and b, Ivica and Marica play the following game: Given two piles of a and b cookies, on each turn a player takes $2n$ cookies from one of the piles, of which he eats n and puts n of them on the other pile. Number n is arbitrary in every move. Players take turns alternatively, with Ian going first. The player who cannot make a move, loses. Assuming both players play perfectly, determine all pairs of numbers (a, b) for which Marica has a winning strategy.

Hints: 614 623 419

Problem 3 (Pranjal Srivastava). There are two piles of n plates each. Two players in turns, remove plate(s) from a pile. In a player's turn they can either remove a plate from the first pile or remove a plate from the second pile.

If a pile does not have any plates, you cannot remove plates from that pile. The only move would be to remove one stone from the other pile. The player who removes the last plate wins. For what n does the first player have a winning strategy?

Hints: 157 49

Problem 4 (Durer Olympiad). At the beginning of the game, there is a table with 10 coins on it. Two players alternate turns. During a player's turn, they can make a move in four ways:

- They can take 1 heads-up coin off the table.
- They can take 2 heads-up coin off the table.
- They can flip 1 tails-up coin (and thereby make it heads-up).
- They can flip 2 tails-up coins (and thereby make them heads-up).

The objective is to remove the last coin from the table. Find all configurations where the first player wins.

Hints: 231

Problem 5. Prove Zermelo's theorem as mentioned at the start of the chapter.

In any finite-stage two-player game with alternating moves, either

1. player 1 has a winning strategy.
2. player 2 has a winning strategy.
3. Both have a strategy to force a draw.

Hints: 208 640

Problem 6 (Saint Petersburg 1997). You are given a composite positive integer N. Alice and Bob take turns writing composite divisors of N on a board, according to the following rules. One may not write N. Also, there may never appear two coprime numbers or two numbers, one of which divides the other. The first player unable to move loses. If A starts, who has the winning strategy?

Hints: 145 632

Problem 7 (USAMO 1999/5). The Y2K Game is played on a 1×2000 grid as follows. Two players in turn write either an S or an O in an empty square. The first player who produces three consecutive boxes that spell SOS wins. If all boxes are filled without producing SOS then the game is a draw. Prove that the second player has a winning strategy.

Hints: 124 426 435

Problem 8 (Double Chess). Consider the following variant of chess: on your turn: you're allowed to make two moves instead of one. Players alternate, pieces move the same way as normal chess, and white goes first. Show that white can always win or draw.

Hints: 582 611 13

> **Remark.** This problem is one of my all-time favourites.

Problem 9 (Durer Olympiad). The number N is written on a blackboard. 2 players take turns reducing the number on the blackboard by an integer. On the first turn of the game, the first player can reduce the number (which is 56 at the start) by any one of $1, 2, 3, \ldots, (N-1)$. In the following turns, each player can only reduce it by some number which is strictly less than twice of how much was reduced in the previous term. (for example, if in the previous turn the number on the blackboard was reduced by 3, the player whose turn it is now, can reduce it by one of $1, 2, 3, 4, 5$). The goal of the game is to write 0 on the board. Who has the winning strategy?

Hints: 637 51

Problem 10 (Sophic Fellowship Application 2021). Alice and Bob play a game. There are some piles (> 1) of stones on a table. In the first move, Alice chooses two piles and asks Bob to remove some stones from one of the piles. Then Bob chooses two of the remaining piles (1 if only 1 pile remains) and Alice removes a positive number of stones from one of the piles and they keep alternating as such until one of them removes the last stone from the table. The one who removes the last stone wins. Find who has the winning strategy based on the initial position of the game.

Hints: 503 44 639

Problem 11 (2004 IMOSL C5). A and B play a game, given an integer N, A writes down 1 first, then every player sees the last number written and if it is n then in his turn he writes $n+1$ or $2n$, but his number cannot be bigger than N. The player who writes N wins. For which values of N does B win?

Hints: 385 139 633 569

Problem 12 (IMOSL 2015 C4). Let n be a positive integer. Two players A and B play a game in which they take turns choosing positive integers $k \leq n$. The rules of the game are:

(i) A player cannot choose a number that has been chosen by either player on any previous turn.

(ii) A player cannot choose a number consecutive to any of those the player has already chosen on any previous turn.

(iii) The game is a draw if all numbers have been chosen; otherwise the player who cannot choose a number anymore loses the game.

The player A takes the first turn. Determine the outcome of the game, assuming that both players use optimal strategies.

Hints: 603 625 361 511 358

Problem 13 (AoPS). Two players, A and B play the following game: Two piles are given, each one contains some stones (not necessarily the same amount). In the first round, B tells A k, a positive integer and A takes k stones from one of the piles (if k is greater than the number of stones in A's chosen pile, then he takes the whole pile). In the second round, A tells B a positive integer l and B takes l stones from one of the piles, etc. The one who takes the last stone loses. Who has a winning strategy and what is it?

Hints: 402 222 18 141 472 91

Problem 14. Two players A and B are given a set of numbers $a_1, a_2, \ldots, a_n$ and a number N. They play alternatingly, and A begins. On their move, they can choose one of the a_i and subtract a_i from N under the condition that $N - a_i \geq 0$. The player who can't make a move loses. Show that if N is large enough, A wins if and only if N is one of $b_1, b_2, \ldots, b_k$ modulo m, for some k, m, b_i. In other words, show that for large enough N, the set of N that work are basically ones that leave specific remainders modulo some number.

Hints: 4 366 498

6 The Pigeon-Hole Principle

6.1 Introduction

Firstly, what even is this name?! Pigeon - hole?! Well, you'll see soon as to how and why this is an actual principle worth dedicating an entire chapter to.

The pigeon-hole principle states the following fairly obvious fact:

Theorem 1 (The pigeon-hole principle)
If you place $n+1$ pigeons among n holes, some hole will have more than 1 pigeon.

More generally, if you place $nk+1$ pigeons among n holes, some hole will have more than k pigeons.

The proof is easy. Let's say that for the sake of contradiction, every hole has at most k pigeons. Then the total number of pigeons can be $\leq k+k+\cdots+k = nk$ but we were given $nk+1$ pigeons, and so we've reached a contradiction.

Also note that $nk+1$ is the best bound possible - if you had nk pigeons you *could* place k into each hole and get away with it.

Okay but this was kind of obvious, no? If the theorem is this *simple*, why would it be useful? To answer this, ask yourself: who said that the most *useful* theorem is the one that is hardest to prove? Sure, there are a lot of theorems that are useful but hard to prove, but that doesn't imply anything about the theorem we have given above.

The idea is that usually when we solve a problem with the pigeon-hole principle, *applying the theorem* isn't really the hard step. Coming up with the appropriate equivalents of pigeons and holes in the given problem is the part which is tricky and important.

Exercise. To get the idea of the principle, try out the following problem.

How many people do you need so that you can be sure that some 3 of them have the same birthday.

Let's start off with the following problem.

Example 2

In a group of n people, some pairs of people know each other (and are friends). It is known that each person knows at least one person. Show that there are two people who know the same number of people

Let's consider each person - their number of friends belongs to the set $\{1, 2, \ldots, n-1\}$. You can think of these as the holes. The people themselves are the pigeons, and they go into the holes according to the number of friends they have. We have $n-1$ holes and n pigeons so some hole has two pigeons which here means that two people have the same number of friends.

It turns out, the slightly stronger statement as follows also holds.

Example 3

In a group of n people, some pairs of people know each other (and are friends). Show that there are two people who know the same number of people

All I did is remove the condition that *each person knows at least one person.* Well now the possibilities for number of friends is $\{0, 1, 2, \ldots, n-1\}$ which has n elements - so pigeon-hole doesn't really give us anything. Or does it? The idea is that the only way all holes have less than 2 pigeons is that all of them have exactly one. In particular there should be one person with 0 friends, and one person with $n-1$ friends. Hmm.

Exercise. Find the finish!

6.2 Example problems

Example 4

Nine lattice points are chosen in three dimensional space. Prove that you can choose two of these points so that their midpoint is a lattice point too. (Here a lattice point refers to a point with integer coordinates).

Let's say the coordinates are $(x_1, y_1), (x_2, y_2), \ldots, (x_5, y_5)$ what we want is something like

$$\left(\frac{1}{2}(x_3 + x_4), \frac{1}{2}(y_3 + y_4)\right)$$

to be a lattice point but this requires $x_3 + x_4$ and $y_3 + y_4$ to be even. This in turn is possible if x_3 and x_4 are both even or both odd, and y_3 and y_4 are also both even and both odd.

In particular we use the term *parity* and say that x_3 and x_4 have the same parity if they're both even or both odd. Then what we need is x_i and x_j to be same parity, and y_i and y_j also to be of the same parity. Notice that the actual values of x_i and y_i don't really matter any more - all we care about is whether they're even or odd. So the coordinate in this new viewing sense looks more like one of

(odd, odd)
(odd, even)
(even, even)
(even, odd)

There are only four possibilities, but five points... so one possibility repeats!

Exercise. Check that this finishes

Example 5

Among six persons there are always three who know each other or three who are mutual strangers.

Let's say I'm one of the six people. Then there are five other people. I either know some 3 of them, or I don't know some 3 of them.

Without loss of generality, let's say that I know some 3 of them (the other case follows in an identical manner). So I know A, B, C. If A and B know each other, we'd have a complete triangle of people who know each other (how?). Similarly we'd be done if B and C know each other, or A and C know each other.

But if none of the three pairs know each other then A, B, C are mutual strangers and we're done. Boom.

Example 6

Let $a_1, a_2, \ldots a_{10}$ be 10 not necessarily distinct integers. Show that there exists a subset of these numbers with sum divisible by 10.

This is a pretty tricky example of the pigeon-hole principle - and it contains an idea that is often quite useful. The idea is to consider *prefix sums*. Essentially,

$$S_k = a_1 + a_2 + \cdots + a_k$$

for each $k \in \{1, 2, \ldots, 10\}$. Now clearly all of these S_i are subsets so if any of them has a sum divisible by 10, we're done already. The cooler observation though, is that their differences also form subsets. In particular, $S_6 - S_3 = a_4 + a_5 + a_6$ so now 2 S_i should have a difference divisible by n.

We now put the S_i into ten buckets: numbers of the form $10k, 10k+1, 10k+2, \ldots, 10k+9$. Clearly every number is of exactly one of the above types. Moreover, we can't have any number of the type $10k$, so we have nine buckets. Nine buckets and ten numbers, hmm.

Example 7

A chess master has 77 days to prepare for a tournament. He wants to play at least one game per day, but not more than 132 games. Prove that there is a sequence of successive days on which he plays exactly 21 games.

A cool thing about this problem is that the constants seem absolutely random, but they turn out to be perfectly relevant.

The basic idea is to consider prefix sums once again. The hope is that two prefix sums differ by exactly 21, that is $S_a = S_b + 21$. So consider

$$(S_1), (S_2), (S_3), \ldots, (S_{77}), (S_1 + 21), (S_2 + 21), (S_3 + 21), \ldots, (S_{77} + 21)$$

Since at least one game is played each day. Clearly if $i \neq j$,

- $S_i \neq S_j$
- $S_i + 21 \neq S_j + 21$

We now also have that $S_i \neq S_j + 21$ (for the sake of contradiction) so these are 154 distinct numbers. But $S_{77} + 21 = 132 + 21 = 153$. So we have 154 positive integers each at most 153. Hmm.

Remark. Isn't it amazing how

$$77 \times 2 = 132 + 21 + 1$$

works perfectly?

Example 8 (Erdos)

The positive integers from 1 to 101 are written down in some order. Show that you can remove 90 of the numbers so that the remaining 11 numbers are either in increasing order or decreasing order.

This is one of the coolest (and trickiest) examples of the pigeon-hole principle. I recommend you to give it a try yourself before coming back here.

So, let's give it a try! We basically need some subsequence (not necessarily contiguous) which is of length 11 and consists of numbers in increasing or decreasing order. To solve this question, let's ask ourselves the following question instead:

How would you find the longest increasing subsequence?

The trick is to maintain some extra information. Consider the length of the longest increasing subsequence X_i ending at each element a_i and use recursion. If we want to find X_i, it's equivalent to finding the best previous subsequence to which we can append a_i. In other words, we consider all j such that $a_j < a_i$ and $j < i$. Now for all such j, we can consider the subsequence as

$$\dots \quad a_j \quad a_i$$

which has length $X_j + 1$. So we just take the max X_j over all such j, and define X_i to be that value plus one. Finally our answer is going to be the maximum of all X_i, since the longest increasing subsequence could have ended anywhere.

Anyway, to this end, let's consider the lengths of the largest increasing subsequence A_i and decreasing subsequence B_i ending at i. The key idea is the following: if $a_i < a_j$ then then $A_j \geq A_i + 1$ and if $a_i > a_j$ we have that $B_j \geq B_i + 1$. In particular, this means that the pair $(A_i, B_i) \neq (A_j, B_j)$ for all $i \neq j$.

Exercise. Find the finish! You'll probably need that $101 > 10 \times 10$.

6.3 Number Theoretic Examples

The pigeon-hole principle also has a ton of number theoretic examples. For the problems that follow, we'll need a bit of number theory background, but don't worry too much, it's nothing too fancy.

Essentially, we say that

$$a \equiv b \pmod{c}$$

(read out as a is congruent to b modulo c) if the difference between a and b is divisible by c. So, for instance:

$$5 \equiv 3 \pmod{2}$$

$$17 \equiv 5 \pmod{6}$$

since $5 - 3$ is divisible by 2 and $17 - 5$ is divisible by 6.

Exercise. Show that if $a \equiv b \pmod{c}$ and $x \equiv y \pmod{c}$ we have

- $a + x \equiv b + y \pmod{c}$
- $ax \equiv by \pmod{c}$

Example 9

Show that there exists a positive integer n such that

$$2^n \equiv 1 \pmod{2023}$$

Note that each number is congruent to one of the 2023 remainders, i.e. $\{0, 1, 2, \ldots, 2022\}$ modulo 2023. Consider the sequence

$$\{2^1, 2^2, 2^3, \ldots, 2^{100000}\}$$

This has more than 2023 numbers, so some remainder repeats. In other words,

$$2^a \equiv 2^b \pmod{2023}$$

for some $a \neq b$. Now let's say $a > b$. We have that $2^a - 2^b = 2^b(2^{a-b} - 1)$ is divisible by 2023. Clearly 2^b shares no factor with 2023 so $2^{a-b} - 1$ is divisible by 2023. Hence, $n = a - b$ works!

Example 10

Consider the Fibonacci series $1, 1, 2, 3, 5, 8, \ldots$. Show that for any n there is a Fibonacci number ending with n zeroes.

Let $F_1 = 1, F_2 = 1, F_3 = 2, \ldots$.

So we essentially want to consider the series modulo 10^n. If we can show that some term $F_t \equiv 0 \pmod{10^n}$ we would be done. In fact, we show that we can find a term t for each N such that

$$F_t \equiv 0 \pmod{N}$$

. The idea is the following. Consider $(F_a, F_{a+1}) \pmod{N}$, there are only N^2 possibilities for this (both are in $\{0, 1, 2, \ldots, n-1\}$), and so at some point we have a repeat, i.e.

$$(F_a, F_{a+1}) \equiv (F_b, F_{b+1}) \pmod{N}$$

In particular, this means that

$$F_{a+2} = F_{a+1} + F_a \equiv F_{b+1} + F_b = F_{b+2} \pmod{N}$$

and so

$$(F_{a+1}, F_{a+2}) \equiv (F_{b+1}, F_{b+2}) \pmod{N}$$

This in turn gives us $F_{a+3} \equiv F_{b+3} \pmod{N}$.

We could also do this backwards, and get that $F_{a-1} \equiv F_{b-1} \pmod{N}$ and so on. Thus we essentially get that the entire sequence is periodic mod N with period $a-b$, that is,

$$F_n \equiv F_{n+a-b} \pmod{N}$$

for all n.

But... how did that help us? We know that if there's one zero there'll be infinite of them, but who said that there should be a zero. We defined $F_1 = F_2 = 1$. The magic trick is to define F_0. So as to satisfy the recurrence, $F_0 = F_2 - F_1 = 0$, which is 0 modulo N (obviously). So later in the series too, there'll be a term which is zero modulo N.

Example 11

Let α be an irrational number. Show that there exist m, n with $m \leq 100$ so that

$$|m\alpha - n| < 0.01$$

We consider $x_i = i\alpha - t_i$ where t_i is an integer so that $x_i \in [0, 1)$ (so we're taking the fractional part of $i\alpha$ essentially). Now list $x_0, x_1, x_2, \ldots, x_{100}$. They all lie between 0 and 1. The trick is to split the gap between 0 and 1 into hundred buckets:

$$[0, 0.01), [0.01, 0.02), \ldots, [0.99, 1)$$

Since we have 101 x_i, some two of them fall into the same bucket, so

$$|x_i - x_j| < 0.01$$

(assuming $i \geq j$). But this means that

$$|(i-j)\alpha - (t_i - t_j)| < 0.01$$

and $(i-j) \leq 100 - 0 = 100$ so we can set $m = i - j, n = t_i - t_j$ and conclude!

Problem 12. As a bonus: prove that there exist infinitely many m, n such that

$$|\alpha - \frac{n}{m}| < \frac{1}{m^2}$$

6.4 Problems

Problem 1. Show that there exists some four digit sequence that occurs infinitely often as the first four digits of a power of two.

Hints: 628

Problem 2 (LMAO 2023/4). Let $\pi(n)$ denote the number of primes at most n. Show that if we have a subset of $\{1, 2, 3, \ldots, n\}$ of size $\pi(n) + 1$, some number divides the product of all others.

Hints: 328 134 254 **Soln:** Page 276, Solution 13

Problem 3. From 52 positive integers, show that we can select two distinct numbers a and b so that one of $a + b$ and $a - b$ is divisible by 100.

Hints: 209 29

Problem 4. Each of ten segments is longer than 1 cm but shorter than 55 cm. Prove that you can select three sides of a triangle among the segments.

Hints: 132 444

Problem 5. Show that one of $n + 1$ numbers from $\{1, 2, 3, \ldots 2n\}$ is divisible by another.

Hints: 100 211

Problem 6. Given a set M of 1985 distinct positive integers, none of which has a prime divisor greater than 26, show that there exists at least one subset of four distinct elements in M whose product is the fourth power of an integer.

Hints: 441 117

Problem 7. Each of 17 scientists correspond with all others. They correspond about only three topics and any two treat exactly one topic. Prove that there are at least three scientists who correspond with each other about the same subject.

Hints: 634 562

Problem 8. Given a set of 5 numbers, show that there exists a subset of them with 3 elements with sum divisible by 3.

Further, given a set of 17 numbers, show that there exists a subset of them with 9 elements with sum divisible by 9.

Hints: 155 175 539 599

> **Remark.** Any guesses on what the general statement could look like? The general statement is in fact true and is known as the Erdös Ginzburg Ziv theorem.

Problem 9. Find the maximum number of colored squares in an $n \times n$ grid if the centres of no four colored squares form a parallelogram.

Hints: 6 297 572 153

Problem 10. Among nine persons, there are three who know each other or four persons who don't know each other.

Hints: 178 644 326

Problem 11. A plane is colored so that each point in it is one of red and blue. Prove that there exists a rectangle with all vertices of the same color.

Hints: 657 516

Problem 12. Let $a_1, \ldots, a_{100}$ and $b_1, b_2, \ldots, b_{100}$ be two permutations of the integers from 1 to 100. Prove that among the products $a_1b_1, a_2b_2, \ldots, a_{100}b_{100}$ there are two which have difference divisible by 100.

Hints: 474 627

Problem 13. The numbers from 1 to 81 are written on the squares of a 9×9 board. Prove that there exist two neighbours which differ by at least 6.

Hints: 330 50

Problem 14 (2003 IMOSL N1). Let m be a fixed integer greater than 1. The sequence x_0, x_1, x_2, ... is defined as follows:

$$x_i = \begin{cases} 2^i & \text{if } 0 \leq i \leq m-1; \\ \sum_{j=1}^{m} x_{i-j} & \text{if } i \geq m. \end{cases}$$

Find the greatest k for which the sequence contains k consecutive terms divisible by m .

Hints: 61 336 391

Problem 15. Each of m cards is labelled by one of the numbers $1, 2, \ldots, m$. Prove that if the sum of the labels of any subset of the cards is not a multiple of $m+1$, all cards are labelled by the same number.

Problem 16. A sequence of m positive integers contains exactly n distinct terms. If $2^n < m$ show that there exists a block of consecutive terms so that their product is a perfect square.

Hints: 588 447 342

Problem 17. Suppose five of the nine vertices of a regular nine-sided polygon are arbitrarily chosen. Show that one can select four among these five such that the form the vertices of a trapezium.

Hints: 264 520

Problem 18. There are n consecutive integers $m, m+1, \cdots, m+n-1$. Prove that you can pick some nonempty subset of these numbers whose sum is divisible by $(1 + 2 + \cdots + n)$.

Hints: 242 481 610 553

7 Soft Techniques

7.1 Introduction

"Soft techniques" in olympiads are techniques that, well, prove nothing.

What...? Alright, I skipped a couple of details. Soft techniques, more specifically, refer to methods you can use to help build a greater *understanding* of the problem. They can build your intuition on what the answer is expected to be. While they don't "prove" anything on their own, they can often give you hints on methods that can be used to prove so. Induction from the previous chapter comes in a lot here too.

So what exactly are these soft techniques. The number one and most important one is:

Check base cases!!

The idea is that whenever a problem gives you a variable such as n, k, or a number that looks absolutely random (such as the current year), see what happens when you replace it with a small number like 1, 2, or 3. Most of the time, looking into these base cases can give you some insight into the problem - or at the very least convince you that the problem "seems" to be true (which by the way, may not always be obvious otherwise!).

Sometimes solving the problem for $n = 4$ or 5 might get you really close to the complete solution, while you might end up doing $n = 20$ without any considerable *understanding* of what really is going on. It is therefore important to find a balance between trying a lot of base cases and understanding when you're *mentally ready* to try solving the problem.

Here are some other examples of soft techniques:

- Trying to find equality cases when you need to prove a certain bound.
- Coming up with a random complicated example (not a base case anymore) and just playing with this to get a hang of the tools you're given
- Considering variants of the problem: what happens if this condition was removed?
- Trying to disprove the problem (!) - if you're not convinced yet, maybe trying (and failing) to disprove the problem by trying to construct a counterexample

can help you find out why such a 'counterexample' can't exist

- Finding out the use of each of the conditions. For example if you're given that n is odd, it may be useful to find out why the problem is false if n was even.

7.2 Examples

Example 1 (United Kingdom 2011)

You have been given two positive integers a, b with $a \leq b$ written on a blackboard. You are allowed to perform the following 2 operations:

(1) Subtract a positive integer k from each of the two integers on the board.

(2) Double any one of the two positive integers.

Show that these two operations can be applied in some order so that both the integers on the board end up as 0.

Okay, so we have two tools in our hands and we need to transform some (a, b) into $(0, 0)$.

A question to consider: how strong are each of these tools on their own? The first one is quite useful, we can subtract a from both the numbers in the first step itself to end up with $(0, b - a)$. In fact, if $a = b$ we're done already! However if say, $b - a = 1$, converting $(0, 1)$ into $(0, 0)$ is quite hard (in fact it's impossible).

The issue is that if one of the two numbers becomes negative, they can never be made positive after this (why?). So if we use tool 1 at any point, this would make the 0 in $(0, 1)$ negative and we would be stuck forever. On the other hand the second tool on its own wouldn't be useful either as it alone can't make a positive integer like 1 in $(0, 1)$ reach 0.

But this does give us some intuition on what we need to do. For starters, we now know that if $a = b$ at any point we would be done. So maybe that's a better objective - get both numbers to be equal. If we start off with equal numbers, we're happy so lets say we start with some $a \neq b$.

At such points in a problem, it is super useful to work out a strategy for small (a, b) and find patterns which (hopefully) generalise.

So lets say $(a, b) = (1, 2)$. This is easy, just double a and now they're equal!

What about $(1, 3)$? This one is slightly tricky. We can't subtract because that'll make the first term ≤ 0 which we've seen earlier to be bad. Doubling 3 seems dumb as we want to reach equality while $(1,6)$ are just further apart. So lets say we double a and get $(2,3)$. Now what? Aha, we can subtract 1 from each and end up with $(1,2)$ which we did earlier.

What about say, $(1,7)$? Lets try the same thing. So we get to $(2,7)$ and then we subtract one from each to get $(1,6)$. Wait so we can just keep reducing the latter number by one! Indeed, if we began with $(1,n)$ we can multiply the first number by 2 to get $(2,n)$ and then subtract one from each to get $(1,n-1)$. We repeat this to get $(1,n-2)$ and $(1,n-3)$ and so on until we get to $(1,1)$ at which point we can subtract one from each and finish!

Exercise. Find out a way to finish using this idea if we begin with a general (a,b) and not $(1,n)$.

Remark. This problem sums up how one can try to approach a wide range of combinatorial problems.

- Try and *understand* the tools you've been provided with.
- Use the tools on base cases and small values.
- Develop an intuition on what is going on and what should be done.
- Formalise your intuition and finish the problem.

Example 2 (2020 ISL C1)

Let n be a positive integer. Find the number of permutations $a_1, a_2, \dots a_n$ of the sequence 1, 2, ... , n satisfying

$$a_1 \leq 2a_2 \leq 3a_3 \leq \cdots \leq na_n$$

Aha, we have at our hands a n. Let's exploit this as much as possible, by setting $n = 1, 2, 3$ and so on and trying to find patterns. Just to clarify, the problem is essentially saying that we have some permutation of $1, 2, 3, \dots, n$, so say $n = 5$ and we have that the permutation is $3, 1, 2, 5, 4$. Then $a_1 = 3, a_2 = 1, a_3 = 2$ and so on.

- $n = 1$: In this case there is only one permutation and that trivially works. That didn't really give us much of an idea of what is going on but let's go on.
- $n = 2$: There are two possible permutations $1, 2$ and $2, 1$ and both work (since $1 \leq 4$ and $2 \leq 2$).
- $n = 3$: Here's where stuff starts getting messy, so let's write down all the permutations:

1. $1, 2, 3$: This works since $1 \le 4 \le 9$.
2. $1, 3, 2$: This works since $1 \le 6 \le 6$.
3. $2, 1, 3$: This works since $2 \le 2 \le 9$.
4. $2, 3, 1$: This doesn't work since $2 \le 6 \ge 3$.
5. $3, 1, 2$: This doesn't work since $3 \ge 2 \le 6$.
6. $3, 2, 1$: This doesn't work since $3 \le 4 \ge 3$.

In conclusion three permutations work. By now we're getting a slight hang of what's going on. And a hunch on the answer too? For $n = 1$, the answer was 1. For $n = 2$, it was 2, and for $n = 3$, it does seem to be 3.

The guess, then, is that the answer is n itself.

However it's important to be double sure about such a guess so that you don't end up trying to prove a false claim for the next two hours. So let's try $n = 4$.

- $n = 4$: After a bit of effort, notice that the following permutations work:

  1. $1, 2, 3, 4$
  2. $1, 2, 4, 3$
  3. $1, 3, 2, 4$
  4. $2, 1, 3, 4$
  5. $2, 1, 4, 3$

  There are five of them! And this is the important realisation. It's very easy to get carried away when you see four solutions and miss the fifth one because you're already somewhat convinced that the answer is going to be four.

  While it's important to trust your intuition, it's important to also have a voice in your head that is just *never convinced* until it's really done.

So it turns out that the pattern is $1, 2, 3, 5$. But the answers are not the only cool part of the information we've found when solving these base cases. Let's list down some of the observations (mostly from $n = 3, 4$)

- **Observation 1**: $1, 2, 3, \ldots, n$ will always work since $1 \le 4 \le \cdots \le n^2$.
- **Observation 2**: Let's bring back the five answers for $n = 4$ and see where 4 occurs.

  1. $1, 2, 3, 4$

2. $1, 2, 4, 3$

3. $1, 3, 2, 4$

4. $2, 1, 3, 4$

5. $2, 1, 4, 3$

The idea is that 4 *seems* to be near the end of the sequence only. Let's, for a moment, consider the ones that have 4 at the end.

1. $1, 2, 3, 4$

2. $1, 3, 2, 4$

3. $2, 1, 3, 4$

Notice that if we ignore the 4, these are actually the solutions for $n = 3$ (!). And that does make sense if you think about it: If your previous sequence was say $1, 3, 2$ that worked because you had the inequality $1 \leq 6 \leq 6$. Now when you add in 4 at the end, you're just adding in ≤ 16. And 16 is definitely going to be bigger than the last number in your inequality which is at most $3 \times 3 = 9$.

Exercise. Show in general that if you have a working permutation for $n-1$, appending n at the end of the permutation still allows it to satisfy the inequalities.

- **Observation 3**: Okay, so if 4 is at the end we're all good. But seems like there are cases where 4 isn't at the end. Let's take those now.

  1. $1, 2, 4, 3$

  2. $2, 1, 4, 3$

  Okay I have to admit that the sample set here is not very big - and in general it is better not to make too many conclusions based on a small amount of data - but trying to do the problem for $n = 5$ does not sound very appealing either, so let's work with this.

  In both of these 4 is in the second last position. Let's say x exists on the final position then. We need $4 \times 3 \leq x \times 4$ to be bigger than this. Note that this forces $x \geq 3$ but we also know $x \leq 3$ since 4 is taken. So in particular 3 must be at the last spot. What remains? Just the first two elements. And if we can create a working permutation for these two then we can add in a 4, 3 at the end of it and it still works (check this!).

- So in particular what we have got is that we can use solutions for $n-1$ to create one set of solutions, and the solutions for $n-2$ to create another set of distinct solutions. In particular, if the answer for n is $f(n)$, then

$$f(n) \geq f(n-1) + f(n-2)$$

Let's get back to our answers now: $1, 2, 3, 5$. It seems as if we in fact have

$$f(n) = f(n-1) + f(n-2)$$

To show this, we essentially want to confirm that n has to appear in the last two positions and can't appear anywhere else.

Problem. Show this! (This is not that easy, you'll essentially need to show that n can't appear at any position k such that $k \notin \{n-1, n\}$ so assume that it appears at such a position, and try to create a contradiction)

Hints: 36 171

By the way, this sequence, as you may know, is actually the *Fibonacci*: defined as

$F_0 = 0, F_1 = 1$ and for all $n \geq 2$,

$$F_n = F_{n-1} + F_{n-2}$$

Remark. One of the key things to take-away from this is that you should have from this problem I believe is the following:

> The point of seeing base cases is not necessarily figuring out the answers and finding the pattern in the answer. A lot of times, these answers can end up being misleading (there are a fair few examples of problems where I have gotten sidetracked into proving false claims because of very convincing base case answers!). But the sorts of observations we made for this problem would have been quite hard to make had we not played with the base cases a fair bit.

Example 3 (IMO 2012/3)

The liar's guessing game is a game played between two players A and B. The rules of the game depend on two fixed positive integers k and n which are known to both players.

At the start of the game A chooses integers x and N with $1 \leq x \leq N$. Player A keeps x secret, and truthfully tells N to player B. Player B now tries to obtain information about x by asking player A questions as follows: each question consists of B specifying an arbitrary set S of positive integers (possibly one specified in some previous question), and asking A whether x belongs to S. Player B may ask as many questions as he wishes. After each question, player A must immediately answer it with yes or no, but is allowed to lie as many times as she wants; the only restriction is that, among any $k+1$ consecutive answers, at least one answer must be truthful.

After B has asked as many questions as he wants, he must specify a set X of at most n positive integers. If x belongs to X, then B wins; otherwise, he loses. Prove that:

(a) If $n \geq 2^k$, then B can guarantee a win.

(b) For all sufficiently large k, there exists an integer $n \geq (1.99)^k$ such that B cannot guarantee a win.

For now, let's work with part (a) of the problem.

A reaction I've seen from a fair few people about this problem is

> "Why is the statement so long?"

I don't really have a great response but I assure you that a little extra reading is definitely worth it.

So moving on, let's try and understand what the problem is saying. So two players are playing a game. B needs to guess the number given by A. All you're given is a number N such that the number belongs to the set $\{1, 2, 3, \ldots, N\}$. Now you're going to ask him questions - "does the number belong to the set $\{1, 3, 4\}$?" for example. Now if he just keeps answering truthfully, you'll easily figure out what the number is (just check if it belongs to each of the single element subsets - or even better, binary search!)

The main issue is that A is allowed to lie - and not just lie but lie A LOT. In particular some k is fixed, and he has to say the truth only once in $k+1$ turns. So the problem in fact is telling you that it is highly unlikely that you'll be able to figure out the exact answer - but what you can seem to do is give out a subset of a "small" number of integers and guarantee that x belongs to that subset. In particular, you need to show that you can create a subset of size at most 2^k.

First thoughts: what if we just keep asking the same question again and again - he'll have to tell us the truth eventually, right? Well, not exactly. If you hear the same answer for the same question $k+1$ times consecutively, you're sure that it's the truth. But that's not going to happen. If $k = 3$, for example, there's a good chance you'll hear "Yes" twice and then "No" twice - you have no idea whether the "Yes" was a lie or the "No" was a lie. In fact, even if you have something like "No" thrice and "Yes" once - he could very well have chosen to say the truth thrice to confuse you - so you don't actually get any information either way (he has to say the truth at least once in $k+1$ turns, not exactly once).

Okay, now what? Too many variables! Let's get back to what we're best at - analysing the *base cases.* In particular, let's try to figure out what happens when $k = 1$ (as it turns out, this is not as easy as it may look.)

The upside to having harder base cases is that they generally leave you with a lot more intuition about the problem itself. Anyway, so $k = 1$. We have a set of N numbers, and we need to reduce it to two numbers. If $N \leq 2$ we're done already - so let's just say $N = 3$. The idea is that you want to somehow eliminate one of these three numbers. Meanwhile A needs to make sure that he says the truth at least once in any two consecutive turns.

So let's ask some sets of questions

- $\{1,2,3\}$ at any point: We all know that the answer has to be Yes, so he might as well say the truth (in fact he must, or you would know that he's going to have to say the truth on the next move). Similarly if you ask $\{\}$ at any point, A would just say No.
- $\{1\}$, and then $\{1\}$ again: He's likely to say Yes once and No the other time - so this doesn't really help us much at all, does it?
- $\{1\}$, and then $\{2,3\}$: Asking $\{2,3\}$ is essentially equivalent to asking $\{1\}$ (why?) so this doesn't help us much either.
- $\{1\}$ and then $\{2\}$: If you think about it, this (and symmetric cases of this) is the only pair from which you can ask hoping to extract some information. There are a couple of possibilities of what could happen here

  1. $\{1\}$: Yes, $\{2\}$: Yes

     We know that one of these two has to be the truth - so in either case, the answer can't be 3 (or else both of these would have been lies). So we can confirm that the answer is in set $\{1,2\}$

  2. $\{1\}$: Yes, $\{2\}$: No

If the answer was 2, both of these statements are lies - so we can confirm that the answer is in $\{1,3\}$

3. $\{1\}$: No, $\{2\}$: Yes

   If the answer was 1, both of these statements are lies - so we can confirm that the answer is in $\{2,3\}$

4. $\{1\}$: No, $\{2\}$: No

   And... we're stuck. Indeed, whatever the answer is, at least one of the above two statements is going to be the truth so we can't confirm anything.

So we've gotten somewhat close, but we're stuck in this case where he keeps saying No to you. The key idea is: can you force him to say yes?

Exercise. Show that by asking him $\{1\}$ twice, you can either force him to say Yes as an answer for it or essentially give us the information that 1 isn't the answer.

So we're done when $N = 3$. What if $N = 10000$? As it turns out, that's not hard at all.

Exercise. Show that you can do the above algorithm repeatedly until you're left with only two possibilities.

So we've gotten a fair bit of intuition of what's going on, but it's probably not time yet to try and create some big generalisation - let's try $k = 2$. This time, we need to reach a set of at most 4 numbers - so let's say $N = 5$ and we need to eliminate one of the five numbers.

- Once again, we can force the first answer for a non-empty subset to be "Yes" (use the same set repeatedly, if he says "No" 3 times in a row, you've eliminated all the numbers in the set).
- So we essentially have our first question as something with the answer Yes, and then two other questions - creating four possibilities. Our hope is that we eliminate a number in each of those four cases.
- What if we ask $\{1\}, \{2\}, \{3\}$. Here's a bit of intuition on why this can't work: you don't have the capability of forcing the second answer to be Yes (convince yourself about this!).

  In particular, notice that the number surely cannot be both 2 and 3, so if you receive "No" from each of $\{2\}$ and $\{3\}$ you can't hope to gain some information since one of the statements is guaranteed to be true anyway.
- Let's anyway fix our first subset as $\{1\}$ and try to handle this double "No" case. We want some number to get eliminated if both the answers are No.

How do we do this? You want that if x is the answer then all the statements are lies. So $x \neq 1$, and x must belong to both the subsets. Fine, so we put in, say 2, into both the second and third subset.

- So how about $\{1\}, \{2\}, \{2,3\}$. The issue is the Yes, No, Yes case. It is easy to see that no matter what x is, at least one statement is true. In particular, if $x = 2$ then the third statement is true else the second statement is true.
- Let's once again try to manually fix this: the Yes, No, Yes case. We want some x such that $x \neq 1$, x is in the second subset, but it is not in the third: so let's try $\{1\}, \{2,3\}, \{2,4\}$.

  Exercise. Check that in each case, at least one number gets eliminated (if all are "Yes", 5 gets eliminated).

Woah, so we've solved $k = 2$ too! Not bad? I think $k = 2$ gives a pretty big chunk of intuition of what's going on - especially if you didn't brute force your way to the answer. You essentially want to fix the first set as $\{1\}$ with answer yes, and then for each of the 2^k possibilities of the next k answers, you want to be able to reject one element. In particular, if the answers are say Yes, No, No, Yes, No: you need an element that is in the first and fourth subsets but not the other three - you need such an element for each such "subset of subsets".

Exercise. Read this again until it starts making sense and use this to construct the final solution.

7.3 Problems

Not too many problems here, but the ideas in this chapter are ones you want to apply *everywhere*.

Problem 1 (2016 ISL C1). The leader of an IMO team chooses positive integers n and k with $n > k$, and announces them to the deputy leader and a contestant. The leader then secretly tells the deputy leader an n-digit binary string, and the deputy leader writes down all n-digit binary strings which differ from the leader's in exactly k positions. (For example, if $n = 3$ and $k = 1$, and if the leader chooses 101, the deputy leader would write down 001, 111 and 100.) The contestant is allowed to look at the strings written by the deputy leader and guess the leader's string. What is the minimum number of guesses (in terms of n and k) needed to guarantee the correct answer?

Hints: 341 309 394 250 440

Problem 2. (India EGMO TST 2022/5) Let k be a positive integer. A sequence of integers $a_1, a_2, \cdots$ is called k-pop if the following holds: for every $n \in \mathbb{N}$, a_n is equal to the number of distinct elements in the set $\{a_1, \cdots, a_{n+k}\}$. Determine, as a function of k, how many k-pop sequences there are.

Hints: 334 3 78 456 23

Problem 3 (2022 ISL C1). A ± 1-sequence is a sequence of 2022 numbers $a_1, \ldots, a_{2022}$, each equal to either $+1$ or -1. Determine the largest C so that, for any ± 1-sequence, there exists an integer k and indices $1 \leq t_1 < \ldots < t_k \leq 2022$ so that $t_{i+1} - t_i \leq 2$ for all i, and

$$\left| \sum_{i+1}^{k} a_{t_i} \right| \geq C.$$

Hints: 271 532 133 357

Problem 4 (2022 IMO/1). The Bank of Oslo issues two types of coin: aluminum (denoted A) and bronze (denoted B). Marianne has n aluminum coins and n bronze coins arranged in a row in some arbitrary initial order. A chain is any subsequence of consecutive coins of the same type. Given a fixed positive integer $k \leq 2n$, Gilberty repeatedly performs the following operation: he identifies the longest chain containing the k^{th} coin from the left and moves all coins in that chain to the left end of the row.

For example, if $n = 4$ and $k = 4$, the process starting from the ordering $AABBBABA$ would be $AABBBABA \to BBBAAABA \to AAABBBBA \to BBBBAAAA \to \ldots$

Find all pairs (n, k) with $1 \le k \le 2n$ such that for every initial ordering, at some moment during the process, the leftmost n coins will all be of the same type.

Hints: 580 316 528 35 105

Problem 5 (2019 ISL C1). The infinite sequence $a_0, a_1, a_2, \ldots$ of (not necessarily distinct) integers has the following properties: $0 \le a_i \le i$ for all integers $i \ge 0$, and

$$\binom{k}{a_0} + \binom{k}{a_1} + \cdots + \binom{k}{a_k} = 2^k$$

for all integers $k \ge 0$. Prove that all integers $N \ge 0$ occur in the sequence (that is, for all $N \ge 0$, there exists $i \ge 0$ with $a_i = N$).

Hints: 285 59 373 400 329 63

Problem 6 (LMAO 2023/5). There is a row of 2022 identical boxes, each with one coin inside it. On your turn, you are allowed to open any box and if it contains a coin, take it. After you are done, the game director will secretly swap the box you opened with one of its neighboring boxes. Determine the maximum number of coins you can guarantee to collect in finite time.

Hints: 541 396 **Soln:** Page 277, Solution 14

8 Algorithms

An algorithm is a set of finite steps to perform a certain procedure and reach a certain end goal. In particular, most algorithms require some input data, which is processed into a certain form of required output. Algorithms surround us wherever we go - google search, finding the optimal way to do something, minimising financial risks and so on. Finding efficient algorithms has always been a very important question in mathematics and computer science and forms a significant part of olympiad combinatorics as well.

8.1 Introduction

In general, many algorithmic problems in olympiads look like the following:

> **Example**
> You have been given ... and are allowed to do one of the following operations: .... Show that you can perform some sequence of operations with each operation being one of the above ones, so that at the end, you end up with ....

Some problems can have more conditions, e.g. your sequence must contain at most $n-1$ operations or you can use the first operation only once. Some problems also ask you to find a necessary and sufficient condition for the algorithm to exist. These problems have 3 parts

- Convincing yourself that the condition is some C (!!)
- Proving that for all cases where C doesn't hold - an algorithm cannot exist
- Finding an algorithm whenever C holds.

Our biggest tool to solving such problems is going to be experimentation, and trying to find patterns in strategies that are initially made by trial and error. That being said, there are a fair few general ideas and techniques that we should keep in mind while trying to solve such problems and can help you solve the problem - greedy algorithms for example.

Here's an example to get things going.

Example 1 (Based on United Kingdom 2011)

You have been given an $m \times n$ grid of positive integers on a board. You are allowed to perform the following 2 operations:

(a) Subtract a positive integer k from all the numbers in a single row

(b) Double all the numbers in a particular column.

Show that these two operations can be applied in some order so that all the integers on the board end up as 0.

If you remember the problem we did in the chapter on Soft Techniques, this is actually quite similar to that. There were only two numbers there but you have an entire grid, with somewhat similar options but not really. In particular, there we solved the problem for the special case: $m = 1, n = 2$.

Exercise. In case you've forgotten how to do it for a 1×2, try solving it on your own this time before moving forward!

Before trying to handle something like a 100×100 grid, let's try something simpler: in particular, how about 1×3. Let's take some examples: how about $(1, 1, 5)$. The idea is that you can multiply the first two numbers by 2, and then subtract 1 from all of them to get $(1, 1, 4)$. Repeating this we can get to $(1, 1, 1)$ at which point subtracting one from each of them finishes. Okay, what about $(1, 2, 5)$? Let's temporarily ignore the bigger number (5 in this case) and try to bring 2 to 1. In particular we can do

$$(1, 2, 5) \to (2, 2, 5) \to (1, 1, 4)$$

And then finish like earlier.

So in general if you have, say, $(1, 3, 5, 7)$ the idea is to make the second number equal to 1. Once this is done you make both of them 2 and subtract one from everything until the third becomes 1 too. Repeat until all become 1, and then subtract 1 from everything. Tada.

For a more concrete example:

$$(1, 3, 5, 7) \to (2, 3, 5, 7) \to (1, 2, 4, 6) \to (2, 2, 4, 6) \to (1, 1, 3, 5) \to (2, 2, 3, 5) \to$$
$$(1, 1, 2, 4) \to (2, 2, 2, 4) \to (1, 1, 1, 3) \to (2, 2, 2, 3) \to (1, 1, 1, 2) \to (2, 2, 2, 2) \to$$
$$(1, 1, 1, 1) \to (0, 0, 0, 0)$$

Exercise. Show that you can actually solve the rows one by one! So for a general $m \times n$, if you make the first row filled with zeroes, you can do think of it as a $(m - 1) \times n$ as multiplying by 2 does nothing to it and we're not going to be subtracting anything from that row anymore.

8.2 Beyond Invariants

Example 2 (Tuymaada 2018/J6)

The numbers $1, 2, \ldots, 1024$ are written on a blackboard. The following procedure is performed ten times: partition the numbers on the board into disjoint pairs, and replace each pair with its nonnegative difference. Determine all possible values of the final number.

The first thing to notice is 1024 and 10. Note that each time the operation halves the number of elements on the board, and so it makes sense that after 10 partitions there is only $\frac{1024}{2^{10}} = 1$ number left.

If you read carefully, we're essentially replacing a and b with $|a - b|$ a lot of times. Does this remind you of anything? Aha, the $\pm$ trick from the invariants chapter. We know that

$$a + b \equiv |a - b| \pmod 2$$

and so the sum of all numbers on the board never changes parity at any point. Thus we can conclude that the final number has to have the same parity as 512×1025 which is even. So the final number has to be even. After the first turn, the maximum value on the board is going to be ≤ 1023 so this even is ≤ 1022 and obviously non negative differences, are well, non negative. So we've reduced our possibilities to only even numbers from 0 to 1022. The question, though, is: can all such even numbers be *achieved*?

This is one of those necessary and sufficient problems: we need to determine all possible values - so we need to show that some values can be achieved, and that all others cannot be achieved.

But the funny thing is that often, neither of those is the key step to solving the problem. The key step is in fact the guess you make on what the answer is. You can't begin proving that some values can/cannot be achieved without first trusting your intuition and making a guess on what the answer is. This is a big leap of faith, and can only be found after a fair bit of experimentation like in the previous problem.

And the leap of faith here is that all even numbers from 0 to 1022 can actually be achieved.

At this point, we need to try and construct such numbers. So we need to actually perform some operations on these numbers from 1 to 1024 and end up with, say, 66. But 1024 is a pretty big number, no? Building on what we learnt in the previous chapter, let's see what happens if we try and solve the same problem but with 1024 replaced by 4.

So we start off with $\{1,2,3,4\}$. Our *guess* is that the we can reach all even numbers less than 4 - so 0 and 2. If we pair off adjacent terms, we get $\{1,1\}$ and end up with 0.

Notice that this technique works for 0 in general too. In the first turn, pair up adjacent numbers - and then you'll have $\{1,1,\ldots,1\}$ which will have to end up at 0 (why?). What if we pair off 1 with 4 instead? We get $\{1,3\}$ and this leads to 2.

But maybe 4 was a little too simple to give much insight, so let's try 8 numbers instead.

So we begin with $\{1,2,\ldots,8\}$. Clearly 0 is achievable by taking adjacent numbers. The key theme here is going to be that we want to keep things simple. Pairing off *almost everything* with adjacent numbers gives us almost all numbers equal to 1, and in the next step, almost all numbers are going to be zero. But this is vague - what does almost even mean here? Well, that's for you to figure out I guess.

Exercise. Find the algorithm for construction $2,4,6$ for $\{1,2,\ldots 8\}$ along the above lines. Once you do this, you should be pretty close to a solution that works for numbers till 1024 as well.

Hints: 54 294 445

Example 3 (Classic)

You are given a tokens of type A, b tokens of type B and c tokens of type C. You are allowed to perform the following operation on the tokens:

- Choose 2 chips of different types and replace them by a single chip of the third type.

Find a necessary and sufficient condition so that we can reach the configuration where exactly one token remains and that token is of type A.

If you've read the chapter on Invariants, I'm sure this problem feels familiar. However, the question there was to show that for a certain (a,b,c) the process cannot be completed - here we're asking precisely which (a,b,c) can work.

This problem is of the *if and only if* type - we need to show that in certain cases the required ending state cannot be reached, as well as that in certain cases the ending state can definitely be reached. Consider the tuple (a,b,c). The allowed algorithm is basically to subtract 1 from two of the numbers here and add 1 to the third. We want to end up at $(1,0,0)$.

The key observation we used in the Invariants chapter is to consider the process modulo 2, and notice that after a single move,

$$(a,b,c) \to (a+1,b+1,c+1)$$

(as $c - 1 \equiv c + 1 \pmod 2$)

Aha! so the only possible ending states are (a, b, c) and $(a+1, b+1, c+1)$ (why?). Since we want to end at $(1,0,0)$ the only possibilities are if either of the following hold

$$(a, b, c) \equiv (0, 1, 1) \pmod 2$$
$$(a, b, c) \equiv (1, 0, 0) \pmod 2$$

Here's where the algorithmic part of the problem starts. We have discarded 6/8 of the possible answers but we now need to show that for these cases an algorithm does in fact exist.

