Codes and Fields

Part 2
Finite Fields

Oliver Pretzel

Chapter 0

Introduction
The need for fields other than $\mathbb{B}$

1. Constructing codes for correcting multiple errors.

We have seen that extending the check matrix of a linear code in a linear manner does not change the code. But it is obvious that if we wish to correct more than one error we must add checks (or start completely afresh). To add non-linear checks in a structured way we need 'good' non-linear functions. Powers in a finite field are particularly promising. This is the idea behind the most frequently used codes: the BCH-codes. But in $\mathbb{B}$ there are no non-trivial powers.

2. Correcting error bursts

In many situations the assumption that errors occur entirely independently of each other is not valid. On many storage devices faults when they occur are likely to affect several neighbouring bits. Such errors are called error bursts. One approach to designing codes for correcting error bursts is to collect the signal bits together into blocks. If we can give these blocks a suitable field structure they can be viewed as the letters of the alphabet of the code. A single error correcting code over that field will then correct any burst lying inside a block, and a double error correcting code will correct any burst of length not greater than a block (because such a burst is covered by at most two adjacent blocks). Applying this idea to BCH-codes yields the frequently used RS-codes.

3. Finding new codes

The structure of the base field controls and limits the possible codes. Allowing a larger range of base fields may yield new codes with special properties. For instance there is a 2–perfect ternary (11,6)–code, also discovered by Golay. This code cannot be constructed over $\mathbb{B}$.

Chapter 2.1

Constructing Finite Fields I
An Example

In this section we shall try and construct a field of 16 elements so that we can manipulate blocks of four bits as single digits. Because human beings find strings of 0's and 1's difficult to distinguish, we shall denote each string by the integer it represents, but this does not mean it should be manipulated like that integer. Thus (1,1,0,1) will be denoted by 13.

1. Try the Integers Modulo 16.

Here we do consider the symbol 13 to represent the number 13. We do ordinary arithmetic, except whenever the result is larger than 15 we subtract off 16's until it lies in the range 0 to 15 (we can also do this by division with remainder). It is not difficult to check (see Exercise Sheet 2) that this makes the set into a commutative ring, which is denoted $\mathbb{Z}/16$ ('zed mod 16'). $\mathbb{Z}$ is the standard symbol for the integers.

Unfortunately the cancellation law fails because $4 \times 4 = 0$. So this structure is not even a domain. Note that the failure is due to the fact that 16 is not a prime number. The techniques we shall develop will enable us to show that if p is a prime number then $\mathbb{Z}/p$ is a field (there is an example on Exercise Sheet 5), but then p is not a power of 2, so this construction is no use for our purposes.

It has a further weakness. The addition does not correspond to the one we have used for our codewords because for example $4 + 4 = 8$, whereas for codewords $u + u = \underline{0}$. We use this plus the idea of taking remainders as the starting point to find an alternative construction.

2. Polynomials over $\mathbb{B}$.

To begin with we need a structure that looks a bit like $\mathbb{Z}$ but uses XOR as its addition. Such a structure is the set of polynomials $a_0 + a_1x + \cdots a_nx^n$ in a 'variable' x with coefficients in $\mathbb{B}$. It is easy to check that these form a commutative ring with the

usual addition and multiplication (see Exercise Sheet 5). And in fact the cancellation law holds, because if we multiply two polynomials the degrees add.

To be a bit more specific if the highest coeficients a_n of $a_0 + a_1x + \cdots + a_nx^n$ and b_m of $b_0 + b_1x + \cdots + b_mx^m$ are both 1, then the highest coefficient of the product is a_nb_m is also 1, so the product is not $\underline{0}$.

3. Now we use $a_0 + a_1x + \cdots + a_nx^n$ to represent the word $(a_n, \ldots, a_1, a_0)$. So the words we are interested in are represented by polynomials of degree ≤ 3. Addition is fine but multiplication could increase the degree, so we copy the idea of $\mathbb{Z}/16$ and divide by a suitable polynomial and take remainders.

But in contrast to $\mathbb{Z}$, there is a choice for polynomials: any polynomial of degree 4 has as its set of remainders the polynomials of degree ≤ 3. For the same reason as with $\mathbb{Z}/16$ we cannot take a polynomial of degree 4 that can be split into the product of two polynomials of smaller degree. So we look for polynomials which do not split. These are called *irreducible*. Trial and error will tell us that there are three choices. One of these is $1 + x^3 + x^4$. If we use this we get the table on the next page which is supposed to be self-explanatory.

This table contains several mysterious nice properties, one of which is the existence of a kind of logarithms. In this part we shall sketch a theory, based on the ideas of this example,which allows us to construct all possible finite fields and develop their beautiful properties. We shall develop computational techniques which will enable us to construct and decode multiple error correcting codes efficiently.

GF(16) based on $x^4 + x^3 + 1$

The number n represents its binary 4-tuple a, b, c, d, which in turn represents the polynomial $ax^3 + bx^2 + cx + d$. Thus $13 \sim 1, 1, 0, 1 \sim x^3 + x^2 + 1$. Addition (the lower half of the table) is ordinary addition over $\mathbb{B}$ (i.e. XOR). In particular $\alpha + \alpha = 0$ for any α, so this does not appear in the table. Multiplication (the upper half of the table) is multiplication in $\mathbb{B}[x]$, followed if necessary, by taking the remainder after division by $x^4 + x^3 + 1$. Every element has an inverse but how do you find it?

For later use we note that every element is a power of $2 \sim x$, and use this to provide 'logarithms'. Using the fact that $2^{15} = 1$ (called Fermat's Little Theorem) we can multiply easily and also find inverses easily. Does such a technique exist for other finite fields? How do you find the 'base' for the logarithms?

log	0	1	12	2	9	13	7	3	4	10	5	14	11	8	6
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	3	4	6	8	10	12	14	9	11	13	15	1	3	5	7
3	2	1	5	12	15	10	9	1	2	7	4	13	14	11	8
4	5	6	7	9	13	1	5	11	15	3	7	2	6	10	14
5	4	7	6	1	8	7	2	3	6	9	12	14	11	4	1
6	7	4	5	2	3	13	11	2	4	14	8	3	5	15	9
7	6	5	4	3	2	1	12	10	13	4	3	15	8	1	6
8	9	10	11	12	13	14	15	15	7	6	14	4	12	13	5
9	8	11	10	13	12	15	14	1	14	12	5	8	1	3	10
10	11	8	9	14	15	12	13	2	3	11	1	5	15	8	2
11	10	9	8	15	14	13	12	3	2	1	10	9	2	6	13
12	13	14	15	8	9	10	11	4	5	6	7	6	10	7	11
13	12	15	14	9	8	11	10	5	4	7	6	1	7	9	4
14	15	12	13	10	11	8	9	6	7	4	5	2	3	2	12
15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	3

Below diagonal $a + b$, on or above $a \times b$,
$0 + a = a$, $a + a = 0$, $0 \times a = 0$

Chapter 2.2

Constructing Finite fields II Euclid's Algorithm

The main technique we shall use to construct finite fields and also to decode BCH codes is *Euclid's Algorithm.* This was invented about 2000 years ago to find highest common factors of integers without first splitting them into their prime factors. We shall apply it mainly to polynomials but it is also useful for integers. The basic tool we need is division with remainder. The first task is to limit the size of the remainder in some way.

