Math 75 notes, Lecture 10

P. Pollack and C. Pomerance

Finite fields: existence and uniqueness

In Lectures 7 and 8, we established the following result on the possible sizes of finite fields:

Theorem A. *If F is a finite field with q elements, then $q = p^d$, where p is a prime and $d \geq 1$.*

As of the last lecture, we now know the complementary result that all these sizes do actually occur:

Theorem B. *If $q = p^d$, where p is a prime and $d \geq 1$, then there is a finite field K with q elements.*

Actually the proof of the last theorem gave us a procedure for constructing such fields. Start with $F = \mathbb{Z}/(p)$. Then search for an irreducible polynomial $f(x)$ of degree d over F. We know from the estimates for $I(f, d)$ in the last lecture that there is always at least one such polynomial, so our search is not in vain! Then $F[x]/(f(x))$ is a field of the size we want.

In many mathematical problems there are two distinct questions: *existence* and *uniqueness*. Theorem B tells us that finite fields with q elements exist for every prime power q. It doesn't tell us whether there is just one finite field of size q, or whether there are millions of them.

Example 1. *Suppose first that $q = 4$. As we have seen, we can construct a field of size 4 by starting with $F = \mathbb{Z}/(2)$ and forming the quotient $K = F[x]/(x^2 + x + 1)$. We've seen in homework that this is the* only *example of a field of size 4, up to relabeling (isomorphism).*

Example 2. *Suppose now that $q = 9$. To construct a field of 9 elements we can start with $F = \mathbb{Z}/(3)$ and take a monic irreducible polynomial $f(x) \in F[x]$ of degree 2. It's easy to check that there are then only three choices here:*

$$x^2 + 1, \qquad x^2 + 2x + 2, \qquad x^2 + x + 2.$$

So we get (ostensibly) three 9-element fields, namely

$$F[x]/(x^2 + 1), \qquad F[x]/(x^2 + 2x + 2), \qquad F[x]/(x^2 + x + 2).$$

It is perhaps not obvious whether a closer look would show these these fields to be really the same (up to isomorphism). Nor is it obvious whether we've captured all finite fields of 9 elements: maybe there is as an example that doesn't look at all like $F[x]/(f(x))$.

There are numerous troubling possibilities implicit in this discussion. It turns out that the truth of the matter is quite clean and simple: the three finite fields of order 9 constructed in the last example are all the same (up to relabeling). And in fact, any finite field of order 9 can be relabeled to look like one of them (and so any of them). More generally, for any prime power

$q = p^d$, there is exactly one field (up to isomorphism) of order q. Establishing this important uniqueness result is the goal of today's lecture.

The notion of 'isomorphism'

It is time to make precise this word 'isomorphism' that we have been throwing around.

Suppose that R and S are commutative rings. The precise definition of a commutative ring was reviewed in Lecture 2; briefly, it is a system where one has addition and multiplication obeying the usual laws of arithmetic. However, unlike a field, we do not require that multiplicative inverses exist. Thus fields are a special case of commutative rings. Other examples of commutative rings include $\mathbb{Z}$, $F[x]$, and $\mathbb{Z}/(10)$.

By an *isomorphism* $\phi\colon R \to S$, we mean a bijective map which respects both addition and multiplication, i.e., satisfies

$$\phi(r_1 + r_2) = \phi(r_1) + \phi(r_2) \quad \text{for all } r_1, r_2 \in R,$$
$$\phi(r_1 r_2) = \phi(r_1)\phi(r_2) \quad \text{for all } r_1, r_2 \in R.$$

Intuitively, ϕ is a rule for relabeling the elements of R by elements of S, in such a way that under this relabeling the addition and multiplication tables of R and S look exactly the same. If there is an isomorphism from R to S, we say that R and S are *isomorphic*.

Because isomorphism is just a way of formalizing the intuitive notion of relabeling, a number of properties are more or less obvious. For example, isomorphism is an equivalence relation:

- (Reflexive) Every ring R is isomorphic to itself,
- (Symmetric) If R is isomorphic to S, then S is isomorphic to R,
- (Transitive) If R is isomorphic to S and S is isomorphic to T, then R is isomorphic to T.

Just as obviously, isomorphism preserves any properties that can be described just in terms of the addition and multiplication tables. In particular, it preserves the property of being a field: For example, if $\phi\colon R \to S$ is an isomorphism and we already know R is a field (i.e., has inverses for all its nonzero elements), then S must also be a field.

