MTH309: Counting and Graph Theory

draft

Coral Gables, Sep 2024

Abstract

Some notes for MTH309

Contents

0 Functions, inductions, and algorithms

You are probably all familiar with the infinite set of natural numbers (also known as "nonnegative integers")

$$\mathbb{N} = \{0, 1, 2, 3, \ldots, n, n+1, \ldots\}$$

It is usually stated in textbooks that natural numbers "come from Nature". This is not entirely true: To accept them, three important abstraction steps are necessary. These steps are non-trivial, as throughout the history of mankind, not all populations have accepted them:

- the notion of *cardinality*, i.e. the realization that two finite sets in bijection with one another have something in common; whence the *names* of numbers. This is not universal: Even today, the Pirahã people in Amazonas, Brazil, have no names for numbers, and have no linguistic way of expressing exact quantity, not even "one". [1]
- the notion of *zero*, as the cardinality of an "empty set". The ancient Greeks had no symbol for zero, for example; Mayas did have a symbol for zero around the year 36 BC, using it as placeholder in their base-20 numerical system; arithmetic operations with zero were first introduced by the Indian mathematician Brahmagupta[2], around 650 AD.
- the existence of an *infinite set*, that is, a set that can be in bijection with a proper subset of itself. The bijection in this case is the *successor* map, that is, the map that adds one to each element; so an equivalent way of formulating this principle is, "the belief that every number has a successor". Once again, this intuition is not universal, and in logic there is a movement of logicians from around 1900, called *(strict) finitists*, who rejected it.

Given natural numbers a and b, we say that "*a divides b*" (or that "*a is a divisor of b*", or equivalently that "*b is a multiple of a*") if there exists a natural number k such that

$$b = k \cdot a.$$

Definition 1 (Prime numbers)**.** Prime numbers are the natural numbers with exactly two distinct divisors (so 1 is not prime!):

$$2, 3, 5, 7, 11, 13, \ldots$$

Given two natural numbers a and b, their *greatest common divisor*, denoted by $\gcd(a, b)$ is the largest integer that divides both a and b.

Example 2. If p is prime, then it has only 1 and p as divisors. Thus for any natural number n

$$\gcd(p, n) = \begin{cases} p & \text{if } p \text{ divides } n, \\ 1 & \text{otherwise.} \end{cases}$$

0.1 Functions

Let X, Y be any two sets. Recall that their Cartesian product is defined by

$$X \times Y \stackrel{\text{def}}{=} \{(x, y) \text{ such that } x \in X, y \in Y\}.$$

Definition 3. A *function* $f : X \longrightarrow Y$ consists of two sets X, Y and a subset $F \subseteq X \times Y$, such that for each element $x \in X$ there is always exactly one element y of Y for which $(x, y) \in F$. Usually we denote this y by $f(x)$, and we say it is the *image of x (under f)*. We also call X (resp. Y) the *domain* (resp. the *codomain*) of the function. The *image of the set X* is the set $\operatorname{Im} X \stackrel{\text{def}}{=} \{f(x) \text{ such that } x \in X\}$.

[1] Frank et al., *Number as a cognitive technology: Evidence from Pirahã language and cognition*, Cognition 108 (2008), 819–824.

[2] Wallin, Nils-Bertil,"The History of Zero". YaleGlobal online, 19 November 2002

Functions are typically described by a formula that tells us how to find $f(x)$ given x. For example, given any set X, the *identity function on* X (usually denoted by id_X) is the function whose output is always identical to the input. In this case, our notation to "explain the function" is

$$\begin{aligned} id_X : X &\longrightarrow X \\ x &\longmapsto x. \end{aligned}$$

Sometimes one does not have an explicit formula, but there is still a clear general method to associate x with its image: for example,

$$\begin{aligned} f : \mathbb{N} &\longrightarrow \mathbb{N} \\ x &\longmapsto \text{the } x\text{-th prime number.} \end{aligned}$$

In the worst case scenario, if we do not see a general pattern, we can always express f by specifying all its values:

$$\begin{aligned} f : \{0,1,2\} &\longrightarrow \{0,1,2\} \\ 0 &\longmapsto 1 \\ 1 &\longmapsto 0 \\ 2 &\longmapsto 2. \end{aligned}$$

Definition 4. A function $f : X \to Y$ is *injective* if for each $x \neq x'$ one has $f(x) \neq f(x')$.

We assume familiarity with logic and quantifiers ($\forall$, $\exists$) and logical equivalence (contrapositives, etc.) For example, it should be clear that an equivalent way to define injectivity is:

$$\forall x, x' \in X, \quad f(x) = f(x') \Rightarrow x = x'.$$

Injectivity depends not only on the "formula", but also on the domain involved. For example, the function "first letter of" is injective on the set { Alba, Bruno }, but not injective on the set { Alba, Alice, Bruno }.

Definition 5. A function $f : X \to Y$ is *surjective* if the image of X coincides with the codomain Y; or in other words, if for each $y \in Y$ there is some $x \in X$ (not necessarily unique) such that $y = f(x)$.

Surjectivity depends not only on the "formula" for f, but also on the domain and the codomain. For example, let E be the set of even natural numbers:

$$\begin{aligned} f : \mathbb{N} &\longrightarrow \mathbb{N} \\ x &\longmapsto 2x \end{aligned}$$ is not surjective, $$\begin{aligned} f : \mathbb{N} &\longrightarrow E \\ x &\longmapsto 2x \end{aligned}$$ is, $$\begin{aligned} f : \mathbb{N} \setminus \{0\} &\longrightarrow E \\ x &\longmapsto 2x \end{aligned}$$ is not.

Definition 6. A function $f : X \to Y$ is *bijective* if it is both injective and surjective. That is, if for each $y \in Y$ there exists exactly one $x \in X$ such that $y = f(x)$.

Given a function $f : X \to Y$ and a function $g : Y \to Z$, their *composite* is the function

$$\begin{aligned} g \circ f : \quad X &\longrightarrow Z \\ x &\longmapsto g(\, f(x)\,). \end{aligned}$$

Proposition 7. *Let $f : X \to Y$ be a function between two non-empty sets.*
(1) *f is surjective $\iff$ there exists $g : Y \to X$ (called "right inverse") such that $f \circ g = id_Y$.*
(2) *f is injective $\iff$ there exists $g : Y \to X$ (called "left inverse") such that $g \circ f = id_X$.*
(3) *f is bijective $\iff$ there exists $g : Y \to X$ (called "inverse") such that $g \circ f = id_X$ and $f \circ g = id_Y$.*

Remark 8. Before starting with the proof, note that two functions are equal when they have same domain, same codomain, and they yield same outputs when given same inputs. So to verify an equality of functions like $g \circ f = id_Y$, both going from Y to Y, we'll need to check that $g \circ f(y) = id_Y(y)$ for all $y \in Y$.

Proof of Proposition 7.

(1), "⇒". Define

$$\begin{aligned} g : Y &\longrightarrow X \\ y &\longmapsto \text{some } x \text{ such that } f(x) = y. \end{aligned}$$

(If there is more than one x such that $f(x) = y$, we simply choose one.) Then by construction, $f \circ g(y) = f(x) = y$ for all $y \in Y$. Hence $f \circ g = id_Y$.

(1), "⇐". For each $y \in Y$, we know that $id_Y(y) = f \circ g(y)$, so $y = f(g(y))$, which means $y \in \operatorname{Im} f$.

(2), "⇒". Choose a point x_0 of X. Define

$$\begin{aligned} g : Y &\longrightarrow X \\ y &\longmapsto \begin{cases} x_0, & \text{if } y \notin \operatorname{Im} f \\ \text{the unique } x \text{ such that } f(x) = y, & \text{if } y \in \operatorname{Im} f. \end{cases} \end{aligned}$$

Then for all x in X, $g \circ f(x) = g(f(x)) = x$. So $g \circ f = id_X$.

(2), "⇐". Suppose $f(x) = f(x')$. Applying g, and remembering that $g \circ f = id_X$, we get

$$x = id_X(x) = g \circ f(x) = g(f(x)) = g(f(x')) = g \circ f(x') = id_X(x') = x'.$$

(3), "⇒". This does not follow immediately from items (1) and (2), because a priori it could be that the two g's (right inverse and left inverse) are different. However, if f is bijective we can simply define

$$\begin{aligned} g : Y &\longrightarrow X \\ y &\longmapsto \text{the unique } x \text{ such that } f(x) = y. \end{aligned}$$

and it is easy to see that it does the trick.

(3), "⇐". This follows from (1) and (2). (Why?) □

0.2 Induction and factorization

Induction is a standard technique to prove a statement for infinite subsets of $\mathbb{N}$. It is based on the fact that every natural number is obtained from 0 by adding 1 sufficiently many times. So if we show that a property P holds for zero and is maintained when we move from any number to its successor, then P holds also for one, for two, for three.... and eventually is shared by all natural numbers.

Formally, a proof by induction consists of two parts:

- ("Basis") We prove that the statement holds true for a specific integer n_0.
- ("Step") We prove that, if there exists a natural number n such that the statement holds for n, then the statement must hold also for $n + 1$.

Once again, the validity of the statement for n_0 implies the validity for $n_0 + 1$, which in turn implies the validity for $n_0 + 2$, and so on: A domino effect, which eventually proves the statement for all integers $n \geq n_0$. If our basis was $n_0 = 0$, then we end up proving the statement for the whole of $\mathbb{N}$.

Example 9. Let us prove by induction that

$$\sum_{i=0}^{n+1} i = \binom{n+2}{2} \text{ for all } n \in \mathbb{N}.$$

- ("Basis") For $n = 0$, the formula above boils down to $0 + 1 = \binom{2}{2}$, which is correct. This brings good luck!
- ("Step") Let us assume that $\sum_{i=0}^{n+1} i = \binom{n+2}{2}$ holds true for some n. Then

$$\sum_{i=0}^{n+2} i = (n+2) + \sum_{i=0}^{n+1} i \overset{!}{=} (n+2) + \binom{n+2}{2} = \binom{n+2}{1} + \binom{n+2}{2} = \binom{n+3}{2}.$$

Non-Example 10. Here is a "fake proof" by induction that

$$\sum_{i=0}^{n+1} i = \frac{(2n+3)^2}{8} \text{ for all positive integers } n.$$

Let us assume that $\sum_{i=0}^{n+1} i = \frac{(2n+3)^2}{8}$ holds true for some n. Then

$$\sum_{i=0}^{n+2} i = (n+2) + \sum_{i=0}^{n+1} i \overset{!}{=} (n+2) + \frac{(2n+3)^2}{8} = \frac{4n^2+20n+25}{8} = \frac{(2n+5)^2}{8} = \frac{(2(n+2)+1)^2}{8}.$$

What did we do wrong? Induction consists of *two* parts, a step and a basis... The basis is not superfluous! We need a domino tile where our domino effect can start.

Remark 11. A common mistake is to memorize the induction step as follows, "Let us assume that the statement holds for all n; then let us prove it for $n + 1$". This makes no sense: If we already know that the statement holds for all n, then we are done!, no need to think about $n+1$. That's not how induction works. What instead we are assuming is much less, namely, that the statement holds for *one* specific n; from there we want to be able to say the same thing about its successor, $n + 1$.

We should pay special attention to whether our induction step is imposing extra conditions on n. If the induction step works only for $n \geq n_0$, then the basis for the induction should be its verification at n_0, and not at 0.

Example 12. Let us prove that $n^2 - 5n + 6 \geq 0$ for all integers n. Let's assume it for n; then

$$(n+1)^2 - 5(n+1) + 6 = (n^2 + 2n + 1) - 5n - 5 + 6 = (n^2 - 5n + 6) + 2n - 4 \geq 2n - 4.$$

Now to conclude we would like to say that $2n - 4 \geq 0$. But this is true only when $n \geq 2$. So we are not done yet; it is incorrect for us to "make the domino tiling start" at $n = 0$ because as far as we know, the validity at 0 might not imply the validity at 1. So we proceed as follows:

- First, we ask ourselves whether the claim holds true for $n = 2$, which is the correct induction basis. Since $2^2 - 10 + 6 = 0$, the answer is "yes". Together with the induction step, this does prove

$$\text{"}n^2 - 5n + 6 \geq 0 \text{ for all integers } n \geq 2\text{."}$$

- This is not what we were asked to do, but almost: we are left with only finitely many cases to consider!, namely, $n = 0$ and $n = 1$. We can check them by hand: for $n = 0$ we have $0 - 0 + 6 > 0$, for $n = 1$ we have $1^2 - 5 + 6 = 0$. This concludes the proof.

Non-Example 13. Here is a "fake proof" (by induction on the number n of students) that

"all students will receive the same grade in the final".

For $n = 1$, we are considering a class consisting of only one student, so the claim is clear. Now suppose we have proved the claim for classes with n students. Let C be any class with $n + 1$ students. Let x, y be any two students enrolled in the class, and let z be any other student. Consider $S \setminus \{x\}$: this is a set of n students, so we can apply the inductive assumption: Everybody in $S \setminus \{x\}$ is going to get the same grade. In particular, y and z will get the same grade. Analogously, by the inductive assumption everybody in $S \setminus \{y\}$ will get the same grade, so in particular x and z. Summing up, x, y, z will all get the same grade. But then *any* two students x, y will get the same grade. What is wrong here?

(Hint: If in a proof you pick three different elements from a set, then you are implicitly assuming that the set has at least three elements. Note that if we have a class of 3 students, and any pair of them gets the same grade, then indeed they all get the same grade...)

There is a variant of induction which sometimes is easier to use than the one above. It is called **strong induction**, or sometimes "complete' or "generalized induction". Essentially, it is just normal induction plus "bookkeeping", i.e. keeping track of everything you have proven before. It consists of two parts:

- ("Basis") Prove a statement for a specific integer n_0.
- ("Step") Prove that, if there is a natural number n such that the statement holds for *every* natural number k such that $n_0 \leq k \leq n$, then the statement must hold also for $n + 1$.

Let us use the shortening $P(n)$for "the property holds for n". It is easy to see how generalized induction works: First of all, $P(n_0)$ implies $P(n_0 + 1)$. Now that we know $P(n_0)$ and $P(n_0 + 1)$, we can infer $P(n_0 + 2)$. But then we know $P(n_0)$ $P(n_0 + 1)$, and $P(n_0 + 2)$, which together imply $P(n_0 + 3)$. And so on: Another domino effect, which results in a proof of the statement for all integers $n \geq n_0$. The difference with classical induction is that instead of proving

$$P(n) \Rightarrow P(n + 1),$$

where $P(n)$ stands for "the property holds for n", we keep track at each step of what we have proven already and show

$$[P(n) \text{ and } P(n-1) \text{ and } P(n-2) \ldots \text{ and } P_(n_0 + 1) \text{ and } P_{n_0}] \Rightarrow P(n + 1).$$

Here is an example in which we need to assume not just $P(n)$, but also $P(n - 1)$:

Definition 14 (Fibonacci numbers)**.** The *n-th Fibonacci number* F_n is defined inductively as follows: $F_0 = 0$, $F_1 = 0$, and $F_n = F_{n-1} + F_{n-2}$ for all $n \geq 2$.

Proposition 15. $F_0 + F_1 + \ldots + F_n = F_{n+2} - 1$.

Proof. Indeed for $n = 0$ we have $F_0 = 0 = F_2 - 1$. Suppose now the statement has already been proven for all integers up to n. In particular, not just for n, but also for $n - 1$Then

$$F_0 + F_1 + \ldots + F_n + F_{n+1} = (F_{n+2} - 1) + F_{n+1} = F_{n+3} - 1. \qquad \square$$

Proposition 16. *Any integer $n \geq 2$ can be "factored", i.e. written as product of primes.*

Proof. The statement is true for $n = 2$, because "$2 = 2$" writes 2 as products of primes. Now let n be an integer, and suppose the claim has already been proven for any integer in $\{2, 3, \ldots, n-1\}$. If n is prime, then

$$n = n$$

is a valid way to write down n as a product of primes, and we are done. If n is composite, then

$$n = n_1 \cdot n_2,$$

with $2 \leq n_i < n$ (for $i = 1, 2$). By strong induction, both n_i can be written as product of primes. But then so can n. □

0.3 The fundamental theorem of arithmetics

Here are a few famous results by Euclid, all derived using induction.

Theorem 17 (Euclid)**.** *There are infinitely many primes.*

Proof. Let $p_1, \ldots, p_r$ be the complete list of the first r primes. Consider the number

$$n \stackrel{\text{def}}{=} 1 + p_1 p_2 \cdots p_r$$

and let p be any prime factor of n; the existence of one such p is guaranteed by Proposition 16. Were p belonging to $\{p_1, \ldots, p_r\}$, then

$$1 \;=\; p_1 \cdots p_r - n$$

would be a difference of two multiples of p; and thus 1 itself would be a multiple of p, a contradiction. So p does not belong to $\{p_1, \ldots, p_r\}$, and in particular, p is larger than all of $p_1, \ldots, p_r$. This already suffices to prove that primes are infinitely many, because for any r, given a complete list of the first r primes, we showed how to produce yet another prime. □

Remark 18. In some textbooks, the theorem above is often misquoted as follows: *given the set of the first k primes, their product plus one is a larger prime.* This argument is incorrect:

$$1 + (2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13) = 30031$$

is divisible by 59.

Lemma 19 (Euclid)**.** *Let a and b be natural numbers. If a prime number p divides ab, it divides either a or b.*

Proof. [3] We proceed by strong induction on the minimum of the pair $\{a, b\}$. Up to relabeling, we can assume $a \leq b$, so that a is the smallest of the pair. If $a = 0$, or $a = 1$, then the claim is clear. So we assume $a \geq 2$ and distinguish two cases: either a is prime, or not.

- Suppose a is prime. Let p be a prime that does not divide a but divides ab for some b. Using the classical division of integers with remainder, write $p = aq + r$ with $0 \leq r < a$. Since $ab = pc$ for some integer c, we have that

  $$ab = pc = (aq + r)c = acq + rc.$$

  This implies that $rc = a(b - cq)$, so the prime a divides rc. Since $r < a$, by strong induction the theorem holds for the pair $\{r, c\}$; that is, when a prime divides rc, it divides either r or c. But the prime a divides rc and does not divide r, because $r < a$. Hence, a divides c. Writing $c = ad$ for some integer d, we get

  $$ab = pc = pad,$$

  and canceling a we get $b = pd$. So p divides b.

