Factorising Polynomials: a background to some algorithms

John Coffey, Cheshire, UK.

2022

Key words: polynomial factorisation, factor, gcd, Chinese remainder theorem, finite field, Fermat's Little Theorem, discriminant, cyclotomic, Hensel lifting, square-free, distinct degree, equal degree, Berlekamp, Cantor & Zassenhaus algorithms

This article is my survey and re-investigation of some of the means by which a general polynomial in one variable x with integer coefficients

$$F(x) = c_n x^n + c_{n-1}x^{n-1} + c_{n-2}x^{n-2} + ... + c_1 x + c_0 = 0 \quad (1)$$

can be split into a product of irreducible factors in the integers $\mathbb{Z}$ or rational number field $\mathbb{Q}$. Most computer-based algorithms achieve this by first factoring over finite fields $\mathbb{F}_p$ or $\mathbb{F}_{p^k}$. The treatment is at the undergraduate level and for the interested amateur.

This study arose from my personal attempt to understand Galois theory and in particular the algorithms by which the Galois group can be determined without the roots of its parent polynomial $F(x)$ first being known. These methods make heavy use of the ability of algebraic software to factorise a large polynomial quickly over the integers. Unless the computational cost of such factorisation is low, finding a Galois group would remain an intractable problem for all but the most simple polynomials. Since much development and ingenuity has gone into integer and polynomial factorisation since computing became widespread in the late 1960s, I have been interested in seeing at first hand how these work. Most rely on classical theorems in number theory, such as Euclid's method for the greatest common divisor (gcd) of two integers, the Chinese remainder theorem (~AD 300) for finding the congruence of a composite integer, and the 18th century investigations of Euler and Gauss.

§1 reviews much school-level knowledge about factors of integers and polynomials. The following sections cover material underpinning computer implementation of fast algorithms.

- §2 discussed some naive and inefficient approaches including Kronecker's early and not very successful attempt to find a factorisation algorithm.
- §3 introduces congruences, the Chinese Remainder theorem for multiple simultaneous congruences, and quadratic residues.
- §3 and 4 introduce respectively the related topics of finite fields and Fermat's little theorem.
- §6, supported by Appendices 2 and 3, describes the powerful discrete degree factorisation algorithm which is applied prior to the Cantor-Zassenhaus algorithm.

- This is followed in §7 by Berlekamp's algorithm, the first really successful factorisation method for computer implementation.
- Berlekamp's algorithm works over a finite field, and obtaining a factorisation over the infinite field $\mathbb{Q}$ is obtained by the process called Hensel lifting, described in §8.
- §9 outlines with an example the more recent algorithm by Cantor and Zassenhaus (C-Z) which has to some extent superseded Berlekamp's. Today's computer software packages mainly use a judicious mix of Berlekamp's and the C-Z algorithm, with refinements to speed them up depending on the precise nature of the given polynomial.
- The final section of the body of the article, §10, is about computing the discriminant.
- Three Appendices give tangential details, though Appendix 3 is a substantial discussion of the factorisation of $x^n - 1$.

1 Elementary methods

1.1 Working with integers

We start with a review of 'by hand' methods for factoring integers and low degree polynomials, such as we were probably taught at secondary school. The fundamental theorem of arithmetic states that every integer is either a prime or a unique product of primes. The primes start 2, 3, 5 7, 11, 13, ... and further primes can be found by the 'sieve of Eratosthenes'. A list of all integers is made, in order up to a convenient limit N, and we strike out every multiple of 2, every multiple of 3, of 5, of 7 and so on. When all multiples of primes have been deleted, the numbers remaining are also primes. Using this enlarged list a further search can be made and the table of primes further extended. The method is efficient in that division by primes from 2 to p_k will reveal all further primes up to $p_{k+1}^2 - 2$. For example, testing with only 2 and 3 identifies all primes up to 23, and testing with 2, 3 and 5 reveals them up to 47.

Any given positive integer N may be factorised by a similar sieve method in which it is divided first of 2 until no further division without remainder is possible, then by 3, then 5, ... until only a single larger prime p remains. N is then in the unique factorisation $2^{n_2}.3^{n_3}.5^{n_5}....p^{n_p}$.