Uniqueness

Let K be an arbitrary finite field. Then K has size $q = p^d$, where p is prime and $d \geq 1$. We first prove a simple theorem characterizing the subfields of K.

Theorem 1. *For every j dividing d, there is a unique subfield of K of size p^j. This subfield is exactly the set of roots in K of the polynomial $x^{p^j} - x$. Conversely, every subfield of K has size p^j for some j dividing d.*

Proof. Suppose first that j divides d. If F is a subfield of K of size p^j, then every β in F satisfies $\beta^{p^j} = \beta$. But the polynomial $x^{p^j} - x$ has at most p^j roots in K, and so F must be exactly the set of roots of $x^{p^j} - x$ in K. So it's clear that if there is a subfield of size p^j, then it is unique. Moreover, this discussion implies that the only candidate for such a subfield is the set F of roots of $x^{p^j} - x$ in K.

Let's verify that $x^{p^j} - x$ does have p^j distinct roots in K, so that F has p^j elements. Using the hypothesis that j divides d, we can prove (and indeed this was done in Lecture 6) that $x^{p^j} - x$ divides $x^{p^d} - x$. Over K, we have the factorization

$$x^{p^d} - x = \prod_{\alpha \in K} (x - \alpha).$$

So unique factorization in $K[x]$ forces $x^{p^j} - x$ to also split into linear factors in $K[x]$, and hence the set F has p^j elements. So far so good. Let's show that F is a field. Since F is a subset of the field K, to prove that F is a subfield it's enough to observe that F contains 0 and 1, that F is closed under addition and multiplication, and that every nonzero element of F has a multiplicative inverse in F. (Note that the element $-\alpha$ is equal to the $(p-1)$-fold sum $\alpha + \alpha + \cdots + \alpha$, so we get additive inverses without any more work.) To check closure under $+$ and $\cdot$, notice that if if $\beta_1, \beta_2 \in F$, then

$$\beta_1^{p^j} = \beta_1 \quad \text{and} \quad \beta_2^{p^j} = \beta_2.$$

Thus

$$(\beta_1 \beta_2)^{p^j} = \beta_1^{p^j} \beta_2^{p^j} = \beta_1 \beta_2,$$

so that $\beta_1 \beta_2$ is a root of $x^{p^j} - x$ and so belongs to F. Moreover, by the 'bad student's binomial theorem', we have

$$(\beta_1 + \beta_2)^{p^j} = \beta_1^{p^j} + \beta_2^{p^j} = \beta_1 + \beta_2,$$

so that $\beta_1 + \beta_2$ is also in F. Lastly, suppose $\beta \in F$ is nonzero. Then it has a multiplicative inverse β^{-1} in K; moreover, since $\beta^{q^d} = \beta$, we have (taking inverses on both sides) that $(\beta^{-1})^{q^d} = \beta^{-1}$, and so β^{-1} belongs to F.

It remains to show that every subfield of K has size p^j for some j dividing d. Let F be an arbitrary subfield of K. Then we can view K as a vector space with F as the field of scalars. We now mimic the proof of Theorem 2 in Lectures 7 and 8: Suppose this vector space has dimension r, and let $\alpha_1, \ldots, \alpha_r$ be a basis for K over F. Then every element of K has a unique representation in the form

$$s_1 \alpha_1 + s_2 \alpha_2 + \cdots + s_j \alpha_j,$$

where the s_i come from F. As a consequence, the number of elements of K, which is p^d, must be the rth power of the number of elements of F. By unique factorization, this is only possible when r divides d and F has p^j elements for $j = d/r$. Clearly j divides d, so we are done. $\square$

Theorem 2. *Let K be a field of size $q = p^d$, where p is prime and $d \geq 1$. Suppose j divides d, and let $f(x)$ be an irreducible polynomial of degree j over $F = \mathbb{Z}/(p)$. Then the unique subfield of K of size p^j is isomorphic to $F[x]/(f(x))$.*

Suppose we take $j = d$ in Theorem 2. The unique subfield of K of size p^d is obviously K itself, and so we find that K is isomorphic to $F[x]/(f)$, for every irreducible polynomial f of degree d. Since K was arbitrary, it follows that any two fields of size p^d are isomorphic. We record this important consequence here for future reference:

Corollary 1. *Any two finite fields of the same size are isomorphic.*

Proof of Theorem 2. Since $f(x)$ has degree j over the field $F = \mathbb{Z}/(p)$, and j divides d, we know that $f(x)$ divides $x^{q^d} - x$. As we noted in the proof of Theorem 1 above, the polynomial $x^{q^d} - x$ splits entirely into linear factors in K. So by unique factorization, $f(x)$ must have j distinct roots in K. Fix one of these and call it α. Note that since $f(x)$ divides $x^{q^d} - x$, the element α is a root of $x^{q^d} - x$, and so α belongs to the unique subfield M (say) of order p^j guaranteed by Theorem 1. Define a map ϕ from $F[x]$ to M by taking the polynomial $g(x)$ to the element $g(\alpha)$.