[3] Proof due to Barry Cipra, math.stackexchange.com/questions/1581173/proof-of-euclids-lemma, 2015

- Suppose a is not prime. Then $a = d_1 \cdot d_2$, with both $d_1, d_2 < a$. Let p be a prime that does not divide a, and divides ab for some b. Then p divides neither d_1 nor d_2 (or else it would divide a), but

$$p \text{ divides } d_1 d_2\, b.$$

By strong induction (since $d_1 < a$) the statement of the theorem holds for the pair $\{d_1, d_2 b\}$: so either p divides d_1 (which is false), or p divides $d_2 b$. Hence,

$$p \text{ divides } d_2\, b.$$

By strong induction (since $d_2 < a$) the statement of the theorem holds for the pair $\{d_2, b\}$: so either p divides d_2 (which is false), or p divides $d_2 b$. Hence, p divides b. □

Lemma 20. *Let $a_1, \ldots, a_n$ be natural numbers. If a prime number p divides their product, then p divides (at least) one of $\{a_1, \ldots, a_n\}$.*

Proof. The case $n = 2$ is Lemma 19. By induction, suppose p divides $a_1 \cdot \ldots \cdot a_n \cdot a_{n+1}$. Call $b \stackrel{\text{def}}{=} a_1 \cdot \ldots \cdot a_n$. Since p divides $b \cdot a_{n+1}$, by Lemma 19 either p divides a_{n+1}, or p divides b, in which case by inductive assumption p divides one of $\{a_1, \ldots, a_n\}$. □

Theorem 21 (Euclid)**.** *Any integer $n \geq 2$ has a **unique** factorization into weakly-increasing primes.*

Proof. By Proposition 16, n has a factorization into primes; after possibly reordering such primes, we get a factorization into weakly-increasing primes. We want to show uniqueness. So, suppose there are natural numbers r, s for which

$$p_1 \cdots p_r = q_1 \cdots q_s,$$

where $p_1 \leq p_2 \leq \ldots \leq p_r$ are prime numbers, and $q_1 \leq q_2 \leq \ldots \leq q_s$ are also prime numbers.

Since p_1 divides the product of the q_j's, by Lemma 20 it must divide at least one of the q_j's. Because they are both primes, this actually means that p_1 is *equal* to one of the q_j's. Since q_1 is the smallest of the q_j's, this means that

$$p_1 \geq q_1.$$

Symmetrically, q_1 divides the product of the p_i's, so it must divide one of them by Lemma 20. By primality, q_1 is equal to one of the p_i's, so in particular

$$p_1 \leq q_1.$$

Now we cancel p_1 and q_1, and proceed recursively. Because

$$p_2 \cdot \ \ldots \ \cdot p_r = q_2 \cdot \ \ldots \ \cdot q_s,$$

we get that $p_2 = q_2$; and so on. It follows that $r = s$ and $p_i = q_i$ for each i. □

0.4 Algorithms

There is a way of phrasing inductive proofs in terms of *algorithms*, which are "procedures to solve a mathematical problem that are guaranteed to *terminate*, i.e. to succeed in a finite number of steps". The proof above for example can be viewed as an algorithm: Suppose you wish to factor n. Initialize $p = 2$ and ask yourself: Does p divide n? If yes, we can write $n = pn'$ and move on to factor n'. Note that factoring n' is an easier task (in the proofs by induction, in fact, it is a task that you assume already solved.) If instead p does not divide n, increment p by 1, and go back to the step where you asked yourself if p divides n.

Algorithm 22 (Factorization by trial division)**.** INPUT: an integer $n > 1$.
Initialize $A \stackrel{\text{def}}{=} [\quad]$ (the empty list), $f \stackrel{\text{def}}{=} 2$.
while $n > 1$:
if f divides n, append f to the list A, and replace n with $\frac{n}{f}$.
else: replace f with $f + 1$.
return A.

The one above is a valid algorithm, because the "while" loop cannot go on forever: At each iteration, either n gets smaller (and the while loop gets blocked as soon as n goes below 2) or f increases; in this second case, the worst case scenario is when n is a prime number, and f must be incremented $n - 2$ times until it reaches n.

Here is another, more famous example:

Theorem 23 (Euclidean Algorithm)**.** *Let n, m be positive integers. Let $n = qm + r$, with $0 \leq r < m$ (the classical division with remainder). Then*

$$\gcd(n, m) = \gcd(m, r).$$

Proof. Set $d_1 \stackrel{\text{def}}{=} \gcd(n, m)$ and $d_2 \stackrel{\text{def}}{=} \gcd(m, r)$. Since d_1 divides both n and m, it divides also $r = n - qm$; so it's a common divisor of m and r. So $d_1 \leq d_2$. On the other hand, d_2 divides m and r, so it divides also $n = qm + r$. So it's a common divisor of n and m. So $d_2 \leq d_1$. □

This result allows you to compute the greatest common divisor of two numbers.

Algorithm 24. INPUT: a, b positive integers, with $b < a$.
def $\gcd(a, b)$:
Do the Euclidean division $a = qb + r$.
if $r = 0$:
return b.
else:
return $\gcd(b, r)$.

1 Counting

1.1 Additive and multiplicative principles

See the book, Chapter 1.1. At the level of sets, if $|A|$ denotes the size of a set A, we have the two following, intuitive facts:

- (*additive principle*) If A and B are disjoint, $|A \cup B| = |A| + |B|$.
- (*multiplicative principle*) $|A \times B| = |A| \cdot |B|$.

Note: If A and B are not disjoint, one has $|A \cup B| \leq |A| + |B|$. In fact, the correct formula, called *principle of inclusion-exclusion*, is the following:

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

There is also a formula for n sets instead of two, but it's more complicated and we will see it later. See Chapter 1.6 of the book, or the wikipedia voice for "Principle of inclusion/exclusion".

1.2 Binomials

We define the factorial function inductively, as

$$0! \stackrel{\text{def}}{=} 1, \qquad \text{and} \qquad n! \stackrel{\text{def}}{=} n \cdot (n-1)!$$

Definition 25. Let $0 \leq k \neq n$ be natural numbers. Define

$$\binom{n}{k} \stackrel{\text{def}}{=} \frac{n!}{k! \cdot (n-k)!}$$

Remark 26. It is obvious from the definition that for all $0 \leq k \leq n$,

$$\binom{n}{k} = \binom{n}{n-k}.$$

Also, it is obvious that $\binom{n}{0} = \binom{n}{n} = 1$ for all n. It is instead *not* obvious from the definition whether this quantity $\binom{n}{k}$ is always an integer: We will prove it below.

Lemma 27 (Recurrence relation for binomial coefficients)**.** *For all $1 \leq k \leq n$,*

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}.$$

Proof. Let us compute the right hand side above:

$$\frac{(n-1)!}{k!(n-1-k)!} + \frac{(n-1)!}{(k-1)!(n-k)!} = \frac{(n-k)(n-1)!}{k!(n-k)!} + \frac{k(n-1)!}{k!(n-k)!} = \frac{n(n-1)!}{k!(n-k)!} = \binom{n}{k}. \qquad \square$$

Lemma 28 (Counting interpretation)**.** *For all $0 \leq k \leq n$, $\binom{n}{k}$ is a natural number: It counts the number of size-k subsets of $[n] \stackrel{\text{def}}{=} \{1, \ldots, n\}$.*

Proof. Let $B(n,k)$ be the number of size-k subsets of $[n] \stackrel{\text{def}}{=} \{1, \ldots, n\}$. It is clear that

$$B(0,0) = 1 = \binom{0}{0}. \tag{1}$$

Now if $1 \leq k \leq n$, a size k-subset S of $[n]$ can either contain the element n, or not. If it does, then $S \setminus \{n\}$ is a subset of $[n-1]$; and in fact, the function

$$\begin{array}{rcl} f : \{\text{size-}k\text{-subsets of } [n] \text{ containing } n\} & \longrightarrow & \{\text{size-}(k-1)\text{-subsets of } [n-1]\} \\ S & \longmapsto & S \setminus \{n\} \end{array}$$

is a bijection, its inverse being

$$\begin{array}{rcl} g : \{\text{size-}(k-1)\text{-subsets of } [n-1]\} & \longrightarrow & \{\text{size-}k\text{-subsets of } [n] \text{ containing } n\} \\ T & \longmapsto & T \cup \{n\}. \end{array}$$

Thus the number of size-k-subsets of $[n]$ containing n is just $B(n-1, k-1)$. But what is the number of size-k-subsets of $[n]$ **not** containing n? Well, this is simply $B(n-1,k)$, because any size-k subset of $[n]$ avoiding n is a subset of $[n-1]$. Thus we showed

$$B(n,k) = B(n-1,k) + B(n-1,k-1). \tag{2}$$

But then we can prove our claim by induction on n. The basis ($n = 0$) is given by Equation 1. Now we are ready to prove the inductive step, suppose that there exists an n such that, for all k such that $0 \leq k \leq n$, we have $B(n,k) = \binom{n}{k}$. Then for all k such that $1 \leq k \leq n$, we have

$$B(n+1,k) = B(n,k) + B(n,k-1) = \binom{n}{k} + \binom{n-1}{k-1} = \binom{n+1}{k}.$$

where the first equality is by Equation 2, the second is by inductive assumption, and the third is by Lemma 27. Note that the claim above is valid when $1 \leq k \leq n$, so there are still the two cases $k = 0$ and $k = n+1$ to consider; but it is clear from the definitions that for all n

$$B(n+1,0) = 1 = \binom{n+1}{0} \quad \text{and} \quad B(n+1,n+1) = 1 = \binom{n+1}{n+1}.$$

Thus for all k such that $0 \leq k \leq n+1$, we have $B(n+1,k) = \binom{n+1}{k}$, as desired. □

Theorem 29 (Newton's formula)**.** *For any natural number n,*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Proof. By induction. When $n = 0$, k is forced to be equal to 0, and the claim boils down to $1 = 1$. Now suppose the formula above holds for some n. Then

$$\begin{aligned}
(x+y)^{n+1} &= (x+y)(x+y)^n = \\
&= x(x+y)^n + y(x+y)^n = \\
&= \textstyle\sum_{k=0}^{n} \binom{n}{k} x^{k+1} y^{n-k} + \sum_{k=0}^{n} \binom{n}{k} x^k y^{n+1-k} = \\
&= \textstyle\sum_{h=1}^{n+1} \binom{n}{h-1} x^h y^{n-(k-1)} + \sum_{k=0}^{n} \binom{n}{k} x^k y^{n+1-k} = \\
&= x^{n+1} + \textstyle\sum_{i=1}^{n} \binom{n}{i-1} x^i y^{n+1-i} + \sum_{i=0}^{n} \binom{n}{i} x^i y^{n+1-i} = \\
&= x^{n+1} + \textstyle\sum_{i=1}^{n} \binom{n}{i-1} x^i y^{n+1-i} + \sum_{i=1}^{n} \binom{n}{i} x^i y^{n+1-i} + y^{n+1} = \\
&= x^{n+1} + \textstyle\sum_{i=1}^{n} \left[\binom{n}{i-1} + \binom{n}{i}\right] x^i y^{n+1-i} \quad + y^{n+1} = \\
&= x^{n+1} + \textstyle\sum_{i=1}^{n} \binom{n+1}{i} x^i y^{n+1-i} \quad + y^{n+1} = \\
&= \textstyle\sum_{i=0}^{n+1} \binom{n+1}{i} x^i y^{n+1-i}. \quad \square
\end{aligned}$$

Corollary 30. $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.

Proof. Plug in $x = y = 1$ into Newton's formula. □

Corollary 31. *For all $1 \leq k \leq n$, one has $\binom{n}{k} < 2^n$.*

Proof. $\binom{n}{k}$ plus other binomial coefficients, which are positive numbers, yields 2^n. □

Proposition 32. *Let $0 \leq k \leq n$ be integers. Then*

$$\binom{n}{k} < \binom{n}{k+1} \quad \textit{if and only if} \quad k < \frac{n-1}{2}.$$

Proof. It follows from the fact that $\binom{n}{k+1} = \frac{n-k}{k+1}\binom{n}{k}$ (exercise). □

There is another very interesting interpretation of binomial coefficients (see Section 1.5 of the book).

Theorem 33 ("Stars and bars")**.** *Let k, n be positive integers. The total number of nonnegative integer solutions of the equation $x_1 + \ldots + x_n = k$ is*

$$\binom{k+n-1}{k}.$$

'Woah' proof by the book. Imagine that we have to distribute a capital of k coins among n people. (The coins are not distinguishable from one another, in the sense that it does not matter whether person 1 gets the first or the third coin; but the people are distinct.) That's the same as placing exactly $n-1$ bars into a string of k 'stars' or if you want, coins: then the first person gets as many coins as to the left of the leftmost bar, and so on. That's the same as picking $n-1$ positions for the bars within a range of $k+n-1$ possible positions. That's the same as choosing a size-$(n-1)$ set out of a set of size $k+n-1$. This can be done in exactly $\binom{k+n-1}{n-1}$ ways. But $\binom{k+n-1}{n-1} = \binom{k+n-1}{k}$. □

Numeric proof by induction. Let $f(k,n)$ be the total number of nonnegative integer solutions of the equation $x_1 + \ldots + x_n = k$. For $k = n = 1$ indeed $f(k,n) = 1 = \binom{k+n-1}{k}$, as there is exactly one solution to $x_1 = 1$. So we proceed by induction. The nonnegative integer solutions of the equation $x_1 + \ldots + x_n + x_{n+1} = k$ are of two disjoint types: Those with $x_{n+1} = 0$, which are exactly $f(k,n)$ (corresponding to "distributing the k coins only among the first n people"); and those with $x_{n+1} \geq 1$, which are in bijection (by "giving one coin to the $(n+1)$-st guy already") to the set of nonnegative integer solutions of the equation $x_1 + \ldots + x_n + y_{n+1} = k-1$. Thus

$$f(k, n+1) = f(k,n) + f(k-1, n+1).$$

At the same time

$$\binom{k+n}{k} = \binom{k+n-1}{k} + \binom{k+n-1}{k-1},$$

because of the recursion of binomial coefficients. So $f(k,n) = \binom{k+n-1}{k}$ by induction. □

Remark 34. Suppose I want to know the total number of integer solutions of the equation $x_1 + x_2 + x_3 = 100$ with $x_1 \geq 2$, $x_2 \geq 5$, and $x_3 \geq 13$. Then I could already distribute 2 units to claimant 1, 5 units to claimant 2, and 3 units to claimant 3, and focus on the the total number of integer solutions of the equation $y_1 + y_2 + y_3 = 100 - (2+5+13) = 80$. So the answer is

$$\binom{82}{80} = \frac{82 \cdot 81}{2} = 41 \cdot 81.$$

Definition 35. A *lattice path* is a sequence of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in $\mathbb{Z}^2$ such that every consecutive difference $\mathbf{v}_i - \mathbf{v}_{i-1}$ is in $\{(1,0),(0,1)\}$. That is, we are only allowed to move right (=East) or upwards (=North).

Proposition 36. *If a, b are nonnegative integers, then from $(0,0)$ to (a,b) there are exactly $\binom{a+b}{b}$ lattice paths.*

Proof. Here is a numerical inductive proof. The point (a,b) can either be accessed from $(a,b-1)$ or from $(a-1,b)$, but not both. Thus if $f(a,b)$ counts the number of lattice paths from $(0,0)$ to (a,b), by partitioning those paths into those that get to $(a,b-1)$ and those that get to $(a-1,b)$, we obtain

$$f(a,b) = f(a,b-1) + f(a-1,b).$$

By induction on $n \stackrel{\text{def}}{=} a+b$ (the base case being $n=1$) we conclude that $f(a,b) = \binom{a+b}{b}$.

A bijective proof is obtained by noticing that any such path can be encoded with a string of Ns and Es, corrisponding to North or East moves. The string has length $a+b$ and is completely determined by the position of the a Ns. (Or, by the position of the b Es). Thus there are as many strings as there are a-element subsets of $[a+b]$ (or respectively, as there are b-element subsets of $[a+b]$). This leads to the count $\binom{a+b}{a}$ (or, equivalently, to $\binom{a+b}{b}$). □

Remark 37. There is a notion of *multinomial coefficient*, e.g. if $n = k_1 + \ldots + k_m$ we can define

$$\binom{n}{k_1, k_2, \ldots, k_m} \stackrel{\text{def}}{=} \frac{n!}{k_1! k_2! \cdots .k_m!}.$$

(When $m = 2$, this is the classical binomial coefficient.) There is also natural generalization to higher-dimensions of lattice paths: Let m be a positive integer. Let $i \in \{1, \ldots, m\}$. Let $\mathbf{e}_i$ be the vector of m integers with a 1 in position i, and zeroes elsewhere. For example, when $m = 3$, we have

$$\mathbf{e}_1 = (1, 0, 0), \quad \mathbf{e}_2 = (0, 1, 0), \quad \mathbf{e}_3 = (0, 0, 1).$$

An *m-dimensional lattice path* is a sequence of vectors in $\mathbb{Z}^m$ such that every consecutive difference $\mathbf{v}_i - \mathbf{v}_{i-1}$ belongs to the set $\{\mathbf{e}_1, \ldots, \mathbf{e}_m.\}$. It is easy to see that lattice paths on a grid $\mathbb{Z}^m$ from the origin to a point $(a_1, \ldots, a_m)$ are exactly $\binom{a_1+\ldots+a_m}{a_1, a_2, \ldots, a_m}$. The argument is the same as above: for simplicity, let us do it for $m = 3$. Let $f(a, b, c)$ count the number of lattice paths from The point (a, b, c) can only be reached through $(a-1, b, c)$ or through $(a, b-1, c)$ or through $(a, b, c-1)$. Thus

$$f(a, b, c) = f(a-1, b, c) + f(a, b-1, c) + f(a, b, c-1).$$

Now by induction on $n \stackrel{\text{def}}{=} a + b + c$, the basis being $n = 1$, we conclude $f(a, b, c) = \binom{a+b+c}{a,b,c}$.

Proposition 38 (Chu–Vandermonde's identity)**.** *For any non-negative integers r, m, n,*

$$\binom{m+n}{r} = \sum_{k=0}^{r} \binom{m}{k} \binom{n}{r-k}.$$

Proof. Here is a simple bijective proof: Given m men and n women, how can I select a committee of r people? The answer is $\binom{m+n}{r}$. But I could also pick k men in $\binom{m}{k}$ ways and then $r-k$ women, in $\binom{n}{r-k}$ ways. □

Remark 39. This proof also generalizes to p genders, as follows:

$$\binom{n_1 + n_2 + \ldots + n_p}{m} = \sum_{k_1 + \ldots + k_p = m} \binom{n_1}{k_1} \binom{n_2}{k_2} \cdots \binom{n+p}{k_p}.$$

The idea is, set $n = n_1 + \ldots + n_p$. To form a committee of m people out of n, we could equivalently select k_1 out of the n_1 of gender 1, then select k_2 out of the n_2 of gender 2, and so on.

