MATH 154 NOTES

ARUN DEBRAY
APRIL 6, 2015

These notes were taken in Stanford's Math 154 class in Spring 2015, taught by Brian Conrad. I TEXed these notes up using `vim`, and as such there may be typos; please send questions, comments, complaints, and corrections to adebray@stanford.edu.

Contents

1. The Fermat Equation: 3/30/15

> *"There may be a Japanese film crew here on Wednesday... they're filming a documentary on the demise of the Hagaromo chalk company, and my supplier in Oakland outed me as her biggest customer."*

Homeworks will be assigned Wednesdays and due Wednesdays; see the website for office hours, etc.

In this class, we will assume familiarity with group theory and Galois theory (which we'll only use about halfway through the course), so Math 120 and 121. Specifically, you should be comfortable with the Galois theory of finite fields.

Let's begin with a warm-up example.

Theorem 1.1 (Fermat)**.** *$y^2 = x^3 - 2$ has only $(3, \pm 5)$ as solutions over $\mathbf{Z}$.*

What makes this interesting is that this does have infinitely many rational points, though this involves certain ideas from the theory of elliptic curves, which we won't discuss in this course. Discussing this will lead to several of the important ideas in this course.

Proof. Note that first, x and y must both be odd: if one were even, then the other would be, but then $-2 = y^2 - x^3 \in 4\mathbf{Z}$, but -2 isn't a multiple of 4, so that doesn't work.

Fermat's key idea was to bring the 2 to the other side, and, even though we care about the integers, factor it in a larger number system. Specifically, $x^3 = y^2 + 2 = (y + \sqrt{-2})(y - \sqrt{-2})$. Thus, we're working in $\mathbf{Z}[\sqrt{-2}] = \left\{a + b\sqrt{-2} \mid a, b \in \mathbf{Z}\right\}$.

We would like for $\mathbf{Z}[\sqrt{-2}]$ to have similar arithmetic properties to $\mathbf{Z}$ (in particular, that it's a UFD), then perhaps $y \pm \sqrt{-2}$ have to be cubes in $\mathbf{Z}[\sqrt{-2}]$. In particular, we need to know if they're "coprime," whatever that means in this ring.

Recall that a *unit* u in a commutative ring R is an element with a multiplicative inverse. The set of units of R is denoted $R^\times$. For example, the units of $\mathbf{Z}$ are ± 1, but the units of $\mathbf{C}[t]$ are the nonzero constants (by thinking about the degrees of polynomials).

Lemma 1.2. *$\mathbf{Z}[\sqrt{-2}]^\times = \{\pm 1\}$, and in fact for any nonsquare $d > 0$, $\mathbf{Z}[\sqrt{-d}]^\times = \{\pm 1\}$.*

Proof. There's a map called *conjugation* $\alpha \mapsto \overline{\alpha}$: $\mathbf{Z}[\sqrt{-d}] \to \mathbf{Z}[\sqrt{-d}]$. Specifically, send $a + b\sqrt{-d} \mapsto a - b\sqrt{-d}$. This should feel a lot like complex conjugation: $\overline{\alpha + \beta} = \overline{\alpha} + \overline{\beta}$, $\overline{\alpha\beta} = \overline{\alpha} \cdot \overline{\beta}$, and $\alpha\overline{\alpha} \in \mathbf{Z}$, and is denoted the *norm* of α, $a^2 + db^2$.

Thus, there's a norm function $N : \mathbf{Z}[\sqrt{-d}] \to \mathbf{Z}$ sending $1 \mapsto 1$. Moreover, $N(\alpha\beta) = N(\alpha)N(\beta)$. This will be useful for finding units, since they're not always obvious (e.g. $1/(1 + \sqrt{2}) = -1 + \sqrt{2}$). Now, if u is a unit,

there's a v such that $uv = 1$, i.e. $N(u)N(v) = N(uv) = N(1) = 1$, so $N(u) = \pm 1$. Conversely, if $N(\alpha) = 1$, then $\alpha\overline{\alpha} = 1$, so α is a unit (its conjugate is its inverse).

But since $N(u) = a^2 + db^2 > 0$, then $N(u) = 1 = a^2 + db^2$, which is pretty nice! In particular, this forces $b = 0$ and $a = \pm 1$. ⊠

Conveniently, ± 1 are both cubes, so the units don't matter much in this proof. In more general situations, where there are units which aren't cubes, this can be a major headache.

We want to claim $y \pm \sqrt{-2}$ should both be perfect cubes, ignoring units. But we need them to be relatively prime: can $y + \sqrt{-2}$ and $y - \sqrt{-2}$ have a *non-unit* common factor $\delta \in \mathbf{Z}[\sqrt{-2}]$?

Well, we know $y + \sqrt{-2}$ and $y - \sqrt{-2}$ are both δ times something, so subtracting them, $2\sqrt{-2} = -(\sqrt{-2})^3$ is also δ times something.

Claim. If δ is not a unit, then it must be divisible by $\sqrt{-2}$.

Fact. $\mathbf{Z}[\sqrt{-2}]$ is in fact a UFD, which we'll prove on Homework 1.

Thus, for the claim, it's enough to show that $\sqrt{-2}$ is irreducible in $\mathbf{Z}[\sqrt{-2}]$, or what you might call prime. Suppose not, so that $\sqrt{-2} = \alpha\beta$, where $N(\alpha), N(\beta) \neq \pm 1$. Then, $2 = N(\alpha)N(\beta)$ in $\mathbf{Z}$ (since $N(\sqrt{-2}) = 2$), but this is impossible, so $\sqrt{-2}$ is irreducible.

Next, can $y + \sqrt{-2}$ be $\sqrt{-2}$ times something? That is, it is of the form $2b + a\sqrt{-2}$. We saw that y must be odd, so this cannot happen; thus, $y + \sqrt{-2}$ and $y - \sqrt{-2}$ have no common factor in $\mathbf{Z}[\sqrt{-2}]$.

