Lecture Notes for Math 696
Coding Theory

Michael E. O'Sullivan
mosulliv@math.sdsu.edu
wwww-rohan.sdsu.edu/ ~ mosulliv

March 4, 2002

1 Introduction

The goal of this course is to cover the fundamental constructions and algorithms that are used in two recent approaches to error correcting codes. The first type of code generalizes Reed-Solomon codes, which are ubiquitous in modern communications systems. Reed-Solomon codes require an understanding of material from modern algebra: e.g. polynomial algebras and finite fields. The recent generalization that we will study is Hermitian codes, which are created using techniques from algebraic geometry. Algebraic geometry is a complex and deep area of mathematics that can take a great deal of effort to learn. But we have very well defined goals and I think that we can learn enough about certain essential ideas, many of which appear in the development of finite fields, to be able to understand Hermitian codes.

The second type of code that we will study is low density parity check codes, which are generating a great deal of excitement among coding theorists. Actually, the codes themselves are not the central issue. Rather, the decoding algorithm is the key. The algorithm uses probabilistic reasoning in a manner similar to algorithms in artificial intelligence and other areas.

This is an applied mathematics course, so the goal is to use certain mathematical results correctly and intelligently. Our goal is not to rigorously develop a mathematical theory. Of course, to correctly and intelligently use mathematical results one needs to have a reasonable grounding in the theory of the subject. My goal with these notes is to make the theoretical development of the material seem reasonable and believable and clear. In particular, I want to show how similar techniques are used to create several algebraic structures. Starting with an algebraic object, we set up an equivalence relation to define a new structure called the quotient modulo the equivalence. We next show that the algebraic structure of the original object carries over nicely to the quotient. We then study the new properties of the quotient structure.

The presentation here is quite rigorous, though I skip the more tedious details. I want the skeleton to be soid but I don't want us to get lost in minutae. You may find the proofs are quite terse. Remember that your goal is to use the results intelligently, not to follow all of the justification for the theory.

The first section on commutative rings should be seen as an appendix read over the course of a couple of weeks.

2 Commutative Rings

I will use the following notation.

- The positive integers: $\mathbb{N}$.
- The nonegative integers: $\mathbb{N}_0$.
- The integers greater than or equal to a: $\mathbb{N}_a$.
- All integers: $\mathbb{Z}$.
- The rationals: $\mathbb{Q}$.
- The real numbers: $\mathbb{R}$.
- The complex numbers: $\mathbb{C}$.
- The finite field of q elements: $\mathbb{F}_q$.
- The integers modulo n: $\mathbb{Z}/n$.

Now to review a few definitions.

Definition 2.1. A *ring* is a set R with two binary operations—addition, $+$, and multiplication, $*$—such that

- both operations are commutative,
- both operations are associative,
- both operations have an identity (0 for $+$ and 1 for $*$),
- there is an additive inverse for each element of R, and
- multiplication distributes over addition.

A *field* is a ring with the extra property that there is a multiplicative inverse for each nonzero element.

Remark 2.2. Strictly speaking, the definition above is for a commutative ring with identity. Since that is a mouthful, and we will only use commutative rings with identity, I opt for a four letter term. Some authors allow fields to have noncommutative multiplication. Others call such a structure a skew field or division ring. We always assume multiplication is commutative.

Definition 2.3. Let R be a ring. An element $a \in R$ is called a *unit* if there is a multiplicative inverse for a. It is called a *zero-divisor* if it is nonzero itself and there is some nonzero $b \in R$ such that $a * b = 0$.

Here are some quick exercises:

1. The zero element, 0, is unique.
2. For any $a \in R$, $0 * a = 0$
3. The unity element, 1, is unique.
4. The inverse of an element (additive or multiplicative) is unique.
5. No element of a ring can be both a zero-divisor and a unit.
6. In $\mathbb{Z}/n$ every element is either 0, a zero-divisor, or a unit.
7. Every nonzero element of a field is a unit.

Most of the time we are interested in rings with no zero-divisors.

Definition 2.4. A ring R is called an *integral domain* if it has no zero-divisors.

The only examples of rings which are not integral domains that you may know are $\mathbb{Z}/n$ for n composite.

Exercises 2.5.
1. Prove the cancellation law for integral domains. If a, b, c are elements of an integral domain with $a \neq 0$ and $ab = ac$ then $b = c$.

Definition 2.6. Let A and R be rings. A *homomorphism* from A to R is a function f that preserves the operations of A and R. That is, for $a, b \in A$, $f(a +_A b) = f(a) +_R f(b)$, $f(1_A) = 1_R$ and $f(a *_A b) = f(a) *_R f(b)$.

