Math 75 notes

P. Pollack and C. Pomerance

1 Euclid's algorithm and $\mathbb{Z}/(p)$

We start at a place that might seem a bit far afield: Euclid's algorithm from elementary number theory for finding the greatest common divisor of two integers. Take the example of the integers 91 and 143. We start Euclid's algorithm by taking the smaller of the numbers, 91, into the larger; this goes in once and leaves a remainder of 52. We then take the remainder, 52, into the last divisor, 91. This goes in once, and we find a new remainder of 39. We keep this process up until one of these divisions works out without any remainder. According to Euclid, the last nonzero remainder (13 in this case) is the gcd:

$$\begin{aligned} 143 &= 91 \cdot 1 + 52 \\ 91 &= 52 \cdot 1 + 39 \\ 52 &= 39 \cdot 1 + 13 \\ 39 &= 13 \cdot 3 + 0. \end{aligned}$$

For another example we could take the numbers 7 and 17:

$$\begin{aligned} 17 &= 7 \cdot 2 + 3 \\ 7 &= 3 \cdot 2 + 1 \\ 3 &= 1 \cdot 3 + 0, \end{aligned}$$

so that in this case Euclid's algorithm tells us that 7 and 17 are *relatively prime*: their only common positive divisor is the number 1.

It is tempting to suppose that Euclid's algorithm was wasted on our last example. Indeed, 17 is a number which is immediately recognizable as a prime, so that it is obvious it has no factors in common with the smaller integer 7. But our effort is not in vain! Indeed, for our purposes it is not the final answer of Euclid's algorithm (which we could have predicted was 1 immediately) that is of interest, but the intermediate steps to reach that outcome which are of most interest.

These computations allow us to express 1 as a linear combination of 17 and 7. We start by writing

$$3 = 17 - 7 \cdot 2,$$

which comes from the first of our Euclidean steps above. Thus we have 3 as a linear combination of 7 and 17. Next we write

$$1 = 7 - 3 \cdot 2,$$

which is the result of transposing the second Euclidean step above. This gives 1 as a combination of 7 and 3. But we determined that $3 = 17 \cdot 1 + 7(-2)$, and so

$$1 = 7 \cdot 1 - (17 \cdot 1 + 7(-2))2 = 7 \cdot 5 + 17(-2).$$

And now if we look at this equation in the system $\mathbb{Z}/(17)$, where $17 = 0$, we find that

$$1 = 7 \cdot 5.$$

So our computation with Euclid's algorithm has earned its keep: it allowed us to *construct* an inverse of 7 modulo 17.

More generally, Euclid's algorithm gives one a constructive proof of the following important result:

Theorem 1. *Let a, b be integers, not both zero, and let d be the greatest common divisor of a and b. Then there are integers x and y with*

$$ax + by = d.$$

In particular, if a and b are relatively prime, then one can solve the equation

$$ax + by = 1.$$

This allows us to settle one of our outstanding debts. We can now prove the following result, mentioned in the introductory lecture:

Theorem 2. *Let p be a prime. Then $\mathbb{Z}/(p)$ is a field.*

The proof is now easy. We already know everything we need to know about $F = \mathbb{Z}/(p)$, except for the existence of multiplicative inverses for everything in $F \setminus \{0\}$. But $\mathbb{Z}/(p) = \{0, 1, \ldots, p-1\}$, and so $F \setminus \{0\} = \{1, 2, \ldots, p-1\}$. Each of $a = 1, 2, 3, \ldots, p-1$ are relatively prime to p, so that we can solve

$$ax + py = 1.$$

Reading this equation in $\mathbb{Z}/(p)$, we see that $ax = 1$, so that a has an inverse modulo p. Done!

It is worth emphasizing that the way we have proved this result not only shows the existence of inverses but gives an algorithm for computing them.

