4000 Day one **Induction and Well ordering**

The first topic we study is the natural numbers and the integers. Many things can be deduced about them just from one simple principle, the **"well ordering principle".** This says that in any non empty set of natural numbers, there is always a smallest number. This is intuitively obvious, since if our set of natural numbers is non empty then it contains some number, say 149. Then the smallest natural number in our set is one of the numbers 1,2,3,.....,148, 149. After examining all of these numbers to see which ones are in the set, which only takes a finite amount of time, we can determine the smallest number in our set. Obvious or not, we will assume this as a fundamnental fact about the natural numbers. It should be memorized.
By the natural numbers we mean all positive integers, 1,2,3,4,5,.....

Well ordering principle: If S is a non empty subset of the natural numbers then there is a smallest element in S, i.e. there is a number in S that is smaller than every other number in S.

IMPORTANT: The "contrapositive" statement, which is equivalent to it, hence also true, is that if a set S of positive integers has no smallest element, then S is empty.

Lets apply this to show that all positive integers ≥ 2 can be factored into primes.
Definition: If n is an integer, we say another integer a is a factor of n, or that a divides n, if there is another integer b such that $n = ab$. (I.e. the integer a divides then integer n if the answer after dividing n/a is also an integer.)

Definition: A natural number p is prime if and only if $p \geq 2$, and the only positive integers that divide p, are p and 1.

Note that 1 is not prime, because it is not ≥ 2. (For reasons we will explain later, when we prove uniqueness of prime factorization, we do not want an integer to be prime if its reciprocal is also an integer. This rules out 1 and -1, although they have no factors other than themselves and 1.)

Theorem: Every integer $n \geq 2$ can be written as a product of (one or more) prime integers.
proof: The well ordering principle gives us a way to prove a set of positive integers is empty, so we try to prove the set of positive integers for which the theorem is false is empty.

I.e. let S be the set of all those integers ≥ 2 which cannot be written as a product of prime integers. Since a prime number is already written as a product of one prime number, the theorem is true for all prime numbers, so there are no prime numbers in the set S. By the well ordering principle, either S is empty or S has a smallest number in it. Suppose S is not empty and let n be the smallest number in S. Then $n \geq 2$, but n is not prime, so n can be expressed as a product $n = ab$, where a,b, are positive integers and neither one equals 1. Thus they are both larger than 1, and both are smaller than n. Hence neither a nor b belongs to the set S, i.e. both a and b CAN be written as products of prime numbers. If $a = p_1 \ldots\ldots p_r$ and $b = q_1 \ldots\ldots q_s$, where p_i and q_j are all prime, then multiplying them together we get
$n = ab = p_1 \ldots\ldots p_r\, q_1 \ldots\ldots q_s$, which expresses n as a product of prime numbers. But this contradcits the assumption that n was in S. So in fact S cannot have a smallest element, so then S cannot have any elements. I.e. there are no integers ≥ 2 that cannot be written as a product of primes. **QED.**

Note this is a very theoretical argument. I.e. it does not tell you how to actually factor any integer into primes. For example, here is a nice 28 digit integer: 1111222223333444455556666789, can you factor it into primes? Fortunately I have copy of the program mathematica, (Maple will do this too), and in three seconds it gives:
{{1289,1},{68636479,1},{12560097320271619,1}},
which means that the number 1111222223333444455556666789 = (1289)(68636479)(12560097320271619), and that each of those three numbers is prime. (The 1's in the parentheses above mean each factor occurs once in the factorization.)
Now that might take quite a while to do by hand, and in fact if I put in a slightly longer number, say 40 digits long, then my Mathematica won't do it either, in any reasonable amount of time. This fact, that actual numbers are quite hard to factor explicitly, is the basis for security codes used in credit cards and espionage communications by the government.

Remark:
We also used in the proof above, the fact that if $n = ab$, and all the integers a,b,n are ≥ 2, then n is larger than both a and b. It gets tedious sometimes to prove everything, but we could prove this too as follows. We assume the product of two positive numbers is positive as a basic property. If $a > 1$, and $b > 0$ then $(a-1) > 0$ so the product $b(a-1) > 0$. I.e. $ba - b > 0$ so $ba > b$. Thus if both $a,b > 1$, then $ab > a$ and $ab > b$, which is what we claimed.

4000/6000 Day 2
There are infinitely many primes (due to Euclid); binomial thm.

Review: We have learned the fundamental "well ordering" principle, that a non empty set of natural numbers must have a smallest element. Equivalently, a set of natural numbers with no smallest element has no elements at all. As a consequence, if a statement about natural numbers cannot have a smallest number for which it is false, then it cannot be false for any natural numbers. We used this to prove every natural number ≥ 2 is either prime, or can be expressed as a product of primes.

So the existence of a factorization into primes is relatively easy for positive integers. The factorization is also unique, i.e. the prime factors are uniquely determined by the integer n, but this is more work to prove, and involves the concept of gcd. We will take this up soon.

Recall, assuming we know which integers are positive, that we can "order" the integers as follows:
Definition of Ordering
Given integers a,b, we say $a > b$ if and only if $a-b > 0$, i.e. $a > b$ means $a-b$ is positive.

Here is a basic property we will need.
Positivity property:
If a,b are positive integers, then a+b and ab are also positive integers.

As we saw, it follows then that
If a, b, c are integers and $a > b$ and $c > 0$, then $ac > bc$.
It is also true that $a+c > b+c$, whether c is positive or not.
exercise: derive these facts from the positivity property.

Now we give an application of the existence of prime factorization, to the proof, given two thousand years ago by Euclid, that there must be an infinite number of prime numbers.

Recall the definition of "divides": We say an integer a divides an integer n if there is another integer b such that $n = ab$.

The following simple principle is very useful.
Three term principle: If $N = a + b$, all integers, and if an integer d divides two of the integers a, b, N, then d also divides the third.
proof: If d divides N and a, say $N = dr$ and $a = ds$, then we have $dr = N = a + b = ds + b$, thus $dr = ds + b$, so by subtracting ds from both sides, b

= dr - ds = d(r-s), and thus d divides b. The other cases are similar.

