LECTURE NOTES ON MATHEMATICAL INDUCTION

PETE L. CLARK

Contents

1. Introduction 1
2. The (Pedagogically) First Induction Proof 4
3. The (Historically) First(?) Induction Proof 5
4. Closed Form Identities 6
5. More on Power Sums 7
6. Inequalities 10
7. Extending binary properties to n-ary properties 12
8. Miscellany 13
9. One Theorem of Graph Theory 15
10. The Principle of Strong/Complete Induction 17
11. Solving Homogeneous Linear Recurrences 19
12. The Well-Ordering Principle 22
13. Upward-Downward Induction 24
14. The Fundamental Theorem of Arithmetic 25
14.1. Euclid's Lemma and the Fundamental Theorem of Arithmetic 25
14.2. Rogers' Inductive Proof of Euclid's Lemma 27
14.3. The Lindemann-Zermelo Inductive Proof of FTA 27
References 28

1. Introduction

Principle of Mathematical Induction for sets
Let S be a subset of the positive integers. Suppose that:
(i) $1 \in S$, and
(ii) $\forall\ n \in \mathbb{Z}^+, n \in S \implies n+1 \in S$.
Then $S = \mathbb{Z}^+$.

The intuitive justification is as follows: by (i), we know that $1 \in S$. Now apply (ii) with $n = 1$: since $1 \in S$, we deduce $1 + 1 = 2 \in S$. Now apply (ii) with $n = 2$: since $2 \in S$, we deduce $2 + 1 = 3 \in S$. Now apply (ii) with $n = 3$: since $3 \in S$, we deduce $3 + 1 = 4 \in S$. And so forth.

This is not a proof. (No good proof uses "and so forth" to gloss over a key point!) But the idea is as follows: we can keep iterating the above argument as many times as we want, deducing at each stage that since S contains the natural number which is one greater than the last natural number we showed that it contained. Now it

is a fundamental part of the structure of the positive integers that every positive integer can be reached in this way, i.e., starting from 1 and adding 1 sufficiently many times. In other words, any **rigorous definition** of the natural numbers (for instance in terms of sets, as alluded to earlier in the course) needs to incorporate, either implicitly or (more often) explicitly, the principle of mathematical induction. Alternately, the principle of mathematical induction is a key ingredient in any axiomatic characterization of the natural numbers.

It is not a key point, but it is somewhat interesting, so let us be a bit more specific. In Euclidean geometry one studies points, lines, planes and so forth, but one does not start by saying what sort of object the Euclidean plane "really is". (At least this is how Euclidean geometry has been approached for more than a hundred years. Euclid himself gave such "definitions" as: "A point is that which has position but not dimensions." "A line is breadth without depth." In the 19th century it was recognized that these are descriptions rather than definitions, in the same way that many dictionary definitions are actually descriptions: "cat: A small carnivorous mammal domesticated since early times as a catcher of rats and mice and as a pet and existing in several distinctive breeds and varieties." This helps you if you are already familiar with the animal but not the word, but if you have never seen a cat before this definition would certainly not allow you to determine with certainty whether any particular animal you encountered was a cat, and still less would it allow you to reason abstractly about the cat concept or "prove theorems about cats.") Rather "point", "line", "plane" and so forth are taken as **undefined terms**. They are related by certain **axioms**, or abstract properties that they must satisfy.

In 1889, the Italian mathematician and proto-logician Gisueppe Peano came up with a similar (and, in fact, much simpler) system of axioms for the natural numbers. In slightly modernized form, this goes as follows:

The undefined terms are **zero**, **number** and **successor**.

There are five axioms that they must satisfy, the **Peano axioms**. The first four are:

(P1) Zero is a number.
(P2) Every number has a successor, which is also a number.
(P3) No two distinct numbers have the same successor.
(P4) Zero is not the successor of any number.

Using set-theoretic language we can clarify what is going on here as follows: the structures we are considering are triples $(X, 0, S)$, where X is a set, 0 is an element of X, and $S : X \to X$ is a function, subject to the above axioms.

From this we can deduce quite a bit. First, we have a number (i.e., an element of X) called $S(0)$. Is $0 = S(0)$? No, that is prohibited by (P4). We also have a number $S(S(0))$, which is not equal to 0 by (P4) and it is also not equal to $S(0)$, because then $S(0) = S(S(0))$ would be the successor of the distinct numbers 0 and $S(0)$, contradicting (P3). Continuing in this way, we can produce an infinite

sequence of distinct elements of X:

$$0, S(0), S(S(0)), S(S(S(0)), \dots. \tag{1}$$

In particular X itself is infinite. The crux of the matter is this: is there any element of X which is *not* a member of the sequence (1), i.e., is not obtained by starting at 0 and applying the successor function finitely many times?

The axioms so far do not allow us to answer this question. For instance, suppose that the "numbers" consisted of the set $[0, \infty)$ of all non-negative real numbers, we define 0 to be the real number of that name, and we define the successor of x to be $x + 1$. This system satisfies (P1) through (P4) but has much more in it than just the natural numbers we want, so we must be missing an axiom! Indeed, the last axiom is:

(P5) If Y is a subset of the set X of numbers such that $0 \in Y$ and such that $x \in Y$ implies $S(x) \in Y$, then $Y = X$.

Notice that the example we cooked up above fails (P5), since in $[0, \infty)$ the subset of natural numbers contains zero and contains the successor of each of its elements but is a proper subset of $[0, \infty)$.

Thus it was Peano's contribution to realize that mathematical induction is an axiom for the natural numbers in much the same way that the parallel postulate is an axiom for Euclidean geometry.

On the other hand, it is telling that this work of Peano is little more than one hundred years old, which in the scope of mathematical history is quite recent. Traces of what we now recognize as induction can be found from the mathematics of antiquity (including Euclid's Elements!) on forward. According to the (highly recommended!) Wikipedia article on mathematical induction, the first mathematician to formulate it explicitly was Blaise Pascal, in 1665. During the next hundred years various equivalent versions were used by different mathematicians – notably the methods of infinite descent and minimal counterexample, which we shall discuss later – and the technique seems to have become commonplace by the end of the 18th century. Not having an formal understanding of the relationship between mathematical induction and the structure of the natural numbers was not much of a hindrance to mathematicians of the time, so still less should it stop us from learning to use induction as a proof technique.

Principle of mathematical induction for predicates
Let $P(x)$ be a sentence whose domain is the positive integers. Suppose that:
(i) $P(1)$ is true, and
(ii) For all $n \in \mathbb{Z}^+$, $P(n)$ is true $\implies$ $P(n+1)$ is true.
Then $P(n)$ is true for all positive integers n.

Variant 1: Suppose instead that $P(x)$ is a sentence whose domain is the natural numbers, i.e., with zero included, and in the above principle we replace (i) by the assumption that $P(0)$ is true and keep the assumption (ii). Then of course the conclusion is that $P(n)$ is true for all natural numbers n. This is more in accordance

with the discussion of the Peano axioms above.[1]

Exercise 1: Suppose that N_0 is a fixed integer. Let $P(x)$ be a sentence whose domain contains the set of all integers $n \geq N_0$. Suppose that:
(i) $P(N_0)$ is true, and
(ii) For all $n \geq N_0$, $P(n)$ is true $\implies$ $P(n+1)$ is true.
Show that $P(n)$ is true for all integers $n \geq N_0$. (Hint: define a new predicate $Q(n)$ with domain $\mathbb{Z}^+$ by making a "change of variables" in P.)

2. The (Pedagogically) First Induction Proof

There are many things that one can prove by induction, but the first thing that everyone proves by induction is invariably the following result.

Proposition 2.1. *For all $n \in \mathbb{Z}^+$, $1 + \ldots + n = \frac{n(n+1)}{2}$.*

Proof. We go by induction on n.
Base case ($n = 1$): Indeed $1 = \frac{1(1+1)}{2}$.
Induction step: Let $n \in \mathbb{Z}^+$ and suppose that $1 + \ldots + n = \frac{n(n+1)}{2}$. Then

$$1 + \ldots + n + n + 1 = (1 + \ldots + n) + n + 1 \stackrel{\text{IH}}{=} \frac{n(n+1)}{2} + n + 1$$

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

Here the letters "IH" signify that the induction hypothesis was used. □

Induction is such a powerful tool that once one learns how to use it one can prove many nontrivial facts with essentially no thought or ideas required, as is the case in the above proof. However thought and ideas are good things when you have them! In many cases an inductive proof of a result is a sort of "first assault" which raises the challenge of a more insightful, noninductive proof. This is certainly the case for Proposition 2.1 above, which can be proved in many ways.

Here is one non-inductive proof: replacing n by $n-1$, it is equivalent to show:

$$\forall n \in \mathbb{Z}, n \geq 2 : \; 1 + \ldots + n - 1 = \frac{(n-1)n}{2}. \tag{2}$$

We recognize the quantity on the right-hand side as the **binomial coefficient** $\binom{n}{2}$: it counts the number of 2-element subsets of an n element set. This raises the prospect of a **combinatorial proof**, i.e., to show that the number of 2-element subsets of an n element set is *also* equal to $1 + 2 + \ldots + n - 1$. This comes out immediately if we list the 2-element subsets of $\{1, 2, \ldots, n\}$ in a systematic way: we may write each such subset as $\{i, j\}$ with $1 \leq i \leq n-1$ and $i < j \leq n$. Then:

The subsets with least element 1 are $\{1,2\}, \{1,3\}, \ldots, \{1,n\}$, a total of $n-1$.
The subsets with least element 2 are $\{2,3\}, \{2,4\}, \ldots, \{2,n\}$, a total of $n-2$.
$\vdots$
The subset with least element $n-1$ is $\{n-1, n\}$, a total of 1.

[1] In fact Peano's original axiomatization did not include zero. What we presented above is a standard modern modification which is slightly cleaner to work with.

Thus the number of 2-element subsets of $\{1, \ldots, n\}$ is on the one hand $\binom{n}{2}$ and on the other hand $(n-1) + (n-2) + \ldots + 1 = 1 + 2 + \ldots + n - 1$. This gives a combinatorial proof of Proposition 2.1.

For a very striking pictorial variation of the above argument, go to `http://mathoverflow.net/questions/8846/proofs-without-words` and scroll down to the first diagram.