1. In this chapter D will denote *either* the set of integers $\mathbb{Z}$ *or* the set of polynomials $F[x]$ in a variable x over a finite field F such as $\mathbb{B}$. In both these setups we can perform *divison with remainder.* For every a and $b \neq 0$ in D there exist a *quotient* q and *remainder* r in D such that $a = qb + r$, and $r = 0$ or r is 'smaller' than b.

For $\mathbb{Z}$ we measure size by the modulus so r smaller than b means $|r| < |b|$, which also works for $r = 0$. For polynomials we measure size by the degree so r smaller than b means $\deg r < \deg b$.

For the integers q and r are not unique: eg. $25 = 3 \cdot 7 + 4 = 4 \cdot 7 + (-3)$. For polynomials q and r are unique.

Domains in which this type of division with remainder is possible are called *Euclidean Domains.*

We say b *divides* a if the remainder r is 0. We denote this by $b \mid a$. Elements with an inverse divide everything. They are called *invertible.* For $\mathbb{Z}$ they are the numbers ± 1 for polynomials they are the non-zero constants.

2. DEFINITION. In D as above we say d is a *Highest Common Factor* (US terminology: Greatest Common Divisor) of $a \neq 0$ and $b \neq 0$ denoted by (a, b) if:
HCF1. $d \mid a$ and $d \mid b$, that is d divides both a and b, and
HCF2. if $c \mid a$ and $c \mid b$ then $c \mid d$, that is any other common divisor divides d.

Highest common factors of a and b are only determined up to multiplication by invertible elements (± 1 in $\mathbb{Z}$ any non-zero constant for polynomials). We deal with

this by 'normalising' the HCF to be positive in $\mathbb{Z}$ and to have highest coefficient 1 for polynomials. However, Euclid's algorithm does not do this normalisation automatically.

We say a and b are relatively prime if $1 = (a, b)$. This means that any element dividing both a and b has an inverse.

3. Euclid's Algorithm

We wish to calculate the highest common factor of a and b. To do this we produce a table with four columns headed Q, R, U, V. Q and R stand for quotient and remainder. We number the rows of the table starting with -1. Each row is calculated from its two predecessors.

ALGORITHM Euclid's Algorithm.

Step 1. Set up a table with 4 columns (5 if you count the row number) headed Q, R, U, V. Fill in the first two rows (numbers -1 and 0) as follows:

ROW	Q	R	U	V
-1	–	a	1	0
0	–	b	0	1

Step 2. Calculate the Q and R entries for row 1 by dividing a by b: $a = q_1 b + r_1$. The U entry is 1 and the V entry is $-q_1$.

1	q_1	r_1	$u_1 = 1$	$v_1 = -q_1$

Note that $r_1 = 1a + (-q_1)b$.

Step 3. Suppose we have calculated up to row k and the last two rows are as follows

$k-1$	q_{k-1}	r_{k-1}	u_{k-1}	v_{k-1}
k	q_k	r_k	u_k	v_k

If $r_k = 0$ *Stop.*

Otherwise divide r_{k-1} by r_k: $r_{k-1} = q_{k+1} r_k + r_{k+1}$.

That gives the Q and R entries of row $k+1$.

Using the Q entry just calculated, put $u_{k+1} = u_{k-1} - q_{k+1} u_k$ and $v_{k+1} = v_{k-1} - q_{k+1} v_k$.

EXAMPLE. Calculate the HCF of 104 and 12.

ROW	Q	R	U	V
–1	–	104	1	0
0	–	12	0	1
1	8	8	1	–8
2	1	4	–1	9
3	2	0	3	–26

Note that 4 is the highest common factor of 12 and 104 and also that $4 = -1 \times 104 + 9 \times 12$. Before we go on let's note in passing that the cancelled form of 12/104 is 3/26. That is no accident, but we will not prove it.

4. PROPOSITION. (a) *The algorithm stops after a finite number of steps. The last non-zero element of the R-column is a highest common factor of a and b.*

(b) *For any k, $r_k = u_k a + v_k b$.*

Proof. (a) Let $\varphi(n) = |n|$ for integers and $\varphi(f(x)) = \deg f$ for polynomials. At each step the value $(\varphi(r_k)$ decreases by at least one. So the number of steps is at most $\varphi(b) + 1$.

Let the last non-zero element be r_n. Then $r_{n-1} = q_{n+1} r_n + 0$. So $r_n \mid r_{n-1}$. Next, $r_{n-2} = q_n r_{n-1} + r_n$. Since r_n divides both summands on the right hand side it divides r_{n-2}. Now suppose we have shown r_n divides r_{k+1} and r_k. As $r_{k-1} = q_{k+1} r_k + r_{k+1}$, it follows in the same way that r_n divides r_{k-1}. Finally, r_n divides $r_0 = b$ and $r_{-1} = a$. So r_n is a common factor of a and b.

Conversely, it will follow from (b) that if $c \mid a$ and $c \mid b$, then $c \mid r_n = u_n a + v_n b$. So r_n is indeed (a, b).