Example 1. *Let $p = 101$ and $a = 11$. Find a^{-1} in $\mathbb{Z}/(p)$.*

Solution. We begin by performing Euclid's algorithm on the integers 11 and 101:

$$\begin{aligned} 101 &= 11 \cdot 9 + 2 \\ 11 &= 2 \cdot 5 + 1 \\ 2 &= 1 \cdot 2 + 0. \end{aligned}$$

Now we solve for 1 as a linear combination of 11 and 101:

$$\begin{aligned} 2 &= 101 + 11 \cdot (-9) \\ 1 &= 11 + 2 \cdot (-5) \\ &= 11 + (101 + 11 \cdot (-9)) \cdot (-5) \\ &= 11 \cdot 46 + 101 \cdot (-5). \end{aligned}$$

Reading this last equation in $\mathbb{Z}/(101)$, we see $1 = 11 \cdot 46$. So 46 is the inverse of 11. □

2 Constructing new fields from old

The insight that $\mathbb{Z}/(p)$ is a field is from the viewpoint of generalization a rather fundamental one. Cast in broad terms, we started with a system $\mathbb{Z}$ (not itself a field) and created a field by supplementing the arithmetic of $\mathbb{Z}$ with a new equation '$p = 0$' for a prime p. In order to understand all finite fields, we will use a similar construction, but with systems other than $\mathbb{Z}$ as our starting points. We now introduce these systems.

Let F be a field. We define the *univariate polynomials over F in the indeterminate x* as the set of expressions of the form

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_k x^k,$$

where all the a_i belong to F. The *degree of a polynomial* is the highest power of x that appears with a nonzero coefficient; for example, over $\mathbb{R}$, the polynomial $3x^2 + 1$ has degree 2 and the constant polynomial $1 = 1x^0$ has degree 0. The polynomial 0 (corresponding to choosing all the a_i above to equal zero) has no powers of x which appear with a nonzero coefficient; we leave its degree undefined.

When the field F is the field of real or complex numbers, these definitions are familiar, but they work just as well over any field (e.g., $\mathbb{Z}/(p)$ for a prime p). Moreover, the usual operations of addition and multiplication for polynomials make sense in this general context. For example, let $F = \mathbb{Z}/(5)$, and define $A, B \in F[x]$ by

$$A = 3x^3 + x + 1, \quad B = 4x + 2.$$

Then

$$A + B = 3x^3 + 5x + 3 = 3x^3 + 3$$

(recalling that $5 = 0$ in $\mathbb{Z}/(5)$) while

$$\begin{aligned} AB &= 12x^4 + 6x^3 + 4x^2 + 2x + 4x + 2 \\ &= 12x^4 + 6x^3 + 4x^2 + 6x + 2 \\ &= 2x^4 + x^3 + 4x^2 + x + 2. \end{aligned}$$

With these definitions of $+, \cdot$, the system $F[x]$ becomes what is called a *commutative ring*:

- it is an abelian group under addition (with additive identity the polynomial 0),
- '$\cdot$' distributes over '+': for all $A, B, C \in F[x]$, we have $A(B + C) = AB + AC$,
- there is a multiplicative identity (the constant polynomial 1),
- multiplication is commutative ($AB = BA$ for all $A, B \in F[x]$) and associative (i.e., $A(BC) = (AB)C$ for all $A, B, C \in F[x]$).

These properties look very close to the properties singled out in our definition of a field. Nevertheless, $F[x]$ is *not* a field. In a field every nonzero element has an inverse, but it turns out that there are relatively few elements of $F[x]$ which are invertible. We call invertible elements *units*.

Proposition 1. *The only units in $F[x]$ are the nonzero constants in F.*

The proof is simple: Suppose that A is an element of $F[x]$ that has an inverse. Write

$$A = a_0 + a_1 x + \cdots + a_k x^k$$

where we assume that a_k is nonzero in F. If A has an inverse, then $AB = 1$ for some polynomial $B \in F[x]$. Write

$$B = b_0 + b_1 x + \cdots + b_j x^j,$$

where b_j is nonzero. In the product AB, the term x^{k+j} shows up with a coefficient $a_k b_j$ of F, and one can show (homework!) that $a_k b_j$ is nonzero since both a_k and b_j are. So AB has degree $k + j$. But we were supposed to have that $AB = 1$. Thus we need $k + j = \deg 1 = 0$, which means that $k = j = 0$, and so both A and B are constant polynomials.