Theorem: There are an infinite number of prime numbers.
proof:
The proof is by contradiction, i.e. we shall show the assumption there are only a finite number of prime numbers leads to a contradiction.
Assume there are only a finite number of primes, say $p_1,....,p_n$. Then consider the integer $N = p_1.....p_n + 1$, obtained by multiplying together all the given primes, and then adding 1.

Claim: None of the prime numbers $p_1,.....,p_n$ divides N.
proof of claim: If p_i divides N and p_i also divides $p_1.....p_n$, then by the three term principle, p_i also divides 1, which is false since $p_i \geq 2$.

Now no p_i divides N, but $N \geq 2$, so by the result of yesterday, N is a product of primes. Thus there are primes that divide N, but none of the primes p_i do so. Hence there must exist primes other than the p_i, a contradiction to the assumption that the only primes in existence are the $p_1,....,p_n$. Thus there must be an infinite number of primes.
QED.

Remark: This is proposition 20 of Book IX of Euclid's Elements. It can be found on the web (with the rest of the Elements) at

http://aleph0.clarku.edu/~djoyce/java/elements/bookIX/propIX20.html

Binomial theorem
Next we discuss a topic that is useful in algebra, the binomial formula for expanding powers of two terms. We want a formula for the product $(a+b)^n$, where n is a positive integer. We begin, as is always advisable, by looking at some examples. Recall the "distributive law" for multiplication, that a(b+c) = ab + ac.

$(a+b)^1 = a^1 + b^1$. no comment.

$(a+b)^2 = (a+b)(a+b) = a(a+b) + b(a+b)$ (why?)

$= aa + ab + ba + bb = aa + (ab + ba) + bb$

$= a^2 + 2\,ab + b^2$. So this is the binomial theorem for n = 2.

What we want to do is determine why there is a 2 in front of the term ab,

and only a 1 in front of the other two terms. Apparently it is because there are two ways to get ab, as ab and as ba, whereas there is only one way to get a^2, namely as aa, and only one way to get b^2 as bb.

So what do we expect when we expand $(a+b)^3$?

Well let's see:

$(a+b)^3 = (a+b)(a+b)^2 = a(a+b)^2 + b\,(a+b)^2 =$

$a(aa + ab + ba + bb) = b(aa + ab + ba + bb) =$

$aaa + aab + aba + abb + baa + bab + bba + bbb =$

$aaa + (aab + aba + baa) + (abb + bab + bba) + bbb$

$a^3 + 3a^2b + 3ab^2 + b^3$. This is the binomial theorem for n = 3.

Here the first 3 is because there are three ways to get a^2b, namely as aab, aba, and baa. Namely there are three positions, and only one of them should be a "b", so the number of terms corresponds to the number of ways to choose which position to put the "b". I.e. either the b is first, second, or third.

Now what about $(a+b)^4$? There should be terms for a^4, a^3b, ab^3, and b^4.

How many ways are there to get a^4? Only one, and also one for b^4.

So we have terms a^4 and b^4.

Now what about a^3b? There are four places here in the product and only one should be a "b", so how many ways can this happen? Four ways it seems, since there are four ways to choose where to put the b.

I.e. aaab (last, i.e. fourth), aaba (next to last, i.e. third), abaa (second) , or baaa (first). So we should have

$a^4 + 4a^3b + ??a^2b^2 + 4ab^3 + b^4$.

Now what is the ?? coefficient of a^2b^2?

We have four places to fill and two of them should be "a's". So how many ways are there to choose where the two a's go?

I.e. aabb, abab, abba, baab, baba, bbaa. This seems to be all possibilities.

So we get $a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$.

Now for $(a+b)^5$ we should get $a^5 + 5a^4b + ?a^3b^2 + ?a^2b^3 + 5ab^4 + b^5$.

And the coefficient of a^3b^2 should be the number of ways to choose two places from 5 possible choices (where to put the b's). And the coefficient of a^2b^3 should be the number of ways to choose 3 places from among 5 choices (where to put the b's). But it is equivalent to know where to put the two a's, so these coefficients are the same.

Definition: Now define the symbol $\binom{n}{k}$ to be the number of ways to choose k objects from among n possible choices, where $n \geq 1$ and $0 \leq k \leq n$, and k, n are both integers. This symbol is called the binomial coefficient "n choose k", and we have seen it is the coefficient of both a^kb^{n-k} and also $a^{n-k}b^k$ in the expansion of $(a+b)^n$. Notice there is only one way to choose zero objects, and only one way to choose n objects, from among n things, so $\binom{n}{0} = 1 = \binom{n}{n}$. In general $\binom{n}{k} = \binom{n}{n-k}$.

Then we have the formula.

Binomial theorem 1:

$$(a+b)^n = a^n + na^{n-1}b + \ldots\binom{n}{k}a^kb^{n-k} + \ldots\ldots + nab^{n-1} + b^n$$

$$= \sum_{k=0}^{n}\binom{n}{k}\, a^kb^{n-k}.$$

Important: This is only half the binomial theorem. To make it useful, we need an explicit formula for the symbol $\binom{n}{k}$.

Definition: If n is any positive integer, then $n! = n(n-1)(n-2)(\ldots\ldots)(2)(1)$,

read "n factorial" is the product of n with all smaller positive integers. If n = 0, then 0! = 1.

Then the other half of the binomial theorem is this.

Binomial theorem 2: For any integer $n \geq 1$, and any integer k with $0 \leq k \leq n$, we have $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

We have two concerns, 1) how do you come up with such a formula, and 2) how do you prove it? The good part about induction is it is useful for proving things even if you have no idea how they were thought of or why they are true. It just works like magic. So we try to prove the second part of the theorem using induction on the number n. First we introduce another version of induction, not because our old one wouldn't work here, but because we need to review the other method anyway.

Principle of induction: Suppose we have a statement we can make about natural numbers. Suppose the statement is true for the number 1, and suppose that whenever it is true for any one natural number, it is also true for the next. Then the statement is true for all natural numbers.
Proof: This follows, by contradiction, from the well ordering principle as follows. If there were any natural numbers for which the statement is false, then there is a smallest one, say n. But the statement is true for 1, so $n \geq 2$. Since n is the smallest number for which the statement is false, it must be true for n-1. But then by the assumption above, the statement is also true for the next number, namely n, a contradiction. Since there cannot be a smallest number for which the statement is false, there cannot be any numbers for which the statement is false. **QED.**

I.e. to use this, check the statement holds for n=1, and then assume n is any integer ≥ 1, and the statement holds <u>for this one integer</u>. then try to show it also holds for the next integer n+1.