(b) It is clear that the statement is true for $k = -1, 0$. Suppose it is true for r_{k-1} and r_k. We shall show it is true for r_{k+1}.

$$\begin{aligned} r_{k+1} &= r_{k-1} - q_{k+1} r_k \\ &= u_{k-1} a + v_{k-1} b - q_{k+1}(u_k a + v_k b) \\ &= (u_{k-1} - q_{k+1} u_k) a + (v_{k-1} - q_{k+1} v_k) b \\ &= u_{k+1} a + v_{k+1} b. \end{aligned}$$

□

5. CONSTRUCTION. Let $D = \mathbb{Z}$ or $D = F[x]$ for a field f, and let $a \in D$. We denote by D/a ('D modulo a') the set of remainders of elements of D when divided by a. If $D = \mathbb{Z}$ we 'normalise' the remainders by choosing them to be positive. We add and multiply as in D except that, if necessary, we divide the result by a to obtain the remainder. It is easy to check that the result is a commutative ring. The hardest axioms to check are the associative laws:

A1: Let $x + y + z = q_0 a + r_0$ and $x + y = q_1 a + r_1$ and $r_1 + z = q_2 a + r_2$. Then $q_0 a + r_0 = (q_1 + q_2) a + r_2$. Since we have made remainders unique, $r_0 = r_2$. The same argument adding $y + z$ first proves the associative law in D/a.

M1: Let $xyz = q_0 a + r_0$ and $xy = q_1 a + r_1$ and $r_1 z = q_2 a + r_2$. Then $q_0 a + r_0 = (q_1 z + q_2) a + r_2$. Again this implies $r_0 = r_2$. The same argument multiplying yz out first proves the associative law in D/a.

If $a = 0$, then the set of remainders is the whole of D, and if a has an inverse $a^{-1} = b$, then $x = xba + 0$ for all $x \in D$. So there is only one remainder: 0. We exclude these two 'trivial' cases.

6. DEFINITION. We call $a \neq 0$ *irreducible* in D if it does not have an inverse and whenever $a = xy$ with $x, y \in D$, then x or y has an inverse in D.

EXAMPLE. For $\mathbb{Z}$ irreducibles are just the ordinary prime numbers $2, 3, 5, 7, 11, \ldots$.

The key property of an irreducible element a is that if x is any element of D, then either a divides x or (a, x) is invertible. To see this consider $d = (a, x)$. We have $a = de$ and $x = df$. If d is not invertible, then e is. So $x = ae^{-1}f$.

THEOREM. *Let a be a non-zero element of D without an inverse.*

(a) *If a is not irreducible, then D/a does not satisfy the Cancellation Law.*

(b) *If a is irreducible, then D/a is a field.*

EXAMPLES. This tells us that Z/n is a field iff n is prime. Z/n has exactly n elements $\{0, 1, \ldots, n-1\}$. So we get fields with any prime number of elements. For $n = 2$, we get $\mathbb{B}$, and for $n = 3$ we get the ternary field. That does not help us find fields with 2^n elements, but using polynomials over $\mathbb{B}$ does, because the set of remainders of a polynomial of degree n in $\mathbb{B}[x]$ is just the set of polynomials of degree $\leq n - 1$. That set has exactly 2^n elements. Of course, we still have the difficulty of finding irreducible polynomials. But to check that, say, GF(16) is a field, all we need to do is to check that $x^4 + x^3 + 1$ is irreducible. We do that below with other sample calculations derived from the proof of the theorem.

Proof. (a) If a is not irreducible we can find x and y in D such that $xy = a$ and neither x nor y is invertible. Let $x = qa + x'$ and $y = q'a + y'$. If $x' = 0$, then $qay = a$, or $a(qy - 1) = 0$. As the Cancellation Law holds in D, $qy = 1$ and y has inverse q. This contradicts our hypothesis, so $x' \neq 0$. Similarly $y' \neq 0$.

What is the remainder of $x'y'$ when divided by a? Well, $x' = x - qa$ and $y' = y - q'a$, so

$$x'y' = xy - (q + q' - qq'a)a = (1 - q - q' + qq'a)a.$$

Thus the remainder is 0. Hence x' and y' violate the Cancellation Law in D/a.

(b) Let $x \neq 0$ be an element of D/a. We must find y in D/a so that xy leaves remainder 1 when divided by a. To do this we deduce from Euclid's algorithm that $1 = ua + vx$ for some u and v in D. Then we show that if y is the remainder of v when divided by a, then $xy = 1$ in D/a.

Step 1. As x is a remainder we must have $\varphi(x) < \varphi(a)$. So a does not divide x. Hence $d = (a, x)$ as calculated by Euclid's algorithm is invertible.

Step 2. Let d be as above and $d = r_n$ in the calculation for Euclid's Algorithm. Then $d = r_n = u_n a + v_n x$. We know that d has an inverse, say b, in D. Thus

$$1 = db = r_n b = u_n ab + v_n bx.$$

Let $u = u_n b$ and $v = v_n b$. Then $1 = ua + vx$.

Step 3. Keeping the notation of Step 2, let $v = qa + y$. Then $y \in D/a$ and

$$xy = vx - qax = 1 - ua - qa = -(u + qx)a + 1.$$

So we have found our inverse.

7. EXAMPLE. $D = \mathbb{Z}$, $a = 787, x = 53$.

ROW	Q	R	U	V
-1	-	787	1	0
0	-	53	0	1
1	14	45	1	-14
2	1	8	-1	15
3	5	5	6	-89
4	1	3	-7	104
5	1	2	13	-193
6	1	1	-20	297
7	2	0	53	-787

So according to the algorithm $297 \times 53 = 1 + 20 \times 787$, which is true. In other words, calculating 'modulo' 787 the inverse of 53 is 297.

8. To show that our construction of GF(16) works we must only show that $x^4 + x^3 + 1$ is irreducible.

Does it have a factor of degree 1? The only ones are x and $x + 1$. Clearly x does not divide it. And $x + 1$ divides $x^4 + 1$, so if it divided $1 + x^3 + x^4$ it would have divide x^3.

Could it be the product of two polynomials of degree 2? These are x^2, $x^2 + x$, $x^2 + 1$, and $x^2 + x + 1$. x divides the first two but not $1 + x^3 + x^4$ so these are out. $(x + 1)^2 = x^2 + x + x + 1 = x^2 + 1$, so the third is out. So we are left with the possibility of the last. The second factor would also have degree 2 so it would also have to be that same polynomial. But

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

As $1 + x^3 + x^4$ has no factors of degree 1 or 2, it is irreducible.

Let us calculate the inverse of $9 = 1001 = 1 + x^3$, by the method above.

ROW	Q	R	U	V
-1	-	11001	00001	00000
0	-	01001	00000	00001
-	10	01011	00001	00010
1	01	00010	00001	00011
2	100	00001	00100	01101
3	10	00000	01001	11001

Thus the inverse is $1101 = 13$ which agrees with our table.