Definition. Say that nonzero $\alpha, \beta \in \mathbf{Z}[\sqrt{-2}]$ are *associate* if $\alpha = \beta u$ for a unit u.

Since we're in a UFD, every nonzero α factors as

$$\alpha = u \cdot \prod_i \pi_i^{e_i},$$

where the π_i are pairwise non-associate irreducibles. This is nice in $\mathbf{Z}[\sqrt{-2}]$, since we have no nontrivial units, but in larger number fields, this will be trickier, and arguments don't generalize.

Returning to $x^3 = (y+\sqrt{-2})(y-\sqrt{-2})$ in the UFD $\mathbf{Z}[\sqrt{-2}]$, then since the left side is a cube, so factorizing both sides, unique up to units and rearrnangements, each irreducible factor must be a cube, since the units ± 1 are both cubes (this is very nice, and does *not* generalize!) — thus, $y \pm \sqrt{-2}$ must be cubes in $\mathbf{Z}[\sqrt{-2}]$. Great!

Now, what does that tell us? If $y + \sqrt{-2} = (a + b\sqrt{-2})^3$ for $a, b \in \mathbf{Z}$, then after grinding out the algebra, it becomes

$$y + \sqrt{-2} = a(a^2 - 6b^2) + b(3a^2 - 2b^2)\sqrt{-2}.$$

This severely limits the possibilities for b: it must be ± 1, and thus $3a^2 - 2 = \pm 1$. Then, looking at this mod 3, it must be $+1$, so $b = 1$ and therefore $3a^2 - 2 = 1$, or $a = \pm 1$. Plugging this into $y = a(a^2 - 6b^2) = \pm(1-6) = \mp 5$. Then, $x^3 = y^2 + 2 = 27$, so $x = 3$. ⊠

This seems like an involved argument: well, we have no theory yet, so of course. But the key idea is to take a statement in $\mathbf{Z}$ and talk about it in a larger number system.

The moral of the story is:

(1) To solve a problem in $\mathbf{Z}$, it can be useful to exploit "arithmetic" in larger number systems, e.g. the UFD property, using norms, etc.
(2) Beware of units:[1] we got lucky in this problem.

There's a historically important example where the units don't work, though they can be fixed.

Example 1.3. Let's look at $\mathbf{Z}[\sqrt{-3}]$; this comes up in some problems, though not the one we were looking at earlier. This ring has $\{\pm 1\}$ as its units again, but it isn't a UFD. For example,

$$4 = 2 \cdot 2 = (1 + \sqrt{-3})(1 - \sqrt{-3}).$$

Of course, none of these are units, and one can show 2 is irreducible. Note that this is not the same as prime when we're not in UFDs. Well, suppose $2 = \alpha\beta$ with $\alpha, \beta \in \mathbf{Z}[\sqrt{-3}]$; then, $N(\alpha)N(\beta) = 4$ with

[1] Well, of course; this is only a 3-unit class!

$N(\alpha), N(\beta) \neq 1$, then $N(\alpha), N(\beta) = 2$; however, there's no way to write $2 = a^2 + 3b^2$, so oops. Thus, 2 is not irreducible, and $1 \pm \sqrt{-3}$ are not unit multiples of 2, so we really have distinct factorizations.[2]

In some sense, $\mathbf{Z}[\sqrt{-3}]$ is almost a UFD; it sits inside $\mathbf{Z}[(-1+\sqrt{-3})/2]$ (i.e. $\mathbf{Z}[\zeta_3]$), which we'll later see is a UFD. However, it has other units, e.g. ζ_3 (since its inverse is its square). In some sense, if $\mathbf{Q}(\sqrt{-3})$ corresponds to $\mathbf{Q}$, the correct analogue of $\mathbf{Z}$ isn't $\mathbf{Z}[\sqrt{-3}]$, but rather $\mathbf{Z}[(-1+\sqrt{-3})/2]$.

This leads to a question: in a finite extension $K/\mathbf{Q}$, what is the "correct" analogue of $\mathbf{Z} \subset \mathbf{Q}$? In what sense are we looking for an analogue? If $d \in \mathbf{Z}$ is square-free, then $\mathbf{Z}[\sqrt{-d}]$ might be the right thing, but not always.

Remark. When Euler proved Fermat's last theorem for $n = 3$, he factored it over $\mathbf{Z}[\sqrt{-3}]$; thus, he assumed that it was a UFD. This, of course, is not true, but was a common assumption back then. However, his argument works nearly unchanged in $\mathbf{Z}[\zeta_3]$, which has units $\{\pm 1, \pm\zeta_3, \pm\zeta_3^2\}$ (which we will prove on one of the homeworks). However, ζ_3 isn't a cube, which is kind of annoying.

It seems less canonical to use ζ_3, because it has a denominator, but, well, $3 = 6/2$, so honestly it's not that important.

In any case, the UFD property is hard to salvage, so one has to use more sophisticated methods.

Just as a linear algebra problem over $\mathbf{R}$ can be made more complicated, but more possible, over $\mathbf{C}$ (e.g. we now have eigenvalues!), Fermat's great idea was to work in a larger number field.

Let's talk about Fermat's last theorem in general, i.e. that $x^n + y^n = z^n$, when $n \geq 3$ and $x, y, z \neq 0$, has no solutions. If $m \mid n$ and $m \geq 3$, then it's enough to treat m, since any n^{th} power is an m^{th} power. Then, every $n \geq 3$ is either divisible by 4 or an odd prime; the case $n = 4$ was done by Fermat himself, and is in §1.2 of the textbook. There are multiple proofs of it, including one with elliptic curves.