Abstract Principle of induction:
If S is a subset of natural numbers, and S contains 1, and if whenever S contains any integer k, then S also contains the next integer k+1, then S contains all natural numbers.

Now we use this to prove the formula $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. First check it for n =

1. If n = 1, then the only possible values of k are 0 and 1, so we must prove that $\binom{1}{0} = \frac{1!}{0!(1-0)!}$, and $\binom{1}{1} = \frac{1!}{1!(1-1)!}$. We leave it as an exercise in reviewing the meanings of these symbols, that these numbers are all 1.

Now assume that n ≥ 1 and that we have proved the formula $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ for all values of k such that 0 ≤ k ≤ n. Then consider the case n+1, and 0 ≤ k ≤ n+1.

We want to prove $\binom{n+1}{k} = \frac{(n+1)!}{k!(n+1-k)!}$.

Somehow we need to reduce this to a case we already know, for the number n and k ≤ n. Now the case k = n+1 is easy since we know $\binom{n+1}{n+1} =$ 1. (Why??) Also the case k = 0 is easy since $\binom{n+1}{0} = 1$.

So assume 1 ≤ k ≤ n.

Now recall that $\binom{n+1}{k}$ <u>means</u> the number of ways to choose k things from among n+1 choices, and $\binom{n}{k}$ means the number of ways to choose k things from only n choices. How can we relate these two? One way is to recall that n+1 is ≥ 2, and to designate one special object in our set to ignore, leaving only n other things to choose from. Then we can choose k objects from among these. That gives us $\binom{n}{k}$ choices. But these choices are the ones that do not include our one special object, so now we must count the choices that do include that object. How many are there of those? Well we choose that one special object first, and then we have to make the other choices from among the remaining objects, so there are k-1 more objects to choose from among n objects. So this gives $\binom{n}{k-1}$ more choices. Thus the total number of ways to choose k things from among n+1 things, is the sum of the number of ways which do <u>not</u> include the special object, that's $\binom{n}{k}$ ways, plus the number of ways that <u>do</u> include

the special object, and that is $\binom{n}{k-1}$ more ways. This proves the useful

Lemma: For all positive integers n, and all integers k with $1 \le k \le n$, we have $\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$.

Now using this lemma we can finish the argument. I.e. by induction we may assume the formulas we want have been proved for n and all inetegers k with $1 \le k \le n$, so we have $0 \le k-1 \le n$, so we know both $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, and also $\binom{n}{k-1} = \frac{n!}{(k-1)!(n-(k-1))!} = \frac{n!}{(k-1)!(n-k+1)!}$.

Now all we have to do is add these formulas and see if we get the desired one. That is the beauty of induction, once we come up with the formula, even if we do not understand why it is true, we can check it anyway!

So here we go, purely by calculating we get $\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$

$= \frac{n!}{k!(n-k)!} + \frac{n!}{(k-1)!(n-k+1)!}$, now we are adding fractions so we must find common denominators, so we must multiply the left side bottom by (n-k+1) and the right side bottom by k, getting $= \frac{n!(n-k+1)}{k!(n-k+1)!} + \frac{n!k}{(k)!(n-k+1)!}$

now we can add the tops to get $\frac{n!(n-k+1)+n!k}{k!(n-k+1)!} = \frac{n!(n-k+1+k)}{k!(n-k+1)!}$

$= \frac{n!(n+1)}{k!(n-k+1)!} = \frac{(n+1)!}{k!(n-k+1)!} = \frac{(n+1)!}{k!(n+1-k)!}$. This what we wanted. **QED.**

Problem: Where does this formula come from? I.e. how do we <u>understand</u> why it is true, and how would one discover it?

4000. Day 3, Toward uniqueness of prime factorization

We return to the question of factoring natural numbers into primes. We have proved it is possible for all $n \geq 2$, but we left open the question of uniqueness of the prime factors that are obtained. I.e. for all we know, an integer might have several different prime factorizations. The precise statement is this.

Uniqueness of prime factorization

If $n \geq 2$ is an integer, and $n = p_1....p_r = q_1.....q_s$ are two factorizations of n into prime factors, i.e. if all the p's and all the q's are primes, then the prime factors $p_1,....,p_r$ are the same as the prime factors $q_1,...,q_s$, except possibly for the order in which they occur. Moreover each prime factor occurs among the p's the same number of times as among the q's.
I.e. if $n = p_1....p_r = q_1.....q_s$ are two factorizations of n into prime factors, then r = s, and the factors q_j can be renumbered so that for every subscript i, we have $p_i = q_i$.

proof: (This proof also appears in the Euclid.) Suppose we try to prove this by induction say on the number s of factors q_j. Let's begin with s = 1, so that we have $n = p_1....p_r = q_1$. But since q_1 is prime, it cannot be written as a product of two or more factors all ≥ 2. Since all $p_i \geq 2$, there is only one of them, so r = s = 1, and we have $n = p_1 = q_1$ and we are done.

We need an elementary fact you no doubt already believe.

The cancellation property.

If a,b,c are positive integers, and ab = ac, then b = c.

proof: If not then say b > c. Then (b-c) > 0, so by the positivity property, also a(b-c) > 0, so then ab-ac > 0, so ab > ac, a contradiction, since we assumed ab = ac. **QED.**

Remark: The cancellation property also holds for any integers a,b,c such that $a \neq 0$.

Now assume that we have proved the uniqueness theorem for a given value of $s = k \geq 1$, and we try to prove it for s = k+1. So assume $n = p_1....p_r = q_1.....q_{k+1}$. If we could only cancel one common prime factor from both sides we would be done by induction. I.e. if we could prove that p_r equals one of the $q's$, then we could renumber the q's until the one equal to p_r is q_{k+1}, i.e. then we could assume that $p_r = q_{k+1}$.

Then we would have $p_1....p_{r-1}p_r = q_1.....q_kp_r$, and after cancelling p_r on both sides, by the cancellation property, we would have $p_1....p_{r-1} =$

$q_1 \ldots q_k$.