3. The (Historically) First(?) Induction Proof

Theorem 3.1. *(Euclid) There are infinitely many prime numbers.*

Proof. For $n \in \mathbb{Z}^+$, let $P(n)$ be the assertion that there are at least n prime numbers. Then there are infinitely many primes if and only if $P(n)$ holds for all positive integers n. We will prove the latter by induction on n.
Base Case ($n = 1$): We need to show that there is at least one prime number. For instance, 2 is a prime number.
Induction Step: Let $n \in \mathbb{Z}^+$, and assume that $P(n)$ holds, i.e., that there are at least n prime numbers $p_1 < \ldots < p_n$. We need to show that $P(n+1)$ holds, i.e., there is at least one prime number different from the numbers we have already found. To establish this, consider the quantity

$$N_n = p_1 \cdots p_n + 1.$$

Since $p_1 \cdots p_n \geq p_1 \geq 2$, $N_n \geq 3$. In particular it is divisible by at least one prime number, say q.[2] But I claim that N_n is not divisible by p_i for any $1 \leq i \leq n$. Indeed, if $N_n = ap_i$ for some $a \in \mathbb{Z}$, then let $b = \frac{p_1 \cdots p_n}{p_i} \in \mathbb{Z}$. Then $kp_i = p_1 \cdots p_n + 1 = bp_i + 1$, so $(k-b)p_i = 1$ and thus $p_i = \pm 1$, a contradiction. So if we take q to be, for instance, the smallest prime divisor of N_n, then there are at least $n+1$ prime numbers: $p_1, \ldots, p_n, q$. □

Remark: The proof that there are infinitely many prime numbers first appeared in Euclid's *Elements* (Book IX, Proposition 20). Euclid did not explicitly use induction (no ancient Greek mathematician did), but in retrospect his proof is clearly an inductive argument: what he does is to explain, as above, how given any finite list $p_1, \ldots, p_n$ of distinct primes, one can produce a new prime which is not on the list. (In particular Euclid *does not* verify the base case, and he must have regarded it as obvious that there is at least one prime number. And it is – but it should be included as part of the proof anyway!) What is strange is that in our day Euclid's proof is generally *not* seen as a proof by induction. Rather, it is often construed as a classic example of a proof by contradiction – which it isn't! Rather, Euclid's argument is perfectly contructive. Starting with any given prime number – say $p_1 = 2$ – and following his procedure, one generates an infinite sequence of primes. For instance, $N_1 = 2 + 1 = 3$ is prime, so we take $p_2 = 3$. Then $N_2 = 2 \cdot 3 + 1 = 7$ is again prime, so we take $p_3 = 7$. Then $N_3 = 2 \cdot 3 \cdot 7 + 1 = 43$ is also prime, so we take $p_4 = 43$. But this time something more interesting happens:

$$N_4 = 2 \cdot 3 \cdot 7 \cdot 43 + 1 = 13 \cdot 139$$

[2]Later in these notes we will prove the stronger fact that any integer greater than one may be expressed as a product of primes. For now we assume this (familiar) fact.

is *not* prime.[3] For definiteness let us take p_5 to be the smallest prime factor of N_4, so $p_5 = 13$. In this way we generate an infinite sequence of prime numbers – so the proof is unassailably constructive.

By the way, this sequence of prime numbers is itself rather interesting. It is often called the **Euclid-Mullin sequence**, after Albert A. Mullin who asked questions about it in 1963 [Mu63]. The next few terms are

$$53, 5, 6221671, 38709183810571, 139, 2801, 11, 17, 5471, 52662739, 23003,$$

$$30693651606209, 37, 1741, 1313797957, 887, 71, 7127, 109, 23, \ldots.$$

Thus one can see that it is rather far from just giving us all of the prime numbers in increasing order! In fact, since to find p_{n+1} we need to factor $N_n = p_1 \cdots p_n + 1$, a quantity which rapidly increases with n, it is in fact quite difficult to compute the terms of this sequence, and as of 2010 only the first 47 terms are known. Perhaps Mullin's most interesting question about this sequence is: does every prime number appear in it eventually? This is an absolutely open question. At the moment the smallest prime which is not known to appear in the Euclid-Mullin sequence is 31.

Remark: Some scholars have suggested that what is essentially an argument by mathematical induction appears in the later middle Platonic dialogue *Parmenides*, lines 149a7-c3. But this argument is of mostly historical and philosophical interest. The statement in question is, very roughly, that if n objects are placed adjacent to another in a linear fashion, the number of points of contact between them is $n - 1$. (Maybe. To quote the lead in the wikipedia article on the *Parmenides*: "It is widely considered to be one of the more, if not the most, challenging and enigmatic of Plato's dialogues.") There is not much mathematics here! Nevertheless, for a thorough discussion of induction in the *Parmenides* the reader may consult [Ac00] and the references cited therein.

4. Closed Form Identities

The inductiive proof of Proposition 2.1 is a prototype for a certain kind of induction proof (the easiest kind!) in which $P(n)$ is some algebraic identity: say $LHS(n) = RHS(n)$. In this case to make the induction proof work you need only (i) establish the base case and (ii) verify the equality of successive differences

$$LHS(n+1) - LHS(n) = RHS(n+1) - RHS(n).$$

We give two more familiar examples of this.

Proposition 4.1. *For all $n \in \mathbb{Z}^+$, $1 + 3 + \ldots + (2n-1) = n^2$.*

Proof. Let $P(n)$ be the statement "$1 + 3 + \ldots + (2n-1) = n^2$". We will show that $P(n)$ holds for all $n \in \mathbb{Z}^+$ by induction on n.
Base case $n = 1$: indeed $1 = 1^2$.
Induction step: Let n be an arbitrary positive integer and assume $P(n)$:

$$1 + 3 + \ldots + (2n-1) = n^2. \tag{3}$$

Adding $2(n+1) - 1 = 2n + 1$ to both sides, we get

$$(1 + 3 + \ldots + (2n-1)) + 2(n+1) - 1 = n^2 + 2(n+1) - 1 = n^2 + 2n + 1 = (n+1)^2.$$

[3]Many mathematical amateurs seem to have the idea that $N_n = p_1 \cdots p_n + 1$ is always prime, but clearly it isn't.

This is precisely $P(n+1)$, so the induction step is complete. □

Proposition 4.2. *For all* $n \in \mathbb{Z}^+$, $1^2 + 2^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}$.

Proof. By induction on n.
Base case: $n = 1$.
Induction step: Let $n \in \mathbb{Z}^+$ and suppose that $1^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}$. Then

$$1 + \ldots + n^2 + (n+1)^2 \stackrel{\text{IH}}{=} \frac{n(n+1)(2n+1)}{6} + (n+1)^2 =$$
$$\frac{2n^3 + 3n^2 + n + 6 + 6n^2 + 12n + 1}{6} = \frac{2n^3 + 9n^2 + 13n + 7}{6}.$$

On the other hand, expanding out $\frac{(n+1)((n+1)+1)(2(n+1)+1)}{6}$, we also get $\frac{2n^3+9n^2+13n+7}{6}$. □

Often a non-inductive proof, when available, offers more insight. Again returning to our archetypical example: $1 + \ldots + n$, it is time to tell the story of little Gauss. As a child of no more than 10 or so, Gauss and his classmates were asked to add up the numbers from 1 to 100. Most of the students did this by a laborious calculation and got incorrect answers in the end. Gauss reasoned essentially as follows: put

$$S_n = 1 + \ldots + (n-1) + n.$$

Of course the sum is unchanged if we we write the terms in descending order:

$$S_n = n + (n-1) + \ldots + 2 + 1.$$

Adding the two equations gives

$$2S_n = (n+1) + (n+1) + \ldots + (n+1) = n(n+1),$$

so

$$S_n = \frac{n(n+1)}{2}.$$

This is no doubt preferable to induction, so long as one is clever enough to see it.

Mathematical induction can be viewed as a particular incarnation of a much more general proof technique: try to solve your problem by reducing it to a previously solved problem. A more straightforward application of this philosophy allows us to deduce Proposition 4.1 from Proposition 2.1:

$$1+3+\ldots+(2n-1) = \sum_{i=1}^{n}(2i-1) = 2\sum_{i=1}^{n} i - \sum_{i=1}^{n} 1 = 2\left(\frac{n(n+1)}{2}\right) - n = n^2 + n - n = n^2.$$

5. More on Power Sums

Suppose now we want to find a formula for $\sum_{i=1}^{n} i^3 = 1^3 + \ldots + n^3$.[4] A key point: we can't use induction yet because we don't know what the answer is! (As we will see again and again, this is, like Kryptonite for Superman, induction's only weakness.)

So let's try to actually think about what's going on. We previously found a formula

[4]Why might we want this? For instance, such sums arise in calculus as Riemann sums for the integral $\int_a^b x^3 dx$. Of course there is a better way to evaluate such integrals, via the Fundamental Theorem of Calculus. Perhaps it is safest to say that finding closed formulas for sums is an intrinsically interesting, and often quite challenging, endeavor.

for $\sum_{i=1}^n i$ which was a quadratic polynomial in n, and then a formula for $\sum_{i=1}^n i^2$ which was a cubic polynomial in n. We might therefore guess that the desired formula for $\sum_{i=1}^n i^3$ is a fourth degree polynomial in n, say

$$a_4n^4 + a_3n^3 + a_2n^2 + a_1n + a_0.$$

If we think more seriously about Riemann sums, the fundamental theorem of calculus and the fact that $\frac{x^4}{4}$ is an antiderivative for x^3, this guess becomes more likely, and we can even guess that $a_4 = \frac{1}{4}$. Also by looking at the other examples we might guess that $a_0 = 0$. So we are looking for (presumably rational?) numbers a_1, a_2, a_3 such that

$$1^3 + \ldots + n^3 = \frac{1}{4}n^4 + a_3n^3 + a_2n^2 + a_1n.$$

Now, inspired by the partial fractions technique in calculus, we can simply plug in a few values and solve for the coefficients. For instance, taking $n = 1, 2, 3$ we get

$$1^3 = 1 = \frac{1}{4} + a_3 + a_2 + a_1,$$

$$1^3 + 2^3 = 9 = 4 + 8a_3 + 4a_2 + 2a_1,$$

$$1^3 + 2^3 + 3^3 = 9 + 3^3 = 36 = \frac{81}{4} + 27a_3 + 9a_2 + 3a_1.$$

This gives us the linear system

$$a_1 + a_2 + a_3 = \frac{3}{4}$$

$$2a_1 + 4a_2 + 8a_3 = 5$$

$$3a_1 + 9a_2 + 27a_3 = \frac{63}{4}.$$

I will leave it to you to do the math here, in what way seems best to you.[5] The unique solution is $a_1 = 0,\ a_2 = \frac{1}{4},\ a_3 = \frac{1}{2}$, so that our conjectural identity is

$$1^3 + \ldots + n^3 = \frac{n^4}{4} + \frac{n^3}{2} + \frac{n^2}{4} = \frac{n^2}{4}(n^2 + 2n + 1) = \left(\frac{n(n+1)}{2}\right)^2.$$

Exercise 2: Prove (by induction, of course) that this identity is in fact correct.