Firstly notice that if we started off with $(5, 0, 0)$ there's nothing we can do. So we need to make a slight modification to our answer - atleast one of b, c must be greater than 0 for it to work (or we must begin with $(1,0,0)$ itself).

Note that in every move, $a + b + c$ reduces by exactly 1 so if we can keep applying the algorithm for long enough, we *must* reach $(k, 0, 0)$ for some k odd.

Exercise. Check that the previous sentence is true.

Hints: 263 121

A configuration like $(3, 0, 0)$ is the main one we need to be scared about. If not for it, we could have probably just done any random legal moves and everything would have worked out as we would eventually have reached $(1, 0, 0)$.

So we need to avoid reaching something like $(3, 0, 0)$. To reach $(3, 0, 0)$ we must have been at $(2, 1, 1)$ on the previous step. But here we could have instead chosen to convert this to $(1, 0, 2)$ (by reducing the first two and increasing the third) and now it all works out. In general, if we got stuck at $(2n + 1, 0, 0)$ for some $n \geq 1$ - on the previous step we would've been at $(2n, 1, 1)$. Then instead we can convert this to $(2n - 1, 0, 2)$.

There are many ways to finish using this idea. The most direct is to just solve the problem manually in the case that we reach $(2n, 1, 1)$. Now go to $(2n - 1, 0, 2)$ and then $(2n - 2, 1, 1)$. Eventually we'll reach $(0, 1, 1)$ at which point we can change this to $(1, 0, 0)$.

A cooler trick is to consider the following greedy algorithm: at any stage increase the minimum value while reducing the other 2 (break ties arbitrarily). Then if we were ever at $(2n, 1, 1)$ with $n \geq 1$ we would never choose to increase $2n$ and so we will eventually end up with $(1, 0, 0)$ - yay!

Remark. The *convincing yourself* step we just skip while writing a solution is actually really hard. What's the guarentee that the algorithm exists when the configuration is $(0, 1, 1)$ modulo 2? Often its just a leap of faith - a general idea is to try finding an algorithm for some time. If nothing seems to be working out, its probably because the algorithm doesn't exist at all and you need to discard more cases.

8.3 Induction

While solving algorithms, it is often very useful to try and *induct*, in some sense similar to how we approached the first problem. In general, if you see that solving the problem ever becomes equivalent to solving a smaller case of the same problem, induction should come to your mind.

Example 4 (China Girls Math Olympiad 2011)
There are n boxes $B_1, B_2, \ldots B_n$ from left to right, and there are n balls in these boxes. You are allowed to perform the following operations:

- If there is at least 1 ball in B_1, we can move it to B_2
- If there is at least 1 ball in B_n, we can move it to B_{n-1}
- If there are at least 2 balls in B_k, $2 \leq K \leq n-1$ we can move one to B_{k-1} and one to B_{k+1}

Prove that for any arrangement of the n balls, we can achieve that each box has one ball in it.

The main idea here is that if we can get B_n to have exactly one ball, we can ignore that ball and try to solve the problem for $n-1$ boxes. So if B_n originally has one ball, we'd be done.

In fact, if it has any $k \geq 1$ balls, we can just move $k-1$ to B_{n-1} using operation 2, and induct. So the only case left to handle is if B_n originally has zero balls.

So all we need to do is force a ball into B_n. For this, intuitively we want to keep applying operations 1 and 3 so as to move balls rightwards while we can. If we can't push balls at some point this would mean that the first $n-1$ boxes contain at most $0+1+1+\cdots+1 = n-2$ balls so in fact B_n has atleast 2 balls and we can push the extra ones back.

But there is one key detail that we've missed. What if we can just keep applying operations 1 and 3 forever, and the ball never reaches B_n? The question now is *why does the process even have to end?* This is a question we have previously answered in the invariants and monovariants chapter - and the trick was to find a monovariant which is bounded.

The total number of balls is invariant, but that isn't too useful. We need something that increases (or reduces) each time we make a move. Like some problems we saw in that chapter, we assign a *weight* w_i to the balls in each box B_i. So every time we move balls, the hope is that the summation of weights increases. For this to happen, we would need

$$w_2 > w_1$$

$$w_{k+1} + w_{k-1} > 2w_k$$

(why?). But this is easy to ensure, we can just start off with $w_1 = 1$ and then keep picking weights that satisfy the previous conditions. If you want an actual construction, note that $w_k = 2^k$ works.

What does this imply? Well, the quantity $\sum_i 2^i \cdot b_i$ increases by atleast 1 (why?) at every step (where b_i is the number of balls in B_i) and is bounded above by say $2^n \cdot n$ and so the process has to end as required.

Exercise. Complete the induction. There can be a slight issue when you try to apply the $n-1$ boxes algorithm and try to operate on B_{n-1} - in the original algorithm it only passed a ball into B_{n-2} but now it passes a ball into B_n as well. How can we fix this? (it's not that hard)

Hints: 288

> **Remark.** The idea of considering a weighted sum is actually really useful. In particular, thinking of binary representations and making inequalities which the w_i must satisfy are two ways to motivate many such sums.

Example 5 (ISL 1998 C3)

Cards numbered 1 to n are arranged at random in a row. In a move, one may choose any block of consecutive cards whose numbers are in ascending or descending order, and switch the block around. For example, if $n = 9$, 9 1 6 5 3 2 7 4 8 may be changed to 91 3 5 6 2 7 4 8. Prove that in at most $2n - 6$ moves, one can arrange the n cards so that their numbers are in ascending or descending order.

We have a total of $2n - 6$ moves to convert the number, and so if you're trying to induct and have sorted n of the numbers with the $n + 1$th being left, you would have to put this into place within 2 moves. Let's take an example for this: $n = 9$ and we have $\{9, 1, 6, 5, 3, 2, 7, 8, 4\}$. Lets say we sort the first 8 numbers in order (and say it appears in ascending order) so we have $\{1, 2, 3, 5, 6, 7, 8, 9, 4\}$. We need to fix this in two moves.

Instead of discussing the entire solution, in what follows I'll just discuss the key steps, and leave you to work out the details.

- Find a way to put 4 in place in two moves (this is not so hard).
- See what happens when the first 8 numbers appeared in descending order instead, and generalise these ideas for n and k.
- Prove the induction base case $n = 5$. This is tricky and a not-so-fun bash so you can skip it if you like.

8.4 Efficient algorithms

Till now, we've only been obsessed with being able to *find* an algorithm. In real life, there's no point finding an algorithm if it takes ages to actually perform it. In particular, there are a variety of problems which have a very simple solution - bruteforcing through all possibilities until you find something that works. Unfortunately, this usually takes too much time and is exponential in the size of the constraints.

As an example, let's consider the sudoku. There are 81 squares on it, and each can have 9 possibilities. So theoretically, checking each of the 9^{81} possibilities is enough to solve any given sudoku. However a computer can only do around 10^8 operations in a second (at the time of writing this, at least) and so actually computing that value would take far more centuries than there are atoms in the solar system!

In general, an efficient algorithm is one which is *polynomial* in the size of the constraints. (In particular, something like 81^2 or 81^3 would have been far more reasonable than 9^{81}.)

While these ideas are not often *directly* relevant in math olympiads, they form the crux for informatics olympiads and do help a fair bit with intuition to algorithms in math olympiads as well.

8.5 Big-O notation

Efficient algorithms is still a pretty *qualitative* term and we would like something that helps us understand the complexity of the algorithm quantitatively. For this, we take help of the notation $\mathcal{O}(f(n))$.

Intuitively, we say that an algorithm takes $\mathcal{O}(f(n))$ steps if the number of steps it requires to run is bounded by $C \cdot f(n)$ for some constant C independent of n. Here n is the size of the input (number of bits in the input to be precise but you can ignore these details).

So if your algorithm takes say, $2n^2 + 3n + 5$ steps to run, we can say that this algorithm has a complexity $\mathcal{O}(n^2)$ because the steps it takes is bounded by

$$10n^2 = 2n^2 + 3n^2 + 5n^2 \geq 2n^2 + 3n + 5$$

We're still not done though. What exactly do we really mean by a *step*? Well, the usual idea is that you can consider a *simple operation* as a single step. For example, adding two numbers is a step, finding the maximum of two numbers is a step (or just checking which is greater), finding the remainder when one number is divided by another is a single step and so on.

Example 6

You're given a list of n positive integers: $a_1, a_2, \ldots a_n$ and m queries such that each query is of the following form:

- Find the sum $a_i + a_{i+1} + \cdots + a_j$ for a particular $1 \leq i, j \leq n$

Find an algorithm that answers all of these queries in $\mathcal{O}(n + m)$ steps.

A naive algorithm would be the following: every time you see an i and j, sit down and calculate the sum $a_i + a_{i+1} + \cdots + a_j$. The issue is that if i is close to the start and j is close to the end of the list you're adding close to n numbers and therefore spending around $\mathcal{O}(n)$ time on this query. If you get unlucky and this happens every time, you'd have to spend a total of $\mathcal{O}(mn)$ steps, which is quite inefficient.

Instead, you want some algorithm which spends close to constant time answering any query. The key observation is that we can precompute certain sums, using which all sums become easy to compute. Indeed, if we calculate all the *prefix sums*, ie. the sum S_k of the first k elements of the list for all k, all we need to do now is $S_j - S_{i-1}$ which takes close to constant time. Computing the n sums takes $\mathcal{O}(n)$ time at the start (how?) and so all queries can be answered in $\mathcal{O}(n + m)$ steps.

Remark. Often in these types of questions, it is really easy to get a naive algorithm that solves the problem - finding one that is efficient enough is the tricky part of the problem.

8.6 Divide and Conquer

This section deals with problems where the main idea is to split a list of numbers in half, solve the problem independently for each of the halfs and then combine the results. This can be surprisingly useful for a variety of different problems.

Example 7 (Binary search)

Given a sorted list of positive integers, we need to check whether a given positive integer is present in the list. Find an algorithm to do so in at most $\mathcal{O}(\log_2 n)$ steps.

We could do this by just checking each element one by one and this takes $\mathcal{O}(n)$ steps. However, this isn't what we want. The important idea here is that set of integers is given to be sorted and we can use this to our advantage.

We check if the middle element in the list matches. If it does, we're done. If not, we check whether the element we're looking for is smaller or larger than the middle element. Either way, we have eliminated half of the list (how?) and so the number of steps we took is $f(n) = f(\lfloor n/2 \rfloor) + 1$ steps.

Exercise. Check by induction that the number of steps required is $\mathcal{O}(\log_2 n)$.

> **Remark.** We could have just written $\mathcal{O}(\log n)$ as the constant factor of $\frac{1}{\log 2}$ is irrelevant anyway.

Example 8 (Merge sort)

Given a list of n positive integers, find an algorithm that sorts it using at most $\mathcal{O}(n \log n)$ steps.

Okay, so going with the usual theme - let's say we sorted the first half and the second half of the list independently. The objective now, is to *merge* the two sets (hence the name merge sort).

So let's say I have lists $[1, 3, 7, 9, 11]$ and $[2, 5, 10, 14]$ and I want to merge these lists.

The first element has to be one of the first elements in each list - it's 1. The trick is that now you need to compare the second element of the first list with the first of the other - this time 2 is smaller, so you compare the second element of each list this time. And we repeat this exact process until all the numbers are done.

Exercise. Check that this takes $\mathcal{O}(n)$ steps.

Now all that is left is a simple induction. Assume that sorting a list of size $n/2$ takes $f(n/2) \leq c(n/2)(\log(n/2))$ steps. Hence,

$$f(n) \leq 2f(n/2) + cn = n \log n$$

And the induction is complete!

8.7 Problems

Problem 1 (2019 USAJMO/1). There are $a+b$ bowls arranged in a row, numbered 1 through $a+b$, where a and b are given positive integers. Initially, each of the first a bowls contains an apple, and each of the last b bowls contains a pear. A legal move consists of moving an apple from bowl i to bowl $i+1$ and a pear from bowl j to bowl $j-1$, provided that the difference $i-j$ is even. We permit multiple fruits in the same bowl at the same time. The goal is to end up with the first b bowls each containing a pear and the last a bowls each containing an apple. Show that this is possible if and only if the product ab is even.

Hints: 535 62 253 60 **Soln:** Page 278, Solution 15

Problem 2 (IMOSL 1989). A natural number is written in each square of an $m \times n$ chess board. The allowed move is to add an integer k to each of two adjacent numbers in such a way that non-negative numbers are obtained. (Two squares are adjacent if they have a common side.) Find a necessary and sufficient condition for it to be possible for all the numbers to be zero after finitely many operations.

Hints: 362 319 219 193

Problem 3 (Argentina TST 2011/2). A wizard kidnaps 31 members from party A, 28 members from party B, 23 members from party C, and 19 members from party D, keeping them isolated in individual rooms in his castle, where he forces them to work. Every day, after work, the kidnapped people can walk in the park and talk with each other. However, when three members of three different parties start talking with each other, the wizard reconverts them to the fourth party (there are no conversations with 4 or more people involved).

1. Find out whether it is possible that, after some time, all of the kidnapped people belong to the same party. If the answer is yes, determine to which party they will belong.

2. Find all quartets of positive integers that add up to 101 that if they were to be considered the number of members from the four parties, it is possible that, after some time, all of the kidnapped people belong to the same party, under the same rules imposed by the wizard.

Hints: 364 571 87 189

Problem 4 (Russia 2004, generalised). On a table there are $2n$ boxes, where n is a positive integer, and in each box there is one ball. Some of the balls are white and the number of white balls is even and greater than 0. In each turn we are allowed to point to two arbitrary boxes and ask whether there is at least one white ball in the two boxes (the answer is yes or no). Show that after $(4n-3)$ questions we can indicate two boxes which definitely contain white balls.

Hints: 140 658 81 381

Problem 5 (Canada 2018/1). Consider an arrangement of tokens in the plane, not necessarily at distinct points. We are allowed to apply a sequence of moves of the following kind: select a pair of tokens at points A and B and move both of them to the midpoint of A and B. We say that an arrangement of n tokens is collapsible if it is possible to end up with all n tokens at the same point after a finite number of moves. Prove that every arrangement of n tokens is collapsible if and only if n is a power of 2.

Hints: 557 642 149 84

Problem 6 (ISL 1994 C3). Peter has three accounts in a bank, each with an integral number of dollars. He is only allowed to transfer money from one account to another so that the amount of money in the latter is doubled. Prove that Peter can always transfer all his money into two accounts. Can Peter always transfer all his money into one account?

Hints. 648 201 448 354

Problem 7 (ISL 2022 C6). Let n be a positive integer. We start with n piles of pebbles, each initially containing a single pebble. One can perform moves of the following form: choose two piles, take an equal number of pebbles from each pile and form a new pile out of these pebbles. Find (in terms of n) the smallest number of nonempty piles that one can obtain by performing a finite sequence of moves of this form.

Hints: 521 347 53 514 579

Problem 8 (2018 IOI/1). Let $1 \leq N \leq 2000$ be a given integer. In a video game, a secret word S from the alphabet $\{A, B, X, Y\}$, is hidden from you. However, you know the length N of the string, and you also know that the first character of S appears only once.

In a *combo move*, you may press a sequence p of up to $4N$ buttons. The video game then responds immediately by giving you the length of the longest prefix of S which appears in p. Give an algorithm which, by making at most $N + 2$ combo moves, determines the sequence S.

Hints: 417 42 137 143 20

Problem 9 (IMO 2010/5). Each of the six boxes B_1, B_2, B_3, B_4, B_5, B_6 initially contains one coin. The following operations are allowed:

Type 1) Choose a non-empty box B_j, $1 \leq j \leq 5$, remove one coin from B_j and add two coins to B_{j+1} Type 2) Choose a non-empty box B_k, $1 \leq k \leq 4$, remove one coin from B_k and swap the contents (maybe empty) of the boxes B_{k+1} and B_{k+2}.

Determine if there exists a finite sequence of operations of the allowed types, such that the five boxes B_1, B_2, B_3, B_4, B_5 become empty, while box B_6 contains exactly $2010^{2010^{2010}}$ coins.

Hints: 551 311 194 496

Problem 10. You are playing a game with the Devil. There are n coins in a line, each showing either H (heads) or T (tails). Whenever the rightmost coin is H, you decide its new orientation and move it to the leftmost position. Whenever the rightmost coin is T, the Devil decides its new orientation and moves it to the leftmost position. If you can make all coins face the same way within 2^n moves, you win - else the devil wins. Show that you have a winning strategy, i.e, no matter what the devil does, you can find a way to make all coins heads or all coins tails.

Hints: 574 2 414 320 **Soln:** Page 279, Solution 16

Problem 11 (USACO 2022 US Open Contest, Silver P3). Bessie finds a string S of length n containing only the three characters 'C', 'O', and 'W'. S is *nice* if it's possible to turn this string into a single 'C' (her favorite letter) using the following operations:

- Choose two adjacent equal letters and delete them.
- Choose one letter and replace it with the other two letters in either order.

Find a $\mathcal{O}(n)$ algorithm that checks if S is nice.

Hints: 605 421 213 46

Problem 12. For the previous problem, let's say you now also have to process q queries, each of which contains two integers l and r, and output whether the substring S' of S from the lth to rth character is *nice*. Find an algorithm of complexity $O(n+q)$ that solves this problem.

Hints: 332 473

Problem 13 (Russia 2000). Tanya chooses a natural number $X \leq 100$, and Sasha is trying to guess this number. She can select two natural numbers M and N less than 100 and ask for the value of $\gcd(X+M, N)$. Show that Sasha can determine Tanya's number with at most seven questions (the numbers M and N can change each question).

Hints: 220 129 307 237

Problem 14 (Codechef Starters 92 P4, easy version). You are given a positive integer k and an array of N positive integers, with the maximum value being M. In a move, you're allowed to subtract one from each of k integers in your array. If at most $k-1$ of the integers in the array are positive, the process ends. Find a necessary and sufficient condition on the original array which allows you to perform M such moves.

Hints: 273 116 475

Problem 15 (Codechef Starters 92 P4, harder version). You are given a positive integer k and an array of N positive integers from the range 1 to M. In a move, you're allowed to subtract one from each of k integers in your array. If at most $k-1$ of the integers in the array are positive, the process ends. Find a $O(N \log M)$ algorithm that outputs the maximum possible number of moves you can make.

Hints: 268 281 433

Problem 16 (HMIC 2019/2). Annie has a permutation $\pi \in S_{2019}$ which Yannick wants to guess. Each turn Yannick provides a function $f\colon \{1, 2, \ldots, 2019\} \to \{1, 2, \ldots, 2019\}$ and Annie tells Yannick the number of indices $i \in \{1, 2, \ldots, 2019\}$ such that $\pi(i) = f(i)$. Show that Yannick can always guess Annie's permutation with at most 24000 guesses.

Hints: 70 73 216 229

Problem 17 (2016 IOI/5). Let $n = 128$. A permutation $p\colon \{1, \ldots, n\} \to \{1, \ldots, n\}$ is fixed but hidden from you. You may do the following interactive procedure:

- First, you may specify up to $w = 896$ binary strings of length n, to obtain a set S.
- Then, the permutation p is applied on every element of S to the bits of the string. For example, if $p(1) = 3$ then the first bit of each string becomes the third bit, etc. The resulting set is denoted S'.
- Finally, you may make up to $r = 896$ queries about whether a given binary string is in S'. Responses are immediate and you can base future queries on the results of preceding queries.

Give an algorithm to determine p.

Hints: 423 96 494

9 Greedy Algorithms and the Extremal Principle

9.1 Greedy Algorithms

Greedy algorithms form a very interesting field in combinatorics and computer science. The main ideology by which greedy algorithms work are:

> "I don't really care about whether it was a good decision long term, this option *seemed* nice so I chose it."

Mind you, when I say *seemed nice*, I may not be going to be picking some random element - we'd land ourselves in the probabilistic chapter then. Instead, I'm going to pick a very specific object which is some sort of maximum or minimum in something.

There are multiple *types* of problems that one can categorise greedy problems into. Some of them are ones where you get a method and it gives you an answer and the method is pretty efficient - but you have no idea whether what you did was *optimal*. Proving that your greedy algorithm works (in addition to finding it!) forms the crux of the problem in such cases. In other places, you don't even know if your greedy algorithm is ever going to terminate. If it does terminate, you're happy but you need to show that you don't go into some sort of endless cycle and never reach a result at all.

The main thing to keep in mind while learning greedy algorithms is that they need not always work. There is a large set of problems where you just *cannot* solve the problem using a greedy algorithm.

"Why should I bother learning them then? I should just go and do some more useful technique then, shouldn't I?" I never said they're not useful. Infact, they're really useful. Besides, even when they don't work, they often provide us with some intuition about the problem and what ideas could work.

9.2 Proving Optimality with Greedy Algorithms

Example 1

You are given a set of n activities. Performing the ith activity takes a_i minutes. Find an $\mathcal{O}(n \log n)$ algorithm which finds the maximum number of activities that you can attend if you only have a total of N minutes.

We want to pick as many activities as possible, but we also want to spend at most N minutes on these activities. Logically, the shortest activity seems like the most reasonable choice. The reasoning is that if you're taking an event, you might as well take the shortest event as all others are giving you the same result but *wasting* some extra time. Okay, let's say I'm done with the shortest event. Now what? Seems like the next best bet is to pick the second shortest event. In general, it seems like picking the k shortest events is the most reasonable way of picking k events.

So we set the following greedy algorithm: keep taking the shortest currently available event under the condition that the total time spent is at most N. If the event you want to select makes the total time exceed N, stop right there and output the number of events you've visited now.

This has given an answer. And sure, it seems logical - intuitively seems like the best thing to do. But without proof, how can we be sure that the answer we got was optimal?

So here comes the important (and often neglected) part: let's prove that the answer we got *has to be* optimal. The general idea to do this is assume for the sake of contradiction that the true answer was $\geq t+1$ while our answer was t.

Let's say we've sorted the array so that $a_1 \leq a_2 \leq a_3 \leq \cdots \leq a_n$. Then we must have had $a_1 + a_2 + \cdots + a_t \leq N$ but $a_1 + a_2 + \cdots + a_t + a_{t+1} > N$. Now apparently the answer is $\geq t+1$ so there is some set of more than t a_i with sum S at most N. The idea is that $a_1 + a_2 + \cdots + a_t + a_{t+1} \leq S$ (why?!). Clearly this means that $S > N$, a contradiction.

Exercise. Check that $a_1 + a_2 + \cdots + a_t + a_{t+1} \leq S$.

> **Remark.** The general idea is that you want to show that the *official answer* cannot be better than our answer - as we do "at least as well."

A much harder but somewhat similar greedy problem is as follows:

Example 2 (Oxford MAT)

You are given a list of items. Each item has a weight, and a strength. You have been tasked with placing these items one above another. However the condition is that the strength of each item should be at least the sum of weights of all items above it. Find an efficient method to check if such an ordering exists, and output one such ordering.

First guess: why not put the heaviest at the bottom? And the next heaviest above and so on. The issue is that there's a possibility where the heaviest item has very low strength, while the other items are relatively lighter but have strength high enough to handle the heaviest item. In particular, imagine a scenario with just 2 bags. One bag has weight 100 and strength 1. The other bag has weight 2 and strength 150. Clearly the right thing to do would be to put the bag with weight 2 below, contrary to our guess.

Okay, seems like the greatest strength is what matters then? How about we place the the bag with greatest strength at the bottom? There are still a couple of issues. Imagine a scenario with 2 bags again: one bag has weight 1 and strength 10. The other bag has weight 20 and strength 2. You clearly can't put the first bag at the bottom since $10 < 20$, but you could have put it at the top since strength 2 is enough to handle its weight 1.

Okay, so we have two failed guesses now. At this point we begin to question life and whether there was a greedy algorithm in the first place at all. The key insight is that we were not wrong in thinking that weight and strength mattered. We were just not *right enough*. In particular, *both* of them matter - so your metric for the greedy algorithm has to depend on both of them - not just one.

In such cases where a metric isn't directly clear - we can try and reverse motivate the entire argument. Ideally, our solution would be: "Look, if you sort by this it all works out." But what do we really mean by "it all works out"? A general cool way to prove the optimality of your greedy algorithm goes as follows:

Let's say my algorithm isn't optimal: so there's a better "solution" that didn't involve sorting by the function I was thinking of, or perhaps doesn't involve sorting at all. Compare "your" solution to the "correct" one. Let's say the bottom k items are the same in both, but the $k+1$th item is different.

So "our" $(k+1)$th item appears somewhere above. If we can somehow manage to swap this item with items directly below repeatedly while maintaining the niceness of the configuration (no item should be stressed more than it can handle), we can bring this to the $(k+1)$th place, and the configuration is still nice. Repeating this, we can essentially get all the items to the "right" positions: showing that our solution satisfies the constraints as well as the "correct" one.

The main step however is this swapping step. Obviously, you can't always maintain the niceness by swapping any two (or else literally every configuration would have worked). This is where the *metric* comes in. So you essentially want to create some inequality that ensures this swap is possible: and then use this inequality to *define* your metric. Since the $(k+1)$th element has a higher value for this metric than all the ones around it so you can keep doing this swap until it comes next to the kth highest element - at which point you can't repeat it.

It's okay if this was all a bit unclear, try going forward and then coming back to this part after you're done with the problem.

So essentially the key take away for now is that we want to find out when it is possible to swap two adjacent items while maintaining the *niceness*. Let's say you have n items with item 1 at the bottom and item n at the top, and you want to swap items i and $i+1$. Firstly, for items above $i+1$, this change makes no difference: they don't care about things below them. For the items below i, only the sum of the weights matters so this exchange makes no difference to them too. It's therefore sufficient to check that i and $i+1$ still work out nicely.

So let's say the sum of weight of items $i+2$ to n is W. Let w_k denote weight of the kth item and s_k denote the strength of the kth item. Since the original configuration was nice, we have that

$$s_{i+1} \geq W \text{ and } s_i \geq W + w_{i+1}$$

On the other hand, we need

$$s_{i+1} \geq W + w_i \text{ and } s_i \geq W$$

The second part works out easily - it's the first one that may not always hold. We need

$$s_{i+1} \geq W + w_i$$

We know that

$$s_i \geq W + w_{i+1} \implies s_{i+1} \geq W + w_i + (w_{i+1} - w_i + s_{i+1} - s_i)$$

So it's sufficient to have that

$$(w_{i+1} - w_i + s_{i+1} - s_i) \geq 0 \implies w_{i+1} + s_{i+1} \geq w_i + s_i$$

And there we have it, our metric: $w_i + s_i$. If we sort by $X_i = w_i + s_i$ and put the item with largest X_i at the bottom and so on, this is optimal - so if there is *any* solution - this has to work too. However it is quite possible that there's no solution at all - so you just need to check if this ordering you got using sorting by X_i works or not.

This was quite messy, so I've put down the solution below once again more neatly.

Solution: Sort the items in order of $w_i + s_i$. The claim is that if this configuration does not work (which can be checked quite easily), no ordering of items can work. To prove this, let's say the items in order of $w_i + s_i$ be $x_1, x_2, x_3, \ldots, x_n$ where x_1 has maximum $w_i + s_i$.

Now let's say for the sake of contradiction that the ordering with $x_1, x_2, \ldots, x_n$ with x_1 at the bottom doesn't work, while some other ordering $x_{a_1}, x_{a_2}, \ldots, x_{a_n}$ works where the sequence a_i forms a permutation.

So let's say x_1 appears at position t in this sequence. Note that we can exchange it with the box below it repeatedly since it has a higher $w_i + s_i$ (and that is the condition needed to swap 2 items). We do this until it is now at the bottom.

Repeat this until $t = a_t$ and then start swapping the $(t+1)$th item and so on. We end up with $x_1, x_2, \ldots x_n$ so if there was a solution, this should be a solution too.

> **Remark.** The two key things to take away here as per me are:
>
> - Sometimes when the path isn't too clear, try working backwards.
> - Swapping "adjacents" often does not end up impacting too many objects, so finding out the condition under which objects can be swapped could be pretty useful.

Example 3 (Activity Selection Problem)

On a certain day, there are n events that you would like to attend. Each of these events has a start time s_i and a finish time f_i. Some of these events may overlap and you are only allowed to attend events with no overlap. Provide an efficient algorithm that lets you maximise the number of events you attend (completely).

Clearly we could have just brute-forced through each of the 2^n possibilities and seen which works with maximum number of events. However this is exponential in the number of events and is therefore very inefficient. Instead, we need an algorithm that can at most require a number of steps that is polynomial in n.

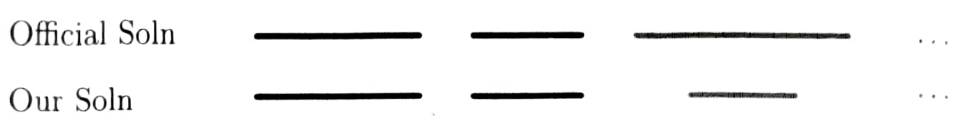
The first guess is a greedy algorithm that picks the shortest event followed by the next shortest event that can still be picked (and so on). However, this doesn't work.

Exercise. Try to figure out why (for a hint, see the diagram above)

Okay, next guess: how about a greedy algorithm that tries to pick the event with minimum s_i (that can be picked) at each point (so the even that starts first, etc.)?

Exercise. Show that this doesn't work either

Let's work backwards once again. Assume that the first k events picked by us and picked by the "official solution" are the same. We need to figure out a condition under which we can easily swap their $(k+1)$th event with ours. If we want to do this without troubling the events after us, what we need is that our event finishes before theirs. That way, we can always take the events they took. In particular, we can exchange their $(k+1)$th event with ours in the official solution and everything should still work.



Aha, so at every point, take the event that finishes first (and can be picked).

Exercise. Prove formally that this works using the official solution trick.

9.3 Is it good enough?

The idea here is going to be the following: the problem asks us to show that we can select some ten objects satisfying some property. Now we create a greedy algorithm that keeps selecting objects while making sure that the property is preserved. Since we kept making arbitrary choices at each point, there's a good chance that the list of objects we end up with before getting stuck isn't optimal. All we need, though, is for it to be *good enough*. Not very clear? Let's understand through the following problem.

Example 4

Suppose 4951 distinct points in the plane are given such that no four points are collinear. Show that it is possible to select 100 of the points for which no three points are collinear.

Pick two arbitrary points. Now look for a point that isn't collinear with these two. Once you've found that, look for a fourth point not on any of the $\binom{3}{2}$ possible lines. Then look for a fifth and so on, until you get stuck - aka, *repeat until stuck* (as Evan Chen says in the OTIS excerpts).

We now try to understand the condition that made us "get stuck". So let's say we currently have n points with us. We know that these form $\binom{n}{2}$ lines and that on each of these lines at most one other point can exist (or four points become collinear). So,

$$4951 - n \leq \binom{n}{2}$$

This is enough to force $n \geq 100$ as required. Done!

Example 5 (IMO 2003/1)

Let A be a 101-element subset of $S = \{1, 2, \ldots, 10^6\}$. Prove that there exist numbers $t_1, t_2, \ldots, t_{100}$ in S such that the sets

$$A_j = \{x + t_j \mid x \in A\}, \qquad j = 1, 2, \ldots, 100$$

are pairwise disjoint.

Let's say

$$A = \{a_1, a_2, \ldots, a_{101}\}$$

Once again, keep picking t_i one by one while possible. Let's say we got stuck. So we have $t_1, t_2, \ldots, t_n$ and we know that no matter what t we pick now, it gives some overlap, or more formally

$$a_i + t = a_j + t_k \implies t = a_j - a_i + t_k$$

There are $101 \times 100 \times n$ possibilities for the RHS but $10^6 - n$ for the LHS so

$$10^6 \leq 10101 \times n$$

But if $n \leq 99$ the RHS is at most 999999, a contradiction!

Remark. A slight optimisation of sorting the t_i and then picking them can improve this bound of 100 by a factor of two!

Example 6 (HMMT 2018 T6)

Let $n \geq 2$ be a positive integer. A subset of positive integers S is said to be *comprehensive* if for every integer $0 \leq x < n$, there is a subset of S (possibly empty) whose sum has remainder x when divided by n. Show that if a set S is comprehensive, then there is some (not necessarily proper) subset of S with at most $n-1$ elements which is also comprehensive.

We'll create a new S' which is a subset of S. It originally starts off as an empty set - so currently only the remainder 0 can be created. At each point we decide whether to add in a particular element from S. If at least one new remainder can now be created modulo n we add it into our set, else we don't.

Exercise. Find the finish!

9.4 Does it end?

Example 7 (All-Russian Olympiad, 1961)

You are given an $m \times n$ grid of real numbers. Call an individual row or column a line. In a move you are allowed to select any line, and flip the signs of each number in this line. Prove that you can make a finite sequence of moves such that the sum of entries in any line is non negative at the end.

Okay, so we're given a grid, the operation is that you can flip all signs of a line, and you want to end up with a scenario where each line has non negative sum. There are 2 main steps to solving a problem with a greedy algorithm:

- Keep repeating some simple step while you can.
- Finding a monovariant which shows that the algorithm is finite and you don't end up repeating that step forever.

It's okay if this doesn't make much sense to you just yet, it'll be a lot more clear after a problem or two. So the first step is to just *repeat* a step. For us, this step is to flip the signs of a line.

In particular, the idea is that we can come up with the following procedure: if you see a line with negative sum, flip it, and keep doing this until there is no line with negative sum. The only issue is, there is no real reason as to why this algorithm should end. The problem is that when we flip a row, we have converted the row-sum to positive but we may have cause problems to a bunch of columns which earlier had a positive sum.

To solve this issue, our objective is to find some quantity which only increases or only decreases when we make an operation. Since there are only finitely many possible values for this quantity (one for each of the 2^{m+n} possibilities on the grid), we cannot keep increasing it forever, and so the algorithm must end. But if the algorithm ends, this means that there is no line left with negative sum and we're done.

In particular, that monovariant here is the sum of all numbers on the grid (!!). This increases each time (why?) but there are only finitely many possible values for the sum of all numbers and so at some point this process must end (when we encounter the largest possible sum across all possibilities, say).

> **Remark.** The key idea here is to realise that the sum of all numbers always increases. Finding the monovariant is usually the hardest part of solving such a question. Even though the monovariants end up seeming quite motivated later on, coming up with them is quite tricky and often takes a lot of time and skill.

> **Remark.** These sorts of problems and ideas are also closely related to the *extremal principle*, which rephrases these arguments to form much neater solutions (but also seemingly less motivated). The key idea is that you begin with the quantity being its maximum or minimum, and show that if you aren't done, you could have used your greedy algorithm to increase the quantity further. As an example, here is how an extremal solution to the above problem would look.

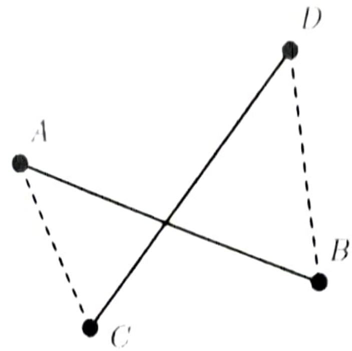
Solution. Across all possible 2^{m+n} choices of flippings, choose the flipping so that the sum of elements in the grid is maximised. If some row or column has negative sum, flipping it would have increased the sum further which is impossible. Hence each row and column has non negative sum in this choice of flipping. □

Example 8 (Putnam 1979)

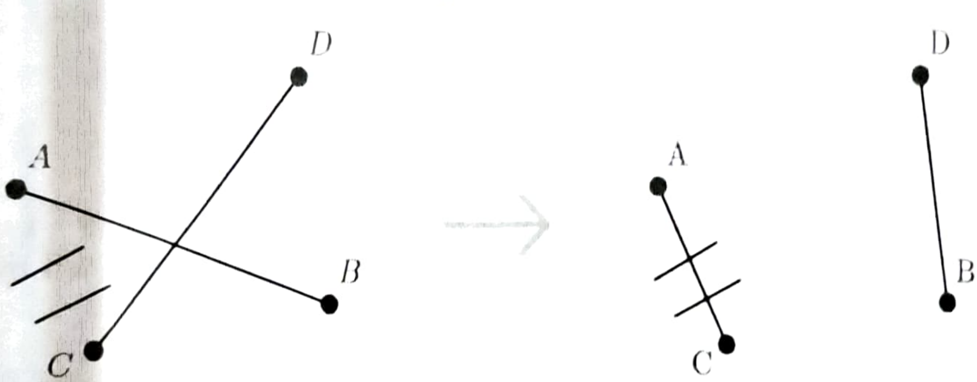
Given n red points and n blue points in the plane, no three collinear, prove that we can draw n line segments, each joining a red point to a blue point such that no segments intersect.

While this does not *appear* to be an algorithmic problem, it turns out to use identical ideas to the previous problem. Lets say, we begin by making these pairs randomly, and unfortunately some of the line segments end up intersecting.

The main idea is that if lines AB and CD intersect (where A, D are red) we can replace these with AC and DB, which cannot intersect.



The issue however, is the same as last time - yes, you fixed this problem, but how do you know you didn't make an additional 100 problems because AC ends up intersecting with some 100 other segments?



Once again, we are tasked with finding a monovariant. Well, what is the most famous inequality in the entirety of geometry? Of course, the triangle inequality. Notice that lengths $AC + BD < AB + CD$ (why?) and so we can set our monovariant to be the sum of the lengths of the n line segments!

> **Remark.** This solution is dangerously similar to the previous one - both of them follow a very similar theme. In general, if you see olympiad problems following a common theme, it is a good idea to note this down in your head, as it is quite possible the idea appears again, and this time you would be a lot more ready for it.

Exercise. How would you frame this solution using the extremal principle?

Example 9 (Graph Partition Lemma)

There are n students in a zoom call, some pairs of which are friends with each other. Show that the teacher can break them into two break out rooms so that each student has at least as many friends in the opposite room as in their own room.

Okay, so let's say we just begin with a random choice of breakout rooms, and then try to optimise our selection using a greedy algorithm.

By now, the idea should've fit into your head so I'll just leave the questions here, and you should try to figure it out.

- What should we keep *repeating*? **Hints**: 523
- The issue though is that while doing this repetition, we're changing the conditions for other people. To get around this, we need a monovariant. Find it. **Hints**: 525

9.5 Extremal Principle

As we've seen above, the extremal principle is a sort of *trick* to represent a normal greedy algorithmic solution in a slick way, and the solution is motivated by the algorithm anyway.

However, this is not entirely and necessarily true. For the problems we saw above, using the extremal principle seemed *unnatural* and wasn't intuitive. There are also many problems of the opposite flavour: where often the algorithm itself is motivated by guessing the quantity you're trying to maximise/minimise first.

This seems to be a little more risky as there is a much higher chance that the quantity you took may leave you with no information, but on the flip side, you get to skip the hard step of *finding* the monovariant by guessing it before even thinking of the algorithm - and so *solving* the problem takes a lot lesser time if you get lucky. Moreover, you don't end up wasting too much time even if your idea didn't work.

Example 10

There are 2000 points on a circle, and each point is given a number that is equal to the average of its neighbours. Show that all the numbers must be equal.

Let's consider the smallest number on the circle. It's the average of its two neighbours. Hmm.

Exercise. Finish the proof.

Hints: 324 This proof can be written in a greedy algorithm way too: something like for each number keep going to the neighbour that is at most the current value and repeat but this is a little bit messy to deal with and a lot less intuitive.

We'll see a couple more of such *extremal* problems in Graph Theory 2.

9.6 Problems

Problem 1. You are given denominations of $\$1, \$5, \$10, \$50, \$100$. You want to use these notes to get a total of $\$N$. However you want to minimise the number of notes you use. Find an algorithm that takes $\mathcal{O}(1)$ time to output the minimum number of notes required.

Hints: 397 **Soln:** Page 279, Solution 17

> **Remark.** $\mathcal{O}(1)$ means that the process takes at most C operations for some constant C. (We're assuming that performing usual operations like multiplication or division take constant time no matter how big the numbers are, which is slightly flawed but doesn't matter much practically.)

Problem 2. You are given denominations of $\$2, \$3$. You want to use these notes to get a total of $\$N$. However you want to minimise the number of notes you use. Show that you aren't as lucky this time. In particular, show that the greedy algorithm we used last time will not lead to a valid solution.

Hints: 613 **Soln:** Page 280, Solution 18

Problem 3. Let $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ be two sequences of real numbers. You are allowed to permute the two arrays to $c_1, c_2, \ldots, c_n$ and $d_1, d_2, \ldots, d_n$ (so the elements remain the same but their order changes). Find $\mathcal{O}(n \log n)$ time complexity algorithms that find:

1. The minimum and maximum possible values of $\sum_{i=1}^{n} |c_i - d_i|$
2. The minimum and maximum possible values of $\sum_{i=1}^{n} c_i d_i$

Hints: 262 530

> **Remark.** The second one here is the re-arrangement inequality!

Problem 4 (2019 Russia 10.2). Pasha and Vova play the following game, making moves in turn; Pasha moves first. Initially, they have a large piece of plasticine. By a move, Pasha cuts one of the existing pieces into three(of arbitrary sizes), and Vova merges two existing pieces into one. Pasha wins if at some point there appear to be 100 pieces of equal weights. Can Vova prevent Pasha's win?

Hints: 187 104 130 327 529

Problem 5 (2005 Russia). In a $2 \times n$ array we have positive reals s.t. the sum of the numbers in each of the n columns is 1. Show that we can select a number in each column s.t. the sum of the selected numbers in each row is at most $\frac{n+1}{4}$.

Hints: 545 412 16 370 646

Problem 6. Five hundred people attend a party in a mansion with many rooms. Each minute someone walks from one room into a different room with at least as many people. Prove that eventually all the people will be gathered in one room.

Hints: 267 531 583 30 **Soln:** Page 280, Solution 19

Problem 7. Let B and W be finite sets of black and white points with the property that every line segment that joins two points of the same color contains a point of the other color. Prove that both sets must lie on a single line segment.

Hints: 597 526 512

Problem 8. Each vertex of a regular 2023−gon is labelled with real numbers that the sum of reals is positive. Starting at some vertex, we write down the labels of the vertices reading counterclockwise around the polygon. Is it always possible to choose the starting vertex so that the sum of the first k reals written down is positive for $k = 1, 2, 3, \ldots, 2023$.

Hints: 497

Problem 9. Several positive integers are written on a blackboard. One can erase any two distinct integers and write their greatest common divisor and least common multiple. Prove that eventually the numbers stop changing.

Hints: 437

Problem 10. There are 2000 points on a circle, and each point is given a number that is equal to the average of its neighbours. Show that all the numbers must be equal.

Hints: 356 655

Problem 11 (IMO 1983). Is it possible to choose 1983 distinct positive integers, all less than or equal to 100000, no three of which are consecutive terms of an arithmetic progression?

Hints: 372 360 101

Problem 12 (IMOSL 2001, generalized). A set of three nonnegative integers $\{x, y, z\}$ with $x < y < z$ is called historic if $\{z - y, y - x\} = \{a, b\}$ for $0 < a < b$. Show that the set of all nonnegative integers can be written as the union of pairwise disjoint historic sets.

Hints: 284 367

Problem 13 (Romania). Given n points in the plane, show that you can choose at least $\lfloor \sqrt{n} \rfloor$ of them such that no three form an equilateral triangle.

Hints: 359 339 278

Problem 14. Given a set A of positive integers and a positive integer n, there is a subset B of $\{1, 2, \cdots, n\}$ with size at least $\dfrac{n}{2|A| + 1}$ such that the absolute value of the difference of any two distinct elements of B is not in A.

Hints: 258

Problem 15 (Moscow 1957). Let $1 = a_1 \le a_2 \le \cdots \le a_n$ be integers. Assume $a_{i+1} \le 2a_i$ for all $i = 1, \dots, n-1$ and $a_1 + \cdots + a_n$ is even. Prove that the n integers can be split into two piles with equal sum.

Hints: 600 621 314 513

Problem 16 (Putnam 2016 B3). Consider finitely many points in the plane such that, if we choose any three points A, B, C among them, the area of triangle ABC is always less than 1. Show that all these points lie within the interior of some triangle with area 4.

Hints: 65 110 94

Problem 17 (India TST 2003). Let n be a positive integer and $\{A, B, C\}$ be a partition of $\{1, 2, 3, \dots, 3n\}$ such that $|A| = |B| = |C| = n$. Prove that there exist $x \in A, y \in B, z \in C$ such that one of x, y, z is the sum of the other two.

Hints: 85 602 274 265 15

Problem 18 (Shortlist 2019 C2). You're given n blocks each with weight at least 1 and total weight $2n$. Prove that for every real number $r \in [0, 2n-2]$ there is a subset of the blocks whose total weight is between r and $r+2$ inclusive.

Hints: 275 468

Problem 19 (IMO 2014/5). For every positive integer n, the Bank of Cape Town issues coins of denomination $\frac{1}{n}$. Given a finite collection of such coins (of not necessarily different denominations) with total value at most $99 + \frac{1}{2}$, prove that it is possible to split this collection into 100 or fewer groups, such that each group has total value at most 1.

Hints: 549 71

Problem 20 (The Sylvester–Gallai theorem). Prove that every finite set of points in the Euclidean plane has a line that passes through exactly two of the points or a line that passes through all of them.

Hints: 576 612 152 643

10 Counting in Two Ways

10.1 Introduction

Counting in two ways is a powerful combinatorial principle that relies on being able to (as the name suggests) count the number of possible values of a quantity in two different ways.

However, the following question arises. Lets say we've done some counting and we got an answer. Why on earth would you want to use another method when you *know* the answer you're going to end up with. The trick is that it was never really the *quantity* (which we were counting) that we really cared about. In particular sometimes we want to find the value of some x. To do this, we define a quantity that depends on x, and figure out that the number of possible values of the quantity is (as an example) $10x$. We now evaluate the quantity in a different way that allows us to find its value directly, as let's say 50. We can then conclude that $x = 5$, a conclusion that may have been difficult without defining the *quantity*.

Like most olympiad combinatorics topics, there is one step which forms the make-or-break for this technique: being able to define the right quantity. The issue is that nobody is going to come and tell you that this *is* the correct thing to introduce, or that this entire principle is even useful to solve the problem. But this issue is what makes the technique and olympiad combinatorics all the more fun and challenging! If you already knew what to do, would it be fun at all?

10.2 Example Problems

Example 1

Let $U = \{1, 2, \ldots, 10\}$, and $A_1, A_2, \ldots A_{20}$ be 5-element subsets of U. It is given that each element is contained in the same number of subsets. In how many subsets is the element 1 contained?

Note that since each element is contained in the same number of subsets given to us, we may assume this number to be x. Since 1 is contained in x subsets too, we essentially wish to figure out x.

We define the following as our quantity: pairs of the form (subset, element) such that the element is contained in the subset and the subset is one of the 20 given to us.
Note that there are 20 subsets, and 5 elements in each of them so we get that the number of possible values for the pair is 20×5. On the other hand, we know that we can first fix our element in one of 10 ways. Now this element is contained in x subsets so the total number of ways is $10x$. These together give us that

$$x = \frac{20 \times 5}{10} = 10$$

Now that we have solved the problem, it is time to raise the following question.

> How on earth are we going to come up with an obscure quantity like pairs of the form (subset, element)?

Honestly, this is fairly motivated once you've seen enough problems from this chapter. Most of the expressions we use include pairs or triplets, which give us the following advantage: we can fix the first element and vary the second one, or we can fix the second element and vary the first one. This powerful idea allows us to equate different looking expressions. As for what the pairs are of, the best hope is to try and find elements that allow you to use up as much *information* provided to you as possible.

Example 2

Let $U = \{1, 2, ..., n\}$, and $A_1, A_2, \dots A_m$ be k-element subsets of U. It is given that for every pair of elements in U there are exactly x subsets such that contain both of them. Find x. Also, in how many subsets is the element 1 contained?

Step one: what is the crucial condition.

> "It is given that for every pair of elements in U there are exactly x subsets such that contain both of them."

So we want to pick some pair of elements in U, and then we want to pick a subset that contains them. So the following triplet can be seen to be very useful: (subset, element from subset, another element from subset).

If we fix the two elements first, we get $n(n-1)x$ ways (order matters - but it's okay to ignore this as long as you ignore the order in the other expression as well) since each pair belongs to x subsets. On the other hand, if we fix our subset first, there are $k(k-1)$ ways to pick the two elements, so the number of possible values is $k(k-1)m$. So we conclude that

$$x = \frac{k(k-1)m}{n(n-1)}$$

We now wish to find out how many subsets 1 is contained in. The trick is to let the first element just be equal to one. In particular now your triplet is (subset, 1, another element from subset) and we're only including the subsets that contain 1. Let's say there are t of these. Then we can first fix the subset in t ways, and then find another element in $(k-1)$ ways. On the other hand, we can first fix our second element in one of $n-1$ ways and then find the subset in one of x ways. So

$$t = \frac{(n-1)x}{k-1} = \frac{(n-1)k(k-1)m}{(k-1)n(n-1)} = \frac{mk}{n}$$

The surprising thing is not that we ended up with an expression very similar to the previous problem (we had the same expression with $m = 20, k = 5, n = 10$), but the fact that the expression exists in the first place. Confused? The magical thing is that this time, nobody told you that each element will be contained in the same number of subsets. All we knew is that pairs of elements were contained in the same number of subsets and that was magically enough to conclude what we needed.

