MA542 Lecture

Timothy Kohl

Boston University

February 24, 2025

Divisibility in Integral Domains

What is the connection between the concept of 'prime' (as in $\mathbb{Z}$) vs. 'irreducible' (as in $F[x]$) in a domain?

We need to give some formal definitions.

Definition

Let D be an integral domain.
Elements a, b of D are <u>associates</u> if $a = ub$ for $u \in U(D)$.

A non-zero element $a \in D$ is <u>prime</u> if a is not a unit, and $a \mid bc$ implies $a \mid b$ or $a \mid c$.

A non-zero element $a \in D$ is <u>irreducible</u> if a is a non-unit, and if $a = bc$ for $b, c \in D$, then either b or c are units.

One might see that the definition of irreducible seems to match up with the usual definition of 'prime number', namely:

Definition

An natural number p is prime if it is only divisible by itself and 1.

i.e. it cannot be factored $p = ab$ for integers $a > 1$ and $b > 1$.

However, a prime number *does* satisfy the primality definition we just gave in that if $p \mid ab$ then either $p \mid a$ and/or $p \mid b$.

And, although this may be a confusing thing to note at this point, it seems as if, for integers at least, prime and irreducible are the same thing.

This *is* true in $\mathbb{Z}$, but for general integral domains, we want to look at these definitions separately since, in other domains, they are not equivalent notions.

We begin with a relationship that does hold between primes and irreducibles.

Theorem

In an integral domain D, every prime is irreducible.

Proof.

Let $p \in D$ be a prime and suppose $p = ab$ for a, b **non-units**.
Since $p = ab$ then obviously $p \mid ab$ so either $p \mid a$ or $p \mid b$, so suppose $a = cp$.

Thus $p = cpb$ and since D is a domain we can 'cancel' the common factor of p from both sides to get $1 = cb$ which implies that b is a unit,which is a contradiction.

Thus p is irreducible. $\square$

So in a domain PRIME $\longrightarrow$ IRREDUCIBLE so it begs the question, are there irreducible elements in a domain which aren't prime?

The answer is YES.

The reason that these definitions are distinct has to do with the nature of the units in the domain.

Consider the ring $\mathbb{Z}[\sqrt{d}] = \{a + b\sqrt{d}\}$ where $d \neq 1$ is an integer which is not divisible by the square of any integer, what we term 'square free', e.g. $d = -1, 2, 3, 5, 6, \dots$, etc.

Define the norm $N : \mathbb{Z}[\sqrt{d}] \to \mathbb{Z}^+$ (non-negative integers) by

$$N(a + b\sqrt{d}) = |a^2 - db^2|$$

which has several important properties which make it *almost* like a homomorphism.

Properties of $N(a+b\sqrt{d}) = |a^2 - db^2|$

- $N(x) = 0$ iff $x = 0$
- $N(xy) = N(x)N(y)$
- $N(x) = 1$ iff x is a unit of $\mathbb{Z}[\sqrt{d}]$.

These aren't too difficult to verify.

For example if $x = a + b\sqrt{d}$ then $N(x) = |a^2 - db^2|$ so $N(x) = 0$ implies that $a^2 = db^2$, namely that $\left(\frac{a}{b}\right)^2 = d$.

But this implies that $\frac{a}{b} = \pm\sqrt{d}$ which is impossible since $a, b \in \mathbb{Z}$ and $\sqrt{d}$ is irrational. (Why?)

That $N(xy) = N(x)N(y)$ is a pure calculation, to wit:

$$(a + b\sqrt{d})(A + B\sqrt{d}) = (aA + dbB) + (aB + Ab)\sqrt{d}$$

and

$$(aA + dbB)^2 - d(aB + Ab)^2 = a^2A^2 + d^2b^2B^2 - da^2B^2 - dA^2b^2$$

while $(a^2 - db^2)(A^2 - dB^2) = a^2A^2 + d^2b^2B^2 - da^2B^2 - dA^2b^2$

a tedious but straightforward verification.

And $N(a+b\sqrt{d})=1$ implies that $a^2-db^2=\pm 1$ so if you look at $\frac{1}{a+b\sqrt{d}}$ you see that it equals

$$\frac{a}{a^2-db^2}+\frac{-b}{a^2-db^2}\sqrt{d}$$

which is an element of $\mathbb{Z}[\sqrt{d}]$ only if $\frac{a}{a^2-db^2}\in\mathbb{Z}$ and $\frac{-b}{a^2-db^2}\in\mathbb{Z}$, which means the denominator a^2-db^2 must be ± 1, i.e. $N(a+b\sqrt{d})=1$.