I used $+_A$ and $+_R$ to emphasize that the two operations take place in different rings. Normally we won't be so fussy. You might wonder why we didn't require that $f(0_A) = 0_R$. The reason is that you can prove it from the preservation of addition and the existence of inverses (exercise). That doesn't work for 1, since there may not be an inverse for $f(1)$. The more general ring theoretic approach does not require a homormorphism to satisfy $f(1_A) = 1_R$, but from a commutative algebra viewpoint it is natural to do so.

We next want to introduce polynomial rings over a given ring R. This is actually a bit tricky, and you can find an interesting discussion in [1, p.81]. We will give an informal definition and refer to [1, Appendix G] for one way to formally define polynomials. See also [2, pp. 248-]

Definition 2.7. Let R be a ring and let x be a variable. For any $n \in \mathbb{Z}$ and $a_1, \ldots, a_n \in R$ we will call an expression of the form $a_0 + a_1 x + a_2 x^2 + \ldots a_n x^n$ a *polynomial in x over R*. The *polynomial ring in x over R*, denoted $R[x]$ is the set of all such polynomials with addition and multiplication defined as you might expect using the operations in R.

Note that the zero element of $R[x]$ is the polynomial whose coefficients are all 0, and the unity element is the polynomial such that $a_0 = 1$ and all other coefficients are 0.

Some funny things can happen when the ring R has zero divisors. For example in $\mathbb{Z}/4[x]$, the product of $2x$ and $2x$ is 0.

Definition 2.8. For a nonzero polynomial $f(x) = a_0 + a_1x + \cdots + a_nx^n$ over R let d be the largest integer such that $a_i \neq 0$. We will call d the *degree* of $f(x)$, a_dx^d the *leading term*, and a_d the *leading coefficient*. The polynomial is *monic* if the leading coefficient is 1.

Provided that R has no zero divisors, the degree of the product of two polynomials is the sum of the degrees.

It's worthwhile at this point to collect our known examples of rings. The first ring is $\mathbb{Z}$. It is actually first in a very precise sense. There exists exactly one ring homomorphism from $\mathbb{Z}$ to any other ring, R. It is defined by $1_Z \to 1_R$. There is also a "last ring." See if you can find it!

Other examples of rings are $\mathbb{Z}/n$ and the polynomial rings defined from it: $\mathbb{Z}/n[x]$, $\mathbb{Z}/n[x, y]$ etc. We also have the fields $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ and the polynomial rings defined from them. If you took number theory, you may have seen the Gaussian integers, $\mathbb{Z}[i]$, or other quadratic extensions of $\mathbb{Z}$, $\mathbb{Z}[\sqrt{m}]$.

Example 2.9. Here are two rings sandwiched between $\mathbb{Z}$ and $\mathbb{Q}$ that are very important in algebraic geometry and commutative algebra.

Let $\mathbb{Z}_2$ be the set of elements of $\mathbb{Q}$ that can be expressed as a/b with b a power of 2. So $\mathbb{Z}_2 = \{a/2^i : a \in \mathbb{Z}, i \in \mathbb{Z}, \gcd(a, 2^i) = 1\}$. This ring is called the localization of $\mathbb{Z}$ away from 2.

Let $Z_{(2)}$ be the set of elements of Q that can be expressed as a/b with b odd. So $\mathbb{Z}_{(2)} = \{a/b : a \in \mathbb{Z}, b \in \mathbb{N}, b \text{ odd }, \gcd(a, b) = 1\}$. This ring is called the localization of $\mathbb{Z}$ at 2.

The names for these rings will be explained in a later example.

The first step in analyzing many mathematical objects is to understand the elements that can't be broken down effectively.

Definition 2.10. Let R be an integral domain. A non-unit $p \in R$ is called *irreducible* if whenever p can be expressed as a product $p = a * b$, either a or b is a unit. A non-unit $p \in R$ is *prime* if, whenever p divides ab, either p divides a or p divides b.

There is a subtle difference between the two definitions. Note first of all that our usual definition of a prime integer corresponds to the definition for irreducible above, not the one for prime. Fortunately for integers every irreducible is really prime. In fact that is one of the theorems in number theory, to show that any irreducible integer is prime (according to the definitions here).

An important exercise is to show:

Theorem 2.11. *In any integral domain R, any prime element is irreducible.*

The converse is not always true! Examples can be found in $\mathbb{Z}[\sqrt{-5}]$ where $1 + \sqrt{-5}$, 2, and 3 are all irreducible but not prime. Why are they not prime?