Notation. When n is a positive integer, we denote by $[n]$ the set $\{1, 2, \ldots, n\}$.

Theorem 40. *The number of increasing functions $f : [k] \to [n]$ is*

$$\binom{n}{k}.$$

The number of non-decreasing functions $f : [k] \to [n]$ is

$$\binom{k+n-1}{k}.$$

Proof. For the first part: either by induction, or with the following bijective proof: we are basically selecting $\{f(1), \ldots, f(k)\}$, a size-k-subset of $[n]$. (For any such subset it is clear from the order who is $f(1)$, who is $f(2)$ and so on.) As for the second part: It can be proven by induction. Here is a bijective proof: any non-decreasing function from $[k]$ to $[n]$ is completely determined once we specify, for each integer $0 \leq m \leq n$, how many elements of $[k]$ are assigned value m. For example, consider a non-decreasing function f from [5] to [7]. If we know that zero elements are mapped to 1 or 2, two elements are mapped to 3, one element is mapped to 4, zero elements to 5 and 6, and two elements are mapped to 7, then f is the map

$$f(1) = f(2) = 3, \quad f(3) = 4, \quad f(4) = f(5) = 7.$$

So basically determining any such function f uniquely is like picking a non-negative integer solution for the equation $x_1 + \ldots + x_n = k$. □

Proposition 41. *If n is a nonnegative integer,*

$$\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}.$$

Proof. The ordinary proof is by induction on n. Here is a simple bijective proof from Wikipedia: suppose you have $2n$ objects arranged in a row, and you want to select n of them. There are $\binom{2n}{n}$ ways to do this. But also, you could select k objects from the first n and $n-k$ objects from the second n. Thus

$$\binom{2n}{n} = \sum_{k=0}^{n} \binom{n}{k}\binom{n}{n-k}.$$

Since $\binom{n}{n-k}$ is the same as $\binom{n}{n-k}$, we are done. □

Proposition 42. *For any non-negative integer n, the $(n+1)$-st Fibonacci number is*

$$F_{n+1} = \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n-k}{k}.$$

Proof. By induction (simple exercise). Alternatively, here is a relatively simple bijective proof: First show that F_n counts the number of ways to cover a $n \times 1$ strip of squares using only 2×1 and 1×1 tiles. But any such cover uses k tiles of the type 2×1 and $n-2k$ tiles of type 1×1, for some $k \leq \lfloor n/2 \rfloor$. The way to order k double tiles and $n-2k$ single tiles are $\binom{n-k}{k}$, corresponding to where you position the single tiles ("stars and bars"). □

Proposition 43 (Hockey-Stick identity)**.** *If $0 \leq r \leq n$ are integers,*

$$\sum_{m=r}^{n} \binom{m}{r} \stackrel{\text{def}}{=} \binom{r}{r} + \binom{r+1}{r} + \binom{r+2}{r} + \ldots + \binom{n}{r} = \binom{n+1}{r+1}.$$

Proof. By induction on n (simple exercise). For a bijective proof, see Wikipedia. □

Proposition 44. *If $0 \leq r \leq n$ are integers,*

$$\sum_{m=r}^{n} \binom{n}{m}\binom{m}{r} = 2^{n-r}\binom{n}{r}.$$

Proof. By induction on n (exercise). Bijective proof: Say we have n marbles. On the left we count the ways to first select a subset $S \subseteq [n]$ with at least r marbles, and then color in red r of the elements of S. Another way to get the same result is this: First we pick r marbles from $[n]$, and color them red; then we form a set S containing those r marbles, by independently deciding, for each of the other $n-r$ marbles, whether they belong to S or not. □

1.3 Permutations and k-permutations

Theorem 45. *The number of injective functions $f : [k] \to [n]$ is*

$$n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!} = \binom{n}{k} \cdot k!$$

Proof. There are n options for $f(1)$, then $n-1$ options for $f(2)$, and so on. □

Corollary 46. *The number of bijective functions $f : [n] \to [n]$ is*

$$n!$$

Proof. By the pigeonhole principle, any injective function $f : [n] \to [n]$ is automatically surjective, and thus a bijection. □

Sometimes the number $n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}$ is denoted by $P(n,k)$.

Remark 47. It is not a coincidence that

$$P(n,k) = \binom{n}{k} \cdot k!$$

In fact, $P(n,k)$ counts **the ways to arrange k objects chosen from $[n]$**. (Proof: an arrangement is precisely an injective function $f : [k] \to [n]$.) That's equivalent to first **choosing k objects from $[n]$**, and then given any such choice, **arrange them**, which can be done in $k!$ ways.

Example 48. In how many ways can I choose a committee of 3 from 10 people? The answer is $\binom{10}{3}$. But if the order is specified, i.e. "in how many ways can I choose a chair, a vicechair and a treasurer, out of 10 people") the answer becomes $\binom{10}{3}3!$, or equivalently, $10 \cdot 9 \cdot 8$.

The structure of permutations

Let n be any positive integer. Let $[n] \stackrel{\text{def}}{=} \{1, \ldots, n\}$. Let

$$\mathcal{S}_n \stackrel{\text{def}}{=} \{\sigma : [n] \longrightarrow [n] \text{ bijective}\}.$$

The elements of $\mathcal{S}_n$ are called *permutations*. There are three types of notation to write down the same permutation:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 3 & 5 & 6 & 4 \end{pmatrix}, \quad \sigma = (12)(456), \quad \text{and } \sigma = (12)(45)(56).$$

The second notation writes σ as product of disjoint cycles; the third, as product of non-disjoint swaps. We will explain them in a few minutes. The first notation is called *two-line notation* and it is the most intuitive: the rule is, σ maps each elements of the first row into the element of the second row immediately below. (In this case the first row is ordered, but it does not matter: What matters is that below each i sits $\sigma(i)$.) For example, $\sigma(3) = 3$. To compose two functions, we write them on top of one another, remembering that when we write $\tau \circ \sigma(1)$ the first permutation to be applied to 1 is σ, so σ should be on top. The two-line notation of $\tau \circ \sigma$ is then obtained by looking at only the first and the last row, ignoring all intermediate ones. For example, if

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 4 & 3 & 5 & 6 \end{pmatrix} \qquad \text{and} \qquad \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 3 & 5 & 6 & 4 \end{pmatrix}$$

then

$$\tau \circ \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 3 & 5 & 6 & 4 \\ 2 & 1 & 4 & 5 & 6 & 3. \end{pmatrix} \qquad \text{and} \qquad \sigma \circ \tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 4 & 3 & 5 & 6 \\ 2 & 1 & 5 & 3 & 6 & 4. \end{pmatrix}$$

En passant, notice that $\tau \circ \sigma$ and $\sigma \circ \tau$ are different, so $\circ$ is not commutative.

Definition 49. Let $k \leq n$ be positive integers. A *cycle (of length k)* in a permutation $\sigma \in \mathcal{S}_n$ is a k-tuple

$$(a_1, a_2, \ldots, a_k),$$

such that $\sigma(a_k) = a_1$ and $\sigma(a_i) = a_{i+1}$ for all $i \in \{1, \ldots, k-1\}$. Cycles of length two are called *swaps* (or *transpositions*).
Any cycle g (of length k) is naturally associated to a permutation $\gamma \in \mathcal{S}_n$, as follows: $\gamma = \sigma$ on the elements of the cycle, and $\gamma = id$ otherwise.

Example 50. In the permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 3 & 5 & 6 & 4 \end{pmatrix},$$

there is a cycle of length 3, namely, $g = (4, 5, 6)$. Its associated permutation is

$$\gamma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 3 & 5 & 6 & 4 \end{pmatrix}.$$

Theorem 51. *Every permutation different than the identity can be written as product of disjoint cycles.*

"Proof by example". We sketch the algorithm with the help of an example. Suppose

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 8 & 4 & 6 & 5 & 7 & 9 & 2 & 1 \end{pmatrix}.$$

To find the first cycle, we start with 1 and apply iteratively σ, until you get back to 1. In our case

$$\sigma(1) = 3, \quad \sigma(3) = 4, \quad \sigma(4) = 6, \quad \sigma(6) = 7, \quad \sigma(7) = 9, \quad \sigma(9) = 1.$$

So the first cycle is $(1, 3, 4, 6, 7, 9)$. Now let us consider the smallest integer not contained in this cycle, and start over again, applying iteratively σ until we get back to it. In our case, we re-start with 2:

$$\sigma(2) = 8, \quad \sigma(8) = 2.$$

So $(2, 8)$ is the second cycle. By construction, it is disjoint from the first cycle, because σ is injective. Now we continue inductively, until all numbers have been listed. In our example, the smallest integer that belongs to neither if the previous cycles is 5. Since $\sigma(5) = 5$, our third cycle has length one. Now all numbers have been listed. Our final result is

$$\sigma = (1, 3, 4, 6, 7, 9)(2, 8)(5). \qquad \square$$

Now, technically what we found is just a *list* of disjoint cycles. But if we interpreted every cycle as its associated permutation in $\mathcal{S}_n$, the list can actually be interpreted a product of permutations. More precisely, if $\gamma, \gamma_1, \gamma_2, \gamma_3$ are the permutations of $\mathcal{S}_n$ associated respectively to σ, to $(1, 3, 4, 6, 7, 9)$, to $(2, 8)$, and to (5), then it is clear that

$$\gamma = \gamma_1 \circ \gamma_2 \circ \gamma_3.$$

For this reason, we speak of "product of cycles".

Remark 52. By convention, one usually omits writing down the cycles of length one, because the associated permutation of $\mathcal{S}_n$ is the identity. So in the example above, we would simply write

$$\sigma = (1, 3, 4, 6, 7, 9)(2, 8).$$

Lemma 53. *Every cycle of length $k \geq 2$ can be written as product of $k-1$ non-disjoint swaps.*

Proof. We claim that

$$(a_1, a_2, \ldots, a_k) = (a_1, a_2)(a_2, a_3) \cdots (a_{k-2}, a_{k-1})(a_{k-1}, a_k).$$

By this we mean that if γ is the permutation of $\mathcal{S}_n$ associated to $(a_1, \ldots, a_k)$, and γ_i is the permutation of $\mathcal{S}_n$ associated to (a_i, a_{i+1}), then

$$\gamma = \gamma_1 \circ \gamma_2 \circ \ldots \circ \gamma_{k-1}.$$

As a warm up, let us check this first for the element a_1. By definition, $\gamma(a_1) = a_2$. On the other hand, γ_i swaps a_1 with a_{i+1}, so it has no effect on a_1 if $i \geq 2$. Formally,

$$\gamma_i(a_1) = \begin{cases} a_1 & \text{if } i \geq 2 \\ a_2 & \text{if } i = 1. \end{cases}$$

So

$$\gamma_1 \circ \gamma_2 \circ \ldots \circ \gamma_{k-1}(a_1) = \gamma_1 \circ \gamma_2 \circ \ldots \circ \gamma_{k-2}(a_1) = \ldots = \gamma_1(a_1) = a_2,$$

as desired. Now let us check the effect on the generic element a_j, with $j < k$. Clearly $\gamma(a_j) = a_{j+1}$, with the exception of a_k, for which $\gamma(a_k) = a_1$. On the other hand,

$$\gamma_i(a_j) = \begin{cases} a_i & \text{if } i \geq j+1 \\ a_{j+1} & \text{if } i = j \\ a_{j-1} & \text{if } i = j-1 \\ a_i & \text{if } i \leq j-2. \end{cases}$$

So if $j < k$, we have

$$\gamma_1 \circ \ldots \circ \gamma_{k-1}(a_j) = \ldots = \gamma_1 \circ \ldots \circ \gamma_j(a_j) = \gamma_1 \circ \ldots \circ \gamma_{j-1}(a_{j+1} = \ldots = \gamma_1(a_{j+1}) = a_{j+1}.$$

For a_k instead we have

$$\gamma_1 \circ \ldots \circ \gamma_{k-1}(a_k) = \gamma_1 \circ \ldots \circ \gamma_{k-2}(a_{k-1}) = \gamma_1 \circ \ldots \circ \gamma_{k-3}(a_{k-2}) = \ldots = \gamma_1(a_2) = a_1. \qquad \square$$

Corollary 54. *Every permutation can be written as product of non-disjoint swaps.*

Proof. In case the permutation is the identity, write it as $(12)(12)$. First write the permutation as product of disjoint cycles, and then break each cycle into swaps. $\square$

Example 55. Let us verify that $(1, 3, 4, 6, 7, 9) = (1, 3)(3, 4)(4, 6)(6, 7)(7, 9)$. In fact, the right hand side is given by the first and the last row of the matrix

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 1 & 2 & 3 & 4 & 5 & 6 & 9 & 8 & 7 \\ 1 & 2 & 3 & 4 & 5 & 7 & 9 & 8 & 6 \\ 1 & 2 & 3 & 6 & 5 & 7 & 9 & 8 & 4 \\ 1 & 2 & 4 & 6 & 5 & 7 & 9 & 8 & 3 \\ 3 & 2 & 4 & 6 & 5 & 7 & 9 & 8 & 1 \end{pmatrix}$$

and the left hand side is precisely

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 2 & 4 & 6 & 5 & 7 & 9 & 8 & 1 \end{pmatrix}$$

1.4 Inversion, even and odd permutations

Definition 56. An *inversion* in a permutation $\sigma \in \mathcal{S}_n$ is a pair of integers (i, j) such that $i < j$ but $\sigma(i) > \sigma(j)$. A permutation is called *even* if it contains an even number of inversions, and *odd* if it contains an odd number of inversions.

Example 57. For $n = 3$, the identity permutation

$$e = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$$

has zero inversions, so it is even. The swaps

$$\gamma_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \quad \gamma_2 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \quad \gamma_3 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$$

have one inversion each, so they are odd. Finally, the cycles of length 2

$$\sigma_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \quad \text{and} \quad \sigma_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

have two inversions each, so they are even.

Lemma 58. *Let γ be a swap. If σ is even then $\gamma \circ \sigma$ is odd, and if σ is odd, $\gamma \circ \sigma$ is even. (In other words, each additional swap changes the parity).*

Proof. Let $a < b$ be the two elements of $[n] = \{1, \ldots, n\}$ that are swapped by γ. Let $i < j$. We distinguish six cases:

1. if $\{\sigma(i), \sigma(j)\}$ and $\{a, b\}$ are disjoint, then γ has no effect on $\sigma(i)$ and $\sigma(j)$. So
$$(i, j) \text{ is an inversion in } \sigma \iff (i, j) \text{ is an inversion in } \gamma \circ \sigma.$$
2. if $\{\sigma(i), \sigma(j)\} = \{a, b\}$, then γ creates an inversion if there is none, or removes the inversion if there is one. Hence
$$(i, j) \text{ is an inversion in } \sigma \iff (i, j) \text{ is NOT an inversion in } \gamma \circ \sigma.$$
3. if $\sigma(i) \in \{a, b\}$ and $a < \sigma(j) < b$, then again γ creates an inversion if there is none (the case $\sigma(i) = a$), or removes the inversion if there is one (the case $\sigma(i) = b$). Hence
$$(i, j) \text{ is an inversion in } \sigma \iff (i, j) \text{ is NOT an inversion in } \gamma \circ \sigma.$$
4. if $\sigma(i) \in \{a, b\}$, and $\sigma(j) < a$ or $\sigma(j) > b$, then the swap given by γ does not affect the order. So
$$(i, j) \text{ is an inversion in } \sigma \iff (i, j) \text{ is an inversion in } \gamma \circ \sigma.$$
5. if $\sigma(j) \in \{a, b\}$ and $a < \sigma(i) < b$, symmetrically to case (3) we have
$$(i, j) \text{ is an inversion in } \sigma \iff (i, j) \text{ is NOT an inversion in } \gamma \circ \sigma.$$
6. if $\sigma(j) \in \{a, b\}$, and $\sigma(i) < a$ or $\sigma(i) > b$, then the swap given by γ does not affect the order, and symmetrically to case (4) e have
$$(i, j) \text{ is an inversion in } \sigma \iff (i, j) \text{ is an inversion in } \gamma \circ \sigma.$$

So if we compare the number of inversions of σ versus $\gamma \circ \sigma$, there is an odd number of changes, because cases (3) and (5) are equally frequent, and case (2) occurs exactly once. □

Theorem 59. *Let $n \geq 2$ be an integer. Let $\sigma \in \mathcal{S}_n$. The following are equivalent:*
(A) *σ is even;*
(B) *σ can be written down as product of an even number of swaps;*
(C) *σ cannot be written down as product of an odd number of swaps.*

Proof. The identity permutation is even, because it has no inversion. Now suppose we have a decomposition of σ into a product of s swaps. (One possible such decomposition is given by By Corollary 54.) This means that σ is obtained with s swaps from an even permutation. Lemma 58 tells us that each swap changes the parity. So σ is even if and only if s is even. □

Corollary 60. *A k-cycle is an even permutation if and only if k is odd.*

Proof. A k-cycle is the product of $k-1$ swaps, and k is even if and only if $k-1$ is odd. □

Corollary 61. *The set*

$$A_n \stackrel{\text{def}}{=} \{\textit{even permutations}\}$$

is a subset of $\mathcal{S}_n$ with $\frac{n!}{2}$ elements.

Proof. Consider the following function between sets

$$\begin{array}{rccc} \psi: & A_n & \longrightarrow & (\mathcal{S}_n \setminus A_n) \\ & \sigma & \longmapsto & \sigma \circ (1,2). \end{array}$$

This function is well-defined by Lemma 58. Moreover, it is invertible, the inverse being

$$\begin{array}{rccc} \phi: & (\mathcal{S}_n \setminus A_n) & \longrightarrow & A_n \\ & \tau & \longmapsto & \tau \circ (1,2). \end{array}$$

This proves that A_n and its complement within $\mathcal{S}_n$ have the same number of elements. Hence, A_n has $\frac{n!}{2}$ elements. □

Corollary 62. *For each $n \geq 3$, every permutation $\sigma \in A_n$ can be written as product of 3-cycles.*

Proof. By theorem 59 the permutation σ can be written as product of an even number $2s$ of swaps.The trick is to simply to pair them. We claim in fact that the product of any two swaps is either the identity or a 3-cycle. In fact,

- if $a = c$ and $b = d$, then $(a,b)(a,b) = id$;
- if $a = c$ but $b \neq d$, then $(a,b)(a,d) = (adb)$;
- if a, b, c, d are all distinct, then $(a,b)(c,d) = (ab)(ac)(ac)(c,d) = (abc)(cda)$. □

2 Sequences and their difference

Definition 63. A *(real-valued) sequence* in a set S is a function $a : \mathbb{N} \to \mathbb{R}$. We usually denote by a_n the element $a(n)$.