The prime factors of any N combine to give the complete list of its factors. Thus if $N = 36 = 2^2.3^2$, its factors are $\pm 36, \pm 18, \pm 12, \pm 9, \pm 6, \pm 4, \pm 3, \pm 2$ and trivially ± 1. This can sometimes be used to spot integer roots of simple quadratic expressions such as $x^2 + 5x - 36$. If this does split into linear factors $(x - \rho_1)(x - \rho_2) = x^2 - (\rho_1 + \rho_2)x + \rho_1\rho_2$, the constant term is the product of the roots and the coefficient of x is the negative of their sum. So here we are looking for two factors of -36 which add to 5. 9 and -4 come readily to mind giving $x^2 + 5x - 36 = (x + 9)(x - 4)$. Secondary school students learn to apply this to exercises about sketching the graph of a quadratic function since the roots 4 and -9 are the two positions when the curve cuts the x-axis. Quadratics occur frequently in physics, such as describing the trajectory of a projectile under gravity. In these real world situations the variable x represents a physical quantity such as distance, velocity or area.

Still considering integers, the greatest common divisor (gcd), also called the highest common factor, is important in much of number theory and crucial in computer algorithms for factorising polynomials. It is obvious that if a divides b (written $a|b$), $\gcd(a, b) = a$. Otherwise the gcd of two integers can be found by Euclid' algorithm. It is based on the following facts. If $A < B$, division $B \div A$ is performed by finding quotient q and remainder r, $0 \le r < A$ such that $B = qA + r$. If $\gcd(A, B) = d$,

then $A = ad$, $B = bd$ for coprime integers $a < b$. Then $bd = qad + r$. Since d divides the left side, it must also divide the right so $r = jd$ for some integer j. So A and r have the same gcd as do B and A. The problem has been reduced to finding the gcd of the smaller numbers A and r, which in turn can be further reduced by the same process of quotient and remainder. Eventually the downwards ladder must end with d as the final non-zero remainder.

Take $B = 657$, $A = 225$. The steps to find $d = \gcd(A, B)$ are set out in the left panel below. This is read from top to bottom and proves that $d = 9$, the last non-zero remainder. Euclid's algorithm can be run in reverse to give an expression for d in terms of the given integers A and B. In the right panel, to be read from bottom to top, the remainder at each stage is substituted by a difference of numbers in the row above so that d eventually becomes a linear combination of A and B. This presentation is called Bézout's identity.

$$\begin{array}{rcrcr} 657 & = & 2\times 225 & + & 207 \\ 225 & = & 1\times 207 & + & 18 \\ 207 & = & 11\times 18 & + & 9 \\ 18 & = & 2\times 9 & + & 0 \end{array}$$

$$\begin{array}{rcrcr} 9 & = & 12\times 657 & - & 35\times 225 \\ & = & -11\times 225 & + & 12\times(657-2\times 225) \\ & = & -11\times 225 & + & 12\times 207 \\ & = & 207 & - & 11\times(225-207) \\ 9 & = & 207 & - & 11\times 18 \end{array}$$

1.2 Basic operations with polynomials

In Britain A-Level maths exam papers usually contain a question such as

> "Given that $(x+2)$ and $(x-1)$ are factors of $ax^3+5x^2+bx-6=0$, find the values of a and b".

The examiners expect students to use the 'factor theorem' or 'root theorem' which states that if $x-\rho$ is a linear factor of polynomial $F(x)$, then $F(\rho)=0$. Clearly $F(x)=(x-\rho)G(x)$ where the degree of $G(x)$ is one less than that of $F(x)$. Then $F(x)=0$ if either $x-\rho$ or $G(x)$ is zero, or both are. This exam question is solved by substituting the given roots $x=-2$ and $x=1$ in turn into the polynomial to obtain two simultaneous equations in a and b. In this problem $a=2$, $b=-1$ so $2x^3+5x^2-x-6=0=(x-1)(x+2)(cx-d)$. In this example the third factor can be found by inspection to match the coefficients; it is $2x+3$.