Exercise 3: Use a similar technique to find a closed form expression for $\sum_{i=1}^n i^4$.

The above method is a useful one for solving many types of problems: make a guess as to the general form the answer may take, plug that guess in and fine tune the constants accordingly. In this case the method has two limitations: first, it involves a rather large amount of calculation, and second we cannot find out whether our general guess is correct until after all the calculations have been made. In this case, there is a better way to derive formulas for the power sums

$$S_d(n) = 1^d + \ldots + n^d.$$

We begin with the sum

$$S = \sum_{i=1}^n \left((i+1)^{d+1} - i^{d+1}\right),$$

[5]Yes, this is an allusion to *The Return of the King*.

which we evaluate in two different ways. First, writing out the terms gives

$$S = 2^{d+1} - 1^{d+1} + 3^{d+1} - 2^{d+1} + \ldots + n^{d+1} - (n-1)^{d+1} + (n+1)^{d+1} - n^{d+1} = (n+1)^{d+1} - 1.$$

Second, by first expanding out the binomial $(i+1)^{d+1}$ we get

$$S = \sum_{i=1}^{n} \left((i+1)^{d+1} - i^{d+1} \right) = \sum_{i=1}^{n} \left(i^{d+1} + \binom{d+1}{1} i^d + \ldots + \binom{d+1}{d} i + 1 - i^{d-1} \right) =$$

$$\sum_{i=1}^{n} \left(\binom{d+1}{1} i^d + \ldots + \binom{d+1}{d} i \right) = \binom{d+1}{1} \sum_{i=1}^{n} i^d + \ldots + \binom{d+1}{d} \sum_{i=1}^{n} i + \sum_{i=1}^{n} 1 =$$

$$\sum_{j=0}^{d} \binom{d+1}{d+1-j} S_j(n) = \sum_{j=0}^{d} \binom{d+1}{j} S_j(n).$$

Equating our two expressions for S, we get

$$(n+1)^{d+1} - 1 = \sum_{j=0}^{d} \binom{d+1}{j} S_j(n).$$

Solving this equation for $S_d(n)$ gives

$$S_d(n) = \frac{(n+1)^{d+1} - \left(\sum_{j=0}^{d-1} \binom{d+1}{j} S_j(n) \right) - 1}{(d+1)}. \tag{4}$$

This formula allows us to compute $S_d(n)$ recursively: that is, given exact formulas for $S_j(n)$ for all $0 \leq j < d$, we get an exact formula for $S_d(n)$. And getting the ball rolling is easy: $S_0(n) = 1^0 + \ldots + n^0 = 1 + \ldots 1 = n$.

Example ($d = 1$): Our formula gives

$$1 + \ldots + n = S_1(n) = (\frac{1}{2})((n+1)^2 - S_0(n) - 1) = (\frac{1}{2})(n^2 + 2n + 1 - n - 1) = \frac{n(n+1)}{2}.$$

Example ($d = 2$): Our formula gives $1^2 + \ldots + n^2 = S_2(n) =$

$$\frac{(n+1)^3 - S_0(n) - 3S_1(n) - 1}{3} = \frac{n^3 + 3n^2 + 3n + 1 - n - \frac{3}{2}n^2 - \frac{3}{2}n - 1}{3} =$$

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

Our formula (4) also has theoretical applications: with it in hand we can apply induction to a more worthy goal, namely the proof of the following result.

Theorem 5.1. *For every positive integer d, there exist $a_1, \ldots, a_d \in \mathbb{Q}$ such that for all $n \in \mathbb{Z}^+$ we have*

$$1^d + \ldots + n^d = \frac{n^{d+1}}{d+1} + a_d n^d + \ldots + a_1 n.$$

Proof. Exercise 4. □

6. Inequalities

I remind the reader that *for me*, $\mathbb{N}$ denotes the non-negative integers $\{0, 1, 2, 3, \ldots\}$.

Proposition 6.1. *For all $n \in \mathbb{N}$, $2^n > n$.*

Proof analysis: For $n \in \mathbb{N}$, let $P(n)$ be the statement "$2^n > n$". We want to show that $P(n)$ holds for all natural numbers n by induction.
Base case: $n = 0$: $2^0 = 1 > 0$.

Induction step: let n be an arbitrary natural number and asusme $P(n)$: $2^n > n$. Then

$$2^{n+1} = 2 \cdot 2^n > 2 \cdot n.$$

We would now like to say that $2n \geq n + 1$. But in fact this is true if and only if $n \geq 1$. Well, don't panic. We just need to restructure the argument a bit: we verify the statement separately for $n = 0$ and then use $n = 1$ as the base case of our induction argument. Here is a formal writeup:

Proof. Since $2^0 = 1 > 0$ and $2^1 = 2 > 1$, it suffices to verify the statement for all natural numbers $n \geq 2$. We go by induction on n.
Base case: $n = 2$: $2^2 = 4 > 2$.

Induction step: Assume that for some natural number $n \geq 2$ we have $2^n > n$. Then

$$2^{n+1} = 2 \cdot 2^n > 2 \cdot n > n + 1,$$

since $n > 1$. □

Exercise 5: Use calculus to show that in fact $2^x > x$ for all real x. (To see what's going on, it will be very helpful to graph the two functions. Of course, merely drawing a picture will not be a sufficient proof.)

Proposition 6.2. *There exists $N_0 \in \mathbb{Z}^+$ such that for all $n \geq N_0$, $2^n \geq n^3$.*

Proof analysis: A little experimentation shows that there are several small values of n such that $2^n < n^3$: for instance $2^9 = 512 < 9^3 = 729$. On the other hand, it seems to be the case that we can take $N_0 = 10$: let's try.
Base case: $n = 10$: $2^{10} = 1024 > 1000 = 10^3$.

Induction step: Suppose that for some $n \geq 10$ we have $2^n \geq n^3$. Then

$$2^{n+1} = 2 \cdot 2^n \geq 2n^3.$$

Our task is then to show that $2n^3 \geq (n+1)^3$ for all $n \geq 10$. (By considering limits as $n \to \infty$, it is certainly the case that the left hand side exceeds the right hand side for all sufficiently large n. It's not guaranteed to work for $n \geq 10$; if not, we will replace 10 with some larger number.) Now,

$$2n^3 - (n+1)^3 = 2n^3 - n^3 - 3n^2 - 3n - 1 = n^3 - 3n^2 - 3n - 1 \geq 0$$

$$\iff n^3 - 3n^2 - 3n \geq 1.$$

Since everything in sight is a whole number, this is in turn equivalent to

$$n^3 - 3n^2 - 3n > 0.$$

Now $n^3 - 3n^2 - 3n = n(n^2 - 3n - 3)$, so this is equivalent to $n^2 - 3n - 3 \geq 0$. The roots of the polynomial $x^2 - 3x - 3$ are $x = \frac{3 \pm \sqrt{21}}{2}$, so $n^2 - 3n - 3 > 0$ if $n > 4 = \frac{3+\sqrt{25}}{2} > \frac{3+\sqrt{21}}{2}$. In particular, the desired inequality holds if $n \geq 10$, so by induction we have shown that $2^n \geq n^3$ for all $n \geq 10$.

We leave it to to the student to convert the above analysis into a formal proof.

Remark: More precisely, $2^n \geq n^3$ for all natural numbers n *except* $n = 2, 3, 4, 6, 7, 8, 9$. It is interesting that the desired inequality is true for a little while (i.e., at $n = 0, 1$) then becomes false for a little while longer, and then becomes true for all $n \geq 10$. Note that it follows from our analysis that if for any $N \geq 4$ we have $2^N \geq N^3$, then this equality remains true for all larger natural numbers n. Thus from the fact that $2^9 < 9^3$, we can in fact deduce that $2^n < n^3$ for all $4 \leq n \leq 8$.

Proposition 6.3. *For all* $n \in \mathbb{Z}^+$, $1 + \frac{1}{4} + \ldots + \frac{1}{n^2} \leq 2 - \frac{1}{n}$.

Proof analysis: By induction on n.
Base case ($n = 1$): $1 \leq 2 - \frac{1}{1}$.

Induction step: Assume that for some $n \in \mathbb{Z}^+$ we have $1 + \frac{1}{4} + \ldots + \frac{1}{n^2} \leq 2 - \frac{1}{n}$. Then

$$1 + \frac{1}{4} + \ldots + \frac{1}{n^2} + \frac{1}{(n+1)^2} \leq 2 - \frac{1}{n} + \frac{1}{(n+1)^2}.$$

We want the left hand side to be less than $2 - \frac{1}{n+1}$, so it will suffice to establish the inequality

$$2 - \frac{1}{n} + \frac{1}{(n+1)^2} < 2 - \frac{1}{n+1}.$$

Equivalently, it suffices to show

$$\frac{1}{n+1} + \frac{1}{(n+1)^2} < \frac{1}{n}.$$

But we have

$$\frac{1}{n+1} + \frac{1}{(n+1)^2} = \frac{n+1+1}{(n+1)^2} = \frac{n+2}{(n+1)^2}.$$

Everything in sight is positive, so by clearing denominators, the desired inequality is equivalent to

$$n^2 + 2n = n(n+2) < (n+1)^2 = n^2 + 2n + 1,$$

which, at last, is a true inequality. Thus we have all the ingredients of an induction proof, but again we need to put things together in proper order, a task which we leave to the reader.

Remark: Taking limits as $n \to \infty$, it follows that $\sum_{n=1}^{\infty} \frac{1}{n^2} \leq 2$. In particular, this argument shows that the infinite series converges. The exact value of the sum is, in fact, $\frac{\pi^2}{6}$. A proof of this requires techniques from advanced calculus.

7. Extending binary properties to n-ary properties

Example: All horses have the same color.

Proposed proof: There are only finitely many horses in the world, so it will suffice to show that for all $n \in \mathbb{Z}^+$, $P(n)$ holds, where $P(n)$ is the statement that in any set of n horses, all of them have the same color.

Base case: In any set S of one horse, all of the horses in S have the same color!