Example 3 (IMO 1998)

In a contest, there are m candidates and n judges, where $n \geq 3$ is an odd integer. Each candidate is evaluated by each judge as either pass or fail. Suppose that each pair of judges agrees on at most k candidates. Prove that

$$\frac{k}{m} \geq \frac{n-1}{2n}.$$

Read carefully. We've been given some candidates and some judges. The only real condition that we're given is

"Suppose that each pair of judges agrees on at most k candidates."

This is actually all we need to figure out what to count. Take a pair of judges and count the number of students they agree on. In particular, we define our triplet as,

(judge, other judge, candidate they agree on)

Then there are at most $n(n-1)k$ ways of picking this. We are now looking for another way to figure this out. Let's say we fix one judge and a candidate - but we have no good way of figuring out how many other judges award the candidate pass fail. Okay, not this then. What if we just fix a candidate. Now we need the number of judges that agree on him/her.

So either we need both judges to evaluate the candidate as a pass or both to evaluate as a fail. In particular the number of ways of doing this is

$$a(a-1)+b(b-1)$$

Here a is the number of judges who award the candidate pass and b is the number of judges who award the candidate fail. In particular, $a+b=n$, and the expression is minimised when a is closest to b (why?)

$$\begin{aligned} a(a-1)+b(b-1) &\geq \left(\frac{n-1}{2}\right)\left(\frac{n-1}{2}-1\right)+\left(\frac{n+1}{2}\right)\left(\frac{n+1}{2}-1\right) \\ &= \frac{(n-1)^2}{2} \end{aligned}$$

So we have that that

$$n(n-1)k \geq X \geq \frac{m(n-1)^2}{2} \implies \frac{k}{m} \geq \frac{n-1}{2n}$$

Example 4

A school has n students, and each student can take any number of classes. Every class has at least two students in it. We know that if two different classes have at least two common students, then the number of students in these two classes is different. Prove that the number of classes is not greater that $(n-1)^2$.

Once again, what's the condition?

> "We know that if two different classes have at least two common students, then the number of students in these two classes is different."

Take a pair of students - they cannot both belong to 2 classes of the same size. So take a pair of students - (student, another student) and fix the size of the class as k. If you fix a pair of students in $(n)(n-1)$ ways, there's at most one class for this size k. On the other hand, if we fix the class as k first, we now have $k(k-1)C_k$ ways of picking the two students, where C_k is the number of classes of size k. So we have that

$$\implies C_k \leq \frac{n(n-1)}{k(k-1)}$$

In particular, we can sum this over all k to get

$$\text{Number of classes} = \sum_{k=2}^{n} C_k \leq \sum_{k=2}^{n} \frac{n(n-1)}{k(k-1)} = n(n-1) \sum_{k=2}^{n} \frac{1}{k(k-1)}$$

Exercise. Evaluate this (hint: the sum telescopes!).

Example 5

In an $n \times n$ matrix where $n = m^2$ for some positive integer m, each of the numbers $1, 2, \ldots, n$ appear exactly n times. Show that there is a row or a column in the matrix with at least m distinct numbers.

This time, there's no information on pairs, so it seems unlikely you want to take 2 numbers together - or two rows/columns together. Let's keep it simple and try the following counting the following (a number, total number of rows + columns it belongs to).

For the sake of contradiction, all rows and columns have at most $(m-1)$ distinct numbers. Then we can choose a row/column in $2n$ ways and then pick a number from it in at most $(m-1)$ ways. So,

$$\text{Possibilities} \leq 2(m-1)n$$

On the other hand, let's say we focus on a number k. If it appears in only a_k rows and b_k columns, this means that all of its appearances are in the $a_k b_k$ intersection points of those rows and columns. In particular $a_k b_k \geq n$. By the AM-GM inequality,

$$\text{Possibilities} \geq \sum_{k=1}^{n} a_k + b_k \geq \sum_{k=1}^{n} 2\sqrt{a_k b_k} \geq 2nm$$

In particular, we have that

$$2mn \leq \text{Possibilities} \leq 2(m-1)n$$

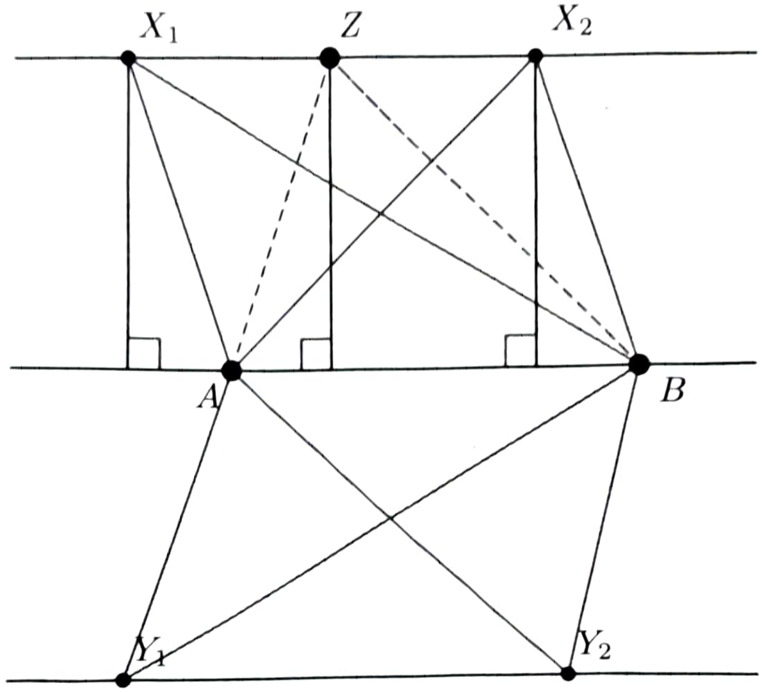
Which doesn't look very likely, does it?

10.3 Examples from Combinatorial Geometry

Example 6 (Iran 2010)

There are n points in the plane such that no three of them are collinear. Prove that the number of triangles whose vertices are chosen from these n points and whose area is 1 is not greater than $\frac{2}{3}(n^2 - n)$.

So we're given n points and we want to make as many triangles as possible have area equal to 1. The key idea is to fix the base of our triangle as AB. Since area depends on the product of base and height lengths and base length is fixed, the height length is now fixed as shown below. In particular, if we have points Z, X_1, X_2 all above, they're collinear - which isn't allowed. So there can be at most 4 triangles per pair (two above and two below the base).



Now let's count! We know that for each pair of vertices there are at most 4 triangles. So let's count the following pairs:

(pair of vertices, triangles with that pair having area 1)

Now if we fix our triangle first, we get 3 pairs of vertices so the number of possibilities is $3T$. On the other hand, it is also at most $4\binom{n}{2}$ and so we have that

$$T \leq \frac{2}{3}(n^2 - n)$$

Example 7 (IMO 1987)

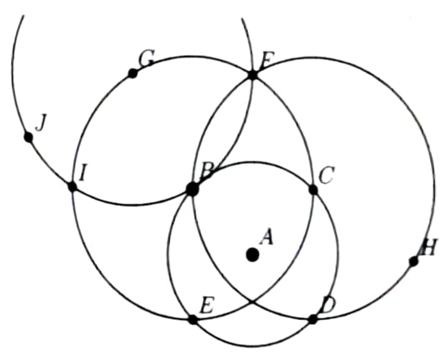
Let n and k be positive integers and let S be a set of n points in the plane such that

1. No three points of S are collinear.
2. For every point P of S there are at least k points of S equidistant from P.

Prove that:

$$k < \frac{1}{2} + \sqrt{2 \cdot n}$$

So we have a situation somewhat as shown below, where we can draw a circle through each vertex that contains at least k other points.



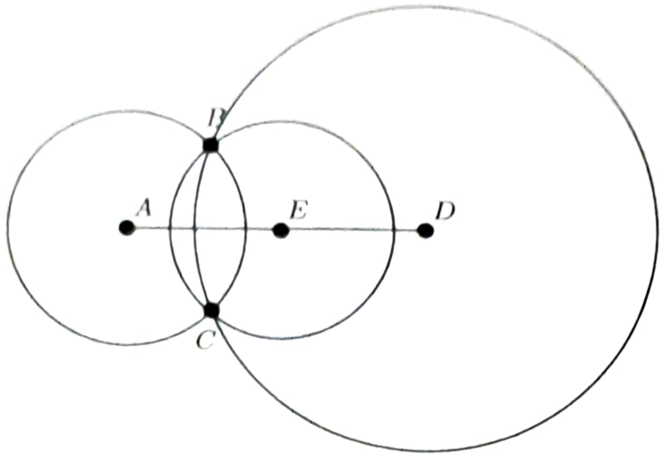
We could try counting the number of pairs of (circles, points on it). If we fix our circle, there's at least k ways of picking the point. The issue is the other half though - it is very much possible to have some point on almost every circle so it is hard to count.

At this point, we're slightly stuck. The key idea is to go back to the problem and see what condition you haven't used yet.

Aha,

No three points of S are collinear.

But... how does this relate with the other condition. The trick is to use the following geometry fact: if two points lie on 3 different circles, then the centres of those circles are collinear, as shown below.



So the idea is that if we pick any pair of points, this should lie on at most 2 circles. So we count (circles, pairs of points lying on the circle).

Exercise. Finish from here!

10.4 Problems

Problem 1 (China 1993). A group of 10 people went to a bookstore. It is known that

1. Everyone bought exactly 3 books;

2. For every two persons, there is at least one book that both of them bought.

What is the least number of people that could have bought the book purchased by the greatest number of people?

Hints: 313 57

Problem 2. Let $A_1, A_2, \ldots, A_7$ be subsets of $M = \{1, 2, \ldots, 7\}$, such that each pair of elements of M belongs to exactly one of the subsets, and $|A_i| \geq 3$ for each i. Show that $|A_i \cap A_j| = 1$ for all $i \neq j$. $|A_i \cap A_j| \le 1$ and $|A_i| = 3$

Hints: 252 88 174 333

Problem 3 (Italy TST 2005). A stage course is attended by $n \geq 4$ students. The day before the final exam, each group of three students conspire against another student to throw him/her out of the exam. Prove that there is a student against whom there are at least $\sqrt[3]{(n-1)(n-2)}$ conspirators.

Hints: 257 567

Problem 4 (1995 IMOSL C5). At a meeting of $12k$ people, each person exchanges greetings with exactly $(3k+6)$ others. For any two people, the number of people who exchange greetings with both of them is the same. How many people are at the meeting?

Hints: 233 99 266

Problem 5 (China TST 1992). Sixteen students took part in a math competition where every problem was a multiple choice question with four choices. After the contest, it is found that any two students had at most one answer in common. Determine the maximum number of questions.

Hints: 306

Problem 6 (China TST 1996). Eight singers participate in an art festival where m songs are performed. Each song is performed by 4 singers, and each pair of singers performs together in the same number of songs. Find the smallest m for which this is possible.

Hints: 337 418

Problem 7 (USA TST 2005). Let n be an integer greater than 1. For a positive integer m, let $S_m = \{1, 2, \ldots, mn\}$. Suppose that there exists a $2n$-element set T such that

(a) each element of T is an m-element subset of S_m;

(b) each pair of elements of T shares at most one common element;

(c) each element of S_m is contained in exactly two elements of T.

Determine the maximum possible value of m in terms of n.

Hints: 17 1 406

Problem 8 (2000 IMOSL C3). Let $n \geq 4$ be a fixed positive integer. Given a set $S = \{P_1, P_2, \ldots, P_n\}$ of n points in the plane such that no three are collinear and no four concyclic, let a_t, $1 \leq t \leq n$, be the number of circles $P_iP_jP_k$ that contain P_t in their interior, and let

$$m(S) = a_1 + a_2 + \cdots + a_n.$$

Prove that there exists a positive integer $f(n)$, depending only on n, such that the points of S are the vertices of a convex polygon if and only if $m(S) = f(n)$.

Hints: 577 24

Problem 9 (IMO 2004 C1). There are 10001 students at an university. Some students join together to form several clubs (a student may belong to different clubs). Some clubs join together to form several societies (a club may belong to different societies). There are a total of k societies. Suppose that the following conditions hold:

1. Each pair of students are in exactly one club.
2. For each student and each society, the student is in exactly one club of the society.
3. Each club has an odd number of students. In addition, a club with $2m + 1$ students (m is a positive integer) is in exactly m societies.

Find all possible values of k.

Hints: 594 591 609 **Soln:** Page 281, Solution 20

Problem 10 (Iberoamerican). Let X be a set with n elements. Given $k > 2$ subsets of X, each with at least r elements, show that we can always find two of them whose intersection has at least

$$r - \frac{nk}{4(k-1)}$$

elements.

Hints: 223 315 393

Problem 11 (Corradi's Lemma). Let $A_1, A_2, \cdots A_n$ be r-elements sets such that the intersection of any two of these sets has size at most k. Show that the union of these sets has size at least

$$\frac{nr^2}{r+(n-1)k}$$

Hints: 480 384 542

Problem 12. On the first day, there are a hundred students who are divided into five groups. On the second day, the same one hundred students are divided into four groups. Prove that there exists a student who belongs to a larger group on the second day than the first.

Hints: 460 556 495 **Soln:** Page 281, Solution 21

III

Graph Theory

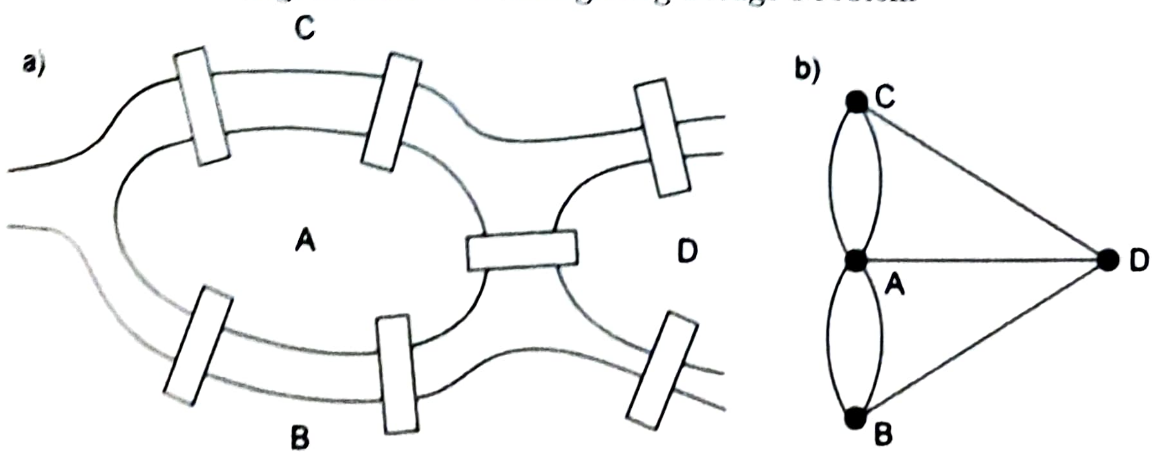
11 Graph Theory I

Graph Theory is a branch of mathematics and computer science that deals with modeling various relationships using vertices and edges. Graph theory as a field is crucial today, and is used everywhere we go - from searching the internet to artificial intelligence, and from DNA sequencing to Google Maps. The power of graph theory emerges from its simplicity. Despite being a concept that is fairly simple to understand, it is an active field of research with hundreds of open problems that mathematicians work on today.

11.1 The Königsberg Bridge Problem

The city of Königsberg is located on the Pregel River in Prussia. As shown in image (a) below, the river divided the city into 4 landmasses which were connected by seven bridges. The citizens often wondered if it was possible to start from home, travel through the city cross every bridge exactly once and return home.

Figure 11.1: The Königsberg Bridge Problem



Exercise. Find a round trip crossing each bridge exactly once, or try to prove that no such trip exists. (It's okay if you're not able to do this, just give it a try!)

The problem was first solved by Euler who gave rise to the technique of Graph Theory using this problem.

What Euler realised is that the entire imagery of cities and bridges is almost irrelevant. All that matters is that you have some landmasses, and there are some ways of going from one way to another. So he imagined each landmass as a point (vertex) and joined the points with lines (edges) based on the bridges as shown in figure (b). The round trip of the city, in this graph-theoretical form is now famously known as an *Eulerian Circuit*. Finding out which graphs have an Eulerian circuit is within your problems for this chapter!

Anyway, we describe a graph using its set of vertices and edges which connect pairs of vertices.

11.2 Why Graphs?

The concept of graph theory from the point of view of vertices and edges is fairly simple and raises questions as to how this could actually be useful. Graph theory is a topic that can often throw in surprises, coming up in unexpected ways and allowing you to analyze conditions in a new and often much simpler way.

A couple of places we usually expect to see Graph Theory in Math Olympiads include:

- Any problem that involves people - them knowing or not knowing each other - These usually have a direct conversion into a graph theory problem.
- Tournaments - graphs which represent round-robin tournaments where each edge is given a *direction* based on which of the 2 people that the vertices are representing beat the other.
- Cities, and their being connected by a road, bridge, or an airway.

To give you a much more obscure example, once in the past I coded up a program that gave the optimal method to solve a 2 by 2 Rubik's cube, given any initial scramble. How is graph theory involved? The idea is that you can represent each possible position on the Rubik's cube as a *vertex*. Now join two vertices using an edge if it is possible to go from one of the positions to the other in a single move. All that is left is finding the *shortest path* from the original scramble to the solved cube - and an algorithm known as the breadth-first search can be done to compute this path.

There are also a certain set of graph problems that come under the *Hall's Theorem*. These can come in handy in a variety of different problems, but in itself, the concept is pretty loosely connected to the rest of graph theory. I believe the hardest graph theory problems are the ones where you wouldn't know you had to use graph theory, and so this list of uses of graph theory is in itself nowhere close to complete.

Beyond the scope of Mathematical Olympiads, graph theory is used in various fields as described at the start of the chapter, and forms an integral part of research in mathematics and computer science today.

11.3 The Definitions

We are faced with one of my (and maybe yours too) oldest enemies - the boring task of knowing the meaning of some new terms. While this may feel useless or too abstract at first, stay with me - because putting these terms together can create an entire world of math as we'll see through this chapter and the next one.

Definition 1

Mathematically, we describe a graph G by its set of vertices V and edges E. i.e,

$$G = (V, E)$$

As described above, vertices are essentially used to describe certain objects (like landmasses on a map or just people) which are bound together by certain relationships described by edges (being connected by a bridge, or being friends or enemies). In the image below,

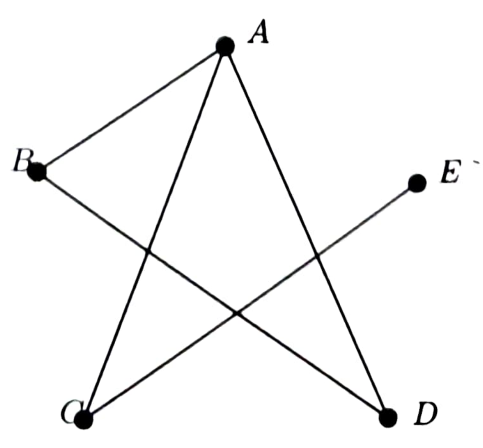
$$\{A, B, C, D, E\}$$

represents V, the set of vertices.

The set of edges E is

$$\{(A, B), (A, C), (A, D), (B, D), (C, E)\}$$

since those are the pairs of vertices which are joined by edges.



It's important to note that the *geometry* here is completely irrelevant to us (most of the time). So changing the positioning of the vertices, or the shape of the line joining two vertices (curved or anything) is irrelevant to us as long as the only 2 things we care about are the same, i.e., we have the same set of vertices and the same set of edges.

Usually, the graphs we deal with are *simple* and *undirected*, but it's worth noting that sometimes you might need to have multiple edges between the same two vertices (multigraph), and sometimes you give each edge a *direction* to make the graph a *directed one* (as if the roads are now one-way)

1. **Adjacent vertices**: Two vertices are said to be adjacent if there is an edge connecting them. So in the above graph, A and B are adjacent, while B and C aren't adjacent.

   Exercise. Are each of the pairs (C, D) and (D, A) adjacent?

2. **Degree**: The degree of a vertex d_i is the number of vertices it is adjacent to.

   Exercise. What is the degree of each of the 5 vertices in the given graph?

Example 2 (The Handshake Lemma)
Show that the sum of degrees of the n vertices of a graph is even.

This is a really important result, and you're bound to see a lot of applications of this throughout the chapter. This problem also introduces us to one of the most powerful ideas in graph theory - counting in two ways. It's often really powerful to use the fact that thinking of something vertex by vertex, or edge by edge should lead to the same result. This idea works especially well in graph theory and is something that should be on our mind every time we try a graph problem.

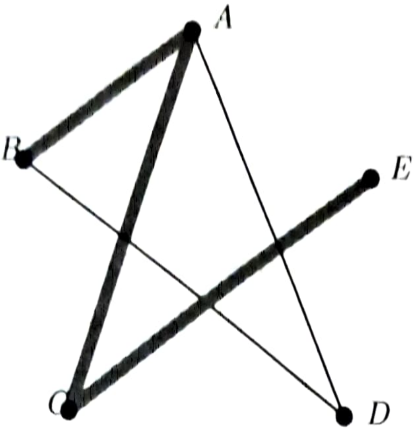
In particular, what happens if we consider this problem *edge by edge*?

Notice that each edge contributes exactly one to the degree of two vertices. In particular, every edge is counted twice in the summation as required!

In particular,

$$\sum_{i=1}^{n} d_i = 2|E|$$

where $|E|$ denotes the number of edges in the graph and d_i denotes the degree of the ith vertex.



3. **Path**: A path is what you expect it to be - some way of getting from one vertex to another using the edges of the graph.

   More formally, a path in a graph G is defined to be a sequence of *distinct* vertices $v_0, v_1, \cdots, v_t$ such that v_i is adjacent to v_{i+1} for all $i \in \{0, 1, \ldots, t-1\}$

   So above in figure 3, $B - A - C - E$ forms a path as highlighted.

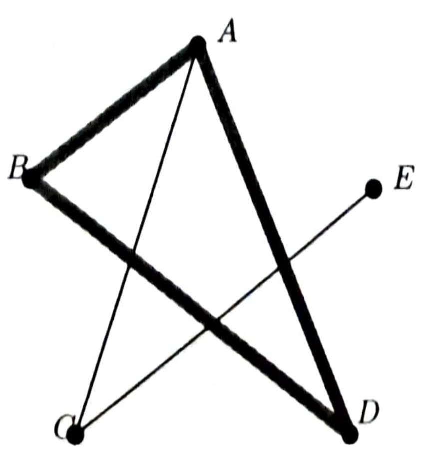
4. **Connected graphs**: A graph is said to be connected if there exists a path from each vertex to each other vertex of the graph.

   Exercise. Check that the graph drawn above is connected

5. **Cycles**: Once again, a cycle is what you expect it to be: a path that ends up at the same vertex as the one we started from.

   More formally, a cycle in a graph G is defined to be a sequence of distinct vertices $v_0, v_1, \cdots, v_t$ such that such v_i is adjacent to v_{i+1} for all $i \in \{0, 1, \cdots, t-1\}$ and v_t is adjacent to v_0. Here $A - B - D - A$ is a cycle.

   Exercise. Is there any other cycle in this graph?



By now I'm sure you're thinking:

> "Okay, I'm done. Enough definitions. This is way too abstract, and it's hard to believe this could be useful at all. I can accept that connectivity

is probably useful - but why on earth would I care if my graph had a cycle."

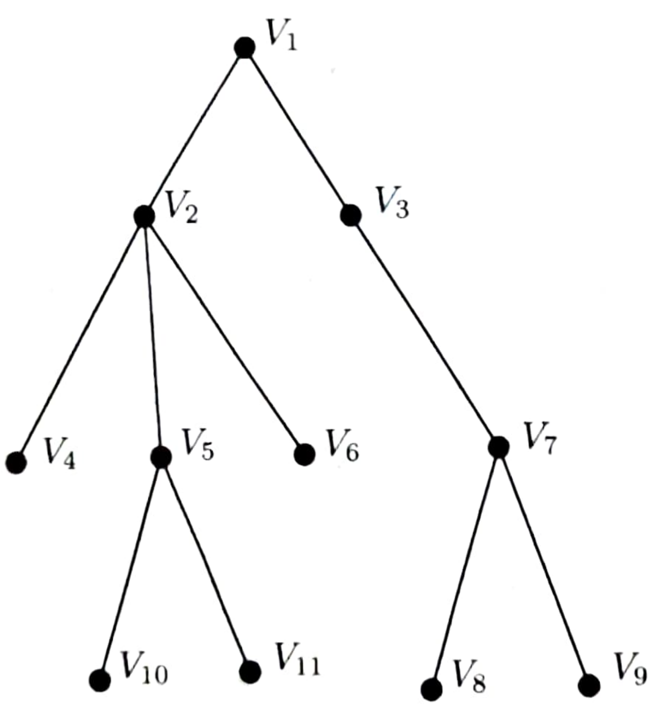
As it turns out, the existence of a cycle in the graph (or it's absence) can be a huge help in a variety of different problems, as we'll see in the upcoming sections. This motivates thinking of the *worst possible graphs* as the ones that *almost* have cycles but not really. Confused? The idea is to figure out the structures that have as many edges as possible while still not having a cycle.

It turns out that these structures have a name - **Trees**.

11.4 An In(tro)duction to trees

What is a tree?

Okay, I lied. Trees aren't the *worst possible graphs*. On the contrary, they're probably the most useful and powerful graphs. The biggest advantage of having a tree around is that it has a very strong condition attached to it.



You can probably make out why we decided to call the structure a tree from the above image - it does look like a family tree (I guess?). You could also argue that it is like the root of a tree extending downwards, but on the contrary, people decided to call the vertex at the top the root of the tree so that's unfortunate. On the other hand, the vertices at the bottom who have just one vertex connected to them are called leaves. It's okay if these names don't sit in your head for too long and we won't be too bothered by them - you can just check what the term meant if you forget at some point.

We can define a tree in many ways. Three of them are below:

- A connected graph with no cycles.
- A graph such that there is exactly one path between any two vertices.
- A *minimally* connected graph (so deleting any edge disconnects the graph).

Note that a connected graph has at least one path between any two vertices. And if there are two paths, there must've been a cycle in the graph, so definitions one and two are equivalent.

Problem 3. Show that the third definition is equivalent too.

However, we're yet to announce the most useful property of a cycle. Let's see if you can guess it first.

Exercise. Count the number of vertices and edges in the tree drawn above.

If you managed to count it correctly, you would've gotten 16 vertices and 15 edges. You're probably not convinced yet. But it's true. The number of edges in any tree is $|V| - 1$ (!!)

Let's try and prove this. One of the most useful techniques when we deal with trees is the method of induction (hence the in(tro)duction joke). The idea is that all trees have a *leaf* (a vertex with degree one, we'll show that this exists soon), and deleting them from the graph leads us to an almost identical sort of tree structure - with one vertex and one edge less.

More formally, we assume as our induction hypothesis that any tree with $n-1$ vertices has $n-2$ edges. Now if every tree on n vertices has a leaf, we can delete this leaf along with the single edge. This must leave us with $n-1$ vertices and $n-2$ edges so we must originally have had $n-1$ edges. For the base case, you can take $|V| = 2$ and there'll have to be exactly one edge.

Now let's ask ourselves - why should a leaf exist in the first place? There are several ways to prove this as well, but the idea is to consider some sort of greedy algorithm. Essentially, keep going *down* the tree while you can. If you somehow manage to go *up*: show that there must have been a cycle. On the other hand, the tree has a finite number of vertices so you can't keep going *down* forever. You can also frame this extremally by considering an endpoint in the longest path of the tree. If it was connected to anything outside the path, we'd have a longer path, and if it was connected to anything inside we'd have a cycle. So it has degree exactly one.

Exercise. Find another way to induct where we do not delete a leaf.

Hints: 125 403

Exercise. Find yet another way to induct where we do not delete a leaf, and we do not delete what we deleted in the above exercise.

Hints: 170 166

Remark. Notice that I've been stressing the fact that we're inducting by *deleting* something. Would the same induction work, if we tried to *add* a leaf to a $n-1$ vertex tree instead? Not exactly. The issue is that when you *add* a leaf, you cannot be sure that you can reach any possible tree on n vertices by this method. The only way to be sure is to start with your n vertex tree in the first place, delete a leaf, and then add it back in the end.

Also, in general, Induction works really well with trees because of these leaves!

11.5 The n edge trick

(Yes, I coined the above due to a lack of better ideas)

In the previous section, we showed that a connected graph without cycles has $n-1$ edges. The powerful trick comes in, when you think of this slightly differently though. Any graph which has *more* than $n-1$ edges must have a cycle.

Lemma 4 (The n edge trick)
Any graph on n vertices with at least n edges has a cycle.

Let's quickly prove this. If the graph is connected, we're done already since a connected graph with no cycles has exactly $n-1$ edges. But what if the graph isn't connected? The idea is that this only makes matters worse. In particular, the following statement holds. The proof is fairly easy, and you can try showing this on your own.

> Any graph on n vertices with k connected components and no cycles has $n-k$ edges.

Hints: 616

This trick that any graph with n vertices and n edges has a cycle comes up a lot, and one of the things that makes it important as well as hard is that the problems where it is used may not appear to have any connection whatsoever to graph theory.

Example 5 (CodeForces)

Consider n integers $a_1, \ldots, a_n$ satisfying

$$i - n \leq a_i \leq i - 1$$

for all i from 1 to n. Show that there always exists a nonempty subset of these integers with sum 0.

The really magical thing about this problem is that it does not appear to be connected to graph theory at all - which is why coming up with the proof to this is really hard if you're not familiar with the n edge trick (that being said, it is very difficult even if you know it because you also need to *know* that it'll be useful).

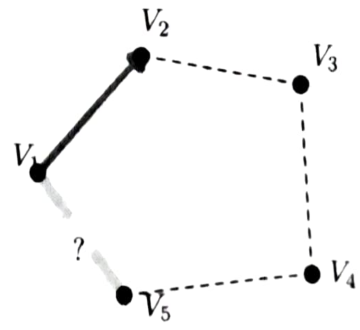
Anyways, let's first simplify the expression given a bit: we get

$$1 \leq i - a_i \leq n$$

for all i from 1 to n. Lets define $f(i) = i - a_i$, so we now have a function from $\{1, 2, \ldots, n\}$ to $\{1, 2, \ldots, n\}$.

The cool idea here is to think of this as a functional graph. So consider a graph with vertices labeled $\{1, 2, \ldots, n\}$, and draw an arrow (an edge which now also has a direction) from i to $f(i)$. This graph has n vertices and n edges (one originating from each vertex) - so it has a cycle (!!!).

We first show that this is in fact a directed cycle, i.e. if v_1 points to v_2, then v_2 points to v_3, v_3 points to v_4, and so on. To prove this, let's take an example of a cycle of length 5, and let's first assume that V_1 points to V_2. each vertex points to exactly one other vertex (i points only to $f(i)$) and so V_1 can't point to V_5 as well. This means V_5 points to V_1, which in turn means V_4 points to V_5, etc.



The question though, is that what have we achieved, using this directed cycle? Well, if the labels of the vertices are $b_1, b_2, \ldots b_k$ so that $f(b_1) = b_2$ and so on, we have some equations now. For instance,

$$b_2 = b_1 - a_{b_1}$$

In general, we have the following

$$
\begin{aligned}
b_2 &= b_1 - a_{b_1} \\
b_3 &= b_2 - a_{b_2} \\
b_4 &= b_3 - a_{b_3} \\
&\vdots \\
b_k &= b_{k-1} - a_{b_{k-1}} \\
b_1 &= b_k - a_{b_k}
\end{aligned}
$$

adding up all of these, the b_i cancel and we're done!

Example 6 (Based on 2005 IMOSL C4)
Let $n \geqslant 3$ be an integer. Prove that if the edges of K_n are colored with n colors (and each color is used at least once), then some triangle has three edges of different colors.

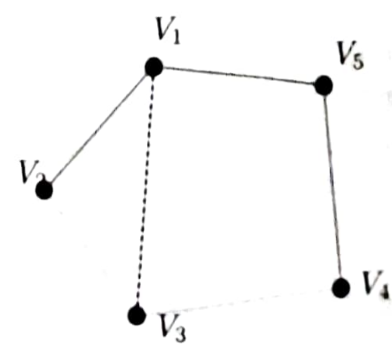
Note that a K_n is a complete graph on n vertices - so every pair of vertices has been joined together by an edge. In particular, each of these $\binom{n}{2}$ edges has a color now, so that a total of n colors are used.

There are multiple ways to solve this problem - one of them is a simple induction, but we'll be looking at how we can apply the n edge trick to the problem.

So we have n vertices already, but we don't have n edges. We have a lot more of them. But we do have n colors. Aha! What if we take one edge of each color? Now this graph must have a cycle - and all the edges in this cycle have different colors. It remains to show that if we can construct such a cycle of length k, we can use it to create cycles of smaller and smaller lengths.

So let's say our cycle had vertices

$$v_1 - v_2 - v_3 - \ldots v_k - v_1$$



The idea is that we can consider the edge between two non-adjacent members of this cycle - say v_1 and v_3. If this edge has a color different from $v_1 - v_2$ and $v_2 - v_3$, we're done already (why?). If not, the cycle

$$v_1 - v_3 - \cdots - v_k - v_1$$

also has unique colors (check this). So we can keep shortening the cycle and eventually end up with a triangle.

Exercise. Find a direct way to induct, i.e. assume that a K_{n-1} colored with $n-1$ colors has a rainbow triangle and use that hypothesis to prove the problem.

Example 7 (Taiwan 2001)
Let $n \geq 3$ be an integer and $A_1, A_2, \ldots, A_n$ be n distinct subsets of $S = \{1, 2, \ldots, n\}$. Show that there exists an element $x \in S$ such that the subsets

$$A_1 \backslash \{x\}, A_2 \backslash \{x\}, \ldots, A_n \backslash \{x\}$$

are also distinct.

Let's assume for the sake of contradiction that no matter what x you pick, some two subsets become the same. So for example let's say when you pick 1, A_1, A_2 become the same. This essentially means that $A_2 = A_1 \cup \{1\}$ or the other way round (why?). Similarly two sets become the same when you pick $x = 2$, or $x = 3$, or x as anything from 1 to n.

Any ideas on how to continue from here?

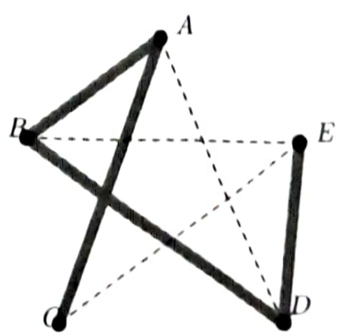
Problem 8. Find the graph, and the cycle!
Hints: 188 544

11.6 Back to trees!

Example 9
Show that every connected graph has a spanning tree

We must first define what a spanning tree is. The objective here is to delete some edges from the original graph so that the graph formed by the vertices and remaining edges (known as a subgraph) is a tree. These remaining edges, along with the original set of vertices form the spanning tree of the graph.

For instance, in the graph shown on the right, we have deleted the dashed edges to end up with only the highlighted edges - which form a spanning tree.



So in general, we want to be able to delete some edges from this graph, while making sure we're not *disconnecting* the graph. Essentially, at every point we're looking for an edge that we can get away with deleting. The question is, how do we find such an edge?

Let's rephrase this more formally. We want an edge $u - v$ so that even after deleting it, there's a path from u to v. Aha, this means that this edge is the part of some cycle!

So the trick is to consider what happens if we just simply delete a random edge from a cycle in the graph (Note that if a cycle doesn't exist, the graph is already a tree!). The hope would be that the graph is still connected. In that case, we can just keep deleting edges from cycles until only $|V| - 1$ edges remain, at which point we would be done since the graph is connected, and is a tree!

Exercise. Check that if we delete an edge from a cycle in a connected graph, the graph is still connected.

A question to ask here:

> "Why does a spanning tree have any significance?"

A lot of problems in graph theory are framed in the following way:

Problem. You are given a connected graph. You are allowed to delete edges. You want to end up at the following ...

In such a case, note that it is always possible to reduce any connected graph to a tree. So if we can prove the result for a tree, we would be done for all connected graphs. Moreover, a tree is a connected graph - so we need to prove the result for trees anyway.

This idea of spanning trees is therefore super useful in a lot of problems. Trees have much more structure than general graphs (like having leaves), which gives us more tools to play with.

Example 10 (BAMO 2005/4)

There are N cities in the country of Euleria, and some pairs of cities are linked by dirt roads. It is possible to get from any city to any other city by traveling along these roads. Prove that the government of Euleria may pave some of the roads so that every city will have an odd number of paved roads leading out of it if and only if N is even.

Firstly let's convert the problem into graph theoretic terms. We have a connected graph G on N vertices. We want to be able to delete some edges so that the remaining graph has each degree d_i odd.

There are two parts to this problem - proving that N odd cannot work and proving that N even works. The first part is relatively simple and intuitive and follows from Example 2 (Why!).

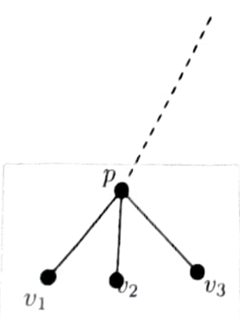
We now use the trick from the previous problem to prove that even N do in fact work - we can assume that our graph is a tree by converting any connected graph into its spanning tree.

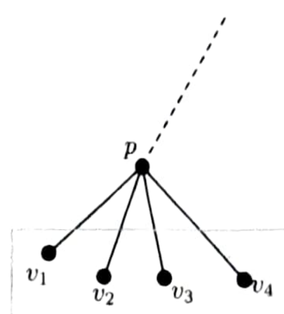
Okay, so now we have a tree to deal with. That's nice, but what now? Well, as we've done earlier - what's that one technique that works really well with trees? Induction! By deleting leaves. Notice that any leaf in the tree has degree 1 at the start. We need the leaf to have an odd degree at the end, so this means that we must keep this edge in our final graph.

Wait... So can we not just delete this leaf and induct? Well, not exactly. The problem we are facing is that the graph now has an odd number of vertices, and the vertex the leaf was adjacent to (say v) now needs to have an even rather than odd degree (because at the end we'll have to add back 1 edge which came from the leaf).

To solve this, we can take two cases - based on whether v, the vertex the leaf was adjacent to (known as its parent), has an even number of leaves adjacent to it, or an odd number of them.

Technically we can run into issues if this vertex v has some non-leaves under it but this is easy to fix - just pick the *lowest* parent (think about this until it makes sense).





1. Case 1: The parent had an odd number of leaves under it. In this case, we must keep all of those leaf edges. After doing this, notice that the parent currently has an odd number of edges originating from it as well, so we delete the parent along with all those leaves.

   Exercise. Check that all the conditions are satisfied.

2. Case 2: The other case is where the parent has an even number of leaves under it. Once again, we must keep all the leaf edges. After this, the parent has an even degree so we can't delete it just yet. What we can do, though, is delete all the leaves under it - and once again, we've deleted an even number of edges!

Remark. It's important to raise the following question here.

"What motivates inducting by deleting leaves?"

Induction by deleting leaves is just the most common thing you would do with trees. The idea is that leaves have a lot of rigidity, by having a fixed degree and always existing in a tree. This makes it a powerful and useful technique in graph theory.

T

Theory $\boxed{y}$ y

y

y

y

11.7 Problems

Well known results

Problem 1. Let G be a connected graph with n vertices having a cycle. We delete an arbitrary edge of the cycle. Show that the new graph (called an induced subgraph) is connected too.

Hints: 430 561

Problem 2. Show that a connected graph has an Eulerian Circuit iff all its vertices have even degree.

> **Remark.** An Eulerian Circuit is a sequence of vertices $v_{i_1}, v_{i_2}, \dots, v_{i_n}$ that starts and ends at the same vertex ($i_1 = i_n$), v_{i_k} is adjacent to $v_{i_{k+1}}$, and each edge is used exactly once in the sequence.

Hints: 280 190 47

Problem 3. Show that a graph with minimum degree $\delta \geq 2$ has a path of length at least $\delta + 1$.

take longest path

Hints: 298 575

Problem 4. Let T be a tree with n edges. Prove that any graph G with a minimum degree at least n has a subgraph isomorphic to T. (Two graphs are isomorphic if their vertices can be labelled in a way that makes the set of edges in the graphs the same.)

greedy

Hints: 43 177

Problem 5. Consider a graph G where each degree is equal to 2. Show that G is a disjoint union of cycles.

Hints: 235 232

Problem 6. Let G be a finite simple graph with $m > 0$ edges and $n > 1$ vertices. Show that one can delete some number of vertices of G to obtain a graph with at least one vertex whose minimum degree is at least $\frac{m}{n}$.

Hints: 501 352 156

The n edge trick

Problem 7. Suppose $2n$ points of an $n \times n$ grid are marked. Prove that there exists a $k > 1$ and $2k$ distinct marked points

$$a_1, \cdots, a_{2k}$$

such that, for all i, a_{2i-1} and a_{2i} are in the same row, while a_{2i} and a_{2i+1} are in the same column.

Hints: 369 620 568

Problem 8. You're given an $n \times n$ matrix of numbers so that no two rows are identical to each other (so for each 2 rows, there exists at least one column in which their values are different). Show that you can delete a column such that the rows are still pairwise distinct.

Hints: 138 11 388

$^nC_2$

Problem 9 (INMO 2023/5). Euler marks n different points in the Euclidean plane. For each pair of marked points, Gauss writes down the number $\lfloor \log_2 d \rfloor$ where d is the distance between the two points. Prove that Gauss writes down less than $2n$ distinct values. Note: For any $d > 0$, $\lfloor \log_2 d \rfloor$ is the unique integer k such that $2^k \leq d < 2^{k+1}$.

Hints: 161 76 415 424 203

Problem 10 (ISL 2020 C4). The Fibonacci numbers $F_0, F_1, F_2, \ldots$ are defined inductively by $F_0 = 0, F_1 = 1$, and $F_{n+1} = F_n + F_{n-1}$ for $n \geq 1$. Given an integer $n \geq 2$, determine the smallest size of a set S of integers such that for every $k = 2, 3, \ldots, n$ there exist some $x, y \in S$ such that $x - y = F_k$.

Hints: 245 491 318 181 622

Induct/Tree

Problem 11 (PUMAC). Given the graph G and cycle C in it, we can perform the following operation: add another vertex v to the graph, connect it to all vertices in C, and erase all the edges from C. Prove that we cannot perform the operation indefinitely on a given graph.

Hints: 249 12 **Soln:** Page 281, Solution 22

Problem 12 (Iran). Consider a tree with n vertices, labeled with $1, \ldots, n$ in a way that no label is used twice. We change the labeling in the following way - each time we pick an edge that hasn't been picked before and swap the labels of its endpoints. After performing this action $n - 1$ times, we get another tree with its labeling a permutation of the first graph's labeling. Prove that this permutation contains exactly one cycle.

Hints: 279 276

Problem 13. Consider a graph G on n vertices. At any stage, a vertex can be either red or blue. Initially, all the n vertices have color blue. In an operation, you are allowed to switch the colors of a vertex along with the colors of all of its neighbors. Show that these operations can be performed in a way so that all the vertices end up red.

Hints: 221 469 407 461

Problem 14 (MOP 2008). Prove that if the edges of a K_n, the complete graph on n vertices, are colored such that no color is assigned to more than $n-2$ edges, there exists a triangle in which each edge is a distinct color.

Hints: 638 64 14 343

Other problems

Problem 15 (PUMaC 2013). Let G be a graph and let k be a positive integer. A *k-star* is a set of k edges with a common endpoint and a *k-matching* is a set of k edges such that no two have a common endpoint. Prove that if G has more than $2(k-1)^2$ edges then it either has a k-star or a k-matching.

Hints: 454 148

Problem 16. A building consists of 4004001 rooms arranged in a 2001×2001 square grid. Is it possible for each room to have exactly two doors to adjacent rooms?

Hints: 506 650 74 75 142

Problem 17 (Russia 1999). In a country, there are N airlines that offer two-way flights be- tween pairs of cities. Each airline offers exactly one flight from each city in such a way that it is possible to travel between any two cities in the country through a sequence of flights, possibly from more than one airline. If $N-1$ flights are cancelled, all from different airlines, show that it is still possible to travel between any two cities.

Hints: 486 490 67 483 410

Problem 18 (Komal A. 722.). The Hawking Space Agency operates $n-1$ space flights between the n habitable planets of the Local Galaxy Cluster. Each flight has a fixed price which is the same in both directions, and we know that using these flights, we can travel from any habitable planet to any habitable planet.

In the headquarters of the Agency, there is a clearly visible board on a wall, with a portrait, containing all the pairs of different habitable planets with the total price of the cheapest possible sequence of flights connecting them. Suppose that these prices are precisely $1, 2, \ldots, \binom{n}{2}$ monetary units in some order. prove that n or $n-2$ is a square number.

Hints: 624 380 601

12 Graph Theory II

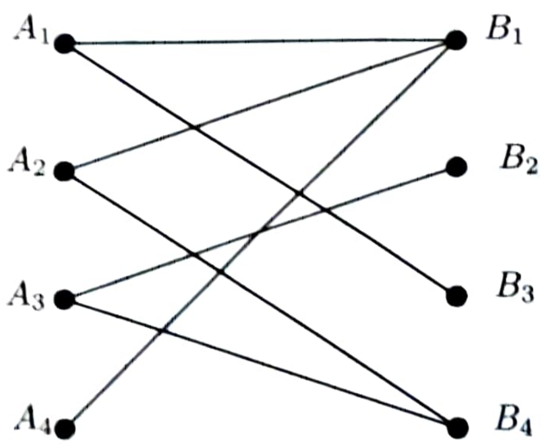
Graph Theory is by no means limited to cycles and trees. We still have bipartite graphs, chromatic numbers, cliques, independent sets, Hamiltonian paths, Turan's theorem, Ramsey numbers, and so on in a list that hardly ever ends.

12.1 Bipartite Graphs

This is a fairly different concept from the ones we have discussed till now, yet really useful in various olympiad problems.

Definition 1

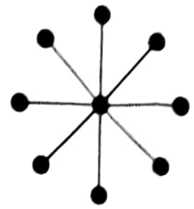
A graph is called Bipartite if its vertices can be partitioned into two subgroups A and B so that any edge in the graph has one vertex in A and one in B. In other words, there should be a partition of the vertices into A and B so that no two members of A or two members of B are adjacent.



As shown, the ovals represent the subsets A and B. Notice that there is no edge between any of A_1, A_2, A_3, A_4 or between B_1, B_2, B_3, B_4. The only edges have one vertex from each.

In general, the size of the subsets U and V can be different.

So even a graph where one vertex is connected to all others and the others are connected only to it (called a star graph) is bipartite since we can set U to that single vertex and V to all the others. The main problem is, for an n vertices graph, there are $2^{n-1} - 1$ possible choices for U and V, and it may not always be so obvious whether the graph is bipartite or not. It turns out that there is a very strong yet simple condition about the existence of bipartite graphs.



Lemma 2

A graph is bipartite if and only if it has no odd cycles. (An odd cycle is a cycle with an odd number of edges.)

Let's try and prove this. First, we show that a graph with an odd cycle cannot be bipartite. For the sake of contradiction, let's say the graph was bipartite. Then, we can group the vertices into A and B so that there is no edge between 2 vertices of A or 2 vertices of B.

Let the odd cycle be $v_1 - v_2 - v_3 - \cdots - v_{2k+1} - v_1$.

Let's say $v_1 \in A$ (without loss of generality). This would mean $v_2 \in B$ (or we have an edge within A). This, in turn, means $v_3 \in A$ and so on. In particular, we get $v_{2k+1} \in A$ so both v_{2k+1} and v_1 are in A, which isn't allowed.

The second part is slightly trickier. We want to show that if the graph has no odd cycles, then it is indeed bipartite.

Note that if the graph has several connected components, we can solve the problem independently for each component and put them together at the end (how?). Thus, we can assume that the graph is connected for the time being.

Let's say we put v_1 in A. If you think about it, the group for each of the other vertices is just forced. Anything adjacent to v_1 must be in B, anything adjacent to one of those must be in A, and so on. In particular, if we consider the path length (take an arbitrary path) between v_1 and v_k, we want to put v_k in B if this path length is odd and in A otherwise (check this!). This can only go wrong if there are two conflicting path lengths - one path is of odd length, and another is of even length.

Exercise. Show that the graph must have had an odd cycle in this case. (you'll have to be slightly careful since these paths must have common vertices.)

Remark. There's a slightly neater way of phrasing this (which avoids the common vertices in path issue): take a spanning tree of the graph and color it using the path lengths from v_1 (there is only one path, so there's no problem). Now, when you're adding back edges, if an edge you're trying to add creates a problem at any point, it must have also created an odd cycle.

Odd cycles may seem like a pretty random thing right now, but they come up surprisingly often, sometimes in the form of *triangles* too - which are just cycles of length 3.

Let us now see a couple of olympiad problems around bipartite graphs.

Example 3 (Baltic Way 2021)

Let n be a positive integer, t be a non-zero real number and $a_1, a_2, \ldots, a_{2n-1}$ be (not necessarily distinct) real numbers. Prove that among these $2n-1$ numbers, we can find n such that no two of them have difference t.