Thus, let's assume $n = p$ is an odd prime:

$$x^p = z^p - y^p = \prod_{j=0}^{p-1} (z - \zeta_p^j y),$$

where ζ_p is a *primitive* p^{th} *root of unity*, i.e. $\zeta_p = e^{2\pi i/p} \in \mathbf{C}$. Thus, this is an important motivation for understanding the arithmetic of $\mathbf{Z}[\zeta_p]$.

Now, there's good news and bad news. Which do you want to hear first? The good news? Ok, bad news first: once $p \geq 5$, $\mathbf{Z}[\zeta_p]^\times$ is infinite, though it is a finitely generated abelian group; thus, it has a nontrivial free part. That means there are units of infinite (multiplicative) order that are not themselves e^{th} powers for any $e > 1$. Thus, the unit problems get really painful. Moreover, when $p > 19$, $\mathbf{Z}[\zeta_p]$ is *never* a UFD.

The good news is, Kummer basically invented number theory to deal with this.

Of course, in the end, algebraic number theory didn't completely solve this problem; it succeeded for many exponents, and was useful for a wide variety of other problems. Nonetheless, there are ways of dealing with units, and ways of getting around the UFD property.

2. Euclidean Domains: 4/1/15

Today there is actually a camera crew here. No joke.

Recall that last time, we looked at the Fermat equation $y^2 = x^3 - 2$, and solved it with the arithmetic of $\mathbf{Z}[\sqrt{-2}]$; it is a UFD and has units $\{\pm 1\}$.

This is reasonable, but one concern is that a small change makes a huge difference: it is hard to generalize this method. For example, if one simply changes it to $y^2 = x^3 + 2$, then we want to solve it in $\mathbf{Z}[\sqrt{2}]$. This is still a UFD, but its unit group is infinite: ± 1 are units as before, but so is $1 + \sqrt{2}$, as $(1+\sqrt{2})(1-\sqrt{2}) = -1$. Instead, *Pell's equation* will later tell us that these generate all of the units: $\mathbf{Z}[\sqrt{2}]^\times = \{\pm 1\} \times (1+\sqrt{2})^{\mathbf{Z}}$.

One reasonably obvious solution is $(1, \pm 1)$. It turns out this is the only solution. The same technique as before shows that in $\mathbf{Z}[\sqrt{2}]$, a solution must satisfy $y + \sqrt{2}$ is a cube or of the form $(1+\sqrt{2})^{\pm 1}(y')^3$ for some y'.

The next question is, we've stated that these are UFDs, but how should we prove it? One way is the pretty lame notion[3] of a Euclidean ring (or Euclidean domain), which axiolatizes long division.

[2] The same argument works to show $1 \pm \sqrt{-3}$ is irreducible, or more generally, anything with norm 4 is irreducible in $\mathbf{Z}[\sqrt{-3}]$, since all we used about 2 was that $N(2) = 4$.

[3] These are the professor's words, not mine.

Definition. A domain R is *Euclidean* if there is a function $\nu : R \to \mathbf{Z}_{\geq 0}$ such that $\nu(x) = 0$ iff $x = 0$ and for all $a, b \in R$ with $b \neq 0$, there exist $q, r \in R$ such that $a = bq + r$ such that $\nu(r) < \nu(b)$.

It's more suggestive to think of $a/b = q + r/b$ (quotient plus remainder), though this is an abuse of notation. Also, keep in mind that ν need not be multiplicative, though in many cases it is. ν should be thought of as the "size" of an element.

Example 2.1.

(1) The canonical example is $R = \mathbf{Z}$, with $\nu(r) = |r|$.
(2) Another good example is $R = \mathbf{C}[t]$, with $\nu(f) = \deg(f)$. Notice that this example isn't multiplicative.

The key point, and the reason we care about $\mathbf{Z}_{\geq 0}$, is that as ν keeps dropping as division goes on, it terminates at some point. Euclidean domains are UFDs and PIDs, which is nice, but it only helps in a few cases; in most cases, even in algebraic number theory, it's not all that helpful, and the Euclidean property doesn't lead to any interesting theorems outside of a few small cases. (For another example, notice that there are interesting UFDs that aren't PIDs, e.g. $\mathbf{C}[x, y]$, but there aren't any cases where something being a PID but not Euclidean is interesting or helpful.)

Nonetheless, the Euclidean notion is useful for $\mathbf{Z}[\sqrt{d}]$ for small d.

Theorem 2.2. $\mathbf{Z}[i]$*, the* Gaussian integers*, is a Euclidean domain, and therefore a UFD and a PID.*

Proof. Let $\nu(\alpha) = N(\alpha) = \alpha\overline{\alpha}$, as in the last lecture. Then, we want to explicitly figure out the division algorithm.

Translate this to $\mathbf{C}$, where given $a, b \in \mathbf{Z}[i]$, we want to write $a/b = q + r/b$. Then, the norm $\alpha\overline{\alpha}$ is perfectly reasonable on $\mathbf{C}$ too, so we can require that $\nu(r/b) < 1$, since in this case ν is multiplicative.

Thus, the idea is to take $a/b = t_1 + t_2 i$, where $t_1, t_2 \in \mathbf{Q}$. Then, we want to find the nearest point in $\mathbf{Z}[i]$: write $t_j = q_j + \varepsilon_j$, where $q_j \in \mathbf{Z}$ and $|\varepsilon_j| \leq 1/2$. If $q = q_1 + q_2 i$ and $\varepsilon = \varepsilon_1 + \varepsilon_2 i$, then $a/b = q + \varepsilon$.