Induction step: We suppose that for some positive integer n, in any set of n horses, all horses have the same color. Consider now a set of $n+1$ horses, which for specificity we label $H_1,\ H_2, \ldots, H_n,\ H_{n+1}$. Now we can split this into two sets of n horses:

$$S = \{H_1, \ldots,\ H_n\}$$

and

$$T = \{H_2, \ldots, H_n,\ H_{n+1}\}.$$

By induction, every horse in S has the same color as H_1: in particular H_n has the same color as H_1. Similarly, every horse in T has the same color as H_n: in particular H_{n+1} has the same color as H_n. But this means that $H_2, \ldots, H_n, H_{n+1}$ all have the same color as H_1. It follows by induction that for all $n \in \mathbb{Z}^+$, in any set of n horses, all have the same color.

Proof analysis: Naturally one suspects that there is a mistake somewhere, and there is. However it is subtle, and occurs in a perhaps unexpected place. In fact the argument is completely correct, except the induction step is not valid when $n = 1$: in this case $S = \{H_1\}$ and $T = \{H_2\}$ and these two sets are disjoint: they have no horses in common. We have been misled by the "dot dot dot" notation which suggests, erroneously, that S and T must have more than one element.

In fact, if only we could establish the argument for $n = 2$, then the proof goes through just fine. For instance, the result can be fixed as follows: if in a finite set of horses, any two have the same color, then they all have the same color.

There is a moral here: one should pay especially close attention to the smallest values of n to make sure that the argument has no gaps. On the other hand, there is a certain type of induction proof for which the $n = 2$ case is the most important (often it is also the base case, but not always), and the induction step is easy to show, but uses once again the $n = 2$ case. Here are some examples of this.

The following is a fundamental fact of number theory, called **Euclid's Lemma**.

Proposition 7.1. *Let p be a prime number, and $a, b \in \mathbb{Z}^+$. If $p \mid ab$, $p \mid a$ or $p \mid b$.*

Later in these notes we will give a proof of Euclid's Lemma (yes, by induction!). For now we simply assume it to be true. Our point is that we can swiftly deduce the following useful generalization.

Proposition 7.2. *Let p be a prime number, $n \in \mathbb{Z}^+$ and $a_1, \ldots, a_n \in \mathbb{Z}^+$. If $p \mid a_1 \cdots a_n$, then $p \mid a_i$ for some $1 \leq i \leq n$.*

Proof. This is trivial for $n = 1$. We show that it holds for all $n \geq 2$ by induction.
Base case: $n = 2$: This is precisely Euclid's Lemma.

Induction Step: We assume that for a given $n \in \mathbb{Z}^+$ and $a_1, \ldots, a_n \in \mathbb{Z}^+$, if a prime p divides the product $a_1 \cdots a_n$, then it divides at least one a_i. Now let $a_1, \ldots, a_n, a_{n+1} \in \mathbb{Z}$, and that a prime p divides $a_1 \cdots a_n a_{n+1}$. Then $p \mid (a_1 \cdots a_n)a_{n+1}$, so by Euclid's Lemma, $p \mid a_1 \cdots a_n$ or $p \mid a_{n+1}$. If the latter holds, we're done. If the former holds, then by our inductive hypothesis, $p \mid a_i$ for some $1 \leq i \leq n$, so we are also done. □

Comment: In this and other induction proofs of this type, it is the base case which is nontrivial, and the induction step is essentially the same argument every time.

Corollary 7.3. *Let p be a prime, $n \in \mathbb{Z}^+$, $a \in \mathbb{Z}^+$ such that $p \mid a^n$. Then $p \mid a$.*

Exercise 6: Use Corollary 7.3 to show that for any prime p, $p^{\frac{1}{n}}$ is irrational.

Proposition 7.4. *Let $n \geq 3$ be an integer, and let $f_1, \ldots, f_n : \mathbb{R} \to \mathbb{R}$ be differentiable functions. Then*

$$(f_1 \cdots f_n)' = f_1' f_2 \cdots f_n + f_1 f_2' \cdots f_n + \ldots + f_1 \cdots f_{n-1} f_n'.$$

Proof. We argue by induction on n.
Base case ($n = 2$): The assertion is $(f_1 f_2)' = f_1' f_2 + f_1 f_2'$, which is the product rule from differential calculus.

Induction step: We assume the result is true for any n differentiable functions. If $f_1, \ldots, f_{n+1}$ are all differentiable, then

$$(f_1 \cdots f_n f_{n+1})' = ((f_1 \cdots f_n) f_{n+1})' \stackrel{*}{=} (f_1 \cdots f_n)' f_{n+1} + f_1 \cdots f_n f_{n+1}' =$$

$$(f_1' f_2 \cdots f_n) f_{n+1} \stackrel{**}{=} f_1 f_2' f_3 \cdots f_n f_{n+1} + \ldots + f_1 \cdots f_{n-1} f_n' f_{n+1} + f_1 \cdots f_n f_{n+1}'.$$

Note that in the first starred equality we have applied the usual product rule and in the second starred equality we have applied the induction hypothesis. □

Corollary 7.5. *For any positive integer n, if $f(x) = x^n$, then $f'(x) = nx^{n-1}$.*

Proof. Exercise 7. □

When teaching freshman calculus, it is very frustrating not to be able to prove the power rule by this simple inductive argument!

8. Miscellany

Proposition 8.1. *Let S be a finite set. Then $\#\mathcal{P}(S) = 2^{\#S}$.*

Proof. Let $n = \#S$. We go by induction on n.
Base case ($n = 0$): If $\#S = 0$, then $S = \varnothing$ and $\mathcal{P}(S) = \{\varnothing\}$, so $\#\mathcal{P}(S) = 1 = 2^{\#S}$.

Induction step: assume the result holds for any finite set with n elements, and let S be a set with $n + 1$ elements. In particular S is nonempty, so choose $x \in S$. Define

$$\mathcal{P}_1 = \{T \subset S \mid x \in T\},$$
$$\mathcal{P}_2 = \{T \subset S \mid x \notin T\}.$$

First observe that $\#\mathcal{P}_1 = \#\mathcal{P}_2$. Indeed, to every subset T of S which contains x as an element, we can associate the subset $T' = T \setminus \{x\}$. This gives a one-to-one correspondence from $\mathcal{P}_1$ to $\mathcal{P}_2$. (More later on such correspondences!) Secondly, $\mathcal{P}_2$ is precisely the power set of $S \setminus \{x\}$. Since $\#(S \setminus \{x\}) = n$, by induction $\#\mathcal{P}_2 = 2^n$. Therefore

$$\#\mathcal{P}(S) = \#\mathcal{P}_1 + \#\mathcal{P}_2 = 2\mathcal{P}_1 = 2 \cdot 2^n = 2^{n+1} = 2^{\#S}.$$

□

Proposition 8.2. *Let $f(x) = e^{x^2}$. Then for all $n \in \mathbb{Z}^+$ there exists a polynomial $P_n(x)$, of degree n, such that*

$$\frac{d^n}{dx^n} f(x) = P_n(x) e^{x^2}.$$

Proof. By induction on n.
Base case ($n = 1$):
$\frac{d}{dx} e^{x^2} = 2xe^{x^2} = P_1(x)e^{x^2}$, where $P_1(x) = 2x$, a degree one polynomial.

Inductive step: Assume that for some positive integer n there exists $P_n(x)$ of degree n such that $\frac{d^n}{dx^n} e^{x^2} = P_n(x)e^{x^2}$. So $\frac{d^{n+1}}{dx^{n+1}} e^{x^2} =$

$$\frac{d}{dx}\frac{d^n}{dx^n} e^{x^2} \stackrel{\text{IH}}{=} \frac{d}{dx} P_n(x)e^{x^2} = P_n'(x)e^{x^2} + 2xP_n(x)e^{x^2} = \left(P_n'(x) + 2xP_n(x)\right) e^{x^2}.$$

Now, since $P_n(x)$ has degree n, $P_n'(x)$ has degree $n-1$ and $2xP_n(x)$ has degree $n+1$. If f and g are two polynomials such that the degree of f is different from the degree of g, then $\deg(f+g) = \max(\deg(f), \deg(g))$. In particular, $P_{n+1}(x) := P_n'(x) + 2xP_n(x)$ has degree $n+1$, completing the proof of the induction step. □

Exercise 8: Use induction and L'Hôpital's rule to show that for all $n \in \mathbb{Z}^+$, $\lim_{x\to\infty} \frac{x^n}{e^x} = 0$.

Proposition 8.3. *For all $n \in \mathbb{N}$, $\int_0^\infty x^n e^{-x} dx = n!$.*

Proof. By induction on n.
Base case ($n = 0$): $\int_0^\infty e^{-x} = -e^{-x}|_0^\infty = -e^{-\infty} - (-e^0) = -0 - (-1) = 1 = 0!$.

Induction step: let $n \in \mathbb{N}$ and assume $\int_0^\infty x^n e^{-x} dx = n!$. Now to make progress in evaluating $\int_0^\infty x^{n+1} e^{-x} dx$, we integrate by parts, taking $u = x^n + 1,\ dv = e^{-x}dx$. Then $du = (n+1)x^n dx,\ v = e^{-x}$, and

$$\int_0^\infty x^{n+1} e^{-x} dx = (n+1)x^n e^{-x}|_0^{oo} - \int_0^\infty (-e^{-x}(n+1)x^n) dx$$

$$= (0-0) + (n+1)\int_0^\infty x^n e^{-x} dx \stackrel{\text{IH}}{=} (n+1)n! = (n+1)!.$$

Note that to evaluate the improper integral at ∞ we used $\lim_{x\to\infty} \frac{(n+1)x^n}{e^x} = 0$, as established in Exercise 8. □

9. One Theorem of Graph Theory

In this section we will give a micro-introduction to the field of mathematics known as **graph theory**, which is relatively new but is rapidly becoming one of the most active branches of contemporary mathematical research.

We will consider graphs which are **simple**, **undirected** and **finite**. Such a graph is given by a finite set V of **vertices**, together with a set $E \subseteq 2^V$ of two-element subsets $\{x, y\}$ of V. Most practicing graph theorists would roll their eyes a little at the formality of this definition. What it really means is: given any pair of distinct vertices, they are either connected by an **edge** or they are not. (Because are graphs are "simple," we do not allow an edge to connect a vertex to itself, and we do not allow more than one edge between the same pair of vertices.) Thus to give a graph on a finite set V with n elements, you range over all $\frac{n(n-1)}{2}$ pairs of elements of V, and for each one you get to decide, independently, whether you have an edge or not. When two vertices x and y are connected by an edge we call them **adjacent**, and when we are feeling lazy we will abbreviate "x is adjacent to y" by "$x \sim y$."