Indeed,

$$U(\mathbb{Z}[\sqrt{d}]) = \{a + b\sqrt{d} \mid |a^2 - db^2| = 1\}$$

and it's interesting to see how the choice of d affects the size and structure of this group.

For example, when $d = -1$ we get the Gaussian integers whose only units are $\{\pm 1, \pm i\}$.

But for $d = 2$ it turns out that the units are of the form

$$\{\pm(1 + \sqrt{2})^n\} \cong \mathbb{Z}_2 \times \mathbb{Z}$$

specifically $\langle -1 \rangle \times \langle (1 + \sqrt{2}) \rangle$.

For those interested in this subject in general, Dirichlet's Unit theorem describes these unit groups in terms of a 'torsion' component (i.e. finite), and a 'free' component, namely a product of a certain number of copies of $\mathbb{Z}$.

We now demonstrate the existence of an irreducible which is not a prime.

Consider $D = \mathbb{Z}[\sqrt{-3}]$ where $N(a + b\sqrt{-3}) = |a^2 + 3b^2| = a^2 + 3b^2$.

Observe that $N(1 + \sqrt{-3}) = 4$ and suppose $1 + \sqrt{-3} = xy$ for some $x, y \in D$ which *aren't* units, then $N(1 + \sqrt{-3}) = N(xy) = N(x)N(y)$ where $N(x) \neq 1$ and $N(y) \neq 1$.

Since $N(x), N(y)$ are positive integers, (both greater than 1) then it must be the case that $N(x) = 2$ and $N(y) = 2$.

However, if $x = a + b\sqrt{-3}$ then $N(x) = a^2 + 3b^2$ and since a and b are not both zero then it is impossible for $N(x) = 2$.

As such $1 + \sqrt{-3} = xy$ for x, y non-units is impossible, so $1 + \sqrt{-3}$ is irreducible.

So now, let's observe that $(1+\sqrt{-3})(1-\sqrt{-3}) = 4 = 2\cdot 2$ which means that

$$(1+\sqrt{-3}) \mid 2\cdot 2$$

so does $(1+\sqrt{-3}) \mid 2$?

Well if $2 = (a+b\sqrt{-3})(1+\sqrt{-3}) = (a-3b)+(a+b)\sqrt{-3}$ then $a-3b = 2$ and $a+b = 0$ which implies that $2b = 1$ which is impossible since $b \in \mathbb{Z}$!

So what we've demonstrated is that $1+\sqrt{3}$ is an irreducible which is **not** a prime.

We still have the case of $\mathbb{Z}$ where prime and irreducible mean the same thing, so when *does* this occur?

Definition

An integral domain R is a principal ideal domain (PID) if for every ideal $I \subseteq R$, there is an $a \in R$ such that $I = \langle a \rangle$.

There are two fundamental examples, which we've already encountered.

We already proved that $\mathbb{Z}$ is PID by using the division algorithm to show every member of an ideal $I \subseteq \mathbb{Z}$ is a multiple of a fixed positive integer.

What's nice about this example is that the proof of it (i.e. the use of the division algorithm) also implies that $F[x]$ is a PID for F a field.

Basically, if $I \subseteq F[x]$ is an ideal, then let $f(x)$ be a polynomial of minimal degree in I, then for any $g(x) \in I$ one can write $g(x) = q(x)f(x) + r(x)$ where $r(x) = 0$ (meaning $f(x) \mid g(x)$) or $deg(r(x)) < deg(f(x))$.

But then $r(x) \in I$ too, so if $r(x) \neq 0$ then we have a contradiction of the fact that $deg(f(x))$ is the smallest degree of any polynomial in I!

There are other examples of PID's we can mention now, some of which will be proven to be so later on.

- $F[[x]]$ - the ring of 'formal power series' under addition and multiplication where F is a field
- $\mathbb{Z}[i]$
- $\mathbb{Z}[\sqrt{2}]$
- $\mathbb{Z}[\zeta] = \{a + b\zeta \mid a, b \in \mathbb{Z}\}$ where $\zeta = e^{\frac{2\pi i}{3}}$, namely a complex cube root of unity

The latter three examples, all use the norm idea developed earlier, but there are some extra conditions needed.

A non-example to consider is $\mathbb{Z}[\sqrt{-5}]$, which we'll discuss in a broader context later on.