4000/6000 Rational numbers

Remember the Greeks used numbers as a means of measuring lengths, in terms of a given unit, i.e. of comparing two lengths. Thus given a unit length, multiples of it were assigned whole numbers, 1,2,3,4,,,,, according to how many copies of the unit it takes to measure the desired length. On the other hand, given two lengths, in order to compare them, they sought out a common unit, a unit length short enough to measure both given lengths. For instance given two lengths, say the first one 12 feet long, and the second one 20 feet long, one might use a unit of one foot to measure both of them. Then the two lengths are measured by the two integers (12,20) and together these give the "ratio" 12:20 of the two given lengths. In this way two lengths give us a pair of integers 12:20. However this pair of integers depends on the unit chosen to measure the lengths. It is possible in this case to choose say a measuring stick of length 2 feet, and then it takes 6 copies to measure the first length and 10 copies to measure the second length, giving us the pair of integers 6:10. This pair represents the same ratio as the previous pair. Other units are possible, e.g. we could choose a stick of length 4 feet, and get the pair 3:5. This is the largest unit we can use to measure both these lengths by whole numbers, and 4 is indeed the "gcd" i.e. greatest common divisor, or "greatest common measure" for the two lengths 12 and 20. When we use the greatest common unit, the two integers we get, in this case (3,5), are always relatively prime.

Thus if we want to represent the ratio between the two lengths as a "number", then we need to use two integers to represent this new type of number. Moreover there may be many pairs of integers representing the same number. These new numbers, are represented by pairs of integers a:b, where a:b and na:nb represent the same ratio. I.e. any "multiple" of the pair a:b represents the same ratio. Thus suppose a:b and c:d are two pairs of integers, and suppose some multiple of the pair a:b equals some multiple of the pair c:d. I.e. suppose there are integers n,m such that an:bn is the same pair as cm:dm, i.e. such that an = cm, and bn = dm. Then we must have a:b = an:bn = cm:dm = c:d. Thus we should say two pairs a:b and c:d represent the same number if there are integers n,m such that an = cm and bn = dm. This is a little "clunky" to use as a rule, so we relate it to the usual "cross multiplication rule" as follows:

Lemma: Given two pairs of integers a:b and c:d such that $b \neq 0$ and $d \neq 0$, we have ad = bc if and only if there are non zero integers n,m such that an = cm and bn = dm.

Proof: First assume the cross multiplication rule holds. I.e. assume that ad = bc. The just take n = d and m = b, i.e. multiply both a and b by d, replacing the pair a:b by ad:bd. Then multiply c and d by b, replacing the pair c:d by the pair cb:db. Then compare the pairs ad:bd and cb:db. These two pairs have the same second entry by commutativity, and the same first entry by the assumption ad = bc, hence they are the same.

Conversely assume there are integers n,m such that an = cm and bn = dm. Then we claim that ad = bc. Well we need to get a and d on the same side. So starting from an = cm, multiply both sides by d, to get

adn = cdm. Now we need b together with c so multiply through by b, to get adnb = cdmb. Now use the hypothesis that bn = dm, so we can cancel the nb on the left, and the dm on the right. (This is where we need that the integers b,d,n,m are non zero.) This leaves ad = cb = bc, as desired.

QED.

Equivalence of ratios: Now we can define two pairs a:b and c:d where a,b,c,d, are integers, and $b \neq 0$, $d \neq 0$, to be equivalent if and only if ad = bc, or what is the same, if and only if there are non zero integers n,m such that an =cm, and bn = dm. We denote the equivalence class of the pair a:b by the symbol a/b or $\frac{a}{b}$.

These new numbers are called "rational numbers", or "numbers to represent ratios" between two (commensurable) lengths. Notice that two lengths can only be compared in this way if a common unit can be found for both of them. (The Greeks assumed at first this was always true, until they discovered "incommensurable" lengths and were forced to invent irrational numbers. We will discuss those in the a later section. For now we stick to rational numbers.)

Another way to view these rational numbers, is as numbers needed to measure lengths in terms of a given unit. I.e. fix a unit of measure once and for all, say a yard. Then given another length we try to measure it using our unit measure. We want to assign a number to the new length, which represents the ratio of our new length to the fixed unit length. If the new length is a whole number of copies of our unit we get an integer for its length. If not, we start trying to subdivide our unit until we have a new unit so short that it evenly measures the new length. (Again the Greeks assumed this was always possible.)

For example, suppose our unit is a yard, and we want to measure a length that happens to be 14 feet. Then we can divide can divide our yard stick into 3 equal pieces and then we see that if we use one of these pieces as new unit, the new length is exactly 14 of these pieces, while our fixed unit took exactly 3 of these pieces. Then we assign the number pair 14:3 to the new length, to represent the ratio between the new length and the fixed unit. Thus once we choose a unit, we expect every length to be measured by one of these number pairs a:b where both a and b are integers, and b is not zero, since it represents the non zero length of the fixed measuring stick. (A measuring stick of length zero would not be of much use after all.)

If we had not foreseen just the right number of pieces to cut our yardstick into, we might have chosen to subdivide it into 36 pieces, i.e. into inches. Then we would have needed 12(14) = 168 of these pieces to measure our new length, and thus would have gotten the number pair 168:36. Since this is just a further subdivision of our previous case, i.e. 168 = 14(12), and 36 = (3)(12), the number pairs 14(12):3(12) and 14:3 should again be considered equal, since they represent the same length. I.e. 168/36 = [(14)(12)]/[(3)(12)].

Notice that if we subdivide our original unit into say n parts, to get a new unit, and then it only takes exactly one of these new units to measure our new length, we assign the new length the measure 1:n. Thus our new unit has length 1/n. In this sense, if it took m of these pieces to measure some other length, that new length would be represented by the number m/n which means m copies of the new unit length 1/n. Thus m multiplied by 1/n should be m/n. But if we want multiplication to be commutative, we also need m/n to be 1/n multiplied by m. Thus multiplying a length by 1/n means subdividing the length into n equal parts. A length of 1/m, represents our unit divided into m equal parts, so to divide it again into n equal parts, meansa we have a new length which is equal to our original unit divided into mn equal parts. E.g. if we take 1/3 of a yardstick and subdivide it again into 5 equal parts, it takes 5 of these parts to make each third of our original unit, hence it takes 15 of these parts to make one original unit. Thus multiplying 1/n by 1/m should give 1(mn). Similarly if we two lengths shoiuldf mean the length we get by putting them end to end, so if they are measured in the same units, we should justa dd the number of these units. Since a/n represents a length of a copies of the unit 1/n, and b/n represents b copies of that unit, we should define a/n + b/n to be (a+b)/n. Guided by all this geometric motivation, we just define rational numbers now formally, and check they make sense.

Definition of rational numbers
A “rational number” is represented by an ordered pair (a,b) of integers where the integer b is non zero, and the pairs are subject to the following equivalence relation (written as equality): $(a,b) = (c,d)$ if and only if $ad = bc$. The equivalence class represented by the pair (a,b) is called a rational number and is denoted by a/b, or $\frac{a}{b}$. Thus a/b makes sense if and only if a,b are integers, and $b \neq 0$, and $a/b = c/d$ if and only if $ad = bc$.

Arithmetic of rational numbers.
The set of all rational numbers is denoted by the capital letter “Q”, or a special one like \$, (which stands for “quotients” I suppose).

Basic fact: The rational numbers form a field. Moreover, Q is the smallest field containing the integers Z.

Definition of addition in Q.
Recall that a/b meant that we subdivided our original unit into b parts creating a new unit of length $1/b$, and then we took a copies of that unit. To add measurements they need to be in the same units, so to add a/b to c/d we could put them in the same units by changing their representatives to be ad/bd, and bc/bd, since $a/b = ad/bd$, and $c/d = bc/bd$. Thinking in terms of adding ad copies of the unit $1/bd$, to bc copies of the unit $1/bd$, gives $(ad+bc)$ copies of that same unit, i.e. given a/b and c/d,
define: $a/b + c/d = (ad)/(bd) + (bc)/(bd) = (ad+bc)/bd$.

Does our addition make good sense?
We must check this definition gives the same answer if we use different representatives for our rational number. I.e. we must check that if $a/b = u/v$, and $c/d = r/s$, then also $a/b + c/d = u/v + r/s$. This kind of thing is tedious, like doing finger exercises on the piano. But we want to learn to play, so here it is. Besides if this were not true we might as well stop here.
Assume if $a/b = u/v$, and $c/d = r/s$. I claim that $a/b + c/d = u/v + r/s$, i.e. that $(ad+bc)/bd = (us+vr)/vs$. Let’s see now, to check that means checking that $(ad+bc)vs = (us+vr)bd$. I don’t have any great ideas, so let’s just multiply out and see what happens. We want to show

(*) $(advs+bcvs) =? (usbd+vrbd)$.

We get to use the hypotheses that $a/b = u/v$, i.e. that $av = bu$, and also that $c/d = r/s$, i.e. that $cs = rd$. Substituting with these equations into the left side of (*) gives $(advs+bcvs) = buds + bvrd$. Hey that does it! That equals the right side of (*), after using commutativity to rearrange the factors. **QED.**

It seems to me that proof was made harder by using the cross multiplication version of equivalence. Let’s see if it looks easier with the “scaling” version of equivalence. I.e. if $a/b = u/v$, and $c/d = r/s$, then there are non zero integers n, m, x, y such that $an = um$, $bn = vm$, and $cx = ry$, $dx = sy$. Then look at the two sums $(ad+bc)/bd$ and $(us+vr)/vs$. Just multiply the first one top and bottom by xn, and the second one top and bottom by ym, and they become the same. I.e. using all the equations, one checks that $[(ad+bc)xn]/[bdxn]$ is the same as $[(us+vr)ym]/[vsym]$. Lets check they have the same bottoms. By assumption, $bn = vm$, and $dx = sy$, so multilying these equations gives $bndx = vmsy$, i.e. the bottoms of the two fractions agree. Similarly the tops agree. Well, this was really not much better than the other way. OK, I admit it, this proof is not a pretty sight. See if you like the computation at the bottom of page 46 any better.