Let's first consider this seemingly algebraic problem into graph theory - so let's say we consider a graph on $2n-1$ vertices, and we join two vertices if the numbers associated with them have difference t. We want to find n vertices so that no two are connected (an independent set).

Here comes the powerful use of bipartite graphs: we know that there is no edge between vertices of the same subgroups in a bipartite graph. By the pigeon-hole principle, if we have $2n-1$ vertices, one of these two subgroups has size n - so we have n vertices such that no two are connected.

Of course, having such an independent set is not equivalent to the graph being bipartite. However, a powerful tool we have on our hands is showing that the graph is bipartite, as that would directly imply that it has such an independent set.

The good part about having to prove a certain graph bipartite is that you don't need to do it directly - you can decide to show that the graph has no odd cycle (which is often a lot more reasonable). In particular, let's assume that the graph has a certain odd cycle - say of length 3. so we have $|a_1 - a_2| = |a_2 - a_3| = |a_3 - a_1| = t$.

Notice that this means,

$$\begin{aligned} a_1 &= a_2 \pm t \\ a_2 &= a_3 \pm t \\ a_3 &= a_1 \pm t \end{aligned}$$

Adding the equations gives that $a_1 + a_2 + a_3 = a_1 + a_2 + a_3 + kt$ where $k \in \{3, 1, -1, -3\}$ according to the number of negative signs. But this means that $t = 0$, a contradiction.

In general, we have $2n-1$ equations, and we'll end up with $kt=0$ with k is an odd number between $-(2n-1)$ and $2n-1$, which implies that $t=0$. Thus, this graph can not have any odd cycles and is thus bipartite.

Remark. Whenever you want to show that there exists an independent set of around half the size, it is useful to consider the possibility of the graph being bipartite.

Example 4 (Greece National Olympiad 2023)

A class consists of 26 students with two students sitting on each desk. Suddenly, the students decide to change seats such that every two students that were previously sitting together are now apart. Find the maximum value of the positive integer N such that, regardless of the students' sitting positions, at the end there is a set of N students such that every two of them have never been sitting together.

Let's first try to guess the value of N. Obviously, if we try $N=26$, every student's partner on each day is present, so that isn't a good idea at all. In fact, let's consider just day 1. We can split the students into 13 pairs. Thus, if we take 14 students, we'll definitely have some two from the same pair, so $N \leq 13$.

Let's consider the graph created by drawing edges between two vertices iff the corresponding students sat together on at least one of the days. So each vertex has degree 2.

Exercise. As a fun exercise (not too relevant to the problem), try characterizing all graphs that have all degrees equal to 2.

Now, if we can show that this graph is bipartite, that would mean that there's a group of 13 students such that no pair of them sat together on either of the days.

Let's say there's some odd cycle - once again, let's consider a cycle of just three vertices first: A, B and C. Let's say A and B sat together on day 1. Then B and C must have sat together on day 2 (B was sitting with A on day 1). So C must have partnered A on day 1 (C was sitting B on day 2). So A was partnering C and B both on day 1. That doesn't sound right.

In general, we get that if v_1-v_2 was on day 1, then v_2-v_3 was on day 2, v_3-v_4 was on day 1, and so on. In particular, any odd cycle would lead to a contradiction, so the graph is bipartite - and thus has a subgroup of at least 13 vertices so that no two students in that group sat with each other on either of the days.

Example 5

Some of the cities in a country are connected by direct two way routes (there is at most one direct route between two cities) so that there are exactly three direct routes from each city and every two cities connected have different population. On each direct route, the least common multiple of the population of the two cities the route connects is written. It turns out that the sum of the written integers on the routes is equal to twice the overall population in the country. Prove that the cities can be divided into two groups so that there is no direct route which connects two cities from the same group.

So we have a graph, and each vertex is assigned a number so that any two neighbouring vertices have different values. Each vertex has degree 3, and each edge is assigned a value equal to the LCM of its two endpoints. It is given that the sum of the edge values is equal to twice the sum of the vertex values. We somehow need to show that this limited information proves that the graph is bipartite.

Before starting off with some odd cycle, let's consider this LCM condition. Can the sum of LCMs even be equal to twice the vertex sum?! Well, apparently, it can - but we'll need to figure out how.

Let's say the endpoints of an edge are a and b with $a > b$ (the problem condition says that $a \neq b$). Now, we want to find some condition on the LCM in terms of the sum of $a + b$ (so that we can sum this over all the edges). We know that the LCM cannot be b, so it's at least $2b$. On the other hand, it is also at least a. We can't really say a lot more than this because in the case where $a = 2b$, both equalities hold. So we get that the LCM (say k) satisfies

$$\begin{aligned} \frac{k}{2} &\geq b \\ k &\geq a \\ \implies \frac{3k}{2} &\geq a + b \end{aligned}$$

So $k \geq \frac{2}{3}(a+b)$. Now let's try to sum this over all edges, and say X_u denotes the value of vertex u

$$2\sum_{v \in V} X_v \geq \sum_{(u,v) \in E} \frac{2}{3}(X_a + b) = \sum_{v \in V} 3 \cdot \frac{2}{3} X_v = 2\sum_{v \in V} X_v$$

(We multiply by the factor of 3 since each vertex has degree 3 and thus appears in $X_u + X_v$ thrice.)

This may look like we have reached nowhere, but notice the inequality sign in between. If we ended up with $a \geq a$, equality must have held everywhere - and in particular, this means that the LCM must have been equal to $\frac{3}{2}(a+b)$ in every case. So for each edge, its two endpoints are $2b$ and b or b and $2b$. Yay!

Exercise. Now use this $2x$ and x condition to show that we can't have an odd cycle (this actually becomes equivalent to the first problem we did if you take logs everywhere!)

Example 6 (2004 ISL C3)

The following operation is allowed on a finite graph: Choose an arbitrary cycle of length 4 (if there is any), choose an arbitrary edge in that cycle, and delete it from the graph. For a fixed integer $n \geq 4$, find the least number of edges of a graph that can be obtained by repeated applications of this operation from the complete graph on n vertices (where each pair of vertices is joined by an edge).

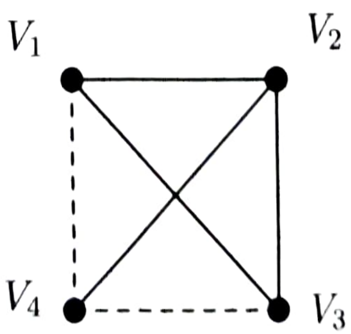
Remark. The complete graph on n vertices is often referred to as K_n.

So we're picking an edge from a cycle of length four and deleting it. This may seem similar to something we tried to do in the previous chapter to prove the existence of a spanning tree - take an edge from a cycle and delete it while you can. The idea was that the graph never gets disconnected at any point.

Exercise. Check that the graph cannot become disconnected if you keep removing single edges from existing cycles.

So, the guess? Well, we know that the worst case is a tree - so the answer must be $n-1$? Not quite.

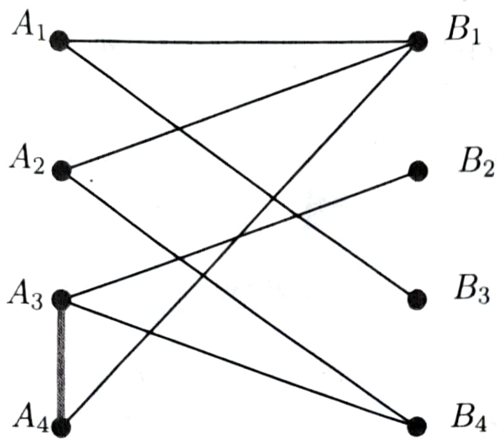
Let's try out a couple of base cases. For $n = 4$, we can delete two edges like below. First we delete an edge from cycle $V_1 - V_2 - V_3 - V_4 - V_1$ and then one from $V_1 - V_3 - V_2 - V_4 - V_1$. However, no matter how hard we try, it seems impossible to delete three edges. So the answer for $n = 4$ is 4 and not 3 (!!)



So the answer isn't $n-1$ (at least for $n = 4$). What's the issue? The only possible issue is that the cycle of length four condition is somehow preventing us from getting rid of all cycles.

At this point, we find ourselves a bit stuck. How do we use the fact that the cycle is of length four? The trick is to notice that we're always getting rid of "even cycles" so the graph shouldn't ever become bipartite. Let's try to prove this formally.

Clearly, the graph at the start isn't bipartite. So let's say deleting a certain edge made the graph bipartite at some point. What happened just before this edge was deleted? Well, the graph was almost bipartite except for one faulty edge (as shown below).



But... we just deleted this from a cycle of length four. That doesn't sound possible, does it? (Why not!)

So the graph always has an odd cycle - and so the answer is always $\geq n$

Problem 7. Show that n is, in fact, achievable - by creating a construction that leaves only n edges at the end.

> **Remark.** This problem emphasizes the need to try base cases!! It's easy to get carried away thinking that the answer is $n-1$, but a simple check on $n = 4$ can make sure we don't get drawn along that path.

12.2 Chromatic numbers

Chromatic numbers, in some sense, form a generalization to a bipartite graph. You might ask: What is so special about 2? Well, a lot, but not much in this context. In particular, the chromatic number of a graph is essentially the minimum number of divisions you must make in the graph so that there is no edge between two vertices of the same component. This can also be thought of as coloring a graph.

and you don't want an edge between two vertices of the same color, which gives origin to the *chromatic* part of the name.

Interestingly, 2 *was* quite special. Finding whether a graph is bipartite or not (i.e. checking if its chromatic number is 2) is quite easy using the odd cycle trick. However, there is no efficient way (a polynomial time algorithm) to find whether a general given graph has chromatic number 3 or not.

While we can't figure out the exact chromatic number of most graphs too easily, we can create bounds on this chromatic number. In particular, the following is a way to bound it.

Example 8

The max degree of a vertex in the graph G is Δ. Show that the chromatic number of the graph is at most $\Delta + 1$.

Note that this is equivalent to showing that you can split the graph into $\Delta + 1$ divisions properly. In general, we think of this as coloring the vertices with $\Delta + 1$ colors so that no edge consists of vertices of the same colors.

So any vertex in the graph has at most Δ neighbours, but the number of colors we can afford is even more than that. This sort of *weak* bound suggests that a greedy algorithm that is "good enough" may be the way to go. In particular, let's say we have assigned the first five vertices a color. Now, we're trying to assign the sixth vertex a color. The only issue can be if none of the colors can be used for it - but in that case, it must have a neighbour of each color and thus a total of $\Delta + 1$ neighbours. In particular, this means that the following greedy algorithm works: order the vertices arbitrarily and color them one by one, making sure that the vertex you just colored doesn't share a color with any of it's colored neighbors. And we're done!

Let's try a fun problem which asks us to show that the chromatic number is at most four!

Example 9 (Russia)

You are given a connected graph G which exhibits the following property: deleting any odd cycle from G disconnects the graph. Show that the chromatic number of G is at most 4.

Wow, how do we even end up at that with such a random condition? We'll not concern ourselves with the number four just yet. Let's try and understand the condition. It says that if you delete an odd cycle, the graph isn't connected.

So let's say I take a spanning tree of this graph. If there's any odd cycle completely outside the graph, we would have a contradiction. This means that at least one edge of any odd cycle must be in this spanning tree. Now what? The idea is to think of our graph G as the spanning tree T unioned with the remaining G'. This G' doesn't have any odd cycles and so it is bipartite. On the other hand, trees are anyway bipartite (they don't have cycles at all, duh). So each of G' and T are bipartite.

Exercise. Find the finish!

Hints: 126

12.3 Cliques and Independent Sets

Cliques and independent sets relate closely to the *extremal* ideas we saw in the extremal principle and greedy algorithms chapter. Essentially, a clique is a graph with a subset of the vertices in your original graph so that every two vertices in that subgraph are joined by an edge. In some sense, it is like a hidden K_n in your graph. On the other hand, an independent set is the complete opposite - it is a subset of the vertices so that no two of the vertices from that subset are adjacent to each other (i.e. have an edge connecting them).

Why are these so useful? It's often a very powerful condition to have the *largest independent set or clique* ready with you. These can give you very strong conditions on the remaining graph and help you give structure to general graphs.

Exercise. Show that a tree has an independent set of size $\geq |V|/2$.

Hints: 653

Example 10 (Croatian TST 2011)
There are n people at a party, among whom some are friends. Among any 4 of them there are either 3 who are all friends with each other or 3 who aren't friends with each other. Prove that the people can be separated into two groups A and B such that A is a clique (that is, everyone in A knows each other) and B is an independent set (nobody in B knows anyone else in B).

Okay, so we want to split this group of people into a group where everyone knows each other and a group where nobody knows anyone. Here's the idea: let A be the largest clique. We just want to prove that the remaining vertices (call this set B) form an independent set.

- Let's say there's an edge between u and v in the set B of vertices. Also we can't add in u to A so there exists $t \in A$ such that t, u is not an edge.

- We still need to apply the main condition that among 4 of them we have three who are all friends or 3 out of whom none are. So let's take u, v, t and some other $x \in A$. We know $u - v$ and $x - a$. Note that this rules out the possibility that there are 3 out of whom none are friends (why?). So we have one of the following cases:

  1. u, v, t is a complete triangle: We know that t, u is not an edge so this is impossible.

  2. u, t, x is a complete triangle: once again this is impossible since t, u is not an edge.

  3. v, t, x is a complete triangle - this means that v is connected to t and x.

  4. u, v, x is a complete triangle - this means that u and v are both connected to x.

  So the last two cases are the only ones we need to look at. Note that in both of them, v is connected to x, and this is true for all $x \neq t$ in A, so if x, t is an edge, we can just take x into A - which is a contradiction to maximality. On the other hand, if v, t is not an edge note that u and v are both adjacent to all the vertices other than t in A. But then... (and this is the smart idea) you can kick out t and add in u and v into A - contradicting maximality again. And we're done!

Remark. The key idea here is finding ways to contradict maximality. It's important to keep looking for places where you can increase the size of a set you defined to be maximal as it can often help you get rid of cases and provide you with a path towards solving the problem.

12.4 Hamiltonian Paths and Cycles

Hamiltonian paths and cycles are another very interesting part of graph theory. We often concern ourselves with taking the *longest* path in a graph. The question is: how special does your graph have to be so that you can find a path with all vertices in it? Or a cycle which contains all the vertices? Well, the answer is: we're not really sure! There is no simple and complete characterization of *which graphs* have Hamiltonian paths - and even for a computer, there is currently no efficient method (polynomial time algorithm) to find out whether a graph has a Hamiltonian path.

Why are we bothered then? Well, if you add in certain constraints on your graph - it definitely contains a Hamiltonian path. Moreover, the existence of such a path can be really useful for a varied number of reasons! To add to that, you can also spend your time working on finding a polynomial time algorithm for checking if a graph has such a path - if you manage to do so you'd probably be the Nobel prize winner in a couple of days!

Example 11 (Ore's Theorem)

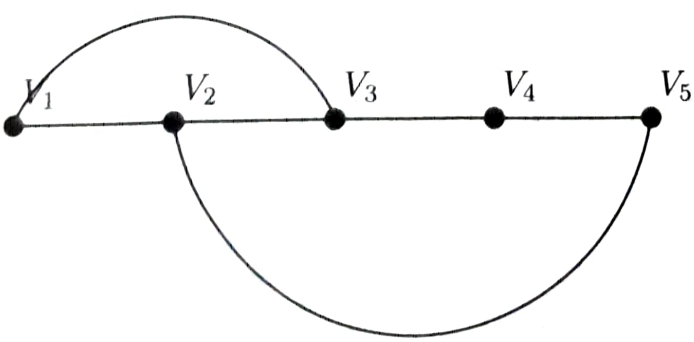
If in a graph on n vertices, for any two vertices u and v, there exists an edge between u and v or the sum of the degrees of u and v is $\geq n$, the graph has a Hamiltonian cycle.

Let's say there does exist such a graph G. We perform the following cool trick: keep adding edges to G until any edge you add will create a Hamiltonian cycle. The idea here is to essentially make a very strong condition on the graph you have which allows you to work with it more easily. So we have reached a point where adding any edge that does not already exist would create a Hamiltonian cycle. In particular, this means that the current graph has a Hamiltonian path.

Exercise. Check that the above statement is true.

So now what? Well let's say this Hamiltonian path in our graph is $v_1 - v_2 - \cdots - v_n$. We definitely don't have an edge between v_1 and v_n in our graph (or we have a Hamiltonian cycle already). We want to somehow show that a humiliation cycle already exists in the graph (that would contradict our assumption).

Let's say we have an edge from say v_1 to v_3. Then we can go from v_1 to v_3, and then go $v_3 - v_4 - \cdots - v_n$. The hope is that we can now go from v_n to v_2 because then we can go from v_2 to v_1 and complete the cycle.



In general we're looking to get $v_1 - v_r$ and $v_{r-1} - v_n$.

Exercise. Show that this must exist for some r. (Pigeon-hole! Use the condition they have given.)

Hints: 225

12.5 Tournaments

A tournament is a special type of graph where there is an edge between any two vertices, but now each such edge has an orientation too, i.e it points from one of the vertices to the other. You can imagine this as a round-robin tournament where each player plays with all others, and then this orientation helps describe whether they won the match or they lost.

Tournaments have some cool properties, and they come up a lot. Let's try proving the following:

Example 12

Consider a tournament where the ith vertex wins w_i games and loses l_i games. Show that in such a tournament,

- $\sum_i w_i = \sum_i l_i$
- $\sum_i w_i^2 = \sum_i l_i^2$

The first one is quite simple, the idea is that every game contributes one to the LHS and one to the RHS. If a certain player one a certain game, someone must have lost that game and vice versa.

More concretely, consider pairs (u, v) of vertices such that u beat v. One way of evaluating this is to fix u first as, say, v_1. We now have w_1 of picking the second vertex since v_1 won w_1 games. So the total number of pairs comes out to be $w_1 + w_2 + \cdots + w_n$. On the other hand, if we fix v first, we get $l_1 + l_2 + \cdots + l_n$, and so the two quantities must be equal.

For the second part, one simple way is to do it algebraically. Note that

$$\sum_i (w_i^2 - l_i^2) = \sum_i (w_i + l_i)(w_i - l_i) = (n-1)\sum_i (w_i - l_i) = 0$$

since $w_i + l_i$ is the total number of games each player played.

However there's another method that is closer to the way we did the first part and it also introduces a useful way of thinking about things when working with graphs. The idea is that $\binom{x}{2}$ often makes a lot more sense in a combinatorics background compared to x^2, so it is often useful to think of n^2 as $2\binom{n}{2} + n$. In particular, here we know that $\sum w_i = \sum l_i$, so it becomes equivalent to showing that

$$\sum_i (w_i^2 - w_i) = \sum (l_i^2 - l_i)$$

which in turn is equivalent to proving that

$$\sum_i \binom{w_i}{2} = \sum_i \binom{l_i}{2}$$

We now want to create some quantity, which on double counting leads us to $\binom{w_i}{2}$ and $\binom{l_i}{2}$.

So let's take some vertex a, and two of the players it beat - b and c. These triplets can be chosen $\sum_i \binom{w_i}{2}$. How do we involve l_i? Well, let's say b beat c: then c lost to both a and b. So we can also fix c first and take two players it beat. In particular, we're counting the number of triplets (a, b, c) such that a beat b and c, and b beat c. This quantity is equal to both the expressions we desire, and so the expressions themselves must be equal, as required.

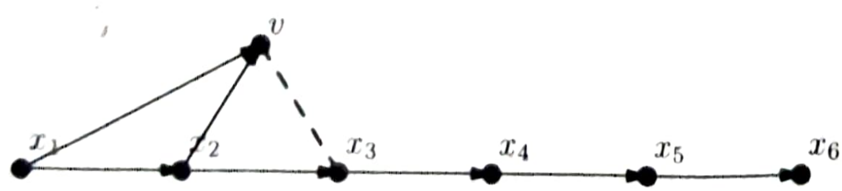
Example 13

There are n teams in a tennis tournament and every pair of teams faces off in a single match. If there are no ties, prove that the teams can be arranged in a line so that each team beats the team to its left.

Like a lot of the previous problems in this chapter, we're going to think extremally. Let's take the largest directed path in this tournament. We want to show that this contains all the vertices. So for the sake of contradiction, let's say there's some vertex v outside it. Let the path be

$$x_1 \to x_2 \to \cdots \to x_n$$

if $v \to x_1$ we can push it in at the start, so we must have that $x_1 \to v$. Now if $v \to x_2$, we can push it in between x_1 and x_2 so we must have that $x_2 \to v$. We repeat this to get that $x_i \to v$ for all i (check that this works!). So we have that $x_n \to v$, but then we can just add v at the end!



Example 14

Show that for all tournaments on n vertices, one of the following statements is true:

- The graph has a Hamiltonian cycle
- There exists a partition of the vertices of the graph into subsets A and B so that both subsets are non empty and whenever $a \in A$ and $b \in B$ the edge is from a to b.

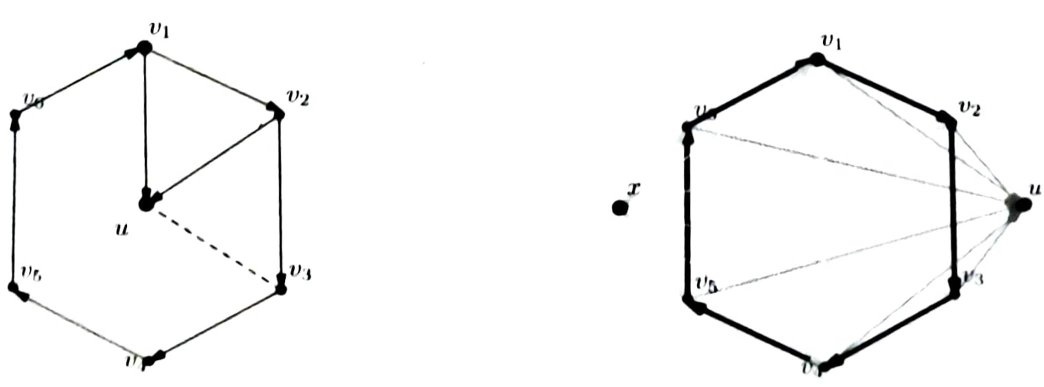
The general idea we've been following and we continue to follow is:

> "When you want to prove that an extremal condition is, well, extreme: take the extremal condition and show that it indeed is quite extreme."

Let's take the directed cycle with the maximum number of vertices in our tournaments. If this is Hamiltonian, we're done.

If not, we want to show that the second condition is true. Let's try to use the fact that there is some vertex outside and it *can't* be pushed in. Let the cycle by $v_1 \to v_2 \to \cdots \to v_k$ and an edge outside be u. Let's say $v_1 \to u$. We must avoid $v_1 \to u \to v_2$ (or else u can be pushed in) so we have that $v_2 \to u$. This in turn means that $v_3 \to u$ and so on. In particular $v_i \to u$ for all i, so u beat everyone in the cycle. In the other case where $u \to v_1$ we get similarly that u lost to everyone in the cycle.

Exercise. Show that if this isn't the case. we can rearrange the vertices slightly to add both u and x into the cycle.



At this point, we're almost done. We can essentially set u to be in one group and the cycle to be in the other group. But... what about other vertices? Nobody said there'll be only one vertex, no?

To get rid of this issue let's first for the sake of simplicity assume that u had beat everyone in the cycle. Then if x (some other vertex not in the cycle) beat everyone in the cycle, we add it to the same group as u. On the other hand, if it lost to everyone in the cycle, we show that it must have lost to u as well, and can thus be put in the cycle group.

Example 15 (Twitch Lemma)

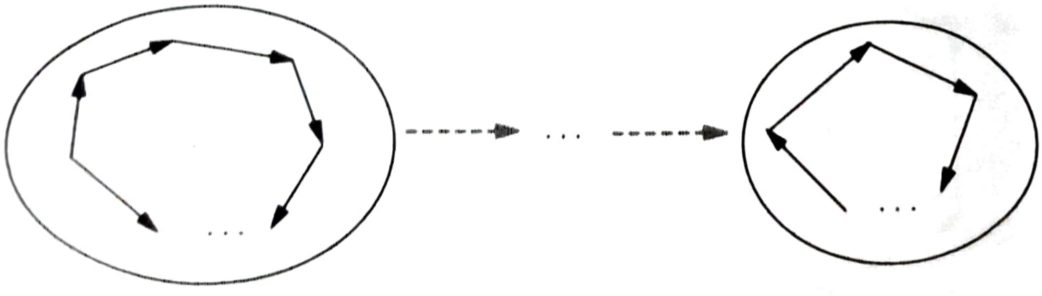
Show that if a tournament has a cycle of length $k+1$, it has a cycle of length k.

Sticking to what we're best at, let's take the largest cycle amongst those $k+1$ vertices. Wait, that's just the cycle containing all of them. Oops, never mind. Then just take the largest cycle amongst the $k+1$ vertices which, uh, does not contain all $k+1$ of them. In particular, if there are exactly k vertices here, we're done. So we need to handle the case where there are at least two vertices outside this cycle. Now apparently we can't add in just one of these.

Exercise. The arguments that follow are quite similar to the previous problem, so I leave you to fill in the details.

- By an argument similar to the previous problem, show that each vertex must beat everyone in the cycle, or lose to everyone in the cycle.
- Show (using the extremal principle again) that if u lost to everyone in the cycle and v beat everyone in the cycle then v beat u
- Can this graph even have a cycle containing all $k+1$ vertices?!

We now actually have enough information about tournaments to comment on their general structure. As it turns out, all tournaments look like the following image:



Here the dashed edges between the groups represent that everyone in the cycle on the left beat each person in all the groups on their right. Note that a group could be just one vertex. So there's a case where there are n different groups (this graph has no cycles at all), and a case when there's just one group (this graph has a Hamiltonian cycle).

Exercise. Show that this sort of representation can be created for all tournaments!

12.6 Counting in two ways

All graphs can be categorized in terms of their vertices, as well as in terms of their edges. This dual representation enables double counting to work well with graphs. In particular, a large variety of graph theory problems require you to count a certain expression in multiple ways using different interpretations. As an example consider the following problem.

Example 16 (APMO 1989)
Show that a graph with n vertices and k edges has at least

$$\frac{k(4k-n^2)}{3n}$$

triangles.

Consider any edge $u-v$ in this graph. We want to figure out how many triangles there are in this graph - so we need to figure out the number of x that are adjacent to both u and v.

So let's say temporarily that any x can be adjacent to at most one of u and v. This means that the degrees of u and v sum upto at most $1+1+(n-2)$. Here, the $1+1$ comes from the degrees contributed by edge $u-v$, and each of the other vertices contributes to at most one of the two degrees so we get $(n-2)$. So if we have that the sum of degrees is more than n, this means that there must be some triangle.

In particular, if the sum of degrees is, say, $n+5$ we get get at least 5 triangles with this single base. In general an edge $(u,v) \in E$ contributes at least $(\deg(u)+\deg(v)-n)$ triangles. However note that a triangle may get counted three times (once from each edge) so we should divide by 3.

So we have that

$$3T \geq \sum_{(u,v)\in E} (\deg(u)+\deg(v)-n)$$

We now think of the RHS differently. The idea is to think of it in terms of vertices. The degree of a certain vertex u is counted $\deg(u)$ times on the RHS (once for each edge). Hence, we can conclude that

$$\sum_{u\in V} \deg(u)^2 = \sum_{(u,v)\in E} \deg(u)+\deg(v)$$

However, by the Cauchy Schwarz Inequality (you can assume the inequality for now if you aren't aware of it).

$$\sum_{u\in V} \deg(u)^2 \geq \frac{1}{n}\left(\sum_{u\in V} \deg(u)\right)^2$$
$$= \frac{4k^2}{n}$$

Exercise. How did we get $4k^2$?

In particular, we can now conclude that

$$\frac{4k^2}{n} - nk \leq \left(\sum_{u\in V} \deg(u)^2\right) - nk = \sum_{(u,v)\in E} (\deg(u) + \deg(v) - n) \leq 3T$$

which gives us the desired result!

Problem (Mantel's Theorem). Consider a graph with no triangles on $2n$ vertices. Show that this graph has at most n^2 edges, and find the case where equality holds.

Example 17 (India TST 2011)
Consider an undirected graph with n vertices that has no cycles of length 4. Show that the number of edges is at most

$$\frac{n}{4}(1 + \sqrt{4n-3})$$

.

So we want to work with cycles of length 4 instead of triangles now. How do we do this? Let's consider some two vertices like last time - v_1 and v_2. Let's say some x and y are adjacent to both v_1 and v_2.

Then we have a cycle of length 4: $x - v_1 - y - v_2 - x$, and that's not good! So essentially, for any v_1, v_2 there can be at most one such pair.

We call such a pair $v_1 - x - v_2$, a V-pair, and we'll try to figure out the number of V pairs in a different way now. The idea is to fix x first. We can now pick v_1 and v_2 in $\binom{\deg(x)}{2}$ ways since we can take any 2 vertices that it is connected to.

On the other hand, if we pick v_1 and v_2, there should be at most one way of picking x so the number of V's is at most $\binom{n}{2}$. Thus,

$$\sum_{v\in V} \binom{\deg(v)}{2} \leq \binom{n}{2}$$

Exercise. Manipulate the LHS to get in terms of $\deg(v)$ and $\deg(v)^2$, and then convert that into E using hand-shake lemma and the inequality used in the previous problem.

Example 18 (STEMS 2023 P6)

Let $K_{n,n}$ denote the bipartite graph with two sets of S_1, S_2 of n vertices each, such that each vertex of S_1 is adjacent to each vertex of S_2. For a positive integer n, let $f(n)$ denote the largest positive integer satisfying the following property: Given any colouring of the edges of $K_{n,n}$ into two colours, red and blue, we can always find a subgraph of $K_{n,n}$, isomorphic to $K_{f(n),f(n)}$ such that all its edges are of the same colour. Determine whether the following statement is true or not: For each positive integer m, we can find a natural number N, such that for any integer $n \geq N$, $f(n) > m$.

Let's first decipher the statement a bit. We essentially want some N such that in any complete bipartite graph on N, N vertices (N vertices in each subgroup) so that no matter how we color the edges of this graph, there exists some subgraph of it which consists of m vertices from each side and all the edges are of the same color. The question asks us whether such an N exists. Intuitively at least, it should - if you have a huge graph but there are only two colors, you'll probably get lucky at some point. But this is hard to concretely draw because of the big graphs involved even for something like $m = 2$, and we need to prove things formally, so let's get down to it.

Let's actually focus on $m = 2$. We essentially want v_1, v_2, w_1, w_2 such that all four edges among them are of the same color. The idea is to fix v_1 and v_2. We're now looking for a w such that $w - v_1$ and $w - v_2$ have the same color. In particular, if we find 2 such w, we'll be happy... or would we? The issue is that we also need both the w_i to share the same colors. In particular, it is possible that $v_1 - w_1$ and $v_2 - w_1$ are red while the other two are blue. This is still fixable though, let's try and find three such w. Then we can guarentee two of the same color.

So we should have at most two such w for each pair of vertices (v_i, v_j). We can actually label this shape a V, and then we are essentially counting the number of V's. From our calculation we know that

$$\text{Number of V's} \leq 2\binom{n}{2}$$

On the other hand, we can also fix w first, and pick 2 neighbours with whom it shares the same color. For example, if there are 3 red edges and 4 blue edges, we can pick our v_i and v_j in $\binom{3}{2}+\binom{4}{2}$. In general, notice that

$$\text{Number of V's} = \sum_i \binom{B_i}{2} + \binom{R_i}{2}$$

We're going to bound this in a rather dumb way. Notice that at least one of B_i and R_i is at least $\frac{n}{2}$. (To be more precise we should probably do the following argument with floors and ceilings, but I'm just going to use $\frac{n}{2}$ for the sake of simplicity.)

So we can conclude that

$$2\binom{n}{2} \geq \text{Number of V's} = \sum_i \binom{B_i}{2} + \binom{R_i}{2} \geq n \cdot \binom{n/2}{2}$$

But the LHS is a polynomial of degree 2 and the RHS is a polynomial of degree 3, and the inequality won't hold for large enough n, and we're done!

Exercise. The general case of m replaced with 2 is almost identical to the above - you get polynomials of degree m and $m+1$. Work out the details!

12.7 Problems

Problem 1. Let $m \geq 2$ be a positive integer. At most how many elements can a set of irrational numbers contain so that among any m of them there are two with rational sum?

Hints: 202 508

Problem 2. In a sequence $a_1, a_2, \ldots, a_{2021}$ of integers, determine the maximal number of subsequences $a_i, a_{i+1}, \ldots, a_j$ $(1 \leq i \leq j \leq 2021)$ with sum 2021.

Hints: 608 122

Problem 3 (APMO 2004). Let a set S of 2004 points in the plane be given, no three of which are collinear. Let $\mathcal{L}$ denote the set of all lines (extended indefinitely in both directions) determined by pairs of points from the set. Show that it is possible to colour the points of S with at most two colours, such that for any points p, q of S, the number of lines in $\mathcal{L}$ which separate p from q is odd if and only if p and q have the same colour.

Note: A line ℓ separates two points p and q if p and q lie on opposite sides of ℓ with neither point on ℓ.

Hints: 458 382

Problem 4. In a graph G, every odd cycle is a triangle. Prove that $\chi(G) \leq 4$.

Hints: 312 564

Problem 5. In the fictional country of Mahishmati, there are 50 cities, including a capital city. Some pairs of cities are connected by two-way flights. Given a city A, an ordered list of cities $C_1, \ldots, C_{50}$ is called an antitour from A if every city (including A) appears in the list exactly once, and for each $k \in \{1, 2, \ldots, 50\}$, it is impossible to go from A to C_k by a sequence of exactly k (not necessarily distinct) flights.

Baahubali notices that there is an antitour from A for any city A. Further, he can take a sequence of flights, starting from the capital and passing through each city exactly once. Find the least possible total number of antitours from the capital city.

Hints: 546 38 647 144

Problem 6 (CMO 2019/5). A 2-player game is played on $n \geq 3$ points, where no 3 points are collinear. Each move consists of selecting 2 of the points and drawing a new line segment connecting them. The first player to draw a line segment that creates an odd cycle loses. (An odd cycle must have all its vertices among the n points from the start, so the vertices of the cycle cannot be the intersects of the lines drawn.) Find all n such that the player to move first wins.

Hints: 310 305 169 205

Problem 7 (ISL 2013 C3). A crazy physicist discovered a new kind of particle which he called an imon, after some of them mysteriously appeared in his lab. Some pairs of imons in the lab can be entangled, and each imon can participate in many entanglement relations. The physicist has found a way to perform the following two kinds of operations with these particles, one operation at a time.

- If some imon is entangled with an odd number of other imons in the lab, then the physicist can destroy it.
- At any moment, he may double the whole family of imons in the lab by creating a copy I' of each imon I. During this procedure, the two copies I' and J' become entangled if and only if the original imons I and J are entangled, and each copy I' becomes entangled with its original imon I; no other entanglements occur or disappear at this moment.

Prove that the physicist may apply a sequence of much operations resulting in a family of imons, no two of which are entangled.

Hints: 353 552 630 450

Problem 8 (Korea 2020/2). There are 2020 groups, each of which consists of a boy and a girl, such that each student is contained in exactly one group. Suppose that the students shook hands so that the following conditions are satisfied:

- boys didn't shake hands with boys, and girls didn't shake hands with girls;
- in each group, the boy and girl had shake hands exactly once;
- any boy and girl, each in different groups, didn't shake hands more than once;
- for every four students in two different groups, there are at least three handshakes.

Prove that one can pick 4038 students and arrange them at a circular table so that every two adjacent students had shake hands.

Hints: 261 127 518 131 363

Problem 9. Let G be a tournament with its edges colored red and blue. Prove that there exists a vertex v of G such that other vertex, there is a monochromatic directed path from v to it.

Hints: 147 517 485

Problem 10. A Magician and a Detective play a game. The Magician lays down cards numbered from 1 to 52 face-down on a table. On each move, the Detective can point to two cards and inquire if the numbers on them are consecutive. The Magician replies truthfully. After a finite number of moves, the Detective points to two cards. She wins if the numbers on these two cards are consecutive, and loses otherwise.

Prove that the Detective can guarantee a win if and only if she is allowed to ask at least 50 questions.

Hints: 652 108 180 626

Problem 11. $n \geq 4$ players participated in a tennis tournament. Any two players have played exactly one game, and there was no tie game. We call a company of four players *bad* if one player was defeated by the other three players, and each of these three players won a game and lost another game among themselves. Suppose that there is no bad company in this tournament. Let w_i and l_i be respectively the number of wins and losses of the i-th player. Prove that

$$\sum_{i=1}^{n} (w_i - l_i)^3 \geq 0.$$

Hints: 431 548

Problem 12. Let $x_1, x_2, \ldots, x_n$ be real numbers. Prove that there are at most $\frac{n^2}{4}$ pairs (i, j) with $1 \leq \imath < j \leq n$ such that $1 < |x_i - x_j| < 2$

Hints: 160 464

Problem 13 (CGMO 2013/3). In a group of m girls and n boys, any two persons either know each other or do not know each other. For any two boys and any two girls, there are at least one boy and one girl among them,who do not know each other. Prove that the number of unordered pairs of (boy, girl) who know each other does not exceed $m + \frac{n(n-1)}{2}$.

Hints: 167

Problem 14. Let G be a triangle free graph with n vertices and m edges. Show that the complement of the graph G (where an edge exists iff it didn't in G) contains at least

$$\frac{n(n-1)(n-5)}{24} + \frac{2}{n}\left(m - \frac{n^2-n}{4}\right)^2$$

Hints: 573 224

13 Hall's Theorem

13.1 Introduction

Before talking about any sort of technique or idea or concept, let's first consider a problem in front of us.

> **Example** (Kazakhstan 2003)
> We are given two square sheets of paper with area 2023. Suppose we divide each of these papers into 2023 polygons, each of area 1. (The divisions for the two papers may be distinct.) Then we place the two sheets of paper directly on top of each other. Show that we can place 2023 pins on the pieces of paper so that all 4046 polygons have been pierced.

If I was to tell you, this problem was connected to finding pairings of vertices in a bipartite graph, would you believe me?

Had I not known of the Hall's theorem, the above statement would feel really weird to me. Firstly, how even did we bring graph theory into the picture? As far as I can see, we were given some nice and simple polygons... or were we?

Hall's theorem, as a chapter is the epitome of the idea that

> If you knew the technique you needed to use, the problem would've been so much easier

The issue with the problems we face in this chapter is that, well, I've essentially *told* you that Hall's theorem is the way to go. Once that has been done, the problems aren't going to be half as hard as they would be if you saw them in a contest with no hints.

How do we deal with that? Well, the hope is that if you do enough problems around Hall's theorem here, you could end up with a fairly strong intuition on when you can use the theorem and what you should look out for. As we go on, we'll begin to see a lot of patterns in the type of problems we face, and that'll help us out. That being said, there will always be problems where you'd have no idea that you were meant to use Hall's theorem when you first approach the problem (which is one of the beauties of olympiad combinatorics!)

But enough talk. Let's get right into it.

13.2 The theorem

Hall's theorem is a theorem which essentially tells us when a certain bipartite graph has a perfect matching. Don't worry if that went over your head, let's first figure out what a perfect matching is.

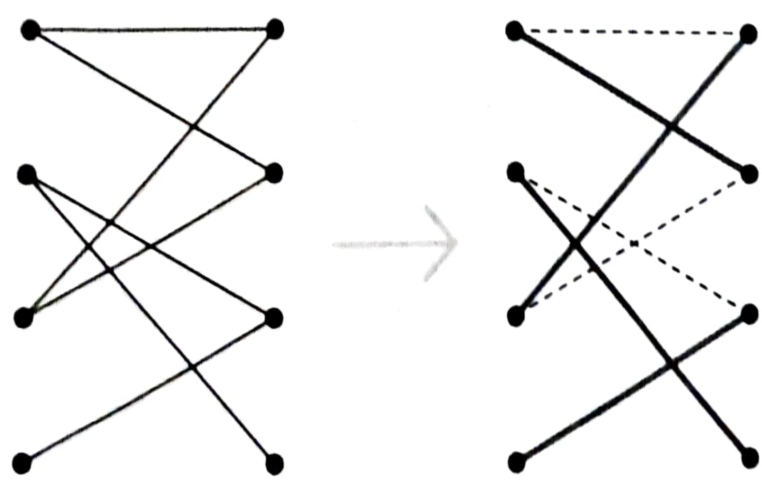
Imagine a set of n teachers and n students. Now each teacher can teach at most one student. But there are a couple of issues. Some teachers just don't like certain students. In particular, you're given a list of students each teacher is comfortable with teaching. Students don't really have a preference, they'll go on with any teacher that is ready to take them. The question:

> Can you assign a student to each teacher so that each teacher gets one of the students they were comfortable with?

Lets put that a little more formally. Imagine a bipartite graph with $2n$ vertices - n in each component. One component represents the teachers and the other one represents the students. A teacher and student are joined by an edge iff the teacher is comfortable with teaching the student. We want to match each teacher with a unique student from among her neighbours in the graph (set of students with whom she shared an edge).

Take a moment and read the set up again. The image below gives an example where there are 4 students: S_1 to S_4 and 4 teachers: T_1 to T_4

One *perfect matching* of the given graph has been shown below.

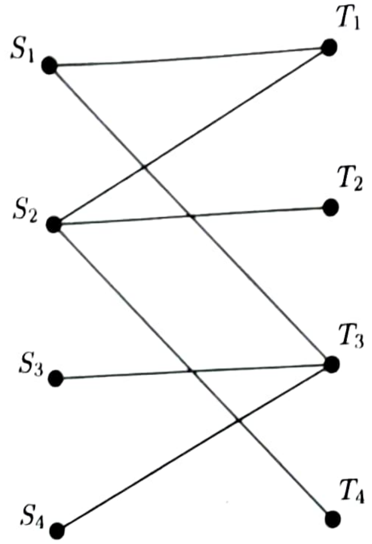


The first question to consider:

> Does there always exist a perfect matching in such a bipartite graph on $2n$ vertices?

Well, one possibility is that none of the teachers like any of the students, so there are no edges in the graph. Obviously this doesn't have a perfect matching. In particular, if a certain teacher doesn't like any of the students we would be stuck.

So let's say each teacher likes at least one student. Do we have a matching now? Unfortunately, still no - what if all the teachers like only S_1? You wouldn't be able to pair up S_2 with anyone. Okay, so let's say each student is liked by at least one teacher, and each teacher likes at least one student. Surely it works now? Actually, not really. To convince yourself, look at the graph below:



As hard as you may try, this graph just doesn't have a perfect matching. Oops! The broad reason as to why this doesn't work is that teachers T_2 and T_4 both like only S_2. Now you can assign S_2 to only one of them, so the other one wouldn't be happy. This motivates the following more general observation:

> If a certain set of k teachers are connected to less than k students, there can be no perfect matching.

"Connected" refers to all students who are liked by at least one teacher in the set. For instance, for the set T_1, T_4, students S_1, S_2 are the appropriate students.

If, like above, 2 teachers end up liking a single student, obviously there can't be a perfect matching. Similarly if there are 5 teachers but they together are connected to only 4 students, there can't be a perfect matching. And here comes the big step.

Theorem (Hall's theorem)

Let G be a bipartite graph with bipartite sets X and Y of equal size. Then there exists a perfect matching if and only if for each subset W of X,

$$|W| \leq |N(W)|$$

In a broad sense, the theorem says that if there is no set of k teachers connected to less than k students for any set of teachers, our graph has a perfect matching.

Here $N(W)$ represents the neighborhood of W - which in our case was the students that set W of teachers was connected to. So if the size of the neighborhood is at least as much as the size of the set W of teachers for every set, our graph must have a perfect matching.

We'll come to the proof of this in just a bit, but let's first see the value of this theorem. We took a very simple example of students and teachers. But Hall's theorem is in no way restricted to this. Anytime we want to pair up 2 things - a row and column, a polygon with another, or children with gifts - Hall's theorem comes right into the picture.

13.3 Proof of Hall's theorem

We'll discuss 2 proofs - one by induction and one using this idea of "alternating paths"

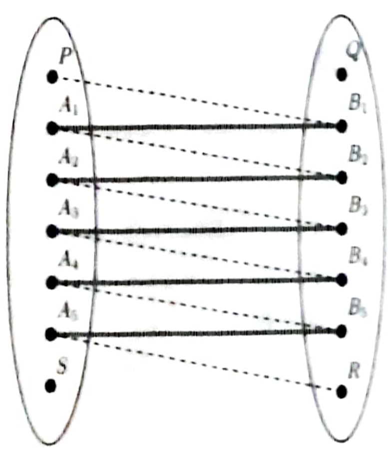
Proof using alternating paths

The idea behind this proof is essentially the following. Let's say the graph $G = A \cup B$ satisfied Hall's condition but didn't have a perfect matching after all.

Nevertheless, let's take the maximal matching in the graph - so the matching which has the maximum number of vertices involved has, say, $A_1 - B_1, A_2 - B_2, \ldots, A_k - B_k$. The issue is that there is some vertex $P \in A$ outside it.

Now P has to have *some* neighbour, but it can't be adjacent to any vertex $Q \notin \{B_1, B_2, \ldots, B_k\}$ (or else we can just add in this edge to the matching).

Now let's say P is adjacent to B_1. If A_1 is adjacent to any vertex not in the matching, we can instead select edges $P - B_1$ and $A_1 - Q$ to get a bigger matching. In general, we consider an alternating path as a path where every alternate edge belongs to the maximal matching.



If we have an alternating path that ends outside the matching, we can replace the original edges with the dashed ones to get a larger matching.

So all such alternating paths must end inside the matching. Consider all the vertices in B reachable by such a path. The idea is that no vertex outside B_i is allowed in this set of reachable vertices.

Notice that we haven't used Hall's criteria at all just yet, so it is time to use it. Let's say some i vertices $B_1, B_2, \ldots, B_i$ are reachable by an alternating path.

Apply Hall's criteria to $P, A_1, A_2, \ldots A_i$. If any of these are adjacent to anything except $B_1, B_2, \ldots B_i$, we can add that to our list of reachable vertices, which shouldn't be possible - and so Hall's criteria isn't satisfied.

Proof using induction

Let $G = A \cup B$. We use strong induction on the size of A. The base case is quite obvious, since if $|A| = 1$, this must be adjacent to the vertex in B and we have a perfect matching. Let's see how we can now prove the induction step.

The idea is the following: consider all proper subsets S of A. We know by Hall's criteria that $|S| \leq |N(S)|$. If equality doesn't hold anywhere, we can just pick a vertex u in A, pair it up with any of its neighbours v, and consider the graph G' without v and v'.

Exercise. Since we assumed that $|S| > |N(S)|$, show that $|S| \geq |N(S)|$ now so we're done by induction.

So the only case is where equality *does* hold for some subset S. The idea is that we might as well solve the problem separately for S and the remaining vertices in A in that case.

Exercise. Show that Hall's criteria must hold in each set - the first one is easy, but the second one is a little tricky! Be careful.

13.4 Example problems

Example 1 (AMSP C3 2014)

Let $k \in \{1, 2, 3, \ldots, 8\}$. Consider an 8×8 chessboard with the property that on each column and each row there are exactly k pieces. Prove that we can choose 8 pieces such that no two of them are in the same row or same column.

he craziest part about this problem is that it finds itself as an example in this apter. How on earth does this have any connection with Hall's theorem. And w on earth would I realise that Hall's theorem is the way to go?

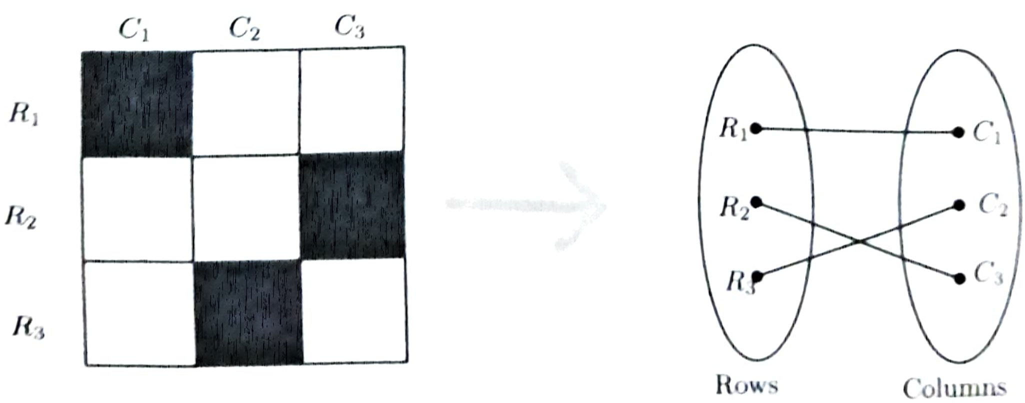
et's currently keep the second one on hold - and see if we can even solve the oblem if someone told us that we had to use Hall's theorem.

he key condition for Hall's theorem is the two sets. You need to have 2 sets of ojects, which are mutually connected. So we need two objects connected to the iessboard and the marked pieces. What should we do?

e could try connecting the chessboard with different pieces, but a "matching" akes no sense here. We could try connecting neighbouring cells of a grid and eate some sort of bipartite graph based on that but this seems far fetched since e don't really care too much about cells neighbouring each other.

kay, so we're looking for two things connected to a chessboard, their intersections ould be meaningful, and they should ideally have the same size. Two things iat together form a chessboard... of course! rows and columns!

o imagine one set to be the set of 8 rows and the other set to be the set of 8 lumns. Now what? We need to define what it means for a row to be connected a column. The most obvious interpretation of the edge between row 1 and lumn 1 is to concern ourselves with the cell $(1,1)$. In particular, that's exactly hat we use. If the cell (a,b) has a piece on it, we join row a to column b. Nice! nd what do we need to show? We need to show that we can select 8 pieces dges) such that no two are in the same row or column (share a vertex). In other ords, we're looking for a matching!



ut of course, this can't be it. After all, as we know, not all graphs have a iatching. The key piece of information we need to use is that in each row and ach column there are exactly k pieces - so the degree of each vertex in the graph k. Let's now rephrase the problem.

roblem. You are given a bipartite graph G composed of sets A and B of vertices such that $|A| = |B| = n$. Each vertex in both sets has degree k where k is a fixed integer in the range $\{1, 2, 3, \ldots, n\}$. Prove that this graph has a perfect matching.