For exams the polynomials are chosen to have simple and fairly obvious coefficients and roots, but rarely do 'real world' problems produce such amenable polynomials. Consider, for instance, $x^2-2316x+998739$. The constant term is $998739 = 3^2.7.83.191$ so there are 30 non-trivial factors of 998739 formed as products of these five primes, together with their negatives. We could make a list, but there is the alternative of using the well known formula for the roots of a quadratic '$ax^2=bx+c=0$'

$$\rho_{1,2} = \frac{1}{2a}\left(-b \pm \sqrt{b^2-4ac}\right).$$

This is derived by the process of 'completing the square' which compares

$$x^2+\frac{b}{a}x+\frac{c}{a} = x^2+\frac{b}{a}x+\frac{4ac}{4a^2} \quad \text{with} \quad \left(x+\frac{b}{2a}\right)^2 = x^2+\frac{b}{a}x+\frac{b^2}{4a^2}.$$

Here $b^2-4ac = 1,368,900 = 1170^2$ so $\rho_1 = 1743$, $\rho_2 = 573$. The quantity 'b^2-4ac' is called the discriminant Δ of the quadratic and carries significant information about the roots. Specifically, if $\Delta=0$, there are two coincident roots, if $\Delta>0$ there are two real roots, and if $\Delta<0$, the roots are a complex conjugate pair. Moreover, there are no integer or fractional roots unless Δ is a perfect square, n^2 or n^2/d^2, of some n and d.

Discriminants are defined for polynomials of all degrees, and carry equivalent information about the nature of the roots. Here we note that for a monic polynomial (leading coefficient = 1) with integer coefficients, Δ is always either a positive or a negative integer. It can be expressed as a polynomial in the coefficients of the given $F(x)$. An algorithm is given in §10.

In looking to factorise polynomials an obvious step is to factor out the gcd of the coefficients. A polynomial whose coefficients have no common factor is called 'primitive'. There is a theorem by Gauss about primitive polynomials which places the integers $\mathbb{Z}$ and the rationals $\mathbb{Q}$ on the same footing. It states that if F and G are two primitive polynomials, then their product FG is also primitive. A corollary, proved in textbooks, is that a primitive polynomial is irreducible over $\mathbb{Z}$ only if it is also irreducible over $\mathbb{Q}$. Much of the theory is developed for 'monic' polynomials which are those whose leading coefficient is 1. Eq 1 can be transformed to a monic polynomial with the same degree and essential structure by the substitution $x = u/c_n$. For example putting $x = u/3$ in $3x^4 + 2x - 5$ transforms it to $(u^4 + 18u - 135)/27$, and the common constant factor $1/27$ does not affect the operations for finding the product, quotient, gcd with another polynomial. An alternative method is used when working modulo an integer as explained in Appendix 1.

The fundamental theorem of algebra states that every polynomial will split into linear factors over the complex numbers $\mathbb{C}$. However this article is about factorisation over integers $\mathbb{Z}$ and rationals $\mathbb{Q}$, and here many polynomials are irreducible. A polynomial of degree n may split over $\mathbb{Q}$ into several irreducible factors of degrees d_1, d_2, d_3, ..., d_k where $d_1 + d_2 + ... + d_k = n$. The degrees are a partition of n. A cubic, therefore, could be irreducible (degree = 3), or split with degrees [1,2] or [1,1,1]. In the case [1,2] there is one real root and a complex conjugate pair. In the [1,1,1] case it has split into linear factors, though not necessarily over $\mathbb{Q}$ – the roots could lie in some algebraic number field extension of $\mathbb{Q}$, the subject of Galois theory. In the general case of factorising a high degree polynomial, we hope for clues as to how the total degree is partitioned.

One important set of polynomials is the cyclotomic ('circle-cutting') polynomials; they can point to the factorisation of general polynomials. For all n

$$x^n - 1 \;=\; (x-1)(x^{n-1} + x^{n-2} + x^{n-3} + ... + x + 1). \qquad (2)$$

These polynomials, written Φ_n, are the irreducible factors of $x^{n-1} + x^{n-2} + x^{n-3} + \;....\; + x + 1$ over $\mathbb{Z}$. They are called circle-cutting because if their roots in $\mathbb{C}$ are plotted in the complex plane, they lie on the unit circle. The totality of roots of $x^n - 1$ lie at equal angels around the unit circle with one point at $x = 1$, and each cyclotomic polynomial contributes its own particular subset of these complex roots. Over $\mathbb{Z}$, if n is an odd prime p, $x^p - 1$ does not factor any further. If n is even, however, there is a second real root at $x = -1$. Appendix 3 examines the factors of $x^n - 1$ in some detail.