Now with only k factors on the right side, and r-1 on the left, we could use the induction hypothesis to conclude that k = r-1, so we would get $p_1 \ldots p_k = q_1 \ldots q_k$ and also that the primes $q_1, \ldots, q_k$ can be renumbered until $q_1 = p_1, \ldots, q_k = p_k$. Then since also $q_{k+1} = p_r = p_{k+1}$ we would be done.

This is what we need for the inductive step: We need to show that whenever $p_1 \ldots p_r = q_1 \ldots q_{k+1}$, then p_r equals one of the primes q_j on the right side.

The following important result due to Euclid would do the trick.
Divisibility property of primes.

Proposition: If p is a prime number, and p divides a product $\prod_{i=1}^{n} a_i = a_1 \ldots a_n$ of integers, then p divides at least one of the factors a_i.

I.e. if we assume this proposition, and let $n = p_1 \ldots p_r = q_1 \ldots q_s$, be two prime factorizations of n, then since p_r divides n, then p_r divides the product $q_1 \ldots q_s$, so then by the divisibility property, we could conclude that p_r must divide one of the factors q_j. But since q_j is prime, the only way p_r could divide it is if $q_j = p_r$. Then we could cancel p_r from both sides and finish with the inductive proof above.

Thus to finish off the proof of the uniqueness of prime factorization, we need to prove the previous divisibility lemma for primes.

We claim it suffices to prove the divisibility property for two factors.
Divisibility Lemma: If p is a prime number, and p divides the product ab of any two positive integers, then p divides at least one of the factors a or b, (possibly both).

Corollary: If p is a prime number, and p divides a product $\prod_{i=1}^{n} a_i = a_1 \ldots a_n$ of integers, then p divides at least one of the factors a_i.
Proof: Assuming the lemma, the corollary is proved by induction on the number of factors. I.e. if there are two factors in the product $\prod_{i=1}^{n} a_i$ then the corollary follows immediately from the lemma. Now assume the corollary proved for n-1 factors and assume p divides $a_1 \ldots a_n$. Then write

this product as a product of two factors, i.e. then p divides the product $a_1....a_n = a_1(a_2....a_n)$, where one factor is the product $(a_2....a_n)$. It follows from the lemma that either p divides a_1, and we are finished, or else p divides $(a_2....a_n)$. Since $(a_2....a_n)$ is a product of n-1 factors, then by the inductive hypothesis, p divides one of the a_i with $2 \leq i \leq n$, so we are done again. **QED.**

Remarks: In future we will assume the reader can do the inductive argument to extend such results from the case of two factors to the case of any finite number of factors, so it is wise to practice by learning this one. But the hard part is to prove it for two factors. I.e. all our previous proofs were pretty much automatic just using induction. This one is not so easy. We need to use something clever to help out with the inductive step. Fortunately the following clever argument was thought of a couple thousand years ago by the Greeks, and recorded in Euclid's book.

Greatest common divisors.
To prove the divisibility lemma, which is not at all trivial, we need the concept of the greatest common divisor of two integers.

Definition: Given two positive integers a,b, their greatest common factor, or gcd, is the largest integer that divides both of them.

Since each factor of a is less than or equal to a, there are only finitely many factors to consider, as possible common factors. Moreover 1 is always a common factor. So there are some common factors but only a finite number, so looking at all common factors we eventually can find the largest one, so the gcd does always exist and is unique. (In fact any two integers, not necessarily positive, have a gcd, as long as they are not both zero. What integers divide into zero?)

Now we need an interesting fact about gcd's which to me is not at all obvious. Recall the old problem about measuring water in buckets of various sizes. I.e. suppose you have two buckets that hold 3 quarts and 8 quarts respectively, and you want to measure exactly one quart of liquid. How would you do it? Well you could fill up the 3 qt bucket three times, and then pour off into the 8 qt bucket, filling the 8 qt bucket and leaving exactly 1 qt in the 3 qt bucket. This says that $3x3 - 8 = 1$.

Now if you had a 2 quart bucket and an 8 qt bucket, how would you do it? It is not hard to believe after a while, that no matter what you do, you will

always have an even number of qts left in each bucket, so you can never get 1 qt. But you can get 2 qts, of course.

What about a 14 qt bucket and a 20 qt bucket? What is the smallest amount you can measure? Well, you could pour 14 qts from the 14 qt bucket into the 20 qt bucket, refill the14 qt bucket and pour 6 qts into the 20 qt bucket, filling it and leaving 8 qts in the 14 qt bucket. Then empty the 20 qt one and pour in the 8 qts from the 14 qt bucket. Then refill the 14 qt bucket and pour off 12 more qts into the 20 qt bucket, filling it and leaving 2 qts in the 14 qt bucket. Since both buckets hold an even number of qts, this is the smallest number you can get.

Following through the procedure we used, gives the formula

$14 - \{20 - [2(14) - 20]\} = 2$, or $14 - 20 + 2(14) - 20 = 2$, i.e. $3(14) - 2(20) = 2$.

What if you had a 12 qt bucket and a 21 qt bucket? What do you think is the smallest number you can get? I claim it is 3 qts. Why? Do you have a guess as to the general pattern? Can you use the equation $2(12) - 21 = 3$ to help you? (Fill the 12 twice and pour off as much as possible into the 21 both times, leaving 3 in the 12 the second time.)

OK the thing to notice is that the smallest quantity you can get from the two buckets is their greatest common divisor! Then the second thing to notice is that this means that whenever d is the gcd of a and b, then we should be able to solve the equation $d = ax + by$, for some integers x,y, which are not necessarily positive.

We are ready to prove this, and this even gives us a clue how to proceed, using the well ordering principle.

Greatest common divisor Proposition.

1. If d is the gcd of a,b, then there exist integers n, m, not necessarily positive, such that $d = an+bm$. In fact d is the smallest positive number of form an+bm, when n, m range over all possible integer values.
2. Every common factor of a and b divides gcd(a,b). I.e. if $d = \gcd(a,b)$, and if some integer e divides both a and b, then e divides d.