As it turns out this is a well known result and is worth keeping in mind - any bipartite graph with all degrees equal (and non zero) has a perfect matching.

Exercise. How did I get away with not mentioning that the two components have an equal number of vertices?

Let's try proving this now. It's time to use Hall's theorem. We want to show that for each set $W \in A$, $|N(W)| \geq W$, i.e the neighbourhood of the set is at least as large as the original set.

The general idea to do this is to use the number of edges in a smart manner. When we consider a set W of size $|W|$, a total of $k|W|$ edges are coming out of it.

On the other hand, if you consider $N(W)$, each of those edges must be received by one of the members here so we know that

$$|N(W)| \cdot k \geq \text{Number of incoming edges}$$

Notice that we're using $\geq$ rather than equality because vertices in $N(W)$ may also receive edges from vertices other than W, this is only a lower bound.

Thus we have that

$$|N(W)| \cdot k \geq |W| \cdot k \implies |N(W)| \geq |W|$$

Example 2

There are twelve candies arranged in a circle, four of which are rare candies. Chad and Eric want to collaborate on a strategy for the following act. First, Eric comes and is told which four candies are rare candies, then removes four non-rare candies from the circle. Then Eric leaves, and Chad comes and must determine which of the four candies (of the eight remaining candies) are rare. Decide whether this is possible or not.

Woah. This problem has seemingly nothing to do with Hall's theorem, and yet - here we are.

It is in fact a common theme for hall's theorem to appear in such - "complete the magician and assistant trick" problems. The idea is that the two people essentially want to collaborate so that every possibility is uniquely mapped to some result (and the existence of such a mapping is shown using Hall's theorem).

Now when the magician returns, he can map back the original event from the outcome they are seeing now by using the key fact that each original event is matched with a specific outcome they are observing.

So in such problems, it's a decent idea to consider the set of original events (which is also the event that needs to be predicted) and the possible set of outcomes that the assistant creates and the magician would be looking at.

In particular, here. there are $\binom{12}{4}$ possibilities of which candies are rare. Now when the magician comes back in, he sees 8 candies remaining, and there are $\binom{12}{8}$ possibilities for what he is seeing now. As it turns out, these two numbers are equal so we just need to orchestrate some key which the assistant can value so that when the magician sees a particular set of candies remaining he can map it back to which four original candies were rare.

The cool part, though, is that we're not going to do the heavy-lifting. We'll not create this sort of mapping - we'll just show its existence as that is all the problem requires us to create. In particular, consider a graph $G = A \cup B$.

The vertices in A represent the $\binom{12}{4}$ possibilities for the rare candies. The vertices in B represent the $\binom{12}{8}$ possibilities for what the magician sees. All that is left is the edges. We want to join a vertex in A with a vertex in B if we can delete the appropriate set of four edges without deleting any edge which was rare. So in particular, each vertex in A is joined to $\binom{8}{4}$ vertices in B. Also every vertex in B is joined to $\binom{8}{4}$ edges in A since there are that many possibilities of which candies could have been rare.

So all the degrees in this graph are equal, and we're done thanks to the previous problem!

Example 3 (AMSP C3 2014)

An $n \times n$ chessboard has some of its squares painted blue. Assume that for every n squares chosen, no two in the same row or column, at least one of the squares is blue. Prove that one can find a rows and b columns whose intersection contains only blue squares, so that $a + b \geq n + 1$.

We have a chessboard once again, but this time we're prepared! So take a bipartite graph $G = A \cup B$ where A represents the set of rows and B represents the set of columns.

Okay now let's actually read the question, oops. So we have some squares painted blue. The main condition seems to be the following:

> Assume that for every n squares chosen, no two in the same row or column, at least one of the squares is blue.

So firstly, like the previous chessboard problem, let's define edges: we join a row and column if, let's say, the cell of board corresponding to the row and column is blue.

How do we handle the condition? The n squares must be such that no two are in the same row or column. What does this mean? Aha! It means that the edges in the bipartite graph corresponding to those cells would form a perfect matching.

So the problem says that if you take any perfect matching of the complete bipartite graph having n vertices on either side, at least one of the edges involved is blue. So the contrapositive is essentially that we never get a case where none of all the edges are non-blue: so there's no perfect matching of the non-blue edges in some sense.

In particular, imagine we change our graph slightly: so you create an edge if the cell corresponding to that row and column is *not* coloured blue! Now we have that this graph has no perfect matching (why!).

So Hall's criteria must be violated in this graph. This means that there exists some set of S vertices in A such that the neighbourhood $N(S)$ contains less than $|S|$ vertices. This means that if you consider any edge between a vertex in S and a vertex in $B \setminus N(S)$ it must be blue (check this!). We now use these sets of rows and columns to generate the required condition.

Exercise. Check that the bound $a + b \geq n + 1$ is satisfied since $|N(S)| < |S|$

⋆ $|X| + |C(G')/N(X)| = (|X| - |N(X)|) + |C(G')| \geq N+1$

> **Example 4** (Putnam 2012 B3)
> A round-robin tournament among $2n$ teams lasted for $2n - 1$ days, as follows. On each day, every team played one game against another team, with one team winning and one team losing in each of the n games. Over the course of the tournament, each team played every other team exactly once. Can one necessarily choose one winning team from each day without choosing any team more than once?

This is one of the problems where it is fairly clear that we want to be matching something. The problem says:

> Can one necessarily choose one winning team from each day without choosing any team more than once?

So we're essentially looking for a perfect matching between the days and the teams. But here comes the issue. There are $2n - 1$ days and $2n$ teams. So... obviously there can't be a perfect matching, right? Well, yes - but nobody asked us for a perfect matching. All we need is a matching that maps each day to a unique team - it's entirely okay for one team to be left out. In particular, we're looking for a one-one/injective function from the set of days to the set of teams.

As it turns out, the actual statement of Hall's theorem discusses injective functions. In particular, the condition that $|S| \leq |N(S)|$ is true for all subsets S of A implies that there is a matching that covers all edges in A and maps each vertex in A to a unique vertex in B. It just turns out that usually the bipartite sets end up having the same cardinality and when they do, this matching must end up covering all vertices in B as well. The proof of this version of the theorem is completely identical, and I encourage you to give it a try.

Anyway, so we just want to show that for every subset of days, there are at least as many winners over those days. Take a set S of the first 5 days, for instance. We would be in trouble if at most 4 teams won over these days. But we know that on any given day n teams win - so if $n \geq 5$ then we're happy. Specifically, if we take any subset of at most n days, there'll be n winners over any one of those days and so the cardinality of neighbourhood is at least n.

Okay, so now we just need to bother about subsets of size $n+1$ or more. Let's say we took a subset of $n+k$ days, and T teams won over those days, and we want to show that $n+k \leq T$. So we know that the remaining teams lost all of their games over these $n+k$ days. In particular, each of the remaining $2n-T$ teams lost all their games. If there were two members out of these $2n-T$ that played against each other during these $n+k$ days, one of them must have won that game - and that is clearly not possible. So any game any of these teams played must have been against one of the T teams. Thus any of these teams played at most T matches. But we know that they played $n+k$ games so we have that $n+k \leq T$, as required.

$\# \text{games} = n(n+k) \leq T(2n-T) = \# \text{possible games}$

$n^2+kn \leq 2nT - T^2 \Rightarrow T^2 - 2nT + n^2 + nk \leq 0 \Rightarrow \boxed{(T-n)^2 + nk \leq 0}$ wtf???

> **Example 5** (Brazil National Olympiad 2020)
> There are $2n$ people in a room where each pair of persons is classified as friends or strangers. Two game players from the outside play a game where they alternate turns picking one person in the room such that this person was not picked before and this person is friends with the person previously picked. The last player who can make a legal move wins. The player that moves first can pick anyone he/she wants. Prove that the player that moves second has a winning strategy if and only if the $2n$ people can be used to form n disjoint pairs such that the two people in each pair are friends.

This is a very cool problem that is built around the same ideas as Hall's theorem and its proof although it doesn't actually use it directly at any point.

Let's begin by taking the obvious graph theoretical interpretation of the problem - consider $2n$ vertices, and join two vertices by an edge if the 2 relevant people were friends. It's important to note that this time the graph isn't bipartite - but the problem still seems to have something to do with perfect matchings.

Let's try and figure out first what happens if the graph does have a perfect matching. We need to show that the second player wins in this case. So let's says the perfect matching in this graph is $v_1 - v_2, v_3 - v_4, \ldots, v_{2n-1} - v_{2n}$. Imagine A begins by moving to v_5. Where do we move? Well, we know for sure that v_5 is adjacent to v_6, so let's just move there. Now A isn't allowed to pick a vertex again, so he picks another vertex - say v_2, and we just pick v_1.

In particular, A needs to move to a new pair each time and we complete the pair. So B can make a move no matter what A does, and thus must win.

Let's now consider what happens in the other case where there is no perfect matching. We need to show that A wins in such a case.

The idea is to consider the largest matching in the graph: and let's say it has vertices $A_1 - B_1, A_2 - B_2, \ldots, A_k - B_k$. Now if you start at a vertex inside the matching, you're making life for B very easy - they just move to the other end-point.

So the idea is to start at a vertex u *not* in this matching. We now use an idea very similar to the alternating paths we used to prove Hall's theorem. The idea is that B shouldn't be able to move to a vertex v which isn't in the largest matching since we can then add in $u - v$ to create a larger matching. So B moves to some A_1, and you just move to its other end point: B_1. If B now moves to a vertex that is outside the matching, we can create a large matching using $u - A_1$ and $B_1 - x$ instead (just like the alternating path idea).

In particular, B must always keep moving to new vertices within the matching, and we can move to the other endpoint of the vertex in the matching. Thus A can always make a move, and can therefore guarentee a victory.

13.5 Problems

Problem 1. Let $G = A \cup B$ be a bipartite graph on $2n$ vertices with minimum degree $n/2$ and $|A| = |B| = n$. Show that G has a perfect matching.

Hints: 115 615

Problem 2. We have a regular deck of 52 playing cards, with exactly 4 cards of each of the 13 ranks. The cards have been randomly dealt into 13 piles, each with 4 cards in it. Prove that there is a way to take a card from each pile so that after we take a card from every pile, we have exactly a card of every rank. Also prove that, in fact, we can go further: after taking a card of every rank, there are 3 cards left in each pile. We can then take a card of every rank once more, leaving 2 cards in each pile. Finally, we do it once more, and the remaining card in each pile must be of every rank.

Hints: 554 317 185

Problem 3 (Tuymaada 2018/7). A school has three senior classes of M students each. Every student knows at least $\frac{3}{4}M$ people in each of the other two classes. Prove that the school can send M non-intersecting teams to the olympiad so that each team consists of 3 students from different classes who know each other.

Hints: 559 186

Problem 4 (Kazakhstan 2003). We are given two square sheets of paper with area 2003. Suppose we divide each of these papers into 2003 polygons, each of area 1. (The divisions for the two piece of papers may be distinct.) Then we place the two sheets of paper directly on top of each other. Show that we can place 2003 pins on the pieces of paper so that all 4006 polygons have been pierced.

Hints: 617 52

Problem 5 (Bulgaria 1997). Let $n \geq 2$ be a positive integer, and consider ordered n-tuples of distinct integers in the set $\{1, \ldots, n+1\}$. Two such tuples $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ are called *disjoint* if there exists $1 \leq i, j \leq n$ such that $i \neq j$ and $a_i = b_j$. What is the maximum possible number of pairwise disjoint n-tuples?

Hints: 206 10 492

Problem 6 (Shortlist 2010 C2). On some planet, there are 2^N countries ($N \geq 4$). Each country has a flag N units wide and one unit high composed of N fields of size 1×1, each field being either yellow or blue. No two countries have the same flag. We say that a set of N flags is diverse if these flags can be arranged into an $N \times N$ square so that all N fields on its main diagonal will have the same color. Determine the smallest positive integer M such that among any M distinct flags, there exist N flags forming a diverse set.

Hints: 595 467 504 586

Problem 7. A table has m rows and n columns with $m, n > 1$. The following permutations of its mn elements are permitted: any permutation leaving each element in the same row (a "horizontal move"), and any permutation leaving each element in the same column (a "vertical move"). Find the smallest integer k in terms of m and n such that any permutation of the mn elements can be realized by at most k permitted moves.

Hints: 283 478 34

Problem 8. Let G be a bipartite graph on $A \cup B$ with no isolated vertices. Assume that for each edge ab with $a \in A$ and $b \in B$, we have $\deg a \geq \deg b$. Prove that G contains a matching using all vertices in A.

Hints: 452 519 325 39 **Soln:** Page 282, Solution 23

Problem 9 (Shortlist 2006 C6). An upward equilateral triangle of side length n is divided into n^2 cells which are equilateral triangles of unit length. A *holey triangle* is such a triangle with n upward unit triangular holes cut out along gridlines. A diamond is a $60^\circ - 120^\circ$ unit rhombus. Prove that a holey triangle T can be tiled with diamonds if and only if the following condition holds: Every upward equilateral triangle of side length k in T contains at most k holes, for $1 \leq k \leq n$.

Hints: 291

Problem 10 (RMM 2017/5). Fix an integer $n \geq 2$. An $n \times n$ sieve is an $n \times n$ array with n cells removed so that exactly one cell is removed from every row and every column. A stick is a $1 \times k$ or $k \times 1$ array for any integer $k \geq 1$. For any sieve A, let $m(A)$ be the minimal number of sticks required to partition A. Find all possible values of $m(A)$, as A varies over all possible $n \times n$ sieves.

Hints: 79 476

IV

Advanced Topics

14 Processes

14.1 Introduction

This chapter will feel a lot like the chapter we had on soft techniques. A major part of it involves being able to understand what is going on in a certain procedure (without necessarily figuring out every detail). To do so, we'll apply our tools of simplifying the problem and trying to understand base cases. We'll also learn about some techniques which we can use to work with such processes and find out-of-the-box ways to prove our desired conclusions.

14.2 Understanding processes

Example 1 (USA TST 2017/4)

You are cheating at a trivia contest. For each question, you can peek at each of the $n > 1$ other contestants' guesses before writing down your own. For each question, after all guesses are submitted, the emcee announces the correct answer. A correct guess is worth 0 points. An incorrect guess is worth -2 points for other contestants, but only -1 point for you, since you hacked the scoring system. After announcing the correct answer, the emcee proceeds to read the next question. Show that if you are leading by 2^{n-1} points at any time, then you can surely win first place.

So there's some sort of quiz and there is some weird marking system. Note that if you get the answer right in any round, nobody can get closer to your score (the people who got it right are also at the same score and the others have had their scores reduced). So in fact, we can assume that we get the answer wrong every time, discarding any round where we get the answer right. Just like the soft-techniques chapter, we'll try and work with the base cases of the problem.

- If $n = 1$, the problem is not a lot of fun. We can just guess the same as the other contestant, and we'll obviously never get into trouble.
- Let's now work with $n = 2$. We want to decide whether we want to pick the contestant A's answer or contestant B's answer. (It is pointless to pick something other than these two - if the answer is something else, both of the contestants get it wrong and you're happy anyway.)

So the first time, let's say we pick arbitrarily - and so we just pick A's answer. Now from what we discussed earlier, we should be able to assume that we got it wrong, and so A got it wrong too, and B got it right. So B has gotten closer to us, but A has gotten further apart. Now for the next question, it makes sense to pick B's answer. If it's right, the lead remains unchanged and if it's wrong it cancels out with the first round (A gets back a point and B loses a point).

So the general idea of what we're doing is: begin by picking an arbitrary answer picked by subset of people. Now if you have the option of neutralising the advantage gained by a subset of people, we use it - and if not we'll just pick arbitrarily again. Let's try to understand this more formally.

There are now n contestants, and there are subsets $S_1, S_2, \dots, S_k$ such that members of subset S_k gave answer a_k. The first time, we just pick a_1, and some subset of people X get it right where X is one of the S_i. Now if we the exact same subset X appears later in another round as one of the S_i', we pick the same answer as them - they get it wrong and we've neutralised their advantage - but what about the set S_j' of contestants who got it right this time? Well, they all got it wrong when S_i got it right in the earlier round, so they haven't achieved any advantage either. In particular, we can now ignore both of these rounds.

Let's formalise a little more: consider a family of subsets T, initially empty. In round r, if any of the S_i matches a subset in T, pick this S_i, then delete that subset from T. If not, pick S_1 and add the S_k which got it right (not S_1) to T.

The idea is that this family has at most $2^n - 1$ subsets as it never contains the set $\{c_1, c_2, \dots, c_n\}$ (why?) and in particular, it contains at most $2^{n-1} - 1$ subsets that contain a particular contestant.

Exercise. Finish the problem!

14.3 Rules are meant to be broken

Example 2

There are 10 ants on a rod. Each of them moves at a speed 1 mm/s and the length of the rod is 10 cm. They have been given some starting positions and directions, and they begin to move with their given speed. However, here's the catch: every time two ants bump into each other, they reverse their directions. If the ant reaches the end of the rod, it falls off and dies. Find the maximum time across all possible starting positions and directions, for which at least one ant stays alive.

Okay, wow. There are a billion different things that can happen. Note that if there are very few bumps, all the ants should fall off fairly quickly. The issue is that if one ant undergoes lots of bumps, it's may stay for longer. Or will it?

The idea here is to focus ourselves on one bump. What exactly is happening? Well, an ant comes from the left, and another ant comes from the right. They bump, and then the first ant goes to the left and one ant goes to the right. Wait... so one ant is moving leftwards and one ant is moving rightwards before the bump. And after the bump, we still have one ant moving leftwards and one ant moving rightwards. Note that since the ants look identical to you, it's as if no ant changed directions at all!!

Consider the following alternate problem by breaking the rules. Ants are moving at a given speed on a rod of given length. If ants bump each other, they just keep moving on along their path as if nothing happened. Now find the maximum time for which at least one ant stays alive. But this is quite easy. Every ant is just moving in a straight line now, so they each spend at most 100 seconds on the rod, which is achieved when they are all moving leftwards and start at the left corner.

The cool part is that these problems were actually equivalent - so the answer for the actual problem is 100 as well!

Example 3

Numbers 1 to n are written in some order. At any point, if number k is at the start, we remove it, and place it so that it is now in the kth position. Show that eventually 1 is at the start. (At which point the process ends because 1 stays at the start.)

The idea is to induct on n. If n ever comes at the start, it goes to the end, and then it stays there because nothing can go after it. Thus, if n ever comes at the start, it would be as if we only have $n-1$ numbers after that - and it's all good.

But what if n *never* comes at the start? Then either it's already at the end - and we're happy, or some other number k is at the end. Note that k stays at the end since the only one who could displace it was n and that never happens. Hence we can conclude that n and k were never touched.

It would've been great if it was n at the end... but if k and n are never *operated on*, does it matter if we exchange the two? It does not!

In particular, let's say there is some permutation for which 1 never comes at the start and n is never operated on. Then we can exchange k and n and 1 should not be able to come to the start of this new permutation as well. But now that n is at the end, we can use induction to finish.

Example 4 (Shortlist 2018 C3)

Let n be a given positive integer. Sisyphus performs a sequence of turns on a board consisting of $n+1$ squares in a row, numbered 0 to n from left to right. Initially, n stones are put into square 0, and the other squares are empty. At every turn, Sisyphus chooses any nonempty square, say with k stones, takes one of these stones and moves it to the right by at most k squares (the stone should say within the board). Sisyphus' aim is to move all n stones to square n. Prove that Sisyphus cannot reach the aim in less than

$$\left\lceil \frac{n}{1} \right\rceil + \left\lceil \frac{n}{2} \right\rceil + \left\lceil \frac{n}{3} \right\rceil + \cdots + \left\lceil \frac{n}{n} \right\rceil$$

turns. (As usual, $\lceil x \rceil$ stands for the least integer not smaller than x.)

You have some piles of stones. In a move, you pick a pile, and take a stone from it - but you can move it to a distance that is at most the number of stones in the pile it started off in. And we need to bound on the number of moves taken to shift everything from pile 0 to pile n.

Notice that we can shift the first stone in just one move... and that seems to resemble the $\lceil \frac{n}{n} \rceil$ term, perhaps?

Let's look at the bottom end of the pile. In particular, look at the stone at the bottom of pile 0. Eventually, this is the only stone in pile 0 and at that point you can move this, but you can move it at most one pile to the right - so pile 1. If pile 1 didn't already have a stone, this needs to be shifted to the right by 1 again.

The hope is that we can show that the bottom stone thus contributes at least $\lceil \frac{n}{1} \rceil$ to the sum. But here comes in the issue: let's say when we move to pile 2, this pile already had a stone. We can now move the stone by 2 if we wish.

However, pile 2 still has a *bottom* stone. So when we get rid of the other stones in pile 2, we'll have to make a jump by exactly one to pile 3. Then pile 3 has a bottom stone (which may or may not be this stone) so this means that eventually that bottom stone will need to be moved - and it can be moved by only one to pile 4. Repeating this we get in our factor of $\lceil \frac{n}{1} \rceil$. But what have we even counted in the process? We have not counted the number of moves of a pebble, because we were changing our pebble to the bottom one every time.

Let's keep these questions on hold. Similar to the situation for the bottom stone, this time we are hoping that the second stone from the bottom in the original pile contribute $\lceil \frac{n}{2} \rceil$ to the summation.

Let's wait until there are only two stones in the first pile. At this point, that stone must move - and it can move by either 1 or 2. If the stone falls into a pile with at least one stone, this pile will now have a second last stone, which will be able to move by a distance 1 or 2. If it falls into an empty pile, it'll be able to move by only 1. Either way, this second last stone is not moving by more than 2 and so we should have got $\left\lceil \frac{n}{2} \right\rceil$ moves here.

There's still a lack of clarity on what these numbers are. Since we're changing the pebble we're looking at pretty often, there's a pretty scary chance that we're over counting - what if we counted a certain pebble's move in the $\left\lceil \frac{n}{1} \right\rceil$ as well as $\left\lceil \frac{n}{2} \right\rceil$?

The only real way to avoid this would be to somehow show that the bottom stone actually moves at least $\left\lceil \frac{n}{1} \right\rceil$ times. But... it just doesn't, no? How are you going to prove something that isn't true. Well, by making it true. What if I just push that bottom stone to the bottom of the new pile! Clearly everything's identical, so this is obviously legal.

How does this generalise for the second stone, though? Let's try placing it at the bottom as far as possible (unless the first stone is at the bottom, in which case we place it second from bottom). Then it is always one of the bottom 2 stones, so it always moves in steps of at most 2.

In general, label the stones in the pile 0 at the start as 1 to n with 1 being at the bottom. If you're placing stone k into a different pile, place it so that the pile is sorted in increasing order from bottom to top. Then the stone labelled k moves in steps of size at most k, and we're done!

Exercise. Convince yourself that this should work, and then prove it.

Also note that the bound is $\mathcal{O}(n \log n)$. Show that this can be achieved. (Hint: Divide and conquer).

Example 5 (IMO 2021/5)

Two squirrels, Bushy and Jumpy, have collected 2021 walnuts for the winter. Jumpy numbers the walnuts from 1 through 2021, and digs 2021 little holes in a circular pattern in the ground around their favourite tree. The next morning Jumpy notices that Bushy had placed one walnut into each hole, but had paid no attention to the numbering. Unhappy, Jumpy decides to reorder the walnuts by performing a sequence of 2021 moves. In the kth move, Jumpy swaps the positions of the two walnuts adjacent to walnut k.

Prove that there exists a value of k such that, on the kth move, Jumpy swaps some walnuts a and b such that $a < k < b$.

Let's try and understand the problem. So you have some numbers written down in a circle, and in the kth move you swap the positions of the numbers adjacent to k. We want to show that at some point the two numbers we swap will be such that the value k belongs to the range (a, b).

Let's try and see what happens for say, $n = 4$ ($n = 3$ is not as fun - you'll always swap 3 and 1 on the second turn). Let's position the numbers as:

$$\begin{array}{cc}\underline{1} & 4\\ 3 & 2\end{array} \to \begin{array}{cc}1 & 3\\ 4 & \underline{2}\end{array} \to \begin{array}{cc}1 & 4\\ \underline{3} & 2\end{array} \to \begin{array}{cc}2 & \underline{4}\\ 3 & 1\end{array} \to \begin{array}{cc}1 & 4\\ 3 & 2\end{array}$$

Huh, this doesn't seem to work at all. In fact, all even numbers don't work (!) - the idea is that you can place the numbers from 1 to n in between gaps of numbers from $n + 1$ to $2n$. So for the first n moves, you'll be swapping numbers $\geq n + 1$ and for the next n moves, you'll be swapping numbers $\leq n$.

Hence the problem must depend on the fact that 2021 is odd. Let's try and understand the process in a little more depth.

Notice that we only really care about the neighbours of a number are both bigger/smaller - we don't actually care about their magnitude: 18 and 21 are no different to you if $k = 5$. Thus, the trick is to think of numbers smaller than k as one type of pebbles, and numbers bigger than k as another type of pebble. In particular imagine the following *process*.

> Start by coloring all the walnuts yellow. On the kth move, color the walnut that has k written on it using the color red.

Now if both the neighbours are more than k, neither of them is going to be red. Similarly, if both neighbours are less than k, both are going to be yellow. Let's say the problem is false - so there is some starting arrangement so that each move consists of one of the above two types. Then as far as the colors are concerned, exchanging the two doesn't really make a difference as either both are red or neither is. Thus all we're doing, is picking a yellow walnut which has both neighbours of the same color, and making it red. We want to show that we cannot color all walnuts in this process.

Remember that we haven't used the fact that 2021 is odd just yet.

Exercise. Use the fact appropriately to conclude that after the first move we'll always have a contiguous segment of an even number of yellow walnuts. Further, if we reach a stage $R - Y - Y - R$, the yellow's can never be turned into red.

> **Remark.** The biggest question that should be on your mind is:
>
> "How do *I* come up with the coloring idea?"
>
> If you think about it after having solved the problem, it feels fairly natural. However coming up with it is a lot harder than it may seem after having read the solution. It is very easy to get drawn out by the vastness of the

problem - and as it turns out, it is quite hard to understand what is going on in its entirety. It's also possible to get drawn to ideas such as

> "What happens if 1 and 2 are adjacent?" "What happens if 1 and 3 are adjacent?" and so on.

While these help a fair bit in solving base cases, you'll realise that these ideas don't generalise too nicely - it's hard to figure out what happens if say 1 and 27 are adjacent in the 2021 walnuts case.

The way I look at it is that there are two types of problems.

- There is too much distinction, which just creates a distraction while trying to understand the problem
- Too many things are identical, and it's hard to concretely imagine what is going on.

We're given a huge sample space to analyse, and we're finding that pretty tough but if you pay attention, you only really care about whether the number is bigger than the current k, or not. So when there's too many things to handle, it's often useful to *purposely lose track.* We only care about the color of a number. This contrasts the previous problem where all the pebbles were originally given to you identically, but you label them differently to solve the problem.

Either way, the key trick is to look *beyond* the problem. The problem calling stones identical shouldn't stop you from deciding that you'd like them to be distinct. On the other hand, the problem naming them differently doesn't prevent you from coloring them somewhat identically. Think out of the box!

14.4 The superposition trick

Example 6 (CIIM 2017 Problem 6)

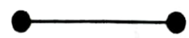
Let G be a simple, connected graph on n vertices. An invisible rabbit and a hunter play a game on this graph. The rabbit starts at an arbitrary vertex, unknown to the hunter. They take turns alternatively. At his turn, the rabbit must move from his current vertex to an adjacent one. At the hunter's turn, he picks a vertex and checks if the rabbit is there. Characterize the graphs G such that the hunter has a strategy to capture the rabbit in a finite number of turns, regardless of the rabbit's initial position and movement.

The problem says that there is some graph where a rabbit is moving. Each day the hunter picks a vertex and checks if the rabbit is over there. If it is, the rabbit is caught. Else, the rabbit has to move to one of the neighbouring vertices. After this the hunter picks a vertex again. We want to categorize all graphs such that the hunter can guarentee catching the rabbit after a finite number of moves.

"All graphs" is a pretty big statement. In fact, it is hard to believe that you can catch the rabbit in any graph whatsoever. So the first step:

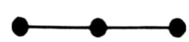
> Convincing ourself that graphs where the hunter has a strategy do exist.

- Let's start off with the simplest possibility: a graph on two vertices (which must be connected by an edge since graph is connected).

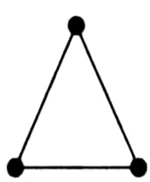
  

  So let's say we check v_1 on the first day. The key insight is that if he wasn't on v_1, he must have been on v_2, and since he *must* move he will have to be at v_1 on the next turn. So check v_1 again, and you're guarenteed to have caught him this time.

- Okay, slightly harder: what about a graph on three vertices with $v_1 - v_2 - v_3$. The trick is to check v_2. If the rabbit wasn't at v_2, it would be on the next day (check this), and it's caught once again!

- How about a triangle? so v_1, v_2, v_3 and they're pairwise connected.

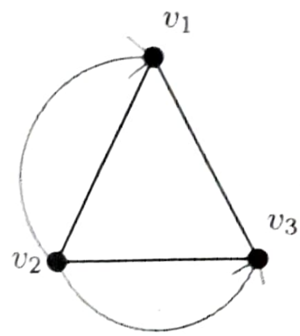
  

  Let's say we check v_1 on the first day - so either we catch it, or we get to know that it was on one of v_2 and v_3. If it was on v_2, it can now be on either of v_1 and v_3 and in the other case we know that it has to be on one of v_1 and v_2. All in all, it can now be on any of v_1, v_2, v_3: so it seems like we *gained* zero information - and thus the rabbit can't be caught. We'll need a way to formalise this further though.

Enter the technique of superpositioning! Don't worry, this has nothing to do with quantum physics, I just thought the name sounded cool. Anyway, the idea is that we let the rabbit exist in a superposition of two states: being in v_2 and being in v_3. We only decide which of the two only after the hunter makes his guess so that we aren't caught (somewhat like the Schrodinger's cat story except we make the hunter unlucky by force rather than leaving anything to luck).

I'm sure you're not convinced. If you are - well, read again. The question is "Why on earth is that allowed?!"

The idea is that apparently the hunter is so confident that he can win, that he says that I have a strategy that works regardless of where you started and how you moved. In particular, if such a strategy does exist, he might as well show it to the rabbit - so the rabbit can essentially know beforehand what the hunter is going to do. If he is so doomed that he'll be caught regardless, this shouldn't help the rabbit at all.



So he sees the first move of the strategy - let's say it's that the hunter is going to check v_1. He positions himself at v_2. Now the next day, he can move to either of v_1 and v_3. The hunter can check at most one of these (and the rabbit *knows* which one the hunter is going to check, if any) so he positions himself in a way that he avoids being caught. In general, at any point if he is at v_i, he can be on either of the other two on the next turn and can decide which one he likes more based on the hunter's strategy list.

A slightly funnier way of saying this (which makes life easy in some cases) is the following:

> As it turns out, the rabbit has an ego of his own, and decides to give the hunter the following rather useful (and true) information before the hunter makes his first move,
>
> "Hey, I'm at one of v_2 and v_3"
>
> He gives this additional information just to bully the hunter knowing that he has no hope of having a strategy for catching you either way. If he checks one of v_2 and v_3 - you just position yourself at the other one (superpositioning!). Once again, you only have the freedom to do this because the hunter's strategy should have worked no matter what you did. So he knows that you were at, say, v_3 because he checked v_2 but now you're at one of v_1 and v_2. So he has gained *no information*, or in other words, we're back to a position isomorphic to the starting one: so it's as if the game just started and you can repeat this entire superpositioning trick.
>
> Also if for some reason the hunter decided to check v_1 (he thought you were lying), he'll see that you aren't there and further, you decide to

tell him.

"Fine, I was at v_2, so now I am at one of v_1 and v_3"

So once again we're back to scratch.

Wow, that was a lot of effort just to solve the problem for $n \leq 3$. However it does give us the intuition necessary to solve a big chunk of this problem. Let's begin with the following, rather big, question

Can the graph have a cycle?

Expectedly, the answer is no. If the graph does have a cycle we can implement a technique similar to the one for a triangle with a slight modification to create a contradiction.

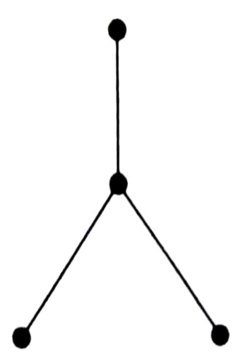
In particular, what we can do this time is say we're at one of v_1 and v_3 (where the cycle is $v_1 - v_2 - \cdots - v_n - v_1$). Now we position ourselves on the one the hunter didn't check on the first turn (if he didn't check either, tell him where you were) - so the hunter knows that you were at, say, v_3. But then you're now at one of v_2 and v_4, a situation isomorphic to the one before and we can repeat the same argument repeatedly.

So the graph must be a tree! Do all trees work? Let's find out.

- For $n = 2$ and $n = 3$, there is only one possible tree and it works, but let's dig further.
- For $n = 4$, there are two possible graphs - a star and a line (a star graph in general is a tree where one vertex is connected to all other vertices and a line is a single path from v_1 to v_n). Let's try these out one at a time.

  1. The star graph: This is actually not so hard at all: we can check the center node twice, and we're done.

     Exercise. Check that this works for a star graph with any number of nodes

  2. The line: This is a lot trickier than the previous case, but as it turns out solving it is still possible. Let's try and discover how we may do this.

First guess: why don't we just keep choosing a single vertex? Let's say we keep picking v_2 in our graph $v_1 - v_2 - v_3 - v_4$. Then the rabbit could keep alternating between v_3 and v_4 and thus avoid being caught.

Second guess: Alright never mind, what if I alternate between say v_2 and v_3? Well the issue is that the rabbit could always alternate between them in a way that it is never caught. For example if you check in the order v_2, v_3, v_2, v_3, the rabbit could position itself as v_3, v_2, v_3, v_2. Oops.

Third guess: So we want to avoid this sort of repetitive structure - how about I try something like v_2, v_3, v_3, v_2

Exercise (Tricky). Show that this checking pattern works for the hunter.

(Don't read forward until you've done the above.)

Let's now try to generalise this to a big line $v_1 - v_2 - \cdots - v_n$. The idea was that repeating v_3 allows us to break the parity-based cheat code that the rabbit was using (it would stay on an odd numbered vertex when we were on an even one and vice versa).

Note that a tree is bipartite and we can therefore color the vertices with two colors so that any edge is between red and blue vertices. We construct a strategy that works if the rabbit begins on a red vertex.

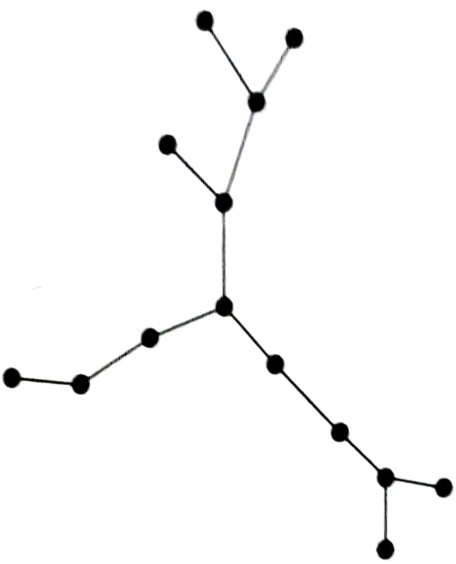
If our strategy completes without the rabbit being caught, we know that the rabbit must have started on a blue vertex, and since we know the number of moves that have taken place, we know the color of the vertex the rabbit is now. In particular, we can waste a move if needed to make sure that the rabbit is going to move to a red vertex next, and then repeat the initial strategy.

Essentially this would look something like v_2, v_3, v_1(wasted move), v_2, v_3.

Exercise. Show that if we have assumed that the rabbit begins on an even numbered vertex in a line graph, we can just say $v_2, v_3, \ldots, v_{n-1}$ one by one and we're guaranteed to have caught the rabbit!

- Lines and stars both work, and since they are pretty different from each other, a conjecture we may come up with is that all trees work. Unfortunately and very surprisingly that is false (!!).

So what now? Let's try and find a place where the rabbit could possibly have some sort of strategy. One thing with a line/star is that either they have either high degree but short paths, or long paths but low degrees. In particular, let's see what happens if we have multiple neighbours, and large paths from each of them. Imagine a scenario something like the below image.



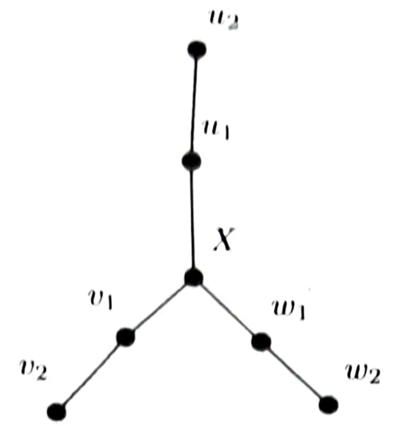
The idea for the hunter possibly remains similar - he can pre-decide the parity of the starting point of the rabbit without any issue. What if he just treats this like three big lines?

There is an issue, though. Let's label the three regions as A, B and C. It is possible that when the hunter was moving around in region A, the rabbit was sitting in region B. When the hunter is exiting and coming towards region B, the rabbit moves to region C instead. Then when the hunter comes to C the rabbit has already moved to region A, and so on. There are some issues, though. For starters, the hunter doesn't have to move linearly, he can always just jump from one region to the next - at the cost of not completing a complete scan of the region, though.

So what do we do? Who wins? Hunter? Or Rabbit? Urgh. This is definitely the trickiest part of every problem, and more so this problem - making a good guess and trusting your intuition.

There's a good chance this is doomed to fail but let's try to help the hunter win. So firstly, we can color the graph with two colors as before and assume that the rabbit starts off on a certain color we like.

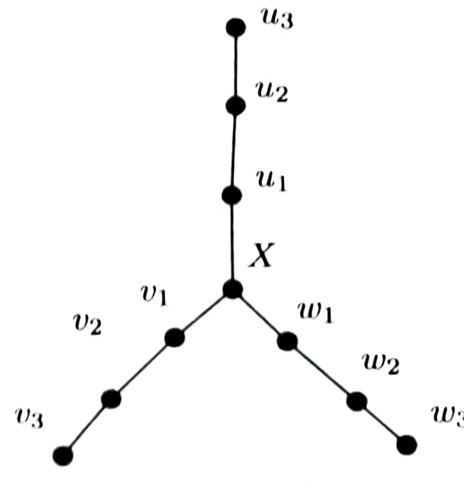
Okay, so let's say we have 2 vertices in each region as follows.



Exercise. Assume that the rabbit begins at one of u_1, v_1, w_1 and find a way to catch it by eliminating all possibilities in a series of steps.

Solution. Check u_1. Not there? It must have been in v_1 or w_1 so it's in X, v_2, w_2 now. Check X. Not there? Must have been in v_2 or w_2, so v_1 or w_1 now. Check v_1. Not there? Must have been in w_1 then, so X or w_2 now. Check X and then check w_1, done. □

But what if I make this slightly more complicated, like so:



We want to eliminate some possibilities - so let's say once again we assume as the hunter that the rabbit began on one of the vertices X, u_2, v_2, w_2.

The hunter can eliminate a possibility by selecting v_2 (the rabbit can't be on v_3 after this move). Now the hunter can pick v_1 - if the rabbit isn't there, that means we won't be on v_2 next! So the hunter now knows that the rabbit on one of X, u_2, w_2. As it turns out, however, this is not as useful as you may hope.

In fact, we now use the superposition trick again, to show that we can evade the hunter forever. The rabbit says: "Oh, so apparently you can figure out that I'm not at v_2 after some effort. No worries, I'll just tell you that I'm not there. In particular, let's say I'm at one of X, u_2, w_2.

If the hunter picks X, I position myself at one of the other two so that I am now at one u_1, u_3, w_1, w_3. Note that no matter where the hunter checks now, I can find a way to survive and end up at each of X, u_2, w_2 on the next turn (Show this!) - so you've gained no information.

And if you check u_2, I'm now at one of u_1, v_1, w_1, w_3. If you check w_1 or w_3, I can now end up at each of X, u_2, v_2. If you check v_1, I can end up at u_2, w_2, X and if you check u_1 I can end up at v_2, w_2, X - and all of these are isomorphic to the starting position. And if the hunter picks something else at the start, we just tell him we weren't at X and the game goes on the same way.

Finally, let's finish the problem.

Exercise. Characterise all trees avoiding this sort of configuration, and show that there is a solution to catch a rabbit on all such graphs.

Soln: Page 282, Solution 24

That... was one long example. The entire solution has been attached again as linked above, feel free to read that if you need to grasp things more concretely.

The biggest take away from this problem is the superposition idea. It may not even make it into the final solution, but the application of superposition in the 3 region case is what shows its true power: the simplification of a problem. The main idea behind superpositioning on a specific set of possibilities rather than over all of them is the following.

> It is possible, that by performing a set of very smart moves the hunter *does* infact gain some information about what is going on like when he figured out that you aren't at v_3. However, after the hunter gains a certain amount of information - he becomes clueless as to how to move forward. The idea is to save the hunter that pain by giving him that information at the start itself. The reason this helps is that you can simplify your working a lot - rather than looking at possibilities where you could be at a 100 different places and the hunter could do a 100 different things, you reduce to a case where both you and the hunter can perform a very limited set of moves. In particular, this makes the conditions of the problem easier to deal with.

In fact, this trick is not just limited to this problem. It can come up whenever the problem looks something like

> Show that player A can do ... no matter how B plays.

If A has such an unbreakable strategy, he might as well show it to B, giving rise to this entire argument. Let's do another problem which incorporates this idea.

Example 7 (USA TST 2015/3)

A physicist encounters 2015 atoms called usamons. Each usamon either has one electron or zero electrons, and the physicist can't tell the difference. The physicist's only tool is a diode. The physicist may connect the diode from any usamon A to any other usamon B. (This connection is directed.) When she does so, if usamon A has an electron and usamon B does not, then the electron jumps from A to B. In any other case, nothing happens. In addition, the physicist cannot tell whether an electron jumps during any given step. The physicist's goal is to isolate two usamons that she is 100% sure are currently in the same state. Is there any series of diode usage that makes this possible?

Once again, we have a player who wants to guarentee something - he is essentially playing against the *usamons* themselves (or some evil player who decided how the usamons were initialised). No matter how the usamons are initialised, you need to isolate two of them which eventually end up in the same state.

The move you're allowed to do is convert $(1,0) \to (0,1)$. In all other cases the pair remains the same. So essentially you know at the end that either B has an electron, or neither have an electron - which is definitely non zero information.

What we need to look for, is a way to *neutralise* this information by somehow already telling the physicist that it is going to be true before hand, while making sure that the physicist can't use this information to solve the problem. Okay, let's say that for each pair of electrons I give you the information that either both of them are zero or a specific one of them is guarenteed to be 1.

The issue, though, is the following: "How are we going to decide which of them we want to to guarentee to be 1?"

- For instance let's say that in the pair (a,b) we have said that either usamons a and b are both zero or b is 1. A decent way of interpreting these conditions is the following:

$$b = 0 \implies a = 0$$

- Similarly for pair (b,c) let's say we have said that

$$c = 0 \implies b = 0$$

What should we do for (a,c)? We know that $c = 0 \implies b = 0 \implies a = 0$ (check this!). So if we set the condition for (a,c) as: either usamons a and c are both zero or a is 1, this would mean that $a = 0 \implies c = 0$, so a and c would be guarenteed to be the same - that's not good at all for us. So we define the condition as

$$c = 0 \implies a = 0$$

for these two.

Generalising a bit, essentially we have for each a_i, a_j either $a_i = 0 \implies a_j = 0$ or the other way round. All we want is that there should be no cycle in this directed graph, i.e there should be no way of guaranteeing $a_i = 0 \implies a_j = 0$ and $a_j = 0 \implies a_i = 0$.

A simple way of deciding which condition you want is the following: let's say $i < j$. Then we say that

$$a_i = 0 \implies a_j = 0$$

Clearly this can't have any cycles (why?) - so we've gotten somewhere, but we're still not exactly sure what we've achieved. First let's interpret the above condition slightly differently. Essentially it says that if $a_i = 0$ then everything to the right of it is zero as well, so our sequence looks like

$$1, 1, \ldots, 1, 0, 0, \ldots, 0$$

The only missing information is the transition point: the point where we start getting zeroes. It could be the first one itself, and be the last one too.

So here comes the magic: you go and say to the physicist

> "The usamons exist in one of the following states:
>
> 1. $1, 1, \ldots, 1, 1$
> 2. $1, 1, \ldots, 1, 0$
>
> $\vdots$
>
> 2015. $1, 0, \ldots, 0, 0$
> 2016. $0, 0, \ldots, 0, 0$
>
> Good luck."

From now, the usamons stay in superposition of these 2016 states and they only take a deterministic value after the physicist makes her guess. If at a certain point these are the possible states clearly the physicist cannot guarentee anything since if she picks i and j, it's very much possible that one of them is the transition point and both of them end up with different values.

We show that no matter what the physicist does, she cannot change her fate (by changing the possibilities).

Now if the physicist makes a move which attempts to move an electron from a_i to a_j with $i > j$ that's clearly not going to work (why?).

For the final blow (and this is pretty tricky to come up with!), if the physicist attempts to move an electron from a_j to a_i with $i > j$, we perform the move if required and then swap the labels of i and j. Done!

Exercise. Part one: Convince yourself that this works. Part two: Prove that it does.

14.5 Problems

Problem 1 (USAMO 2016/6). Integers n and k are given, with $n \geq k \geq 2$. You play the following game against an evil wizard. The wizard has $2n$ cards; for each $i = 1, \ldots, n$, there are two cards labeled i. Initially, the wizard places all cards face down in a row, in unknown order. You may repeatedly make moves of the following form: you point to any k of the cards. The wizard then turns those cards face up. If any two of the cards match, the game is over and you win. Otherwise, you must look away, while the wizard arbitrarily permutes the k chosen cards and then turns them back face-down. Then, it is your turn again.

We say this game is *winnable* if there exist some positive integer m and some strategy that is guaranteed to win in at most m moves, no matter how the wizard responds. For which values of n and k is the game winnable?

Hints: 566 446 422 598

Problem 2 (ELMO SL 2019 C3). In the game of Ring Mafia, there are 2019 counters arranged in a circle, 673 of these which are mafia, and the remaining 1346 which are town. Two players, Tony and Madeline, take turns with Tony going first. Tony does not know which counters are mafia but Madeline does.

On Tony's turn, he selects any subset of the counters (possibly the empty set) and removes all counters in that set. On Madeline's turn, she selects a town counter which is adjacent to a mafia counter and removes it. (Whenever counters are removed, the remaining counters are brought closer together without changing their order so that they still form a circle.) The game ends when either all mafia counters have been removed, or all town counters have been removed.

Is there a strategy for Tony that guarantees, no matter where the mafia counters are placed and what Madeline does, that at least one town counter remains at the end of the game?

Hints: 238 184 659

Problem 3 (IMO 2016/6). There are $n \geq 2$ line segments in the plane such that every two segments cross and no three segments meet at a point. Geoff has to choose an endpoint of each segment and place a frog on it facing the other endpoint. Then he will clap his hands $n - 1$ times. Every time he claps, each frog will immediately jump forward to the next intersection point on its segment. Frogs never change the direction of their jumps. Geoff wishes to place the frogs in such a way that no two of them will ever occupy the same intersection point at the same time.

(a) Prove that Geoff can always fulfill his wish if n is odd.

(b) Prove that Geoff can never fulfill his wish if n is even.