A sequence may be defined by a closed formula, like the sequence of cubes $q_n = n^3$, or inductively. For example, the Fibonacci sequence is defined by

$$F_0 = 0, \quad F_1 = 1, \qquad F_n = F_{n-1} + F_{n-2} \text{ for all } n \geq 2.$$

2.1 The k-th difference of a sequence

Definition 64. Let $a : \mathbb{N} \to \mathbb{R}$ be an arbitrary sequence. The *first difference of* a is the sequence

$$(\Delta a)_n \stackrel{\text{def}}{=} a_{n+1} - a_n.$$

For brevity, we simply write Δa_n instead of $(\Delta a)_n$.

Definition 65. Let $X = \{$ functions from $\mathbb{N}$ to $\mathbb{R}\}$. The *difference operator* is a function $\Delta : X \longrightarrow X$ defined by $\Delta(a) = \Delta a$. Recursively, we define, ffor all positive integers k,

$$\Delta^k \stackrel{\text{def}}{=} \Delta \circ \Delta^{k-1}.$$

Example 66. The sequence q of cubes is

$$0,\ 1,\ 8,\ 27,\ 64,\ 125,\ 216,\ 343,\ 512, \ldots$$

Its first difference Δq is

$$1,\ 7,\ 19,\ 37,\ 61,\ 91,\ 127,\ 169, \ldots$$

The second difference $\Delta^2 q$ is

$$6,\ 12,\ 18,\ 24,\ 30,\ 36,\ 42, \ldots$$

The third difference $\Delta^3 q$ is

$$6,\ 6,\ 6,\ 6,\ 6,\ 6, \ldots$$

and the forth difference is zero.

Lemma 67 ("Linearity of Δ^k")**.** *For all sequences a, b and for all real numbers r, one has*

$$\Delta(a + b) = \Delta a + \Delta b \quad \text{and } \Delta(r \cdot a) = r \cdot \Delta a.$$

The same is true replacing Δ with Δ^k, for any k.

Proof. Element-wise:

$$\begin{aligned} \Delta(a+b)_n &= (a+b)_{n+1} - (a+b)_n = (a_{n+1} + b_{n+1}) - (a_n + b_n) = \\ &= (a_{n+1} - a_n) + (b_{n+1} - b_n) = \Delta a_n + \Delta b_n, \end{aligned}$$

$$\Delta(r \cdot a)_n = (r \cdot a)_{n+1} - (r \cdot a)_n = r \cdot a_{n+1} - r \cdot a_n = r \cdot \Delta a_n.$$

For Δ^k the proof is left to you (it's a simple induction on k). □

Theorem 68. *If $a_n = n^k$ for some positive integer k, then $\Delta^k a$ is a constant nonzero sequence, equal to $k!$ for all n, and thus for all $h > k$ the sequence $\Delta^h a$ is the zero sequence.*

Proof. Set $f_k(n) = n^k$. We proceed by induction on k. The case $k = 1$ is clear:

$$\Delta^1 a \stackrel{\text{def}}{=} a_{n+1} - a_n = (n+1)^1 - n^1 = 1 = 1!$$

Now suppose we proved the statement for $k - 1$, and let's show it for k. By the inductive assumption, we know that

$$\Delta^{k-1}(f_i) = \begin{cases} (k-1)! & \text{if } i = k-1, \\ 0 & \text{if } i < k-1. \end{cases} \tag{3}$$

But then

$$\begin{aligned}\Delta^k a_n &= \Delta^{k-1}\Delta a_n = \Delta^{k-1}\left((n+1)^k - n^k\right) = \Delta^{k-1}\left(\sum_{i=0}^{k-1}\binom{k}{i}n^i\right)\\ &= \sum_{i=0}^{k-1}\binom{k}{i}\Delta^{k-1}(n^i) = \binom{k}{k-1}(k-1)! = k!,\end{aligned}$$

where the first step is by definition of Δ^k, the second step is by definition of a_n, the third step is by Newton's formula (plus the cancellation of the term n^k), the forth step is by the linearity of Δ, the fifth step is by Equation 3 above (only one summand survives!, namely, the summand corresponding to $i = k-1$), and the sixth step is a simple computation. □

Corollary 69. *If a_n coincides with a degree-k polynomial in the variable n, then $\Delta^k a$ is constant nonzero, and thus $\Delta^{k+1}a$ is zero.*

Proof. By the previous theorem plus the linearity of Δ. □

I will now argue that the converse is also true. First we need a fact connecting $\Delta^k a$ with the consecutive $k+1$ terms $a_n, a_{n+1}, \ldots, a_{n_k}$ of the sequence.

Theorem 70. *For any positive integer k,*

$$\Delta^k a_n = \sum_{t=0}^{k}(-1)^t\binom{k}{t}a_{n+k-t}.$$

Proof. By induction. The case $k=1$ is easy: it reads $\Delta a_n = a_{n+1} - a_n$. Suppose the statement holds for some k; let us prove it for $k+1$.

$$\begin{aligned}\Delta^{k+1}a_n = \Delta^k(\Delta a_n) &= \Delta^k a_{n+1} - \Delta^k a_n = \textstyle\sum_{t=0}^{k}(-1)^t\binom{k}{t}a_{n+1+k-t} - \sum_{t=0}^{k}(-1)^t\binom{k}{t}a_{n+k-t}\\ &= a_{n+1+k} + (-1)^k a_n + \textstyle\sum_{t=1}^{k}(-1)^t\binom{k}{t}a_{n+1+k-t} - \sum_{t=0}^{k-1}(-1)^t\binom{k}{t}a_{n+k-t}\\ &= a_{n+1+k} + (-1)^k a_n + \textstyle\sum_{t=1}^{k}(-1)^t\binom{k}{t}a_{n+1+k-t} + \sum_{t=0}^{k-1}(-1)^{t+1}\binom{k}{t}a_{n+k-t}\\ &= a_{n+1+k} + (-1)^k a_n + \textstyle\sum_{t=1}^{k}(-1)^t\binom{k}{t}a_{n+1+k-t} + \sum_{u=1}^{k}(-1)^u\binom{k}{u-1}a_{n+k-(u-1)}\\ &= a_{n+1+k} + (-1)^k a_n + \textstyle\sum_{t=1}^{k}(-1)^t\binom{k}{t}a_{n+1+k-t} + \sum_{t=1}^{k}(-1)^t\binom{k}{t-1}a_{n+1+k-t}\\ &= a_{n+1+k} + (-1)^k a_n + \textstyle\sum_{t=1}^{k}(-1)^t\left[\binom{k}{t} + \binom{k}{t-1}\right]a_{n+1+k-t}\\ &= a_{n+1+k} + (-1)^k a_n + \textstyle\sum_{t=1}^{k}(-1)^t\binom{k+1}{t}a_{n+1+k-t}\\ &= \textstyle\sum_{t=0}^{k+1}(-1)^t\binom{k+1}{t}a_{n+1+k-t}. \quad \square\end{aligned}$$

Remark 71. The previous formula can be inverted, though we omit the inductive proof because it's too complicated: One can prove that

$$a_{n+k} = \sum_{i=0}^{k}\binom{k}{i}\Delta^i a_n.$$

So fundamentally, understanding the first $k+1$ terms after a_n is the same as understanding subsequent differences. Understanding relations among the terms $a_n, \ldots, a_{n+k}$ (e.g. for $k=2$ the Fibonacci relation $F_{n+2} = F_{n+1} + F_n$), is the same as understanding "differential equations", i.e. equations involving the original sequence and its first, second, ..., k-th differences. For example, going back to the Fibonacci sequence: Since $\Delta F_n = F_{n+1} - F_n = F_{n-1}$ and $\Delta^2 F_n = \Delta \circ \Delta F_n = \Delta(F_{n-1}) = F_{n-2}$, the Fibonacci relation gives rise to the differential equation

$$F_n = \Delta F_n + \Delta^2 F_n.$$

For the next result, we need a classical lemma from calculus.

Lemma 72 (Polynomial interpolation)**.** *Given $k+1$ points*

$$(a_0, b_0),\ (a_1, b_1),\ \ldots,\ (a_k, b_k) \in \mathbb{R}^2$$

such that no two a_i's are the same, there exists a unique polynomial g with coefficients in $\mathbb{R}$ and degree at most k, such that

$$g(a_0) = b_0,\ g(a_1) = b_1,\ \ldots,\ g(a_k) = b_k.$$

Partial proof. We only show existence, which is all we need for our theorem below. (For more details, see e.g. Wikipedia, en.wikipedia.org/wiki/Polynomial_interpolation.)
For every $j \in \{1, \ldots, n\}$, let P_j be the product of all $(x - a_i)$'s, except for $(x - a_j)$. Being a product of k binomials of degree 1, this P_j is a polynomial of degree k; we can evaluate it a any number, so it makes sense to compute $P_j(a_j)$, by plugging in a_j for x. This $P_j(a_j)$ is a *nonzero* real number, because by assumption $a_i \neq a_j$ when $i \neq j$. So we can define

$$L_j(x) \stackrel{\text{def}}{=} \frac{P_j(x)}{P_j(a_j)} = \frac{(x-a_0)\cdot(x-a_1)\cdots(x-a_{j-1})(x-a_{j+1})\cdots(x-a_k)}{(a_j-a_0)\cdot(a_j-a_1)\cdots(a_j-a_{j-1})(a_j-a_{j+1})\cdots(a_j-a_k)}.$$

Since the denominator is a (nonzero!) real number, and the numerator is a degree-k polynomial, also L_j is a degree-k polynomial. Moreover, when we evaluate it at some a_i, we get that

$$L_j(a_i) = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases}$$

But then if we define

$$g(x) \stackrel{\text{def}}{=} b_0 \cdot L_0(x) + b_1 \cdot L_1(x) + \ldots + b_k \cdot L_k(x),$$

this is a sum of degree-k polynomials with the property that $g(a_i) = b_i$. A sum of degree-k polynomials has degree at most k. □

Theorem 73. *A sequence a is the restriction to the integers of a degree-k polynomial function $\iff$ $\Delta^k a$ is a nonzero constant.*

Proof. '$\Rightarrow$' is already done, it's Corollary 69.
'$\Leftarrow$': Let a be a sequence with $\Delta^k a$ constant but not zero. By polynomial interpolation, there exists a polynomial g with coefficients in $\mathbb{R}$ and degree $\leq k$ such that

$$g(0) = a_0,\ g(1) = a_1,\ \ldots,\ g(n) = a_n.$$

Now consider the sequence b defined by

$$b_n \stackrel{\text{def}}{=} g(n) - a_n.$$

We know that $0 = b_0 = b_1 = \ldots = b_k$, by definition of g. On the other hand, since $\Delta^k a$ is constant, $\Delta^{k+1} a$ is the zero sequence. $\Delta^{k+1} g(n)$ is also zero for all n, because of part '$\Rightarrow$' , which has already been proved. So by the linearity of Δ

$$\Delta^{k+1} b_n = \Delta^{k+1} g(n) - \Delta^{k+1} a_n = 0 - 0.$$

But then if we apply Theorem 70 to the sequence b, we get

$$0 = \Delta^{k+1} b_n = b_{n+k+1} + \sum_{t=1}^{k+1} (-1)^t \binom{k+1}{t} b_{n+1+k-t}.$$

But for all $t \geq 1$, obviously $1 + k - t \leq k$, so automatically $b_{1+k-t} = 0$. (Because it's one element of the list $b_0, \ldots, b_k$.)

- for $n = 0$ we get $0 = \Delta^{k+1} b_0 = b_{k+1} + \sum_{t=1}^{k+1} (-1)^t \binom{k+1}{t} 0$, so $b_{k+1} = 0$;
- for $n = 1$ we get $0 = \Delta^{k+1} b_1 = b_{k+2} + \sum_{t=1}^{k+1} (-1)^t \binom{k+1}{t} 0$, so $b_{k+2} = 0$;
- and so on.

Thus $b_n = 0$ for all n, which means that the polynomial function g restricted to the integers coincides with a_n. But then g cannot have degree $\leq k-1$, otherwise we would have $0 = \Delta^k g(n) = \Delta^k a_n$ for all n, contradicting the assumption. So the degree of g is exactly k. □

2.2 The sequence of partial sums

Like for derivatives and integrals, there is an operator on sequences that is somewhat the "inverse" of Δ.

Definition 74. Let $a : \mathbb{N} \to \mathbb{R}$ be an arbitrary sequence. The *sequence of partial sums of* a, or shortly the *sum of* a, is the sequence

$$(\Sigma a)_n \stackrel{\text{def}}{=} a_0 + \ldots + a_n.$$

For brevity, we simply write Σa_n instead of $(\Sigma a)_n$.

Definition 75. Let $X = \{$ functions from $\mathbb{N}$ to $\mathbb{R}\}$. The *sum operator* is a function $\Sigma : X \longrightarrow X$ defined by $\Sigma(a) = \Sigma a$. Recursively, we define, for all positive integers k,

$$\Sigma^k \stackrel{\text{def}}{=} \Sigma \circ \Sigma^{k-1}.$$

Example 76. The "initial-zero-then-constantly-one" sequence c is

$$0,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1, \ldots$$

Its sum Σc is the sequence of natural numbers

$$0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8, \ldots$$

The second sum $\Sigma^2 c$ is the sequence of *triangular numbers* $t_n = \binom{n+1}{2}$, namely,

$$0,\ 1,\ 3,\ 6,\ 10,\ 15,\ 21,\ 28,\ 36 \ldots$$

The third sum $\Sigma^3 c$ is the sequence of *tetrahedral numbers*

$$0,\ 1,\ 4,\ 10,\ 20,\ 35,\ 56,\ 84,\ 120 \ldots$$

A closed formula for $\Sigma^3 c$ is $\binom{n+2}{3}$ - easy prove by induction. Can you guess a closed formula for $\Sigma^4 c$? Can you prove it?

Lemma 77 ("Linearity of Σ^k")**.** *For all sequences a, b and for all real numbers r, one has*

$$\Sigma(a+b) = \Sigma a + \Sigma b \quad \text{and } \Sigma(r \cdot a) = r \cdot \Sigma a.$$

The same is true replacing Σ with Σ^k, for any k.

Proof. Element-wise; left to you. □

Remark 78. Note that for any natural number n

$$(\Delta(\Sigma(a))_n = \Sigma(a_{n+1}) - \Sigma(a_n) = (a_0 + \ldots + a_n + a_{n+1}) - (a_0 + \ldots + a_n) = a_{n+1},$$

while

$$(\Sigma(\Delta(a))_n = \Delta a_0 + \ldots + \Delta a_n = (a_1 - a_0) + (a_2 - a_1) + \ldots + (a_{n+1} - a_n) = a_{n+1} - a_0.$$

So strictly speaking, neither $\Delta\Sigma$ nor $\Sigma\Delta$ is the identity. However, almost: $\Delta\Sigma a$ is basically the original sequence, "shifted to the left"; and *when applied to sequences that start with a zero*, then also $\Sigma\Delta$ is just a shift to the left. For example, if $a = 0, 1, 2, 3 \ldots$, then $\Sigma\Delta a = \Delta\Sigma a$ is the sequence $1, 2, 3, 4 \ldots$

2.3 The geometric sequence and its sum

If "first differences" are a discrete analog of derivatives and "partial sums" of integrals, is there a discrete analog of the exponential function? The natural guess would be to study the following:

Definition 79. Let v, r be real numbers. The *geometric sequence of initial value v and ratio r* is the sequence $g = g(v, r)$ defined by

$$g_n \stackrel{\text{def}}{=} v \cdot r^n.$$

Note that $g_0 = v$ and $\frac{g_{n+1}}{g_n} = r$, which explains the choice of words "initial value" and "ratio" for v and r, respectively.

Proposition 80. *The first difference of a geometric sequence is again a geometric sequence (with same ratio). In particular, any polynomial of any degree, when restricted to the integers, does not coincide with the geometric sequence.*

Proof. From the definition,

$$\Delta g_n = v \cdot r^{n+1} - v \cdot r^n = v(r-1) \cdot r^n.$$

Iterating,

$$\Delta^k g_n = v(r-1)^k \cdot r^n.$$

Since this is not constant (for any k), the geometric sequence is not polynomial. □

What about the sum? We want to understand the value $f(n) = v + vr + vr^2 + \ldots + vr^n$. The key is that multiplying this quantity by r, one gets

$$rf(n) = v + vr + vr^2 + \ldots + vr^n + vr^{n+1} = f(n) - v + vr^{n+1}.$$

This tells us that

$$f(n) = v\frac{r^{n+1} - 1}{r - 1}.$$

So the sum of a geometric sequence is not a geometric sequence, although it resembles one. In fact, $(r-1)f(n) + 1$ is a geometric sequence (shifted).

Example 81. The geometric sequence $g_n = 3^n$ of ratio 3 and initial value 1 is

$$1,\ 3,\ 9,\ 27,\ 81,\ 243\ldots$$

Its first difference is

$$2,\ 6,\ 18,\ 54,\ 162\ldots$$

which is twice g. If instead we try to sum g, we get the mysterious sequence

$$1,\ 4,\ 13,\ 40,\ 121\ldots$$

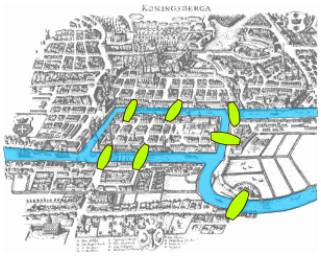
But if we double the latter sequence, and add one to it, we get something much less mysterious:

$$3,\ 9,\ 27,\ 81,\ 243\ldots$$

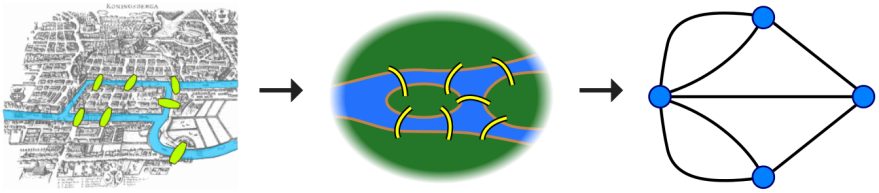
which is the original geometric sequence, shifted by one. (In other words, it's the sequence 3^{n+1}.)