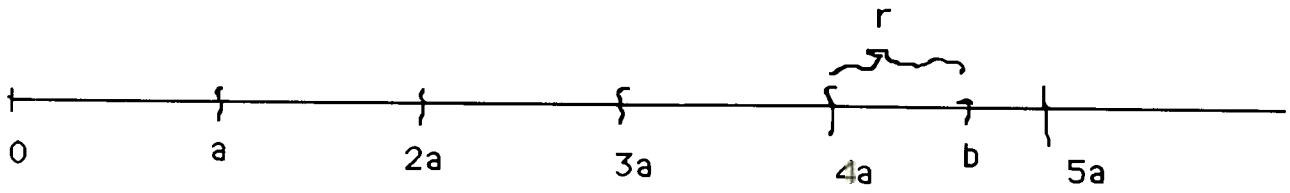
Note: Property 2. says that the gcd of a and b is not only larger than all other common factors of a and b, but it is a <u>multiple</u> of all other common factors. Notice also that property 2. is an immediate corollary of property 1. using the three term principle. I.e. if $d = an+bm$, and if e divides both a and b, then e divides the two terms on the right side of the equation $d = an+bm$, so e also divides the third term, i.e. e divides d.

Math 4000/6000. Day 4, GCD's and their properties.
Recall the main result.
Greatest common divisor Proposition.
1. If d is the gcd of a,b, then there exist integers n, m, not necessarily positive, such that d = an+bm. In fact d is the smallest positive number of form an+bm, when n, m range over all possible integer values.
2. Every common factor of a and b divides gcd(a,b). I.e. if d = gcd(a,b), and if some integer e divides both a and b, then e divides d.

So the main thing is to prove part 1., that gcd(a,b) can be obtained as the smallest number of form an+bm, with n,m integers. To prove this we need to understand division of integers, and the concept of a remainder.

Recall that division is often called "repeated subtraction". I.e. to divide a smaller integer a into a larger one b, we subtract the smaller integer a from the larger one b, as many times as possible, and when the result is too small to subtract again, i.e. when the result r of the subtraction, is less than a, we call that result r = the "remainder". Then we have written b as a certain number of copies of a, plus a remainder r. So we get b = ak+r, where $0 \leq r < a$. We allow r = 0, because this is the important special case where a divides b "evenly".

From looking on the Euclid web site mentioned earlier, one learns that the Greeks used the word "measures" instead of the word "divides". They were apparently thinking in terms of using a as a unit, and using it to measure b. Thus if a divides b evenly, then b could be exactly measured in terms of the unit a.
This leads to a geometric picture of the process of division, or repeated subtraction, as follows.



This picture illustrates dividing "b", or measuring b, using a length "a", as measuring unit. We lay off copies of a repeatedly, and stop just before we would go past b, and in this example we get the equation b = 4a + r. We have not shown the length "1" in our picture, but assuming a and b are integers means they are both multiples of "1", i.e. there is a length "1"

that can be used to measure both a and b, and it follows from the three term principle, that the remainder r in that case is also an integer, i.e. r can be measured by the same length 1. That is, if b = 4a+r, then any unit which measures a and b, i.e. which divides both a and b, also divides, i.e. measures, r as well.

Thus the Greeks were interested in knowing which lengths could measure other lengths. If we start with only a single length, say a, then the only lengths which can be measured with it are multiples na of a. But what if we start with two lengths say a and b, and we ask which lengths we can measure using both of them? First of all assume that these are both integer lengths. i.e. that there is some unit length called u or 1, such that both a and b can be measured by multiples of u. Then we can ask two types of questions.

1) Which lengths can be measured using both a and b, (but not using u). E.g. what is the <u>smallest</u> length that can be measured using both a and b?

This is the kind of problem we looked at before, of measuring water amounts with two buckets.

2) Which lengths can be used to measure both a and b? In particular what is the <u>largest</u> length which can be used to measure both a and b?

This is equivalent to asking for the greatest common divisor, or greatest common measuring unit, valid for both a and b.

The discovery of the Greeks was that these two questions have the same answer! I.e. It turns out then that the smallest length d, which can be measured using both a and b, can also be used to measure both of them! Moreover d is the largest length that will measure both a and b. So there is a kind of reciprocity here. I.e. the smallest length d that can be measured using both a and b, is itself the largest length that will measure both a and b. Note at least that if d is the gcd of a and b, i.e. if both a and b can be measured using d, then anything that can be measured using a and b can also be measured using just d. Thus every length of form ax+by, is some multiple of d. So question 1 becomes "what multiples of d occur in the form ax+by"? I.e. what is the smallest multiple of d that we can express as ax+by for some integers x and y? It is plausible to guess that all multiples of d occur, i.e. that d itself occurs as such a multiple, and this is in fact the case. (What other multiple would be more likely?

6d? ad? bd? there really is no more natural guess, and in mathematics the most "natural" guess is often correct, at least if you have a well developed sense of what is natural.)

It follows then that if d is the smallest length that can be measured using both a and b, then the other lengths that can be measured using both a and b, are just the various multiples of d. I.e. since d can be used to measure a and b, it can also be used to measure anything they will measure. So anything you can measure using a and b could be measured using d, i.e. it is a multiple of d. But d itself can be measured using a and b, so things that can be measured using both a and b, are the same as those things that can be measured using just d. Thus using two measuring sticks is no different from using one, if you pick the right one.

All this assumes however that our two measuring sticks were known to have integral length, i.e. they could both be measured using some common unit length. What if we take any two lengths, is it true that there is a common measuring unit? I.e. do any two lengths have a greatest common measuring unit? Given any two lengths a and b, is there always a unit length u, such that a is an integer number of copies of u, and also b is some integer number of copies of u?

The Pythagoreans initially assumed this was true, and were shocked, even resorting to murder it is said, to keep it quiet, when they learned otherwise. I.e. it was learned that the side of a square and the diagonal of that same square cannot be measured by any common unit length, i.e. these two lengths are "incommensurable". This led to the discovery of irrational numbers in algebra.

Lets get back to our current train of thought, proving the two questions above have the same answer. First we need to discuss what happens when trying to measure one length using another (commensurable) length, i.e. what happens when we try to divide one integer by another?.

Here then is a statement of the fundamental property of division.
Division theorem for integers. If a, b are any two positive integers, then there are non negative integers k, r such that $0 \leq r < a$, and $b = ak+r$.
Proof:
Just let r be the smallest non negative integer which can be written as b-ak, where k is a non negative integer. Certainly the set of such integers of form b-ak contains some non negative integers since we can take $k = 0$ and then $b-ak = b > 0$. If 0 can be written as b -ka, then take $r = 0$. if not then the set of numbers that can be so written is a non empty set of

natural numbers, so a smallest one exists by the W.O. principle.