At the risk of trying your patience, let me make another explanation of why we should expect that addition of rational numbers will give equivalent answers, i.e. the "same" answer, when you use equivalent integer pairs. Most of us probably believe that those rational numbers represent lengths and that adding two lengths gives a unique answer. So since we set up the addition to always give the length obtained by setting two lengths end to end, we believe the length should not depend on how many parts we subdivide it into. But this argument depends on believing in the concept of length. The algebraic proof above says that our addition makes sense on its own, without reference to geometry. But of course no one would have thought of this goofy definition of addition of fractions if it were not motivated from the geometry of measurement.

Definition of multiplication in Q.
As we argued above, we want b(1/b) = 1, and (a/b) = a(1/b). Thus we want (a/b)(c/d) = (ac)(1/b)(1/d) = x. Thus multiplying both sides by bd, we get (ac) = (bd)x. Now multiply both sides by (1/bd), to get
x = (ac)/(bd). I.e. (a/b)(c/d) = (ac/bd). This is our definition.

Definition: The product of two rational numbers a/b and c/d is defined to be (a/b)(c/d) = (ac/bd).

Notice the definitions of adding and multiplying both use the "domain" property of Z, since we need to know that bd is $\neq 0$ whenever $b\neq 0$ and $d\neq 0$, for both operations to make sense. I.e. ac/bd would not be a rational number unless bd $\neq$ 0. In fact this property is all that is needed. I.e. anytime you come across a domain you can construct a smallest field this way which contains it. But so far, we do not need this fact, since the other domains we know, Z_p with p prime, are already fields.

Does our multiplication make sense?
For this we should again check this definition is well defined independent of representation of the numbers. Without giving the complete proof, the idea is that if we multiply (a/b) by (c/d) we get (ac)/(bd), and essentially the only way to change representatives of these fractions is to "scale" them, i.e. to replace a/b by (an)/(bn), and c/d by cm/dm. Then notice that using these reporesentatives, gives (ancm)/(bndm) = (ac)(nm)/[(bd)(nm)] = (ad)/(bd), i.e. the same answer. I do not choose to do use the official cross product definition method. You might try it for practice. It should be easier than the addition case.

ARRGHHH.......
Then we should check that these definitions make Q into a field. I.e. we need to check associativity, commutativity, identities, inverses, and distributivity. I will check only the easiest ones. (The implication is that as infinitely strong and indefatigable students, who have nothing else to do, you will do all the harder ones, I guess. Just kidding!!)

Identities: The additive identity is 0/1, since then (0/1) + (a/b) = (0b+1a)/1b = a/b. The multiplicative identity is 1/1 since then (1/1)(a/b) = (1a)/(1b) = a/b.

Inverses: Lets see, the additive inverse for a/b is (-a)/b, since then a/b + (-a)/b = (ab + (-ab))/b^2 = 0/b^2 = 0/1.

The multiplicative inverse for a/b exists only when a/b $\neq$ 0/1, i.e. when a $\neq$ 0, and then $(a/b)^{-1}$ = (b/a), since this makes sense and (a/b)(b/a) = ab/ab = 1/1.

AHA> another easy one: associativity for multiplication: (a/b)[(c/d)(e/f)] = (a/b)(ce/df) = [a(ce)]/[b(df)] = [(ac)e/[(bd)f] = [(a/b)(c/d)][e/f].
Commutativity for multiplication looks easy too.

Let me show good faith by doing a harder (?) one,
distributivity. I need to show that
(a/b)[c/d + e/f] = (a/b)(c/d) + (a/b)(e/f). The left side is (a/b) [(cf+de)/ df] = (acf+ade)/bdf, and the right side is (ac/bd) + (ae/bf) = (acbf + aebd)/bdbf = (acf+aed)/bdf. This seems to check but you must help me out here. Did I make any mistake?? It is moderately interesting that there is the need to cancel the b here. I wonder why that is?

What is the geometric "reason" that distributivity is true? I guess it says that if you measure off n copies of a given unit, and then cut the unit into two pieces a and b, then to get the same length, you have to measure off n copies of a, and also n copies of b. I.e. n(a+b) = na + nb.

Ordering rational numbers.
Since rational numbers measure lengths on a line, it should not be too shocking that one can order the rational numbers. I.e. starting from an origin, one length is longer than another if it reaches further to the "right" from the origin, i.e. if the shorter endpoint falls "between" the origin and the longer one. To express this in terms of symbols, i.e. rational numbers, instead of lengths, look at a/b and c/d, with a,b,c,d all positive, and ask which is greater. First put them in the same units, i.e. ad/bd and bc/bd. then a/b is larger than c/d, if and only if ad > bc. I.e.
a/b > c/d, if and only if ad > bc. I can never remember this last rule myself, so I recommend to always change a/b and c/d, into the form ad/bd and bc/bd, before deciding which is larger.

WARNING: This rule does <u>not</u> work for negative rational numbers!! unless you are careful. I.e. notice that any positive number is larger than any negative one, so 1/3 > 2/(-3), but 1(-3) is not greater than 2(3). Thus when using this rule, you must always represent your rational number by a fraction with positive denominator. This is always possible. Then it does seem to work. (Do you agree?)

Alternate definition via Positivity:
The definition in the book of ordering is nicer, since he just defines "positive". I.e. a rational number a/b is positive, i.e. a/b > 0, if and only if ab > 0. Then a/b > c/d if and only if (a/b)-(c/d) > 0. Notice this explains our rule above, since (a/b)-(c/d) = ad/bd - bc /bd = (ad-bc)/bd, and saying this is positive is the same as saying ad-bc is positive, provided b and d are positive.

Properties of ordering:
A number is positive if it is greater than 0. Then the sum of two positive numbers is positive and the product of two positive numbers is positive.
Moreover every number satisfies exactly one of these three properties: either it is positive, or it is

zero, or its inverse is positive.

Multiplying by a positive number preserves inequalities and multiplying by a negative number reverses them. I.e. $a > 0$ and $b > c$ implies $ab > ac$, while $a < 0$ and $b > c$ implies $ab < ac$.

A field with an ordering is called an ordered field.

Theorem: Q is an ordered field.

Rational numbers in "Lowest terms"

A very useful fact, is that every rational number has a unique representation in "lowest terms". I.e. each rational number has a unique representation in the form a/b where b > 0, and where gcd(a,b) = 1. Try proving this yourself. For example, 224/336 is not in lowest terms, but if we factor the top and bottom we get 2(112)/2(168) = 112/168 = 2(56)/2(84) = 56/84 = 2(28)/2(42) = 28/42 = 2(14)/2(21) = 14/21 = 2(7)/3(7) = 2/3. This last representation is in lowest terms.

To get the unique "smallest" or "lowest terms" representation for a given rational number a/b we could factor both integers a and b into primes, and then cancel all primes which occur in both integers. I.e. let a,b be two given natural numbers, and let their gcd be d. Then we can write a = dn, and b = dm, where n,m cannot have any common prime factors, i.e. n,m are relatively prime. [Otherwise, if k were a common prime factor of both n and m we could write n = kr and m = ks, and then we would have a = dn = dkr, and b = dm = dks, and then dk would be a common divisor of a and b, and dk is greater than d, a contradiction.]

Thus every (positive) rational number can be written as a pair n/m where m > 0 and where gcd(n,m) = 1. In fact there is only one way to do this, as we prove next.

Lemma: Assume that a/b and cd are rational numbers with b > 0, d > 0, and that gcd(a,b) = 1 = gcd(c,d), and that a/b = c/d. Then we claim a = c and b = d.

Proof: How are we going to do this? We only have one hypothesis to apply to the equations ad = bc, and that is relatively primeness. So we must use the only fact we know about relatively primeness, namely the division property. Recall that if a,b are relatively prime and a divides bx, then a divides x. We want to prove a =c, so let's try at least to use this argument to show that a divides c. If we are lucky that will be enough. Since ad = bc, we know a divides the left, hence a also divides the right side, i.e. a divides bc. But a and b are realtively prime, so a divides c. I.e. c = an, for some integer n.

Similarly d divides both sides of the equation ad = bc, and gcd(c,d) = 1, so d divides b, say b = dm. Since b,d, are both positive, m > 0 also. Thus we have ad = bc = (dm)(an) = (ad)(nm). Then by the cancellation property we get nm = 1. Since m > 0, hence n > 0 also, and n = m = 1. Thus b = dm = d, and c = an = a. **QED.**

This says there is only way to put a given rational number into lowest terms.

Remark: Often bright students like yourselves can just memorize a complicated proof like this, cold turkey. But even then you will probably forget it promptly. Thus I do not advise this. The only way to really remember it, say next summer, or when you teach the subject three years from now, is to understand how it works. That is why I tried to comment above on how one could think of it, using relative primeness. To learn it you should get out a clean sheet of paper now and try to reproduce it. Probably you will not be able to at first, until you go through some thought process like that above, that involves recreating the argument for yourself. Once you do that, it will stay with you a long time, and if it goes away, you will be able to get it back more quickly the second time than the first. I am speaking from experience since when I was a student I tried to memorize these proofs without understanding them, and it was very hard to do. I understood them only after I started trying to explain them to a class. Preparing for class was very rewarding, but also time consuming because I had not really understood the ideas while taking the courses. Thus you should try explaining these to each other or to someone outside the class. You cannot understand how someone thought of a proof just by reading the proof over and over. You have to try to find the argument yourself, and then explain to someone how you found it, or at least why it works.