The inserted line with – as row number is an intermediate step in the long division of r_{-1} by r_1. In general, if we have to do long division to divide the R entry one row (the target) by another (the divisor), we just use left shifts and match the highest coefficients of the R entries. We apply this operation to 1 and place the result in the Q column, then we apply it to the R, U, V entries subtract from the entries of the target row. That corresponds to one step of long division. We give the new row a – sign as row number. Then we continue using this row instead of the target row. When the degree of in the R column drops below the degree of the divisor r, that row gets the next number $k+1$ and we continue as though the '–' rows did not exist[1].

Note that in both examples the last entries of U and V are x and a respectively. That is always the case when a is irreducible and forms a useful calculation check.

[1]The correct value of q is obtained by adding all the q–values in the – rows to the value in row $k+1$.

Chapter 2.3

The Structure of Finite Fields

The proofs of the results of this section are not needed for the actual manipulation of codes. So we state the results first and then present the proofs for the mathematically inclined.

0. Results

THEOREM A. *Any finite field F contains a copy of $\mathbb{Z}/p$, where p is a prime number, called the prime field of F.*

The prime p is called the *characteristic* of the field.

The case we shall be concerned with is $p = 2$. In that case $\mathbb{Z}/2 = \mathbb{B}$ is the binary field. The reader should verify that all the statements we make hold for GF(16) with $p = 2$.

THEOREM B. *If F has characteristic p, then it has exactly p^n elements for some positive index n.*

So the characteristic is unique.

THEOREM C. *If F has characteristic p, then for $a, b \in F$, $(a + b)^p = a^p + b^p$.*

The map $a \mapsto a^p$ is called the *Frobenius Automorphism* of F.

THEOREM D. *If F has $p^n = q$ elements, then for any element $a \in F$, $a^q = a$.*

This is called *Fermat's Little Theorem.*

1. DEFINITION. Let F be any field, $e \in F$, and let n be a positive integer. We define the element $n.e$ to be obtained by adding the element e to itself n times, $0.e = 0$, and $(-n).e = -(n.e)$.

PROPOSITION. *For $m, n \in \mathbb{Z}$, and 1 the identity element of F, $(m + n) \cdot 1 = m \cdot 1 + n \cdot 1$ and $(mn) \cdot 1 = (m \cdot 1)(n \cdot 1)$.*

Proof. Notice that there is something to prove. $(m+n).1$ means 'first add the two integers m and n and then take the appropriate multiple of 1', whereas $m.1 + n.1$ means 'take appropriate multiples of 1 and then add them in F'.

The most confusing thing about the proof is the notation. We are just generalising the fact that the laws of arithmetic imply that

$$(1+1+1)+(1+1) = (1+1+1+1+1)$$
$$\text{and } (1+1+1)(1+1) = (1+1+1+1+1+1).$$

These are the statements for $m = 3$ and $n + 2$.

We shall assume in the proof that m and n are positive. The other statements follow by manipulating the signs of the terms. We write the terms $m \cdot 1$ and $n \cdot 1$ as $\sum_1^m 1$ and $\sum_1^n 1$. Then from the associative law of addition it follows that

$$m \cdot 1 + n \cdot 1 = \sum_1^m 1 + \sum_1^n 1 = \sum_1^{m+n} 1 = (m+n) \cdot 1.$$

From the distributive law of multiplication we get

$$(m \cdot 1)(n \cdot 1) = (\sum_1^m 1)(\sum_1^n 1) = \sum_1^{mn}(1 \cdot 1) = (mn) \cdot 1. \qquad \square$$

Now let F be finite. Then $m \cdot 1 = n \cdot 1$ must hold for some pair $m = n$. Suppose $m > n$. Subtracting $n \cdot 1$ form both sides we get $(m-n) \cdot 1 = 0$. Let p be the smallest positive integer for which $p \cdot 1 = 0$ holds. We call p the characteristic of F. Then it follows that $p \cdot e = 0$ for any $e \in F$.

THEOREM. *Let F be a finite field of characteristic p. Then*

(a) *p is prime and*

(b) *the set of multiples $m \cdot 1, m \in \mathbb{Z}$, is a copy of $\mathbb{Z}/p$.*

DEFINITION. We call this set the *prime field* of F.

Proof. (a) If $p = ab$ with $0 < a, b < p$, then $a \cdot 1 \neq 0 \neq b \cdot 1$, but $(a \cdot 1)(b \cdot 1) = (ab) \cdot 1 = p \cdot 1 = 0$. That cannot happen in a field.

(b) If $m = qp + r$, then $m \cdot 1 = (q \cdot 1)(p \cdot 1) + r \cdot 1 = (q \cdot 1)0 + r \cdot 1 = r \cdot 1$. So the elements of the prime field are in 1-1 correspondence with the elements of $\mathbb{Z}/p$ and they add and multiply in exactly the same way. $\square$

2. DEFINITION. Suppose E is a field and F is a subset of E containing 0 and 1, such that F is a field with the addition and multiplication inherited from E. Then F is a *subfield* of E and E is an *extension* of F.

EXAMPLES. The real numbers are a subfield of the complex numbers.
$\mathbb{B}$ is a subfield of GF(16).
Every finite field is an extension field of its prime field.

PROPOSITION. *If E is an extension field of F, then E is a vector space over F.*

EXAMPLE. Remember the complex plane. The complex numbers are a two-dimensional vector space over the real numbers.

Proof. The vector space addition axioms are the same as the field addition axioms. The vector space axioms for multiplication by scalars follow from the field axioms for E by restricting the left hand multipliers to elements of F. □

THEOREM. *Let F be a finite field of characteristic p, then E has p^n elements for some integer n.*

Proof. F is a vector space over its prime field P. If F is finite, then the dimension of this space must be finite. Now every vector in a vector space of dimension n over the field P can be represented by a coordinate sequence $(x_1, \ldots, x_n)$ and every such sequence determines a unique vector. So the number of vectors is the same as the number of coordinate sequences.