Example 2.12. The irreducible elements of $\mathbb{Z}_2$ are the ordinary odd prime integers. For example, 3 is irreducible, for if $3 = (a/2^i) * (b/2^j)$ then $3 = ab/2^{i+j}$. But since a and b are odd (we may assume the original factors were in lowest terms) we must have

$i + j = 0$ so $ab = 3$. then one of a or b is ± 1. But then one of $a/2^i$ or $b/2^j$ is a unit in $\mathbb{Z}_2$.

Since the only primes of $\mathbb{Z}_2$ are the primes of $\mathbb{Z}$ not equal to 2, we say that $\mathbb{Z}_2$ is $\mathbb{Z}$ localized away from 2.

The only irreducible element of $\mathbb{Z}_{(2)}$ is 2. It is clear that the odd primes integers are not irreducible in $\mathbb{Z}_{(2)}$ since they are in fact units! Since 2 is the only prime of $\mathbb{Z}$ that remains prime in $\mathbb{Z}_{(2)}$ we say that $\mathbb{Z}_{(2)}$ is $\mathbb{Z}$ localized at 2.

Exercises 2.13.

1. Define what we mean by $\mathbb{Z}_n$ and by $\mathbb{Z}_{(n)}$ for a positive integer n.
2. Find the irreducibles in $\mathbb{Z}_n$ and $\mathbb{Z}_{(n)}$.
3. Show that any irreducible in $\mathbb{Z}_n$ is an irreducible in $\mathbb{Z}$. Same thing for $\mathbb{Z}_{(n)}$.

3 Review of Properties of the Integers

In this section we will review the main theorems that lead to the unique factorization of integers. We will establish analogous results for polynomials over a field in the next section. I recommend Rosen's Chapter 3 for a more detailed treatment of this material.

Theorem 3.1 (Division). *Given integers a, b with $b > 0$ there exist unique integers q, r such that*

1. *$a = bq + r$, and*
2. *$0 \leq r < b$.*

PROOF: For existence, use well ordering of the integers. Take the smallest nonnegative integer of the form $a - bq$ with $q \in \mathbb{Z}$. For uniqueness, suppose there are two, take the difference and show the two must be equal. □

Theorem 3.2. *Given integers a, b, not both 0, the greatest common divisor of a and b may be written as a linear combination of a and b.*

PROOF: Choose the smallest positive element of $\{ra + sb : r, s \in \mathbb{Z}\}$, call it d. It is clear that any common divisor of a and b divides d. Show that the remainder when either a or b is divided by d must be 0, otherwise the definition of d is violated. Thus d must be $\gcd(a, b)$. □

The preceeding result simply says that the gcd of two integers can be written as a linear combination of the two. The next result—2000 years old—describes how to compute the gcd.

Theorem 3.3. *The Euclidean algorithm may be used to find the greatest common divisor of two integers, and to find the the linear combination that gives the greatest common divisor.*

There is a matrix version of the Euclidean algorithm which is very neat and which is easily modified to produce the Belekamp-Massey algorithm for Reed-Solomon codes.

Definition 3.4. A positive integer p is *irreducible* if for any factorization $p = ab$ either a or b is a ± 1. An integer p is *prime* if whenever p divides ab, p divides either a or b.

In a first course in number theory one normally defines a prime as we have defined an irreducible. The reason for the switch here is to conform with the more modern usage in commutative algebra.

If you look for the following theorem in a number theory book it will say: "Suppose p is prime, then if p divides ab either p divides a or p divides b."

Theorem 3.5. *Let $p \in \mathbb{Z}$ be irreducible. Then p is prime.*

PROOF: Suppose p is irreducible and $p|ab$. If $p \not| a$ then the $\gcd(p, a) = 1$, so there are integers r and s such that $pr + as = 1$. Multiplying by b we have $prb + asb = b$. Now p divides the left hand side, so $p|b$. □

Now we have unique factorization!