Let $G = (V, E)$ be a finite graph. The **degree** of a vertex v is the number of vertices which are adjacent to v; equivalently, it is the number of edges coming out of v (formally, the number of $e \in E$ for which v is a member). A vertex v is **isolated** if it has degree 0: i.e., it has no edges coming out of it at all. For instance, if V consists of a single element, we are not allowed to have any edges and that vertex must be isolated. More generally for any finite set V we can make a graph with no edges whatsoever: every vertex is isolated. (This is not a very interesting graph, but it's a legal one.) A vertex v is **pendant** if it has degree 1, i.e., there is exactly one edge coming out of V.

A **path** in a finite graph $G = (V, E)$ is a sequence of *distinct* vertices $v_0, \ldots, v_n$ such that for all $0 \leq i \leq n-1$, v_i is adjacent to v_{i+1}. We say the path is from v_0 to v_n, that v_0 is the initial vertex, and that v_n is the terminal vertex. The **length** of the path is n, i.e., the number of edges in the path. It may seem a bit silly, but it will be convenient to regard a v_0 as giving a path of length 0. A finite graph is **connected** if for any pair of vertices $v_i \neq v_j \in V$, there is a path from v_i to v_j.

A **cycle** in a finite graph $G = (V, E)$ is a sequence of vertices $v_0, \ldots, v_n$ (with $n \geq 2$) such that v_i is adjacent to v_{i+1} for all $0 \leq i \leq n-1$, such that $v_0, \ldots, v_{n-1}$ are all distinct and $v_n = v_0$. (Thus a cycle looks like a path until we traverse the very last edge and discover that we are back where we started.)

I heartily suggest that you draw some pictures of finite graphs: give yourself some points in the plane, and connect some of them with edges. The only subtlety here is that the "edges" are not actually required to be curves in the Euclidean plane, so you should not worry if the edges that you draw cross each other: you really can have as many or as few edges as you like. In particular you should draw enough pictures to convince you that the definition of connected is a good one: it corresponds perfectly to the intuitive idea that a graph should have "one piece."

Here is our final, key, definition.

A **(finite) tree** is a finite, connected graph with no cycles.

Now you should draw all possible shapes of trees with a small number of vertices. The graph with one vertex (and no edges, necessarily) is a tree, quite trivially. There is ("up to isomorphism," a concept we will not formalize here) one tree with two vertices: we connect the two vertices with an edge, getting a path of length 2. There is one tree with three vertices: a path of length 3. At four vertices, things get more interesting: in addition to the path of length 4 there is a "star" obtained by putting one vertex in the middle and drawing edges between it and each of the other three vertices. The middle vertex has degree 3, so this is different from any path. How many trees are there on five vertices?

Our goal now is to prove the following result, which occurs in the first pages of most graph theory texts.

Theorem 9.1. *Let $T = (V, E)$ be a finite tree. Then $\#V = 1 + \#E$. That is, the number of vertices is one more than the number of edges.*

Theorem 9.1 provides a nice example of a certain kind of induction proof, in which we have a problem of discrete "complexity" and we solve it by showing that we can always "reduce the complexity" at every step. We need a preliminary result.

Lemma 9.2. *Let $T = (V, E)$ be a finite tree with more than one vertex. Then T has a pendant (i.e., degree one) vertex.*

Proof. Let G be a finite graph with n vertices. Then G has a path $v_0 \sim \ldots \sim v_k$ of maximal length $k \leq n$. Why? Well, we can start with a path of length 0. Being given a path of length k, to say it is not of maximal length means that we can add one more edge at the beginning or end to get a longer path. But since the vertices of a path are required to be distinct and we have n vertices in the entire graph, clearly this process of extending the path if necessary can occur at most n times.

That was true in any finite graph. Now in our finite *tree* let $v_0 \sim v_1 \sim \ldots \sim v_k$ be our path of maximal length. First note that $k \geq 1$: indeed any edge in a graph gives a path of length 1, and if we have more than two vertices and no edges we cannot have a connected graph! So $v_0 \neq v_k$. I claim that in fact v_0 and v_k are both pendant vertices. The argument is the same, so we look at v_k. If it is not pendant, then there is another edge coming out of v_k making it adjacent to some vertex v'. If v' is different form all the v_i, $0 \leq i \leq k$ then we can take $v_{k+1} = v'$ and we have extended our maximal length path: contradiction. If on the other hand $v' = v_i$ for some $0 \leq i < k$, then $v_i \sim v_{i+1} \sim \ldots \sim v_k \sim v_i$ is a cycle in T: contradiction. □

In fact the arugment showed that any finite tree with more than one vertex has two pendant vertices, and this is the best possible result: for any $n \geq 1$ a path of length n has exactly two pendant vertices.

Proof of Theorem 9.1:
We go by induction on n, the number of vertices of our finite tree T.
Base Case ($n = 1$): If there is one vertex, there are no edges, and $1 = 1 + 0$: OK.
Induction Step: Let $n \in \mathbb{Z}^+$, assume that every finite tree with n vertices has $n - 1$ edges, and let T be a finite tree with $n + 1$ vertices. By Lemma 9.2 we have a

pendant vertex v, so that there is exactly one edge e coming out of v. Here is the crux of the entire proof: we may **prune** the tree by removing the pendant vertex v and its edge e. This leaves us with a tree T'. Certainly T' has no cycles: if we start with a graph with no cycles and remove stuff, certainly we have no cycles. Moreover a pendant vertex cannot occur in the middle of a path (i.e., as neither the initial or terminal vertex of the path), because being a middle vertex in a path means there are at least two edges coming out of it. If v_1 and v_2 are distinct vertices in T', then whatever path connected them in T cannot include the removed pendant vertex v, so it still gives a path in T'. The tree T' has $n-1$ vertices, so by induction it has $n-2$ edges. On the other hand, clearly T' has one less edge than T, so T must have $n-2+1-n-1$ edges. We're done.

10. The Principle of Strong/Complete Induction

Problem: A sequence is defined recursively by $a_1 = 1$, $a_2 = 2$ and $a_n = 3a_{n-1} - 2a_{n-2}$. Find a general formula for a_n and prove it by induction.

Proof analysis: Unless we know something better, we may as well examine the first few terms of the sequence and hope that a pattern jumps out at us. We have

$$a_3 = 3a_2 - 2a_1 = 3 \cdot 2 - 2 \cdot 1 = 4.$$
$$a_4 = 3a_3 - 2a_2 = 3 \cdot 4 - 2 \cdot 2 = 8.$$
$$a_5 = 3a_4 - 2a_3 = 3 \cdot 8 - 2 \cdot 4 = 16.$$
$$a_6 = 3a_5 - 2a_4 = 3 \cdot 16 - 2 \cdot 8 = 32.$$

The evident guess is therefore $a_n = 2^{n-1}$. Now a key point: it is not possible to prove this formula using the version of mathematical induction we currently have. Indeed, let's try: assume that $a_n = 2^{n-1}$. Then

$$a_{n+1} = 3a_n - 2a_{n-1}.$$

By the induction hypothesis we can replace a_n with 2^{n-1}, getting

$$a_{n+1} = 3 \cdot 2^{n-1} - 2a_{n-1};$$

now what?? A little bit of thought indicates that we think $a_{n-1} = 2^{n-2}$. If for some reason it were logically permissible to make that substitution, then we'd be in good shape:

$$a_{n+1} = 3 \cdot 2^{n-1} - 2 \cdot 2^{n-2} = 3 \cdot 2^{n-1} - 2^{n-1} = 2 \cdot 2^{n-1} = 2^n = 2^{(n+1)-1},$$

which is what we wanted to show. Evidently this goes a bit beyond the type of induction we have seen so far: in addition to assuming the truth of a statement $P(n)$ and using it to prove $P(n+1)$, we also want to assume the truth of $P(n-1)$.

There is a version of induction that allows this, and more:

Principle of Strong/Complete Induction:
Let $P(n)$ be a sentence with domain the positive integers. Suppose:
(i) $P(1)$ is true, and
(ii) For all $n \in \mathbb{Z}^+$, if $P(1), \ldots, P(n-1), P(n)$ are all true, then $P(n+1)$ is true.
Then $P(n)$ is true for all $n \in \mathbb{Z}^+$.

Thus, in a nutshell, strong/complete induction allows us to assume not only the

truth of our statement for a single value of n in order to prove it for the next value $n+1$, but rather allows us to assume the truth of the statement for all positive integer values less than $n+1$ in order to prove it for $n+1$.

It is easy to see that PS/CI implies the usual principle of mathematical induction. The logical form of this is simply[6]

$$(A \implies C) \implies (A \wedge B \implies C).$$

In other words, if one can deduce statement C from statement A, then one can also deduce statement C from A together with some additional hypothesis or hypotheses B. Specifically, we can take A to be $P(n)$, C to be $P(n+1)$ and B to be $P(1) \wedge P(2) \wedge \ldots \wedge P(n-1)$.[7]

Less obviously, one can use our previous PMI to prove PS/CI. To most mathematicians this is a comforting fact: one does not want to keep introducing additional "axioms" or "assumptions" in order to solve problems. Again the proof is not hard but slightly tricky. Suppose that we believe in PMI and we wish to prove PS/CI. Let $P(n)$ be any sentence with domain the positive integers and satisfying (i) and (ii) above. We wish to show that $P(n)$ is true for all positive integers n, using only ordinary induction.

The trick is to introduce a new predicate $Q(n)$, namely

$$Q(n) = P(1) \wedge P(2) \wedge \ldots \wedge P(n).$$

Notice that $Q(1) = P(1)$ and that (ii) above tells us that $Q(n) \implies P(n+1)$. But if we know $Q(n) = P(1) \wedge \ldots \wedge P(n)$ and we also know $P(n+1)$, then we know $P(1) \wedge \ldots \wedge P(n) \wedge P(n+1) = Q(n+1)$. So $Q(1)$ holds and for all n, $Q(n) \implies Q(n+1)$. So by ordinary mathematical induction, $Q(n)$ holds for all n, hence certainly $P(n)$ holds for all n.

Exercise 9: As for ordinary induction, there is a variant of strong/complete induction where instead of starting at 1 we start at any integer N_0. State this explicitly.

Here is an application which makes full use of the "strength" of PS/CI.

Proposition 10.1. *Let $n > 1$ be an integer. Then there exist prime numbers $p_1, \ldots, p_k$ (for some $k \geq 1$) such that $n = p_1 \cdots p_k$.*

Proof. We go by strong induction on n.
Base case: $n = 2$. Indeed 2 is prime, so we're good.