We know P has p elements. How many coordinate sequences $(x_1, \ldots, x_n)$ with entries in P are there? Exactly p^n. □

3. THEOREM. *Let F be a field of characteristic p then for $a, b \in F$, $(a+b)^p = a^p + b^p$.*

Proof. First note that $p \cdot a = (p \cdot 1)a = 0 \cdot a = 0$. Next, observe that the binomial theorem allows us to calculate $(a+b)^p$:

$$(a+b)^p = \binom{p}{0}a^p + \binom{p}{1}a^{p-1}b + \cdots + \binom{p}{p}b^p.$$

But the formula $\binom{p}{k} = p!/(k!(p-k)!)$ shows that for $k \neq 0, p$, $\binom{p}{k}$ is a multiple of p. So all the middle terms of the expansion are 0. □

4. THEOREM Fermat's Little Theorem. *If F is a field with exactly $q = p^n$ elements, then for every element $a \in F$, $a^q = a$.*

Proof. Assume $a \neq 0$. Let the elements of F be $f_1 = 0, f_2, \ldots, f_q$. Then by the cancellation law $af_i \neq af_j$ for $i \neq j$. So the elements of F are also precisely $af_1 = 0, af_2, \ldots, af_q$. Hence taking the product of all non-zero elements of F we get:

$$f_2 \cdot \cdots \cdot f_q = (af_2) \cdot \cdots \cdot (af_q).$$

Although, of course the terms of the two products are in a different order. Now we evaluate the right-hand side:

$$f_2 \cdot \cdots \cdot f_q = a^{q-1} f_2 \cdot \cdots \cdot f_q.$$

By the cancellation law, it follows that $a^{q-1} = 1$. So $a^q = a$. □

Chapter 2.4

Roots of Polynomials

In this chapter F will be a field and E an extension field of F (for example F could be $\mathbb{R}$ the reals and E could be $\mathbb{C}$ the complex numbers, or $F = \mathbb{B}$ and $E = \mathrm{GF}(16)$. We shall be interested in the roots in E of polynomials with coefficients in F.

1. DEFINITION. Let $f = \sum a_i x^i \in F[x]$, be a polynomial and $\beta \in E$. We define $f(\beta) = \sum a_i \beta^i \in E$. β is a *root* of f if $f(\beta) = 0$.

PROPOSITION. (a) *The 'evaluation map' $f(x) \mapsto f(\beta)$, respects addition and multiplication (it is a 'homomorphism').*

(b) *$f(\beta)$ is the remainder in $E[x]$ of $f(x)$ when divided by $(x-\beta)$. In particular $f(\beta) = 0$ iff $(x - \beta) \mid f$ in $E[x]$.*

Proof. (a)This follows directly. □

(b) $f(x) = (x - \beta)q(x) + \alpha$, where α is a constant in E. By Part (a) $f(\beta) = 0q(\beta) + \alpha$. □

COROLLARY. *f has at most $\deg(f)$ roots in E.*

Warning. This is false if E does not satisfy *all* the field axioms.

Proof. If $\beta_1, \ldots, \beta_n$ are the distinct roots of f in E, then $f(x) = (x - \beta_1)g(x)$ and for $i > 1$, $(\beta_i - \beta_1)g(\beta_i) = 0$. As $\beta_i - \beta_1 \neq 0$, it follows that $g(x)$ has roots $\beta_2, \ldots, \beta_n$ in E. By induction $\deg(g) \geq n - 1$. Hence $\deg(f) \geq n$. □

2. The formal derivative

We cannot differentiate in the usual way over finite fields, but we can define the formal derivative of a polynomial by copying the formula for real polynomials. This formal derivative retains some of the properties of the original.

DEFINITION. For $f = \sum a_i x^i \in F[x]$, define $f' = \sum i \cdot a_i x^{i-1}$.

PROPOSITION. (a) $(f+g)' = f' + g'$.
(b) $(fg)' = fg' + f'g$.

Proof. (a) This follows directly.
(b) It is sufficient by part (a) to prove this for $g = x^k$.

$$\begin{aligned}(x^k f(x))' &= \sum (i+k) a_i x^{i+k-1} \\ &= \sum i \cdot a_i x^{i+k-1} + \sum k \cdot a_i x^{i+k-1} \\ &= x^k f' + k x^{k-1} f.\end{aligned}$$

□

3. Multiple roots

PROPOSITION. *The multiple roots of f are precisely those that are also roots of f'.*

Proof. If β is a root of f, then $f(x) = (x-\beta)g(\beta)$. So β is a multiple root of f if it is a root of g. Now $f' = (x-\beta)g' + g$. So β is a root of g iff it is a root of f'.

□

4. The Minimal Polynomial of β

If E is finite, the powers of $\beta \in E$ cannot all be distinct so β is a root of a polynomial of the form $x^m - x^n$ with $m > n$. An element of an extension field E that is a root of a non-zero polynomial over F is called *algebraic* over F, the other elements are called *transcendental.* As all the fields we shall be using will be finite, all elements will be algebraic, but eg. e and π are transcendental over the rational numbers.

DEFINITION. If $\beta \in E$ is algebraic, then the *monic* (ie. highest coefficient = 1) polynomial $t(x)$ in $F[x]$ of lowest degree such that $t(\beta) = 0$ is called the *minimal polynomial* of β. It is denoted by $mp_\beta(x)$.

EXAMPLE. $F = \mathbb{B}, E = \mathrm{GF}(16)$.

Element(s)	Minimal Polynomial
0	x
1	$x+1$
10, 11	x^2+x+1
6, 7, 12,13	x^4+x+1
2, 4, 9, 14	x^4+x^3+1
3, 5, 8, 15	$x^4+x^3+x^2+x+1$.

PROPOSITION. *If $t(x) = mp_\beta$ and $f(x) \in F[x]$, then there exists a unique polynomial $g(x)$ with $\deg(g) < \deg(t)$ and $g(\beta) = f(\beta)$.*

Remark. g will be the remainder of f divided by t. This means that the values of polynomials at $x = \beta$ which we denote by $F[\beta]$ are in 1–1 correspondence with the remainders in $F[x]$ on division by t.

Proof. If $f(x) = t(x)q(x)+g(x)$ then $f(\beta) = 0q(\beta)+g(\beta)$. So there is a polynomial of the required type. If $g(\beta) = h(\beta)$, $g \neq h$ and both have degree less than t, then $g - h$ has β as a root and degree less than the degree of t, contradicting the fact that $t = mp_\beta$. □

Important Special Case. f has β as a root iff t divides f. This shows that the minimal polynomial is unique. Because two minimal polynomials for the same element would have to divide each other. Hence they would differ by a constant factor. But they both have highest coefficient = 1, so they're the same.