3 Graph Theory

In 1735 Leonhard Euler, the most famous mathematician of his time, was presented a riddle by the mayor of Danzig. Given the map below of the center of Königsberg (now Kaliningrad, Russia), is there a path that travels across all seven bridges exactly once?



Leonhard Euler was a pioneer of all the main branches of mathematics at the time. However, he could not quite tell what kind of math problem this was. It wasn't algebra or analysis, as there was no symbol involved. But it wasn't a geometry problem either, because the shape of the island was irrelevant. In fact, one could simplify the question just by considering a structure of dots and arcs connecting them (or as we would say today, "a multigraph consisting of vertices and edges between them") and ask: could you walk on all edges once?



In honor of Euler, a path where all edges are traveled across exactly once is called *Eulerian*. Euler solved the problem quickly, with a "No, forget about it" answer. In fact, let's call *degree* of a vertex the number of edges touching it. As we walk around with an Eulerian path, every time we pass a vertex, we use one new edge to get there and one new edge to leave. Hence:

- In a *closed* Eulerian path (that is, if we come back to where we started), all vertices have even degree. In fact, the edges touching any given vertex v can be partitioned into two classes – those used to get there, and those used to leave – of equal size.
- In a non-closed Eulerian path, the starting point and the final point will have odd degree, whereas all other vertices will have even degree, for the same reason as above.

So if an Eulerian path exists, the total number of odd-degree vertices is either 0 or 2. Since Königsberg's map had vertex degrees $3, 3, 3, 5$, don't waste your time looking for Eulerian paths.

Euler presented his solution to the St. Petersburg Academy of Sciences on August 26. 1735. The proceedings, *Solutio problematis ad geometriam situs pertinentis*, were published in Comment. Acad. Sci. U. Petrop. 8 (1736), 128–140. This date is considered by many historians the birthdate of the branch of math called **combinatorics**. Euler himself was a bit embarassed in pointing out to a fellow mathematician that the solution, albeit simple, does not fit any of the conventional branches of mathematics at the time (the "art of counting" is what today we would call Calculus).

mechanics, optics, astronomy, navigation, and hydrodynamics. It is not surprising that Euler felt this problem was trivial, stating in a 1736 letter to Carl Leonhard Gottlieb Ehler, mayor of Danzig, who asked him for a solution to the problem [quoted in Hopkins, 2]:

> *"... Thus you see, most noble Sir, how this type of solution bears little relationship to mathematics, and I do not understand why you expect a mathematician to produce it, rather than anyone else, for the solution is based on reason alone, and its discovery does not depend on any mathematical principle. Because of this, I do not know why even questions which bear so little relationship to mathematics are solved more quickly by mathematicians than by others."*

Even though Euler found the problem trivial, he was still intrigued by it. In a letter written the same year to Giovanni Marinoni, an Italian mathematician and engineer, Euler said [quoted in Hopkins, 2],

> *"This question is so banal, but seemed to me worthy of attention in that [neither] geometry, nor algebra, nor even the art of counting was sufficient to solve it."*

3.1 Graphs, multigraphs, and operations

Definition 82. A *graph* is a pair $G = (V, E)$ of a finite nonempty set V and a set E of two-element subsets of V, called *edges*. The elements of V are called *vertices* and typically labeled by $1, \ldots, n$.

We draw a graph by drawing points in correspondence to vertices. Then we draw an arc between each pair of vertices listed in E. We will come back to this representation later on.

Example 83. The *complete graph* K_n has n vertices and all $\binom{n}{2}$ edges.

Example 84. For $n \geq 4$, the *n-cycle* C_n has vertex set $\{1, \ldots, n\}$ and edges $[1,2], [2,3], \ldots, [n-1, n], [n, 1]$.

Example 85. The graph $K_{a,1}$, with $a+1$ vertices and a edges, is defined as follows: Given a set A of a point and an extra point v, draw an edge between any point of A and v.

Implicit in Euler's work is the following statement:

Lemma 86 (Handshake Lemma)**.** *In any graph G with e edges and n vertices, lebaled by $v_1, \ldots, v_n$,*

$$\sum_{i=1}^{n} \deg v_i = 2e.$$

In particular, any graph has an even number of odd-degree vertices.

Proof. This is a good moment to introduce *double-counting*. Let's form an *incidence matrix* of edges and vertices. This is a table that has n rows, in correspondence with the vertices, and e columns, in correspondence with the edges. At the crossing of row v_i and column e_j we place a 1 if e_j contains v_i, and a 0 otherwise. Now let's count the **total number T of ones** in two ways:

- (by row) each row v_i has exactly $\deg v_i$ ones; so $T = \sum \deg v_i$.
- (by column) each column e_j has exactly 2 ones; so $T = 2e$. □

A remark: The definition of graph is great for its simplicity. But it would not capture Königsberg's map, for example, because the latter contained "multiple edges", with same endpoints. To fix this, we present a second definition, more difficult to learn but more inclusive.

Definition 87. A *multigraph* is a triple $M = (V, E, h)$ of

- a finite, nonempty set V, whose elements are called *vertices*.
- a finite set E disjoint from V, whose elements are called *edges*;
- a function $h : E \to F$, where F is the set of two-element and one-element subsets of V.

We draw a multigraph by first drawing points in correspondence with vertices. Then,

– for each element e of E such that $h(e) = \{i, j\}$, we draw an arc between i and j;
– for each element e of E such that $h(e) = \{i\}$, we draw a "loop" starting end ending at i.
The function h need not be injective: so multiple edges may be mapped under h to the same pair $\{i, j\}$. There can also be several loops around the same point.

Proposition 88. *Graphs are multigraphs without loops or multiple edges.*

Proof. "No loops" means, the image of the map h is completely contained in the set 2^V of two-element subset of V. "No multiple edges" means that h is injective. So there is a bijection between E and its image under h. Using this bijection, we can view E as a subset of 2^V. □

Definition 89. A (finite) *walk (between v_0 and v_k)* in a multigraph G is an alternated sequence

$$v_0, e_0, \ldots, v_{k-1}, e_{k-1}, v_k$$

of vertices and edges, such that $e_i = \{v_i, v_{i+1}\}$ for all i. We are not imposing $v_i \neq v_{i+1}$, so loops can be part of a walk. The integer k is called *length* of the walk. The walk is *closed* if $v_0 = v_k$. A *path* is a walk where all vertices are distinct. (In particular, a path cannot contain loops.) A *cycle* is a closed walk where all vertices, except for v_0 and v_k, are distinct. According to this definition, loops and double edges are cycles.

Definition 90. A multigraph is called *connected* if between any two vertices there is a walk (or equivalently, a path).
A multigraph is called *acyclic* if it contains no closed walks (or equivalently, no cycles).

We can now state formally Euler's solution of the seven-bridge-problem. An *Eulerian closed walk* is a closed walk that touches all edges exactly once.

Theorem 91 (Euler 1735, Hierholzer 1873)**.** *Let G be a connected multigraph.*
(1) *G admits a Eulerian closed walk $\iff$ all its vertices have even degree.*
(2) *G admits a Eulerian walk from u to v $\iff$ the sum of the degrees of u and v is even, and all other vertices have even degree.*

Proof technique: ***algorithmic****.* We only sketch the first part, the second is left to you.
'$\Rightarrow$': explained in the previous section.
'$\Leftarrow$': Choose an arbitrary starting vertex v_1; from it, keep walking on previously untaken edges until you return to v_1. (You cannot get stuck anywhere else, because once you use an edge to arrive at a vertex w, by the even degree assumption there is also another edge leaving w.) Call H_1 this closed walk; if $H_1 = G$, the claim is clear. Otherwise, there is some vertex v_2 in H_1 from which untaken edges depart. (Note: There may very well be vertices in G that are not at all touched by the closed walk H; neglect them for the moment. For sure, by the connectedness assumption, if there are untaken edges, then some of them touch H.) We then start a second closed walk from v_2, always choosing previously untaken edges, until we return to v_2. Then we merge this second closed walk to H_1 to get a larger closed walk H_2; if $H_2 = G$, we conclude. Otherwise we continue. □

Remark 92. For graphs, the theorem above is still true if one replaces in (1) "Eulerian closed walk" with "Eulerian cycle". For multigraphs however, this is false: if M is the multigraph consisting of a trangle with an extra loop at each vertex, then each vertex of M has degree 4, but M contains no Eulerian cycle, because by definition a cycle does not contain loops.

On a multigraph $M = (V, E, h)$ we define the following operations:

- the **deletion of an edge** e is the transition from M to a multigraph (V, E', h') where $E' = E \setminus \{e\}$ and h' is the restriction of h to E'.
- the **deletion of a vertex** v is the transition from M to a multigraph (V', E', h') where $V' = V \setminus \{v\}$, $E' = \{e \in E \text{ such that } h(e) \cap \{v\} = \emptyset\}$, and h' is the restriction of h to E'.
- the **contraction of an edge** f is the transition from M to a multigraph (V', E', h') where $E' = E \setminus \{f\}$, $V' = V \setminus h(f) \cup \{v\}$, where v is a "fresh vertex" not in V, and h' is defined as follows:

$$\begin{cases} h'(e) = h(e) & \text{if } h(e) \cap h(f) = \emptyset, \\ h'(e) = \{v\} & \text{if } h(e) \subseteq h(f), \\ h'(e) = \{v, w\} & \text{if } h(e) = \{v', w\} \text{ with } v' \in h(f). \end{cases}$$

  (Check that this definition covers all possible cases!).

Intuitively, in a contraction we first delete an edge v_1, v_2; then we consolidate together the two vertices, calling v their identification; finally, we make this v the new terminal of all edges that previously arrived either to v_1 or to v_2. Note that:

(1) We are not imposing $v_1 \neq v_2$, so we might also contract a loop. The effect of contracting a loop is the same as deleting it.
(2) It is possible that a vertex w was connected in G both to v_1 and to v_2, in which case the edge contraction produces a double edge.

In particular, contractions are *not* internal operations with respect to the world of graphs. (Start from the triangle K_3: A single edge contraction creates a double edge, and if we do one more edge contraction, we obtain a loop.) In contrast, vertex deletions and edge deletions are internal with respect to the world of graphs, because they do not create loops or double edges.

On a final note, given a subset W of the vertices of a multigraph G, we define the **induced subgraph on** W to be the result of the deletion from G of all vertices that are not in W.

These operations provide us with a third proof technique: **induction**.

3.2 Trees

Definition 93. A multigraph is called a *tree* if it is connected and acyclic.

Note that trees are always graphs, because a loop or a double edge would result in a cycle. Sometimes a disjoint union of trees (i.e. an acyclic graph) is called a *forest* in the literature.

Lemma 94. *A graph is a tree $\iff$ between any two vertices there is exactly one path.*
A multigraph is a tree $\iff$ it has no loops, and between any two vertices there is exactly one path.

Proof. We only need to prove the second statement.
"$\Longleftarrow$". The multigraph is connected and has no loops by assumption. If it contained a cycle formed by at least two edges, such cycle would give you two paths between any two vertices of the cycle, against the assumption.
"$\Longrightarrow$". By definition of connected, between any two points of the graph there is at least one path. Were there two, this would result in a cycle. □

Lemma 95. *Deleting any edge from a tree, one gets a "disconnected" (=not connected) graph. Instead, adding any edge (not already present) to a tree, one gets a graph that contains cycles.*

Proof. Let T be a tree. Suppose by contradiction that deleting the edge $\{i, j\}$ yields a connected graph. Then in such graph there is a path from i to j. Which means that in T there is a path P from i to j that does not use the edge $\{i, j\}$. But then P, together with $\{i, j\}$, yields a cycle in T, a contradiction. The second claim is similar and left to the reader. □

So in a tree the number of edges has reached a delicate balance: One more, and you would lose the acyclicity property; one less, and you would lose connectedness. We'll compute this number explicitly in Proposition 99. First, we need a way to define induction.

Definition 96. In a graph, the vertices of degree one are called *leaves*.

Proposition 97. *Any tree with maximum vertex degree D has at least D leaves. In particular, every tree with at least one edge has at least two leaves.*

Proof. First let us verify that the first statement implies the second one. Indeed, apart from the one-edge tree (for which the second statement can be checked manually: both endpoints are leaves), every tree with two or more edges has vertices of degree ≥ 2, and thus from the first statement we get that there are at least $D \geq 2$ leaves. Let us prove the first statement. If $D = 0$, there is nothing to show. If $D > 0$, let v be a vertex of degree D. Let e be any edge departing from v. We claim that **some path starting with v, e, will end in a leaf**. Since there are D edges leaving v, this implies the conclusion: The D leaves we find are obviously different, because inside a tree, there cannot be cycles. But how to prove the claim? The idea is to use an algorithm. Suppose you are a pilgrim, start at vertex v, and march across e. Let v_1 be the first vertex that you encounter, i.e. the other endpoint of e. Is v_1 a leaf? If so, stop: you have achieved our goal. If not, then there is an edge e_1 different from e that leaves v_1. Take it: this will lead you to a vertex v_2. Is v_2 a leaf? If so, stop; if not, keep going. And so on, always choosing edges untraveled by. Since at every step you meet a new vertex (because of the lack of cycles), and since there are finitely many vertices, at some point the algorithm stops. □

Proposition 98. *The deletion of a leaf from a tree, yields a tree.*
Contracting any edge of a tree, yields a tree.

Proof. Let G be the graph obtained from a tree T by deleting a leaf ℓ. If G contained a cycle, so would T, a contradiction. So we only need to show that G is connected. But between any two vertices in T different from ℓ, the (unique) path connecting them does not pass through ℓ, because it is a "dead end": otherwise the edge containing ℓ would have to be traveled twice, which is against the definition of path. So the same path exists in G.
The second claim is similar, and left as exercise. □

Now you can have fun proving things about trees by induction! You can either contract an edge, or delete a *leaf*.

Proposition 99. (A) *Every tree with n vertices has exactly $n-1$ edges.*
(B) *Every connected multigraph with n vertices and exactly $n-1$ edges, is a tree.*
(C) *Every acyclic multigraph with n vertices and exactly $n-1$ edges, is a tree.*

Proof. Part (A) is true for a single point. If $n \geq 2$, then deleting a leaf results in a tree with $n-1$ vertices, and thus $n-2$ edges by inductive assumption.
Part (B): Let G be a connected graph with n vertices and $n-1$ edges. By contradiction, suppose G contains a cycle or a loop. Let's delete an edge from this cycle – or, respectively, let us delete the loop. The resulting connected graph G' has $n-2$ edges. If G' is a tree, stop, otherwise keep going (find a cycle or loop, choose any edge in it, and delete it). Eventually you'll have to stop, obtaining a tree with n vertices and $\leq n-2$ edges: A contradiction with part (A).
Part (C): left to you as exercise. Hint: when a graph G is not connected, we can pick two vertices that are not joined by any path in G, and "add' to G the edge $\{i, j\}$. This won't create any new cycle... □

A subtree that touches all vertices sits inside any connected graph:

Definition 100. Let $G = (V, E)$ be a connected multigraph. A *spanning tree* for G is a tree (V, E'), with $E' \subseteq E$.

Lemma 101. *Every connected multigraph has spanning trees.*

Proof by induction. Let G be a connected multigraph. Either G is acyclic, or not. If G is acyclic, then G is its own spanning tree, and there is nothing to show. If instead G contains a cycle or a loop, choose an edge e from the cycle and delete it. The resulting graph G' is still connected, because we deleted an edge from a cycle, so any path through e can be "rerouted" using the rest of the cycle. Hence, by inductive assumption, G' has a spanning tree T'. The same T' works for G too. □

3.3 Bipartite graphs

Trees and forests do not contain any cycle. Next, we study the broader class of graphs that do not contain any cycle of odd length.

Definition 102. A *bipartite multigraph with parts X and Y* is a multigraph whose vertex set can be partitioned into two disjoint sets X and Y, such that every edge of the multigraph has one endpoint in X and one endpoint in Y.

In particular, bipartite multigraphs have no loops, which are cycles of length 1. According to our definition, the graph consisting of a single vertex and no edge is bipartite. Also any edge-less graph is bipartite, so bipartite multigraphs need not be connected.

Lemma 103. *A graph is bipartite if and only if all its connected components are bipartite.*

Proof. If G_i is bipartite with parts X_i and Y_i, for $i = 1, 2$, then it is easy to see that $G_1 \cup G_2$ is bipartite with parts $X_1 \cup X_2$ and $Y_1 \cup Y_2$. Conversely: Suppose that $G_1 \cup G_2$ is bipartite with parts X and Y, with each G_i connected, and no edge going from G_1 to G_2. Let X_i and Y_i be the vertices of X belonging to G_i, for $i = 1, 2$. Then every edge of G_i must have one endpoint in X_i and one in Y_i; hence, G_i is bipartite. □

Proposition 104. *Let G be a multigraph. The following are equivalent:*
(1) *G is bipartite;*
(2) *G does not contain any odd cycle;*
(3) *G does not admit any closed walk of odd length.*

Proof. **(3) ⇒ (2)** is trivial.
(1) ⇒ (3): Let G be bipartite with parts X and Y. Since there are no loops, a walk of length k starting at a vertex of X will lead to a vertex in X if k is even, and to a vertex in Y if k is odd. So all closed walks have even length.
(2) ⇒ (1): By the previous Lemma, it suffices to prove the claim for each connected component of the multigraph; in other words, we may assume that the multigraph G is connected. By assumption G has no loops, which are closed walks of length one. Choose a spanning tree T for G (cf. Lemma 101). Choose a *root*, i.e. a preferred vertex r in G. By Lemma 94, every other vertex w of G has a unique path in T connecting it to v. The length of this path, i.e. the "distance between v and w along T", may be odd or even. Divide the vertices w of G into two sets X and Y, according to whether their distance from v along T is odd or even. We claim that G is bipartite with parts X and Y. In fact, clearly *in* T there is no edge from X to X

(or from Y to Y). What about in G? Were there an edge $[x, x']$ in G going from X to X, this edge (not in T, by what we said above) would form a closed walk C together with the unique odd-length paths from x to v and from x' to v in T. So the total number of edges of C is the sum of two odd numbers, plus one: so C is an odd closed walk. We are not sure if C is a cycle or not, because possibly we "backtracked" on some of the edges of T; but then C minus we edge we traveled twice on, yields the desired odd cycle C'. Similarly, were there an edge $[y, y']$ in G connecting two vertices of Y, this would produce a closed walk C with a number of edges that is the sum of two even numbers, plus one: so from this C we could extract an odd cycle C', a contradiction. □

Corollary 105. *All trees and all even cycles are bipartite.*
In contrast, K_n is not bipartite for $n \geq 3$.