Induction step: Let $n > 2$ be any integer and assume that the statement is true for all integers $2 \leq k < n$. We wish to show that it is true for n.
Case 1: n is prime. As above, we're good.
Case 2: n is not prime. By definition, this means that there exist integers a, b, with $1 < a, b < n$, such that $n = ab$. But now our induction hypothesis applies to both

[6]The symbol $\wedge$ denotes logical conjunction: in other words, $A \wedge B$ means "A and B".

[7]I do admit that the underlying logical reasoning here is rather abstract and hence mildly confusing. If you want to follow along, give yourself some time and a quiet place to work it out!

a and b: we can write $a = p_1 \cdots p_k$ and $b = q_1 \cdots q_l$, where the p_i's and q_j's are all prime numbers. But then

$$n = ab = p_1 \cdots p_k q_1 \cdots q_l$$

is an expression of n as a product of prime numbers: done! □

This is a good example of the use of induction (of one kind or another) to give a very clean proof of a result whose truth was not really in doubt but for which a more straightforward proof is wordier and messier.

11. Solving Homogeneous Linear Recurrences

Recall our motivating problem for PS/CI: we were given a sequence defined by $a_1 = 1$, $a_2 = 2$, and for all $n \geq 1$, $a_n = 3a_{n-1} - 2a_{n-2}$. By trial and error we guessed that $a_n = 2^{n-1}$, and this was easily confirmed using PS/CI.

But this was very lucky (or worse: the example was constructed so as to be easy to solve). In general, it might not be so obvious what the answer is, and as above, this is induction's Kryptonite: it has no help to offer in guessing the answer.

Example: Suppose a sequence is defined by $x_0 = 2$, $x_n = 5x_{n-1} - 3$ for all $n \geq 1$.

Here the first few terms of the sequence are $x_1 = 7$, $x_2 = 32$, $x_3 = 157$, $x_4 = 782$, $x_5 = 3907$. What's the pattern? It's not so clear.

This is a case where a bit more generality makes things much clearer: it is often easier to detect a pattern involving algebraic expressions than a pattern involving integers. So suppose that we have any three real numbers a, b, c, and we define a sequence recursively by $x_0 = c$, $x_n = ax_{n-1} + b$ for all $n \geq 1$. Now let's try again:

$$x_1 = ax_0 + b = ac + b.$$
$$x_2 = ax_1 + b = a(ac + b) + b = ca^2 + ba + b.$$
$$x_3 = ax_2 + b = a(ca^2 + ba + b) + b = ca^3 + ba^2 + ba + b.$$
$$x_4 = ax_3 + b = a(ca^3 + ba^2 + ba + b) + b = ca^4 + ba^3 + ba^2 + ba + b.$$

Aha: it seems that we have for all $n \geq 1$.

$$x_n = ca^n + ba^{n-1} + \ldots + ba + b.$$

Now we have something that induction can help us with: it is true for $n = 1$. Assuming it is true for n, we calculate

$$x_{n+1} = ax_n + b \stackrel{IH}{=} a(ca^n + ba^{n-1} + \ldots + ba + b) + b) = ca^{n+1} + ba^n + \cdots + ba^2 + ba + b,$$

which is what we wanted. So the desired expression is correct for all n. Indeed, we can simplify it:

$$x_n = ca^n + b\sum_{i=1}^{n} a_i = ca^n + b\left(\frac{a^{n+1} - 1}{a - 1}\right) = \frac{(ab + ac - c)a^n - b}{a - 1}.$$

In particular the sequence x_n grows exponentially in n.

Let us try our hand on a famous two-term recurrence, the **Fibonacci numbers**:

$$F_1 = F_2 = 1, \ \forall n \geq 1, F_{n+2} = F_{n+1} + F_n.$$

Again we list some values:

$$F_3 = 2,\ F_4 = 3,\ F_5 = 5,\ F_6 = 8,\ F_7 = 13,\ F_8 = 21,\ F_9 = 34,\ F_{10} = 55,$$
$$F_{11} = 89,\ F_{12} = 144,\ F_{13} = 233,\ F_{14} = 377,\ F_{15} = 377,$$
$$F_{200} = 280571172992510140037611932413038677189525,$$
$$F_{201} = 453973694165307953197296969697410619233826.$$

This partial list suggests that F_n again grows exponentially in n. Indeed, if we compare ratios of successive values, it seems that the base of the exponential is somewhere between 1 and 2. Especially,

$$\frac{F_{201}}{F_{200}} = 1.618033988749894848204586834\ldots.$$

If you happen to be very familiar with numbers, you might just recognize this as the **golden ratio** $\varphi = \frac{1+\sqrt{5}}{2}$.

However, let's consider a more general problem and make a vaguer guess. Namely, for real numbers b, c we consider an recurrence of the form

$$x_1 = A_1, x_2 = A_2,\ \forall n \geq 1, x_{n+2} = bx_{n+1} + cx_n. \tag{5}$$

In all the cases we have looked at, the solution was, roughly, exponential. So let's **guess** an exponential solution: $x_n = Cr^n$. By plugging this in, we can get information about r:

$$Cr^{n+2} = x_{n+2} = b(Cr^{n+1}) + c(Cr^n),$$

which simplifies to

$$r^2 - br - cr = 0.$$

Evidently the solutions to this are

$$r = \frac{b \pm \sqrt{b^2 + 4c}}{2}.$$

Some cases to be concerned about are the case $c = \frac{-b^2}{4}$, in which case we have only a single root $r = \frac{b}{2}$, and the case $c < \frac{-b^2}{4}$ in which case the roots are complex numbers. But for the moment let's look at the Fibonacci case: $b = c = 1$. Then $r = \frac{1\pm\sqrt{5}}{2}$. So we recover the golden ratio $\varphi = \frac{1+\sqrt{5}}{2}$ – a good sign! – as well as

$$\frac{1-\sqrt{5}}{2} = 1 - \varphi = -.618033988749894848204586834\ldots.$$

So we have two different bases – what do we do with that? A little thought shows that if r_1^n and r_2^n are both solutions to the recurrence $x_{n+2} = bx_{n+1}cx_n$ (with any initial conditions), then so is $C_1r_1^n + C_2r_2^n$ for any constants C_1 and C_2. Therefore we propose $x_n = C_1r_1^n + C_2r_2^n$ as the **general solution** to the two-term homogeneous linear recurrence (5) and the two initial conditions $x_1 = A_1$, $x_2 = A_2$ provide just enough information to solve for C_1 and C_2.

Trying this for the Fibonacci sequence, we get

$$1 = F_1 = C_1\varphi + C_2(1-\varphi).$$
$$1 = F_2 = C_1(\varphi)^2 + C_2(1-\varphi)^2.$$

Multiplying the first equation by φ and subtracting it from the second equation will give us a linear equation to solve for C_2, and then we plug the solution into either of the equations and solve for C_1. It turns out that

$$C_1 = \frac{1}{\sqrt{5}},\ C_2 = \frac{-1}{\sqrt{5}}.$$

Interlude: This is easily said and indeed involves nothing more than high school algebra. But one cannot say that the calculation is much fun. It is always fun to find some clever way to circumvent a tedious calculation, so in that spirit I present the following alternate argument. Namely, instead of determining the constants by evaluating F_n at $n = 1$ and $n = 2$, it would be much easier algebraically to evaluate at $n = 1$ and $n = 0$, because then we have

$$F_0 = C_1\varphi^0 + C_2(1-\varphi)^0 = C_1 + C_2.$$

But for this to work we need to know F_0, which we have not defined. Can it be defined in a sensible way? Yes! Writing the basic recurrence in the form $F_{n+1} = F_n + F_{n-1}$ and solving for F_{n-1} gives:

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

This allows us to define F_n for all integers n. In particular, we have

$$F_0 = F_2 - F_1 = 1 - 1 = 0.$$

Thus we get

$$0 = C_1 + C_2,$$

whereas plugging in $n = 1$ gives

$$1 = C_1(\varphi) + C_2(1-\varphi) = C_1(\varphi) - C_1(1-\varphi) = (2\varphi - 1)C_1,$$

$$C_1 = \frac{1}{2\varphi - 1} = \frac{1}{2\left(\frac{1+\sqrt{5}}{2}\right) - 1} = \frac{1}{\sqrt{5}},\ C_2 = \frac{-1}{\sqrt{5}}.$$

Now we are ready to prove the following result.

Theorem 11.1. *(Binet's Formula) For any $n \in \mathbb{Z}$, the nth Fibonacci number is*

$$F_n = \frac{1}{\sqrt{5}}\left(\varphi^n - (1-\varphi)^n\right),$$

where $\varphi = \frac{1+\sqrt{5}}{2}$.

Proof. We go by strong/complete induction on n. The base cases are $n = 1$ and $n = 2$, but we have already checked these: we used them to determine the constants C_1 and C_2. So now assume that $n \geq 3$ and that the formula is correct for all positive integers smaller than $n + 2$. Then, using the identities

$$\varphi^2 = \varphi + 1,$$
$$(1-\varphi) = -\varphi^{-1},$$
$$1 - \varphi^{-1} = \varphi^{-2} = (-\varphi)^{-2},$$

we compute

$$F_{n+2} = F_{n+1} + F_n = \frac{1}{\sqrt{5}}(\varphi^{n+1} + \varphi^n - (1-\varphi)^{n+1} - (1-\varphi)^n))$$

$$= \frac{1}{\sqrt{5}}(\varphi^n(\varphi+1) - (1-\varphi)^n(1-\varphi+1) =$$

$$\frac{1}{\sqrt{5}}(\varphi^n(\varphi^2) - (-\varphi)^{-n}((-\varphi)^{-1} + 1)$$

$$= \frac{1}{\sqrt{5}}(\varphi^{n+2} - (-\varphi)^{-n}(-\varphi)^{-2}) = \frac{1}{\sqrt{5}}(\varphi^{n+2} - (-\varphi)^{-(n+2)}) = \frac{1}{\sqrt{5}}(\varphi^{n+2} - (1-\varphi)^{n+2}).$$

□

Exercise 10: Find all $n \in \mathbb{Z}$ such that $F_n < 0$.