BIG NEWS: sqrt(2) is irrational. Note that a fraction a/b is in lowest terms if $b > 0$, and a and b have no common prime factor. Thus if a/b is in lowest terms, then so is a^2/b^2. Now suppose that a/b is a fraction in lowest terms such that $a^2/b^2 = 2$. I.e. assume a/b is a square root of 2. Then a^2/b^2 is in lowest terms, but $a^2/b^2 = 2/1$ and 2/1 is also in lowest terms. Since there is only one lowest terms form of a given rational numbers, we must have $a^2 = 2$, and $b^2 = 1$. Thus a is a natural number whose square is 2. This is impossible since 0^2, 1^2 and 2^2 are not equal to 2, and all the other squares of natural numbers are even larger. This contradiction shows there is no rational number whose square is 2. I.e. sqrt(2) is an "irrational" number, provided any such number exists at all.
History has it that the first person to tell this fact publicly was drowned, for revealing secret Pythagorean society information.

4000/6000 Day 11 "Real" Numbers

Rational numbers do not suffice to measure all lengths

Recall that we regard numbers as measurements of the length of line segments. Given a unit of length, we attempted to measure another length as follows. Subdivide the unit into n equal parts, so that some multiple m of one of those parts exactly measures our new length. Then the new length is assigned the rational number m/n. But what happens if we can never measure the new exactly with any unit obtained by subdividing our original unit into a whole number of equal parts? Then we say the new length is not measurable by the original unit, or that the two lengths are "incommensurable". Two lengths then are incommensurable if and only if one is not a rational multiple of the other.

Irrationality of sqrt(2)

First proof: The first proof that some lengths are not commensurable, apparently occurs in Euclid, Book X, and this is the proof presented in our book. The argument is that the side of a square cannot be commensurable with its diagonal. Using the Pythagorean theorem, this is equivalent to showing that no rational number has square equal to 2. Euclid's proof runs as follows. First one proves (easy exercise) that an even number has an even square, and an odd number has an odd square. Hence a^2 is even, if and only if a is even. Now assume that $(a/b)^2 = 2$, so that $a^2/b^2 = 2$, and hence $a^2 = 2b^2$. Since every rational number can be put in lowest terms, we may assume that a and b were relatively prime. Thus they are not both even. But since $a^2 = 2b^2$, a^2 is even, hence a is even. If say a = 2c, then $a^2 = 4c^2$, so $4c^2 = 2b^2$, hence $2c^2 = b^2$, so b^2 is also even, hence b is even. This is a contradiction so there cannot be integers a,b with $(a/b)^2 = 2$.

Second proof: The proof I gave in the notes is as follows: If a/b is in lowest terms it means a and b have no prime factor in common. Since a prime p divides a^2 if and only if p divides a, the prime factors of a^2 are the same primes as the prime factors of a. [Of course the primes occur with different exponents in a and a^2. E.g. if 14 = (2)(7), then $(14)^2 = (2^2)(7^2)$.] Thus if a/b is in lowest terms, also a^2/b^2 is in lowest terms. Now assume again that $(a/b)^2 = 2$, and that a/b is in lowest terms. Then $a^2/b^2 = 2/1$, and both sides of this equation are in lowest terms. Recall however that we proved there is only one way to put a rational number in lowest terms, so $a^2 = 2$, and $b^2 = 1$. Thus a is an integer whose square is 2. But there is no such integer, since $0^2 = 0$, $1^2 = 1$, $2^2 = 4$, and all the other integers have squares which are even larger.

While teaching a course like this I made up this proof, and I have never seen it in any book, (but as soon as I say that, we will probably find it in a book 250 years old). I am not saying it is better than Euclid's proof, but why keep giving the same proof for 2,000 years? It is nice to have a change once in a while. Besides we don't want to give the impression we have had no new ideas all these centuries.

Third proof: Here is another proof I read somewhere once. Notice that if $a = p_1p_2....p_r$, is the prime factorization of a, then $a^2 = p_1{}^2p_2{}^2....p_r{}^2$ is the prime factorization of a^2. In particular, every prime in the prime factorization of a square occurs an even number of times. But if $(a/b)^2 = 2$, then $a^2 = 2b^2$, and this means there are an even number of occurrences of the prime 2 on the left, but an odd number of 2's occurring in the prime factorization of the number on the right (an even number of 2's in b^2, and one more 2 in front of b^2). I.e. there are an even number of 2's in the prime factorization of a^2, but an odd number of 2's in the prime factorization of $2b^2$. This is a contradiction, so the equation $(a/b)^2 = 2$ cannot occur.

Main point: Mathematics is an ongoing process of building arguments out of basic tools. In the same way a child constructs tinker toy houses out of basic building blocks. The houses are the theorems, like the unique prime factorization theorem, or the irrationality of sqrt(2), and the building blocks are the useful lemmas like the (relatively) prime divisibility property. Some houses, like the ones just mentioned, are so nice they are worth remembering in their entirety, but more important are the tools. If you understand the basic tools you can build your own houses. And there are not as many tools as there are possible houses to build from them. So to give yourself a chance to understand math, and be able to actually remember it, and recognize patterns when you see them, you must practice understanding the proofs, and seeing what tools are used. Notice all the proofs given above of the irrationality of sqrt(2) are different in detail, but all of them use the prime divisibility property somewhere. As you read the proofs in the book, try to see which tools are used to prove each one. You know a theorem is really new and interesting if it cannot be proved using the old tools and you need a new tool, to prove it. Of course a really good theorem, like unique prime factorization, becomes a tool itself, to use to build other results. But some results are rather special, and it is hard to see how to use them to do more, like the irrationality of sqrt(2). What you can do however, is use the proof of this result to do other things. That is why it is useful to have as many proofs as possible of things like this. It is not that we want four ways to prove this one fact, but we want to be able to have a lot of new proof methods available for other uses.

So let me jump ahead now to a theorem proved in chapter 3 of our book, the "rational root theorem". The reason it fits here now, is it does not use a new idea, but just the same relatively prime divisibility property, and a generalization of the three term principle. The only things about polynomials it uses is "what is a polynomial?", and "what is a root of a polynomial?". Since we already know these things we can appreciate this result now. (The new results of chapter 3 concern the structure of the ring of polynomials, which we will treat later, although we could do it now, since it is highly analogous to the theory of integers. I.e. later we will discuss irreducible polynomials like X+4, and they will be analogous to prime numbers, and then we will have a division theorem for polynomials, and a theory of unique factorization of polynomials into irreducible ones.....)

Polynomials with integer coefficients.

Recall that to make a polynomial with integer coefficients, we start with a new symbol "X", and take non negative powers of it, $1 = X^0$, $X = X^1$, X^2, X^3,......, and then we take a finite linear combination of these powers, with integer coefficients, like $16X^6 - 63X^5 + 194X^3 - 11$. That is a polynomial with integer coefficients. Notice negative powers are not allowed, so $1/X = X^{-1}$ is not a polynomial. Thus a general polynomial f(X) looks like $f(X) = a_nX^n + a_{n-1}X^{n-1} + ... + a_1X + a_0$, and to say the coefficients are integers means the a_i are all integers. A "root" of a polynomial f(X) is a number c such that f(c) = 0, i.e. a root of a polynomial f(X) is a solution of the equation f(X) = 0. If c is an integer we call c an integer root, and if c is rational we call c a rational root, etc. Much of algebra is about finding out what number system we must use in order to find a root of a given polynomial. Notice that very few polynomials have integer roots, e.g. not even 2X-3, has an integer root, much less $X^3 + 2$. Although many problems in books are cooked up to have integer solutions, like $X^2-5X+6 = 0$, this is very artificial. The rational root theorem tells us what are the possible rational numbers which could be roots of polynomials with integer coefficients.

Theorem: (rational roots theorem). If $a_nX^n + a_{n-1}X^{n-1} +........+a_1X + a_0 = f(X)$, is a polynomial with integer coefficients, i.e. if all the numbers $a_0,....,a_n$ are integers, and if c/d is a rational root of f(X) expressed in lowest form, then c divides a_0, and d divides a_n.

Corollary: If the leading coefficient of f(X) is 1, i.e. if f(X) = $X^n + a_{n-1}X^{n-1} +........+a_1X + a_0$, then the only possible rational roots are the integers which divide a_0.
Proof: The denominator of a rational root in lowest terms must divide 1, hence the denominator would equal 1 or -1. **QED.**

Corollary: There is no rational number whose square is 2 (4th proof).
Proof: A number whose square is 2, is a root of the polynomial $X^2 - 2$. The only possible rational roots are the integer factors of 2, i.e. 1, -1, 2, -2, but none of these work. **QED.**

Proof of the theorem: Suppose c/d is a root, in lowest terms, of f(X) = $a_nX^n + a_{n-1}X^{n-1} +........+a_1X + a_0$. Thus when we substitute c/d in place of X we get zero. I.e. $a_n(c/d)^n + a_{n-1}(c/d)^{n-1} +........+a_1(c/d) + a_0 = 0$.
Thus $a_n(c^n/d^n) + a_{n-1}(c^{n-1}/d^{n-1}) +........+a_1(c/d) + a_0 = 0$. Now take Auslander's basic advice for dealing with fractions and multiply out the bottoms. The common denominator of these bottoms is d^n. Since $d^n/d^k = d^{n-k}$, when we multiply this whole equation by d^n, we get

$a_n(d^nc^n/d^n) + a_{n-1}(d^nc^{n-1}/d^{n-1}) +........+a_1(d^nc/d) + d^na_0 = 0$, or

$a_n(c^n) + a_{n-1}(dc^{n-1}) +a_{n-2}(d^2c^{n-2})+........+a_1(d^{n-1}c) + d^na_0 = 0$.