Now, b returns to the stage (multiplying $a/b = q + \varepsilon$ by b): since a and b are Gaussian integers, then $r = \varepsilon b$ must be a Gaussian integer as well. But we know $\nu(r) = \nu(\varepsilon)\nu(b)$, but $\nu(b) > 0$ and $\nu(\varepsilon) = \varepsilon_1^2 + \varepsilon_2^2 \leq 1/4 + 1/4 = 1/2$. Thus, $\nu(r) < \nu(b)$. ⊠

So, can we generalize this? The key idea is the geometric motivation: we have a picture of $\mathbf{Z}[i]$ as a lattice within $\mathbf{C}$, and this can be extended to, say, $\mathbf{Z}[\sqrt{-2}]$ relatively nicely. However, what can we do for $\mathbf{Z}[\sqrt{2}]$? This sits in $\mathbf{R}$, so there's no lattice, and if one takes the same steps in the proof, it doesn't always work. Furthermore, in $\mathbf{Z}[\sqrt{-d}]$ for $d \in \mathbf{Z}$, if d is large enough, this method doesn't work, because the bound on $\nu(r)$ quickly becomes greater than $\nu(b)$. That doesn't mean that these aren't PIDs; sometimes they are, and the point is that the method doesn't work.

The units in $\mathbf{Z}[i]$ are ± 1 and $\pm i$. We can see this because $N(\alpha) = a^2 + b^2$ for $\alpha = a + bi$. Thus, the only integer solutions to $a^2 + b^2 = 1$ are $a = 0, b = \pm 1$ or $b = 0, a = \pm 1$.

Needless to say, once we move past quadratics, there will be a norm, but suddenly instead of quadratic equations we'll have higher-degree equations for the units, and this makes finding the units or even proving something is a UFD much harder. This is even true for more complicated quadratic fields: there's an old conjecture of Gauss that there are an infinite number of real quadratic fields that are UFDs. In these cases, UFD and PID tend to be equivalent in algebraic number theory.

So now, norms have shown up in these two cases that are really hard to generalize, but we can talk about factoring. How do we factor in $\mathbf{Z}[i]$ (up to units)? Of course, there will be no algorithm; even over $\mathbf{Z}$, we think this is a difficult algorithm. On $\mathbf{N}$, factoring can ignore units, since it's just $\{1\}$, but here we have to be ore aware of them.

Example 2.3. Let's factor $7 + 4i$. It's clearly not a unit. Suppose for a moment this had a nontrivial factorization: $7 + 4i = \alpha\beta$, with α and β not units; then, $N(7 + 4i) = N(\alpha)N(\beta)$, and $N(\alpha), N(\beta) \neq 1$. Well, $N(7 + 4i) = 49 + 16 = 65 = 5 \cdot 13$.[4] Thus, without loss of generality, $N(\alpha) = 5$ and $N(\beta) = 13$.

Thus, $N(\alpha) = 5 = \alpha_1^2 + \alpha_2^2$, so one is 1 and the other is 4. This gives several possibilities, but many of them are associate, e.g. $-\alpha$ or $\pm i\alpha$. However, α is generally not associate to $\overline{\alpha}$, which makes sense because the original $7 + 4i$ isn't associate to $7 - 4i$. In other words, up to units, α has two possibilities: $1 + 2i$ and

[4] In a real quadratic field, e.g. $\mathbf{Z}[\sqrt{2}]$, the norm could be negative; this isn't terrible, but it means more cases, e.g. (5)(13) but also $(-5)(-13)$.

$1-2i$. (You can check this by listing out all eight combinations, and then seeing which ones come from others by multiplying by -1 or i.)

So, are these irreducible? We know in $\mathbf{Z}[i]$, $5 = (1+2i)(1-2i)$, and then we can't go farther. Here's a nice general criterion.

Lemma 2.4. *If $\alpha \in \mathbf{Z}[i]$ has $N(\alpha) = p \in \mathbf{Z}$ is prime, then α is irreducible.*

Proof. If $\alpha = \beta\gamma$, then $p = N(\beta)N(\gamma)$, so one of $N(\beta)$ or $N(\gamma)$ must be 1, and therefore one of them must be a unit, and therefore α is irreducible. ⊠

Be careful: the converse is false; there are more primes whose norm is reducible.

The second case is β, whose norm is $13 = 4+9$. Thus, β could be $2+3i$ or $2-3i$ (and the other possibilities come from units). In particular, if $7+4i = \alpha\beta$, then α is an irreducible factor of 5, and β is an irreducible factor of 13.

Now we use the established technique of "guess and check." We can multiply stuff and see if the result is $7+4i$, up to a unit.[5]

Ok, so let's try $(1+2i)(2+3i)$, since we like + signs. The result is $-4+7i = -i(7+4i)$. Thus, we have our factorization: $7+4i = i(1+2i)(2+3i)$. Cool.

But suppose you made a different choice, $(1+2i)(2-3i) = 8+i$. This doesn't work at all. Even as we systematically try to eliminate cases, we still have to check.

The key problem is: *how do the prime factors of the norm break down in $\mathbf{Z}[i]$?* We've seen two examples, $5 = (1+2i)(1-2i)$ and $13 = (2+3i)(2-3i)$. It's interesting, but perhaps unsurprising, that primes in $\mathbf{Z}$ aren't always irreducible in a larger ring.

Let's go back to the beginning: how does 2 factor? We can write $2 = (1+i)(1-i) = i(1+i)^2$. This is different from the factorizations of 5 and 13: 2 is a square up to a unit!

3, however, is prime.

Claim. Any prime $p \equiv 3 \ (4)$ remains prime in $\mathbf{Z}[i]$.

Proof. As we go on, we'll see a systematic way to understand this problem, but for now, suppose $p = \alpha\beta$. Then, take the norm: $p^2 = N(\alpha)N(\beta)$, so assuming α and β aren't units, so $N(\alpha) = N(\beta = p$. Thus, $p = \alpha_1^2 + \alpha_2^2 = (\alpha_1 + \alpha_2 i)(\alpha - 1 - \alpha_2 i)$, so a nontrivial factorization exists iff p is a sum of two squares.