Theorem 3.6. *Given any $n \in \mathbb{N}$ there is a unique way to write n as a product of primes p_i such that $p_i \leq p_{i+1}$:*

$$n = \prod_{i=1} p_i$$

PROOF: I will just sketch the proof of uniqueness not existence. Suppose that all integers less than n have unique factorization. Suppose that n has two factorizations $p_1 p_2 \cdots p_r = n = q_1 q_2 \cdots q_s$ with the p_i and q_j in increasing order. By the previous theorem $p_1 | q_j$ for some j. But since q_j is prime, $p_1 = q_j$. Since the q_j are in increasing order, $q_1 \leq p_1$. Similarly one show that $p_1 < q_1$. So in fact $p_1 = q_1$. Thus we have $p_2 p_3 \cdots p_r = q_2 q_3 \cdots q_s$. This integer is smaller than n so is factorable in a unique way. Thus $r = s$ and $p_i = q_i$. □

Another important concept from number theory is modular arithmetic. I will skip the definition of Z/n and the justification for the arithmetic in $\mathbb{Z}/n$. One new bit of terminology: we call $\mathbb{Z}/n$ a *quotient ring*. The main result we need is:

Theorem 3.7. *Let p be a prime number. Then $\mathbb{Z}/p$ is a field.*

PROOF: Let a be an integer whose congruence class is nonzero mod p. Then a is not divisible by p, so in fact it must be coprime to p. By Theorem 3.2 there exist r and s such that $ar + ps = 1$. This means that $ar \equiv 1 \mod p$, so a is a unit mod p. Thus every nonzero element of $\mathbb{Z}/p$ is invertible, and $\mathbb{Z}/p$ is a field. □

Exercises 3.8.

1. Find solution(s) if they exist to $3x + 7 \equiv 5 \mod 7$
2. Find solution(s) if they exist to $3x^2 + 7x + 5 \equiv 0 \mod 7$.
3. Find soultions if they exist to $7x^4 + 3x^2 + x + 15 \equiv 0 \mod 31$. Use Maple!

4 Polynomials over a field

In Section 2 we discuss the basics of commutative rings, define what we mean by a field and introduce polynomial rings. In this section we will show that the polynomial ring

over a field F behaves very much like the integers. In particular, we establish unique factorization and show that $F[x]/p(x)$ is a field when $p(x)$ is irreducible. [The analogy is so strong that I wrote this section by pasting the previous one here and then editing a bit!]

Theorem 4.1 (Division). *Given polynomials $a(x), b(x) \in F[x]$ with $b(x) \neq 0$ there exist unique polynomials $q(x), r(x)$ such that*

1. $a(x) = b(x)q(x) + r(x)$, *and*
2. $\deg r(x) < b(x)$.

PROOF: Existence: Take the smallest degree polynomial of the form $a(x) - b(x)q(x)$ with $q(x) \in F[x]$. (You also have to show that this polynomial has degree $< \deg b(x)$).

Uniqueness: Suppose that $a(x) = b(x)q(x) + r(x)$ with $\deg r(x) < b(x)$. Suppose also that $a(x) = b(x)q'(x) + r'(x)$ with $\deg r'(x) < b(x)$. Taking the difference of the two equations, $0 = b(x)(q(x) - q'(x)) + r(x) - r'(x)$. Then $r'(x) - r(x) = b(x)(q(x) - q'(x))$. Now the degree of the left had side is strictly less than $\deg b(x)$, but the degree of the right hand side is at least $\deg b(x)$ unless it is zero. Therefore both sides must be zero and we have $q(x) = q'(x)$ and $r(x) = r'(x)$. □

Theorem 4.2. *Given polynomials $a(x), b(x) \in F[x]$, the greatest common divisor of $a(x)$ and $b(x)$ may be written as a linear combination of $a(x)$ and $b(x)$.*

PROOF: Choose the smallest degree element of $\{r(x)a(x) + s(x)b(x) : r(x), s(x) \in F[x]\}$, call it $d(x)$. It is clear that any common divisor of $a(x)$ and $b(x)$ divides $d(x)$. Show that the remainder when either $a(x)$ or $b(x)$ is divided by $d(x)$ must be 0, otherwise the definition of $d(x)$ is violated. Thus $d(x)$ must be $\gcd(a(x), b(x))$. □

The preceeding result simply says that the gcd of two polynomials can be written as a linear combination of the two. The next result says that the Euclidean algorithm can be extended to polynomials. You should implement it using Maple!

Theorem 4.3. *The Euclidean algorithm may be used to find the greatest common divisor of two polynomials, and to find the the linear combination that gives the greatest common divisor.*

Definition 4.4. A positive polynomial $p(x)$ is *irreducible* if for any factorization $p = a(x)b(x)$ either $a(x)$ or $b(x)$ is an element of F. A polynomial $p(x)$ is prime if whenever $p(x)$ divides $a(x)b(x)$, p divides either $a(x)$ or $b(x)$.