Hints: 28 251

Problem 4 (ELMO SL). Adithya and Bill are playing a game on a connected graph with $n > 2$ vertices, two of which are labeled A and B, so that A and B are distinct and non-adjacent and known to both players. Adithya starts on vertex A and Bill starts on B. Each turn, both players move simultaneously: Bill moves to an adjacent vertex, while Adithya may either move to an adjacent vertex or stay at his current vertex. Adithya loses if he is on the same vertex as Bill, and wins if he reaches B alone. Adithya cannot see where Bill is, but Bill can see where Adithya is. Given that Adithya has a winning strategy, what is the maximum possible number of edges the graph may have? (Your answer may be in terms of n.)

Problem 5 (Shortlist 2009 C5). Five identical empty buckets of 2-liter capacity stand at the vertices of a regular pentagon. Cinderella and her wicked Stepmother go through a sequence of rounds. At the start of each round, the Stepmother distributes one liter of water arbitrarily over the five buckets. Then Cinderella chooses two neighboring buckets and empties them; the next round then begins. Can Cinderella always prevent the Stepmother from causing a bucket to overflow?

Hints: 607 533 196 584

Problem 6 (Shortlist 2011 C5). Let m be a positive integer, and consider a $m \times m$ checkerboard consisting of unit squares. At the center of some of these unit squares there is an ant. At time 0, each ant starts moving with speed 1 parallel to some edge of the checkerboard. When two ants moving in the opposite directions meet, they both turn 90° clockwise and continue moving with speed 1. When more than 2 ants meet, or when two ants moving in perpendicular directions meet, the ants continue moving in the same direction as before they met. When an ant reaches one of the edges of the checkerboard, it falls off and will not re-appear.

Considering all possible starting positions, determine the latest possible moment at which the last ant falls off the checkerboard, or prove that such a moment does not necessarily exist.

Hints: 429 321 272

Problem 7 (Shortlist 2005 C5). There are n markers, each with one side white and the other side black. In the beginning, these n markers are aligned in a row so that their white sides are all up. In each step, if possible, we choose a marker whose white side is up (but not one of the outermost markers), remove it, and reverse the closest marker to the left of it and also reverse the closest marker to the right of it. Prove that, by a finite sequence of such steps, one can achieve a state with only two markers remaining if and only if $n - 1$ is not divisible by 3.

Hints: 395 420 376 215

Problem 8 (Chip Firing). Let G be a connected graph. There are k frogs, each placed on some vertex of G. At each second, you pick a vertex v that contains at least $\deg(v)$ frogs, and then $\deg(v)$ of the frogs on v jump, one on each of the $\deg(v)$ adjacent vertices. Call each such move a firing and call a configuration stable if no moves can be made.

Prove that if by some finite sequence of firings, we can reach a stable configuration S, then, all sequences of firings will end in S after a finite sequence of moves.

Hints: 618 226 619 151 527

Problem 9 (2001 ISL C7). A pile of n pebbles is placed in a vertical column. This configuration is modified according to the following rules. A pebble can be moved if it is at the top of a column which contains at least two more pebbles than the column immediately to its right. (If there are no pebbles to the right, think of this as a column with 0 pebbles.) At each stage, choose a pebble from among those that can be moved (if there are any) and place it at the top of the column to its right. If no pebbles can be moved, the configuration is called a final configuration. For each n, show that, no matter what choices are made at each stage, the final configuration obtained is unique. Describe that configuration in terms of n.

Hints: 408 510

Problem 10 (Shortlist 2013 N5). Fix an integer $k > 2$. Two players, called Ana and Banana, play the following game of numbers. Initially, some integer $n \geq k$ gets written on the blackboard. Then they take moves in turn, with Ana beginning. A player making a move erases the number m just written on the blackboard and replaces it by some number m' with $k \leq m' < m$ that is coprime to m. The first player who cannot move anymore loses.

An integer $n \geq k$ is called good if Banana has a winning strategy when the initial number is n, and bad otherwise. Consider two integers $n, n' \geq k$ with the property that each prime number $p \leq k$ divides n if and only if it divides n'. Prove that either both n and n' are good or both are bad.

Hints: 502 123

Problem 11 (Shortlist 2012 C4). Annie and Bertolt play a game with $N \geq 2012$ coins and 2012 boxes arranged around a circle. Initially Annie distributes the coins among the boxes so that there is at least 1 coin in each box. Then the two of them make moves with Bertolt going first:

- On every move of his, Bertolt passes 1 coin from every box to an adjacent box.
- On every move of hers, Annie chooses several coins that were not involved in B's previous move and are in different boxes. She passes every coin to an adjacent box.

Annie's goal is to ensure at least 1 coin in each box after every move of hers, regardless of how Bertolt plays and how many moves are made. Find the least N that enables her to succeed.

Hints: 635 540

Problem 12 (ToT Fall 2015 S-A7). Suppose N children, no two of the same height, stand in a line. The following two-step procedure is applied: first, the line is split into the fewest possible number of groups so that in each group all children are arranged from the left to the right in ascending order of their heights (a group may consist of a single child). Second, the order of children in each group is reversed, so now in each group the children stand in descending order of their heights.

Prove that in result of applying this procedure $N-1$ times the children in the line would stand from the left to the right in descending order of their heights.

Hints: 350 248

Problem 13 (USA TST 2020/3). Let $\alpha \geq 1$ be a real number. Hephaestus and Poseidon play a turn-based game on an infinite grid of unit squares. Before the game starts, Poseidon chooses a finite number of cells to be *flooded*. Hephaestus is building a *levee*, which is a subset of unit edges of the grid, called *walls*, forming a connected, non-self-intersecting path or loop. The game then begins with Hephaestus moving first. On each of Hephaestus's turns, he adds one or more walls to the levee, as long as the total length of the levee is at most αn after his nth turn. On each of Poseidon's turns, every cell which is adjacent to an already flooded cell and with no wall between them becomes flooded as well.

Hephaestus wins if the levee forms a closed loop such that all flooded cells are contained in the interior of the loop — hence stopping the flood and saving the world. For which α can Hephaestus guarantee victory in a finite number of turns no matter how Poseidon chooses the initial cells to flood?

Hints: 230 270 192

Problem 14 (Mexico National Olympiad 2020 P3). Let $n \geq 3$ be an integer. Two players, Ana and Beto, play the following game. Ana tags the vertices of a regular n- gon with the numbers from 1 to n, in any order she wants. Every vertex must be tagged with a different number. Then, we place a turkey in each of the n vertices. These turkeys are trained for the following. If Beto whistles, each turkey moves to the adjacent vertex with greater tag. If Beto claps, each turkey moves to the adjacent vertex with lower tag. Beto wins if, after some number of whistles and claps, he gets to move all the turkeys to the same vertex. Ana wins if she can tag the vertices so that Beto can't do this. For each $n \geq 3$, determine which player has a winning strategy.

Hints: 629 405

15 Expected Value

Expected Value, in a very broad sense tells us the average outcome of a certain experiment which can have multiple possible outcomes with varied probabilities.

Expected value is a really useful concept and is widely used in probability and statistics. However, to us the key use of expected value is in the context of the probabilistic method. The idea is that if you gave 5 exams with the score on each exam being an integer, and you ended up with an average score 71.5, there must have been some exam where you scored at least 72 and some exam where you scored at most 71.

We apply the same concept with expected values. If the expected value of a certain random variable is 0.8, there must be some case where the variable takes a value that is at least 1. This is in some sense similar to the pigeon hole principle and you can think of the probabilistic method as some sort of generalization to the pigeon hole principle.

Exercise. Can you spot anything wrong with having your average score equal to 71.5?

15.1 Some definitions

Okay, time for some definitions. I've tried to keep these mostly intuitive and simple, so hopefully I'll not bore you off.

- **Sample Set**: This is in a broad sense the set of possible outcomes of a certain experiment you're doing - say rolling a dice, in which case the sample set is one of 6 outcomes:

$$\{\text{getting a 1, getting a 2}, \ldots, \text{ getting a 6}\}$$

- **Random Variables**: Lets say we want to say that the probability of getting a head on tossing a a fair coin is 0.5. How do we do this? Well, one way for sure is to write

$$\mathbb{P}\left(\text{getting a head on tossing a fair coin}\right) = 0.5$$

But we don't really like to put in english text into math equations. How do we find a way to avoid the text within the equation? We define random variables. A random variable can be thought of as a function from the sample set to a set of real numbers (usually non negative integers suffice). For instance, if we think of a coin as having the sample set

$$\{\text{getting a head, getting a tail}\}$$

We can map this to $\{0,1\}$ using a random variable X so that $X = 0$ for a head and $X = 1$ for a tail. So now we can say

$$\mathbb{P}(X = 0) = 0.5$$

Since $X = 0$ if and only if we get a head. Another example is if we wanted to consider the probability of getting an even number on a dice. We could define

$$X : \{\text{getting a 1, getting a 2}, \ldots, \text{ getting a 6}\} \to \{0,1\}$$

so that $X = 0$ for the first third and fifth outcomes while $X = 1$ for the others. Then $P(X = 1)$ represents the probability of getting an even number.

In particular, if the range of X is exactly $\{0,1\}$ we call X an **indicator** random variable. These are quite useful to us when we learn about the linearity of expectation. However, a random variable need not be like this: we could also just have each of the 6 outcomes of a dice map to $\{1,2,3,4,5,6\}$.

- **Expected Value of a random variable**: A random variable also allows us to define its expected value: which is in a broad sense the weighted average of all the possible values of the random variable where the weights are the corresponding probabilities. Formally,

$$\mathbb{E}(X) = \sum_i \mathbb{P}(X = i) \cdot i$$

So lets say we consider a random variable Y that maps getting a k on a fair dice to k. It's expected value is

$$\mathbb{E}(Y) = \sum_i \mathbb{P}(Y = i) \cdot i = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \cdots + \frac{1}{6} \cdot 6 = 3.5$$

Which makes sense as the *average value* of a fair dice.

Exercise. What is the expected value of the above random variable if the dice isn't fair: instead, it gives 6 half the time, and the other five values with equal probabilities?

Exercise. Show that the expected value of an indicator random variable X is $\mathbb{E}(X) = \mathbb{P}(X = 1)$.

15.2 States

Till now, we know exactly one way to find the expected value of a random variable: the summation. But depending on the problem at hand, the numbers involved in this summation can get extremely messy and even if you get an expression for the final answer, it may be hard to evaluate the summation. The method of states gives us a way around this, allowing us to dramatically reduce the computational effort in certain cases.

Example 1

Malay flips a fair coin repeatedly until he gets a head. What is the expected value of the random variable X that represents the number of flips he had to make?

First, we'll do this the usual way. We know that the probability that $X = 1$ is the probability that Malay gets a head on the first flip. So

$$\mathbb{P}(X = 1) = \frac{1}{2}$$

What about the probability that $X = 2$? In this case, the first flip must have been a tail, and the second one must have been a head. Thus,

$$\mathbb{P}(X = 2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Let's take a small step back to understand something. Why were we able to say that the probability is $\frac{1}{2} \cdot \frac{1}{2}$? It may seem rather obvious, but it is worth understanding in a little more depth: the idea is that these events are *independent*: getting a tail on the first flip does not impact the second one and vice versa.

The idea is that in general

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A)$$

where the last term denotes the probability that event B occurs given that event A has already occurred. When events are independent, $\mathbb{P}(B|A) = \mathbb{P}(B)$, so it's okay to multiply the probabilities. However, we obviously cannot say that the probability of getting a tail on the first flip and a head on the same first flip is $\frac{1}{4}$ by the same logic because those events are not independent and in this case, $\mathbb{P}(B|A) = 0$.

In general, if we want to find the probability that $X = k$, the first $k - 1$ flips must have been tails, and the last one must have been a head. Each of these k events happens with probability $\frac{1}{2}$ and since they are independent, the final probability ends up as

$$\mathbb{P}(X = k) = \frac{1}{2^k}$$

Now we need to deal with the summation.

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k \cdot \mathbb{P}(X = k) = \sum_{k=1}^{\infty} \frac{k}{2^k}$$

The terms of the summation here form an AGP. AGPs are arithmetico-geometric progressions that have general term

$$T_n = (a + nd) \cdot r^n$$

for some constant a, d, r which often come up while trying to find the expected value of certain random variables.

In particular, if the probabilities follow a geometric progression, the summation we consider while computing the expected value has terms of an AGP.

To evaluate the summation of the above AGP (this works in the general case too), we use the following cool trick of multiplying the equation by $\frac{1}{2}$ (or in general, r) and subtracting:

$$\begin{aligned} S &= \frac{1}{2^1} + \frac{2}{2^2} + \frac{3}{2^3} + \dots \\ -\frac{S}{2} &= \quad\; -\frac{1}{2^2} - \frac{2}{2^3} - \dots \\ \hline \implies \frac{S}{2} &= \frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{2^3} + \dots = 1 \end{aligned}$$

And so we get $S = 2$. But this evaluation is not nearly as neat in several other problems, and here's where the Method of states comes in.

The basic idea is that we somehow want to relate the expected value (say x) of the event with itself to form a linear equation which we can solve and use to find the value of x. To do so, consider the first flip. Either you get a head already with probability $\frac{1}{2}$ and in this case $X = 1$. In the other case, it's almost as if the game just started: so it should take you x more flips on average to get a head. But one move has already taken place, so we need $x + 1$ flips in all in this case (think about this for a while until it starts to make sense). So we write the equation:

$$x = (\frac{1}{2}) \cdot 1 + \frac{1}{2} \cdot (x + 1)$$

Solving, we get $x = 2$, done.

This equation is basically relying on the fact that

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{P}(\text{event occurs}) \cdot \mathbb{E}(X \text{ if event occurs}) \\ &\quad + \mathbb{P}(\text{event doesn't occur}) \cdot \mathbb{E}(X \text{ if event doesn't occur})\end{aligned}$$

and our event is that we get a head on our first flip.

This was a slightly simplified version of the general method of states. In particular, we essentially had two states here:

- State 1: We've never gotten a head yet
- State 2: We just got a head!

We start at state 1, and each moment there's a probability of $\frac{1}{2}$ that we move to state 2 and a probability of $\frac{1}{2}$ that remain at state 1.

We know that the expected value of the number of moves to go from state 1 to state 2 is x, and now we build the equation the same way as earlier.

Let's try another problem to understand this method a little better.

Example 2

Malay flips a fair coin repeatedly until he gets three consecutive heads. What is the expected value of the random variable X that represents the number of flips he must make?

If you start making up a summation, you'll realize soon that this is not feasible. The last three moves are definitely heads, but before that the only condition is that we must not have three consecutive heads. You can create a recurrence to find the number of ways to do this, but working with that recurrence in the final summation does not sound particularly exciting.

So enter the method of states. We define the following states (we'll discuss the motivation behind this a little later).

- State 1: Your previous flip was not a head.
- State 2: Your previous flip was a head, but the one before that was not.
- State 3: Your previous 2 flips were both heads, but the one before that was not.
- State 4: Your previous 3 flips were heads.

We want to find the expected value of the number of moves required to reach state 4, starting at state 1. Let's say that this is E_1. But we don't stop there: let's say the expected value of moves required starting at state 2 is E_2 and starting at state 3 is E_3. Now we build our equations.

$$
\begin{aligned}
E_1 &= \frac{(E_2+1)}{2} + \frac{(E_1+1)}{2} \\
E_2 &= \frac{(E_3+1)}{2} + \frac{(E_1+1)}{2} \\
E_3 &= \frac{1}{2} + \frac{(E_1+1)}{2}
\end{aligned}
$$

Solving, we can find E_1, as required.

Now let's talk a little bit about how and when we're supposed to construct these states.

- In all such questions, you start somewhere and repeat a somewhat fixed procedure until you reach a certain end goal.
- Usually, the number of moves that have already occurred is irrelevant - the only thing that matters is the state in which you are.
- To find the states you need to introduce, think about the steps that must take place for you to go from start to destination. Sometimes you'll be able to go in a single step like in the first example, while in others you'll have to construct several in-between states. In particular, to get three heads in a row, I must have had two heads in a row on the previous move, etc.

15.3 Linearity of Expectation

We now learn about the *most* important idea in expectation which allows us to reduce hours of computation into a work of seconds!

Lets say we roll a dice, and are tasked with finding the expected value of the random variable which has value equal to the value on the dice. We solved this earlier, and found the value to be 3.5. What if we didn't have one dice - we had two of them and we want the expected value of the sum. There are now 36 different scenarios and sums anywhere between two and twelve. Still, it isn't impossible to calculate and we get

$$\mathbb{E}(X) = \frac{1}{36}\cdot 2 + \frac{2}{36}\cdot 3 + \frac{3}{36}\cdot 4 + \frac{4}{36}\cdot 5 + \frac{5}{36}\cdot 6 + \frac{6}{36}\cdot 7$$

$$+ \frac{5}{36}\cdot 8 + \frac{4}{36}\cdot 9 + \frac{3}{36}\cdot 10 + \frac{2}{36}\cdot 11 + \frac{1}{36}\cdot 12$$

$$= 7$$

The reason this answer is interesting to us is $3.5 \times 2 = 7$, and this does make sense. If the average value on throwing a dice is 3.5, the average value on throwing it twice should just be twice of it.

More generally, if X, Y are random variables, the following statement holds:

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

Lets try and understand what the above statement really means. First of all how do you even define the random variable $X+Y$? Do X and Y have to be of the same *type* or something? Do they have to be independent from each other?

Lets consider two experiments that you're doing: one is rolling a dice and the random variable X we're considering is the value you get on rolling the dice. The other experiment is tossing a coin, and the random variable Y takes value 0 if you get a head and 1 if you get a tail. Consider the sample sets of these two experiments be S_1 and S_2. When we consider $X+Y$, we're performing both experiments so the sample set is actually $S_1 \times S_2$, i.e. the cartesian product of the two. So here we can imagine the sample set is something like

$$\{(\text{getting a 1, getting a head}), (\text{getting a 2, getting a head}), \dots\}$$

In particular, $X+Y$ is now defined by the sum of the values you get from X and Y based on the first and second element in the ordered pair respectively so that the value of the random variable takes value $1+0=1$ for (getting a 1, getting a head).

In particular, we could have set $X=Y$ to the dice random variable to get the expected value on rolling two of them $E[X+X] = 2\mathbb{E}[X]$

Till now, everything about linearity of expectation has felt very intuitive and logical because the two events we've dealt with were independent from each other. So rolling a single dice did not impact the score on the other dice. What's truly magical about the theorem is that it holds even when the two events are dependent on each other. (!!)

But what kind of scenario are we talking about anyway? Let's consider the following example.

Example 3 (Classic)

There are n people, each wearing a hat, who come to a house for a party. On arriving, they leave their hat at the entrance. There is a power cut at the end of the party, and so each person ends up taking a random hat from the entrance. Find the expected value of the number of people who end up with their own hats.

Consider a random variable X which has value k if k people are wearing their own hats. We are tasked with finding the expected value of X. The key observation here is

$$X = X_1 + X_2 + \cdots + X_n$$

where the X_i are *indicator* random variables: taking value 1 if the ith person ends up with their own hat and 0 if the ith person ends up with someone else's hat.

Thus we have

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i]$$

Note that $\mathbb{E}[X_i] = P(X_i = 1) \cdot 1 + P(X_i = 0) \cdot 0$ The probability that a person gets their own hat is clearly $\frac{(n-1)!}{n!} = \frac{1}{n}$ and so we have that

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} \frac{1}{n} = 1$$

Lets now consider the case $n = 2$. There are only 2 people. Either both end up with their own hat, or they end up with each other's hat. The expected value ends up as $\frac{(2+0)}{2} = 1$ as expected. But notice that if we fix the first person's hat, the second hat is now fixed as well. In particular, if person one gets their own hat, the probability of the second person getting their own hat is no longer $1/2$.

While computing the expected value, we very simply just added up the $1/n$'s but it turns out that when you fix the result for one of the people, all the expected values for others change. It just so happens that all of your mistakes somehow just cancel out and you miraculously end up with the right answer.

Seriously, this should feel like cheating.

To convince ourselves that the theorem of expected value is true and that we didn't just get lucky, lets prove it!

$$\begin{aligned}\mathbb{E}(X+Y) &= \sum_x \sum_y (x+y)P(X=x, Y=y) \\ &= \sum_x \sum_y xP(X=x, Y=y) + \sum_x \sum_y yP(X=x, Y=y)\end{aligned}$$

Note that in the first summation, we can take x common from the sum over y, and the idea is that

$$\sum_y P(X=x, Y=y) = P(X=x)$$

(Convince yourself that this is true). Notice that we now end up with the first term as

$$\sum_x xP(X=x) = \mathbb{E}(X)$$

.

Exercise. Show that the second summation comes to $\mathbb{E}(Y)$ and finish.

> **Example 4**
> Given n coin flips of a biased coin that turns up head with probability p, what is the expected number of heads?

Linearity of expectation for the win! The expected value in one turn is $p \cdot 1 + (1-p) \cdot 0 = p$. So in n turns, the expected value is np.

Exercise. How would you formally write this argument in terms of random variables?

> **Example 5**
> A soda company inscribes one of the numbers $1, 2, \ldots, n$ on the bottle-caps of each one of its bottles uniformly at random. What is the expected number of sodas you must buy to collect at least one bottle cap labelled with each number

Lets say $n = 2$. On the first turn, you get either a 1 or 2, and after that you keep playing until you get the other one. So its equivalent to finding the expected value of the number of moves it takes to get this *other one*. This is equivalent to the following problem.

Keep flipping a fair coin until you get a heads. What's the expected value of the number of turns it'll take?

But this is the first problem we did in the previous section! To solve this,we had assumed that the expected value of turns is x. Then we either get it on move one or it takes us $x+1$ moves at an average. So,

$$x = \frac{1}{2}\cdot 1 + \frac{1}{2}(x+1) \implies x = 2$$

What if $n = 3$ instead? It takes us one move to get our first number. We then struggle a bit until we get a second number, and finally we struggle a fair bit until we get the third number. The idea is that linearity of expectation allows us to say

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n]$$

where X_i is the number of moves it takes to get an ith coin given that you've already collected $i-1$ distinct coins. So $\mathbb{E}[X_1] = 1$ because you can always guarentee a first number in a single move if you don't have any of them yet.

Let us try evaluating $E[X_k] = y$. Using the same logic as last time: there are $n-(k-1)$ numbers yet to be discovered, and so the probability that you get one of them in the first move is $\frac{n-k+1}{n}$. If you don't get one of these, it now takes you $y+1$ moves at an average.

$$y = \frac{n-k+1}{n} + \frac{k-1}{n}(y+1) \implies y = \frac{n}{n-k+1}$$

In particular the total value becomes

$$\sum_{k=1}^{n} \frac{n}{n-k+1} = \sum_{i=1}^{n} \frac{n}{i} = nH_n$$

where H_n is the nth harmonic number.

15.4 Problems

Problem 1. At a nursery, 2005 babies sit in a circle. Suddenly, each baby randomly pokes the baby to its left or to its right. What is the expected value of the number of unpoked babies.

Hints: 449 302

Problem 2 (HMMT 2012). You are repeatedly flipping a fair coin. What is the expected number of flips until the first time that your previous 2012 flips are $HTHT\ldots$?

Hints: 128

Problem 3. There are 100 shoelaces in a box. At each stage, you pick two random ends and tie them together. Either this results in a longer shoelace (if the two ends came from different pieces) or it results in a loop (if the ends came from the same piece). What is the expected value of the number of steps until everything is in loops, and the expected number of loops at the end of this process.

Hints: 172 536 **Soln:** Page 285, Solution 25

Problem 4 (HMMT 2015 C5). For positive integers x, let $g(x)$ be the number of blocks of consecutive 1's in the binary expansion of x. For example, $g(19) = 2$ because $19 = 10011_2$ has a block of one 1 at the beginning and a block of two 1's at the end, and $g(7) = 1$ because $7 = 111_2$ only has a single block of three 1's.

Evaluate $g(1) + g(2) + g(3) + \cdots + g(256)$.

Hints: 8 451

Problem 5. Bob has n stacks of rocks in a row, each with heights randomly and uniformly selected from the set $\{1, 2, 3, 4, 5\}$. In each move, he picks a group of consecutive stacks with positive heights and removes 1 rock from each stack. Find, in terms of n, the expected value of the minimum number of moves he must execute to remove all rocks.

Hints: 379 488 282

Problem 6 (NIMO #14). Ten students are arranged in a row. Every minute, a new student is inserted in the row (which can occur in the front and in the back as well, hence 11 possible places) with a uniform $\frac{1}{11}$ probability of each location. Then, either the front most or the back most student is removed from the row (each with a $\frac{1}{2}$ probability). Suppose you are the eighth in the line from the front. What is probability that you exit the row from the front rather than the back?

Hints: 136

Problem 7. You are given a rod of length 10 cm. An ant begins at the 3 cm mark. Each second, the ant moves to either 1 cm to its left or 1 cm to its right, each with $\frac{1}{2}$ probability. If the ant ever reaches the 0 cm mark or the 10 cm mark, the game ends.

1. What is the probability that the ant ends at the 0 cm mark?
2. What is the expected value of the number of moves required for it to reach any one of the end points?

Hints: 179 463 246 163

16 The Probabilistic Method

Now that we're quite familiar with evaluating expected values of different random variables, lets see how we can put it to use to solve (in fact, often *kill*) a variety of olympiad problems using the probabilistic method.

16.1 Introductory Examples

Example 1 (2002 IMC)

200 students participated in a math contest. they had 6 problems to solve. Each problem was correctly solved by 120 participants. Prove that there must be 2 participants such that every problem was solved by at least one of these two students.

First lets decipher what the problem is actually saying. So there's some contest, and apparently its easy because more than 60% of the people have solved each problem (or maybe the contestants are just very smart). We need to find 2 people so that the union of the problems they solved covers all 6 problems. So for instance if one has solved $1, 2, 3, 6$ and another has solved $2, 4, 5$ this would work. On the other hand something like $1, 2, 3, 4, 5$ and $2, 3, 4$ wouldn't as neither of them managed to solve P6.

Like many techniques in combinatorics there are some problems that call for expected value and probabilistic methods. One scream that you may hear (or rather, see) is *show that there exists some person(s) who have done something*. From your perspective, all people are practically the same - so its not like you can do a lot. Okay, you could try and take the persons with the most solves (a greedy algorithm!) and I have to admit, there is a greedy-ish solution to this problem (it's not as simple as taking the two people with the most solves though). That being said, this often doesn't work out nearly as nicely in many other problems, which means that the probabilistic method is definitely a method to try and keep in mind.

Anyway, so let's say we give in at last - "I have no idea whom I should pick". No problem! Let's just pick two random people and hope for the best. That is the entire concept on which the probabilistic method depends. We want the expected value of their *joint score*. If we can show that this is slightly more than 5, then at some point some two people must have had a joint score of 6 and we would be happy.

And now all that is left to do is find the expected value. Either this would end in a celebration, or we would find ourselves back on square one. Consider the random variable X equal to the joint score of the two people. This is the sum of six random variables $X_1, X_2, \ldots, X_6$. Here X_i represents the random variable of the joint score on the ith problem - which in fact is an indicator random variable and has value equal to 0 if both didn't solve, and 1 if at least one of them solved it.

$$P(X_i = 1) = 1 - \left(\frac{80}{200}\right)^2 = \frac{21}{25}$$

(why?) and so

$$\mathbb{E}(X_i) = \frac{21}{25} \implies \mathbb{E}(X) = \frac{126}{25} = 5.04$$

To reiterate, the idea now is that if no two people had a joint score of more than 5, the expected value cannot be more than 5 - and so some two students must have had a joint score of six as required!

Exercise. Figure out why

$$P(X_i = 1) = 1 - \left(\frac{80}{200}\right)^2$$

Remark. We actually used linearity of expectation here - without perhaps even realising that we did.

Seriously, this *should* feel like cheating!

Example 2

Consider n real numbers, not all zero, with sum zero. Prove that one can label the numbers as $a_1, \ldots, a_n$ in some order such that

$$a_1a_2 + a_2a_3 + \cdots + a_na_1 < 0$$

This problem, along with the next one, really show what I meant when I said that this method can *kill* problems.

So we're given n numbers (assume some random labeling for now - we'll figure out what labeling we need later on) so that

$$a_1 + a_2 + \cdots + a_n = 0$$

We want to deal with stuff of the form $a_i a_j$. How do we end up at terms of that form? Simple, square the equation.

In particular,

$$\sum_{i=1}^{n} a_i^2 + 2\sum_{i,j} a_i a_j = 0$$

We know that our sum of squares are positive (and not zero - why?) and so the sum of the $a_i a_j$ terms is negative. This in particular means that if we pick two distinct indices uniformly at random the expected value of their product is negative (why?). Now lets take a random permutation of the a_i and try and figure out the expected value of

$$a_1a_2 + a_2a_3 + \cdots + a_na_1$$

If we can show that the expected value is negative, then the expression can't always have been positive and so we would be done.

And here comes the big blow, so prepare yourself. The expected value of that expression is just $n \times$ (expected value of a_1a_2) which is negative, and so we're done.

> **Remark.** If you are convinced by what happened just now, I suggest you read the solution again (yes, you read the remark correctly).

Now that the problem is solved, lets try and figure out what actually happened. The expected value of the sum is the same as the sum of the expected value of the individual terms by Linearity of expectation (!!). Now, the expected value of say a_4a_5 is the same as the expected value of a_1a_2 because each pair of numbers coming up is equally likely. The key step, though, is linearity of expectation - the fact that we can break down the sum even though the term a_2a_3 is most definitely not independent of the term a_1a_2.

Example 3

In a class of n students, each boy knows at least one girl. Show that one can choose a group of more than half of the students such that every boy in the group knows an odd number of girls in the group.

We have a set of people, and we want to remove some of them so that every remaining boy knows an *odd* number of remaining girls. Notice that if we remove all but one boy and one girl - who know each other - the condition holds true. In fact, we could just delete all the boys and the condition would hold true (every boy remaining knows an odd number of girls remaining). The issue is that we should be removing at most half of the people. So the first strategy almost always fails and the second one works only if there were more girls than boys.

First guess for the probabilistic method: take every person with probability $\frac{1}{2}$ and hope for the best. There is a key reason as to why this *shouldn't* work. Notice that when we're taking a set into consideration, we haven't tried to control whether the odd condition holds true - and now the only way to fix the issue is to delete some extra people.

But the set we took into consideration originally had an expected size of $n/2$ so now if we try to remove someone, it won't work out too well for us (we want the final expected value of the size to end up as $n/2$).

The key observation is the following: yes, we would have to delete a boy if it ends up knowing an even number of girls in the selection we chose. However, we can also add a boy if it happened to know an odd number of girls in the selection we chose originally, but wasn't included in the list of boys we chose.

So maybe lets just not decide the boys we want to take at the start. Take a moment to sink this in. The idea is that we just select the girls randomly, and then figure out how many boys we can take.

Each girl is taken with probability $\frac{1}{2}$ so the expected value of the number of girls is $G/2$ (where G is the number of girls in the class). So all that is left is the expected value of the number of boys we can take.

Each boy has a 50% probability of knowing an odd number of girls in the selected set (why?). So the expected value of the number of boys is $B/2$ and by the Linearity of Expectation, we conclude that the expected value of the size of the class is $n/2$, so there exists some choice by which the size of the class is at least $n/2$ as required.

Exercise. We needed the size of the subset to be strictly more than half. How can we deal with this?

> **Remark.** This is another one of those problems where linearity of expectation leaves me speechless. Yes, there should be a case where we take at least half of the boys, and there should be a case where we take at least half of the girls - but the existence of a case where we take at least half of the class is an idea that takes time to sink in and is pretty amazing.

The probabilistic method also has wide scale applications in graph theory. Every time we need to show that something *exists* in a graph, it's worth considering a random subgraph of vertices (or something of that sort) and hope that the expected values work out in our favor.

Example 4

Show that for every n, there exists a tournament on n vertices having at least $\frac{n!}{2^{n-1}}$ Hamiltonian paths.

For someone unfamiliar with the terminology: essentially we have a complete graph on n vertices, but each edge now has a direction. We want to find a directed path (a path that *follows* the directions correctly) that covers all the vertices exactly once (a Hamiltonian path). In fact, the question asks us to find a way to assign directions so that there are a large number of such Hamiltonian paths.

The first advice I'd give you is not to get scared of that ridiculous looking bound. Where did $\frac{n!}{2^{n-1}}$ even come from?! For now, leave the bound and try to figure out what's the best you can do.

To be honest, I have no idea how to direct the edges, and if you try the problem long enough you'll realise that you can definitely get one or two or maybe a dozen paths to work out but there's no way you're hitting that bound - so lets just do it off randomly. Assign each edge a direction with probability of each direction being $\frac{1}{2}$.

Now consider any path going through every vertex once. This corresponds to a permutation so there are $n!$ such paths. The idea is that for this permutation to correspond to an actual Hamiltonian path, each edge must work in our favour. There are $n-1$ such edges, and so the probability of them all working out is $\frac{1}{2^{n-1}}$. At this point, you're probably seeing where the weird bound comes from. The expected value of the number of the Hamiltonian paths by linearity of expectation is

$$\frac{1}{2^{n-1}} + \frac{1}{2^{n-1}} + \cdots + \frac{1}{2^{n-1}} = \frac{n!}{2^{n-1}}$$

and so there must exist some way to direct the edges such that the graph actually has that many Hamiltonian paths.

Exercise. Rewrite the linearity part formally in terms of indicator random variables.

16.2 Probability of failing

For the next few problems, we use a different approach. Think of it this way. Let's say we need to prove that we can assign some values a certain way so that the final expression becomes what we would like. Then instead, we find the probability that our expression does not end up the way we want. If the probability is less than one, this means that there is a non zero probability that the expression does end up the way we would like. In particular, there is some way to assign values which allows us to end up the way we would like.

Example 5

Show that

$$R(k,k) > 2^{k/2}$$

for all $k \geqslant 4$

Here $R(k,k)$ is in context to the Ramsey numbers. To give some brief context, we're going to be taking the edges of a complete graph and splitting them into two colors. The idea is that you want to color the edges using 2 colors so that there is no set of k vertices with every pairwise edge being of the same color. $R(k,k)$ is the minimum number of vertices so that no matter what you do, there would have to be a set of k such vertices. We essentially want to show that we can create a pretty big graph with no such group of k vertices.

Exercise. Show that $R(3,3) = 6$.

So lets take a graph on n vertices and give each edge a random color (say, red or blue - each with probability half).

The question: what are we really going to find the expected value of? We can't find the expected value of n, that's where we started. We need to incorporate the no set of k vertices condition, so lets try and work towards that. What if we try and find the probability that this random assignment fails. If we can show that it is strictly less than 1, there is some assignment that would definitely work.

First, note that

$$\mathbb{P}(\text{every edge among first } k \text{ vertices is red}) = 2^{-\binom{k}{2}}$$

So the probability that they are of the same color is $2 \cdot 2^{-\binom{k}{2}}$.

We also know that

$$\mathbb{P}(\text{some group of } k \text{ failing}) \leq \binom{n}{k}\mathbb{P}(\text{first } k \text{ fail}) = \binom{n}{k} \cdot 2^{-\binom{k}{2}+1}$$

So it is sufficient to prove that

$$\binom{n}{k} \cdot 2^{-\binom{k}{2}+1} < 1$$

when $n = 2^{k/2}$.

Exercise. Show this. The bounds you'll need are quite weak. In particular, you might need to use

$$\binom{n}{r} \leq \frac{n^r}{r!} \text{ and } r! \geq 2^{r-1}$$

> **Remark.** There is another equivalent way of thinking about this. Consider the expected value of the random variable X that represent the number of groups that fail. That will in fact be exactly $\binom{n}{k}\mathbb{P}(\text{first } k \text{ fail})$. If you can show that this is less than one (which is exactly what we did), there must be some scenario where the random variable takes the value 0.

Example 6

Each vertex of an n-vertex bipartite graph G is assigned a list containing more than $\log_2(n)$ distinct colors. Prove that G has a proper coloring such that each vertex is colored with a color from it's own list.

First thoughts: let's assign each vertex a random color uniformly from its list and look at the probability of a collision. This can happen if two adjacent vertices somehow had the same color: which in turn happens with probability $\frac{1}{\log_2 n}$. Now a bipartite graph can have upto $n^2/4$ edges, so the bound we end up with for the probability of no collision is $\frac{n^2}{4\log_2 n}$ which is far too weak.

So we need to do something differently. Here are some things to notice

- We have to find a way to use the fact that the graph is bipartite. We could do that by bounding the edges by $n^2/4$, but surely there's something more to the story than that.
- The problem mentions a logarithm to the base 2. Ideally, we would like to do $2^{\log_2 n}$ or something of that sort to end up with a neat expression (wishful thinking!)

Let's begin with the latter point. We somehow need to end up with 2 to the power of the number of colors in a list. So let's try assigning one of two labels to each *color*. Now come to the first point. Bipartite graphs. So let's say the graph is $G = A \cup B$ where A and B are the two subgraphs with no edges among themselves. Think of it this way:

> If you assign label 1 to a color, it can only be used in A. And if you assign the label 2, it can only be used in B.

So this automatically ensures that the coloring is valid. All we need is a way to ensure that each vertex can in fact be assigned a color. To put that more concretely, let's say set S consists of the union of the colors in all the sets. Now

S is split into disjoint sets S_1 and S_2 where each color has equal probability of belonging in either set.

Now we need to figure out the probability that this splitting fails. In particular, if we let the list of colors assigned to vertex v be $c(v)$, there must be some vertex in A for which $c(v) \cap S_1 = 0$ (or such a vertex in B). This happens with probability

$$\mathbb{P}(|S_1 \cap c(v)| = 0) < \left(\frac{1}{2}\right)^{\log_2 n} = \frac{1}{n}$$

So the probability that some vertex fails is

$$\mathbb{P}(\text{a vertex fails}) < \frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n} = 1$$

Thus, the probability is less than one and in particular there must be some assignment that does not fail.

16.3 Alterations

Example 7

Show that if the minimum degree in an n vertex graph G is $\delta > 1$, then G has a dominating set containing at most

$$\frac{n(1 + \ln(\delta + 1))}{\delta + 1}$$

vertices.

This problem introduces two very cool ideas:

- What if we work with probabilities other than $\frac{1}{2}$?
- What if we *allow mistakes*, and fix them later?

If you look at the previous problem, we were taking a color to be a part of each set with probability half. The directions we assigned for the Hamiltonian paths were also done with probability $\frac{1}{2}$. And so were the colors of edges in the Ramsey numbers problem. What if you chose to do it with a probability p instead. Well, we would end up with a bound in terms of p then. The idea is that we can then maximise this probability by finding the p that maximises the given expression.

But before that, let's first *understand* the problem. You may already be familiar with the term "dominating set" but in case the term is new to you it essentially goes as follows:

Consider a subset S of the vertices. The set S is called dominating if for each vertex v in the graph, either $v \in S$ or v is adjacent to a vertex u such that $u \in S$.

So for instance if you consider a set containing all vertices, this set is dominating, but that is hardly interesting. We want a dominating set of size as small as possible - and the problem intends to give us a look into such a bound.

So first let's say we select each vertex with probability p. Then the expected value of the size of the subset is np. But how do we include the dominating condition? Here comes in the second idea: consider the vertices such that neither of it and it's neighbors are in S, and add all such vertices into S. So we now want to find the expected value of the size of this list.

If we let X_{v_i} denote the indicator random variable which is 1 if we need to add the vertex manually, we have that

$$\mathbb{E}(X_{v_i}) = (1-p)^{d_i+1} \leq (1-p)^{\delta+1}$$

So we can evaluate the expected value of the size of the dominating set $|S|$ as

$$\mathbb{E}[|S|] \leq np + n(1-p)^{\delta+1}$$

Here's a cool algebraic trick when you want to end up with ln or e^x:

$$(1+x) \leq e^x$$

For positives this follows directly from the Taylor series expansion of e^x, but here we need it for negative x, and the simplest proof I know of is to define $f(x) = e^x - (x+1)$ and show that its minima is 0 by taking derivatives.

Anyway, we have that

$$\mathbb{E}[|S|] \leq np + n(1-p)^{\delta+1} \leq np + ne^{-p(\delta+1)}$$

Now take the derivative with respect to p, and find the optimal p, and luckily, it's exactly what we needed!

Example 8 (LMAO 2023/6)

Let C be a given positive integer. A graph G over n vertices is called attractive, if no matter how we select $\lfloor 0.99n \rfloor$ vertices from G, the subgraph induced by them contains a cycle. A graph G over n vertices is called charming if it has C distinct cycles of the same length. Prove that all large enough attractive graphs are also charming.

First glance: Woah, we have two seemingly unconnected conditions but they seem to have been related to each other quite strongly. We need to show that all large enough attractive graphs are charming. So let's say we have some graph that is not charming. Then it has at most $C-1$ cycles of any fixed length. We'll show that this graph is not attractive either.

The idea is to take some subset of vertices in this graph and try to create a graph with no cycles - if we can make this graph have a large percentage of vertices we would be quite happy as that would mean that the graph could not have been attractive at all (an attractive graph is one where any subset with a large percentage of vertices must have a cycle).

We want to end up with a percentage more than 99%, so it doesn't make sense to pick up this subset with probability half. Let's instead take each vertex in this subset with probability p - rather like the previous problem. So we have some G' with an expected value of np vertices. We want to get rid of all the cycles. Notice that the probability that all the vertices of a 5 member cycle are still around is p^5, and in this case, we'll have to delete one of those vertices. In particular, there are at most $C-1$ cycles of any particular length, so after making all required deletions in G',

$$\mathbb{E}[|G'|] \geq np - \sum_{3 \leq k \leq n} (C-1)p^k \geq Np - \frac{(C-1)p^3}{1-p}$$

Here we've used the dumb bounds that a cycle has at least three vertices and

$$p^3 + p^4 + \cdots + p^n \geq p^3 + p^4 + \cdots = \frac{p^3}{1-p}$$

From here, our life is quite easy. This G' has no cycles, so we just want it to have at least $0.99n$ vertices. In particular, we need

$$p - \frac{(C-1)p^3}{n(1-p)} \geq 0.99$$

But this is easy to ensure: make p something like 0.995 and n large enough so that the second term becomes less than 0.005.

16.4 Final Examples

Example 9

Let G be a graph with n vertices and E edges. Show that the graph has an independent set with at least $\frac{n^2}{2E+n}$ vertices.

The first guess here is to try and do something along the line of the previous two problems - pick a vertex with probability p and then try to fix any mistakes you made. For each edge that remains in the graph, you'll have to delete one of the vertices. So the expected value of the size of the subgraph G' is given by

$$E[|G'|] \geq np - Ep^2$$

This is optimised when $p = n/2E$ and we get the bound $n^2/4E$.

However, note that this is roughly half as strong as the bound we need. So we will have to try something different this time around.

Pick the vertices one by one in a random order - and add it to your list unless a neighbour has already been added. So the idea is that it is okay to pick a vertex if it appears first in the list of its neighbours and itself. This happens with probability $\frac{1}{d_i+1}$ for vertex v_i, so we have that the expected value of number of elements is

$$\mathbb{E}[|G'|] = \sum \frac{1}{d_i + 1}$$

This time, the inequality works out in our favour and we can replace all the d_i with their average by Jensen's and say,

$$\mathbb{E}[|G'|] = \sum \frac{1}{d_i + 1} \geq \frac{n}{d+1}$$

where d is the average degree. Using $d = \frac{2E}{n}$ gives us the desired result.

Remark. For those unfamiliar with Jensen's inequality, it essentially says that for a convex function $f(x)$,

$$f\left(\frac{x_1 + \cdots + x_n}{n}\right) \leq \frac{f(x_1) + f(x_2) + \cdots + f(x_n)}{n}$$

Example 10 (2012 ISL C7)
There are given 2^{500} points on a circle labeled $1, 2, \ldots, 2^{500}$ in some order. Prove that one can choose 100 pairwise disjoint chords joining some of theses points so that the 100 sums of the pairs of numbers at the endpoints of the chosen chord are equal.

So we have a circle with lots of numbers, and we want to find a lot of non intersecting chords such that no two of them intersect each other, and the sum of the values on the end points on each of the chords should be the same. Let $2^{500} = n$. Then the sum ranges from 3 to $2n - 1$. Let's temporarily focus our attention on exactly one of these sums - say 250. We now want lots of chords that don't intersect each other and each having sum 250.

The idea is that you can think of this as a graph! Imagine the chords as vertices and join two vertices if the corresponding chords intersect each other. Now we want a large independent set. From the previous problem we can find an independent set of size

$$\sum_v \frac{1}{\deg(v) + 1}$$

Now going back, let's call this graph G_{250}. We also have a lot of other such graphs - and all we need is the value of the summation above to be at least 100 in at least one of the graphs. In particular, we have

$$\text{Average size of independent set generated} = \frac{1}{2n-3} \sum_{i=3}^{2n-1} \sum_{v \in G_i} \frac{1}{\deg(v) + 1}$$

We just want this value to work out to be more than 100. Note that all chords are counted exactly once above, and the degree of that chord represents the number of other chords it intersects which have the same sum.

If we consider two points that are adjacent to each other, what'll the degree of the chord between those 2 vertices be? Well, there's no edge that intersects this chord inside the circle, so the degree would just be 0. If instead, there is 1 point between the two endpoints of the chord - there can be at most 1 such intersecting chord. In general, if there are k points on the minor arc between two points, there can be at most k such points (each of these k can contribute at most 1).

Notice that there are n chords with degree 0, n chords with degree at most 1, and so on - until finally n chords with degree at most $\frac{n}{2} - 1$. So the summation we're looking at is actually

$$\frac{1}{2n-3}\sum_{i=3}^{2n-1}\sum_{v\in G_i}\frac{1}{\deg(v)+1}=\frac{1}{2n-3}\sum_{c\in\text{chords}}\frac{1}{\deg(c)+1}\geq\frac{n}{2n-3}\sum_{k=1}^{\frac{n}{2}}\frac{1}{k}$$

Now we use the following inequality (you can try proving this using integrals if you are aware of them).

$$\sum_{k=1}^{\frac{n}{2}}\frac{1}{k}>\ln\left(\frac{n}{2}+1\right)$$

All that is left is a little bit of algebra, and as it turns out the bounds work out quite easily and we end up proving that the average value is more than 172, far more than the desired bound of 100.

Exercise. Check the algebra. You may need that $\ln 2\approx 0.693$.

16.5 Problems

Problem 1. Let A be a set of N residues modulo N^2. Prove that there exists a set B of N residues modulo N^2 such that the set $S = A + B$ contains at least half of the residues modulo N^2. Here,

$$A + B = \{a + b \mid a \in A, b \in B\}$$

Hints: 86 27 645 348 296

Problem 2. Consider n red blue points and n red points. Suppose we connect at least $n^2 - n + 1$ pairs of opposite colors. Prove that we can select n segments, no two of which share an end point.

Hints: 547 37 604

Problem 3 (Russia 1996). In the Duma there are 1600 delegates, who have formed 16000 committees of 80 persons each. Prove that one can find two committees having no fewer than four common members.

Hints: 558 606 378

Problem 4. Prove that one can construct a finite tournament G with at least 101 vertices such that for any subset S of 100 vertices of G, there is some vertex v for which all 100 vertices $s \in S$ have a directed edge $v \to s$.

Hints: 191 299 173

Problem 5 (2006 IMOSL C3). Let S be a finite set of points in the plane such that no three of them are on a line. For each convex polygon P whose vertices are in S, let $a(P)$ be the number of vertices of P, and let $b(P)$ be the number of points of S which are outside P. A line segment, a point, and the empty set are considered as convex polygons of 2, 1, and 0 vertices respectively. Prove that for every real number x

$$\sum_P x^{a(P)}(1-x)^{b(P)} = 1,$$

where the sum is taken over all convex polygons with vertices in S.

Hints: 58 500 48 68 344

Problem 6 (Turan's Theorem). Let G be a graph with n vertices and E edges. Show that if the graph contains no K_{r+1}, we have that

$$E \leq \left(1 - \frac{1}{r}\right)\frac{n^2}{2}$$

Hints: 234 176

Problem 7 (USAMO 2010/6). A blackboard contains 68 pairs of nonzero integers. Suppose that for each positive integer k at most one of the pairs (k, k) and $(-k, -k)$ is written on the blackboard. A student erases some of the 136 integers, subject to the condition that no two erased integers may add to 0. The student then scores one point for each of the 68 pairs in which at least one integer is erased. Determine, with proof, the largest number N of points that the student can guarantee to score regardless of which 68 pairs have been written on the board.

Hints: 198 340 41 589

Problem 8 (IMO 2014/6, easier version). A set of lines in the plane is in general position if no two are parallel and no three pass through the same point. A set of lines in general position cuts the plane into regions, some of which have finite area; we call these its finite regions. Prove that for all sufficiently large n, in any set of n lines in general position it is possible to colour at least $c\sqrt{n}$ (for some constant c) lines blue in such a way that none of its finite regions has a completely blue boundary.

Hints: 308 120

> **Remark.** The actual problem asks you to prove the statement for $c = 1$.
>
> However, allowing any c let's you construct a direct probabilistic proof.

Problem 9 (Paul Erdos). A set S of distinct integers is called sum-free if there does not exist a triple (x, y, z) of not necessarily distinct integers in S such that $x + y = z$. Show that for any set X of distinct integers, X has a sum free subset Y such that

$$|Y| > \frac{|X|}{3}$$

Hints: 9 107 69

Soln: Page 285, Solution 26

Problem 10 (Erdos 1959). For all k, l there exists a graph G with each cycle length at least k and chromatic number at least l.