There is however a way to introduce a graph with "the most edges possible", within the bipartite class:

Definition 106. The *complete bipartite graph* $K_{a,b}$ is the bipartite graph with parts $X = \{1, \ldots, a\}$ and $Y = \{a+1, \ldots, a+b\}$, and one edge $[x, y]$ for every vertex $x \in X$ and for every vertex $y \in Y$.

We have already encountered $K_{a,1}$, which is the $b = 1$ case of $K_{a,b}$, in Example 85. Clearly $K_{a,b}$ has exactly ab edges. So on $2n$ vertices, the complete graph has $\binom{2n}{2} = 2n^2 - n$ edges, whereas a bipartite graph can have at most n^2 edges, a maximum achieved by $K_{n,n}$.

4 Graph drawings, polytopes and planar graphs

Definition 107. Let $d \geq 2$. A *line drawing in $\mathbb{R}^d$* of a graph $G = (V, E)$ on n vertices is an injective map φ from V to $\mathbb{R}^d$, which allows to represent each edge $[i, j]$ of G by means of a straight segment that joins $\varphi(i)$ and $\varphi(j)$, with the property that any two segments either intersect at an endpoint of both, or do not intersect at all.

Definition 108. Let $d \geq 2$. A *(broken-line) drawing in $\mathbb{R}^d$* of a graph G is a representation by means of *broken-line segments*, with the property that no two broken-lines may intersect at any point in the relative interior of any of them. (By "relative interior" of a broken-line joining $\varphi(i)$ and $\varphi(j)$ we mean any point other than $\varphi(i)$ or $\varphi(j)$.) A drawing in $\mathbb{R}^d$ of a graph $G = (V, E)$ on n vertices is not fully determined by the choice of n points in $\mathbb{R}^d$; for any edge $[i, j]$, we also have to choose a finite sequence of points, starting at $\varphi(i)$ and ending at $\varphi(j)$. Details are boring and left to the reader.

It turns out that we might restrict the definitions above to $d \in \{2, 3\}$, because of the following result:

Proposition 109. *Any graph has a line drawing in $\mathbb{R}^3$, and thus in any $\mathbb{R}^d$ with $d \geq 3$.*

Proof. The key idea is *randomness*. Let us choose n points in $\mathbb{R}^3$ that are affinely independent, i.e. no four of which are on the same plane. (Note that this condition is satisfied with probability 1 if the points are picked randomly from the unit cube.) Then any two disjoint edges $[i, j]$ and $[h, k]$ of G will be represented by two segments that are skew to one another; and skew lines do not intersect. □

A much more delicate question is which graphs admit a line or broken-line drawing in $\mathbb{R}^2$. To answer this question, we need some "discrete geometry".

Definition 110. A multigraph is *planar* if it has a broken-line drawing in the plane so that any two broken-lines intersect only at their endpoints, or do not intersect at all. Such a drawing breaks the plane into a finite number of connected components called *regions*.

We will see later that the definition simplifies a bit for graphs.

4.1 A crash course on polytopes, linear optimization

A particular case of planar graphs are the graphs of 3-dimensional polytopes, like cubes, pyramids, and so on. Let me explain a bit better what a polytope is, and then we'll come back to planar graphs.

Polytopes

Definition 111. A subset $A \subseteq \mathbb{R}^d$ is *convex* if for any two points x, y of A, the entire segment from x to y is in A.

Lemma 112. *The intersection of convex sets is convex. Also, $\mathbb{R}^d$, as subset of itself, is convex.*

Proof. Exercise. □

Lemma 113. *Given a set S inside $\mathbb{R}^d$, there exists a smallest convex subsets containing all of them (called "convex hull of S").*

Proof. By the previous lemma, certainly there is a convex subset contaning S, namely, the whole $\mathbb{R}^d$. But that typically is not be the smallest. To find the smallest, just take the intersection of all convex sets that contain S. This set still contains S, and is convex by the previous Lemma. □

Definition 114. A *polytope* is the convex hull of some finite set in $\mathbb{R}^d$.

Definition 115. Let P be a polytope in $\mathbb{R}^d$. A *face of P* is any subset $F \subset \mathbb{R}^d$ of the form

$$F = \{\mathbf{x} \in P \text{ such that } \mathbf{c} \cdot \mathbf{x} = c_0\},$$

where $\mathbf{c} \cdot \mathbf{x} \leq c_0$ is an inequality satisfied by all $\mathbf{x}$ in P.
We also say that the linear inequality $\mathbf{c} \cdot \mathbf{x} \leq c_0$ *supports the face F of P*.

In other words, "faces" are where linear functions are maximized within P. Faces may have different dimensions: For example, P is a face of itself, by taking $\mathbf{c} = \mathbf{0}$ and $c_0 = 0$. But also the empty set is a face of any polytope P, by taking $\mathbf{c} = \mathbf{0}$ and $c_0 = 1$.

Definition 116. The faces of dimension 0, 1 and $\dim P - 1$ are called *vertices*, *edges*, and *facets*, respectively.

If S is a finite set, the vertices are in general a subset of S, as the next Proposition explains:

Definition 117. Given finitely many points $\mathbf{x_1}, \ldots, \mathbf{x_n}$ in A, a *convex combination (of points in A)* is a point

$$\mathbf{x} = \sum_{i=1}^{n} \lambda_i \mathbf{x_i}, \quad \text{where} \sum_{i=1}^{n} \lambda_i = 1 \text{ and } \lambda_i \geq 0 \text{ for all } i.$$

Lemma 118. *The convex hull of A is the set of all possible convex combination of points in A.*

Proof. Exercise. Hint: call X the set on the right and show (1) that X is convex, (2) that any $a \in A$ belongs to X, and (3) that any element of X belongs to any convex set containing A. □

Proposition 119. *Let P be a polytope in $\mathbb{R}^d$. Let $\mathbf{v} \in P$. The following are equivalent:*

(i) $\mathbf{v}$ *is a vertex of* P*;*
(ii) $\mathbf{v}$ *is not in the convex hull of* $P \setminus \{v\}$*;*
(iii) *the only vector* $\mathbf{w}$ *for which both* $\mathbf{v}+\mathbf{w}$ *and* $\mathbf{v}-\mathbf{w}$ *are in* P*, is* $\mathbf{w} = \mathbf{0}$*.*
(iv) *There are* d *constraints* $\mathbf{a}_j^\top \mathbf{x} \le b_j$ *valid for* P *which are tight at* $\mathbf{v}$ *(that is,* $\mathbf{a}_j^\top \mathbf{v} = b_j$ *for all* $1 \le j \le d$*) and the vectors* $\mathbf{a}_1, \ldots, \mathbf{a}_d$ *are linearly independent.*

Proof. (i) $\Rightarrow$ (ii): By contradiction, $\mathbf{v} = \sum \lambda_i \mathbf{x}_i$, with $\mathbf{x}_i \in P$ different from $\mathbf{v}$, with $\lambda_i \ge 0$ for all i, and with $\sum_i \lambda_i = 1$. By the assumption, $\{\mathbf{v}\} = H \cap P$, where H is some hyperplane of the form $\mathbf{c}^\top \mathbf{x} = c_0$, and such that $\mathbf{c}^\top \mathbf{x}_i < c_0$ for all i, because $\mathbf{v}$ is the only vertex in $H \cap P$. But then we reach the contradiction

$$c_0 = \mathbf{c}^\top \mathbf{v} = \mathbf{c}^\top (\sum \lambda_i \mathbf{x}_i) = \sum \lambda_i \mathbf{c}^\top \mathbf{x}_i < \sum \lambda_i c_0 = c_0.$$

(ii) $\Rightarrow$ (iii): Write $\mathbf{v} = \frac{1}{2}(\mathbf{v}+\mathbf{w}) + \frac{1}{2}(\mathbf{v}-\mathbf{w})$.
(iii) $\Rightarrow$ (iv): By contradiction, let $\mathbf{a}_j^\top \mathbf{x} \le b_j$ be any constraint valid for P that is *not* tight at v. That means that for some small vector $\mathbf{w}$, both $\mathbf{a}_j^\top (\mathbf{v}+\mathbf{w}) \le b_j$ and $\mathbf{a}_j^\top (\mathbf{v}-\mathbf{w}) \le b_j$. Hence, both $\mathbf{v}+\mathbf{w}$ and $\mathbf{v}-\mathbf{w}$ are in P.
(iv) $\Rightarrow$ (i): Define a hyperplane by

$$H \stackrel{\text{def}}{=} \{\mathbf{x} \text{ such that } \sum_{j=1}^{d} \mathbf{a}_j^\top \mathbf{x} = \sum_{j=1}^{d} b_j\}.$$

Then every $\mathbf{x}$ in P satisfies $\mathbf{a}_j^\top \mathbf{x} \le b_j$ for all j, so in particular it satisfies $\sum_{j=1}^d \mathbf{a}_j^\top \mathbf{x} \le \sum_{j=1}^d b_j$. Moreover, if $\mathbf{x}$ is any element in $P \cap H$, then all d constraints are tight at $\mathbf{x}$; but since a rank-d system of linear equations in $\mathbb{R}^d$ has only one solution, we conclude that $\mathbf{x} = \mathbf{v}$. □

Corollary 120. *A polytope is the convex hull of its vertices (and of no proper subset thereof).*

Proof. Let A be a finite set. Let V be the list of vertices of the polytope $P \stackrel{\text{def}}{=} \text{conv}(A)$. Let $V' = V \cup A$. Since $A \subseteq V' \subseteq P$, clearly

$$P = \text{conv}(A) \subseteq \text{conv}(V') \subseteq \text{conv}(P) = P,$$

which implies $P = \text{conv}(V')$. Now all points not in V can be written as combination of points in V, so $P = \text{conv}(V)$. □

Polytopes and planar graphs

How are polytopes connected to planar graphs? Well, the intuition is that if you put your nose very close to one of the facets of a $3D$-polytope, what you'll see is a *planar drawing* (drawn actually with straight edges!, not broken-lines) of its graph. The number of regions will be exactly the number of facets.

In fact, one can prove the following beautiful theorem (a proof can be found in Ziegler, Lectures on Polytopes, Springer).

Definition 121. A graph with at least $d+1$ vertices is called is *d-connected* if you cannot disconnect it by deleting less than d vertices (no matter how you choose them).

For example, 1-connected is the same as connected. Cycles are 2-connected. The graph of a cube is 3-connected.

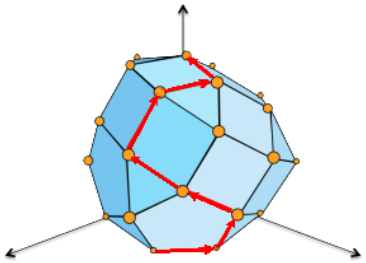
Theorem 122 (Steinitz)**.** *G is the graph of a 3-dimensional polytope $\iff$ G is planar and 3-connected.*

Of course, not all planar graphs arise from 3-dimensional polytopes. For example, a tree is not the graph of any 3-dimensional polytope. (Trees have leaves, whereas all vertices in a 3-dimensional polytope have degree ≥ 3...) We will prove only the $\Rightarrow$ direction of Steinitz' theorem, which has the advantage that it generalizes to higher dimensions. The converse direction, going back to ideas of James Clerk Maxwell and Luigi Cremona, is perhaps more important for applications, but specific for dimension 3.

Linear Optimization in five minutes

What is **linear optimization** about? Let $f(x_1, \ldots, x_d) = x_d$ be the "height" function from $\mathbb{R}^d$ to $\mathbb{R}$. Given a region $P \subset \mathbb{R}^d$, suppose we want to find the highest point, $\max\{f(\underline{x}) : \underline{x} \in P\}$. If P is a *polytope*, two dreams come true:

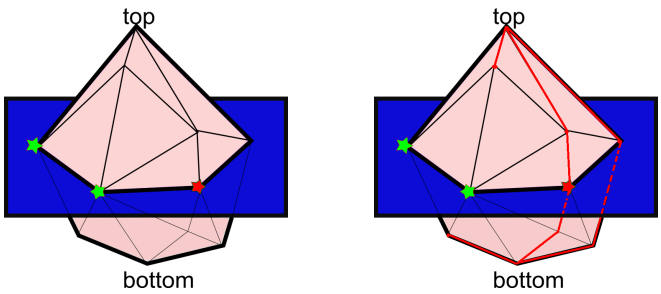
1. $\max\{\, f(\underline{x}) \text{ for } \underline{x} \in P \,\} = \max\{\, f(\underline{v}) \text{ for } \underline{v} \text{ vertex of } P \,\}$;
2. because of convexity, any local max is also a global max.



Naïf Simplex Method: Start at any vertex; move to an adjacent vertex that is higher (under f); keep climbing!, until you get stuck. But when you get stuck, it means you are at a local maximum: so you are at a global maximum, that is, you are at the very top.

OPEN: (*weak Hirsch Conjecture*) Any two vertices of any polytope with n facets are at most $2n$ edges apart. ($\rightsquigarrow$ worst-case running time for Simplex Method). We know this is true for some polytopes.

Theorem 123 (Balinski)**.** *The graph of every d-dimensional polytope is d-connected.*



Sketch of proof. Choose the $k \leq d-1$ vertices that you're losing (green), and another designated survivor x (red). These $k+1 \leq d$ vertices will all belong, within $\mathbb{R}^d$, to some hyperplane. Up to performing a rotation/translation of coordinates, we can imagine that this hyperplane is horizontal. In fact, let's say it's the hyperplane $z = 0$. The hyperplane chops the polytope into two polytopes, both containing x. Apply the simplex method to both polytopes.

From every vertex v of the top polytope, the simplex method gives you a climbing path to the top vertex T. Because this path is climbing, it won't go down to the floor level $z = 0$, and so in particular it won't touch the green vertices. Hence, this path will survive the deletion of the green vertices. Symmetrically from the bottom: from any vertex w of the bottom polytope we can descend to the bottom vertex B with a path that does not touch the green vertices (and thus survives their deletion). And now you understand the importance of the designated survivor! The vertex x belongs to *both* the top and the bottom polytope, and so from x you can climb to T and also descend to B, and both these paths will survive the deletion of the green vertices. □

Unfortunately, Balinski's theorem is the only result we know for graphs of d-polytopes, $d \geq 4$.

4.2 Euler's theorem

A subset $B \subset \mathbb{R}^2$ is called *bounded* if lies in a sufficiently large rectangular box. Because of the existence of Peano's curves, i.e. continuous functions $c : [0,1] \longrightarrow [0,1] \times [0,1]$ that are *surjective* (i.e. their trajectory fills the square!, like a Roomba robot sweeps the room), the next theorem is non-trivial.

Lemma 124 (Jordan's theorem, 1887)**.** *Let c be an* injective *continuous map from the unit circle to $\mathbb{R}^2$. (The image of c in this case is what is called "a simple closed curve".) Then the complement of the image of c consists of exactly two connected components, one bounded and one unbounded.*

Theorem 125 (EULER'S FORMULA: Descartes* 1630, Legendre* 1794; von Staudt 1847)**.** *Given a planar connected graph with v vertices, e edges, and r regions,*

$$v - e + r = 2.$$

Many proofs of the Theorem above appeared throughout the centuries. I recommend browsing https://www.ics.uci.edu/~eppstein/junkyard/euler/interdig.html for a history. A crucial fact to keep in mind is that the first, fully correct proofs (Descartes 1630, Legendre 1794) did not need Jordan's theorem, but work only for planar graphs that come from 3D-polytopes. That's why we marked them with an asterisk. In 1847 von Staudt gave the first proof valid for all planar graphs. Von Staudt's proof (as well as basically all subsequent proofs valid for all planar graphs, including Möbius's 1863 "proof by deluge") relies on Jordan's curve theorem, which was proven by Jordan in 1887; so today we would say that von Staudt and Möbius' proofs "became complete" only in 1887. (To be completely fair to von Staudt, if you start with a planar drawings with *straight* edges, then von Staudt's proof and the proof by induction on r rely only on Jordan's theorem for *polygonal* curves. This is a much simpler theorem, which can be considered an exercise and was certainly known before Jordan's work; a proof is in Tverberg, "A proof of the Jordan curve theorem", 1979. .)

Proof by induction on v. Since what we have in mind is an edge contraction, we are forced to prove a stronger statement. In fact, after an edge contraction, we could get double edges, or in other words, a multigraph that is not a graph; so the inductive assumption would not apply.

To fix this, we try to prove by induction that Euler's formula holds for all planar, connected *multigraphs*.

So let's prove it! Let G be a planar connected *multigraph* with v vertices, e edges, and r regions. We prove by induction on v that $v - e + r = 2$. If $v = 1$, then G consists of e nested loops. Applying Jordan's curve theorem iteratively, from the outer to the inner loop, one can prove that $r = e + 1$. (For this, first you have to prove that given two distinct simple closed curves γ_1, γ_2, the union of their images tiles the space into four connected components, some of which could ne empty: inside both, outside both, inside one but outside the other. Then observe that for *nested* simple closed curves, exactly one of these four regions is empty.) So $v - e + r = 1 - e + (e + 1) = 2$. Good!

Now the induction step. If $v \geq 2$, contract an edge that joins two distinct vertices of G. This yields a connected graph G' with one less vertex, one less edge, and same number of regions. In other words, if G' has v' vertices, e' edges, and r' regions, we have that $v' = v - 1$, $e' = e - 1$, $r' = r$. But since by inductive assumption $v' - e' + r' = 2$, we get

$$v - e + r = (v' + 1) - (e' + 1) + r' = v' - e' + r' = 2. \qquad \square$$

Proof by induction on e. Left as exercise. (Do the $e = 1$ case, distinguishing the $v = 0$ and the $v = 1$ case. As for the induction step, look at the proof by induction on v.) $\square$

Proof by induction on r. The key is to show that $r = 1 \iff$ the graph G contains no cycle. The direction "$\Rightarrow$" is an application of Jordan's curve theorem, because a cycle would create two connected components, and any further cycle drawn on top of it would only increase the number of connected components. As for "$\Leftarrow$", every graph (connected or not) is bounded, because it contains finitely many edges, and every edge is bounded by Weierstrass' theorem (it is the image of a continuous function $\gamma : [0, 1] \to \mathbb{R}^2$). So if an edge of the graph belongs to two regions, at least one of the two regions is bounded. The boundary of this bounded region must be a subset of the graph. Hence, if $r \geq 2$, the graph contains a cycle.