Theorem 4.5. *Let $p(x) \in F[x]$ be irreducible. Then $p(x)$ is prime.*

PROOF: Suppose $p(x)$ is irreducible and $p(x)|a(x)b(x)$. If $p(x) \not| a(x)$ then the $\gcd(p(x), a(x)) = 1$, so there are polynomials $r(x)$ and $s(x)$ such that $p(x)r(x) + a(x)s(x) = 1$. Multiplying by $b(x)$ we have $p(x)r(x)b(x) + a(x)s(x)b(x) = b(x)$. Now $p(x)$ divides the left hand side, so $p(x)|b(x)$. □

Now we have unique factorization!

Theorem 4.6. *Given any $f(x) \in F[x]$ there exists an element $\alpha \in F$, a nonnegative integer r, and for each $i = 1, \ldots, r$, distinct monic irreducibles $p_i(x)$ and integers $a_i > 0$ such that*

$$f(x) = \alpha \prod_{i=1}^{r} p_i^{a_i}$$

The element α and the integer r are unique and the polynomials and there powers are unique, up to reordering.

PROOF: I will just sketch the proof of uniqueness not existence. Suppose that all polynomials of degree less than n have a unique factorization as described in the statement of the theorem. Suppose that $f(x)$ has degree n and two factorizations. Since irreducibles are prime, any irreducible factor of one side must divide (and therefore be equal to) one of the irreducible factors of the other. Since $F[x]$ is an integral domain, we can cancel this factor and get two factorizations for some polynomial $g(x)$ of degree less than n. by assumption, $g(x)$ satisfies the theorem so the two factorizations must be the same. Therefore the two factorizations of $f(x)$ are the same (up to reordering). □

Since the product of two polynomials over F has degree equal to the sum of the degrees of the factors we get the following.

Proposition 4.7. *Let $f(x)$ have the factorization in the Theorem. Then* $\deg f(x) = \sum_{i=1}^{r} a_i \deg p_i(x)$.

Another consequence of unique factorization is that a polynomial of degree d has at most d roots.

Definition 4.8. Let $f(x) =\in F[x]$ and let $f(x) = \sum_{i=1}^{n} a_i x^i$ with $a_i \in F$. For $\alpha \in F$ we define $f(\alpha)$ to be $\sum_{i=1}^{n} a_i \alpha^i$, which is an element of F. We say that α is a *root* of $f(x)$ if $f(\alpha) = 0$.

From the division theorem we see that finding a root of $f(x)$ corresponds to finding a factor of f.

Proposition 4.9. *Let $f(x) \in F[x]$ and let $\alpha \in F$. Then α is a root of $f(x)$ if and only if $(x - \alpha)$ is a factor of $f(x)$.*

PROOF: Show that the remainder when $f(x)$ is divided by $x - \alpha$ is $f(\alpha)$. The remainder is 0 if and only if $x - \alpha$ divides $f(x)$. □

Proposition 4.10. *If* $\deg f(x) = d$ *then $f(x)$ has at most d distinct roots.*

PROOF: Each root gives a factor of $f(x)$, so n distinct roots give n distinct linear factors of $f(x)$. By unique factorization the product of these roots, which has degree n, divides $f(x)$. Thus $n \leq d$. □

Example 4.11. Suppose that F is the field $\mathbb{F}_p = \mathbb{Z}/p$. Fermat's little theorem says that $1, 2, 3 \ldots, p-1$ are all roots of $x^{p-1} - 1$. Therefore the unique factorization of $x^{p-1} - 1$ is $(x-1)(x-2)\cdots(x-(p-1))$.

I'm sure you are well aware that

$$(y-1)(y^{e-1}+y^{e-2}+\cdots+y^2+y+1)=y^e-1$$

Substitute x^d for y and we get

$$(x^d-1)(x^{d(e-1)}+x^{d(e-2)}+\cdots+x^2d+x^d+1)=x^{de}-1$$

In the following result we use this factorization of polynomials in the case where $p-1=de$. The result will be useful in our study of finite fields.

Proposition 4.12. *Let p be prime. If d divides $p-1$ then x^d-1 has d distinct roots in $\mathbb{F}_p$.*

PROOF: Let $e=n/d$. From the example above we know that $x^{p-1}-1$ has $p-1$ distinct roots. From the factorization $(x^d-1)(x^{d(e-1)}+x^{d(e-2)}+\cdots+x^2d+x^d+1)=x^{p-1}-1$, and the fact that $(x^{d(e-1)}+x^{d(e-2)}+\cdots+x^2d+x^d+1)$ has at most $d(e-1)$ roots, we know that x^d-1 must have at least $p-1-d(e-1)=d$ distinct roots. Of course it can have no more than that. □

We now extend the concept of modular arithmetic to polynomial rings. We will only work modulo a prime polynomial. You will soon see why.