Let's check that this makes sense to do. We've already observed that α belongs to M. Moreover, the field $F = \mathbb{Z}/(p)$, which is a subfield of K, is also a subfield of M. (This is clear since F just consists of 1 added to itself a number of times, and M contains 1.) Since M is a subfield, and so closed under addition and multiplication, it must contain everything of the form $g(\alpha)$, where $g(x) \in F[x]$.

So we have a well-defined map. But is this map useful in any way? It's easy to see that it preserves the operations, which is a promising start. For example, to check that it preserves multiplication, just observe that for any $g_1(x), g_2(x) \in F[x]$, we have

$$\phi(g_1(x)g_2(x)) = g_1(\alpha)g_2(\alpha) = \phi(g_1(x))\phi(g_2(x)).$$

The proof for addition is similar. So if ϕ were a bijection, we would have an isomorphism on our hands.

But ϕ isn't a bijection. This is obvious: Since $F[x]$ is infinite while M is finite, there's no way that ϕ can be injective. And it's easy to write down an example where injectivity fails: We have $\phi(0) = 0$ and $\phi(f(x)) = f(\alpha) = 0$, and $f(x)$ isn't the same element of $F[x]$ as the polynomial 0. More generally, if $g_1(x)$ and $g_2(x)$ are any two polynomials which differ by a multiple of $f(x)$, then $g_1(\alpha)$ will be the same as $g_2(\alpha)$, because $f(\alpha) = 0$. Said differently, $\phi(g_1(x))$ will be the same as $\phi(g_2(x))$ whenever $g_1(x) \equiv g_2(x) \pmod{f(x)}$.

In fact, differing by a multiple of $f(x)$ is the only obstacle to injectivity: Suppose that $\phi(g_1(x)) = \phi(g_2(x))$. Then $g_1(\alpha) = g_2(\alpha)$, and so, setting $h(x) = g_1(x) - g_2(x)$, the polynomial $h(x)$ vanishes at α. This implies that h is divisible by the minimum polynomial of α over F. What is this minimum polynomial? From the discussion of Lectures 4 and 5 (see in particular the statement of Theorem 2), we know that it is a monic irreducible in $F[x]$ which divides every polynomial in $F[x]$ that vanishes at α. But $f(x)$ is a polynomial in $F[x]$ that vanishes at α, and $f(x)$ is monic and irreducible! So it has to be that this minimum polynomial is just $f(x)$ itself. Hence $f(x)$ must divide the polynomial $h(x)$ above; since $h(x) = g_1(x) - g_2(x)$, we've shown that $g_1(x) \equiv g_2(x) \pmod{f(x)}$.

We've now isolated the 'reason' why ϕ fails to be injective: it identifies elements of $F[x]$ which are congruent modulo $f(x)$, even if they aren't the same. The map ϕ would be much happier mapping out of a space where elements of $F[x]$ which are congruent modulo $f(x)$ are identified: luckily, such a system is close at hand! We constructed $F[x]/(f(x))$ to be exactly such a system. In other words, if we define a map $\tilde{\phi}$ from $F[x]/(f(x))$ to M by sending $[g(x)]$ to $g(\alpha)$, then ϕ becomes an injective map. Moreover, it still preserves addition and multiplication, because addition and multiplication work the same way in $F[x]/(f(x))$ as in $F[x]$. Moreover, both $F[x]/(f(x))$ and M have p^j elements. So the fact that $\tilde{\phi}$ is an injective map from $F[x]/(f(x))$ to M means that it must also be a surjective (i.e., onto) map.

Thus $\tilde{\phi}$ is a bijective map from $F[x]/(f(x))$ to M preserving all the operations. Hence M is isomorphic to $F[x]/(f(x))$, which is what we set out to prove. $\square$

Algebra enthusiasts will recognize the proof above as an instance of the so-called 'first isomorphism theorem': the map ϕ from $F[x]$ to M is a homomorphism with kernel exactly the ideal $(f(x))$. Thus $F[x]/(f(x))$ is isomorphic to the image of ϕ, which (by size considerations) has to be all of M.