If $p = 3 \bmod 4$, then it cannot be a sum of two squares, because the squares mod 4 are 0 and 1, so a sum of two squares is 0, 1, or 2 mod 4. Whoops. ⊠

What this doesn't address is primes $p \equiv 1 \ (4)$.

Theorem 2.5 (Fermat)**.** *If $p \equiv 1 \ (4)$, then p is a sum of two squares.*

This will be on the homework. But the crucial thing about it is that it uses the arithmetic of $\mathbf{Z}[i]$ (and a very little bit of quadratic reciprocity); next time, we'll take it up a little more systematically.

3. Factoring in $\mathbf{Z}[i]$ and Algebraic Integers: 4/3/15

"What will happen when mathematicians run out of chalk? It'll be just like Mad Max*."*

Last time, we saw that $\mathbf{Z}[i]$ is a UFD and its units are $\mathbf{Z}[i]^\times = \{\pm 1, \pm i\}$. Unlike factoring in the positive integers, where there are no nontrivial units, the units make life a little tricky. For example, if $a \mid b$ and $b \mid a$, then all we know is that $a = bu$ for a unit u (which is generally true in any UFD).

We also talked about the fact that $2 = -i(1+i)^2$, and $1+i$ is irreducible, and associate to $1-i$, $-1+i$, and $-1-i$. If $p \equiv 3 \ (4)$ is prime in $\mathbf{Z}$, then p is still prime in $\mathbf{Z}[i]$.

Then, for $p \equiv 1 \ (4)$, we had Theorem 2.5. Its proof will be in the homework, and really uses the fact that $\mathbf{Z}[i]$ is a UFD; we don't produce the two squares explicitly, just contemplate what it can be, relating to how -1 is a square mod p; then, taking the norm of both sides, $p^2 = N(x+iy)N(x-iy)$, if $p = x^2+y^2$, and therefore $N(x+iy) = p$, so $x+iy$ and $x-iy$ must both be irreducible in $\mathbf{Z}[i]$, but they're *not* unit multiples of each other; they have no factor in common, as one of x or y is greater than 1 (since $p \neq 2$), though their GCD is 1, so they must differ, and therefore cannot be exchanged by -1 or $\pm i$.

[5]Once again, this is hard to generalize; we don't know enough of the structure of Dedekind domains yet to attack more general cases.

Thus, we've produced several different classes of irreducibles: 2 is a special case, and then we have $p \equiv 3 \bmod 4$, and, when $p \equiv 1 \bmod 4$, $p = x^2 + y^2$, and $x - iy, x + iy$ are irreducible. Let's collect this into one result.

Proposition 3.1. *Let $\pi \in \mathbf{Z}[i]$ be a nonzero nonunit. Then, π is irreducible iff one of the following holds:*

- *$N(\pi) = p^2$, where $p \equiv 3 \ (4)$ is prime, and thus $\pi = \pm p$ or $\pi = \pm ip$.*
- *$N(\pi) = p$, where $p \equiv 1 \ (4)$ is prime; then, $\pi = \pm\alpha$, $\pm\overline{\alpha}$, $\pm i\alpha$, or $\pm i\overline{\alpha}$, where $p = x^2 + y^2$ and $\alpha = x + iy$.*
- *$N(\pi) = 2$, so $\pi = u(1 + i)$, where u is a unit.*

Note that as a corollary, there is a unique factorization of π, so when $N(\pi) = p$, the decomposition of $p = x^2 + y^2$ is also essentially unique (well, up to switching them, and multiplying either by -1). This is really slick, and there's no obvious proof by elementary methods.

Proof of Proposition 3.1. The reverse direction is basically the previous discussion. We know $N(\pi) = \pi\overline{\pi}$; if $N(\pi) = \alpha\beta$ where α, β are non-units, then writing α and β as a product of irreducibles. Thus, if $N(\pi) = p^2$, where $p \equiv 3 \ (4)$, then there's no room: we have to get $p \cdot p$, with p prime. Similarly, if $N(\pi) = p$ for $p \equiv 1 \ (4)$, then we know how p factors, and the only way it can go is $(x + iy)(x - iy)$; then, 2 is similar.

Thus, the interesting case is the forward direction. If π is irreducible, then $\pi\overline{\pi} \in \mathbf{Z}_{>1}$. Choose a prime factor p of it; we'll consider separately what different things p could be.

What if $p = 3 \ (4)$? Then, p is irreducible in $\mathbf{Z}[i]$! Nice. So $p \mid \pi\overline{\pi}$, and therefore $p \mid \pi$ or $p \mid \overline{\pi}$, since p is irreducible (and both are true, since multiplicative relations are preserved by conjugation). But since π is irreducible, then $\pi = pu$ for a unit $u \in \mathbf{Z}[i]^\times$.

If there are no factors equal to 3 mod 4, then suppose $p \equiv 1 \ (4)$, so $p = \alpha\overline{\alpha}$, where $\alpha = x + iy$ and $x^2 + y^2 = p$; furthermore, α is irreducible. Then, in $\mathbf{Z}$, $\pi\overline{\pi} = p(-) = \alpha\overline{\alpha}(\overline{--})$; thus, $\alpha \mid \pi\overline{\pi}$. Thus, $\alpha \mid \pi$ or $\alpha \mid \overline{\pi}$, and since π is irreducible, then $\pi = u\alpha$ or $\pi = u\overline{\alpha}$ where $u \in \mathbf{Z}[i]^\times$.

Finally, if $p = 2$, it's the same as $p \equiv 1 \ (4)$, but with $(1 + i)^2$ rather than $\alpha\overline{\alpha}$. ⊠

In summary, since we have a list of primes in $\mathbf{Z}$, we can use it to get a list of primes in $\mathbf{Z}[i]$. Everything is conditional on being able to factor in $\mathbf{Z}$.

We'll see similar results for other (rings of integers of) small quadratic fields, but for other things, such as cubic fields, characterizing irreducibles in terms of the norms will be kind of hopeless. We will be able to talk about irreducibles, but not quite as nicely, and sometimes the UFD property won't hold.