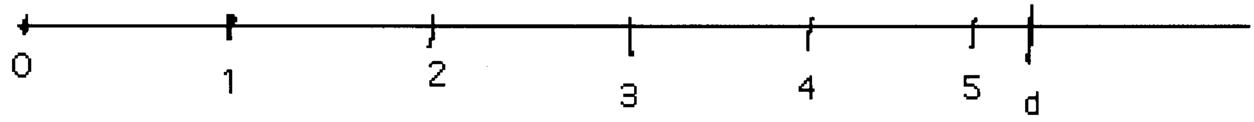
Now we use a generalized "three term principle". I.e. notice that on the left side that d divides every term except the first. Hence it divides also the first. To see this we could put the first term over on the other side by itself. I.e. $a_{n-1}(dc^{n-1}) +a_{n-2}(d^2c^{n-2})+........+a_1(d^{n-1}c) + d^na_0 = -a_n(c^n)$. Thus since d divides the term $a_n(c^n)$, and d is relatively prime to c, hence also to c^n, d must divide a_n, as claimed.

Now see if you can prove that c divides a_0 (warning: if you cannot, you have not understood this argument, and need to read it again). **QED.**

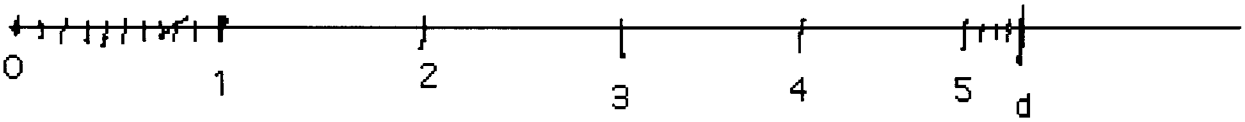
Remark: This theorem will help give us a huge supply of irrational numbers. I.e. next we will discuss real numbers and discuss why all polynomials of odd degree, with real coefficients, do have at least one real root. Then we can conclude that a polynomial whose first and last coefficients are both 1, but for which 1, -1 are not roots, must have an irrational root. E.g. the polynomial $X^5 -3X^3 + 1$ has a real root, but no rational root, hence it has an irrational root.

The Creation of all "real" numbers, both rational and irrational.

The upshot of the results above, is that if we want to use numbers to measure all possible lengths we need some more numbers. Rational numbers will not do the whole job. "Real numbers" can be described as follows: suppose we want to measure a given length d on a line, in terms of a given unit. We will assign a decimal to the length d as follows: Start by laying off as many copies of the unit as possible, starting from the origin 0, until the next one would go past the desired point to be measured. In the case below, in trying to measure the length d, this gives us the first digit of 5.



Next we divide the unit into 10 equal pieces and lay off further as many of these as possible until the next one would go past the point to be measured. Write this number as the "tenths" digit of the decimal being constructed. In the picture below we get 3 in the tenths place, for 5.3.



Then subdivide a tenth part of our unit again into ten equal parts and lay off further as many of these as possible, until the next one would go past the point. I am unable to draw any more of these subdivisions due to lack of manual dexterity and poor eyesight. In this way however you can imagine we get a decimal which may continue forever. For instance, if the length d is 1/3 of our original unit, since 3/10 < 1/3 < 4/10, our number starts out as .3. Then since also 33/100 < 1/3 < 34/100, it continues as .33. This keeps up and our decimal continues as .3333333......... Thus we get an infinite decimal representing our length of 1/3. However now not only simple rational lengths like 1/3 may be represented in this way, but any length. For instance since $(1.4)^2 = 1.96 < (1.5)^2 = 2.25$, for sqrt(2) our decimal starts out as 1.4. Then since $(1.41)^2 = 1.9881 < 2.0164 = (1.42)^2$, the decimal representation of sqrt(2) continues as 1.41. Then again since $(1.414)^2 = 1.999396 < 2.002225 = (1.415)^2$, our decimal representation of sqrt(2) continues as 1.414. In this way we get after an infinite number of steps, an infinite decimal, which represents our length sqrt(2). The decimal does represent the point uniquely in the sense that no other point can be assigned the same decimal. To see this note that any other point would have some finite distance from sqrt(2), hence would differ in its representation at some decimal place from this one.

Now given an infinite decimal how do we describe the point associated to it? We can proceed as follows. Given say .333333........, think of it as equivalent to an infinite sequence of finite decimals: .3, .33, .333, .3333, ... Then we can find each of the points represented by one of the finite decimals on the line. That gives us an infinite sequence of points. Notice that these points are always moving to the right, or at least never moving to the left. Then the actual point d represented by the infinite decimal can be described in words as "the point which is not to the left of any one of these "finite" points, and which is further to the left than any other such point". I.e. our point d is as far to the left as possible without being actually to the left of any "finite" point of our sequence. Note that this does assign the point 1/3 to the decimal .33333......., since the point 1/3 is not to the left of any point of the sequence .3, .33, .333, .3333, ..., but any point to the left of 1/3, is further to the left than most of these points. For example a point q which is say 1/1000 to the left of 1/3 will be to the left of any point which is closer to 1/3 than 1/1000, in particular q will be to the left of all the points .333, .3333, .33333, ...

Thus given a point, it is assigned an infinite decimal, and given that infinite decimal this process assigns back the same point. What about if we begin with a decimal? We do get a unique point. Then when we construct a decimal from that point, is it the decimal we started with? Not always. For example, the decimal .99999999......... is associated to the point 1, since 1 is not to the left of any of these points, but any point to the left of 1 is to the left of most of them. Hence 1 is the furthest point to the left which is still not to the left of any of these points. However when we assign a decimal to the point 1, we get the decimal 1.0000000.........

This is a general pattern. Any decimal which ends in all zeroes is associated to the same point as some decimal ending in all 9's. Thus if we decline to use any decimal ending in all 9's then there is a unique number associated to each point, and a unique point assigned to each decimal.

Definition of real numbers: A real number is an "infinite decimal", (which may end in all 0's), except we do not allow any decimal ending in all 9's.

"Lexicographic" ordering of real numbers: We can order our real numbers by saying an infinite decimal x is larger than another one y, if there is some finite place before which x is larger than y. E.g. x = .3333333222..... is larger than y = 333333222......., because up to 7 decimal places x is approximated by .3333333 and y is approximated by .3333332. We could not say this if we allowed a decimal ending in all 9's since then it would look as if 1.00000.... were larger than .999999..... although they are equal.

Upper bounds: Given a set S of real numbers, we say b is an upper bound for S if (and only if) no element of S is larger than b.

Terminology: A set of real numbers is said to be bounded above if it has an upper bound.

Least upper bounds. Given a set S of real numbers, d is the least upper bound for S, if d is an upper bound for S, and no other upper bound for S is less than d.

Then the main result is the following theorem:
Least upper bound theorem: Every non empty set of real numbers which is bounded above, has a unique real least upper bound.

Adding real numbers.
Then we can add real numbers as follows. Given two positive infinite decimals, we approximate them both up to kth order and add these approximations. We do this for all orders of approximation, and obtain a sequence of approximations to their sum. Then we define the sum to be the least upper bound of these approximations. E.g. to add .333333....... and .44444444........., we form the two sequences .3,.33,.333,.3333,...... and .4,.44,.444,.4444,....... and add these to get the sequence .7, .77, .777, .7777, ... Then the sum is the least upper bound of this last sequence. We can actually evaluate this sum in this case since it is repeating, to get a rational

number as we show next.

Rational numbers as "repeating" decimals.
Notice that all finite decimals are rational, e.g. .3345 = 3345/10,000, but not all rational numbers are finite decimals. E.g. 1/3 = .33333....... is rational. The result is that a decimal is rational if and only if it is eventually "repeating", in the sense that after a while, where it does possibly anything at all, eventually there is a block of numbers that repeats forever. E.g. 129.28736124242424.......where the block 24 repeats forever. This block could be very long. For instance if all I know is that a number starts out as 1.39284673.......it might be repeating if after this it keeps up as 1.392846739284673928467.......repeating 3928467 forever. On the other hand even if a number starts out repeating, it may not be rational if the pattern does not continue forever. E.g. Euler's famous number e starts out as 2.718281828......but the 1828 pattern stops after that, and there is no other pattern that goes on forever. You have an (extra) homework problem to prove in fact that e is not rational.

To compute what rational number a repeating decimal equals is easy. For instance set x = .333333......, and then 10x = 3.3333333...... Thus 10x - x = 9x = 3 (why?) so x = 3/9 = 1/3.
To prove this we can use the famous geometric series formula that $1+r+r^2+r^3+...... = 1/(1-r)$, at least when $|r| < 1$. Thus .3333333...... = (3/10)+(3/100)+(3/1000)+.......
= (3/10)(1 + 1/10 + 1/1000 + 1/10000 +......)
= (3/10)(1/[1-{1/10}])
= (3/10)(1/[9/10]) = (3/10)(10/9) = 3/9 = 1/3.

Recall from page 4 the proof by induction that $1 + r + r^2 + ...+ r^k = \frac{1-r^{k+1}}{1-r}$. If $|r| < 1$, then as $k \dashrightarrow \infty$, the term $r^{k+1} \dashrightarrow 0$, hence the result. I.e. when $0 < r < 1$, the "sum" of the infinite series $1+r+r^2+r^3+......$, i.e. the smallest number not smaller than any of the numbers $1 + r + r^2 + ...+ r^k$, i.e. the smallest number not smaller than any of the numbers $\frac{1-r^{k+1}}{1-r}$, is $1/(1-r)$.