Hints: 349 97 636 **Soln**: Page 286, Solution 27

> **Remark.** You'll need Markov's inequality for the above problem. If X is a nonnegative random variable
>
> $$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$
>
> (Check that this makes intuitive sense based on the usual summation definition)

V

Appendix

A Hints

1. Think of (set 1, set 2, element)
2. Imagine the coins are not actually moving from the right to left, but the coins are being operated on cyclically. (In particular, the first operation will occur on the rightmost coin, but instead of moving it to the left, it stays there and now the next operation is on the second coin from the right, etc)
3. $0 \leq a_{t+1} - a_t \leq 1$
4. It may be a good idea to read the next chapter before doing this problem!
5. The answer is not 30 seconds
6. Guess the answer
7. If Bob's number was 1, Alice would know for sure that her number is 2.
8. Show that the evaluating the summation is equivalent to evaluating the expected value of the number of blocks in $0-1$ sequences of length 8 (you'll have to consider the edge cases separately)
9. This problem has one of the most magical solutions I have ever seen. To start off: the following set is sum-free: $\{k+1, k+2, \ldots, 2k+1\}$.
10. If two tuples are not disjoint, we must have swapped a_{n+1} with some a_k to get the b_i. Use this to create a graph of the non-disjoint tuples, and show that it is bipartite
11. Let's say that deleting column 2 makes row 1 and row 5 the same. Create a graph with n vertices representing the rows and join row 1 and row 5 by an edge.
12. If you have a connected component at the start, can it become disconnected?
13. Let's say black has a winning strategy. Find a way to use it against them.
14. Pick an arbitrary and take a connected component G' in it. Since this can't contain all vertices, there's some v not in it. Show that $v \cup G'$ is a connected component when considering the graph of some other color.
15. If t is in C, conclude that $t-1$ is in A. We also know that the number of elements in A and C are the same.
16. Once you can't take more numbers from the upper row, leave the rest of the work to the bottom row and work out the inequalities (this is tricky).

17. Once again, we double count. Use the second condition to get a good quantity.

18. Binary

19. Let $m = nk - n - 1$

20. Use AB ACB ACC ACD (the spacing to make it easier to understand). This works for each character except the last one, do that in the remaining two moves.

21. Let's say we consider a person A who is lying. Then one of his neighbours is saying the truth and the other is lying. Let's that the left neighbour is saying the truth (the other case is symmetrical). Now consider the person to his right and repeat.

22. The Cauchy Schwarz gives you a bound on the number of numbers on the board at any time. Use this to conclude.

23. If $a_1 = t$, what must be a_{k+1} be? Show that the number of solutions in that case is $\binom{n}{t}$

24. A group of 4 points contributes exactly two to the summation if the four points form a convex quadrilateral, and contributes 1 if i is concave

25. What would have happened just one second before the minute was completed?

26. The product increases by a factor of at least four each time. So we have a bound on the product at the end - now use AM-GM to create a bound on the sum of all numbers using this.

27. Find the probability that a certain residue is not in the set $A + B$.

28. Extend the lines to meet a huge circle.

29. No remainder can appear twice. Can 1 and 99 both appear as remainders?

30. Eventually everyone in the other rooms have come together into a single room. Finish now.

31. Use part 2 from the previous problem

32. We need the equations $1 = x^2 + x^3$ and $x^5 + x = 1$. Show that there is an x satisfying both (the first expression in fact divides the second!)

33. Let's guess that the total sum is $10k + 1$ for some k. Use this to make your guess

34. In your last move, you will just permute the rows. So you can assume that the elements of each row at the end are the same. You now need to go from this configuration to any configuration in 2 moves.

35. The only cases it doesn't occur is when the first chain is moved (show that we're already done in this case) or if the last chain is moved and the first and last chains are different

36. Show that $a_{k+1}, a_{k+2}, \ldots, a_n$ must each be $\geq k$. Use this to conclude that one of $a_{k+1}, a_{k+2}, \ldots, a_n$ is k.

37. The probability that a certain segment exists is $\frac{n^2-n+1}{n^2}$. Find the expected value of the number of pairs for which segments exist (using linearity of expectation)

38. If there's an odd cycle of length $2d+1$, let one of the vertices in it be v. Show that it is possible to go from v to any vertex in the cycle (including v) in n moves for all $n \geq 2d+1$

39. Weight edges with $1/\deg a$

40. Take some (a, b) and try to work out what is going on.

41. The answer is 43

42. Check AB on the first move, and check one of A and C based on the answer on the second move. Use these to find the first letter

43. Greedy.

44. Think about the pile(s) with minimal stones

45. Sum of squares

46. If two equal characters, delete. If not, create two equal characters and finally get a net change of -1.

47. The graph has a cycle - traverse it and use induction on the number of edges

48. We have a summation, which probably means we need to use expected value. Figure out what this expected value signifies

49. Try creating an even number of stones in both piles after your move.

50. Show that the bounds end up working even when 1 and 81 are far apart

51. (As the first player) delete the maximum power of 2 dividing N repeatedly. (First show that the move is legal each time, too.)

52. Use the W and $N(W)$ condition to show that the required graph has a perfect matching

53. This problem is quite connected to the previous problem (ISL 1994 C3)

54. To construct 2, pair 1 with 4, 2 with 3, 5 with 6, 7 with 8 in the first turn

55. Consider the two women as a single block, and multiply by 2 to count for their order (Mrs. X first and then Mrs. Y, or the other way round.)

56. The answer is $2 \times 2 + 2 = 6$

57. Think of (person, another person, book)

58. The trick is to think of x as probability. We'll show that the equation is true when $x \in (0, 1)$ and so the polynomial is equal to 1 at infinitely many points, allowing us to conclude

59. If you're stuck, look for more patterns

60. Let's say a is even. Show that you can swap the fruits at positions $(1, a+1)$ with a sequence of steps

61. This is similar to the Fibonacci example. Try re-reading that and show that the sequence is periodic.

62. Think of odd and even positions

63. Finish!

64. The idea is that if you only consider edges of a single color, this graph wouldn't be connected.

65. Let's say three points A, B, C among the given finite points create a triangle with area exactly 1 (we'll ignore the strictly less than one condition for now). What must D satisfy so that it can be in the set?

66. $\binom{36}{3}$ ways of picking the number card, and $\binom{16}{1}$ for the face card.

67. It says that A must have an even number of vertices

68. The expected value signifies the number of convex polygons such that all the points outside are red while the points outside it are not. How many such convex polygons can even exist?

69. Pick x randomly and look at its intersection with S. Set $p = 3k + 2$ to finish.

70. How are 24000 and 2019 related?

71. Once all coins from 1/1 to 1/2n are placed, you can greedily place the remaining coins wherever there is space.

72. You're adding 10 odd numbers. The sum cannot be odd!

73. $n \log n$

74. We know what graphs with all degrees 2 look like.

75. We have something slightly more special here - show that the disjoint cycles must be of even length.

76. You can have 3 distinct values! (For example 3,6,8). The issue is that 3 is right below a power of 2 - so it is close to 4. What you can't have, though, is these

values differing by too much from each other. In particular, you can't have values $k, k+2, k+4$ in a triangle (check this!)

77. Remember to divide by 6 since we could've picked the same objects in a different order.

78. If $a_t = a_{t+1}$, we have that $a_{t+k} = a_{t+k+1}$. Find something similar in the other case as well

79. The only possible value is $2n - 2$.

80. Find the pattern: $1, 3, 6, 10, 15$

81. If all $2n - 1$ answers are yes, show that a_1 has a white ball (this is not obvious - but it's true, read the conditions again)

82. What happens in the denominators you write out the numbers as $\frac{ab}{x}$

83. We start at $(1, 2, 0)$ but need to end up at $(0, 0, 0)$ since the pile with all stones will have 303 stones (the number of stones does not change) and $303 \equiv 0 \pmod 3$

84. Find an invariant.

85. Set up: try making some assumptions (such as which set has 1 in it). Now think extremally.

86. Pick the elements of B randomly, one by one.

87. When the numbers are $(a, a, a, a+1)$ modulo 4 we need to provide a construction - there's some edge cases that don't work though, find them. Simplify the situation by making one of the numbers close to 0 (don't make them it zero though because we could get stuck!).

88. Count (element, another element, subset containing both) triplets

89. The ants are really all the same

90. If you pick the group of r first, you get the LHS, but if you pick the leader first, you need to pick $r - 1$ others as in the RHS.

91. A wins iff adding $a - 1$ and $b - 1$ in binary involves no carry-forwards. Prove this with induction (show that you can make this property *alternate* like you do with the zero xor in the nim game.)

92. Is there a difference in what you see if the ants choose not to switch directions?

93. The answer for six balls is 120×6.

94. What happens if D lies on the opposite side of B and C of this line?

95. Draw a diagonal to separate the n-gon into a triangle and $n - 1$-gon.

96. $n \log n$ once again, divide and conquer again. You want to end up at $f(n) = 2f(n/2) + n$

97. To work with the cycles, find the expected value of cycles with length less than k with some approximations and set p so that the final expression becomes $c\sqrt{n}$. Then we can make the probability of having more than $n/2$ short cycles less than half.

98. To count the number of numbers, add k to the kth digit from the right - now the numbers will be in strictly increasing order!

99. Double count (set i, set j, common elements) to get an equation. Now we need to solve it in natural numbers

100. Split the numbers into n groups such that you can pick at most one number from each group

101. It jumps near powers of 3

102. Show that the number of inversions changes by ± 1 when you swap adjacents, and any general swap can be broken down into an odd number of adjacent-swaps.

103. If they both guess randomly there's a 25% chance both of them mess up. So we need to be smarter - we must base our guess on what we've seen as our result

104. Pasha can always win

105. The only problem that can occur is if the last chain is picked repeatedly. Show that this can only happen if $k \geq 3n/2$. Find a counter example when $k = 3n/2 + 1$

106. Think about the *complement* of this set (i.e., all the elements not contained in the set)

107. Look modulo a large prime p. Note that if $p \geq 3k + 2$, $\{k + 1, k + 2, \ldots, 2k + 1\}$ is sum free mod p. In particular, we can multiply all numbers in the set by x and it is still sum-free

108. For the second part - create a graph which is initially complete, and when you get to know that two numbers aren't consecutive, the corresponding edge is deleted.

109. Each person makes a different guess on the total sum.

110. Construct a parallel to BC through A

111. Adding a new point creates an extra $n - 3$ diagonals.

112. Inclusion and Exclusion, but it's messy because there are 3 numbers

113. The idea remains the same: try another case like 4 and 17 if you're confused.

114. Let m be fixed. Induct on n, by first showing that the base case of $n = 0$ is true, and then using strong induction.

115. As you would expect, we'll prove that for every subset $W \subset A$, $|N(W)|$ is larger than $|W|$. First resolve the case when $|W| \leq n/2$.

116. If there are at least k numbers equal to M, this is easy, so assume that at every point, less than k numbers are equal to the maximum. Induct on M

117. Consider the possibilities of the remainders modulo 4 for each prime. If we can get four different numbers for whom the remainders all match, we're done - why?

118. $a^2 + b^2 = (0.8a + 0.6b)^2 + (0.6a - 0.8b)^2$

119. The weighted sum remains constant - now find bounds for the start and end, and then work out why that bound should be achievable.

120. Alterations

121. If two of the three numbers are zero, what does the mod 2 equation tell us?

122. For the equality case, put a 2021 in the middle and all other elements as 0

123. Let S be one of the labels. Show that the SMALLEST number at least k of type S has only prime factors which are at most k

124. If you create a $S - -S$ the two $-$'s in between are basically dead - use this to make sure there are no draws.

125. Delete any edge

126. Color the vertices red and blue based on the graph G' and color them orange or green based on the graph T. So each vertex has two colors. Now finish!

127. Think of it as a tournament

128. Define $f(k)$ as the expected value of turns until your last k flips are $HTHT\ldots$. Relate $f(k-1)$ and $f(k)$.

129. $7 = \lfloor \log_2(100) \rfloor$

130. For the 4 case, we want to end up at four blocks of identical size. So let's try and create lots of blocks of equal size.

131. If we find a cycle of length 2019 in this graph, we'll be done. (Although that may not necessarily exist, and we may be done in other cases too)

132. Assume that no such triangle can be formed. Sort the segments in ascending order of length. The first two segments have length > 1. What can we say about the length of the third?

133. Assume that there are more positives than negatives. Show that you can take all the positives and at most half of the negatives. Now find an example where there's nothing better to do than this.

134. If a number doesn't divide the product of the rest, there must be some prime for which its exponent is bigger than everyone else

135. 24 angles. 6 ways to decide what comes at the top, and then four to decide what is in front.

136. Define the function $f(n)$ as the probability if you start at the nth spot. Relate this to other $f(i)$ by considering configurations after one move.

137. We now want to find a new letter roughly every move. We know that the letter at the start doesn't come again though. Let's say this letter is A. Find a way to differentiate B, C, D for the second letter

138. This is similar to the Taiwan 2001 problem.

139. if N is odd, A wins (and is in fact forced to win).

140. What happens if you get a "no" in one of the first two turns

141. You might have to do the base cases till 10 or more to really understand how the binary is working.

142. To show that the cycles are of even length, consider the number of up moves and number of down moves, etc.

143. Check $ABAC$. This doesn't work since we can't tell whether 2 as the answer means that its AB or AC. We want the answer to be different in each case - 1 if its D (we'll not involve D in the string), 2 if it's B and something else if it is C.

144. The answer is $25! \times 25!$

145. A has the winning strategy

146. You'll notice that only alternate negative indices are negative. Also the only way we're going "back" is by taking stones from n and placing them at $n-2$. So find a way to avoid negative indices, and then finish.

147. Induct. Consider a vertex x, and consider the graph G' without x. This has a "good" vertex c, so this forces the direction of the edge from x to c

148. Select an edge between two arbitrary vertices. Now choosing this crosses out less than $2(k-1)$ edges.

149. We want to keep the configuration simple - try placing all points on a single line.

150. Use the formula directly, or create a combinatorial interpretation

151. No vertex can have an infinite sub-chain and a finite sub-chain

152. Consider the distance of D from line AB.

153. Every *distance* appears at most once. A row with k colored squares contributes at least $k-1$ distances.

154. This is quite similar to how we solved the isosceles triangles problem

155. For $n=3$: this is mostly cases on the remainders of each of the five numbers

156. Every time you delete a vertex, the average degree increases!

157. Try solving it manually for small n.

158. The complement of the set contains all the numbers listed before

159. The issue is that we're counting the same grouping multiple times. In particular, the same pairs could have been picked up in different orders. Deal with this.

160. The first attempt is to show that this graph is bipartite, but that may not work out so well. Is there any condition in the chapter that leads to $n^2/4$?

161. Can you have a triangle with three distinct values? (Triangle inequality)

162. Each digit has two options

163. Define $h(i)=g(i+1)-g(i)$ - this is an AP. Use this to find $g(i)$.

164. The sum of squares is constant too

165. What's the answer if we *ignore* the condition that ants need to reverse directions?

166. Let's say you delete a vertex of degree d. Show that you're left with d disconnected trees, and use induction to finish

167. Double count pairs of the form $v-a, v-b$.

168. Previous chapter!

169. When n is even, try to maintain a 2-coloring on all connected components that have more than one color.

170. A graph consists of edges and vertices. We've proven it using an edge.

171. Use the previous hint to conclude that $a_n = k$. If $k \neq n-1$, use a_{n-1} to create a contradiction in the inequalities.

172. Let the expected value of number of loops for n shoelaces be $f(n)$. Relate $f(n)$ to $f(n-1)$

173. Evaluate this expected value, and show that as n tends to infinite, this expression will tend towards zero (and will therefore eventually be less than 1).

174. This is $\binom{7}{2}$ by picking the elements first. What if we pick the subset first? Show that each set has exactly three elements

175. If there is a $0, 1, 2$ or k, k, k we're happy for $n = 3$.

176. Randomly permute the vertices and add the vertex to the graph if all the vertices it is *not* connected to are after it

177. Label your tree with $v_1, v_2, \ldots, v_{n-1}$. Select a random vertex from G and assign v_1 to it. Now use this to define neighbours of v_1 (again broadly arbitrarily). Show that you can't get stuck.

178. Similar to the 17 scientists. Fix one of the people.

179. Let's say the probability that the ant falls of the left side be $f(x)$ if it begins at x cm. Relate $f(x)$ to $f(x-1)$ and $f(x+1)$

180. Show that if this resulting graph still has a Hamiltonian cycle, we can't conclude

181. The issue is that in our equality case the graph has roughly $n/2$ vertices and n edges if you did this correctly - so the n edge trick is out of question. Instead, take only some $n/2$ of the differences as actual differences

182. Break up any currently existing square into four squares, to induct from $n \to n+3$, and then find constructions for the base cases.

183. Find a strategy which works as long as not all hats are the same. Check that this gives us 75% probability.

184. Separate the 2019 counters into blocks of 3 and give Tony some additional information (and use the superposition trick!)

185. Each vertex in the first set has four edges to the second set. The only problem is that some edges might repeat. Modify the proof of the equal-degree bipartite graph problem so that everything works out

186. First group students from two classes and then group this pair of students with students of the third class

187. Try it for 100 replaced with 4? (1,2,3 are direct)

188. Consider a graph where each of the n subsets are vertices. Now if A_1, A_2 become the same when you pick $x = 1$, add an edge between the first two vertices, and do this for each x.

189. Go (safely) from $(1, 4x+1, 4y+1, 4z+2) \to (1, 4x+1, 4y-3, 4z+6)$ (don't make d go under $4k$ as it may only be 2, and don't make b go under $4x$ as it may just be 1.)

190. If a certain vertex has even degree and the graph is connected, the graph has at least n edges (think Handshake Lemma)

191. Let's say the tournament has n vertices. We're not really sure how to direct the edges. What should we do?

192. Suppose the flood is contained after n moves, use the fact that the perimeter of the levee is at most $2n$ to derive contradiction.

193. For the construction: systematically make each row filled with 0s until only the last two rows are left. Now get rid of columns one by one until only a 2 by 2 is non zero.

194. Show that you can go $(a, 0, 0) \to (0, 2^a, 0)$. Now think about how you could improve this even further if you had $(a, 0, 0, 0)$.

195. Let a_n be the number of ways of coming back to A after n moves, and b_n be the number of ways to reach B after n moves. (The second one is also equal to number of ways of reaching C or D - check this.)

196. If two non adjacent buckets both have more than 1 on Cinderella's move she is in trouble. In particular then, if at the start of the stepmother's move, two non-adjacent buckets sum up to more than 1, she can make each of them have more than 1.

197. Try this out - you should find that you're able to do it with 9 coins or 11 coins - but it doesn't "seem" possible to do it with 10.

198. Decide for each k whether you want to take $+k$ or $-k$ with probability p and $(1-p)$ (assuming that (k, k) occurs only for positive k)

199. While $\sum w_i x_i$ remains constant, there are things that change. (Think of the rightmost stone.)

200. If you took the combinatorial interpretation path (good job), think about the number of ways of choosing a group of r, with a leader amongst them

201. Let's say the accounts have money (a, b, c) with $a \leq b \leq c$. We'll try to keep decreasing the minimum of the three. In particular, we want to reduce $b = aq + r$ to $r < a$.

202. Consider the elements of the set as vertices and join two elements by an edge if there sum have rational sum

203. Take only alternate distinct values and consider the edges created by them. There should be no cycle in this graph by our generalised triangle inequality earlier!

204. Keep it simple - what can we infer from the very first statement made by Alice? Is there any case where she *would* know her number?

205. Each player can ensure that each connected component is a bipartite graph with equal number of vertices of either colour

206. Cool trick: If you're taking n tuples of distinct integers from $n+1$ options, you might as well append the remaining one at the end of this sequence as there is a bijection

207. Notice that instead of doing $6+6+6+6$ we could've simply said 6×4. Similarly, for $24+24+24+24+24$ in the previous subpart, we could do 24×5.

208. Think of each position of the game. If a certain position means that a result has occurred we know it's a P, N or D position (D for draw). Try to build up all the way to the starting position inductively.

209. Consider remainders modulo 100

210. For every group of four points, there's an intersect!

211. For $n=4$: $\{(1,2,4,8),(3,6),(5),(7)\}$

212. Bob knows that at least one of him and Charlie is wearing a red hat. But the only way he's sure of his hat color is if Charlie is wearing a blue hat (he'll know that his own color is red)!

213. Now we need to construct. Systematically reduce the size of S

214. $(a+b)^2 \geq 4ab$. Use this to find a bound on the product.

215. Look at the number of white markers with an odd/even number of black markers before

216. Divide and conquer

217. You want to create some weighting based on the location of the chip

218. If you color the cells according to $x+y \pmod 3$, there will be 9 squares that are $0 \pmod 3$, and 8 squares that will $1 \pmod 3$ and $2 \pmod 3$. So both the cells that are covered once must satisfy $x+y \equiv 0 \pmod 3$

219. The difference between the sums on the black and white squares doesn't change

220. How do you think 100 and 7 are related?

221. Induct

222. Solve it for the cases when the min of the piles is $1,2,3,\ldots$ until you get a vague idea of something that's happening

223. We need to create a large intersection of two subsets. Think about the expression we need to end up at: it should help you pick a reasonable quantity to double count.

224. $\sum_v \deg v(n-1-\deg v)$

225. Use the sum of degrees of v_1 and v_n (since these two are not joined by an edge)

226. Set up: Think of the process tree G' where each vertex represents a certain configuration and you can go from one configuration to another by firing a chip.

227. The sum remains constant, while the number of numbers reduces by one each time. Find another expression that remains constant

228. You need to go via Delhi OR you need to go directly

229. Find which of 1 to 2019 are in $\pi(1)$ to $\pi(1015)$. Then solve the first half and second half independently

230. $\alpha > 2$

231. Think modulo 3

232. The graph has a cycle: show that no vertex outside the cycle can be adjacent to a member of the cycle. So ignore this cycle and repeat.

233. Think of the people person i knows as the ith set. Double count!

234. This is very similar to the proof we did for independent sets.

235. Use handshake lemma to find out the number of edges.

236. Take parity of number of squares mod 2 for each color, and find a contradiction using our invariant

237. Use the second move to find the remainder when divided by 4.

238. No winning strategy exists for Tony

239. Let a_n be the number of such sequences that end with 0, and b_n be the number of such sequences that end with 1

240. The people are following a $L-L-T-L-L-T-L-L-T$ pattern (L refers to a liar and T refers to a truthful person). Since the cycle needs to be completed the number of people must be divisible by 3!

241. Think of odd and even numbers

242. Take cases n odd and n even

243. This is a long statement, but the solution is pretty direct - so don't get scared of the problem length

244. You need the same number of breaks, no matter how hard you try

245. Guess the answer by playing with some base cases.

246. For the second part, we do the same thing - define $g(i)$ to be the expected value of moves taken to reach. Remember that you'll have a +1 factor in the equation

247. After one day and one night, the caterpillar has had a net change of 1 inch upwards. Be careful about what happens towards the end, though.

248. Fix a threshold $1 < k < n$, and during the process write 0 for heights at most k and 1 for heights greater than k.

249. How do the number of vertices change when you perform the operation. What about the number of edges?

250. They can only be equal if $n = 2k$. In this case, first show that one guess cannot be enough (if a certain string works, find another that works.)

251. Show that two consecutive points on the circle should never be used together

252. Each pair of elements belongs to exactly one subset. What should you count?

253. In a single move, we either move apples and pears from odd positions to even positions or vice versa. Now frame an invariant!

254. Can two numbers have the same such prime? PHP to finish!

255. Now if we consider $\sum w_i x_i$ this stays perfectly constant - but that's not what we want to prove - we need to show that we reach a "bad position".

256. Who said we must use only two colors?

257. This is not hard, each group of the $\binom{n}{3}$ choose someone, so someone is picked by at least $\frac{1}{n}\binom{n}{3}$ groups.

258. Adding one element to the subset eliminates at most $2|A|$ other elements in B

259. For $N = 15$, take the triangle with vertices $1, 8, 9$ - this splits the circle into three parts having $6, 0, 6$ points each - and now copy!

260. Use the formula we had created. Alternatively, create a bijection.

261. The magic trick is at the start itself. Think beyond the obvious interpretation of this graph theoretically.

262. When does swapping c_i and c_j increase each of the two summations

263. If at most 1 is zero, we can applying the process

264. This is mostly a little bit of case work: for a start, note that if there are two pairs of consecutive vertices among the five we're done

265. If $t-1$ lies in C, think of $t-k$ and $t-k-1$.

266. You should get $(12k-1)t = (3k+6)(3k+5)$ (where t is $|S_i \cap S_j$. Solve this equation by noticing that t is an integer and so $12k-1$ divides the product on the RHS)

267. The room with the largest number of people never reduces in size. (you'll have to be slightly more careful with the ties, though.)

268. If we want to make t moves, number greater than t are just t for our use. We know the condition for any number t to work. What is it?

269. Use an inequality that relates the sum of numbers, sum of squares and the number of numbers

270. To get $\alpha > 2$ to work, leverage the fact that you will place three walls in one turn infinitely often. Try to block off the flood side by side.

271. The main difficulty of the problem is understanding what sort of sequence is the worst case scenario. Play with some cases with 2022 replaced with 10 or something (don't play with numbers that are too small though, the results may be misleading)

272. If a collision occurs late, it must happen close to the centre of the board. Use bounds to verify this. If the last collision happens at time t, we need to show that t + time taken to fall off from here is at most $3m/2 + 1$.

273. We definitely need the total sum to be at least Mk. This is in fact sufficient.

274. Where does $t - 1$ lie?

275. If all blocks have weight at most 2, then a greedy algorithm will work.

276. Figure out explicitly how the permutation changes when you add back the leaf. (It doesn't change much)

277. Instead, first create a recurrence that finds the number of sequences with no consecutive zeroes.

278. $n - k \geq 2\binom{k}{2}$

279. Induct by deleting a leaf

280. First show that each vertex must have even degree (use the fact that there's an edge before and after each vertex in the sequence.)

281. If you can make t moves, you can make $t - 1$ moves too, so if you find the answers for whether each t is achievable, it'll look like $YYYYNNN$. We don't want to do each t as that will take $\mathcal{O}(nm)$

282. Consider $\sum |x_i - x_{i+1}|$. Be careful about the edge cases though.

283. Answer is a fixed constant for $m, n > 1$

284. Set up: we're basically breaking down all nonnegative integers into groups of the form $(k, k + a, k + a + b)$ and $(k, k + b, k + a + b)$. Imagine coloring the integers 3 at a time.

285. Work with some possibilities for $a_0, a_1, a_2, \ldots$.

286. The number of inversions changes very "nicely" when you just swap two adjacent numbers.

287. This is just a generalization to the example we saw in the chapter (where we solved it for $r = 3, n = 10$).

288. First move a ball from B_n into B_{n-1} and then apply step 3 on B_{n-1}. (Notice that the order of moves is important).

289. $\sum ka_k$ works.

290. What are the starting and ending positions modulo 3?

291. Try to match downwards triangles with non-removed upwards triangles.

292. We need to divide by 4!

293. The answer is after 71 days and 70 nights.

294. To construct 4, pair 1 with 6, 2 with 3, 4 with 5 and 7 with 8 in the first turn

295. The procedure is identical to how we approached the sum of squares Example at the start.

296. Do a little bit of algebra to show that the bounds work out. Note that the bound can in fact be improved to $\frac{1}{e}$

297. It is $2n - 1$

298. Take the longest path, and work with the endpoints of this path

299. Pick the directions randomly. Now we want to show that there's a case where there is no *bad* set of 100 vertices - so we want to show that the expected value of the number of bad sets is less than 1.

300. mod 3

301. So you can figure out the leading digit. Now repeat to find out the second digit and so on.

302. Use linearity of expectation to conclude

303. Use the fact that they're roots of that equation

304. Similar to the 1×4 tiles given as an example in the chapter, color the grid with 3 colors.

305. The n odd case should be resolved just by noticing the graph just before the end of the game must be a complete bipartite graph.

306. Think of (student 1, student 2, answers in common)

307. Use the first move to find out if the number is divisible by 2

308. There's a theorem by Euler for Planar graphs: $V - E + F = 2$. Use it to bound the number of finite regions

309. Let's say I want to find the first digit. Find out how many times in the list of strings given, the "correct" digit appears, and how many times a false digit appears. If these numbers are different, your job is pretty easy.

310. Assuming optimal play, figure out what the graph must look like immediately before a player loses.

311. If you have only three boxes $(a, 0, 0)$, how many coins can you create?

312. First delete one edge from each triangle to get a bipartite subgraph. This has chromatic number 2. We will try and show that the remaining graph is usually bipartite too (except one edge case)

313. "For every two persons, there is at least one book that both of them bought."

314. Starting with a_n, repeatedly place each number into the pile with a lower sum currently

315. The $r - \dots$ in the final expression suggests that you think of some sort of complement of the intersection

316. Show that $k \geq n$

317. Consider one set as the set of piles and the other as the set of ranks

318. We now want to show that the bound is optimal. Construct a graph!

319. Remember the dominoes problem in the invariant chapter

320. Try and make the binary number *increase* after one full cyclic rotation (or make the game end at all T's)

321. Similar to the ants on a line problem, assume that some ants only move up and left and others move only move down and right.

322. The idea is that you want to create some sort of isosceles triangle that splits the circle into one small part and two identical ones.

323. Label the hats from 1 to 10. Add up their labels.

324. The two neighbours of the smallest number must be equal to it.

325. Weight edges and consider a summation over a set W

326. The only case at which we're stuck is where we know 3 people and don't know 5 people. Is it possible that each person knows exactly 3 people? (Parity!)

327. Try breaking into $1/N$ groups as far as possible. Then eventually you can guarentee that the only pieces will be of sizes $1/N$ and $2/N$. (Show this)

328. Assume that there is no such number. Think in terms of primes

329. If the possibilities for the next turn are $a+1$ and $b+1$, we must have

$$\{a_0, \ldots, a_n\} = \{0, 1, \ldots, a\} \cup \{0, 1, \ldots, b\}$$

330. If 1 and 81 are somewhat close-by, we can consider numbers on the *path* from 1 to 81. If let's say the path consists of 2 numbers in between $1, a, b, 81$, we can conclude that at least one of these differences is going to be quite big

331. Think euclidean algorithm. (Google this up if you haven't heard of it - it's how we find GCD of numbers.)

332. We care about the number of C's, O's and W's in specific ranges. How do we deal with this?

333. Prove that every element also appears in exactly three sets, and finish

334. The usual. Try the problem for small values of k, and make a guess on the answer, as well as some properties of the sequence.

335. Players are only allowed to subtract one, so the entire game is forced.

336. Find x_{-1}. But don't just stop there!

337. Think of (singer 1, singer 2, song)

338. If Alice is saying the truth, Charlie is lying, but Charlie claimed that Alice said the truth. So Alice must be lying. Now use this to determine whether Charlie and Bob were saying the truth.

339. If you currently have k points, each pair of points can rule out at most 2 other points, so you can build an equation

340. Show that every ordered pair is scored with probability at least $min(p, 1-p^2)$ and pick p to set these equal.

341. Work with some small cases of n and k till you get the hang of what's going on.

342. Show that there are some two prefix products such that all the parities are the same, and divide the two.

343. Repeat this! We should eventually end up with all vertices in a single component for some color - a contradiction.

344. Think convex hull.

345. If $f(k) = f(k+l)$, Bob would've removed $f(k+l) = f(k)$ stones after Alice removed $(k+l)n$ stones, so that position should be N.

346. While applying the formula, be careful of whether you're doing it with natural numbers or whole numbers.

347. The answer is 1 for powers of 2, and 2 for all other numbers

348. Once you have the probability, the expected value of the number of residues is $N^2(1-p)$ by linearity of expectation.

349. Let the graph have n vertices. Pick each edge with probability p. We will try to choose p so that the graph has a low probability of having a large number of short cycles and also low probability of having low chromatic number

350. What can you say about where 1 ends up?

351. Let a_n be such sequences not ending with 1, b_n be the number of sequences ending in 1 but not having a 0 before, and c_n be the number of sequences ending in 01

352. Delete any vertex with degree less than $\frac{m}{n}$ repeatedly. Show that we have a positive number of vertices left at the end

353. The idea here is very cool, but hard to come up with - try to come up with a quantity that you can keep reducing. It's not the number of vertices. (!!)

354. Use the binary representation of q

355. Start by removing an stones, and then remove another $(b-a)n$ stones to construct.

356. Think extremal

357. Consider the sequence

$$+1, -1, -1, +1, +1, -1, -1, +1 \ldots, -1, -1, +1, +1, -1.$$

358. Show that for $n \geq 8$ and n even, B wins - in particular, the only way A survives if they play all the odds/evens, but B can make sure they play at least one of each category.

359. Repeat until stuck

360. Look for patterns: 1, 2, 4, 5, 10, 11, 13, 14, 28, 29 ...

361. B makes their first move as one of 1 and n (at least one of them is empty). Now after k turns, A must have created at least k gaps, while we've made only $k-1$ moves. So B can't lose

362. Find an invariant

363. Using twitch lemma, show that if the graph has a Hamiltonian cycle, we're done. If it doesn't, try to use the Hamiltonian path in the tournament (this always exists!)

364. I'll give hints for part 2 directly (that solves the first as well) - Find an invariant

365. Bob and Diane have contradictory statements. So one of them is lying and the other is saying the truth! Use this to reduce the case work and find the answer

366. Try to connect this to the Fibonacci example in PHP.

367. At any point, set k as the smallest uncolored number. Show that we cannot get stuck.

368. The idea here is fairly unique as well - it is to consider the number of "inversions" in a permutation - the number of (i, j) such that $a_i > a_j, i < j$. (Essentially the number of pairs of numbers where bigger comes before smaller.)

369. Find the right graph (it should have $2n$ vertices and $2n$ edges)

370. Let's say we picked the smallest k elements in the row. Then the $k + 1$th element is at least $\frac{n+1}{4(k+1)}$ (why?)

371. This is just $13 \times 13 \times 13 \times 13$

372. The answer is yes. Try and greedily pick the smallest numbers that can be picked and look for patterns

373. Some possible patterns: any number seems to appear at most twice. We never encounter n without having encountered $n - 1$.

374. The first guess is to find the number of chords, and then pick any two from that, but this is wrong - not all chords meet inside the circle, some meet outside.

375. Find a weighted sum that is monovariant

376. Consider the number of black markers before each white marker and see how it changes after each move.

377. Find the number of such numbers with leading digit 1, 2, ..., until you cross over 2005

378. Do a little bit of algebra and conclude that the answer is more than 3

379. Focus on how you'll find the minimum given some sequence - it's quite easy once you know how to do this.

380. If we want the distance from v_1 to v_2, this is the sum of the distances to the root - twice the distance to the least common neighbour. Just look modulo 2

381. We just want to find one other white ball. Find a way to do this in at most $2n-2$ moves for $2n-1$ possible balls.

382. If there are k_1, k_2 lines separating a, b and b, c, show that the parity of the number of lines separating a, c is the same as the parity of $k_1 + k_2$. Now finish.

383. If the subset contains 1, remove it. If it doesn't contain 1, add it. Use this to create your bijection.

384. $r + (n-1)k$ is the sum of the n intersections a set has with other sets (including itself)

385. Keep solving for small N until you see light at the end of the tunnel.

386. "Perimeter"

387. The second one is also similar, but we have the factors of 12! and 2^12 now.

388. This graph has n edges - one for each column. Show that there shouldn't have been a cycle in this graph though.

389. The LHS counts the number of even element subsets. Create a bijection to show that a set has an equal number of odd-element and even-element subsets

390. Show

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

391. We can't have m consecutive zeroes - or everything would become zero!

392. Pick one arbitrarily first. Now 4 options are excluded for the second one (the one picked already, two neighbours, and opposite). Now if this is "close to" the first one or "close to" diametrically opposite it, we'll have to take some cases, but most of the time, there'll be just be 6 options excluded. Do the casework

393. Think of elements in exactly one of two subsets. In particular, your triplet is (subset, another subset, element in exactly one)

394. When can the two expressions be equal?

395. If $n-1$ is not divisible by 3, use induction to find an algorithm

396. The answer is of the order $3n/4$.

397. The idea is to just keep using the largest note you can at any point in time. So find $\lfloor \frac{N}{100} \rfloor$ to begin with.

398. It is $\binom{n}{2}$

399. Induct on k

400. You will always have to choices for a_i. Try and define these choices in terms of the current sequences more neatly.

401. Define weights on each square so that the expressions work out nicely.

402. Take small cases. Don't give up!

403. Show that there are now two disconnected trees in the graph, and use the induction hypothesis

404. If both of them guess the same as what they've seen, we have an issue. We also have an issue if both of them guess different from what they've seen. Can we mix these up though?

405. Beto wins if n is an odd prime

406. You get $m \leq 2n-1$. To achieve equality, you need each pair of sets to have *exactly* one common element. Use this to create a construction

407. Combine all the processes in the previous part to conclude when n is even. Now we need to find a way to handle n odd.

408. You can solve this problem directly or by using the chip firing lemma in the previous problem

409. Color according to $x-y \pmod 3$. Conclude that the cells covered once must also satisfy $x-y \equiv 0 \pmod 3$. How many cells satisfy conditions $x+y \equiv x-y \equiv 0 \pmod 3$.

410. There must be exactly one flight from A to B. Create a contradiction to the fact that A has even vertices.

411. You're only allowed to subtract odd numbers, so the parity keeps changing.

412. Take as many numbers as possible from the upper row (remember the example 1)

413. Handle the $r=0$ and $r=n$ cases separately.

414. We ideally want to keep all our heads as heads - to maximise our control over the coins, but then the devil keeps his tails as tails and we won't be able to make progress. So we need to force him into making a tails into heads

415. Why stop at triangles? What happens if you have a cycle of values differing from each other by a fair margin?

416. $\binom{4}{2}$ ways of picking the two Aces, and $\binom{4}{2}$ ways of picking the two Kings. We have to pick two Aces "and" two Kings.

417. Find the first letter of S in two moves.

418. For the first way to calculate the expression, fix the two singers first. And for the second, fix the song first. You should get a divisibility constraint that puts a bound on m.

419. The answer is that Malocolm wins iff $|a-b| \leq 1$. Induct to finish!

420. The moves we have at our disposal are $?WWW? \to ?BB?$, $?WWB? \to ?BW?$, etc. We need to look for some sort of pattern modulo 3.

421. (a,b,c) can lead to $(a,b-2,c)$, and $(a+1,b+1,c-1)$ (and its cyclic permutations). How do we make all of these the same?

422. With those two moves, we can be sure of the first card. Repeat this until you're sure of $n+1$ cards

423. Where does 896 come from?

424. Take the n points, and for each distinct value, pick one of the pairs of vertices such that the value is achieved for those two vertices and join them by an edge. This has a cycle - but the values may not differ by at least 2. Now what?

425. One guesses what they saw and the other guesses opposite of what they saw!

426. It matters that 2000 is "large enough"

427. We want to create some coloring. The most obvious one is the usual chess one, but that covers 2 cells of each type. Create a coloring that covers 3 cells of one type and 1 of the other

428. We use a cool idea known as exponential weighting (this actually makes life reasonably more decent on the MOP 1998 problem too). The idea is to give the ith square weight $w_i = x^i$ for some x. The decent part now is that we won't have negative things come up at the end x^i is always positive for integral i if x is positive.

429. The answer is $\frac{3m}{2} - 1$. This is not hard to achieve

430. Let's say we deleted the edge $u - v$ and there's some issue: so there are vertices a and b such that the path must contain that edge. Now what?

431. Try to build on the combinatorial proof of $\sum w_i^2 = \sum l_i^2$

432. Assume that there is no point with two chips on it. So the value of the summation is less than $\sum_x^\infty \sum_y^\infty 2^{x+y}$. Evaluate this, and compare it to the value of the summation at the start.

433. Binary search.

434. Replace 100 with a small number and try to make the trick work until you're able to characterize the types of things that can happen

435. We need to show that the second player always has a legal move - show that this exists

436. If $n = (k+1)t + 1$ for some t, Bob wins

437. The largest number doesn't reduce, and so this falls in a way similar to the "500 people in rooms" problem.

438. If you've assumed that not all hats are of the same color, but you see both the others wearing red hats, you can confidently guess blue.

439. The pattern is $1 + 2 + 3 + \cdots + n$. Check that this makes sense, since we're adding n to the answer for $n - 1$ objects.

440. Assume that the first digit is 1, and figure out the rest in one move.

441. This is pretty similar to the nine points in 3D space problem

442. Conclude that at the end of 2023 moves, we have an odd number of inversions.

443. The total number of "coverings" are 48, so this means that exactly two cells have been covered once, while all others have been covered twice.

444. Triangle inequality. Show that the third segment has length > 2, and the fourth has length > 3

445. To construct $2k$ pair 1 with $2k + 2$ in the first turn

446. To show that the game is winnable when $k \neq n$, start by looking at the first k cards and then looking at cards from 2 to $k + 1$.

447. Consider the number of times each of the n numbers comes in all of the prefix *products*.

448. Let's say $q = 2$. What will you do? (The first move does not involve b). What if $q = 3$?

449. What's the probability that a baby is unpoked?

450. The idea is that deleting only red colored vertices doesn't affect the degree of other red vertices, so you can eliminate a color.

451. Relate the number of $0 - 1$ sequence of length n to those of length $n - 1$.

452. There are two methods to solve this problem. One involves the extremal principle: check hint 2 for more on that. Another follows weighting and counting in 2 ways (hints 3 and 4 for more on that).

453. At the end of each night, some packs of three all have the plague

454. The star condition is basically equivalent to saying that all degrees are at most $k-1$. So try to show that in a graph with degrees at most $k-1$ and $2(k-1)^2$ edges there must be a k matching

455. There are 36 number cards and 16 face cards.

456. The idea is that $a_1, a_2, \ldots, a_{k+1}$ uniquely determine the entire sequence.

457. Color the grid in a way (not hard) that this $(-1,-1,+1)$ happens to cells of different colors

458. Change the definition of "separate" slightly. In particular, let's say two points are also separated by a line passing through one or both of them. Check that this reverses the condition, so if the number of lines is odd, the vertices must have the same color. Show that there are no odd cycles here.

459. If Bob and Charlie were both wearing Blue hats, Alice would know that his hat is red since at least one hat is red. So one of Bob and Charlie is wearing a red hat. Now repeat this kind of argument

460. This problem is in some sense a mix of ideas from the chapter on invariants and monovariants and double counting. While we won't directly think of some pair or triplet, we'll be evaluating a summation in two ways

461. If n is odd, there is a vertex with even degree (handshake lemma!) - use this to your advantage.

462. The total number of possibilities end up as $2 \times 2 \times 2 \cdots \times 2$.

463. Show that $f(i)$ forms an arithmetic progression, and use this to find $f(3)$

464. Show that the graph has no triangles.

465. If you want to go via Delhi, you need to go from Mumbai to Delhi AND from Delhi to Kolkata

466. Consider the total perimeter of the squares that are infected. This never increases.

467. To show that $2^{n-2}+1$ are sufficient, use the injective mapping version of Hall's theorem.

468. Let $k > 2$ be the weight of the heaviest block. Try to use induction on n based on what you do with this heaviest block.

469. If we ignore the first vertex and all the edges incident to it - we should be able to make everything else red. Likewise for all the other vertices, we should be able to make all but one of them red.

470. From the coloring, conclude that there should be an even number of pieces. How many pieces should be there, though?

471. Consider the first two towns. Show that we can get rid of one of them which is definitely not sweepable, and reduce to the case with $n-1$ towns.

472. Consider $a-1$ and $b-1$ (where the piles have $a-1, b-1$ stones)

473. Prefix sums

474. We need to show that some two have the same remainder. We have 100 possible remainders and 100 numbers. So the only case is where these products are $\{0, 1, 2, \ldots, 99\}$ in some order.

475. Choose the k largest numbers on the first turn. This has to be possible by the sum condition, and we have that after this move the maximum will be $M-1$.

476. Each hole splits a row into 1 or 2 parts, call this partition A. Similarly, each hole splits a column into 1 or 2 parts: call this partition B. Try to match the $2n-2$ parts in A with the ones in B using Hall's theorem.

477. Think parity

478. The answer is 3

479. We know that $\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$

480. The first guess is that you want to work with (set, another set, element in both). But we will need to modify this slightly: where do you think $r+(n-1)k$ is coming from?

481. If n odd, we want the sum to be divisible by $(2k+1)(k+1)$. The trick is to pair up some numbers

482. This problem is actually identical to the 2 pile nim game if you think about it.

483. Now consider a company with at least one flight from A to B originally (why must this exist?)

484. This is similar to Problem 4 - you'll need to take some cases and find out what happens

485. Try to create some contradiction in this cycle (i.e show that there is indeed a monochromatic directed path from v_j to v_i even though $v_i \to v_j$ in the cycle)

486. Let's say the final graph isn't connected. Then G has two components A and B such that there's no edge from A to B.

487. If you took the bijection path, notice that each r element subset corresponds to an $n-r$ element subset of elements that were excluded.

488. Let the stones in the piles be $x_1, x_2, \ldots, x_n$. Consider $|x_i - x_{i+1}|$.

489. There are several ways to approach this. The simplest one, though, is to first pick two from the eight, and then pick another two from the remaining six, etc.

490. Consider the company with no flights cancelled. What does it say about the component A?

491. The answer is $\lceil \frac{n}{2} \rceil + 1$. Find a construction (this is not so hard - you can create an S so that all of it's elements are certain Fibonacci numbers.)

492. Since the graph is bipartite, there is an independent set with at least half vertices. To show that this is the best bound, prove Hall's criteria.

493. What is common to each of $1, 3, 5$?

494. Find a way to work with the first and second halves of the string separately.

495. Magic trick: Sum 1/group size over all students

496. You can create towers of 2: something like $2^{2^{2^{2}}}$

497. Let's say I start at an arbitrary vertex. The issue is that the sum of the first k reals probably becomes non-positive at some point. Consider the lowest value you ever got to. Start there!

498. Once a sequence of P's and N's of size a_n repeats, the sequence is essentially just periodic.

499. $\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = 2^n$

500. Imagine coloring each point red with probability x. What does the expression $x^{a(P)}(1-x)^{b(P)}$ signify in terms of probability.

501. Greedy once again

502. Label a number with the set of prime factors less than k it has. We are asked to show the good/badness of n depends only on the type of n. Use induction on n.

503. The main idea here is to keep playing! As long as it takes.

504. Draw a bipartite graph with one side having n vertices representing the n fields in the flag, and the other side having all $2^{n-2}+1$ flags, with edges if there is a zero in the flag at the corresponding field.

505. Any move amounts to doing some sort of $(-1, -1, +1)$. Plus minus trick?

506. Solve the problem for some 3 and 5 until you're convinced about the answer.

507. Set a weight of $1/2^{a+b}$ on the point with coordinates (a, b) and consider the summation of the weights based on the location of the chips.

508. Show that there are no odd cycles

509. The trick is to continuously replace $\binom{n}{r} + \binom{n}{r+1} = \binom{n+1}{r+1}$

510. Play around with different examples until you're clear about what the final configuration should be and then use induction

511. Show that for $n \geq 3$ and n odd, B wins. In particular, assume that it does end in a draw and figure out what A's moves must have been to get a contradiction.

512. If the statement is false, there must be a triangle. If all three vertices have the same color, we're done. If some two are the same and the third is different, use this to create a triangle with all three colors as the same

513. Show that after you place a_k, the difference between the two piles is at most a_k

514. Show that we can always get to at most 2 piles using the previous problem (you can also prove this differently without quoting). It remains to show that one pile cannot be achieved

515. $a^2 = a + 1$, so $a^{n+2} = a^{n+1} + a^n$

516. Construct other lines parallel to ℓ. Show that eventually we'll have the same color configuration on two parallel lines, at which point 2 of those are of the same color.

517. Do this entire procedure for each vertex to end up with a directed cycle (n edge trick!!)

518. Think of each group as a vertex. Now draw an edge from v_1 to v_2 if the girl in v_1 shook hands with the boy in v_2, and from v_2 to v_1 in the other case

519. Take the smallest W that leads to a contradiction. Try and construct a smaller W.

520. Do a little bit of casework if there are 3 consecutive vertices among the five. Once this is done, the only case is where there are at most two consecutive vertices - most of the things are forced here.

521. Guess the answer! Experiment with numbers

522. Replace the $\binom{r}{r}$ with $\binom{r+1}{r+1}$ (both are equal to 1)

523. If someone has too many friends in their room, shift them to the other room.

524. The possibilities are: $\{1\}, \{2, 3, \ldots, n-1\}, \{n\}$ and $\{1, 4, 7, \ldots\}, \{2, 5, 8, \ldots\}, \{3, 6, 9, \ldots\}$ (along with their 6 permutations each). Prove this by induction.

525. Consider the number of "cross-connections" (number of friendships between people in different groups).

526. Try to make a *smaller* such triangle

527. To show that there is only one end point, assume there are at least 2, and take the last common vertex from where the paths leading to two different end points arise

528. Almost any move we make reduces the number of chains, and no move increases it. In what cases does this reduction not occur?