There will also be much more elegant ways of doing things; this style of proof will not be typical in this course! Eventually, we'll use ideals rather than elements to deal with a lack of unique factorization, and a finite group called the class group which is trivial in the case of a UFD, but can be used to prove stuff, e.g. Kummer proved a lot of exponents for Fermat's last theorem (though not all of them) using the class group, even when things aren't UFDs.

So we can use this to (well, theoretically) write a computer program to factor Gaussian integers. Suppose we've been handed an $\alpha = x + iy \in \mathbf{Z}[i]$.

(1) First, use the Euclidean algorithm to extract common factors of x and y. Thus, $\alpha = n(c_1 + c_2 i)$ for some $n \in \mathbf{Z}$ (not necessarily positive), and $\gcd(c_1, c_2) = 1$.

(2) Now, compute the norm of $c_1 + c_2 i$; if it's even, then we can divide by $1 + i$. If $N(c_1 + c_2 i) = 2^e m$ for an odd m, then
$$\frac{c_1 + c_2 i}{(1 + i)^e} = \frac{(1 - i)^e (c_1 + c_2 i)}{2^e} \in \mathbf{Z}[i],$$
and we'll call this $c_1' + c_2' i$.

(3) Now, we've reduced to where the norm is odd, i.e. $e = 0$. Thus, all of the irreducible factors π_j come from some $p_j \equiv 1 \ (4)$. In particular, $c_1^2 + c_2^2 = N(\pi_1) \cdots N(\pi_s)$, and we know that for each j, either π_j or $\overline{\pi}_j$ divides $c_1 + c_2 i$. Thus, we can choose all of the π_j and test which one it is, but then the result might be off by a unit, which is easy to fix.

Thus, factoring in $\mathbf{Z}[i]$ is no worse than factoring in $\mathbf{Z}$,[6] and thanks to the norm.

Example 3.2. Consider $8 + 13i$. Its norm is $64 + 169 = 233$, which is prime. Wait, that means $8 + 13i$ is irreducible.

[6]Not that factoring in $\mathbf{Z}$ is easy, cryptographers hope.

Let's step back more generally; as we saw in the first lecture, it's worthwhile to ask the question, *when is* $\mathbf{Z}[\sqrt{d}]$ *a UFD?* For example, if $d > 1$, then we get $d = 2, 3, 6, 7, 10, 11, 13, 14, \ldots$.

It's a conjecture of Gauss that there are infinitely many $d > 1$ such that $\mathbf{Z}[\sqrt{d}]$ is a UFD; on the other hand, the imaginary quadratic field case was solved in the 1980s. There are finitely many in this case, and the last one is -168.[7] This is a very hard theorem, and is called *Gauss' class number 1 problem*; he framed it in a very different language than we did (both mathematically and literally, in German or Latin!), though it is equivalent.

But there's one more thing to refine: is $\mathbf{Z}[\sqrt{d}] \subset \mathbf{Q}(\sqrt{d})$ even the right ring? At face value, it seems reasonable, but remember that when $d = -3$, $\mathbf{Q}(\zeta_3) \supset \mathbf{Z}[\zeta_3] \supset \mathbf{Z}[\sqrt{-3}]$. $\mathbf{Z}[\zeta_3]$ has $\mathbf{Z}$-basis $\{1, \zeta_3\}$, since $\zeta_3^2 = -1 - \zeta_3$. Then, $\mathbf{Z}[\sqrt{-3}]$ has $\mathbf{Z}$-basis $\{1, \sqrt{-3} = 2\zeta_3 + 1\}$. They're both free $\mathbf{Z}$-modules of the same dimension, but these bases show the latter has index 2 in the former, and isn't a UFD, where $\mathbf{Z}[\zeta_3]$ is. Thus, when we look more carefully, $\mathbf{Z}[\sqrt{-3}]$ isn't actually the best ring to look at.

As another example, when $d = 5$, $\mathbf{Z}[\sqrt{5}]$ is conspicuously not a UFD; however, if $\phi = (1 + \sqrt{5})/2$, then $\mathbf{Z}[\phi]$ is a UFD, and since $\phi^2 = \phi + 1$, then this contains $\mathbf{Z}[\sqrt{5}]$ with index 2; they're both free $\mathbf{Z}$-modules of the same rank.

Another interesting example is $\mathbf{Z}[i/2] = \{(a + bi)/2^e \mid a, b \in \mathbf{Z}, e > 0\}$, which isn't a finitely generated $\mathbf{Z}$-module. We'll leave it as an exercise; the issue is that $(i/2)^2 = -1/4$, so the denominators get worse and worse.

Thus, what we want is an intrinsic notion: when is something integral? Not just "when it looks integral" — denominators are fickle.

Definition. A *number field* is a finite extension of $\mathbf{Q}$. An *algebraic number* is an element of a number field (or sometimes, an element of a more general field that is algebraic over $\mathbf{Q}$).

By the Primitive Element theorem, K is a number field iff $K = \mathbf{Q}(\alpha)$, where α is the root of an irreducible monic $f \in \mathbf{Q}[x]$.

Definition. If K is a number field, then an $\alpha \in K$ is *integral*, or an *algebraic integer*, if there exists a monic $h \in \mathbf{Z}[x]$ such that $h(\alpha) = 0$.

Does the minimal polynomial work? We'll come back to that; the answer is yes.

This notion of integrality feels amorphous and hard to check (we'll find out how to make it easier). We'll eventually check that the sum and product of algebraic integers is algebraic, though this is more about module theory, rather than linear algebra, and then we'll be able to see that this is indeed a ring.

Example 3.3.