5. THEOREM. (a) *The minimal polynomial of an algebraic element β is irreducible.*

(b) *If f is a monic irreducible polynomial in $F[x]$ with β as a root, then $f = mp_\beta$.*

(c) *If $t = mp_\beta$ and $\alpha = \beta^q$ where F has q elements, then $t = mp_\alpha$.*

Proof. (a) Let t be the minimal polynomial of β. If $t = fg$, then $0 = t(\beta) = f(\beta)g(\beta)$, so $f(\beta) = 0$ or $g(\beta) = 0$. Say $f(\beta) = 0$, then $t \mid f$, so g has an inverse and must be a non-zero constant.

(b) By Section 4, t divides f. But f is irreducible so $f = at$, where a is a constant. As both t and f are monic $a = 1$.

(c) Raising elements to the q–th power is linear and leaves elements of F fixed (by Little Fermat). So $t(\alpha) = t(\beta^q) = t(\beta)^q = 0$ The statement now follows from Part (b).

Remark. It can be shown that if starting with β and successively taking q–th powers until we get back to β: $\beta = \beta_1, \ldots, \beta_{n-1}^q = \beta_n, \beta_n^q = \beta$, then in $E[x]$, mp_β factors as $\prod(x - \beta_i)$ (in $F[x]$ it is irreducible).

COROLLARY. *If β is algebraic over F, the set of values of polynomials over F at $x = \beta$, $F[\beta]$, is a field and it is a copy of (ie. isomorphic to) $F[x]/mp_\beta$.* □

CONSTRUCTION. Given a polynomial $f(x) \in F$, we can construct a field containing F and a root of $f(x)$. Let g be an irreducible factor of f. Consider the field $F[x]/g(x)$. It contains the element x which we rename β to avoid confusion. Then $g(\beta)$ is the remainder of $g(x)$ when divided by $g(x)$ which is clearly 0. As $g(\beta) = 0$, it follows that $f(\beta) = 0$.

6. We are now going to construct a field with $q = p^n$ elements for any prime p and any power n. The construction takes place in two steps, but first we note a consequence of Fermat's Little Theorem.

PROPOSITION. *Let E be a field of q elements. Then in $E[x]$, $x^q - x = \prod_{\beta \in E}(x - \beta)$.*

Proof. By Fermat's Little Theorem $(x - \beta)$ is a root of $x^q - x$. The number of distinct linear factors this gives is q. So the left and right hand sides have the same degree and the RHS divides the LHS. Comparing highest coefficients, shows that the quotient must be 1. □

This reduces our problem to finding a field in which $x^q - x$ 'splits' into linear factors and then showing that its roots form the field we are looking for.

7. THEOREM. *Let $f \in F[x]$, then there exists a field E containing F over which f splits into linear factors.*

Proof. The idea is to add one root after another in a systematic way. Formally, we phrase that as an induction on the degree of f.

(1) If $\deg(f) = 1$, F is the field and there is nothing to do.
(2) If $deg(f) > 1$ and $f(x) = (x - \beta)g(x)$ in F, apply the induction hypothesis to $g(x)$ to find E in which g splits into linear factors. Then E does the job for f as well.
(3) If $deg(f) > 1$ and f has no linear factors. Let $g(x)$ be an irreducible factor of f. Apply the construction in Section 5 to obtain $K = F[x]/g(x)$ containing F and the root β of $g(x)$. Now f has a linear factor over K so by Step 2 there is a field E containing K in which f splits into linear factors. Then *a fortiori* E contains F and satisfies the claim. □

8. THEOREM. *Let F be a field with p elements and $q = p^n$. Then there exists a field E containing F with exactly q elements.*

Proof. By Theorem 7, there exists a field K containing F over which $x^q - x$ splits into linear factors. Let E be the set of roots of $x^q - x$ in K. E has the right number of elements, because $(x^q - x)' = qx^{q-1} - 1 = 0 - 1 = 1$ and so by Proposition 3, $x^q - x$ has no multiple roots. We will show that E is a field. To do this we must show it contains 0 and 1, is closed under products, sums, negatives and inverses. The other laws follow because they hold for any subset of K.

0 and 1: Certainly $0^q = 0$ and $1^q = 1$, so 0 and 1 lie in E.

Products: If $\beta^q = \beta$ and $\gamma^q = \gamma$, then $(\beta\gamma)^q = \beta\gamma$.

Sums: By Theorem 2.3B, p and hence q is a power of the characteristic of F which is also the characteristic of K. Hence $(\beta + \gamma)^q = \beta^q + \gamma^q = \beta + \gamma$ By Theorem 2.3C.

Inverses: If $\beta^q = \beta$, then $1/\beta^q = 1/\beta$.

Negatives: If $\beta^q = \beta$, then $(-\beta)^q = (-1)^q\beta$. If q is odd $(-1)^q = -1$. Otherwise the characteristic of E is 2 and $-1 = +1$. □

Chapter 2.5

Primitive Elements

1. DEFINITION. A primitive element of a finite field F is an element $\alpha \in F$, such that for every $0 \neq \beta \in F$, $\beta = \alpha^k$ for some k.

If we find a primitive element α in a field we can define k to be the 'log to the base α' of β when $\beta = \alpha^k$. We can then use a table of logarithms to do multiplication rather than work out the whole multiplication table. Primitive elements also have very useful properties for codes as we shall see later.

EXAMPLE. The primitive elements of GF(16) are 2, 4, 6, 7, 9, 12, 13, 14. You can check this directly (and should do so for two or so elements). We shall always use the element 2.

PROPOSITION. *If F has q elements and α is a primitive element of F then the powers $\alpha, \alpha^2, \ldots, a^{q-1} = a^0 = 1$ are distinct and produce all non-zero elements of F.*

Proof. From the Little Fermat Theorem we know that $\alpha^{q-1} = 1$, thus for $k \geq q$, $\alpha^k = \alpha^{k-q+1}$. So the list in the statement is a full list of the elements that are powers of α. But the list has $q-1$ terms, and as α is primitive the powers must be all $q-1$ non-zero elements of F. □

DEFINITION. We define the value k with $0 \leq k \leq q-2$ and $\beta = \alpha^k$ to be the *logarithm* of β to the base α.

It follows that the logarithm of $\beta\gamma$ is the remainder of the sum of the logarithms of β and γ when they are divided by $q-1$ (*why?*).

EXAMPLE. The top row of the table of GF(16) gives the logarithms to the base 2. Check this.