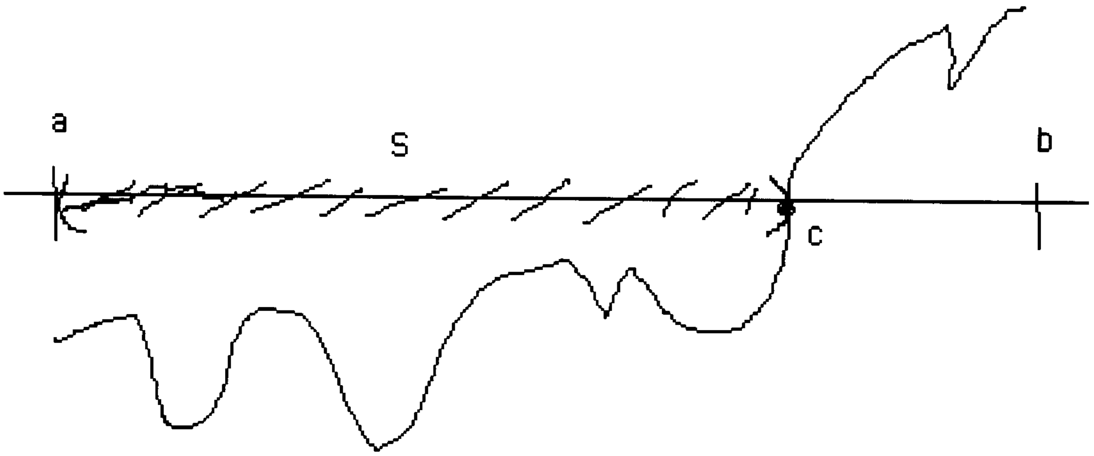
This type of argument shows that every repeating decimal is a rational number. It is a homework problem to show conversely that every rational number has a repeating decimal expansion. To get an idea, I suggest dividing out some rational numbers to see when, and hopefully why, they begin repeating. E.g. 2/7 = .285714285714285714........ Why did it start repeating when it did?

Calculator numbers. Note that a calculator displays only a finite number of decimal places, on a small calculator only 7 or 8. Thus not only are all calculator numbers rational, but often of the extremely special form $n/(10)^7$. Thus calculators always approximate, and cannot be counted on for exact answers. Still they are very useful.

There is a real number whose square is 2.
The modest sounding "least upper bound" property is incredibly powerful, and can be used to produce many, many real number solutions to equations that have no rational solutions at all. The most efficient way is the approach taken in calculus, to introduce the concept of continuity. Let f be a real valued function f of a real variable, i.e. a function whose inputs and outputs are both real. the main property of continuity is this. If f is continuous, and f(a) is

negative, then f(c) is also negative for all c near a. The same holds for positive values of f. Then we can easily prove the famous intermediate value theorem.

Theorem(IVT): If f is continuous and defined on the interval [a,b], and if f(a) < 0, while f(b) > 0, then there is some number c with a < c < b, such that f(c) = 0.
Proof: This is by contradiction. Let S = {all points x such that a ≤ x ≤ b and f(x) < 0}. Then b is an upper bound for the set S which is non empty since a is in S. Let c be the least upper bound of S. See the picture below.



We will show that the assumption f(c) ≠ 0, leads to a contradiction.
Suppose for example that f(c) < 0. Then also for all points x near c, we would have f(x) < 0. In particular for a point x slightly larger than c we would have f(x) < 0, i.e. any element slightly larger than c would belong to S. But this violates the fact that c is larger than all elements of S. Now you prove that f(c) > 0 leads to a similar contradiction. **QED.**

Theorem: All polynomials are continuous.
proof: (Omitted. I.e. it takes work.)

Corollary: There is a real number whose square is 2.
Proof: If $f(X) = X^2-2$, then f(1) = -1 < 0, and f(2) = 2 > 0. Thus there is some real number c between 0 and 1 such that $f(c) = c^2 - 2 = 0$. Thus $c^2 = 2$. **QED.**

Corollary: Every real polynomial of odd degree has a real root.
Proof: (Idea) Take for example a cubic such as $f(X) = X^3-X+1$. Then f is dominated for large values of X by its leading term X^3, so as X-->∞,
f(X) -->∞ also. In particular f(b) > 0 for large enough b. Similarly,
as X--> -∞, f(X) --> - ∞, so for small enough a, f(a) < 0. Thus for some c, we have f(c) = 0.
QED.

4000/6000 Day 12 More real numbers

Recall real numbers are defined as infinite decimals, where we agree either that decimals which end in all 9's are not allowed, or better, that such a decimal equals the corresponding infinite decimal ending in all 0's. E.g. the decimal 6.245999999..... is the same real number as 6.24600000.......

Then we saw that every length on the real line could be represented by one of these numbers, and less clearly, that every one of these numbers represents a length, where the latter statement involved assuming there really is a point which is furthest to the left, among all points which are not further to the left than any one of an infinite sequence of points representing rational lengths. I.e. we had to assume a bounded infinite sequence of points corresponding to rational lengths, has a "least upper bound". Although we had to assume this about points, since we have no axioms for geometry to prove things, one can actually prove that real numbers do have the least upper bound property. I.e. recalling that an "upper bound" for a set S of real numbers is a number A such that A is at least as large as every element of S, with a lot of work, one can prove the next theorem.

Theorem: The real numbers, denoted R, or %, contain the rational numbers as a subset, and also satisfy the "least upper bound property", i.e. every non empty set of reals which has an upper bound, has a smallest real upper bound.

Remember, the "l.u.b. of S" or least upper bound of S, is the smallest number not smaller than any element of S. It may not exist, if S is empty or not bounded, but if it exists, it is unique.

E.g. If a set has a largest element then that largest element is the l.u.b. I.e. if S = the closed interval $[0,1] = \{x : 0 \leq x \leq 1\}$ then lub(S) = 1. But if S = the open interval $(0,1) = \{x : 0 < x < 1\}$ then again lub(S) = 1. This second case is more interesting because it means we can use the concept of l.u.b. to define new numbers we do not have in our set S.

For instance we know there is no rational square root of 2. To prove there is a real square root of 2, consider the set S = {rational x such that $x^2 < 2$}. Then S has a least upper bound say c, and using lemma 2.1, one can show, with a great deal of tedious argument that c^2 must equal 2.

There is a much nicer way using the concept of continuity, as done in calculus courses, which we have sketched in the previous notes.

Now as we stated in those notes, using the concept of l.u.b., one can define addition and multiplication of real numbers and prove they form an ordered field containing the rationals as a subfield. However since the symbols for real numbers involve an infinite number of digits. they are very impractical to deal with explicitly. That is why it is essentially forced on us to start using abstract properties to discuss them. The main fact is the following.

Theorem: The real numbers are the ONLY ordered field which contains the rationals, and which satisfies the l.u.b. property.

We would get very tired of this stuff before finishing the proof of this theorem, as it involves struggling with those infinite decimals in great detail, so we will not prove it. I did once prove all this with a gifted group of high school students age 15-17 at Paideia School in Atlanta, and one of those students is now a math professor at the University of Chicago. If any of you might want to do a project like this, and want my notes, I will be happy to give you a copy of them. I did it as a summer NSF project, and wrote a grant and paid the students, and their teacher, to participate. To relieve the tension at a certain point, two of the students wrote and performed a rap song about real numbers, and the project. Sort of like Sesame Street calculus.

How do the reals contain the rationals?
Given a rational, to change it into an infinite decimal, just divide it out.
E.g. 3/2: Divide 2 into 3, get 1.5000000........

2/3: divide 3 into 2, get .6666666.........

2/7: divide 7 into 2, get .285714285714285714................

Notice these decimals eventually repeat, either 0, or some finite block of numbers, repeats over and over forever. You need to look at some examples until you can explain, i.e. prove, why this is true.

In the other direction, it is easy to change a repeating decimal into a fraction, as follows. Let $x = .2545454......$, then $100x = 25.454545454......$
so $100x - x = 99x = 25.2$, so $x = (25.2)/99 = (252)/990$.

Geometric series:
This can be seen in terms of the famous "geometric series". I.e.
if $0 < r < 1$, the infinite series $1 + r + r^2 + r^3 + ...$ equals $1/(1-r)$.

Now to "see" this in an elementary way that you can show to junior high schoolers, just proceed as above. I.e.
if $x = 1 + r + r^2 + r^3 + ...$, then

$rx = \quad r + r^2 + r^3 + ...$, so that

"subtracting" each term from the one directly above it gives $x - rx = 1$, so $x(1-r) = 1$, so $x = 1/(1-r)$. **"QED".**

[There are quote marks around this QED, because we did not give any reasons for any of these steps, and used subtraction of infinite sums without even discussing what that should mean. Still you can discuss this stuff this way in a junior high class, and in fact a student teacher did present this stuff to my 8th grade class this way in 1955. I thought it was so cool, I have never forgotten it.]

In this context the problem above goes as follows.

$.25454545....... = .2 + .05454545......$

$= .2 + (.054) + (.00054) + (.0000054)..........$

$= .2 + (.054)\ [1 + 1/100 + 1/10{,}000 + ...]$

$= .2 + (0.54)\ [1 + (1/100) + (1/100)^2 + (1/100)^3 + ...]$

$= .2 + (.054)\ [\ 1/(1-1/100)]$

$= .2 + (.054)\ [1/(99/100)] = .2 + (.054)\ [100/99]$

$= .2 + (5.4)/99 = .2 + 54/990 = 2/10 + 54/990 = (198/990) + 54/990$

$= (252/990)$. This is the same answer we got before.

Making precise sense out of the sum of an infinite series

Treating geometric series in this simple way can lead to problems as follows. What if $r = 1$?
Then we get $x = 1 + 1 + 1 +.........$, and then also
$\qquad x = \qquad 1 + 1 + ...$, but then subtracting each term

from the one directly above it gives $0x = 1$!!!????

Equally bad, if $r = 2$, we get
$x = 1 + 2 + 4 + 8 +..........$, so then
$2x = \quad 2 + 4 + 8 + ...$, so if we again subtract each term from the one directly above it we get

$x - 2x = - x = 1$,

so $x = -1 = 1 + 2 + 4 +.....$????