- What if $K = \mathbf{Q}$? Then, the algebraic integers of $\mathbf{Q}$ are just $\mathbf{Z}$; this is the Rational Root theorem from high school: if $f \in \mathbf{Q}[x]$ is monic, then the denominator of a root divides the lead coefficient, which is 1. This is why $\mathbf{Z}$ is sometimes called the *rational integers* if there is ambiguity.
- Let d be a square-free integer not equal to 0 or 1 and $u, v \in \mathbf{Z}$; then, how about $\alpha = u + v\sqrt{d} \in \mathbf{Z}[\sqrt{d}]$? If $v = 0$, we're good, and if $v \neq 0$, the trace shows us that $\alpha^2 - (2u)\alpha + (u^2 - dv^2) = 0$ (which you could also verify by hand, which is less satisfying).

Are there more? We hinted at this for $\mathbf{Z}[\zeta_3]$, and we'll talk more about this next time, including that it's finitely generated. This ring of algebraic integers will be the right thing to study.

4. Algebraic Integers: 4/6/15

Recall that many of the things which we've done with algebraicity use linear algebra. Now, though, we're talking about rings of integers, not fields, so not everything carries over. We'll see that a lot of it does, but we'll need some commutative algebra.

Recall that if $K/\mathbf{Q}$ is a finite field extension, then an $\alpha \in K$ is an *algebraic integer* if $f(\alpha) = 0$ for some monic $f \in \mathbf{Z}[x]$. Note that it's *not* in the definition that the minimal polynomial $m_\alpha \in \mathbf{Q}[x]$ is in $\mathbf{Z}[x]$, even though the minimal polynomial for an algebraic integer is always in $\mathbf{Z}[x]$. The reason we have this more broad definition is that, even though it's equivalent, it's a lot nicer to have this one for getting the theory off of the ground.

[7] "There's a tragic element here involving a German math teacher..."

For example, the Rational Root theorem shows that the algebraic integers of $\mathbf{Q}$ are $\mathbf{Z}$, which are therefore often called the *rational integers*.

Question. If α and β are algebraic integers, what about $\alpha+\beta$ and $\alpha\beta$? We can probably figure out that $-\alpha$ is.

That is, is the set of algebraic integers in K, denoted $\mathscr{O}_K$, a subring?

For those of you live-TEXing these lectures, be sure to use `\mathscr{O}`, not `\mathcal{O}`, which looks like $\mathcal{O}_K$, which is not nearly as pretty.

For a field extension L/K, if $\alpha \in L$ is algebraic (the analogue for fields) iff $k(\alpha)/k$ is finite degree, so if it's degree d, then $\{1, \alpha, \ldots, \alpha^d\}$ must be linearly dependent over k. Thus, any dependence relation gives the algebraic relation over k. This is the setup over fields, but over rings we don't have dimension, we don't have linear algebra, and so on. And by $k^\times$-scaling, this dependence relation can always be made monic. That is, asking for a monic relation in the case of fields doesn't mean anything.

Furthermore, if $\alpha, \beta \in L$ are both algebraic, then

$$k[\alpha,\beta] = \left\{ \sum_{\text{finite}} c_{ij}\alpha^i\beta^j \mid c_{ij} \in k \right\}$$

is a finite-dimensional k-subalgebra of L, and therefore it contains $\alpha+\beta$ and $\alpha\beta$, as are all $(\alpha+\beta)^i$ and $(\alpha\beta)^i$. Thus, these have algebraic relations over k. In particular, using the degree-d dependence relation of α, we don't need anything more than d^{th} powers in α (and the same thing for β).

Of course, $\mathbf{Z}$ is not a field, so a lot of this doesn't work. There are some nice things that happen because it's a PID, but we're going to need a few cases where the base ring isn't a PID, so let's try not to lean on it.

Our goal will be to show that if K is a number field, then $\mathscr{O}_K \subset K$ is a subring, and is a finitely generated $\mathbf{Z}$-module.

You'll always be able to turn something into an integral element by scaling appropriately: if $\alpha^n + q_{n-1}\alpha^{n-1} + \cdots + q_1\alpha + q_0 = 0$, for $q_i \in \mathbf{Q}$, then find a common denominator N for the q_i, and multiply the relation through by N^n.[8] Then, the modified dependence relation shows that $N\alpha \in \mathscr{O}_K$, so $K = \mathbf{Q}(N\alpha)$.

Here, we implicitly had the Primitive Element theorem floating around, but the analogous result for integrality is untrue: given any $n \in \mathbf{N}$, there exists a number field whose ring of integers needs at least $n+1$ generators. In fact, we'll construct examples!

This has a beautiful geometric interpretation for algebraic curves over finite fields, but we'll get to that later.

Example 4.1 (Dedekind)**.** Let α be a root of $\alpha^3+\alpha^2-2\alpha+8=0$. Then, if $K=\mathbf{Q}(\alpha)$, then $\mathscr{O}_K$ isn't generated by any single element. This was one of the first examples found.

So this is quite different from algebraicity; we'll have to do everything differently.

Observe that if $\alpha \in \mathscr{O}_K$ satisfies a relation of degree d, so that $\alpha^d + c_{d-1}\alpha^{d-1} + \cdots + c_1\alpha + c_0 = 0$, with the $c_i \in \mathbf{Z}$, then $\mathbf{Z}[\alpha]$ is generated as a $\mathbf{Z}$-module by $\{1, \alpha, \ldots, \alpha^{d-1}\}$, because any expression in some disgustingly high powers of α can be reduced by the relation: α^d is expressible in terms of lower stuff, and this can keep going. This is why it's crucual to keep it monic: so that α^d is a $\mathbf{Z}$-linear combination of lower-order stuff, and therefore so is α^{d+1}, and so on. This uses no dimensional argument.