529. Take N large enough.

530. For the minimum, sort both a_i and b_i in ascending order (and sort them opposite to each other for maximum).

531. Even with the ties, there'll be at least one room that never reduces in size.

532. Blindly trusting answers from base cases isn't too useful - try and look for patterns in the "worst case sequence". If this is in fact the worst case, showing so is not too tough.

533. What position must Cinderella avoid?

534. We need to make sure we use every piece of information - so first gather them together

535. Use the fact that $i - j$ is even to show that ab odd fails

536. Consider what happens on the first move - either (with some probability) you end up with $n-1$ shoelaces and the same number of loops, or $n-1$ shoelaces with one more loop

537. We have that $w_{i+1} = w_i + w_{i-1}$ - so w follows the Fibonacci recurrence! Show that the other equation is also satisfied if the Fibonacci recurrence is satisfied.

538. Recurse, and remember some extra information

539. For 9, try applying this method twice

540. If there are n boxes and $2n-2$ coins, Annie can win by setting two consecutive boxes with 1 coin and all other boxes with 2 coins

541. Work with some cases - but be prepared, your first guesses on the answer are likely to be wrong. Keep experimenting

542. We allow the two sets to be the same. You might have to use the following inequality (from Cauchy Schwarz)

$$x_1^2 + x_2^2 + \cdots + x_n^2 \geq \frac{1}{n}(x_1 + x_2 + \cdots + x_n)^2$$

543. If you see two distinct colors, just skip.

544. Show that this graph doesn't have any repeating edges (we can't delete either of 1 or 2 and end up at equal sets in each case), and now use n edge trick for your cycle and create a contradiction

545. The trivial thing to do would be to take the smaller number in each column - but that doesn't give us the bound we need (the numbers could be close to equal, with the smaller number in the upper row each time.)

546. Consider the obvious graph interpretation, and show that the graph can contain no odd cycles.

547. Try finding the expected value of the number of segments that exist when you try to pair up the points randomly.

548. Try to create $\binom{w_i}{3}$ somewhere

549. If there are two coins of weight $1/2k$, trade them for a $1/k$ coin. If there are k coins of weight $1/k$, group them together to 1.

550. Try to achieve $f(a,b) = f(a,b+1) + f(a+1,b)$

551. As expected (given the chapter this is in), the answer is yes. Try and find ways to generate a large number of coins from simple starting points

552. You can keep reducing the chromatic number

553. We have $k+1$ pairs if you did this correctly. Use a certain example from the chapter! Deal with the even case similarly (you'll need to be slightly more careful for this case)

554. We need both sets to have the same number of vertices. What number appears twice?

555. To prove the previous statement, create a bijection.

556. We need to take the summation of some expression over each student and think of this in two ways

557. Find a way to bring all the tokens together when n is a power of 2 (this is not so hard.)

558. Take two random committees and find the expected value of common members. (You'll have to consider the number of committees each person is in as a set of variables)

559. Use the first problem twice

560. We cannot cheat, but we can...

561. Find another way of getting from a to b. Note that this isn't necessarily a path (we may have repeated a vertex), but if there's a *way* to get from a to b, there must be a path too (why?).

562. None of the six could've spoken about the topic they did with the fixed scientist, this leaves 2 possibilities. Repeat by fixing one of these six (but only 2 possibilities for the topic now).

563. In all other cases, Alice can use her first move to convert it into an N position.

564. Obviously the other part of the graph can have no odd cycles of length more than 3. If it has a triangle, show that this must have originated from a K_4. Handle this case separately by changing the set of edges we would delete in such a case. (You'll have to be careful that you don't go into some sort of loop - find a simple monovariant)

565. For ten balls, we have $120 \times 6 \times 7 \times 8 \times 9 \times 10$

566. The answer is $k \neq n$

567. If at most k people choose a person, $\left({}_{3\leq}\frac{{}^{k}1}{n}\binom{n}{3}\right)$. Conclude

568. Join two vertices if they represent a row and column respectively that together point to marked a cell.

569. The answer is that Bob wins the N-game are exactly those that, when expressed in binary, have no 1's in the 2^0, 2^2, 2^4, ... positions

570. Take some cases - check what must be going on if Alice is saying the truth, what happens if he is lying, etc.

571. Modulo m trick strikes.

572. The idea is to consider any row and the distance between two colored squares in it. In particular, if you have, say, 2 adjacent colored squares, you can't have 2 adjacent colored squares on any other row.

573. Double count the number of non triangles

574. We've worked with tokens that have two possible states before. Think binary

575. The endpoints can be adjacent to only the members of the path.

576. Let's say we have a set of points so that every line through two of them passes through a third. Let's say there are some three points A, B, C on a line and there is a point D not on this line.

577. Think of only four points for a bit. How much can that contribute to $m(S)$?

578. To find out what information we gain when Alice says that he doesn't know the color of his hat, it's useful to find out in what case he *would* be able to tell the color f his hat

579. Let's say we want to show that for $n = 28$ we cannot reach a single pile. Show that there will in general, always be a pile such that the number of stones in it is not divisible by 7.

580. To avoid slightly messy calculations, we assume that n is even. (The answer changes very slightly for odd n). Do what we're.best at - play!

581. Color alternate columns red.

582. Don't get worried if you aren't very good at chess - this requires close to zero chess skills

583. If that room doesn't eventually contain everyone, then eventually nobody enters or leaves it. Now use induction.

584. Assume that after a certain move by Cinderella, $B_4 = B_5 = 0$, $B_1 + B_3 \leq 1$, $B_2 + B_4 \leq 1$. Show that she can maintain this sort of configuration two moves later too.

585. Play around a bit and get a feel of what is going on.

586. Assume Hall's condition is not satisfied, so there is some set of a fields so that at most $a - 1$ flags have a zero in them. Show that

$$2^{n-2} - a + 2 \geq 2^{n-a}$$

587. Consider number of white squares modulo 2.

588. There are not that many different numbers - so it may not be too far-fetched to hope that there's a block such that each element appears an even number of times

589. To construct, use five pairs each of (i, i) for $1 \leq i \leq 8$ and $(-i, -j)$ for $1 \leq i < j \leq 8$.

590. For the first one, we have 8! ways to permute, 3^8 ways to orient. We have to divide by 3 to account for the reachable cases, and also note that some of these jumbles are the same cube, just from a different angle - figure out the number of angles you could be looking at a cube from.

591. Double count (student 1, student 2, club) and (student, club, society) triplets

592. $a_n = b_{n-1}$ and $b_n = a_{n-1} + b_{n-1}$

593. This is the same idea as the one we used to solve Example 10 in the chapter.

594. You'll need to double count two quantities here and then relate the two expressions. Use the first condition for your first quantity and the second one for the second quantity.

595. The answer is $2^{n-2}+1$. To prove that 2^{n-2} are not sufficient, set the first color as yellow and the second as blue for each flag

596. So from the first statement, we infer that Bob's number is at least 2. Now repeat this argument until we reach the fourth statement.

597. What happen if there are three non collinear points, all of the same color?

598. To show that $k = n$ does not work, assume that there is a strategy and use the superposition trick (you can also do it directly here by assuming that the wizard has access to your strategy)

599. First get lots of subsets with multiples of 3 and theme combine these subsets correctly

600. Surely, you know it by now: you have to be greedy!

601. Color each vertex according to the parity of its distance from the root so that there are a vertices of one type and $n-a$ of the other. Find how many of the $\binom{n}{2}$ pairs should have odd distance based on this. Take cases on the parity of $\binom{n}{2}$

602. Let's say without loss of generality $1, 2, \ldots, k-1$ belong to A and k belongs to B. Consider an element t in C

603. This is one of the places you want to avoid getting too mislead by some of the base cases - the game is not a draw for any $n \geq 7$.

604. The value you get is more than $n-1$. Conclude!

605. Find an invariant.

606. You should have got an expression involving $\sum \binom{a_i}{2}$. This is minimised when all the a_i are equal. What is their sum?

607. Yes, she can prevent the stepmother

608. Draw an edge from i to $j+1$ if $\sum_{k=i}^{j} a_k = 2021$. This graph is bipartite

609. Use the third condition to get a common expression for both the quantities

610. Pair numbers so that the sum of numbers in each pair is divisible by $2k+1$

611. Proof by contradiction!

612. Create some sort of monovariant

613. Just find a value of N where the idea of taking the biggest number each time fails. The difference here is that while $1 \mid 5 \mid 10 \mid 50 \mid 100$, $2 \nmid 3$.

614. Try lots of base cases!

615. If $|W| > n/2$, show that $N(W) = B$.

616. Let's say the connected components have $x_1, x_2, \ldots, x_k$ edges.

617. Think of the 2003 polygons on each sheet as the two sets

618. You need to prove two things: if some sequence leads to a stable configuration, all sequences lead to a stable configuration. Also, if two sequences lead to a stable configuration that stable configuration is the same.

619. Show that if a game is infinite, each vertex is fired infinitely many times.

620. Consider a graph with $2n$ vertices - n representing the rows and n representing the columns. How do we define edges?

621. Start with a_n

622. Consider a graph where each vertex represents and $x \in S$ and join two vertices if their difference is F_{2k+1}. Show that this graph can't have a cycle!

623. If the number of stones are *close* to equal, you can copy

624. Root the tree at a vertex v and consider the distances to each other vertex.

625. The idea is that every time the first player makes a move, they must left a gap (since they couldn't have played at a consecutive number). So the second player tries to play in that gap as far as possible

626. Show by induction/Ore's theorem that that a Hamiltonian cycle does indeed exist.

627. Show that they can't cover all residues (think parity)

628. How many powers of 2 are there? And how many possibilities are there for the leading four digits?

629. For Beto to win, Beto must be able to merge any two given turkeys.

630. Keep deleting vertices until all degrees are even. When you perform step two, all degrees are now odd: Show that the chromatic number didn't increase when you did this, and now try and reduce chromatic number by one.

631. This is the tricky part - look for a similar but different coloring which also involves each stromino using exactly one cell of each of the colors

632. Start by saying pq (where p and q are two primes dividing n), and remember: we can't cheat but we can...

633. For N even - try working modulo 4 and think binary.

634. Fix any one scientist - now it must have spoken about the same topic with at least 6 of its neighbours

635. The answer is 2022

636. To work with chromatic number, we'll work with independent sets instead, and try to make the graph have a low probability of having a large independent set

637. Think about powers of 2, and solve it for small n to guess the answer.

638. Figure out the significance of $n - 2$

639. Consider the parity of the number of piles with minimal stones.

640. If all positions that are at most k moves away from a result have been explored, we can now explore positions that are at most $k+1$ moves away and use our previous results to find the result here

641. $a_n = 3b_{n-1}$ and $b_n = 2b_{n-1} + a_{n-1}$

642. Now we want to show that when n is not a power of 2, not *all* positions are collapsible. So you just want one configuration which you can prove isn't collapsible.

643. Create a smaller distance using one of lines AD, BD, CD.

644. If there are 4 people who our fixed person knows, we're done: none of those 4 people can know each other. On the other hand, if the fixed person doesn't know 6 people - we're happy (use Example 5).

645. If we want a certain residue in $A+B$, say k, we essentially want one of $k-a$ to be in B where a is any one of the n elements in B. Use this to find the probability

646. So we get that each remaining value in the bottom row is at most $1 - \frac{n+1}{4(k+1)}$. Multiply this by $(n-k)$ and do the bounding.

647. So the graph is bipartite. Show that each component has 25 vertices (use the second condition)

648. Show that money can't always be transferred into one account (this is a one liner)

649. The rightmost stone either just sits there (if we're not operating on it) or it moves further right when we operate on it. So we just want to show that the rightmost stone can't reach too far right. The issue is that it's not so obvious - since the weights here can become negative as we're considering Fibonacci for negative indices too! So firstly get a feel of how Fibonacci looks for negative indices $(F_i = F_{i+2} - F_{i+1})$

650. Take the natural graph interpretation with rooms as vertices - you want each vertex to have degree 2

651. There are $24 + 24 + 24 + 24 + 24 = 120$ possibilities.

652. Remember the "if and only if". First show that 50 questions are enough. (Yes, this is easy.)

653. Think of bipartite graphs

654. If A is at the start, there are 24 possibilities

655. Take the minimum number.

656. This is equivalent to $x_1 + x_2 + \cdots + x_8 = 4$, with the condition that $x_1 \geq 1$ and the rest are at least 0.

657. Fix any 3 collinear points on a line ℓ, and take 3 parallel lines perpendicular to ℓ through these points.

658. Ask (a_1, a_2), (a_1, a_3), (a_1, a_4) and so on. The first two answers are yes - so if we ever get a no, finish.

659. Announce that there is exactly one Mafia counter in each block.

B Solutions

Problem 1 (Oxford MAT) — A positive integer is written on the forehead of each of Alice and Bob. They can see the number written on the other person's forehead, but they can't see the number on their head. Naturally, neither of them knows the number written on their head. Now imagine an angel arrives and makes the following statement, "The numbers on your foreheads differ by 1." So the number on Alice's forehead is either one more or one less than the number on Bob's forehead. Imagine the following scenario:

> Alice begins by saying, "I don't know my number."
> Bob replies, "I don't know my number either."
> Alice replies, "I still don't know my number".
> Bob replies, "Aha, I now know my number!"

Find all possible values of Alice and Bob's numbers.
Also, show that in general, given that the numbers are consecutive in some order, Alice and Bob can figure out their numbers by just repeatedly saying "I don't know my number" like above.

Solution 1. Let Alice's number be a and Bob's number be b

When Alice first says that she does not know her number, we can infer that Bob's number is not 1. This is because if Bob's number had been 1, Alice would be sure that her number was 2 (since both numbers are positive integers and they have to be consecutive). So we have that $b \geq 2$.

Now when Bob says that he does not know his number, Alice's number cannot have been 1 or 2, as in the first case Bob's number is guaranteed to be 2 and in the second it is guarenteed to be 3 ($b \neq 1$). So we know that $a \geq 3$.

Now when Alice says that she still doesn't know her number we infer that $b \geq 4$ since if $b \leq 3$ there is only at most one choice for a that satisfies $a \geq 3$.

So we have that $a \geq 3$ and $b \geq 4$. At this point, Bob says that he knows his number. So he either sees that $a = 3$ and concludes that $b = 4$, or he sees that $a = 4$ and concludes that $b = 5$.

For the second part of the problem, we essentially generalise this approach - think about it until you are convinced about its working

Problem 2 — Ten prisoners are given the following situation: they are all given a hat with one of ten colors uniformly at random: it's possible everyone gets the same color, and it's also possible that all ten prisoners get ten different hat colors. Each prisoner can see the hats of others but not their own. Now, each prisoner must

> guess what they think is the color of their hat. If even one of them guesses their hat color correctly, they are all free. Show that by discussing a strategy beforehand, the prisoners can guarantee freedom.

Solution 2. Label the ten colors with numbers from 1 to 10.

Now consider the sum of the numbers of all ten hats. Let this be S.

$$S = 10k + r$$

where r is the remainder on dividing S by 10. There are 10 possible values for r.

The main idea is to have each person make a unique guess on this r - so the first person guesses that $r = 1$, second person guesses that $r = 2$ and so on.

Now based on this guess, they can figure out what number their own hat must be. As an example, if the fourth person (who thinks $r = 4$) looks at all other hats and sees that the sum of the relevant numbers on these hats is 52, then they would guess that their own hat must have 2 on it (to make the total sum 54, which leaves remainder 4).

Now one of the ten people must have made a correct guess on the value of r, and in turn will have guessed the color of their hat correctly! Magic.

> **Problem 3** — Snow White and the seven dwarves are playing a game, and they need to be split into groups of 2. Show that this can be done in exactly 105 ways.

Solution 3. We solve the more general case where there are $2n$ people that need to be split into groups of 2. Note that there are $\binom{2n}{2}$ ways of picking the first group of 2, then $\binom{2n-2}{2}$ of ways of choosing another 2 and so on. However, notice that each such splitting gets counted $n!$ times (the order in which we pick these groups of 2 is irrelevant). In all, the number of ways is

$$\frac{1}{n!}\binom{2n}{2}\cdot\binom{2n-2}{2}\cdots\binom{2}{2} = \frac{(2n)!\cdot(2n-2)!\cdots 2!}{n!\cdot(2!)^n\cdot(2n-2)!\cdot(2n-4)!\cdots 2!} = \frac{(2n)!}{n!\cdot 2^n}$$

Another more direct way of going about this is the following: pick a partner for the first person. $(2n-1)$ ways. Now pick a partner for another unpaired person: $(2n-3)$ ways. Repeat this. We end up with the total number of ways as

$$(2n-1)\cdot(2n-3)\cdots 3\cdot 1$$

Check that this is the same answer!

> **Problem 4** — Consider a $2\times 2\times 2$ Rubik's cube. Each of its 8 pieces can be oriented in three ways, and each of the 8 pieces can be permuted. However, only one third of these can be reached with a solved cube (others require a "twist"). Find the number of possible jumbles of the cube.
>
> On a $3\times 3\times 3$ cube, there are 8 corners and twelve edge pieces. Corners can be oriented in three ways, edges in two, the corners can be permuted amongst

themselves, and so can the edges. 1/6 of these can be reached with a solved cube. Approximate the number of jumbles this time.

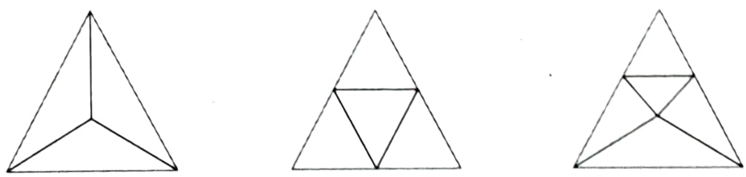
Solution 4. We solve the $3 \times 3 \times 3$ cases. We need to permute the edges, permute the corners, orient the edges, and orient the corners. So there are a total of

$$12! \cdot 8! \cdot 2^{12} \cdot 3^8$$

ways. We now divide by 6 account for the "reachability". (Note that if you twist a corner of a cube once, it becomes impossible to solve - this intuitively explains why such reachability problems occur.) So the final answer we end up at is 86504006548979710000. To put things into perspective, this value is more than double of the number of seconds in a trillion years!

Problem 5 — For any $n \geq 3$, prove that an equilateral triangle can be divided into n isosceles triangles.

Solution 5. Here are the solutions for $n = 3$, 4 and 5.



Now once we have these, note that we can induct from $n \to n+3$ by splitting an isosceles triangle into four triangles by introducing the midpoints of sides (like in the solution for $n = 4$). Since we have a solution for $n = 3, 4, 5$ we now have a solution for $6, 7, 8$ and an easy induction gives that all $n \geq 3$ can be achieved.

Problem 6 (IMO 2000/4) — A magician has one hundred cards numbered 1 to 100. He puts them into three boxes, a red one, a white one and a blue one, so that each box contains at least one card. A member of the audience draws two cards from two different boxes and announces the sum of numbers on those cards. Given this information, the magician locates the box from which no card has been drawn. How many ways are there to put the cards in the three boxes so that the trick works?

Solution 6. We claim that the answer is $2 \cdot 3! = 12$ ways

Note that the solutions $\{1\}, \{2, 3, \cdots, 99\}, \{100\}$

And $\{1, 4, \cdots, 97, 100\}, \{2, 5, \cdots, 98\}, \{3, 6, 9, \cdots 99\}$ along with their permutations provide us with the claimed 12 solutions.

We now prove that these are the only solutions.

Let the boxes be labelled as A, B, C

Note that if for $a \in A$ and $b \in B$, the sum $a+b$ should let the magician guess that the box no card taken from was C and so we call the number $a+b$ unique to C.

We apply induction claiming that the solutions for n are of the 2 forms listed above, the base cases $n = 3, 4$ are quite direct, consider the following 2 sub cases for n+1:

- The number $n+1$ is alone in the of the boxes

  WLOG Let $n+1$ belong to box C and 1 belong to box A

  Thus $n+2$ is unique to B and thus the numbers 2 and n cannot belong to different boxes (i.e one in A and one in B)

  If $2, n \in A$ notice that $n+3 = (n+1)+2$ is unique to B and so $3 \in A$

  This gives $n+4 = (n+1)+3$ is unique to B and so on so in fact all numbers from 1 to n will end up in A but this will leave B empty which is a contradiction

  Thus, $2, n \in B$ and in a similar way, $n+3 = (n+1)+2$ is unique to A and thus we get that all numbers from 2 to n belong to B

  This gives us the solution $\{1\}, \{2, 3, \cdots, n\}, \{n+1\}$ and its permutations.

- There is an element other than $n+1$ in its box.

  We now apply the induction using the fact the removal of the card $n+1$ should result in a working solution of n

  We know that the only 2 working solutions of n (excluding permutations) are

  1. $\{1\}, \{2, 3, \cdots, n-1\}, \{n\}$

  Simple casework shows that wherever you put $n+1$ you will achieve a contradiction

  2. $\{1, 4, \cdots\}, \{2, 5, \cdots\}, \{3, 6, 9, \cdots\}$

  We just consider the 3 different possible remainders of $n+1$ when divided by 3 and show that in each case it must go in the box with all the other numbers of that remainder

Thus we have proven that for any $n \geq 4$ there exist 12 solutions to the given problem!

Problem 7 — We begin with the numbers $1, 2, \ldots 2023$ written on the blackboard. At each step, we can erase any two of the numbers a and b and then write down the number $|a-b|$. We continue until one number remains. Determine whether this final number could be equal to 1.

Solution 7. At the start, the sum of numbers is $1+2+3+\cdots+2023 = 2023 \cdot 1012$ which is even. Notice that the parity of this sum does not change since

$$|a-b| = a+b \pmod 2$$

Thus, at the end the sum of all remaining numbers should be even. Since only one number remains, this number must be even and can therefore not be equal to one.

Problem 8 — A room is initially empty. Every minute, either one person enters or three people leave. After exactly 2023 minutes, could the room contain exactly 201 people?

Solution 8. Note that the number of people modulo 4 changes by exactly $+1$ each minute. So after 2023, minutes, we must have a $+2023 \equiv +3$ modulo 4, and since we started off at 0, the number of people should be of the form $4k+3$, but 201 is not of this form.

Problem 9 — Determine whether it is possible to tile a 10×10 square floor using 1×4 rectangular tiles

Solution 9. Color the cells according to $x + y \pmod 4$. Now any 1×4 covers cells of four different colors - but counting the number of squares of each color we get a contradiction.

Problem 10 (ELMO 2019 SL) — Bella and James are playing a game. Initially, $n \geq 3$ points are given on a circle. On a player's turn, that player must draw a triangle using three unused points as vertices, without creating any crossing edges. The first player who cannot move loses. If Bella goes first and players alternate turns, who wins?

Solution 10. We claim that Bella can win. In the first move, Bella plays according to the following strategy.

If n is odd, take vertices $1, \frac{(n+1)}{2}, \frac{(n+3)}{2}$ and if n is even, take vertices $1, \frac{n}{2}, \frac{n+4}{2}$

Note that the triangle splits the polygon into three regions. One of them has only 0 or 1 vertices and the other two have an equal number of vertices. The main observation is that any move that is made now will have to be within a single region - if points from multiple regions are used, there will be intersections with the initial triangle picked by Bella.

From here, the task is simple. Bella makes sure that the two equal regions are symmetric at the end of her move. Every time James makes a move in one of them, she makes an identical move in the other (by reflecting the move across the y-axis, say).

Problem 11 (Codechef) — Two positive integers: N and k are fixed. A blackboard initially has N written on it. Alice and Bob play the following game: in a move, they must subtract k^x for some non negative integer x from the number on the blackboard, with the added criteria that the number on the board should remain non negative. A player who cannot make a move loses. Who has a winning strategy (in terms of N and k)?

Solution 11. For $k = 4$, the solution is that B wins if and only if $k \equiv 0, 2 \pmod 5$.

In general, if k is even, the idea is to look modulo $k + 1$. Now any move does either a $+1$

or a -1 modulo $k+1$ (since $k^n \equiv (-1)^n \pmod{n+1}$).

Now based on the results, it looks like the answer is

$$0, 2, \ldots, (k-3) \pmod{k+1}$$

are the cases where Bob wins. The trick to proving this is to maintain symmetry - Bob can make sure that at the end of his turn, the number returns to one of these if it originally began at this. In particular, if Alice removes a number which is $-1 \pmod{k+1}$, we subtract 1, and if Alice removes a $1 \pmod{k+1}$, we subtract k. The only time we can run into trouble is if the k is too large to subtract - but in that case the number is already in range 0 to $k-1$ and the only legal move is to subtract 1 - and by doing this repeatedly we can guarentee a win (the parity works out in our favour.)

And if we start at any other remainder, Alice spends their first move reaching one of these (they just need to subtract 1 in most cases, and if they're at $k \pmod{k+1}$, subtract k).

Problem 12 — Alice and Bob are playing a game on an $m \times n$ board - all the cells of which are initially blue. In a move you're allowed to choose any cell (a, b) that is currently blue and change the color to all cells in the rectangle with corners at $(1, 1)$ and (a, b) to red (cells that were already red remain red). If at the end of your turn, no cell is blue - you lose. Show that if $mn > 1$, A has a winning strategy.

Solution 12. Let's say that for the sake of contradiction, B has a winning strategy. We pick $(1, 1)$ on our first turn. Now, Bob picks (x, y) after this and claims a winning strategy. The idea then is that we could have picked (x, y) directly in our first move itself, and use Bob's strategy against him!

Problem 13 (LMAO 2023/4) — Let $\pi(n)$ denote the number of primes at most n. Show that if we have a subset of $\{1, 2, 3, \ldots, n\}$ of size $\pi(n) + 1$, some number divides the product of all others.

Solution 13. Let's say for the sake of contradiction we have a set $S = a_1, a_2, \ldots, a_k$ of size $k = \pi(n) + 1$ such that no number divides the product of others. If this is true, there must be a prime p such that the exponent of p dividing a_1 is greater than the exponent of p dividing the product of the others. In other words there must be a prime p so that

$$\text{Power of } p \text{ dividing } a_1 a_2 \ldots a_k \le 2\times \text{ (Power of } p \text{ dividing } a_1)$$

Now note that the same prime p cannot be associated with a_i and a_j both (check this!). So we have that each a_i is mapped to a unique prime. But there are $\pi(n)$ primes and $\pi(n) + 1$ numbers, contradiction.

Problem 14 (LMAO 2023/5) — There is a row of 2022 identical boxes, each with one coin inside it. On your turn, you are allowed to open any box and if it contains a coin, take it. After you are done, the game director will secretly swap the box you opened with one of its neighboring boxes. Determine the maximum number of coins you can guarantee to collect in finite time.

Solution 14. (Due to Atul Nadig)

The answer is $\boxed{1517}$. In general, the answer is $n - \left\lfloor \frac{n+2}{4} \right\rfloor$, and the proof for general n is the same as for 2022.

To show 1517 is achievable, consider the following sequence of box openings

$$1, 1, 5, 5, 9, 9, \cdots, 2021, 2021, 3, 7, 11, 13, \cdots, 2019$$

where first every 1 (mod 4) box is opened twice and then every 3 (mod 4) box is opened once. This works since opening the 1 (mod 4) boxes twice, gives a coin from the opened box and one adjacent to it. Since no 3 (mod 4) box is adjacent to a 1 (mod 4) box, after we are done with the first step, we can open these and take the coins in them, for a total of 1517 coins.

So it suffices to show you cannot get more than this. Call the following pairs of boxes *friends* $(1,2),(3,4),\cdots,(2021,2022)$ and the following pairs of boxes *neighbors* $(2,3),(4,5),\cdots,(2020,2021)$. Further, call every box that is 0 or 3 (mod 4) to be *good*, and the rest *bad*. Note that every swap of a box is with either its friend or neighbor.

I will use the following strategy:

- If you open a bad box, I will swap it with its friend.
- If you open a good box whose neighbor has a coin, I will swap it with its neighbor.
- If you open a good box (say G) whose friend is say F and whose neighbor is empty, I will either swap with its neighbor or friend depending on whether you open G or F first in the future respectively.

I claim that among any good pair, there will always be either a coin in both boxes, or one coin such that you will not know in which box the coin is (and hence not find the coin if you open any one of the boxes). Call such a position a *nice position*. A nice position always stays nice. Consider any good pair of friends, and say you open a box B_k, WLOG assume B_k and B_{k+1} are friends.

By niceness, B_{k+1} must have a coin. Now there are two cases:

- If B_{k-1} has a coin, then swapping with it makes both B_k and B_{k+1} have a coin, which maintains niceness.
- If B_{k-1} does not have a coin, then swapping with B_{k-1} and B_{k+1} are the exact same configurations apart from the position of the coin, which can be in either B_k or B_{k+1}, and so you have no information on which one it is in.

Since any good pair of friends which is nice always has at least 1 coin in it, and there are 505 such pairs, you can guarantee to get at most $2022 - 505 = 1517$ coins, as desired.

> **Problem 15** (2019 USAJMO/1) — There are $a+b$ bowls arranged in a row, numbered 1 through $a+b$, where a and b are given positive integers. Initially, each of the first a bowls contains an apple, and each of the last b bowls contains a pear. A legal move consists of moving an apple from bowl i to bowl $i+1$ and a pear from

bowl j to bowl $j-1$, provided that the difference $i-j$ is even. We permit multiple fruits in the same bowl at the same time. The goal is to end up with the first b bowls each containing a pear and the last a bowls each containing an apple. Show that this is possible if and only if the product ab is even.

Solution 15. We first prove that ab odd does not work and then we provide an algorithm for a even and for b even

Note that if ab is odd, both a and b must be odd.

Define A_0 as the apples on squares with even numbers at any point and A_1 as the apples on squares with odd numbers at any point.

Similarly define P_0 as the pears on squares with even numbers at any point and P_1 as the pears on squares with odd numbers at any point.

Claim: $(A_0 - A_1) - (P_0 - P_1)$ is invariant under the given process

This is clearly true since by the given condition we either move an apple as well as a pear from an odd to an even number or move both from an even to odd number (since $i-j$ is even).

However, if A and B both are odd we can check that at the start of the process the invariant has value -2 but the final value has value 2. Thus ab cannot be odd.

We now provide an algorithm to convert

$$AAA\cdots APPP\cdots P \to PPP\cdots PAAA\cdots A$$

with a even (and b may be either odd or even). The case where b is even is very similar (you can think of it as applying an identical algorithm from the other end).

We first select i and j as 1 and $a+1$ which moves an apple to 2 and a pear to a. We now select $(2, a)$ and so on until the pear reaches the first position and the apple reaches $a+1$.

We now operate on $(2, a+2)$ and so on so that the apple and pear on positions 2 and $a+2$ get swapped.

If $a \geq b$, continuing this process we finally switch the apple and pear at positions $(b, a+b)$. Notice that at this point, the first b bowls have pears and the last a have apples, so we are done.

If $a < b$, we perform the algorithm until we operate on $(a, 2a)$ (the index $2a$ clearly exists since $a+b > 2a$). This leads us to the position starting with a pears, followed by a apples and then followed by $b-a$ pears. Now ignoring the first a pears, we are left with a apples and $b-a$ pears. Since $a(b-a)$ is even, we can apply strong induction (how?) here to flip the order of these apples and pears and conclude!

Problem 16 — You are playing a game with the Devil. There are n coins in a line, each showing either H (heads) or T (tails). Whenever the rightmost coin is H, you decide its new orientation and move it to the leftmost position. Whenever the rightmost coin is T, the Devil decides its new orientation and moves it to the leftmost position. If you can make all coins face the same way within 2^n moves, you win - else the devil wins. Show that you have a winning strategy, i.e, no matter what the devil does, you can find a way to make all coins heads or all coins tails.

Solution 16. Imagine the coins are not actually moving from the right to left, but the coins are being operated on cyclically. (In particular, the first operation will occur on the rightmost coin, but instead of moving it to the left, it stays there and now the next operation is on the second coin from the right, etc.)

Let there be n coins. Then consider the first n moves as the first round, etc. Let's assume that the Devil can prevent you from winning, so he will never let the position become all heads or all tails. We try to increase the number in binary (where heads are 1 and tails are 0) at the end of each round.

Ironically, the strategy for us, though, is to convert heads into tails from the start of the round until the devil first changes a tail into a head. If the devil does not do this the entire round, we win since all coins now face heads. If the devil does switch a tails to a heads, we now switch our strategy to keeping each heads as a head. Now at the end of the round, we compare the binary numbers. Some of the digits on the right have changed from 1 to 0, but a digit to the left of these has also changed from 0 to 1 (due to the Devil's move). As an example,

$$1010111 \to 1011000$$

So the binary number has increased, and in particular it'll eventually reach all heads!

Problem 17 — You are given denominations of \$1, \$5, \$10, \$50, \$100. You want to use these notes to get a total of $\$N$. However you want to minimise the number of notes you use. Find an algorithm that takes $\mathcal{O}(1)$ time to output the minimum number of notes required.

Solution 17. Use a greedy algorithm that picks the largest note you can at any point in time. In particular, use $\left\lfloor \frac{N}{100} \right\rfloor$ 100 dollar notes.

After this, the remaining money N' is less than 100, so use $\left\lfloor \frac{N'}{50} \right\rfloor$ 50 dollar notes, and just repeat this.

We now show that this has to work. For this, consider the "official solution". Consider both solutions as lists of notes you used in decreasing order: so something like

$$100, 100, 50, 10, 10, 5, 1$$

Let's say our solution matches that one till the kth position. But at the $(k+1)$th position, they used a smaller note than we did. (It couldn't have been a bigger note since our greedy algorithm literally picks the largest note.) Let's say we picked note x, and they picked note y. The idea is to merge notes y and the ones after it until we cross over x.

In particular, for an example:

$$100, 100, 100, 50, \ldots$$
$$100, 100, 100, 10, 10, 10, 5, 5, 5, 5, \ldots$$

So we can merge the $10, 10, 10, 5, 5, 5, 5$ into a single 50. The only issue is if we somehow "jump over" the 50 - but it is easy to see that that can't happen here, since each number divides the next one.

> **Problem 18** — You are given denominations of \$2, \$3. You want to use these notes to get a total of $\$N$. However you want to minimise the number of notes you use. Show that you aren't as lucky this time. In particular, show that the greedy algorithm we used last time will not lead to a valid solution.

Solution 18. $N = 4$. Forget inefficient, you'll get completely stuck if you start with 3, the biggest number that can be picked. In particular, you'll get stuck on all numbers of the form $3k + 1$.

> **Problem 19** — Five hundred people attend a party in a mansion with many rooms. Each minute someone walks from one room into a different room with at least as many people. Prove that eventually all the people will be gathered in one room.

Solution 19. We can induct as discussed through the hints, but here's a cooler solution: consider the tuple $(a_1, a_2, \ldots, a_n)$ where a_i is the number of people in the ith room, and $a_1 \geq a_2 \geq \ldots a_n$. The idea is that on the next day the tuple has become "larger" if you resort the elements. In particular, if you sort these the following way:

> Compare first element, if equal compare second element, if equal compare third element, etc.

You'll notice that the tuple is now larger. There are only finitely many tuples so eventually this ends, and it can only end at one point.

> **Problem 20** (IMO 2004 C1) — There are 10001 students at an university. Some students join together to form several clubs (a student may belong to different clubs). Some clubs join together to form several societies (a club may belong to different societies). There are a total of k societies. Suppose that the following conditions hold:
>
> 1. Each pair of students are in exactly one club.
> 2. For each student and each society, the student is in exactly one club of the society.
> 3. Each club has an odd number of students. In addition, a club with $2m + 1$ students (m is a positive integer) is in exactly m societies.
>
> Find all possible values of k.

Solution 20. Let the number of students in the ith club be $2c_i + 1$. We double count two quantities.

- (Student 1, Student 2, Club with both students). Here there are $\binom{10001}{2} \times 1$ possibilities according to condition 1 (we ignore the order of the students). However, if we fix the club first, we get a total of

$$\sum_i \binom{2c_i + 1}{2} = \sum_i (2c_i + 1)c_i$$

ways

- (Student, Club with given student, Society with given club). If we fix a student and society first in $(10001)k$ ways, there is exactly one club to choose. And if we use the third condition again by fixing the club first, we get

$$\sum_i (2c_i + 1)c_i$$

which is in fact the same expression we got earlier.

In particular, this proves that $\binom{10001}{2} = (10001)k$ so $k = 5000$. It is left to prove that this value of k can indeed be achieved. For this, consider just one club C which belongs to 5000 societies and such that all students are a member of this club.

> **Problem 21** — On the first day, there are a hundred students who are divided into five groups. On the second day, the same one hundred students are divided into four groups. Prove that there exists a student who belongs to a larger group on the second day than the first.

Solution 21. Let the ith student belong to a group of size a_i on day 1 and a group of size b_i on day two. Let

$$A = \sum_i \frac{1}{a_i}$$

$$B = \sum_i \frac{1}{b_i}$$

Note that the term $\frac{1}{a_i}$ repeats a_i times and so each group contributes 1 to the summation, and thus we get that $A = 5$ and $B = 4$. In particular, if $A > B$, at least one of the a_i must be less than b_i, so we're done.

> **Problem 22** — Given the graph G and cycle C in it, we can perform the following operation: add another vertex v to the graph, connect it to all vertices in C, and erase all the edges from C. Prove that we cannot perform the operation indefinitely on a given graph.

Solution 22. Note that the process for any connected component is independent to the process for any other connected component - so we can solve this one component at a time. In particular, the advantage of doing this is that we can now assume that the original graph was connected.

Think of the number of vertices and edges every time we perform this operation. The number of vertices increases by one since we added a new vertex. However the number of edges remains constant since we add k edges and remove k edges when working with a cycle of length k. Also note that if the graph was connected, it continues to stay connected. (For any two vertices in the cycle, there is still a path between them via the new vertex). Since a connected graph with E edges has at most $E+1$ vertices, the number of moves we can make is finite, so the process must end.

Problem 23 — Let G be a bipartite graph on $A \cup B$ with no isolated vertices. Assume that for each edge ab with $a \in A$ and $b \in B$, we have $\deg a \geq \deg b$. Prove that G contains a matching using all vertices in A.

Solution 23. For the sake of contradiction, let us say Hall's criteria was violated, for some subset X of vertices of A. Let $N(X)$ denote it's set of neighbours. This means that $|X| > |N(X)|$. We take the smallest set X for which this is true.

Consider a subset Y with only $|N(X)|$ vertices of X. These vertices must have a perfect matching or we have contradicted the minimality. Now note that these must have been matched with exactly the set $N(X)$. We have that

$$\deg a_i \geq \deg b_i$$

for each pair in the perfect matching, and thus the sum of degrees of the vertices in Y is at least the sum of degrees of the vertices in $N(X)$.

However we also have that the sum of degrees of Y is at most the sum of degrees of $N(X)$ since all the edges of Y are connected to only members of $N(X)$. This forces

$$\sum_{v_i \in X} \deg a_i = \sum_{v_i \in N(X)} \deg b_i$$

However this forces the degrees of all members in $X - Y$ to have degree 0, which is a contradiction since we assumed that there is no isolated vertex!

Problem 24 — Let G be a simple, connected graph on n vertices. An invisible rabbit and a hunter play a game on this graph. The rabbit starts at an arbitrary vertex, unknown to the hunter. They take turns alternatively. At his turn, the rabbit must move from his current vertex to an adjacent one. At the hunter's turn, he picks a vertex and checks if the rabbit is there. Characterize the graphs G such that the hunter has a strategy to capture the rabbit in a finite number of turns, regardless of the rabbit's initial position and movement.

Solution 24. We begin with the following claim

Claim 1: If the hunter has a strategy for a certain graph, he might as well announce that strategy to the rabbit.

Proof: This is by the very definition of a winning strategy. Since the strategy has to work no matter what the rabbit does, it should work even if the rabbit knows this strategy.

Also note that this strategy must be purely deterministic. If it has any decision based on probability, we can assume that the most likely event occurs each time (break ties arbitrarily). Since the strategy works no matter what the rabbit does, it should work when the most likely event occurs each time (in fact, it should work when the least likely event occurs each time too).

Claim 2: G must be a tree.

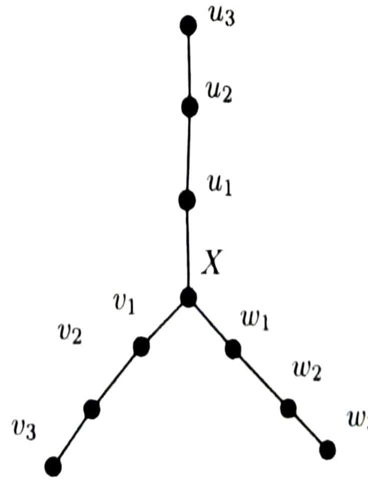
Proof: Let's say the graph G has a cycle $v_1 - v_2 - \cdots - v_n - v_1$ and the hunter has a winning strategy. As discussed in Claim 1, the hunter might as well announce this strategy to the rabbit.

The rabbit announces that it is at one of vertices v_1 and v_3 on the first day. Since it knows the vertex the hunter will check on the first day, he can avoid being captured (since he has two options and he just needs it to be different from the one the hunter checked).

Without loss of generality, let's say the hunter did not check v_3 on the first day (so he checked v_1 or some other vertex). Then the rabbit announces that he was on v_3, so he is now at one of v_2 and v_4.

Note that this position is isomorphic to the position before the first round, so the hunter has not made any progress. In particular, the rabbit can repeat an identical strategy each day to prevent the hunter from catching him.

Claim 3: Consider a graph with a central vertex C with three paths originating from it: $C - u_1 - u_2 - u_3$, $C - v_1 - v_2 - v_3$, $C - w_1 - w_2 - w_3$ as shown below. Call such a graph a trisector. The hunter cannot catch the rabbit on this graph.



Proof: Before the first day, the rabbit announces that it is at one of C, u_2, v_2.

- Hunter checks C: the rabbit places itself at one of u_2 and v_2 and announces that it is now at one of u_1, u_3, v_1, v_3. Now if the hunter checks u_1, notice that it is possible that the rabbit had gone

$$u_2 \to u_3 \to u_2$$
$$v_2 \to v_3 \to v_2$$
$$v_2 \to v_1 \to C$$

so it can now be at any of C, u_2, v_2 once again. In a similar fashion, it can reach each of its original squares for each of the four checks the hunter decides to make on the second day.

- Hunter checks u_2 (the case where he checks v_2 is symmetric to this): The rabbit places itself at one of C and v_2. On the next move, the rabbit can now be at any of u_1, v_1, w_1, v_3.

  In particular, the rabbit can even announce that it is at one of u_1, v_1, w_1.

  The hunter can check at most one of these (say u_1) and then the rabbit can ensure that it can now be at any of C, v_2, w_2, which is isomorphic to the original position.

Now we shall show that for all connected graphs with no cycles and no trisector subgraph, the hunter has a strategy. First let's characterise these graphs differently.

Consider the longest path in graph G. The idea is that any vertex is at distance at most 2 from this longest path. Let's say there is a vertex v in this path, and there are edges $v - u_1 - u_2 - u_3$ with u_1, u_2, u_3 not a part of the path, there is a trisector. (Note that if the actual longest path ended at a distance less than three from v, we could've used $v - u_1 - u_2 - u_3$ in our path instead, so there is definitely a trisector).

So all graphs that we claim to be attainable have a path, and any vertex outside this path is at distance at most 2 from a vertex in the path.

We now show that such a graph has a winning strategy.

Let the longest path in the graph be $v_1 - v_2 \dots v_n$. Obviously, v_1 and v_n don't have any neighbor other than v_2 and v_{n-1} (or we can make the path longer). Call the set of vertices at distance 1 from v_i outside the path S_i and the set of vertices at distance two from v_i outside the path T_i. Also let the union of v_i, S_i, T_i be X_i.

To start off, note that the graph is bipartite, and so we can color the vertices with two colors so that any edge is between red and blue vertices. We assume that v_1 is red, and construct a strategy that works if the rabbit begins on a red vertex.

If our strategy completes without the rabbit being caught, we know that the rabbit must have started on a blue vertex, and since we know the number of moves that have taken place, we know the color of the vertex the rabbit is now. In particular, we can waste a move if needed to make sure that the rabbit is going to move to a red vertex next, and then repeat the initial strategy.

On our first move we check vertex v_1. After this, alternate between v_2 and members of S_2, starting and ending with v_2. Then alternate between v_3 and members of S_3, starting and ending with v_3. Continue this until you reach v_n, by which point you should have caught the rabbit.

To prove the above, we use induction on the following hypothesis

> If you finish the process of alternating between v_i and members of S_i without catching the rabbit, it is now guarenteed to be in $X_{i+1} \cup X_{i+2} \cup \cdots \cup X_n$.

The base case is obvious with $i = 1$. To prove the induction step, let's say it is true for some $i = k$. Now we show that it holds true after we complete the procedure of alternating between v_{k+1} and S_{k+1}.

Notice that if the rabbit wants to enter X_k as we are performing these steps, it must go through v_{k+1}. However, we know that the hunter and rabbit are always on the same color (both flip the color of their square on each move and we assumed that we started off on the same color). So if the rabbit is at v_{k+1}, we cannot be on any of the S_{k+1} vertices, and we would have caught the rabbit.

Moreover, if the rabbit is in S_{k+1} or T_{k+1} it remains stuck there (since it cannot go to another region of the graph without going to v_{k+1}). As we check each member of S_{k+1}, it must get caught (once again due to parity forcing it into a member of S_{k+1}).

Thus, we are done!

> **Problem 25** — There are 100 shoelaces in a box. At each stage, you pick two random ends and tie them together. Either this results in a longer shoelace (if the two ends came from different pieces) or it results in a loop (if the ends came from the same piece). What is the expected value of the number of steps until everything is in loops, and the expected number of loops at the end of this process.

Solution 25. The idea is to define the expected value of number of loops for n shoelaces be $f(n)$. We relate $f(n)$ to $f(n-1)$ as follows.

$$f(n) = \frac{n}{\binom{2n}{2}}(f(n-1)+1) + \left(1 - \frac{n}{\binom{2n}{2}}\right) f(n-1)$$

Note that this means

$$f(n) = f(n-1) + \frac{1}{2n-1} \implies f(100) = \frac{1}{199} + \frac{1}{197} + \cdots + \frac{1}{1}$$

> **Problem 26 (Paul Erdos)** — A set S of distinct integers is called sum-free if there does not exist a triple (x, y, z) of not necessarily distinct integers in S such that $x + y = z$. Show that for any set X of distinct integers, X has a sum free subset Y such that
> $$|Y| > \frac{|X|}{3}$$

Solution 26. Let $p = 3k+2$ be a prime so that $p > 2\max_{m \in X} m$. Note that if a set Y is sum free in the usual sense, it is equivalent for it to be sum free modulo p (since the prime is large).

Notice that $S_a = \{a(k+1), a(k+2), \ldots, a(2k+1)\}$ is sum free modulo p. Consider $Y = X_a \cap S$. We choose a uniformly at random from $\{1, 2, \ldots, 3k+1\}$ and find the expected value of the size of S'.

Notice that the probability a certain element $s \in X$ is also in S_a is $\frac{k+1}{3k+1}$ since $s \cdot a^{-1}$ covers all non zero values modulo p with equal probability. Thus the expected value of Y is

$$\mathbb{E}[|Y|] = \frac{k+1}{3k+1}|X| > \frac{|X|}{3}$$

> **Problem 27** (Erdos 1959) — For all k, l there exists a graph G with each cycle length at least k and chromatic number at least l.

Solution 27. Let a graph have n vertices. Each edge is chosen to be in the graph with probability p uniformly at random. There are two parts to this proof

- Bounding probability of a large number of short cycles.

  Let X denote the number of cycles with size $\leq k$ first.

  $$\mathbb{E}[X] \leq \sum_{3}^{l} n^i p^i$$

  We now pick $p = n^{\frac{1}{2l}-1}$ to end up with $\mathbb{E}[X] \leq c\sqrt{n}$. In particular by Markov's inequality, we have that the probability

  $$\mathbb{P}[X \geq \frac{n}{2}] \leq \frac{k}{\sqrt{n}} < 0.5$$

  for large n.

- Bounding the probability of high independent set size.

  Let the size of the largest independent set be $\alpha(G)$ and x be a value to be fixed later.

  $$\mathbb{P}[\alpha(G) \geq x] \leq \binom{n}{x}(1-p)^{\binom{x}{2}} \leq \left(ne^{-p(x-1)/2}\right)^x$$

  Now by setting $x = \frac{k \ln n}{p}$ for some constant k, we can make the probability above less than half once again for large n.

To finish, note that we can now get a graph with low $\alpha(G)$ and at most $n/2$ small cycles. Delete one vertex for each of those cycles. This does not increase the value of $\alpha(G)$. Finally, the chromatic number l is bounded as

$$\text{Chromatic number} \geq \frac{n/2}{\alpha(G)} \geq \frac{cpn}{\ln n} = \frac{cn^{1/2l}}{\ln n}$$

Since the expression given is unbounded, we can make n large enough so that the chromatic number is at least l.

$f(a,b) = f(a+b, a+b)$

$\Sigma x \leq \Sigma y$ constant

$\Sigma x = \Sigma y$

$f(a,b) = f(a+b$

290

9

2^{2008}

S S S S