(In Euclid's original proof he thought this part was as obvious as most of us think it is, and so he apparently left out this W.O. step in the argument. But today we are extremely thorough, because we have more experience with strange number systems where some of these properties fail, so we have learned to always cultivate a habit of being very clear about what properties follow from which other properties, in order to better know what is true in a new situation.)

Now we claim that if r is the smallest non negative integer of form r = b-ak, then r < a.

If not, then r ≥ a >0, so r-a ≥ 0, so we may write b-ak = r = (r-a) + a.

Then (r-a) = b - ak -a = b - (ak+a) = b - a(k+1). Then r-a can also be written in the form b - ax, and r-a is also non negative.

Moreover, since a > 0, by adding r to both sides, we get r+a > r,]
hence r > r-a ≥ 0.
This contradicts the fact that r was the smallest non negative integer of form b -ax, where x is an integer, since r-a can also be written this way.
QED.

Remark: To picture the last part of the argument, notice that assuming r ≥ a >0, just means that after laying off k copies of a, we still have a remainder bigger than a. That means we have room to lay off another copy of a and still have a non negative remainder.

Exercise: The integers k,r in the division theorem are unique. I.e. if b = ka + r = qa+s, where a,b, k,q are integers, and a > r,s ≥ 0, then r = s, and k = q.

Remark: The division theorem is also true for not necessarily positive integers a,b, as long as a ≠ 0. I.e. given any two integers a,b with a ≠ 0, there exist unique integers k,r, with 0 ≤ r < |a|, such that b = ak + r. (Here k may be negative, but r has to be non negative and less than the absolute value of a, to get the uniqueness statement.)

Now we are ready to prove the GCD proposition. Recall the statement.
Greatest common divisor Proposition, part 1.
1. If d is the gcd of a,b, then there exist integers n, m, not necessarily positive, such that d = an+bm. In fact d is the smallest positive number of

form an+bm, when n, m range over all possible integer values.
Proof. Given a,b, consider the smallest positive integer d of form d = an+bm, where n,m are integers. (Why does such a d exist?)

Then either d divides both a and b or not. If d does not divide say a, then by the division theorem, at least a = dk+r where $0 < r < d$, and k,r are integers.

Then since d = an+bm, by substituting for d, we have a = (an+bm)k+r = ank+bmk+r, so r = a - (ank+bmk) = a(1-nk)+b(-mk).

This shows that r can also be written in the form ax+by where x,y, are integers. But this is a contradiction since $0 < r < d$, and no non negative number smaller than d could be so written.
We have proved that the smallest non negative integer of form d = an+bm, where n,m are integers, does indeed divide a, and the same type of proof shows it also divides b. **QED.**

We know every integer ≥ 2 has at least one prime factor, but two different integers may have no <u>common</u> prime factors. If they do not we call them "relatively prime". This a crucial notion.

Definition: Given two integers a,b, we say they are "relatively prime" if and only if the only common factors they have are 1 and -1.

Remark: Two integers are relatively prime, if and only if they have no common prime factors, if and only if their gcd = 1.
proof of remark: If they have no common factors except 1 and -1, then none of their common factors is prime. Conversely if they do have some common factor other than 1 or -1, then they have a positive common factor, and it in turn has a prime factor, which is then a common prime factor. If their gcd is 1 then they have no common factors larger than 1, hence the only common factors are 1, -1. If the only common factors are 1,-1, then the largest one is 1. **QED.**

Now we get the tool we need.

Corollary: Two integers a,b, are relatively prime, if and only if the equation ax+by = 1, has some integer solutions for x,y.
proof: If one can solve it, then 1 is the smallest positive number of form ax+by, hence 1 = gcd(a,b). Conversely, if 1 = gcd(a,b), then one can solve the equation ax+by = 1, by the GCD proposition. **QED.**

At last this tool enables us to prove the basic divisibility result.
lemma: If a,b are relatively prime, and if n is any integer such that a divides bn, then in fact a divides n.
proof: I always thought this proof was really slick, as a student. All I can say to help remember it is this. Whenever you have two relatively prime numbers, a,b, it helps to write down the equation ax+by = 1, first off.

Now after that, the trick is to introduce the number bn in the theorem, by multiplying this equation by n. I.e. then we get axn + byn = n. Now we are done by the three term principle. I.e. a divides axn, and by hypothesis, a divides bn, hence also byn, so a also divides the third term, namely n. **QED.**

Next we get what we needed for the inductive proof of uniqueness of prime factorization.
Divisibility property of primes: If p is prime and p divides ab, then either p divides a or p divides b.
proof: If p divides a we are done, if not then p and a have no common prime factors, so they are relatively prime. Since p divides ab, by hypothesis, then by the previous lemma, p must divide b. **QED.**

This finishes Euclid's proof of the uniqueness of prime factorization.

4000/6000 Euclidean algorithm

It can be useful not just to know that two numbers have a gcd, but to know exactly what it is, and to have an easy way to compute it. There is an easy way, due to Euclid, based on the idea of measuring lengths using two other lengths. This involves a refinement of the idea of the three term principle.

Suppose we start with two commensurable lengths a and b, i.e. both a and b are integer multiples of some given unit length. We are interested in finding the smallest length that can be measured using both of them. Now a length can be measured using both a and b if that length can be written as a multiple of a, possibly negative, plus or minus a multiple of b, i.e. as an integer multiple of a, plus an integer multiple of b, where the integer multipliers may be negative. So in algebraic terms we are looking at numbers of form ax+by where x and y are integers. Hence we are trying to find the smallest number of form ax+by, for any integers x and y.

Think in terms of measuring. Suppose d = ax+by. I.e. d can be measured using a and b. Then anything that can be measured using d could also be measured using a and b, since we could use a and b to measure of copies of d repeatedly. We want to know when any two of these terms a,b,d, can be used to measure the same lengths. I.e. if d = ax+by, then anything measurable by d can be measured using a and b, and anything measurable using a and d could thus also be measured using a and b. But what about the converse? Can anything measurable by a and b also be measured using a and d? For example can you get b, just using a and d?