This is a finitely generated $\mathbf{Z}$-module.[9] For a non-example, look at $\mathbf{Z}[2/3]$. Then, $2^n/3^n$ is in this ring for $n \gg 0$, and this is independent of all smaller n, so this isn't a finitely generated $\mathbf{Z}$-module: any finite collection has a bounded power of 3 in the denominator. Thus, under $\mathbf{Z}$-linear combinations, we can't make the denominator worse. (This *is* finitely generated as a $\mathbf{Z}$-algebra, where we can multiply, but that's not the goal here.)

This is the device that allows us to bypass dimension: we'll be able to show that $\alpha \in K$ lies in $\mathscr{O}_K$ iff $\mathbf{Z}[\alpha]$ is a finitely generated $\mathbf{Z}$-module. Note that we're *not* claiming at the outset that $\{1, \alpha, \ldots, \alpha^{d-1}\}$ is a basis; this will be shown to be true eventually, but requires more of an argument. (This relates to the discussion above that $\mathbf{Z}[\alpha,\beta]$ is a finitely generated $\mathbf{Z}$-module, since it's generated by terms $\alpha^i\beta^j$, where $i < d$ and $j < d'$, where d and d' are the degrees of the respective monic polynomials.)

[8] Of course, this isn't exactly practical, but it works.

[9] "For those of you taking notes on a computer, you are not on a blackboard, so you should use boldface $\mathbf{Z}$, not blackboard bold $\mathbb{Z}$."

In particular, $\mathbf{Z}[\alpha+\beta]$ and $\mathbf{Z}[\alpha\beta]$ are submodules of $\mathbf{Z}[\alpha+\beta]$, which is a finitely generated module over a PID.

Theorem 4.2. *Submodules of finitely generated modules over a PID are finitely generated.*

This is a corollary of the structure theorem for modules over a PID.[10] There's the other related fact that for finitely generated modules over a PID, torsion-free implies free, which is decidedly not true in the general case (hello, $\mathbf{Q}$).

So now we know $\mathbf{Z}[\alpha+\beta]$ and $\mathbf{Z}[\alpha\beta]$ are finitely generated, and therefore (as we'll prove), they're integral, though we have no idea what the degree is (e.g. what if $\alpha = -\beta$?) The proof will involve an amazing trick with determinants.

Let's generalize! The context for this discussion will be useful later.

Definition. Suppose $A \hookrightarrow B$ is an inclusion of commutative rings; then, a $b \in B$ is *integral over* A if it satisfies a monic relation $b^n + a_{n-1}b^{n-1} + \cdots + a_1 b + a_0 = 0$, with the $a_i \in A$.

The set of such b is called the *integral closure* of A in B.

For example, we've seen that when $A = \mathbf{Z}$ and $B = K$, then the integral closure is $\mathcal{O}_K$. When A and B are fields, this simply expresses algebraicity (since the monic condition doesn't add any more information).

Since B isn't even required to be a domain, there's no sense of minimal polynomial. But nonetheless, the same argument as before shows that if $b \in B$ is integral, the subalgebra $A[b]$ is a finitely generated A-module, generated by $\{1, b, \ldots, b^{n-1}\}$, because, again, b^n can be written in terms of lower powers of b.

When an element is integral, the algebra it generates over A is finitely generated; this will end up characterizing integrality.

Remark. If A is a domain and $B = \mathrm{Frac}(A)$, then the integral closure of A in B is usually denoted $\widetilde{A}$, and one says that A is *integrally closed* if $A = \widetilde{A}$. This is a version of the Rational Root theorem, which says that $\mathbf{Z}$ is integrally closed, and more generally, the ring of integers of a number field is also integrally closed (which is why the term "integral closure" is the correct one.)

Theorem 4.3. *If $A \hookrightarrow B$ is an inclusion of rings, then $b \in B$ is integral over A iff $b \in R \subset B$, where R is an A-subalgebra of B that is finitely generated as an A-module.*

Note that we're not just taking $R = A[b]$; this is broader.

Corollary 4.4. *If b and b' are integral over A, then $b + b'$ and $bb' \in A[b, b'] = R \subset B$, and R is a finitely generated A-module.*

This is exactly what we've been working toward. We're not going to produce the monic relation, especially if you want to compute (it'll be a massive determinant). The corollary is why Theorem 4.3 used a more general R, since we need more than just $A[b]$ in the corollary. And the general formulation of integrality allows us to escape PIDs.

Finally, we get the following: integrality is a very well-behaved notion.

Corollary 4.5. *The integral closure of A in B is a subring.*

So, how do we prove Theorem 4.3? Cramer's rule!

Proof of Theorem 4.3. In the forward direction, take $R = A[b]$.

In the reverse direction, it's possible to write (though by no means uniquely)

$$R = \sum_{i=1}^{N} A r_i,$$

with the $r_i \in R$. Thus,

$$b \cdot r_j = \sum_i a_{ij} r_i,$$

[10] This might not be covered in the prerequisites in the class, since it's from Math 122; thus, if you need to read up, take an algebra textbook to a weekend hackathon or something. Perhaps the result will be more interesting than yet another app.

or in terms of matrices,

$$\begin{pmatrix} b & & & \\ & b & & \\ & & \ddots & \\ & & & b \end{pmatrix} \begin{pmatrix} r_1 \\ \vdots \\ r_N \end{pmatrix} = \begin{pmatrix} & a_{ij} & \end{pmatrix} \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix}.$$

Thus, we can subtract $bI - (a_{ij})$ and call it M. Then, the adjugate matrix M^{adj} times M is diagonal on $\det(M)$ (which, well, is true over fields. Why over rings? Tune in next time!), and therefore we get that $\det(M)I$ times the vector of r_is is zero. Thus, $(\det M)r_i = 0$ for all i, and therefore $(\det M) \cdot 1 = 0$, and the determinant is monic in b over A. ⊠

Well then. This is the determinant trick; it's crucial for bootstrapping integrality, came out of nowhere, and will never be seen again. It's weird and amazing and a little beautiful.