The proposition tells us that we will have to do quite a bit more work to get the fields we are after: starting with F and forming $F[x]$ didn't give us any invertible elements.

Our approach will be the same as the approach taken in elementary number theory to constructing the systems $\mathbb{Z}/(m)$. That approach begins with defining the notion of congruence modulo m. In analogy, we make the following definition: We call two polynomials $A, B \in F[x]$ *congruent modulo M* if M divides $A - B$, i.e., if there is a polynomial $Q \in F[x]$ for which

$$A - B = MQ.$$

In this case we write

$$A \equiv B \pmod{M}.$$

For example, if $F = \mathbb{Z}/(2)$, then $x^3 \equiv x^2 + x + 1 \pmod{x + 1}$ in $F[x]$, because $x + 1$ divides the difference $x^3 - (x^2 + x + 1) = x^3 + x^2 + x + 1 \in F[x]$. (One can check this with the familiar polynomial long division algorithm.)

If F is a field and M is a polynomial in $F[x]$, then congruence modulo M is an equivalence relation. In other words, this relation is...

- reflexive: $A \equiv A \pmod{M}$ for every A,
- symmetric: if $A \equiv B \pmod{M}$, then $B \equiv A \pmod{M}$,
- transitive: if $A \equiv B \pmod{M}$ and $B \equiv C \pmod{M}$, then $A \equiv C \pmod{M}$.

Because of this, the ring $F[x]$ is partitioned into equivalence classes. We define $F[x]/(M)$ as the set of equivalence classes. For each $A \in F[x]$, denote the equivalence class containing A by $[A]$; then

$$[A] := \{B \in F[x] : B \equiv A \pmod{M}\},$$

and $F[x]/(M)$ is exactly the set of these classes $[A]$ as A ranges over $F[x]$.

The systems $F[x]/(M)$ will be our candidates for new fields. For this to make any sense at all, we need to define what it means to add and multiply two elements of $F[x]$. We make the only definition that really is sensible: we define

$$[A] + [B] = [A + B], \quad \text{and} \quad [A] \cdot [B] = [AB].$$

There is something to check here. The definition of addition we've given here basically says 'to add two equivalence classes, pick an element from each, add those two elements, and then take the equivalence class of the sum,' and similarly for multiplication. It is maybe not so obvious that no matter how the elements are chosen, one always winds up with the same result. This is a consequence of the fact that addition and multiplication respect congruence: if $A_1 \equiv A_2 \pmod{M}$ and $B_1 \equiv B_2 \pmod{M}$, then

$$A_1 + A_2 \equiv B_1 + B_2 \pmod{M}, \quad \text{and} \quad A_1 A_2 \equiv B_1 B_2 \pmod{M}.$$

We omit the (pretty easy) proofs of these facts here.

Let us do some examples to get a feel for the arithmetic of the systems $F[x]/(M)$. For the rest of this lecture we take $F = \mathbb{Z}/(5)$ and $M = x^2 + 1 \in F[x]$, and we try to understand the system $F[x]/(M)$.