The explanation is that we have not <u>defined</u> addition of infinite collections of numbers, and this is not always possible. It is only possible if the partial sums do not get too large, i.e. are bounded, and then we can do it using least upper bounds for the partial sums, in the same way we mentioned for adding infinite decimals.

Without going into complete detail, but just to get the idea of how much work this is, here is a sketch of the argument. This is studied in our real analysis course, so we will try not to overdo it here. Still we should know what is involved in discussing real numbers carefully.

First of all define the "sum" of the infinite series $1 + r + r^2 + r^3 + ...$ to be the l.u.b. of the sequence of partial sums $(1, 1+r, 1+r+r^2,.......)$. Thus for this to make sense there must <u>be</u> a least upper bound. Thus first there must be an <u>upper bound</u>. Hence in the cases of $1+1+1+1.......$, or

$1 + 2 + 4 + 8 +$........, there is no upper bound to the partial sums which keep getting larger, so there is no l.u.b. and hence no sum.

Then we have the theorem:

Theorem: If $0 < r < 1$, then the series $1 + r + r^2 + r^3 +$........ has a finite "sum" equal to $1/(1-r)$. I.e. the sequence of partial sums
$(1, 1+r, 1+r+r^2,\ldots\ldots.)$ is bounded above, and has l.u.b. equal to $1/(1-r)$.

Proof sketch: By example 2, page 4, we have a formula for each partial sum as follows: $1+r+r^2+\ldots\ldots+r^k = (1-r^{k+1})/(1-r)$. Since this is less than $1/(1-r)$, and the difference is $r^{k+1}/(1-r)$, all we have to show is that
$r^{k+1}\text{-->}0$, i.e. r^{k+1} approaches zero, as k approaches infinity (provided $|r| < 1$).

This is intuitively obvious, e.g. if $r = (1/2)$, we get $r^2 = 1/4$, $r^3 = 1/8$, ... and these fractions do appear to approach zero. To be precise however, we define "$r^{k+1}\text{-->}0$" to mean the g.l.b., i.e. we claim the "greatest lower bound" of the numbers $\{r, r^2 , r^3, r^4, \ldots\ldots, r^{k+1},\ldots\ldots\ldots\}$ is zero. I.e. we claim there is no positive number which is at least as small as all these numbers. I.e. if $0 < r < 1$, then we claim any positive number is larger than some of these numbers.

This is not trivial to prove. The following argument is the first step.

Lemma: The glb of the numbers $\{1, 1/2, 1/3, 1/4 ,\ldots\ldots\ldots, 1/k,\ldots\ldots. \}$ for all natural numbers k, is zero, i.e. there is no positive number which is smaller than all these numbers.

(In calculus courses this is stated as $1/n \text{-->} 0$, as n approaches infinity.)

proof: It is equivalent to show that there is no positive real number larger than all the natural numbers $\{1,2,3,4,\ldots\ldots,k,..\}$. What if there were? Then by the lub property, there would be a smallest one. Let A be the smallest number not smaller than any natural number. Then what about A-1?? It is smaller than A so it is NOT an upper bound for all natural numbers. I.e. there is a natural number N such that $N > A-1$.

What then?? well then $N+1 > A$. Oops, that says A was not an upper bound for all natural numbers after all, contradiction. **QED.**

Then one shows:

step two: If $0 < r < 1$, then for some N, we have $r < N/(N+1) < 1$.
proof idea: Since we can choose $1/N+1$ as small as we like by step one, choose it smaller than the distnace from r to 1. Then we get $N/N+1$ closer to 1 than r is. QED.

step three: Then $N^k/(N+1)^k \text{-->} 0$, as k approaches infinity.
proof idea: This is equivalent to showing that the reciprocals $(N+1)^k/N^k$ approach infinity, i.e. are unbounded, as k approcahes infinity. But $(N+1)^k/N^k = [(N+1)/N]^k = [1 + 1/N]^k$ = (by the

binomial theorem)
$= 1 + k(1/N) + \dots\dots\dots$ Now this is larger than $1 + k/N$, and as k approcahes infinity k/N is unbounded by the same type of argument that showed k is unbounded in step one. QED.

step four: Since $r^k < N^k/(N+1)^k$, thus also $r^k \dashrightarrow 0$. This is the squeeze principle for limits from calculus.

step five: Thus if $0 < r < 1$, then

$1+r+r^2+\dots\dots+r^k = (1-r^{k+1})/(1-r) \dashrightarrow 1/(1-r)$.
This is because the distance between $(1-r^{k+1})/(1-r)$ and $1/(1-r)$, which is $(r^{k+1})/(1-r)$, approaches zero as k approaches infinity, since the top does. **QED.**

The field Q($\boxed{\sqrt{2}}$).

Now that we see how complicated the field of all real numbers is when taken as a whole, what can we do if all we are interested in is sqrt(2) ? I.e. if all we are interested in is sqrt(2), then maybe we can ignore most of the rest of the reals and just consider this one, and a few more that go with it. I.e. instead of considering the infinite decimal representing sqrt(2) just choose one symbol for it like the old fashioned root symbol $\sqrt{2}$, or maybe the exponential notation $2^{1/2}$. Pick one, and then agree that it represents say the positive square root of two. Then of course $-\sqrt{2}$ is the other one. Now to have a field we also need to add and multiply these things so we will also need to consider the numbers of form $(n/m)\sqrt{2}$, and more generally $(n/m) + \sqrt{2}\,(r/s)$, where n,m,r,s, are integers, and of course m,s are non zero. If we agree that a and b represent rational numbers then we can write one of these numbers as $a+\sqrt{2}\,b$. Then we have a small field, inside the reals, which is easier to work with, and where there is a square root of 2.

We have $a+\sqrt{2}\,b = c+\sqrt{2}\,d$ if and only if $a = c$ and $b = d$, because if $a+\sqrt{2}\,b = c+\sqrt{2}\,d$ then $a-c = (d-b)(\sqrt{2})$, so the left side is rational and the right side is irrational, unless both sides are zero.

Moreover we can add and multiply these as follows, i.e.
$(a+\sqrt{2}\,b) + (c+\sqrt{2}\,d) = (a+c) +\sqrt{2}\,(b+d)$, and

$(a+\sqrt{2}\,b)(c+\sqrt{2}\,d) = (ac + 2bd) + (ad+bc)\sqrt{2}$.

Theorem: Consider the set of real numbers of form $a+\sqrt{2}\,b$, where a,b are rational numbers, and where addition and multiplication are defined as above, i.e. in the only possible way obeying distributivity, and $(\sqrt{2})^2 = 2$. Then this set of numbers is the smallest field containing all rational numbers and also containing a square root of 2.

We call this field the field obtained by “adjoining” a square root of 2 to the rationals. It is denoted Q($\sqrt{2}$). This is much a smaller and more manageable field than the full field of reals, and numbers here can have much shorter names. I.e. this field is only “twice” as big as the rationals, whereas the reals are infinitely many times as large as the rationals.

It is kind of fun to figure out how to divide in this field. I.e. the situation is somewhat like modular arithmetic. We are claiming that we can divide just using the numbers of form $a+\sqrt{2}\,b$. Thus for example $\sqrt{2}$ must have an inverse, but the inverse has to have the form $a+\sqrt{2}\,b$ too. So what is it? I.e. what are a,b such that $(a+\sqrt{2}\,b)(\sqrt{2}) = 1$?

Well try solving the equation $(a+\sqrt{2}\,b)(\sqrt{2}) = 1$. You get $a\sqrt{2} + 2b = 1$. But that means that $b = 1/2$ and $a = 0$. So the inverse is $(1/2)(\sqrt{2})$. I.e. multiply this by $(\sqrt{2})$ and you get $(1/2)(2) = 1$.

How about another one? Say what is $(1+\sqrt{2})^{-1}$? We want $(a+\sqrt{2}\,b)(1+\sqrt{2}) = 1$, so again multiply out and try to solve.

We get $a + a(\sqrt{2}) + (b\sqrt{2}) + 2b = (a+b)\sqrt{2} + a+2b = 1$. Thus we should have $a+2b = 1$, and $a+b = 0$. I think I can guess this one. How about $a = -1$, and $b = 1$? Then $a+b = 0$ and $a + 2b = 1$. Does it work?
I.e. does $(-1 + \sqrt{2})(1+\sqrt{2}) = 1$? Well we get $-1 -\sqrt{2} + \sqrt{2} + 2 = 1$. Yes, it works.

Now there is a clever way to get this as follows. Just write out $1/(1+\sqrt{2})$ as if it existed, and try to manipulate it until it has no denominator involving $\sqrt{2}$, i.e. "rationalize it". To do this we all know we multiply top and bottom by $1/(1-\sqrt{2})$, so we get $[1/(1+\sqrt{2})][(1-\sqrt{2})/(1-\sqrt{2})] = (1-\sqrt{2})/(1-2) = \sqrt{2} -1$, as before.

This gives a general formula for an inverse. I.e. $(a+\sqrt{2}\,b)^{-1} = 1/(a+\sqrt{2}\,b)$
$=(a-\sqrt{2}\,b)/(a^2-2b^2) = a/(a^2-2b^2) - \sqrt{2}\,b/(a^2-2b^2)$. Might as well check it to be sure.
So $(a + \sqrt{2}\,b)[(a-\sqrt{2}\,b)/(a^2-2b^2)] = (a^2-2b^2)/(a^2-2b^2) = 1$.
Notice that the denominator is never zero, because the only way we could have $(a^2-2b^2) = 0$, is if $a^2 = 2b^2$, an equation we know very well has no integer solutions except $a = b = 0$. That of course is the only case where a multiplicative inverse does not exist.

"Integers" in Q($\boxed{\sqrt{2}}$).