This is not so clear. In algebraic terms, we have the equation d = ax+by, but try to solve this equation for b. We get by = d - ax, but how do we get rid of the coefficient y in front of b? But what if we had an equation like d = ax+b. Then we could solve for b = d-ax, i.e. we could measure b using only a and d. Since we could also measure a using a and d, we could then measure anything using a and d that we could measure using a and b. So the point is to have an equation that has a coefficient of 1 or -1 in front of b, if we want to be able to eliminate b, and still measure the same lengths. I.e. if we have the equation d = ax+b, then we can eliminate either d or b and still measure the same lengths. I.e. the three term principle here would say that a and b measure the same lengths as a and d.

So the three term principle would say that if you have an equation like ax+by = cz, then you can eliminate the number whose coefficient is 1 or -1, and not change the lengths you can measure.
Now this is exactly the kind of equation you get when you divide one

length by another. I.e. suppose we start with two lengths a and b, and want to know what lengths we can measure, in particular we want to know the shortest length we can measure. Supposing $a < b$, if we divide a into b we get an equation like $b = ax + r$, where $0 \leq r < a$. Now in this equation, both b and d occur with coefficient 1, so we can eliminate either one and not change what lengths we can measure. Thus the lengths we can measure with a and b, are the same as the ones we can measure using a and r.
So we can forget about b, and consider the two lengths a and r. This is closer to finding the shortest length we can measure since a and r are smaller numbers than a and b. So we repeat the process to get still smaller numbers. I.e. we divide a by r, getting an equation like $a = ry+s$, where $0 \leq s < r$. Now a and r measure the same numbers as r and s, so we can eliminate a, and keep just the smaller numbers r and s. Since the remainder keeps getting smaller, eventually it becomes zero, and we get something like $r = ts + 0$. Then s actually measures r. Since r and s together measure the same numbers as the original pair a and b did, and s measures r, we do not need r at all, and in fact just s measures the same lengths as the pair a and b. In particular s measures both a and b, so it divides them.

On the other hand, a and b together measure the same numbers as r and s, so a and b measure s. I.e. there is some equation of form $s = au+bv$, where u and v are integers, possibly negative. Thus since any number that divides both a and b, must by the old three term principle also divide s, it must also be smaller than s. Thus we have again proved not only that s is the largest common divisor of a and b, but that all other common divisors of a and b divide it. This time into the bargain, we have actually found a procedure for computing the gcd of a and b.

Just try it to see how easy it is. I.e. say $a = 14$ and $b = 19$. Then $19 = 14 + 5$, so replace 14 and 19 by 14 and 5. Now $14 = 2(5) + 4$, so replace 14 and 5 by 4 and 5. Now $5 = 4 + 1$, so replace 4 and 5 by 4 and 1. Now 1 divides 4 so discard 4 and keep only 1, the gcd of 14 and 19.

Or try 68 and 142. We get $142 = 2(68) + 6$. Now $68 = 11(6) + 2$. Now $6 = 3(2)$, so 2 is the gcd of 68 and 142.

Try 8,439 and 615. We get $8,439 = (13)615 + 444$. Then $615 = 444 + 171$. Then $444 = 2(171) + 102$. Then $171 = 102 + 69$. Then $102 = 69 + 33$. Then $69 = 2(33) + 3$. Now $33 = 11(3)$, so 3 is the gcd of 8439 and 615. (I did these in my head, but not too quickly. It is easy to believe that computers can calculate gcd's extremely quickly.)

Five characterizations of GCD's

It is useful to have several different ways of describing the same thing. One description may be useful for calculating efficiently, another for geometric understanding, another for certain applications, and so on. The concept of a greatest common divisor has many such descriptions. The first is given by the name, i.e. the g.c.d. of integers a and b is the largest of the integers that divide both of them. But we have seen other descriptions too, and a different one is given in the book as the "definition" of a gcd. The book says (p.13) a gcd of a and b is a positive integer d such that d divides both a and b, and such that every integer that divides both a and b divides d.

So the simpler definition says only that d is a common divisor, and that every other common divisor is smaller than d. The second definition is more sophisticated, it says not only is every other common divisor smaller than d, but in fact every other common divisor is a factor of d. The problem with this definition is that it is not obvious there is any such number. I.e. it is not obvious that the *largest* common divisor really is *divisible* by every other divisor, so that must be proved in order to justify the existence of the gcd. The book chooses to prove that fact first, and then given the definition afterwards.

I prefer to give the naive definition, and then prove as a theorem that the gcd has the special property given in the book. My hope is that by always doing things in the least sophisticated way, they are easier to understand. It takes longer to do this, but I hope it is justified by better understanding, and longer retention time.

So our definition is the simplest, but the books version is useful in applications. We also need a good way to calculate a gcd in practice. This is where the Euclidean algorithm, built on the division algorithm comes in. We give now all the ways we have of describing a gcd of two natural numbers.

Theorem:

Let a,b be natural numbers (i.e. positive integers). Then a natural number d is the gcd of a and b, if and only if one of these equivalent properties holds:

1) d divides both a and b, and d is the largest integer that divides both a and b.

2) d divides both a and b, and every other integer that divides both a and b, also divides d.

3) d is the smallest positive integer that can be written as a "linear combination" of a and b. I.e. there exist integers x,y, (not necessarily positive) such that d = ax+by, and every other positive integer that can be written in the form an+bm, with n,m, integers, is larger than d. (In the

original Greek conception, using geometry, d is the smallest length that can be measured using two measuring sticks, of lengths a and b.)

4) Consider the following procedure. Begin with two positive integers a and b, with a smaller than b. Divide b by a and let r be the remainder. If r >0, replace a and b, by the pair a and r, where r is less than a. Divide a by r and let s be the remainder. If s > 0, replace a and r by the pair r and s. .....Continue until the remainder is zero. As soon as the remainder of this process becomes zero, the last number divided by (the previous remainder) is the gcd of a and b. For example, if the first remainder r is zero, on dividing b by a, then a is the gcd. If the second remainder s is zero. on dividing a by r, then r is the gcd.