First let us see if we can do some basic arithmetic. Consider the two elements $[x + 1]$ and $[x + 3]$ in $F[x]/(M)$. By definition,

$$[x + 1] + [x + 3] = [(x + 1) + (x + 3)] = [(2x + 4)],$$

and

$$[x + 1][x + 3] = [(x + 1)(x + 3)] = [(x^2 + 4x + 3)].$$

Actually the latter answer can be put in a somewhat simpler form. We have

$$[x^2 + 4x + 3] = [x^2 + 1 + 4x + 2] = [x^2 + 1] + [4x + 2] = [0] + [4x + 2] = [4x + 2].$$

(This calculation of reduction should be viewed as being totally analogous to (say) the calculation that $3 \cdot 5 = 15 = 8$ in the system $\mathbb{Z}/(7)$.) It is somewhat cumbersome to keep having to

write these brackets, and so we omit them when the system we are working with is understood. So, for example, in the future we will just write

$$(x+1)(x+3) = 4x+2$$

when we are understood to be working in $F[x]/(M)$ for $F = \mathbb{Z}/(5)$ and $M = x^2+1$.

How many elements are in our system $F[x]/(M)$? To answer this question it would be good if we had a way of telling when two elements are different. For example, which elements of the list

$$0, 1, x, 2x+2, 3x^3+2, x^5$$

represent distinct elements of $F[x]/(M)$? We can get some insight into this question if we recall that $M = x^2+1$ is zero in $F[x]/(M)$. So

$$3x^3+2 = 3x \cdot x^2 + 2 = 3x(-1) + 2 = -3x + 2 = 2x + 2$$

and

$$x^5 = x \cdot (x^2)^2 = x \cdot (-1)^2 = x.$$

In fact, it's not hard to see that by repeatedly applying the rule that $x^2 = -1$, we can show that every element of $F[x]/(M)$ has the form $a + bx$, where a and b belong to $F = \mathbb{Z}/(5)$.

An alternative proof (which gets this result in one fell swoop) is to use long division for polynomials. That process provides a constructive proof of the following fundamental result:

Theorem 3. *Let F be a field. Let A, B be elements of $F[x]$, and suppose $B \neq 0$. Then there are $Q, R \in F[x]$ with*

$$A = BQ + R \quad \textit{with } R = 0 \textit{ or } \deg R < \deg B.$$

Taking $B = M$, this result shows immediately that every polynomial is congruent to some polynomial which is either zero, or of degree smaller than that of M. In our case, where $F = \mathbb{Z}/(5)$ and $M = x^2+1$, that means that very equivalence class is represented by a polynomial either zero or of degree < 2. Again, these are precisely the polynomials of the form $a + bx$.

At this point it is tempting to assert that $F[x]/(M)$ has precisely $5 \cdot 5 = 25$ elements: there are 5 choices of a and 5 choices for b. But perhaps there is some repetition even among the elements $a + bx$? It seems hard to see how this could occur, since the rule $x^2 = -1$ that defines the system $F[x]/(M)$ doesn't seem to imply any such coincidences. We now prove that all the elements $a + bx$, with $a, b \in F = \mathbb{Z}/(5)$, do define distinct elements of $F[x]/(M)$, so that our system does have 25 elements as claimed. For the proof, notice that if

$$a + bx = a' + b'x$$

for $a, b, a', b' \in F$, then

$$(a - a') + (b - b')x = 0.$$

Expressed in the language of congruence, this asserts that

$$a - a' + (b - b')x \equiv 0 \pmod{x^2 + 1},$$

i.e., that

$$(a - a') + (b - b')x = (x^2 + 1)Q(x)$$

for some polynomial $Q(x) \in F[x]$. But if $Q(x)$ is nonzero, then the degrees of the left and right hand sides will not match up! So it must be that $Q(x) = 0$, which implies that $a - a' + (b - b')x = 0$, which in turn implies that $a = a'$ and $b = b'$.

The proofs we have just outlined work in general. The corresponding result is as follows:

Theorem 4. *Let F be a finite field. Let M be a nonzero polynomial in $F[x]$. Then the number of elements in $F[x]/(M)$ is q^d, where q is the number of elements of F and d is the degree of M. In fact, every element of $F[x]/(M)$ has the form*

$$a_0 + a_1 x + \cdots + a_{d-1} x^{d-1},$$

where $a_0, a_1, \ldots, a_{d-1} \in F$, and the q^d elements defined this way are all distinct.