Now we are ready to set up the induction. If $r = 1$, then G contains no cycle by the claim above. So G is a tree, and $e = v - 1$ by Proposition 99. If $r \geq 2$, then again by the claim above G contains a cycle. Let e be any edge in such cycle. Then the deletion of e yields a graph G' that is connected (because any path through e can be "rerouted" using the rest of the cycle) and planar. Deleting e puts two regions in communication: as a result, G' has one less region than G. Since G' has $v' = v$ vertices, $e' = e - 1$ edges and $r' = r - 1$ regions, by induction

$$v - e + r = v' - (e' + 1) + (r' + 1) = v' - e' + r' = 2. \qquad \square$$

Proof by intertwining trees (von Staudt,1847). Draw the dual graph G^* of G defined as follows: draw a vertex in the middle of each region of G, and then connect the vertices from two adjacent regions of G by a curve e^* that crosses their shared edge e. By the Jordan curve theorem, any cycle in G disconnects G^*. Now pick a spanning tree T for G, and consider the subgraph $(G-T)^*$ of G^* formed by the dual edges of the complement of T. A cycle in $(G-T)^*$ would disconnect T; so since T is connected, $(G-T)^*$ is acyclic. The other way around, since T is acyclic, $(G - T)^*$ is connected. (All these implications make use of Jordan's curve theorem.) Conclusion: $(G - T)^*$ is a tree, so by construction, it is a spanning tree of G^*. By Proposition 99, T has $v - 1$ edges and $(G - T)^*$ has $r - 1$ edges. But every edge of G is either in T, or it is crossed by exactly one edge of $(G - T)^*$. Hence, the number of edges of G is exactly $e = (v - 1) + (r - 1)$. $\square$

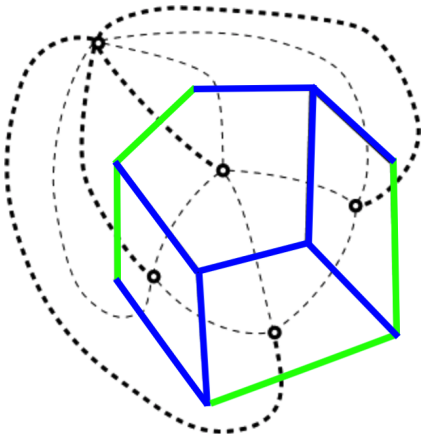


Figure 1: The spanning tree T (in blue) for G (in blue or green) determines a subgraph $(G-T)^*$ (in black, dashed) of G^* (in black or grey, dashed) that is also a spanning tree. (Colorized figure from Eppstein's website, https://www.ics.uci.edu/~eppstein/junkyard/euler/interdig.html).

Proof by deluge (Möbius, 1863). Imagine the graph as drawn on the $z = 0$ plane of $\mathbb{R}^3$. Now lift each vertex vertically upwards, so it reaches a height that is a randomly chosen number in the interval $(0, 1)$. Then deform each edge continuously in the following way: pull the center of the edge further upwards, as when picking up the handle of a suitcase, so that in the end the center of the edge reaches a height that is a randomly chosen number in the interval $(1, 2)$. The endpoints of the edge don't move in the process, so their height is lower, it's still a number in $(0, 1)$. Finally, deform each bounded region in a similar fashion: choose a point in the middle and bump it upwards, until it reaches a height that is a randomly chosen number in the interval $(2, 3)$. Again, this is only a central bump: the region stays a continuous surface, and its boundary edges have not been touched.

Note that the result in the and is a surface with $v + r + e$ critical points: v local minima at the vertices, e saddles at the chosen points of edges, and r local maxima at the chosen points of regions. View the surface as a territory somewhere on Earth. Now a deluge arrives. Ultimately, the deluge completely submerges even the highest peak. So the water level gets from $z = 0$ (first drops) to $z = 3$ (end of the deluge), say.

Let's take a closer look at what happens as the water level goes up. Initially, from $z = 0$ to $z = 1$, the rain will form one lake per each local minimum. In the next phase, from $z = 0$ to $z = 1$, the water will pass saddles. Two things can happen as the water level passes a saddle (depending on whether the two lakes on the sides were already communicating):

(a) either an island is formed because a lake doubles back on itself (like in Mont St. Michel with high tide; to claim an island is formed, we are implicitly using Jordan's curve theorem),
(b) or two different lakes merge.

Let us divide the edges in 'edges of type (a)' and 'edges of type (b)' accordingly. Let e_a and e_b be their number, respectively. Clearly, since all edges are of one of these two types,

$$e_a + e_b = e. \tag{4}$$

Finally, in the third phase, from $z = 2$ to $z = 3$, for each peak passed a landmass will be eliminated.

Let us count the number of land masses: initially there is one, then we add one landmass in the second phase for each edge of type (a), and then we lose one landmass for each peak.

Eventually, it's all under water, so 0 landmasses. Thus

$$1 + e_a - r = 0. \tag{5}$$

Let us now count the number of lakes: initially we add v lakes in correspondence with the vertices. Then we decrease the number of lakes by one for each edge of type (b). In the third phase there is no change in the number of lakes. Eventually, it's all just one lake. Thus

$$0 + v - e_b = 1. \tag{6}$$

Combining Equations 4, 5 and 6, we get

$$1 = 1 - 0 = (0 + v - e_b) - (1 + e_a - r) = v - (e_a + e_b) + r - 1 = v - e + r - 1. \qquad \square$$

Lost-and-found proof by Descartes, 1630. This proof only works for graphs of 3-polytopes, which are only *some* of the connected planar graphs. Choose a facet H and take a straight-line planar drawing of the polytope (graph) by going very close to H. Let us double-count the sum X of all angles in the drawing, including the inner angles of H. Since the sum of all angles in a k-gon is $(k-2)\pi$, and since each edge contributes to 2 faces, we have $\sum_{k\geq 3} \sum_{G \text{ k-gon}} k = 2e$. Note also that $\sum_{k\geq 3} \sum_{G \text{ k-gon}} 2 = 2f$. So counting with respect to facets

$$X = \sum_{k\geq 3} \sum_{G \text{ k-gon}} (k-2)\pi = \pi \left(\sum_{k\geq 3} \sum_{G \text{ k-gon}} k - \sum_{k\geq 3} \sum_{G \text{ k-gon}} 2 \right) = \pi(2e - 2f) = 2\pi(e - f).$$

Now we count the same angle sum X with respect to the vertices. Each interior vertex is surrounded by triangles, so it contributes to X a total of 2π. Each vertex w_H on the outside face H contributes instead twice the angle at w_H. If H is a convex polygon with h vertices and h edges, the sum of the angles of H is $(h-2)\pi$ (prove it by induction!, you know the $h = 3$ case...) So we obtain

$$\begin{array}{lll} X & = \sum_{v \text{ vertex}} \sum_{\alpha \text{ angle at } v} \alpha & \\ & = \sum_{v \notin H} \sum_{\alpha \text{ angle at } v} \alpha & + \sum_{v \in H} \sum_{\alpha \text{ angle at } v} \alpha \\ & = \sum_{v \notin H} 2\pi & + 2(h-2)\pi \\ & = 2\pi(v-h) & + 2\pi(h-2) \qquad = 2\pi(v-2). \end{array}$$

Thus, $2\pi(e-f) = X = 2\pi(v-2)\pi$: dividing by 2π, we get $e - f = v - 2$. $\square$

The last part of this chapter is dedicated to show that some graphs are not planar.

Proposition 126. *Let G be a planar graph with at least three edges. Then*

$$e \leq 3v - 6.$$

More generally, if every region of G has at least m edges, then

$$e \leq \frac{m}{m-2} v - \frac{2m}{m-2}.$$

Proof. Note that the second statement implies the first one by setting $m = 3$: In fact, the assumption "G has at least 3 edges" guarantees that every region of G has at least 3 edges. So let's prove the second statement. First let us show it for *connected* graphs. The idea is double-counting: create an incidence matrix of edges and regions. Let's index rows by regions, and columns by edges. At the intersection of the column that corresponds to the edge e_j and of the row that corresponds to region r_i, we place a 1 in case e_j belongs to r_i, and a 0 otherwise. We count the total number T of ones in two ways:

- (by row) Every region of G has at least m edges, so $T \geq rm$.
- (by column) Every edge is in either 1 or 2 regions, so $T \leq 2e$.

In conclusion,

$$rm \leq 2e.$$

But since G is a planar graph, and we are assuming it's connected, we can apply Euler's theorem: $v - e + r = 2$, or $r = 2 - v + e$. Plugging in,

$$2m - vm + em \leq 2e,$$

or shortly, $(m-2)e \leq vm - 2m$. Dividing by $m-2$, which is positive, we obtain the desired inequality

$$e \leq \frac{m}{m-2}v - \frac{2m}{m-2}.$$

Now suppose that G is *not* connected. Given i, j in G not connected by a path, we can add the edge i, j; and up to moving/rotating/rescaling the connected components, we can assume that such edge does not intersect other existing edges. So from G we obtain a new planar graph G_1 with $v_1 = v$ vertices, $e_1 = e + 1$ edges, and $r_1 = r$ regions. If G' is connected, we stop; otherwise we iterate. Eventually, we will reach a connected planar graph G', with $v' = v$ vertices, $e' > e$ edges, and $r' = r$ regions. To this G' we can apply the present Proposition, which we already proved in the connected case. So we know that

$$e' \leq \frac{m}{m-2}v - \frac{2m}{m-2}.$$

But since $e < e'$, we derive that

$$e < \frac{m}{m-2}v - \frac{2m}{m-2}.$$ □

Corollary 127. *In a planar graph, there is at least one vertex of degree* ≤ 5.
If in addition every region of G *has at least* 4 *edges, then some vertex has degree* ≤ 3.
If in addition every region of G *has at least* 6 *edges, then some vertex has degree* ≤ 2.

Proof. The first claim is clearly true for planar graphs with one or two edges: in that case, there is even one vertex of degree 1. So suppose the graph G has at least 3 edges. Let's compute the *average vertex degree*, using the Handshake Lemma.

$$\text{average degree } = \frac{\sum_{v_i \in V} \deg v_i}{v} = \frac{2e}{v}.$$

But by the previous Proposition,

$$\frac{2e}{v} \leq \frac{6v-12}{v} = 6 - \frac{12}{v} < 6.$$

And when an average of integers is strictly less than 6, one of those integers has to be strictly less than 6, so at most 5. More generally, if every region of G has at least m edges, then by the previous Proposition the average degree is

$$\frac{2e}{v} \leq \frac{2m}{m-2} - \frac{2m}{(m-2)v} < \frac{2m}{m-2}.$$

For $m = 4$, we have $\frac{2m}{m-2} = 4$, so we conclude that there is a vertex of degree ≤ 3. For $m = 6$, we have $\frac{2m}{m-2} = 3$, so we conclude that there is a vertex of degree ≤ 2. □

Corollary 128. *K_n is planar if and only if $n \leq 4$.*

Proof. K_1, K_2 and K_3 are obviously planar. K_4 is the graph of the tetrahedron, a 3-dimensional polytope; so it is planar. To complete the proof, it suffices to show that K_5 is not planar, because K_5 is a subgraph of K_n for $n \geq 5$. (So if K_n coud be drawn in the plane without self-intersections, so could K_5.) But K_5 has 5 vertices and $\binom{5}{2}$ edges, too many to satisfy $e \leq 3v - 6$. □

Corollary 129. *$K_{a,b}$ is planar if and only if one of a and b is ≤ 2.*

Proof. When $a \leq 2$, planarity is clear: draw all points of B on the line $y = 0$ of the plane, and place the points of A in the set $\{(0, 1), (0, -1)\}$. When $b \leq 2$, switch labels of A and B, and do as above. When $a \geq 3$ or $b \geq 3$, clearly $K_{3,3}$ is a subgraph of $K_{a,b}$, so if we prove that $K_{3,3}$ is not planar, we are done. Now, $K_{3,3}$ has $3 + 3 = 6$ vertices and $3 \cdot 3 = 9$ edges, so it does satisfy $e \leq 3v - 6$. However, the crucial observation to conclude is that $K_{3,3}$ contains no triangle: the smallest region is a square. So by 126, applied to $m = 4$, it should also satisfy

$$e \leq \frac{4}{4-2}v - \frac{2 \cdot 4}{4-2},$$

or $e \leq 2v - 4$. But it doesn't: $9 > 12 - 4$. □

5 Networks

Many real-life situations require to streamline "stuff" between two nodes of a graph. "Stuff" could be data to stream through a communication network, liquids to convey through pipelines, electricity to bring over via cables, good to be transported via train tracks, passengers to car-share around through the city streets, and so on. All these problems have two aspects in common that the notion of "graph" does not quite capture:

1. There is often a preferred, or even a unique direction in which edges can be traveled across. For example, streets can be one-way. Clean water comes from the city reservoirs to our faucet, but not the other way around: The water pipes are always one-way.

2. Every edge has a maximum capacity. A pipe has finite diameter, so it can't transport infinite amounts of water. Same for train tracks, electricity cables, or internet connections.

We want to incorporate this into our model of "graph". To this end, we modify the definition as follows:

Definition 130. A *network* is a triple $N = (V, E, c)$ of
- a finite, nonempty set V whose elements are called *vertices*;
- a nonempty subset $E \subset V \times V$, whose elements are called *directed edges*;
- a function $c : V \times V \to \mathbb{Q}_{\geq 0}$, called *capacity*, that is zero on non-edges.

We also assume that E "avoids the diagonal", i.e. contains no element of the type (v, v), with $v \in V$.

We draw a network by first drawing points in correspondence with vertices. Then for each element (i, j) of E, we draw an *arrow* from i to j. Because we imposed E to avoid the diagonal, our networks have no loops. Finally, we write the integer $c(e)$ on top of the edge e.

Remark 131. You may wonder why we defined the capacity also on non-edges, and then set it to zero. The reason will become clear in a few minutes, as we will write down Kirchhoff's law (Remark 136). You may also wonder why we imposed the capacity to be a rational number.

Well, it would be perfectly equivalent to require $c : E \to \mathbb{N}$, since one may rescale. However, there is a surprising difference between rational-capacity networks and real-capacity networks: If we allow edge capacities to be real numbers, then the algorithm we present in the next section (called Ford–Fulkerson's algorithm) might not converge.

Definition 132. With respect to a vertex v, the edges containing v can be partitioned into *outgoing edges*, i.e. those of the type (v, w), for some $w \in V$; and *incoming edges*, i.e. those of the type (w, v), for some $w \in V$.

Definition 133. The vertices of a network can be of four types:
- **isolated vertices** are those without incoming or outgoing edges;
- **sources** are those with outgoing edges, but without incoming edges;
- **sinks** are those with incoming edges, but without outgoing edges;
- **intermediates** are those with both incoming and outgoing edges.

For example, a cycle with all edges oriented clockwise is a network with no source and no sink: all vertices are intermediates. If instead we consider the network $(1, 2), (1, 4), (3, 4)$, it has two sources, two sinks, and no intermediates.

Definition 134. An *(s,t)-network* is a network that has exactly one source, labeled by s, and exactly one sink, labeled by t.

So the (s,t)-flow network is the easiest instance of a transportation problem: There's some good stocked at s, and we want to transport it across the network to t, keeping in mind that each edge has a limited maximum capacity, and that nobody steals across the pipeline.

Definition 135. A *flow* in an (s,t)-network is an assignment of a non-negative rational number $f(u, v)$ to every pair of vertices (u, v), such that
- ("capacity constraint") $f(u, v) \leq c(u, v)$ for every (u, v);
- ("Kirchhoff's law") at every intermediate vertex v, the sum of flows of incoming edges equals the sum of flows of outgoing edges.

The *value* (or *cost*) of the flow is the total amount of flow exiting the source.

Remark 136. Because we defined the capacity of non-edges to be zero, by the capacity constraint we have $f(u, v) = 0$ for every $(u, v) \notin E$. This allows us to write down Kirchhoff's law in a concise form:

$$\text{for every intermediate } i, \qquad \sum_{v \in V} f(v, i) = \sum_{v \in V} f(i, v). \tag{7}$$

And similarly, the value of the flow can be concisely written as

$$\text{val}(f) = \sum_{v \in V} f(s, v).$$

So here's what we want to study:

MAXIMUM FLOW PROBLEM: **Given an (s, t)-network, find a flow of maximum value.**

You may wonder why we are studying the total amount of flow exiting the source, and not the total amount of flow entering the sink, which seems interesting as well. The next Theorem will clarify that the two quantities are equal, precisely because nobody steals across the pipeline.

Definition 137 (Cut)**.** A *cut* in an (s,t)-network is a partition of the vertex set into two sets, A and $V \setminus A$, such that the source s is in A and the sink t is in $V \setminus A$.
The *capacity* of the cut is the quantity

$$c(A) \stackrel{\text{def}}{=} \sum_{a\in A, b\notin A} c(a,b).$$

The *net flow out of the cut* is the quantity

$$f(A) \stackrel{\text{def}}{=} \sum_{a\in A, b\notin A} f(a,b) - \sum_{b\notin A, a\in A} f(b,a).$$

Note that the definition of cut we gave is not symmetric, i.e. A and $V - A$ cannot be exchanged, because only one of them contains the source. So it makes sense to identify the cut with the subset A of vertices that contains the source. Note also that for the capacity of the cut, we are only considering edges directed from A to $V - A$; we don't care about edges traveling in the opposite direction. But we do count them in the net flow out of the cut.

Theorem 138. *Let N be an (s,t)-network. The flow out of the cut $f(A)$ depends only on the network N, but not on the cut A. One has*

$$f(A) = \text{val}(f) \stackrel{\text{def}}{=} \sum_{w\in V} f(s,w).$$

Proof. First, two forewords. The (s,t)-network contains only one source, which is in A, and one sink, which is not in A. Hence any $a \in A$ different than s is either isolated or intermediate. In both cases, using either that the flow is zero on non-edges, or Kirchhoff's law, we obtain

$$\sum_{v\in V} f(a,v) - \sum_{v\in V} f(v,a) = 0 \quad \text{for any } a \in A - \{s\}. \tag{8}$$

The second foreword is that

$$\sum_{a\in A}\sum_{v\in A} f(a,v) = \sum_{a\in A}\sum_{v\in A} f(v,a). \tag{9}$$

This is the consequence of a mere reindexing trick: In fact, the same equality holds for any set A and any function $f : A \times A \to \mathbb{R}$.