5) Suppose a, and b are given integers, and let $p_1,....,p_r$ be the primes that occur as a factor of either a or b. I.e. all these primes are different, and we include a prime p in our list if either p divides a, or p divides b. Then we can write the prime factorization of a, uniquely as $a = p_1^{n_1}\cdots\cdot p_r^{n_r}$, where some of the exponents may be zero. We may also write the prime factorization of b as $b = p_1^{m_1}.....p_r^{m_r}$, where again some of the exponents are zero. For each index j, i.e. each prime p_j, let $s_j = \min(n_j,m_j)$ be the smaller of the two exponents, that with which p_j occurs in a, and that with which p_j occurs in b. Then $d = p_1^{s_1}.....p_r^{s_r}$, is the gcd of a and b.

Remarks:

Notice these descriptions yield completely different ways of calculating a gcd. The first one asks us to locate all common prime factors of a and b, see which is the largest. This could be time consuming, since two numbers can have a lot of common factors. The second condition would be even more difficult to carry out, since we would have to test the largest common factor to see if it were divisible by all the other factors, possibly very impractical. Condition 3 is completely unworkable, since it asks us to look at all the infinitely many linear combinations ax+by that can be formed from all integers x and y, and see which is the smallest positive one. This seems largely hopeless. Condition 4 is very fast and efficient, and yields a result by hand very quickly. So although it is not obvious at first why the result of this computation is the gcd, once one knows it, one can get the result quickly and efficiently. The third description on the other hand is useful in proving the fact that the gcd is divisible by all other common divisors. It also allowed us to prove the divisibility property of primes. So each description has its uses. The 5th description is to be proved by you as an exercise. It is somewhat computable, once you find the prime factorization of both numbers, but is usually not as fast as #4.

Example: Consider the integers 1278 and 371. To find their gcd, by property 4, we divide 371 into 1278 getting 3 with remainder 165. Then we replace the pair 1278 and 371, with the pair 371 and 165. Now divide 165 into 371, and get 2 with remainder 41. Then we could quit since 41 is prime so the gcd of 41 and 165 is 1, since 41 does not divide into 165. But if we continue, we divide 41 into 165 getting 164, with remainder 1. Now we are done. I.e. 1 divides into 41 with remainder 0, so the last number divided by, namely 1, is the gcd. Of course if you get down to 1, you know you are done, since no smaller positive number is possible.

We can also define gcd's for pairs integers that are not positive, as follows (as long as they are not both zero). If a is a positive integer, then the gcd of a and 0, is a. Notice this is the largest of the common divisors of a and 0. If a and b are both positive integers, then the gcd of -a and b, or of -a and -b, is the same as the gcd of a and b. Again gcd(-a,b) = gcd(a,b) is the largest of the common divisors of -a and b. The pair of numbers 0 and 0, do not have a gcd, since every integer, no matter how large, divides both 0 and 0. For this pair, descriptions 1 and 3 would be different, since only 0 can be obtained as a linear combination of 0 and 0. Property 3 still holds for any pair of integers a,b (not both zero). I.e. whether a and b are positive or not, if d is their gcd, then there exist integers x,y, such that d = ax+by.

Why does the Euclidean algorithm (condition #4) work?

The correctness of procedure #4, is based on property 3, of linear combinations. I.e. suppose we know a < b, and that d = gcd(a,b), is the smallest positive number that can be obtained as a linear combination of a and b, with integer coefficients. Then divide b by a, getting b = an+r where r ≥0. If r = 0, then a divides b. So a is a common divisor of a and b, and there can be no larger divisor of both a and b, so gcd(a,b) = a, in that case. If r > 0, then we claim gcd(a,b) = gcd(a,r).

To see this recall that gcd(a,b) is the smallest positive integer which can be written as a linear combination of a and b, and gcd(a,r) is the smallest positive integer that can be written as a linear combination of a and r. So if we could show that every number which can be written as a linear combination of a and b, can also be written as a linear combination of a and r, and vice versa, it would follow that these pairs have the same gcd. I.e. if you were looking for the smallest positive number writeable as a linear combination, both pairs would give the same result.

But we have an equation b = an+r, which shows that b can be written as a linear combination of a and r. And if we solve it as r = b-an, we see that r can be written as a linear combination of a and b. Thus a and b can each be written as a linear combination of a and r. Using substitution,

then any linear combination of a and b can be written in terms of a and r, and vice versa. Thus gcd(a,b) = gcd(a,r).

Now we divide a by r, getting a = rm+s where $0 \leq s < r$. This time if s = 0 then as before r = gcd(a,r) = gcd(a,b). If not we replace and r by r and s, and divide again. Now since the remainders keep getting smaller, eventually the remainder will be zero. Then the last divisor will be the gcd of the last pair. But all the gcd's are equal, so that last divisor is also the gcd of the original pair.

How to use the Euclidean algorithm to actually solve the equation in condition #3.

By substituting backwards, from the various steps of the Euclidean algorithm, we can actually find integers x and y such that gcd(a,b) = ax+by.

For instance in the example above, computing gcd(1728,371), we found the equations, 1278 = 3(371) + 165, 371 = 2(165) + 41, 165 = 4(41) + 1. Thus substituting backwards gives, 1 = 165 - 4(41). Now substitute for 41, getting 41 = 371 - 2(165), so we have 1 = 165 - 4[371 - 2(165)], and simplify getting 1 = 165 - 4(371) + 8(165) = 9(165) -4(371). Now substitute for 165, using 165 = 1278 - 3(371), so we get 1 = 9(165) -4(371) = 9[1278 - 3(371)] - 4(371), and simplifying gives 1 = 9(1278) -31(371). And this seems to check.

Relatively prime integers.

A version of the divisibility property of primes is actually true for "relatively prime" pairs of numbers as follows.

Definition: Two integers a,b, not both zero, are called "relatively prime" if their gcd equals 1.

Theorem: Two integers a,b, not both zero, are relatively prime if and only if one of the following equivalent properties holds.

1) The only positive integer that divides both of them is 1.

2) No prime number divides both of them.

3) There exist integers x,y (not necessarily positive) such that ax+by = 1.

See if you can prove this. It should be very easy, if you review the meanings of the words, and the previous results we know. You should also be able to prove the following divisibility property analogous to that for prime numbers as an exercise.

Theorem: If a and b are relatively prime integers, and a divides the product bn, where n is any integer then a divides n.

Proof hint: Try to use the same proof we gave for the prime divisibility property. **QED.**