2. The Theorem of the Primitive Element

In this paragraph we shall prove that primitive elements always exist. The proof gives a theoretically adequate, but clumsy, method of finding a primitive element. The reason the proof uses this method is that it is easier to show that it works than with more sophisticated methods. In practice finding a primitive element for a large finite field is a difficult problem. For the proof we need a definition and two lemmas.

DEFINITION. If $0 \neq \beta \in F$, then the *order* of β is the smallest positive power k such that $\beta^k = 1$. We denote it by $\mathrm{ord}(\beta)$. A primitive element is characterised by the fact that $\mathrm{ord}(\alpha) = q - 1$.

LEMMA 1. (a) *For every non-zero* β, $\mathrm{ord}(\beta)$ *divides* $q - 1$.
(b) *If* $\mathrm{ord}(\beta) = m$ *and* d *divides* m, *then* $\mathrm{ord}(\beta^d) = m/d$.

Proof. Part (b) is almost obvious, but part (a) needs a short argument. Firstly $\beta^{q-1} = 1$ by Little Fermat, so $\mathrm{ord}(\beta) = m \leq q - 1$. Now let $q - 1 = mn + r$, where $0 \leq r < m$. Then $\beta^r = \beta^{q-1}\beta^{-mn} = 1.1 = 1$. But m is the smallest positive power for which $\beta^m = 1$. Hence $r = 0$. □

LEMMA 2. *If* $\beta, \gamma \in F$ *and* $\mathrm{ord}(\beta) = m$ *and* $\mathrm{ord}(\gamma) = n$, *and their highest common factor* $(m, n) = 1$, *then* $\mathrm{ord}(\beta\gamma) = mn$.

Warning. It is in general not even true that $\mathrm{ord}(\beta)$ divides $\mathrm{ord}(\beta\gamma)$. Take $\gamma = \beta^{-1}$. Even if γ is not a power of β, this may still be the case — suppose $\mathrm{ord}(\beta) = 10$, $\mathrm{ord}(\gamma) = 6$ and $\gamma^3 = \beta^5$. Then $(\beta\gamma)^{15} = \beta^{15}\gamma^{15} = \beta^{15}\beta^{25} = 1$.

Proof. Certainly $(\beta\gamma)^{mn} = 1$. Suppose $(\beta\gamma)^k = 1$. Then $\beta^{kn} = \beta^{kn}\gamma^{kn} = (\beta\gamma)^{kn} = 1$. Hence m divides kn. So n divides km. Now m and n have no prime factors in common, so m divides kn only if it divides k. Similarly $\gamma^{km} = 1$ and n divides km only if it divides k. So both m and n divide k. Thus their least common multiple, mn divides k. □

THEOREM. *Every finite field has a primitive element.*

Warning. Many textbooks contain a false proof of this theorem which assumes that if the highest common factor of m and n is d, then m and n/d have no prime factors in common.

Proof. Let $p_1, \ldots, p_k$ be the prime factors of $q - 1$ and let s_i be the highest power of p_i that divides the order of some element γ_i of F. By Lemma 1b we may assume that $\mathrm{ord}(\gamma_i) = s_i$. Then $\alpha = \gamma_1 \cdots \gamma_k$ has order $s_1 \cdots s_k = u$ by Lemma 2. Now by the construction of u any $\beta \in F$ has order dividing u. So the non-zero elements of F are all roots of $x^u - 1$. Hence $q - 1 \leq u$. But each s_i is a factor of $q - 1$ (by Lemma 1a), and they are powers of distinct primes. Therefore u divides $q - 1$. Thus they are equal. □

3. Primitive Polynomials

DEFINITION. We call a polynomial *primitive* if it is the minimal polynomial of a primitive root of E.

Remark. There is another use of the term primitive polynomial in the mathematical literature, so check if you come across the term in a textbook whether it has the same meaning as here.

You should also note that whether a polynomial is regarded as primitive depends to some extent on the large field E. It may be the minimal polynomial of a primitive element of a field between F and E. However, there are some polynomials that are never primitive such as the minimal polynomial of 3 in GF(16).

PROPOSITION. *If $|F| = q$ and $|E| = q^n = r + 1$, then a monic polynomial $f(x)$ is primitive iff it is irreducible, f divides $x^r - 1$, but f does not divide $x^m - 1$ for any $m < r$.*

Proof. Over E the polynomial $x^r - 1$ splits into linear factors. So any polynomial f dividing $x^r - 1$ has a root in E. If f is also irreducible, then it must be the minimal polynomial of any of its roots. The element α is a primitive root of E iff α has order r. That is the same as saying α is a root of $x^r - 1$ but not of $x^m - 1$ for any $m < r$. The statement follows. □

EXAMPLE. Primitive polynomials for $F = \mathbb{B}$ and $E = \mathrm{GF}(16)$.

x	does not divide $x^{15} - 1$	not primitive.
$x + 1$	divides $x - 1$	primitive for $\mathbb{B}$ but not for E.
$x^2 + x + 1$	divides $x^3 - 1$	primitive for GF(4) but not for E.
$x^4 + x + 1$		primitive for E.
$x^4 + x^3 + 1$		primitive for E.
$x^4 + x^3 + x^2 + x + 1$	divides $x^5 - 1$	not primitive.

COROLLARY. *If $f(x)$ is primitive for E over F, then all its roots lie in E and they are all primitive elements of E.* □

EXAMPLE. The fact that 2 is a primitive element of GF(16) now automatically gives the primitive elements 4, 9, and 14.

4. THEOREM. *Let E and K be two finite fields with $|E| = |K| = q = p^n$, where p is a prime number. Then E and K are isomorphic,* ie. *there is a one-to-one map of E onto K that preserves all the arithmetic operations.*

Remark. This means that there is essentially exactly one finite field for each prime power order.

Proof. Let $F = \mathbb{Z}/p$ which is contained in both fields. Let α be a primitive element of E with minimal polynomial $f(x)$. Then f has a root β in K which is a primitive element of K. So $E = F[\alpha]$ and $K = F[\beta]$. But both these are isomorphic to $F[x]/f(x)$. □

Finite Fields Summary

Construction

- The 'smallest' finite fields are the fields Z/p for p a prime number obtained by taking the remainders on division by p.
- To extend a finite field F, find an irreducible polynomial $f(x)$ in $F[x]$ and take the remainders modulo $f(x)$, that is $F[x]/f(x)$.
- Addition, multiplication and subtraction are the usual operations followed by taking the remainder on division by $f(x)$ (or p in the case of Z/p).
- Division is done by Euclid's Algorithm.