If there is an analogy between the fields Q and Q($\sqrt{2}$), what ring in Q($\sqrt{2}$) corresponds under this analogy to the ring of integers in Q? presumably it would be the numbers of form $a + b\sqrt{2}$, where a and b are integers. We call these integers in Q($\sqrt{2}$), and denote them as Z[$\sqrt{2}$]. Note then that there are more "units" than before, i.e. integers in Z[$\sqrt{2}$] whose inverse are also elements of Z[$\sqrt{2}$]. For example $3 + 2\sqrt{2}$ is a unit since its inverse is $3-2\sqrt{2}$, i.e. $(3 + 2\sqrt{2})(3-2\sqrt{2}) = 9-8 = 1$. Indeed by our formula for inverses, an "integer" $a+b\sqrt{2}$, is a unit in Z[$\sqrt{2}$] if such that $a^2-2b^2 = 1$ or -1. The inverse of 7 is still 1/7, hence 7 is still not a unit, so it makes sense to try to factor it into "integers". Note that $(3+\sqrt{2})(3-\sqrt{2}) = 9-2 = 7$, so the "integer" 7 is no longer prime in our new ring Z[$\sqrt{2}$]! Also $(5+\sqrt{2})(5-\sqrt{2}) = 25-2 = 23$, so 23 is no longer prime either! Which integers that used to be prime in Z do you think are still prime in Z[$\sqrt{2}$]?? Do you think Z[$\sqrt{2}$] has unique factorization into "primes"? Some more: $(7+\sqrt{2})(7-\sqrt{2}) = 47$. $(9+\sqrt{2})(9-\sqrt{2}) = 79$. $(11+\sqrt{2})(11-\sqrt{2}) =$

Question: I suppose the smallest field containing rationals and a square root of 3 would be all numbers of form a+b$\sqrt{3}$, with a and b rational. If we want the smallest field containing rationals, and square roots of both 2 and 3, would it be all numbers of form a+b$\sqrt{2}$ + c$\sqrt{3}$ +d$\sqrt{6}$?? if so, what would be the inverse of a typical number, say the inverse of $\sqrt{2}$ + $\sqrt{3}$??

4000/6000 Day 13 Adjoining square roots, e.g. $Q(\sqrt{2})$, $R(\sqrt{-1})$.

Last remarks on real numbers, concrete versus abstract.

Jen made a good suggestion yesterday, that it is helpful to connect the way we are doing things in class with the way they are done in the book. This is particularly useful in section 2.2 on real numbers. I have finally understood what they are doing. In one sense they are giving an abstract treatment of real numbers as opposed to our concrete treatment. Their approach is very clean, but unfortunately it does not show clearly what the real numbers look like. We have actually described real numbers as infinite decimals subject to a certain equivalence relation that some of them represent the same real number. We have also shown how the reals contain the rationals, and we have stated but not proved that reals have the key l.u.b. property.

The book on the other hand just states the l.u.b. property and claims in Theorem 2.2 there does exist exactly one field satisfying it and containing the rationals. They do not even tell what the elements are. Although they imply that reals are decimals, they do not mention that different decimals can sometimes represent the same real number. Thus their point of view is that you do not need to know what reals look like, you just need to know they have the lub property, and then how to use that property to get everything else you might want to know. This is a very pretty approach but it can leave us wondering whether the real numbers actually exist. I.e. how the existence part of thm. 2.2 is proved.

Although theorem 2.2 is not proved, the book does prove Propositions 2.3 and 2.4, the "Archimedean" property, and the "density property". They deduce these results just from the l.u.b. property. In fact from our point of view, these properties are almost obvious when you know that real numbers are infinite decimals. E.g. given an infinite decimal, it is very easy to write down an integer which is larger. (For instance give me an integer larger than 567.1212343456565432.......) Thus the integers are not bounded above. (Prop. 2.3. follows by choosing an integer larger than y/x, although for some reason the proof in the book makes it look harder than this.)

Also given two positive infinite decimals, one larger than the other, it is almost as easy, if we make sure neither decimal ends in all 9's, to write down a finite decimal, hence a rational number, which lies strictly between the two given real numbers. (For instance give me a finite decimal lying between 234.567234...... and 234.5781923.....)

So why are they going to so much trouble to prove Prop 2.3 and 2.4?
What they are doing is showing how one would prove the uniqueness part of theorem 2.2 that any ordered field with the lub property must be the reals. I.e. they are assuming they have a field with the l.u.b. property, and then proving their field looks like the field of infinite decimals. I.e. the first step is to show that the all their numbers are caught in between the integers, which follows from the fact that the natural numbers are unbounded. This is essentially Prop. 2.3. Then the second step is to prove that in between each integer, each number can be approximated as closely as desired by rationals. This is the density property proved in Prop. 2.4. This essentially says that your numbers must look like the points on the familiar real line.

Notice every infinite decimal is the l.u.b. of a sequence of finite decimals. Once you know that also all the numbers in your field can be approximated by rationals, you can prove that your field just consists of l.u.b.'s of rationals, i.e. that your numbers are essentially just infinite decimals.

Although the properties in Prop 2.3, 2.4 are easier to see just using infinite decimals than using the lub property, the l.u.b. property is still important. As we mentioned, some basic things such as the fact that you can add and multiply reals, and the existence of a square root of 2, and the existence of solutions to some polynomial equations, are easier to prove using the l.u.b. property than explicitly with decimals. I.e. it is much easier to provide solutions of equations like $X^2 = 2$, or $X^3 + X + 1 = 0$, by introducing continuity, and using lub's to prove the intermediate value theorem, than it is to prove there is an infinite decimal solving the equation.

To be honest, I am not a big fan of using the l.u.b. property to do things that are easier using decimals. I prefer saving the l.u.b. property for things that are hard using decimals. Having said this, we do not want to look too ignorant after taking this course, so I do want you to learn the proof that the lub property implies that the natural numbers are unbounded (Archimedean property). Here it is. I hope it looks easy. If not ask me.

Archimedean property:

version I: The set of natural numbers is unbounded (above).

Proof: (Using only the l.u.b. property.) Proof by contradiction. If there were a real number B such that $B \geq n$ for every natural number n, then there would be a smallest such number A. Thus A is an upper bound for all natural numbers but A-1 is not. I.e. there is a natural number N such that $N > A-1$. Then $N+1 > A$. That says A was not an upper bound for all natural numbers after all, contradiction. **QED.**

version II (Prop. 2.3): Given any two positive real numbers $x,y > 0$, there is a natural number n such that $nx > y$.

Proof: Just take $n > (y/x)$, using version I. **QED.**

You should also learn the proof of the intermediate value theorem using l.u.b.'s, in the previous notes. I do not want you to learn the proof of Prop. 2.4, (density), using lub's, although I will give it next for those of you who are going on to analysis courses and will pursue this kind of thing.

The point is to prove that between any two real numbers there is a rational number, but not using the fact that reals are infinite decimals, just using the l.u.b. property and basic facts about ordering. Remember you do not need to learn this proof for my course. I suggest you read it just to see that you can follow it, but you do not need to memorize it. I will not ask you to prove this. I believe it is unhealthy to memorize proofs like this. I have tried however to make it smoother than the one in the book.

Density property of rationals.

Prop. 2.4. Given two positive numbers x,y, there is a rational number r such that $x < r < y$.

Proof: The idea is to get a denominator n so big that 1/n is shorter than the distance y-x between the numbers. Then you just start laying off copies of 1/n until you land between x and y. Actually until you just go past y. Then take the previous copy, which should be between x and y. I will give a slick description of the proof though, that will not make it super clear why it works.

step 1: Choose $n > 1/(y-x)$, possible since natural numbers are unbounded. Hence $n(y-x) > 1$, so $ny - 1 > nx$.

step 2: By unboundedness of natural numbers, and well ordering, there is a smallest natural number m so that $m \geq ny$.

step 3: Combining steps 1,2, we get $m-1 \geq ny-1 > nx$.

step 4: Since m is the smallest natural number greater $\geq$ ny, then

$m-1 < ny$. By step 3 then, $ny > m-1 > nx$.

step 5: Dividing the last inequality by n gives $y > [(m-1)/n] > x$. Thus the rational number $(m-1)/n$ lies between x and y. **QED.**

Square roots and Complex numbers.

There is a good analogy here between the construction of complex numbers, and the construction of the smallest field $Q(\sqrt{2})$ containing the rationals and a square root of 2. I.e. $Q(\sqrt{2})$ is the smallest field containing the rationals and a square root of 2, and the complex numbers are the smallest field containing the reals and a square root of -1. So if you can construct one you should be able to imitate the method and construct the other. To construct $Q(\sqrt{2})$ we know the real number field contains a square root of 2, so we choose one and called it $\sqrt{2}$. Then we take the smallest subfield of the reals which contains the rationals and $\sqrt{2}$. That turns out to give us things of form $r+\sqrt{2}\,s$, where r,s are rational numbers. In particular the multiplicative inverse of a number like that is again a number like that. To be precise, $(r+\sqrt{2}\,s)^{-1}$ = "$1/(r+\sqrt{2}\,s)$" =
$(r-\sqrt{2}\,s)/[(r+\sqrt{2}\,s)(r-\sqrt{2}\,s)] = (r-\sqrt{2}\,s)/(r^2-2s^2)$
$= r/(r^2-2s^2) - \sqrt{2}\ s/(r^2-2s^2)$.
This is in fact a rational number, namely $r/(r^2-2s^2)$, plus a rational multiple of $\sqrt{2}$, namely $s/(r^2-2s^2)$ times $\sqrt{2}$.

Now that used the fact that we had already constructed the reals, and there was a square root of 2 in the reals. But what if we did not have the reals available? Or what if we think the infinite decimals and least upper bounds are too complicated and we do not believe in them, or for some reason do not want to use them. We could still construct the field $Q(\sqrt{2})$ starting just from Q.

Adjoining a square root of 2 to Q.