It is not quite true that any solution (5) must have exponential growth. For instance, consider the recurrence

$$x_1 = 1,\ x_2 = 2,\ \forall n \geq 1, x_{n+2} = 2x_{n+1} - x_n.$$

Then

$$x_3 = 2x_2 - x_1 = 2 \cdot 2 - 1 = 3,\ x_4 = 2x_3 - x_2 = 2 \cdot 3 - 2 = 4,\ x_5 = 2 \cdot 4 - 3 = 5.$$

It certainly looks as though $x_n = n$ for all n. Indeed, assuming it to be true for all positive integers smaller than $n + 2$, we easily check

$$x_{n+2} = 2x_{n+1} - x_n = 2(n + 1) - n = 2n + 2 - n = n + 2.$$

What happened? The characteristic polynomial in this case is $r^2 - 2r + 1 = (r-1)^2$, so that it has repeated roots. One solution is $C_1 1^n = C_1$ (i.e., x_n is a constant sequence). This occurs if and only if $x_2 = x_1$, so clearly there are nonconstant solutions as well. It turns out that in general, if the characteristic polynomial is $(x - r)^2$, then the two basic solutions are $x_n = r^n$ and also $x_n = nr^n$. It is unfortunately harder to guess this in advance, but it is not hard to check that this gives a solution to a recurrence of the form $x_{n+2} = 2r_0 x_{n+1} - r_0^2 x_n$ (which is the most general recurrence whose characteristic polynomial is $(r - r_0)^2$.

These considerations will be eerily familiar to the reader who has studied homogeneous linear differential equations. For a more systematic exposition on "discrete analogues" of calculus concepts (with applications to the determination of power sums as in §3), see [DC].

12. The Well-Ordering Principle

There is yet another form of mathematical induction that can be used to give what is, arguably, an even more elegant proof of Proposition 10.1 (and of course has other uses as well). Namely:

Theorem 12.1. *(Well-Ordering Principle) Let S be any nonempty subset of the positive integers. Then S has a least element, i.e., there exists $s \in S$ such that for all $t \in S$, $s \leq t$.*

Intutitively, the statement is true by the following reasoning: first we ask the question: is $1 \in S$? If so, it is certainly the least element of S. If not, we ask: is $2 \in S$? If so, it is certainly the least element of S. And then we continue in this way: if we eventually get a "yes" answer then we have found our least element. But if for every n the answer to the question "Is n an element of S?" is negative, then S is empty!

The well-ordering principle (henceforth **WOP**) is often useful in its contrapositive form: if a subset $S \subset \mathbb{Z}^+$ does *not* have a least element, then $S = \varnothing$.

Although WOP is, if anything, even more intuitively clear than PMI and PS/CI, it is nevertheless interesting to know that it is logically equivalent to these two principles.

First, we will assume PS/CI and show that WOP follows. For this, observe that WOP holds iff $P(n)$ holds for all $n \in \mathbb{Z}^+$, where $P(n)$ is the following statement:

$P(n)$: If $S \subset \mathbb{Z}^+$ and $n \in S$, then S has a least element.

Indeed, if $P(n)$ holds for all n and $S \subset \mathbb{Z}$ is nonempty, then it contains some positive integer n, and then we can apply $P(n)$ to see that S has a least element. Now we can prove that $P(n)$ holds for all n by complete induction: first, if $1 \in S$, then indeed 1 is the least element of S, so $P(1)$ is certainly true. Now assume $P(k)$ for all $1 \leq k \leq n$, and suppose that $n+1 \in S$. If $n+1$ is the least element of S, then we're done. If it isn't, then it means that there exists $k \in S$, $1 \leq k \leq S$. Since we have assumed $P(k)$ is true, therefore there exists a least element of S.

Conversely, let us assume WOP and prove PMI. Namely, let $S \subset \mathbb{Z}$ and suppose that $1 \in S$, and that for all n, if $n \in S$ then $n+1 \in S$. We wish to show that $S = \mathbb{Z}^+$. Equivalently, putting $T = \mathbb{Z}^+ \setminus S$, we wish to show that $T = \emptyset$. If not, then by WOP T has a least element, say n. Reasoning this out gives an immediate contradiction: first, $n \notin S$. By assumption, $1 \in S$, so we must have $n > 1$, so that we can write $n = m + 1$ for some $m \in \mathbb{Z}^+$. Further, since n is the least element of T we must have $n - 1 = m \in S$, but now our inductive assumption implies that $n + 1 = n \in S$, contradiction.

So now we have shown that PMI $\iff$ PS/CI $\implies$ WOP $\implies$ PMI. Thus all three are logically equivalent.

Let us give another proof of Proposition 10.1 using WOP. We wish to show that every integer $n > 1$ can be factored into primes. Similarly, to the above, let S be the set of integers $n > 1$ which *cannot* be factored into primes. Seeking a contradiction, we assume that S is nonempty. In that case, by WOP it has a least element, say n. Now n is certainly not prime, since otherwise it can be factored into primes. So we must have $n = ab$ with $1 < a, b < n$. But now, since a and b are integers greater than 1 which are smaller than the least element of S, they must each have prime factorizations, say $a = p_1 \cdots p_k$, $b = q_1 \cdots q_l$. But then (stop me if you've heard this one before)

$$n = ab = p_1 \cdots p_k q_1 \cdots q_l$$

itself can be expressed as a product of primes, contradicting our assumption. therefore S is empty: every integer greater than 1 is a product of primes.

This kind of argument is often called proof by **minimum counterexample**.

Upon examination, the two proofs of Proposition 10.1 are very close: the difference

between a proof using strong induction and a proof using well ordering is more a matter of literary taste than mathematical technique.

13. Upward-Downward Induction

Proposition 13.1. *(Upward-Downward Induction) Let $P(x)$ be a sentence with domain the positive integers. Suppose that:*
(i) For all $n \in \mathbb{Z}^+$, $P(n+1)$ is true $\implies$ $P(n)$ is true, and
(ii) For every $n \in \mathbb{Z}^+$, there exists $N > n$ such that $P(N)$ is true.
Then $P(n)$ is true for all positie integers n.

Proof. Let S be the set of positive integers n such that $P(n)$ is false. Seeking a contradiction we suppose that S is nonempty. Then by Well-Ordering S has a least element n_0. By condition (ii) there exists $N > n_0$ such that $P(N)$ is true.

Now an inductive argument using condition (i) shows that $P(N)$ is true for all positive integers less than N. To be formal about it, for any negative integer let $P(n)$ be any true statement (e.g. $1 = 1$). Then, for $n \in \mathbb{N}$, define $Q(n) = P(N-n)$. Then $Q(0) = P(N)$ holds, and for all $n \in \mathbb{N}$, if $Q(n) = P(N-n)$ holds, then by (ii) $P(N-(n+1)) = Q(n+1)$ holds, so by induction $Q(n)$ holds for all n, which means that $P(n)$ holds for all $n < N$.

In particular $P(n_0)$ is true, contradiction. □

It is not every day that one proves a result by Upward-Downward Induction. But there are a few nice applications of it, including the following argument of Cauchy.

Theorem 13.2. *(Arithmetic-Geometric Mean Inequality) Let $n \in \mathbb{Z}^+$ and let $a_1, \ldots, a_n$ be positive real numbers. Then:*

$$(a_1 \cdots a_n)^{\frac{1}{n}} \leq \frac{a_1 + \ldots + a_n}{n}. \tag{6}$$

Equality holds in (6) iff $a_1 = \ldots = a_n$.

Proof. Step 0: We will prove the result by Upward-Downward Induction on n. For $n \in \mathbb{Z}^+$ let $P(n)$ be the statement of the theorem. Then we will show:
• $P(1)$ and $P(2)$ hold.
• For all $n \in \mathbb{Z}^+$, if $P(n)$ holds, then $P(2n)$ holds.
• For all $n > 1$, if $P(n)$ holds then $P(n-1)$ holds.
By Proposition 13.1 this suffices to prove the result.
Step 1 (Base Cases): $P(1)$ is simply the assertion that $a_1 = a_1$, which is indeed true. Now let a_1, a_2 be any two positive numbers. Then

$$\left(\frac{a_1 + a_2}{2}\right)^2 - a_1 a_2 = \frac{a_1^2 + 2a_1a_2 + a_2^2}{4} - \frac{4a_1a_2}{4} = \frac{(a_1 - a_2)^2}{4} \geq 0,$$

with equality iff $a_1 = a_2$. This proves $P(2)$.
Step 2 (Doubling Step): Suppose that for some $n \in \mathbb{Z}^+$ $P(n)$ holds, and let $a_1, \ldots, a_{2n}$ be any positive numbers. Applying $P(n)$ to the n positive numbers $a_1, \ldots, a_n$ and then to the n positive numbers $a_{n+1}, \ldots, a_{2n}$ we get

$$a_1 + \ldots + a_n \geq n\,(a_1 \cdots a_n)^{\frac{1}{n}}$$

and

$$a_{n+1} \ldots + a_{2n} \geq n\,(a_{n+1} \cdots a_{2n})^{\frac{1}{n}}.$$

Adding these inequalities together gives

$$a_1 + \ldots + a_{2n} \geq n\left((a_1 \cdots a_n)^{\frac{1}{n}} + (a_{n+1} \cdots a_{2n}\right)^{\frac{1}{n}}.$$

Now apply P(2) with $\alpha = (a_1 \cdots a_n)^{\frac{1}{n}}$ and $\beta = (a_{n+1} \cdots a_{2n})^{\frac{1}{n}}$ to get

$$n(a_1 \cdots a_n)^{\frac{1}{n}} + n(a_{n+1} \cdots a_{2n})^{\frac{1}{n}} = n(\alpha + \beta) \geq 2n(\sqrt{\alpha\beta})$$
$$= 2n\left(a_1 \cdots a_{2n}\right)^{\frac{1}{2n}},$$

so

$$\frac{a_1 + \ldots + a_{2n}}{2n} \geq (a_1 \cdots a_{2n})^{\frac{1}{2n}}.$$

Also equality holds iff $a_1 = \ldots = a_n$, $a_{n+1} = \ldots = a_{2n}$ and $\alpha = \beta$ iff $a_1 = \ldots = a_{2n}$.
Step 3 (Downward Step): Let $n > 1$ and suppose $P(n)$ holds. Let $a_1, \ldots, a_{n-1}$ be any positive numbers, and put $s = a_1 + \ldots + a_{n-1}$, $a_n = \frac{s}{n-1}$. Applying the result with $a_1, \ldots, a_n$ we get

$$a_1 + \ldots + a_n = s + \frac{s}{n-1} = \left(\frac{n}{n-1}\right)s \geq n\left(\frac{a_1 \cdots a_{n-1}s}{n-1}\right)^{\frac{1}{n}},$$

so

$$s^{\frac{n-1}{n}} \geq (n-1)^{\frac{n-1}{n}}(a_1 \cdots a_{n-1})^{\frac{1}{n}}$$

and thus

$$a_1 + \ldots + a_{n-1} = s \geq (n-1)(a_1 \cdots a_{n-1})^{\frac{1}{n-1}}.$$

We have equality iff $a_1 = \ldots = a_n$ iff $a_1 = \ldots = a_{n-1}$. □

14. The Fundamental Theorem of Arithmetic

14.1. **Euclid's Lemma and the Fundamental Theorem of Arithmetic.**

The following are the two most important theorems in beginning number theory.