We're done with the foreword. Now consider the quantity

$$\sum_{a\in A}\sum_{v\in V} \left[f(a,v) - f(v,a)\right].$$

We'll evaluate it in two ways:

- *By breaking "$a \in A$" into the cases "$a \in A - \{s\}$" and "$a = s$".* Using 8 at the step marked with an exclamation mark, and the fact that there is no edge (v,s), we get

$$\begin{aligned}\sum_{a\in A}\sum_{v\in V}[f(a,v) - f(v,a)] &= \sum_{a\in A-\{s\}}\sum_{v\in V}[f(a,v) - f(v,a)] + \sum_{v\in V}[f(s,v) - f(v,s)] = \\ &\stackrel{!}{=} 0 + \sum_{v\in V}[f(s,v) - f(v,s)] = \sum_{v\in V}[f(s,v) - 0] = \text{val}(f).\end{aligned} \tag{10}$$

- *By breaking "$v \in V$" into the cases "$v \in A$" and "$v \notin A$".* Using 9 at the step marked with an exclamation mark, we obtain

$$\begin{aligned}\sum_{a\in A}\sum_{v\in V}[f(a,v)-f(v,a)] &= \sum_{a\in A}\sum_{v\in A}[f(a,v)-f(v,a)] + \sum_{a\in A}\sum_{v\notin A}[f(a,v)-f(v,a)] = \\ &\overset{!}{=} \quad 0 \quad + \sum_{a\in A}\sum_{v\notin A}[f(a,v)-f(v,a)] = \\ &= \sum_{a\in A, v\notin A} f(a,v) - \sum_{a\in A, v\notin A} f(v,a) \overset{\text{def}}{=} f(A). \quad \square\end{aligned} \tag{11}$$

Corollary 139. *Let N be an (s,t)-network. For any flow f, and for any cut A,*

$$\mathrm{val}(f) \leq c(A).$$

Proof. By Theorem 138, $\mathrm{val}(f) = f(A)$. Using the capacity constraint at the step marked with an exclamation mark, we get

$$f(A) \overset{\text{def}}{=} \sum_{a\in A, b\notin A} f(a,b) - \sum_{b\notin A, a\in A} f(b,a) \leq \sum_{a\in A, b\notin A} f(a,b) \overset{!}{\leq} \sum_{a\in A, b\notin A} c(a,b) \overset{\text{def}}{=} c(A). \qquad \square$$

Unlike $f(A)$, the quantity $c(A)$ depends on the cut A chosen. Remember that we want to find the maximum flow. To get the best possible upper bound for the flow, we need to find cuts of smallest capacity. The next section contains a miraculous theorem, namely, that the maximu of $\mathrm{val}(f)$ over all possible flows always *equals* the minimum of $c(A)$ over all possible cuts.

5.1 Max-flow-min-cut and Ford-Fulkerson's algorithm

Given a flow in a network, we'd like to know if the flow can be improved. For this, it's important to look at every pair of vertices u, v, and understand if "we can ship more from u to v". The *obvious* way to look for improvement, is to check if $c(u,v) - f(u,v) > 0$. If so, then in principle we could push more material from u to v through the edge (u,v). But this is not the only way! There's another way we can improve things, namely, to check if by any chance $f(v,u) > 0$. That is, if the current flow is sending material backwards from v to u. Then we could "improve" simply by reducing the backward flow. (It's a bit like in business: To pay someone, you should figure out ways to wire him money!, but there's also another strategy: if by any chance he owes you money, you could 'virtually' send him money by condoning part of his debt.)

Definition 140 (Residual capacity)**.** Let f be a flow in an (s,t)-network $N = (V,E,c)$. The *residual network* is an (s,t)-network $R_f = (V,E,c_f)$ on the same vertices and edges of N, and with same source and sink, but with *residual capacity*

$$c_f(u,v) \overset{\text{def}}{=} c(u,v) - f(u,v) + f(v,u).$$

Note that c_f and R_f depend on f.

Remark 141. For any pair (u,v) of vertices, $c_f(u,v)$ is a rational number ≥ 0, because $(c(u,v) - f(u,v)) \geq 0$ and $f(v,u) \geq 0$. Moreover, $c_f(u,v)$ is 0 if and only if both $c(u,v) = f(u,v)$ and $f(v,u) = 0$.

Definition 142 (Augmenting path)**.** Let f be a flow in an (s,t)-network $N = (V,E,c)$. An *augmenting path* is a path in the residual network R_f where all edges have positive residual capacity.

Theorem 143 (Max-Flow-Min-Cut)**.** *For a fixed flow f in an (s,t)-network $N=(V,E,c)$, the following are equivalent:*

(1) *The flow has maximum value.*
(2) *There is no augmenting path from s to t.*
(3) *There is a cut A such that* $\mathrm{val}(f)=c(A)$.

Proof. (3) $\Rightarrow$ (1) You can't top that!, since for any other flow g we have $\mathrm{val}(g)\leq c(A)$ by Corollary 139.

(1) $\Rightarrow$ (2) Were there an augmenting path, you could improve the flow further! (In detail: if m is the smallest residual capacity among the edges involved in the augmenting path, send the quantity m from s to t! You'll have increased the flow value by m.)

(2) $\Rightarrow$ (3) The right idea is to define A as **the vertices you can reach from s with an augmenting path.** By definition $s\in A$. By the assumption (2), we know that $t\notin A$. So A is a well-defined cut. Now consider an edge a,b such that $a\in A$ and $b\notin A$. The residual capacity of (a,b) must be zero, otherwise the augmenting path from s to a (which exists because $a\in A$) could be extended one step further, yielding a path from s to b (which cannot exist because $b\notin A$). So

$$0=c_f(a,b)\stackrel{\text{def}}{=}(c(a,b)-f(a,b))+f(b,a).$$

By Remark 141, this means that $c(a,b)=f(a,b)$ and $f(b,a)=0$. But since this holds for every (a,b) across the cut, using Theorem 138 we get

$$\mathrm{val}(f)=f(A)\stackrel{\text{def}}{=}\sum_{a\in A,b\notin A}f(a,b)-\sum_{b\notin A,a\in A}f(b,a)=\sum_{a\in A,b\notin A}f(a,b)=\sum_{a\in A,b\notin A}c(a,b)\stackrel{\text{def}}{=}c(A).$$

□

The previous theorem is due to Lester Randolph Ford Jr. (1927–2017) and Delbert Ray Fulkerson (1924–1976). They also answered the natural, follow-up question: So how do we find the best flow? To see this, first let us introduce a natural operation within flows.

Definition 144. A flow in an (s,t)-network is "simplified' if for any two vertices u,v, one of $f(u,v)$ and $f(v,u)$ is zero. We can "simplify" any flow in the following way: while there are two vertices u,v such that $f(u,v)$ and $f(v,u)$ are both positive, create a new function f' that is identical to f on all other pair of vertices, but such that

$$f'(u,v)\stackrel{\text{def}}{=}\begin{cases}f(u,v)-f(v,u) & \text{if } f(u,v)>f(v,u)\\ 0 & \text{if } f(u,v)\leq f(v,u)\end{cases}$$

$$f'(v,u)\stackrel{\text{def}}{=}\begin{cases}0 & \text{if } f(u,v)>f(v,u)\\ f(v,u)-f(u,v) & \text{if } f(u,v)\leq f(v,u)\end{cases}$$

Clearly one of $f'(u,v)$ and $f'(v,u)$ is zero. Let us verify that f' is again a flow:

(1) Capacity constraint holds because with respect to f, we reduced the flow on both (u,v) and (v,u) (and left untouched the flow on all other edges). So $f'(u,v)\leq f(u,v)\leq c(u,v)$, and same for (v,u).

(2) Kirchhoff law holds because
- if $f(u,v)>f(v,u)$, then $f'(u,v)-f'(v,u)\stackrel{\text{def}}{=}(f(u,v)-f(v,u))-0=f(u,v)-f(v,u)$.
- if $f(u,v)\leq f(v,u)$, then $f'(u,v)-f'(v,u)\stackrel{\text{def}}{=}0-(f(v,u)-f(u,v))=f(u,v)-f(v,u)$.

So either way, $f'(u,v)-f'(v,u)=f(u,v)-f(v,u)$.

Ford–Fulkerson Algorithm (1954).

INPUT: Network $N = (V, E, c)$.
Initialize $f(u, v) := 0$ for all (u, v).

1. Compute the residual capacity c_f.
2. While there is an augmenting path $\wp$ from s to t:
   (i) let m be the minimum of the residual capacities of the edges along that path.
   (ii) Augment the flow at all edges in that path by m units; that is, define

$$f'(u, v) \stackrel{\text{def}}{=} \begin{cases} f(u, v) + m & \text{if } (u, v) \in \wp \\ f(u, v) & \text{otherwise.} \end{cases}$$

   (iii) Then, "simplify" this f'. If f'' is the resulting flow, replace f with f'' and go back to step 1.
3. Output f.

Remark 145. At the moment it is not clear whether this "procedure" eventually stops: A priori, it might keep running forever. However, for the moment notice that *if* this procedure stops, then we are in a situation where there is no augmenting path — so by Theorem 143 the output flow is certainly the optimal one (and equal to the smallest cut capacity).

Indeed, we will now show that the procedure above stops after a finite number of steps. For this it is crucial that the capacities are elements of $\mathbb{Q}$. Should you want to consider real-valued capacities, for whatever reason, then Theorem 143 would still hold!, but the Ford–Fulkerson procedure might keep running forever, and it might not converge to the maximum flow.

Recall that a function is *integral* if its image is contained in $\mathbb{N}$.

Theorem 146. *If the capacity is integral, then at each iteration of Ford–Fulkerson's algorithm the flow f is also integral. In particular, the Ford–Fulkerson's algorithm terminates after at most $c(\{s\})$ steps, returning an optimal flow that is* integral.

Proof. Initially $f(u, v) = 0$ for all (u, v), so in the first "while" cycle the residual capacity is equal to the capacity, so it is integral. Thus m will be an integer. Thus the flow f' will be integral. Finally, "simplifying" an integral flow still gives an integral flow (because the difference of two integers is still an integer), so f'' is also integral. Now at every iteration of the Ford–Fulkerson's algorithm, the flow improves by at least 1. Since we start with a flow of value 0 and the value of the value cannot exceed $c(\{s\})$ (by Corollary 139, applied to $A \stackrel{\text{def}}{=} \{s\}$), the total number of iterations is at most $c(\{s\})$. So the algorithm eventually stops, outputting an integral flow. □

Corollary 147. *Ford–Fulkerson terminates after finitely many steps.*

Proof. Remember that we define capacities to be functions from $V \times V$ to $\mathbf{Q}_{\geq \mathbf{0}}$. Our problem is clearly scale-invariant, i.e. we could change measure unit for the capacities, without changing the problem: the solution would be proportional. So since our capacities are a finite number of fractions, we can always rescale everything (i.e. multiply by the least common multiple M of all denominators) and assume that **all capacities are non-negative integers**. Applying the theorem, we see that Ford–Fulkerson then converges, after at most $M \cdot c(\{s\})$ steps. □

Ford–Fulkerson is an example of **greedy** algorithm: As soon as you see an augmenting path from s to t, you take it!, using it for increasing your flow. Concretely, remember that an augmenting path is a path consisting of:

- non-full forward edges (i.e., edges (u, v) such that $c(u, v) - f(u, v) > 0$) *or*
- non-empty backward edges (i.e., edges (v, u) such that $f(v, u) > 0$) *or*
- both!

Remark 148. Recall that $\text{val}(f) \leq c(A)$ for any cut A, by Corollary 139. So if by any chance, at any stage of the algorithm, you see a cut whose capacity *equals* the value of the flow you are considering, then you might as well stop, because clearly there's no way to beat that flow.

5.2 Consequences on Marriage

Max-flow-min-cut is a wonderful tool to prove other theorems! Let us start with an observation. Suppose we have **a network where the capacity of every edge is 1**, and the capacity of every non-edge is 0. Then by Theorem 146 the flow of maximum value is also integral. But for any edge (u, v), the constraints

$$\begin{cases} 0 \leq f(u, v) \leq c(u, v) = 1 \\ f(u, v) \in \mathbb{Z} \end{cases}$$

imply that $f(u, v)$ is either 0 or 1. So essentially, a flow can either use an edge, or not. Then in a network where the capacity of all edges is 1 a flow is merely a subset of the (directed) edges, walking on which you can get from s to t. And if $\text{val}\, f = t$, this means that our flow inside G consists of t edge-disjoint-paths from s to t.

In this setup, what does *cut capacity* mean? By definition,

$$c(A) \stackrel{\text{def}}{=} \sum_{a \in A, b \notin A} c(a, b) = \sum_{a \in A, b \notin A} 1,$$

so $c(A)$ just counts the number of edges that go out of the cut. So Theorem 143 says something we didn't know, about a maximum number of edge-disjoint path from s to t being equal to the smallest number of edges whose deletion destroys any path from s to t.

The next trick we are going to learn is, **what if we assign capacity L to some edges, where L is a very large integer**. Suppose you want to find a cut of capacity k, with k small. Then the edges of capacity L *cannot be cut*. In other words, if a cut A contains an endpoint of one of these "superedges", then A is forced to contain also the other endpoint. This has interesting applications.

Theorem 149 (Menger, 1927)**.** *Let G be a connected graph, with two distinct vertices s, t.*

- *The maximum number of edge-disjoint $s - t$ paths equals the minimum cardinality of a set of edges whose deletion destroys any path from s to t.*
- *The maximum number of vertex-disjoint $s - t$ paths equals the minimum cardinality of a set of vertices whose deletion destroys any path from s to t.*

Proof. The first statement is basically proven above; the details are left to you. We will focus on the second statement, which is harder to prove. The crucial idea is the following trick.

Let us transform every vertex v of G into a pair of vertices $(v_{\text{in}}, v_{\text{out}})$. Set $c(v_{\text{in}}, v_{\text{out}}) = 1$ and $c(v_{\text{out}}, v_{\text{in}}) = 1$. Let L be a very large number (for example, larger than twice the number of vertices of G). Every edge $\{u, v\}$ in G is transformed into two directed edges: one edge from u_{out} to v_{in}, to which we assign capacity L, and another symmetric edge from v_{out} to u_{in}, to which we also assign capacity L. This way from G we have obtained an $(s_{\text{out}}, t_{\text{in}})$-network N_G, of integral capacity. Now we apply max-flow-min cut. Finding a flow of value k in N_G is the same as finding k vertex-disjoint paths inside G. Moreover, since $k < L$, a cut of capacity k can only be crossed by edges of the type $(v_{\text{in}}, v_{\text{out}})$. So finding a cut of capacity k in N_G is the same as finding a set of k vertices in G whose deletion destroys any path from s to t. □

Corollary 150. *Let G be a graph with $n \geq d+1$ vertices.*
G is d-connected $\iff$ between any 2 non-adjacent vertices there are $\geq d$ vertex-disjoint paths.

Next comes another famous theorem in graph theory, proved by Philip Hall in 1935. It can be described as follows.

Suppose the population of a small, rural town consists of the same number n of men and women. Suppose in this town everybody belongs to just one of these two genders, everybody is heterosexual, and every marriage results in happiness. (Yes, even in 1935 England this was completely not realistic, but they liked to pretend it was.) In this town, each woman keeps a secret list of the men in town, any one of which she would happily marry; and any man would be happy to marry a woman who wants to marry him. Consider whether it is possible to pair up the men and the women (in marriage!, of course) so that every person is happy.

Theorem 151 (Hall's Marriage theorem)**.** *It is possible to pair up everybody happily in this town of freaks, if and only if for every subset S of the women, the total amount of guys written in their lists is at least $|S|$.*

Proof. The direction $\Rightarrow$ of this theorem is rather obvious. Namely, suppose not: suppose there's for example 7 women, say, that have indicated in their lists at most 6 guys altogether. Then there is no chance to subdivide the 6 guys among them: At least one woman will remain without partner (and thus, according to the 1935 axioms, apparently unhappy and inconsolable).

So let's focus on the direction $\Leftarrow$. We are going to create a network with vertex set

$$V = \{s\} \cup W \cup M \cup \{t\},$$

where the points in W represents women and the points in M, men. We are assuming $|W| = |M| = n$, so $|V| = 2n + 2$.

- First thing, for every $w \in W$, we draw a directed edge (s, w). Let's give it capacity 1.
- Next, we draw a directed edge (w, m) whenever a guy m is in the wishlist of the woman w. Let's give it capacity L, where L is a very large number – for example, $L = 2n$.
- Finally, for every $m \in M$, draw a directed edge (m, t) and give it capacity 1.

Okay, so now we are in business. Integral capacities, so the maximal flow is integral. What we want to prove is that **there is a flow of value $\mathbf{n} = |\mathbf{W}| = |\mathbf{M}|$**. In fact, this means finding n edge-disjoint paths from s to t, or equivalently, n disjoint woman-man edges, which would give the desired matching and restore British decency.

So suppose by contradiction that the maximal flow has value k, for some $k < n$. By max-flow-min-cut, this means that there is a cut A of capacity k. Since L is larger than k, you cannot break the women's preferences: all the edges going across the cut A must be of type (s, w) or (m, t). And since the capacities of these edges is 1, a cut A of size k means that

- you are cutting i edges of type (s, w), which means, "i women are not in A";
- you are cutting j edges of type (m, t), which means, "j men are not in $V - A$" (or better, "j men are in A");
- and $i + j = k$.

So summing up,

$$|W \cap (V - A)| + |M \cap A| = k. \tag{12}$$

On the other hand, obviously

$$|W \cap (V - A)| + |W \cap A| = |W| = n \tag{13}$$

and since we are assuming $k < n$, by comparing Equations 12 and 13 we conclude

$$|M \cap A| < |W \cap A|. \tag{14}$$

Now for $S \subset W$, let $\ell(S)$ be the list of men present in the preferences of the women in S altogether. By assumption, $|S| \leq |\ell(S)|$. Applying this to $S = W \cap A$, we conclude

$$|M \cap A| < |W \cap A| \leq |\ell(W \cap A)|.$$

But this is a contradiction: in fact, since no edge (w, m) goes across the cut, $\ell(W \cap A) \subset M \cap A$. And it cannot be that$M \cap A$ has smaller cardinality than a subset of himself. □

If you wish, look up the definition of "matching" and "vertex cover", and try to prove the following theorem by yourself:

Theorem 152 (König–Egervary)**.** *If G is a bipartite graph, the cardinality of a maximum matching in G equals the cardinality of a minimum vertex cover in G.*