Start from the rational field Q, and define a new number system to consist of the set of all pairs of rational numbers (r,s), where we define two pairs (r,s), and (u,v) to be equal if and only if r = u, and s = v. (Think of the pair (r,s) as the real number $r + s\sqrt{2}$, of course.)

To define addition, we set **(r,s) + (u,v) = (r+u, s+v).**
To define multiplication, set **(r,s)·(u,v) = (ru+2sv, rv+su),**
[why did we define it that way?]

Now check that this defines a commutative ring, which is in fact a field. Denote the special number (1,0) by 1, the number (0,0) by 0, and denote (0,1) by $\sqrt{2}$. Then the pair (a,b) can be written a(1,0) + b(0,1) =
$a\cdot 1 + b\cdot\sqrt{2} = a + b\sqrt{2}$. Then note that $(\sqrt{2})^2 = (0,1)(0,1) = (2, 0) = 2$. Moreover
$(a+b\sqrt{2})(c+d\sqrt{2}) = (a,b)(c,d) = (ac+2bd, ad+bc) =$
$(ac+2bd) + (ad+bc)\sqrt{2}$. Now this is what we would have gotten by multiplying out $(a+b\sqrt{2})(c+d\sqrt{2})$ in the usual way, so we never have to remember the stupid multiplication rule above again.

I.e. we just went through this song and dance with pairs of rationals, to prove to the skeptical among you that it really does make sense to just say "take a new symbol $\sqrt{2}$, and set $(\sqrt{2})^2 = 2$, and then make up a new number system with numbers which are symbols of form

$a+b\sqrt{2}$, where a,b are rational numbers, and where $a+b\sqrt{2} = c+d\sqrt{2}$ if and only if a=c and b =d."

Call the new set of numbers $Q(\sqrt{2})$. The point is we have constructed these numbers without knowing in advance that the reals exist, i.e. we did not use them. So even if there was no square root of 2 in our original number system, and we had never seen a number system that had a square root of 2, still we have just built a new system in which there is one. And this new system is much simpler than the real number system. (Mainly because it is much smaller.) Q is contained in $Q(\sqrt{2})$ which is contained in R. [If you remember your linear algebra, the difference is that $Q(\sqrt{2})$ is a 2 dimensional vector space over Q, while R is infinite dimensional over Q.]

Theorem: The set of numbers $Q(\sqrt{2})$ is a field. (It is the smallest field containing Q in which X^2-2 has a root.)
Proof: We already defined addition and multiplication, and you can see by looking at the formulas for a few minutes, that both addition and multiplication are commutative. (The safest thing to do is to just take some scratch paper and write out $(r,s)\cdot(u,v)$ and then $(u,v)\cdot(r,s)$, and compare them.)

Addition is so simple that it is pretty clear that (0,0) is the additive identity, and that (-a,-b) is the additive inverse of (a,b). Multiplicative inverses are not so obvious. They depend on an operation I always make fun of in my calculus class, "rationalizing the denominator". I.e. I think there is no reason to rationalize a denominator, i.e. remove a radical from the bottom, when working with real numbers, but here when radicals are not allowed in our number system, it is crucial.

Multiplicative inverses.
To discover the inverse of $(a,b) = a+b\sqrt{2}$, just write it as if there was one, i.e. write it as $1/(a+b\sqrt{2})$ and then rewrite it as a legitimate number in our system, i.e. "rationalize it!" This means multiply top and bottom by the "conjugate", i.e. by $(a-b\sqrt{2})$. This gives $a-b\sqrt{2}$ in the top, and a^2-2b^2 in the bottom. Thus the inverse of $a+b\sqrt{2}$ is

$$\mathbf{(a+b\boxed{\sqrt{2}})^{-1} = (a/[a^2-2b^2]) - (b/[a^2-2b^2])\boxed{\sqrt{2}}.}$$

Of course I have no idea if I did this right, so let's check it in a few cases. Thus $(1+3\sqrt{2})^{-1} = (1-3\sqrt{2})/(-17)$. Try it. $(1+3\sqrt{2})(1-3\sqrt{2})/(-17) =$
$-17/-17 = 1$.

Here is an interesting one: $(1+\sqrt{2})^{-1} = (1-\sqrt{2})/(-1) = \sqrt{2}-1$. Then $(\sqrt{2}-1)(\sqrt{2}+1) = 2-1 = 1$. Notice the inverse involves no fractions.

Integers in $Q(\boxed{\sqrt{2}})$.
The previous example suggests we distinguish two kinds of numbers in our new system, those involving fractions and those not. I.e. there are "integers" in our new system, consisting of numbers of form $a+b\sqrt{2}$ where both a and b are ordinary integers. For example $\sqrt{2}$ is a new

integer. The inverse of this integer is $\sqrt{2}/2$, which is not an integer. We denote these new integers by $\mathbf{Z}[\boxed{\sqrt{2}}]$.

Then the number $(1+\sqrt{2})$ is interesting because it is an "integer" whose multiplicative inverse $\sqrt{2}-1$ is also an "integer". [The idea here is the rational root theorem, i.e. any solution in Q of an equation with integer coefficients whose leading coefficient is 1, must be an integer in Q. We say the same thing here. I.e. any solution in $Q(\sqrt{2})$ of an equation with ordinary integer coefficients, whose leading coefficient is 1, is an "integer" in $Q(\sqrt{2})$. Thus $\sqrt{2}$ which is a solution of $X^2-2 = 0$, is an integer in $Q(\sqrt{2})$.]

Factoring integers in $\mathbf{Z}[\boxed{\sqrt{2}}]$

Notice that we have more numbers available, so there is the possibility of factoring numbers in ways that could not be done before. For instance $(3+\sqrt{2})(3-\sqrt{2}) = 9-2 = 7$, so 7 is no longer a prime in our new system of integers. It is kind of hard to see if unique factorization holds here since for example we also have $(5+3\sqrt{2})(5-3\sqrt{2}) = 25 - 18 = 7$, is another factorization of 7. But recall that even in the ordinary integers we have more than one prime factorization of integers if we use factors of 1. I.e. -1 is a factor of 1, so we can factor 7 as $(1)(7) = (-1)(7)$. In our new situation, we have the factors $(1+\sqrt{2})(\sqrt{2}-1) = 1$, so from $(3+\sqrt{2})(3-\sqrt{2}) = 7$, we get the equivalent factorization $(1+\sqrt{2})(\sqrt{2}-1)(3+\sqrt{2})(3-\sqrt{2}) = (1+\sqrt{2})(3+\sqrt{2})(\sqrt{2}-1)(3-\sqrt{2}) = (5+4\sqrt{2})(-5+4\sqrt{2}) = -25 + 32 = 7$.

This is not an essentially different factorization, and in fact this new ring has "unique factorization" into primes. I.e. a prime in our new ring is an element $x+y\sqrt{2}$ in $Z[\sqrt{2}]$ such that whenever $x+y\sqrt{2} = (a+\sqrt{2})(c+d\sqrt{2})$, then at least one of the elements $(a+\sqrt{2})$, or $(c+d\sqrt{2})$ must be a unit in $Z[\sqrt{2}]$. Now one can recognize a unit in the ring $Z[\sqrt{2}]$ because it has form $a+b\sqrt{2}$, where a,b are integers, i.e. it is an integer, and also the denominator of its reciprocal is 1 or -1. I.e. $a+b\sqrt{2}$ is a unit if and only if $a^2-2b^2 = 1$ or -1.

Thus a non trivial factorization of an element $x+y\sqrt{2}$ of $Z[\sqrt{2}]$, is a factorization as a product $x+y\sqrt{2} = (a+\sqrt{2})(c+d\sqrt{2})$, where a^2-2b^2 does not equal either 1 or -1, and also $c^2 - 2d^2$ also does not equal either 1 or -1.

Recall that a unit is an integer with an inverse which is also an integer. One fact about this new ring of integers which makes factorization difficult in it is there are infinitely many units in this ring. In the old ring Z, there were only 2 units, 1 and -1, so we could easily tell whether two factorizations were equivalent by looking at them. it is much harder now.

A much easier case is that of a field obtained by adjoining a square root of a negative number like -1. Let's form a field by adjoining a square root of -1 to Q. We get the field Q(i) where i is the usual symbol to denote a square root of -1. To define it carefully, (or blindly?) we let the elements of Q(i) consist of pairs of rational numbers (a,b) where we define addition as usual to be $(a,b)+(c,d) = (a+c,b+d)$, and multiplication to be $(a,b)(c,d) = (ac-bd, ad+bc)$. (Of course we are thinking of numbers of form a+bi when we do this, but we keep that a secret.)

Then denote the special number (1,0) by 1 and the number (0,1) by i, and then every number can be written as $(a,b) = a(1,0) + b(0,1) =$
$a + bi$. Then the multiplication takes the form $(a+bi)(c+di) =$

$(ac-bd) + (ad+bc)i$. We call these complex numbers with rational coefficients. They form a field where the multiplicative inverse of $a+bi$ is $(a-bi)/(a^2+b^2) = (a/a^2+b^2) - (b/a^2+b^2)i$. The integers in this field have form $a+bi$ where a,b are ordinary integers. We call the number system of form $a+bi$ where a,b are real numbers, the complex number field.

Theorem: In the complex number field, every polynomial has a root.
proof: I know how to prove this but it takes a long time. The proof is a generalization of the proof of the intermediate value theorem, but you have to figure out how to generalize the concept of a curve "crossing the x axis", to create a precise concept of a closed curve winding around the origin in the plane. For those of you who have taken math 3500, the key tool is the Green's theorem of several variable calculus. I figured out how to do this during my first big teaching experience in college. I am pretty proud of it, although many other people have done it too. But let's skip the details on this one for now. Again, for those who want to know, I have notes on this. **QED.**

Geometry of complex multiplication, relation with trigonometry.
Factorization of integers using complex integers.
Which primes in Z remain prime in Z[i]?
Connections with sums of two squares.