The first thing to do is to establish the equivalence relation to define the quotient ring $F[x]/p(x)$. Next we show that arithmetic is well defined. I will just state the results. The proofs are exact analogues of those for integers.

Definition 4.13. Let $p(x)\in F[x]$ with F a field. We say $f(x),g(x)$ are congruent modulo $p(x)$, written $f(x)\equiv g(x) \mod p(x)$, if $p(x)$ divides $f(x)-g(x)$.

Proposition 4.14. *Congruence modulo $p(x)$ is an equivalence relation.*

If two unequal polynomials have degree less than $\deg p(x)$ then they cannot be congruent mod $p(x)$. Clearly, every polynomial is equivalent to its remainder when divided by $p(x)$, and the remainder has degree less than $\deg p(x)$. Thus the set of polynomials of degree less than $p(x)$ form a complete system of residues mod $p(x)$.

Proposition 4.15. *If $f(x)\equiv g(x) \mod p(x)$ and $a(x)\equiv b(x) \mod p(x)$ then*

- $a(x)+f(x)\equiv b(x)+g(x) \mod p(x)$.
- $a(x)f(x)\equiv b(x)g(x) \mod p(x)$.

Suppose that $p(x)$ has degree d. Then the discussion above shows that a complete system of representatives for $F[x]/p(x)$ is $\{a_0+a_1x+\cdots a_{d-1}x^{d-1} : a_i\in F\}$. The addition of these polynomials is done component-wise. Multiplication by an element of F is also component-wise. Consequently, $F[x]/p(x)$ is a vector space of dimension d over F. Our next theorem concerns the multiplicative structure.

Theorem 4.16. *Let $p(x)$ be an irreducible polynomial over a field F. Then $F[x]/p(x)$ is a field.*

PROOF: Let $a(x)$ be a polynomial whose congruence class is nonzero mod $p(x)$. Then $a(x)$ is not divisible by $p(x)$, so in fact it must be coprime to $p(x)$. Then by Theorem 4.2 there exist $r(x)$ and $s(x)$ such that $a(x)r(x)+p(x)s(x) = 1$. This means that $a(x)r(x) \equiv 1$ mod $p(x)$. Consequently $a(x)$ is a unit mod $p(x)$, and $F[x]/p(x)$ is a field. □

Exercises 4.17.

1. What is $\mathbb{R}[x]/(x^2+1)$? Can you find any other extensions of the reals?

2. What is $\mathbb{Q}[x]/(x^2+1)$. Construct other extensions of Q of degree 2. Give an example of an extension of degree 3. Does x^3+1 work?

3. Find an irreducible polynomial of degree 2, one of degree 3, and one of degree 4 over $\mathbb{F}_2$. For each, find a system of representatives for $\mathbb{F}_2[x]/p(x)$. Find a polynomial of degree 4 over $\mathbb{F}_2$ which has no roots but is not irreducible.

5 Finite Fields

The last section ended with the result that for a field F and irreducible polynomial $p(x)$ over F, $F[x]/p(x)$ is a field. We are interested in the case when F is $\mathbb{F}_p = \mathbb{Z}/p$. We will use $P(x)$ for an irreducible polynomial to eliminate confusion with the prime p.

The main result of this section is the following theorem. We will prove it in stages.

Theorem 5.1. *Let F be a field with a finite number of elements.*

1. *F has p^n elements where p is a prime.*
2. *F is isomorphic to $\mathbb{F}_p[x]/P(x)$ for some irreducible polynomial $P(x)$ over $\mathbb{F}_p$.*

For any prime p and any positive integer n,

1. *There exists a field with p^n elements.*
2. *Any two fields with p^n elements are isomorphic.*

As a first step we prove

Proposition 5.2. *A finite field is a vector space over $\mathbb{F}_p$ for some prime p Consequently, the number of elements of F is a power of p.*

PROOF: Suppose that F is a finite field. Consider the additive subgroup generated by 1, i.e. 1, $1+1$, $1+1+1$. Let m be the smallest positive integer such that the sum of m 1's is 0. If m where composite, $m = ab$, then we would have

$$0 = \underbrace{1+1+1+\cdots+1+1}_{m \text{ terms}} = (\underbrace{1+1+\cdots+1}_{a \text{ terms}})(\underbrace{1+1+\cdots+1}_{b \text{ terms}})$$