Theorem 14.1. *(Euclid's Lemma) Let p be a prime number and a, b be positive Suppose that $p \mid ab$. Then $p \mid a$ or $p \mid b$.*

Theorem 14.2. *(Fundamental Theorem of Arithmetic) The factorization of any integer $n > 1$ into primes is unique, up to the order of the factors. Explicitly, suppose that*

$$n = p_1 \cdots p_k = q_1 \cdots q_l,$$

are two factorizations of n into primes, with $p_1 \leq \ldots \leq p_k$ and $q_1 \leq \ldots \leq q_l$. Then $k = l$ and $p_i = q_i$ for all $1 \leq i \leq k$.

Let us say that a prime factorization $n = p_1 \cdots p_k$ is in **standard form** if, as above, we have $p_1 \leq \ldots \leq p_k$. Every prime factorization can be put in standard form by ordering the primes from least to greatest, and dealing with standard form factorizations is a convenient bookkeeping device, since otherwise our uniqueness statement would have to include a proviso "up to the order of the factors", which makes everything slightly more complicated.

Remark: When I teach number theory I state the *existence* of prime factorizations as the first part of the Fundamental Theorem of Arithmetic and the above uniqueness statement as the second part. Since we have already proven – twice! – that every integer greater than one may be factored into a product of primes, it

doesn't seem necessary to restate it here. Anyway, the uniqueness of prime factorizations lies much deeper than the existence.

We wish to draw the reader's attention to the following important point: given Proposition 10.1 – i.e., the existence of prime factorizations, Theorems 14.1 and 14.2 are equivalent: each can be easily deduced from the other.

EL implies FTA: Assume Euclid's Lemma. As we have already seen, this implies the Generalized Euclid's Lemma (Proposition 7.2): if a prime divides any finite product of integers it must divide one of the factors. Our proof will be by minimal counterexample: suppose that there are some integers greater than one which factor into primes in more than one way, and let n be the least such integer, so

$$n = p_1 \cdots p_k = q_1 \cdots q_l, \tag{7}$$

where each of the primes is written in nonincreasing order. Evidently $p_1 \mid n = q_1 \cdots q_l$, so by the Generalized Euclid's Lemma (Proposition 7.2), we must have that $p_1 \mid q_j$ for some $1 \leq j \leq l$. But since q_j is also prime, this means that $p_1 = q_j$. Therefore we can cancel them from the expression, getting

$$\frac{n}{p_1} = p_2 \cdots p_k = q_1 \cdots q_{j-1} q_{j+1} \cdots q_l. \tag{8}$$

But now $\frac{n}{p_1}$ is strictly less than the least integer which has two different factorizations into primes, so it must have a unique factorization into primes, meaning that the primes on the left hand side of (8) are equal, in order, to the primes on the right hand side of (8). This also implies that $p_1 = q_j$ is less than or equal to all the primes appearing on the right hand side, so $j = 1$. Thus we have $k = l$, $p_1 = q_j = q_1$ and $p_i = q_i$ for $2 \leq i \leq j$. But this means that in (7) the two factorizations are the same after all! Done.

FTA implies EL: Assume that every integer greater than one factors *uniquely* into a product of primes, and let p be a prime, and let a and b be positive integers such that $p \mid ab$. If either a or b is 1, then the other is just p and the conclusion is clear, so we may assume that a and b are both greater than one and therefore have unique prime factorizations

$$a = p_1 \cdots p_r, \ b = q_1 \cdots q_s;$$

our assumption that p divides ab means $ab = kp$ for some $k \in \mathbb{Z}^+$ and thus

$$ab = p_1 \cdots p_r q_1 \cdots q_s = kp.$$

The right hand side of this equation shows that p must appear in the prime factorization of ab. Since the prime factorization is unique, we must have at least one p_i or at least one q_j equal to p. In the first case p divides a; in the second case p divides b.

The traditional route to FTA is via Euclid's Lemma, and the traditional route to Euclid's Lemma (employed, famously, by Euclid in his *Elements*) is via a series of intermediate steps including the **Euclidean algorithm** and finding the set of all integer solutions to equations of the form $ax + by = 1$. This route takes some time to develop – perhaps a week in an elementary number theory course. It is therefore

remarkable that one can bypass all these intermediate steps and give direct inductive proofs of both EL and FTA. We will give both of these in turn (which is, to be sure, twice as much work as we need to do given the just proved equivalence of EL and FTA).

14.2. **Rogers' Inductive Proof of Euclid's Lemma.**

Here is a proof of Euclid's Lemma using the Well-Ordering Principle, following K. Rogers [Ro63].

As we saw earlier in the course, one can prove Euclid's Lemma for any particular prime p by consideration of cases. In particular we have already seen that Euclid's Lemma holds for all a and b when $p = 2$, and so forth. So suppose for a contradiction that there exists at least one prime such that Euclid's Lemma does not hold for that prime, and among all such primes, by WOP we consider the least one, say p. What this means that there exist $a, b \in \mathbb{Z}^+$ such that $p \mid ab$ but $p \nmid a$ and $p \nmid b$. Again we apply WOP to choose the least positive integer a such that there exists at least one positive integer b with $p \mid ab$ and $p \nmid a$, $p \nmid b$.

Now consider the following equation:

$$ab = (a - p)b + pb,$$

which shows that $p \mid ab \iff p \mid (a - p)b$. There are three cases:

Case 1: $a - p$ is a positive integer. Then, since $0 < a - p < a$ and a was by assumption the *least* positive integer such that Euclid's Lemma fails for the prime p, we must have that $p \mid a - p$ or $p \mid b$. By assumption $p \nmid b$, so we must have $p \mid a - p$, but then $p \mid (a - p) + p = a$, contradiction!
Case 2: We have $a = p$. But then $p \mid a$, contradiction.
Case 3: We have $a < p$. On the other hand, certainly $a > 1$ – if $p \mid 1 \cdot b$, then indeed $p \mid b$! – so that a is divisible by at least one prime (a consequence of Proposition 10.1) q, and $q \mid a < p$, so $q < p$. Therefore q is a prime which is smaller than the least prime for which Euclid's Lemma fails, so Euclid's Lemma holds for q. Since $p \mid ab$, we may write $pk = ab$ for some $k \in \mathbb{Z}^+$, and now $q \mid a \implies q \mid ab = pk$, so by Euclid's Lemma for q, $q \mid p$ or $q \mid k$. The first case is impossible since p is prime and $1 < q < p$, so we must have $q \mid k$. Therefore

$$p\left(\frac{k}{q}\right) = \left(\frac{a}{q}\right) b,$$

so $p \mid \frac{a}{q} b$. But $1 < \frac{a}{q} < a$ and a is the *least* positive integer for which Euclid's Lemma fails for p and a, so it must be that $p \mid \frac{a}{q}$ (so in particular $p \mid a$) or $p \mid b$. Contradiction. Therefore Euclid's Lemma holds for all primes p.

14.3. **The Lindemann-Zermelo Inductive Proof of FTA.**

Here is a proof of FTA using the Well-Ordering Principle, following Lindemann [Li33] and Zermelo [Ze34].

We claim that the standard form factorization of a positive integer is unique. Assume not; then the set of positive integers which have at least two different standard

form factorizations is nonempty, so has a least elment, say n, where:

$$n = p_1 \cdots p_r = q_1 \cdots q_s. \tag{9}$$

Here the p_i's and q_j's are prime numbers, not necessarily distinct from each other. However,we must have $p_1 \neq q_j$ for any j. Indeed, if we had such an equality, then after relabelling the q_j's we could assume $p_1 = q_1$ and then divide through by $p_1 = q_1$ to get a smaller positive integer $\frac{n}{p_1}$. By the assumed minimality of n, the prime factorization of $\frac{n}{p_1}$ must be unique: i.e., $r - 1 = s - 1$ and $p_i = q_i$ for all $2 \leq i \leq r$. But then multiplying back by $p_1 = q_1$ we see that we didn't have two different factorizations after all. (In fact this shows that for all i, j, $p_i \neq q_j$.)

In particular $p_1 \neq q_1$. Without loss of generality, assume $p_1 < q_1$. Then, if we subtract $p_1 q_2 \cdots q_s$ from both sides of (9), we get

$$m := n - p_1 q_2 \cdots q_s = p_1(p_2 \cdots p_r - q_2 \cdots q_s) = (q_1 - p_1)(q_2 \cdots q_s). \tag{10}$$

Evidently $0 < m < n$, so by minimality of n, the prime factorization of m must be unique. However, (10) gives two different factorizations of m, and we can use these to get a contradiction. Specifically, $m = p_1(p_2 \cdots p_r - q_2 \cdots q_s)$ shows that $p_1 \mid m$. Therefore, when we factor $m = (q_1 - p_1)(q_2 \cdots q_s)$ into primes, at least one of the prime factors must be p_1. But $q_2, \ldots, q_j$ are already primes which are different from p_1, so the only way we could get a p_1 factor is if $p_1 \mid (q_1 - p_1)$. But this implies $p_1 \mid q_1$, and since q_1 is also prime this implies $p_1 = q_1$. Contradiction!

References

[Ac00] F. Acerbi, *Plato: Parmenides 149a7-c3. A Proof by Complete Induction?* Archive for History of the Exact Sciences 55 (2000), 57–76.

[DC] P.L. Clark, *Discrete calculus*. In preparation. Draft available on request.

[Li33] F.A Lindemann, *The Unique Factorization of a Positive Integer*. Quart. J. Math. 4, 319–320, 1933.

[Mu63] A.A. Mullin, *Recursive function theory (A modern look at a Euclidean idea)*. Bulletin of the American Mathematical Society 69 (1963), 737.

[Ro63] K. Rogers, *Classroom Notes: Unique Factorization*. Amer. Math. Monthly 70 (1963), no. 5, 547–548.

[Ze34] E. Zermelo, *Elementare Betrachtungen zur Theorie der Primzahlen*. Nachr. Gesellsch. Wissensch. Göttingen 1, 43–46, 1934.