Existence

- There exist fields of order p^n for all primes p and all powers n and no others.

Structure

- The field of order q is the set of roots of the polynomial $x^q - x$.
- There is essentially only one field of each possible order.
- A copy of the field F of order q lies inside the field E of order r iff r is a power of q.
- Every finite field has a primitive element. Using this as a base for logarithms makes it possible to perform arithmetic without recourse to division with remainder or Euclid's algorithm.

Coding Theory
Exercise Sheet 5

1. Find the inverses of 79 and 90 modulo 787.

2. Calculate the inverse of $5 = 0101$ and $7 = 0111$ using Euclid's Algorithm in GF(16).

3. Check thatif $F = \mathbb{B}$, $\mathbb{Z}$ or $GF(16)$ polynomials with coefficients in F satisfy the conditions for a domain (even though $\mathbb{Z}$ is not a field).

The next three exercises will take quite a lot of computation, but they are irreplaceable for obtaining a feeling for the kind of calculations involved in codes. It is not necessary to do the ternary case, but you should at least work through the binary one.

4. Construct fields with 8 and 9 elements respectively, by finding appropriate irreducible polynomials over $\mathbb{B}$ and $\mathbb{T} = \mathbb{Z}/3$. Write down their addition and multiplication tables.

5. Find the minimal polynomials over $\mathbb{B}$ and $\mathbb{T}$ respectively for all the elements of your two fields. Show that these are all the irreducible polynomials of degree dividing 3 and 2 respectively. Check that the products of the polynomials give $x^8 - x$ and $x^9 - x$ in the two cases.

6. For each minimal polynomial check that the complete set of its roots can be obtained from any one of them by successive squaring for the field of order 8 and successive cubing for the field of order 9.

Coding Theory
Exercise Sheet 6

These two exercises will be needed for decoding BCH codes. Question 1 is proved by induction. Question 2 ought to be familiar from mathematical methods courses. It can be accepted on faith if you so wish. It is proved by adding rows and columns cleverly to make the calculation of the determinant easy. As they are used in the course, answers for these exercises are provided.

1. Let the rows of Euclid's algorithm, applied to polynomials $a(x)$ and $b(x)$ be numbered starting with -1 (as in the lectures). Prove:

a. The degrees of the entries in the R column decrease strictly. The degrees of the entries in the U and V columns increase strictly for $k \geq 0$.

b. For any $k \geq 0$, $u_{k-1}v_k - u_k v_{k-1} = (-1)^k$. Deduce that the highest common factor of u_k and v_k is 1.

c. For any $k \geq 0$, $r_{k-1}v_k - r_k v_{k-1} = (-1)^k a$.

2. Let $a_1, a_2, \ldots, a_n$ be elements of a field F, and let A be the matrix with entries (a_i^j). Show that the determinant of A is

$$a_1 a_2 \cdots a_n \prod_{i<j} (a_j - a_i).$$

This is the so-called Vandermonde determinant. Notice that if $a_1, \ldots, a_n$ are non-zero and distinct, the value of the determinant is non-zero.

2001

Coding Theory
Hints and Answers 6

1. a. The equation $r_{k-1} = q_{k+1}r_k + r_{k+1}$ represents division with remainder and the degree of the remainder is always less than the degree of the divisor. Note that this implies that the quotients q_k all have degree ≥ 1.

The proof of the second statement is by induction. We shall do the proof for u_k. It is the same for v_k. $u_0 = 0$, $u_1 = 1$, induction start is given. $u_{k+1} = u_{k-1} - q_{k+1}u_k$ and $\deg(u_k) > deg(u_{k-1})$. So $\deg(u_{k+1}) = \deg(q_{k+1}) + \deg(u_k) > \deg(u_k)$. □The above formula can be used to determine the degrees of u_k and v_k precisely.

b. The proof is again by induction. For $k = 0$: $1.1 - 0.0 = 1$, so the start is given. Now suppose the formula holds for k.

$$\begin{aligned} u_k v_{k+1} - u_{k+1}v_k &= u_k(v_{k-1} - q_{k+1}v_k) - (u_{k-1} - q_{k+1}u_k)v_k \\ &= u_k v_{k-1} - u_{k-1}v_k = -(-1)^k = (-1)^{k+1}. \end{aligned}$$

Hence any polynomial that divides both u_k and v_k must divide $u_k v_{k-1} - u_{k-1}v_k = \pm 1$.

c. The proof is a precise analogue of that of (b).

For $k = 0 : a.1 - 0.0 = a$, so the induction start is given. Now suppose the formula holds for k.

$$\begin{aligned} r_k v_{k+1} - r_{k+1}v_k &= r_k(v_{k-1} - q_{k+1}v_k) - (r_{k-1} - q_{k+1}r_k)v_k \\ &= r_k v_{k-1} - r_{k-1}v_k = -(-1)^k a = (-1)^{k+1}a. \end{aligned}$$

2. First divide each row by a_i. This divides the determinant by $a_1 \cdots a_n$, and it produces an $n \times n$- matrix $B_1 = (a_i^{j-1})$. The idea is to manipulate the matrix B_1 to reduce calculation of its determinant to that of an $n-1 \times n-1$- matrix of the same form. To do this subtract $a_1 \cdot \mathrm{col}(n-1)$ from $\mathrm{col}(n)$, then $a_1 \cdot \mathrm{col}(n-2)$ from $col(n-1)$, and so on until we subtract $a_1 \cdot \mathrm{col}(1)$ from $\mathrm{col}(2)$ (the order these subtractions are performed in is important. Why?). The determinant has remained unchanged and the first row of the resulting matrix is (1, 0, ..., 0). So we expand by the first row. This gives $\det(B_1) = \det(A_2)$, where $A_2 = (a_i^{j-1} - a_1 a_i^{j-2})$, and i and j range from 2 to n.

Divide each row of A_2 by $(a_i - a_1)$. This divides the determinant by $(a_2 - a_1) \cdots (a_n - a_1)$ and leaves a matrix $B_2 = (a_i^{j-2})$, where i and j range from 2 to n. B_2 has the same form as B_1 but one row and column less. We can now appeal to induction (or repeat the process applied to B_1) to get:

$$\det(B_2) = \prod_{2 \leq i < j \leq n} (a_j - a_j).$$

Now $\det(A) = a_1 \cdots a_n \det(B_1) = a_1 \cdots a_n (a_2 - a_1) \cdots (a_n - a_1) \det(B_2)$ proving the formula.