The two factors on the right would then be zero-divisors, contradicting the assumption that F is a field. Thus m is in fact a prime, which we will now call p. We may think of F as containing a copy of $\mathbb{F}_p$. From the field axioms we see immediately that F is a vector space over p. If its dimension over $\mathbb{F}_p$ is n then F must have p^n elements. Thus any field has a prime power number of elements. □

The prime p is called the *characteristic* of the field. Now we have the Freshman's dream:

Proposition 5.3. *Let α, β be elements of a field of characteristic p. Then $(\alpha+\beta)^p = \alpha^p + \beta^p$.*

PROOF: Expand $(\alpha+\beta)^p$ using the binomial theorem and we get terms like

$$\binom{p}{k}\alpha^k\beta^{p-k}$$

The binomial coefficient really means 1 added to itself $\binom{p}{k}$ times. Since p divides the binomial coefficient when $1 < k < p$ the coefficient is 0 unless $k = 1$ or $k = p$. That gives the result. □

Suppose that $q = p^n$ is the number of elements in F By the field axioms, the set of nonzero elements of F is a group under multiplication. There are $q-1$ nonzero elements, so LaGrange's theorem says that for any $\alpha \in F$, $\alpha^{q-1} = 1$ in F. In other words, α is a root of the polynomial $x^{q-1} - 1$. We know that this polynomial can have at most $q-1$ distinct roots. But, we have just shown that all the $q-1$ nonzero elements of F are roots. Therefore $x^{q-1} - 1$ factors into distinct linear factors over F. We have therefore established the following proposition.

Definition 5.4. A *splitting field* for a polynomial $f(x)$ defined over a field F is an extension field of F in which the $f(x)$ factors into linear factors and the field is generated by the roots of the linear factors.

Proposition 5.5. *A field of $q = p^n$ elements is a splitting field of $x^{q-1} - 1$ over $\mathbb{F}_p$.*

We can now justify the final two items of the main theorem. One can show fairly easily that a splitting field always exists. Thus a splitting field for $x^{q-1} - 1$ exists. In fact, the splitting field has q elements, just 0 and the $q-1$ roots of $x^{q-1} - 1$. One can show this by showing that the product of two roots of $x^{q-1} - 1$ is a another root, and, using the Freshman's dream, the sum of two roots is also a root. Thus for any $q = p^n$ there exists a finite field with q elements.

It is a bit harder, but one can also show that a splitting field for a polynomial is unique up to isomorphism. This means that any two finite fields with the same number of elements are isomorphic. Henceforth we will use the symbol $\mathbb{F}_q$ to denote this unique field. For the details on splitting fields I refer you to [1, §10.4, 6], [2, Ch. 20], [3, VII.3].

It remains to justify the statement that a field with $q = p^n$ elements is isomorphic to $F_p[x]/P(x)$ for some irreducible polynomial. Let me make the problem more precise, by giving an example.

Example 5.6. Suppose that $n = 6$. Perhaps our field $\mathbb{F}_{p^6}$ is obtained by taking an extension of degree 2, and then following it by an extension of degree 3. In other words, $\mathbb{F}_{p^2} = F_p[x]/P(x)$ where $P(x)$ is quadratic and $\mathbb{F}_{p^6} = \mathbb{F}_{p^2}[y]/Q(y)$ where $Q(y)$ is cubic with coefficients in $\mathbb{F}_{p^2}$. This is in fact the case; $\mathbb{F}_{p^6}$ can be constructed this way. BUT, it is also true that there exists an element $\alpha \in \mathbb{F}_{p^6}$ such that $1, \alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5$ are all linearly independent over $\mathbb{F}_p$. Remember that $\mathbb{F}_{p^6}$ is a vector space of dimension 6 over $\mathbb{F}_p$, so $1, \alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5$ is therefore a basis for $\mathbb{F}_{p^6}$. That implies that α^6 is a linear combination of $1, \alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5$. This linear combination then gives a polynomial $R(x)$ of degree 6 for which α is a root. Furthermore, $R(x)$ must be irreducible, for

otherwise α would be a root of one of the factors, and that would mean that some linear combination of $1, \alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5$ would be zero, contradicting what I stated above. Thus $\mathbb{F}_{p^6} = \mathbb{F}_p[x]/R(x)$.

We want to show that every finite field $\mathbb{F}_{p^n}$ has an element α such that $1, \alpha, \ldots, \alpha^{n-1}$ are linearly independent over $\mathbb{F}_p$. This implies that α is a root of a degree n polynomial $R(x)$ with coefficients in $\mathbb{F}_p$ and therefore $\mathbb{F}_{p^n} = \mathbb{F}_p[x]/R(x)$. We will prove this as a corollary of Theorem 5.9.

Proposition 5.7. *Let F be a finite field with q elements. If d divides $q-1$ then $x^d - 1$ has d distinct roots in F.*

PROOF: The proof is identical to the proof of Proposition 4.12

□

We now show that the multiplicative group of a finite field is cyclic.

Definition 5.8. Let F be a finite field with q elements. For $\alpha \in F$, the *order* of α is the smallest positive integer r such that $\alpha^r = 1$ (its order in the multiplicative group of F). We say α is *primitive* if $\operatorname{ord} \alpha = q - 1$.

Theorem 5.9. *Let $\mathbb{F}_q$ be the finite field with q elements. Then for each d dividing $q-1$, $\mathbb{F}_q$ has $\phi(d)$ elements of order d. In particular, $\mathbb{F}_q$ has $\phi(q-1)$ primitive elements. Consequently, the multiplicative group of $\mathbb{F}_q$ is cyclic.*

PROOF: Let $g(d)$ be the number of elements of order d in F. By LaGrange's Theorem, every element of F has order dividing $q-1$, so $g(d)$ is zero for d not a divisor of q. One can readily show that if α has order d then α^k for $k = 1 \ldots d$ are all distinct, and they must be all the roots of $x^d - 1$ in F. Furthermore, $\operatorname{ord} \alpha^k = d/\gcd(d,k)$. Therefore if F has an element of order d then it has $\phi(d)$ of them, namely α^k for $\gcd(k,d) = 1$. Summarizing, $g(d)$ is either 0 or $\phi(d)$.

Now every element of F has some order dividing $q-1$ so

$$q - 1 = \sum_{d|q-1} g(d) \leq \sum_{d|q-1} \phi(d)$$

There is a result in number theory [4, Theorem 7.7] that $\sum_{d|n} \phi(n) = n$. This tells us that the $\leq$ in the equation above must be an equality. the only way this can hold is if $g(d) = \phi(d)$. Which is want we wanted to prove. □

Corollary 5.10. *For $q = p^n$, $\mathbb{F}_q$ is isomorphic to $\mathbb{F}_p[x]/P(x)$ for some irreducible polynomial of degree n.*

PROOF: Let α be a primitive element of $\mathbb{F}_q$. Let $P(x)$ be the minimal polynomial for α over $\mathbb{F}_p$. If $\deg P(x) = r$ then the field $\mathbb{F}_p[x]/P(x)$ has p^r elements and is isomorphic to the the subfield of $\mathbb{F}_q$ generated by α. But α generates all of $\mathbb{F}_q$, so $q = p^r$ and therefore $r = n$. □

Exercises 5.11.

1. Construct $\mathbb{F}_{2^6}$ in three ways:

a) by constructing $\mathbb{F}_4$ using an irreducible polynomial of degree 2 over $\mathbb{F}_2$ and then by constructing $\mathbb{F}_6$ using an irreducbile polynomial of degree 3 over $\mathbb{F}_4$.

b) by constructing $\mathbb{F}_8$ using an irreducible polynomial of degree 3 over $\mathbb{F}_2$ and then by constructing $\mathbb{F}_6$ using an irreducbile polynomial of degree 2 over $\mathbb{F}_8$.

c) by using an irreducible polynomial of degree 6 over $\mathbb{F}_2$.

Show that the three representation are isomorphic.

2. Construct $\mathbb{F}_{81}$ in two ways (why just two?). You may notice that x^2+2x+1 and x^2+x+2 are both irreducible over $\mathbb{F}_3$. Can you construct $\mathbb{F}_{81}$ by using one of these polynomials and then the other?

3. Prove that if $r|n$ then $\mathbb{F}_{q^r}$ is a subfield of $\mathbb{F}_{q^n}$.

4. Factor $x^{15}-1$ over $\mathbb{F}_2$. Construct $\mathbb{F}_{16}$ in three ways as a degree 4 extension of $\mathbb{F}_2$ and show isomorphisms between the three representations.

References

[1] T. W. Hungerford, *Abstract Algebra: An Introduction*, 2nd ed., Saunders Harcourt Publishing, 1997.

[2] J. A. Gallian, *Contemporary Abstract Algebra* 3rd ed., D. C. Heath and Co., 1994.

[3] S. Lang, *Algebra* 2nd ed., Addison-Wesley, 1984.

[4] K. Rosen, *Elementary Number Theory and Its Applications* 4th ed. Addison Wesley Longman, 2000.