Given an arbitrary polynomial, $P(x)$, perhaps the first step should be to determine by some test whether it is irreducible over $\mathbb{Q}$. Reliable tests for irreducibility are essential since much theory, such as traditional Galois theory, applies almost entirely to polynomials irreducible over $\mathbb{Q}$. Probably the most familiar test is Eisenstein's. Given $P(x)$, we look for a prime q such that

- q does not divide the leading coefficient, c_n (= 1 for a monic polynomial),
- q does divide all the other coefficients,
- but q^2 does not divide the constant term c_0.

Consider these three cubics over the integers:

$$P_1(x) = x^3 - 14x^2 + 21x + 35, \qquad P_2(x) = x^3 - 14x^2 + 22x + 24. \tag{3}$$

$$P_3(x) = x^3 + 4x^2 + 12x + 37 \quad \rightarrow \quad P_3(x-3) = x^3 - 5x^2 + 15x + 10.$$

Looking at P_1, 7 stands out as a candidate prime. 7 divides 14, 21 and 35 but $7^2 = 49$ does not divide 39. Hence $P_1(x)$ is irreducible over $\mathbb{Q}$. Applying the test to P_2, however, we have 2 dividing 14, 22, 24 but also $2^2|24$. So the criterion is not met and the test tells us nothing. P_3 shows how the scope of Eisenstein's criterion can be extended by making a shift in the variable $x \rightarrow x + h$, $h \in \mathbb{Z}$. $P_3(x)$, as expressed on the left, has no coefficients consistent with Eisenstein's test. However replacing x with $x - 3$ transforms P_3 to a polynomial which does meet Eisenstein's criterion and so is proved irreducible. The original, unshifted version of $P_3(x)$ is necessarily also irreducible.

A polynomial which is reducible over $\mathbb{Z}$ or $\mathbb{Q}$ must also factor in every finite field $\mathbb{F}_p$. Moreover, it factors modulo p into polynomials with the same or smaller degrees. The basic reason is that

$$(a \mod p)(b \mod p) = ab \mod p. \tag{4}$$

As a numerical example, $P(u) = u^4 + 18u - 135$ mentioned above does factorise over $\mathbb{Z}$ as $(u-3)(u^3 + 3u^2 + 9u + 45)$ and factorises as follows modulo several primes:

mod 2	$(u+1)^4$
mod 3	$(u+1)^4$
mod 5	$u(u+3)(u^2+3u+4)$
mod 7	$(u+4)^2(u^2+6u+6)$
mod 11	$(u+8)(u^3+3u^2+9u+1)$
mod 29	$(u+9)(u+24)(u+26)(u+28)$

Note how in each prime factorisation there is at least one linear term. The version for mod 11 is almost the original $P(x)$ with coefficients mod 11. A corollary is that if $P(x) \mod p$ is irreducible for any prime p, then $P(x)$ must be irreducible over $\mathbb{Q}$. This can be a helpful test for an irreducible polynomial, complementing Eisenstein's test. 2 is a good starting prime since there are only a small number of polynomials of any degree mod 2 and each can be test divided into the given $P(x) \mod 2$. For example at Eq 3 $P_1(x) \mod 2 \rightarrow x^3 + x + 1$. The only possible factors are x, $x+1$, x^2, x^2+1, x^2+x, $x^2+x+1 \mod 2$, and of these we only have to try division by x and $x+1$ so see that there is no factorisation. Nor does P_1 factor modulo 3, 7, 11 or 13, giving ample proof that it is irreducible over $\mathbb{Q}$.

The converse in *not* true. There are polynomials such as $x^4 + ax^2 + b^2$ which will factorise modulo *any* prime, but which may be irreducible over $\mathbb{Q}$, depending on a and b.

Look back now to $P_2(x)$ at Eq 3. A search of the factors of 24 reveals that $P_2(12) = 0$ so $x - 12$ is a linear factor. The quadratic factor may be found by polynomial division. Arithmetic operations on polynomials are often done best by regarding the quantity x and its powers merely as place-holders to distinguish the various coefficients. This is an important different way of regarding the quantity x. In physical problems such as equations of motion, we use x to represent a physical quantity. However, there is a philosophically different way to regard x – just as a place-holder in a formal polynomial which has no links to physical reality. The analogy is with calculating in hundreds, tens and units by working in columns, ..., x^2, x^1, x^0. The role of the x^k here is to identify and separate the columns. For this reason in some texts x is referred to as an 'indeterminate'. Provided

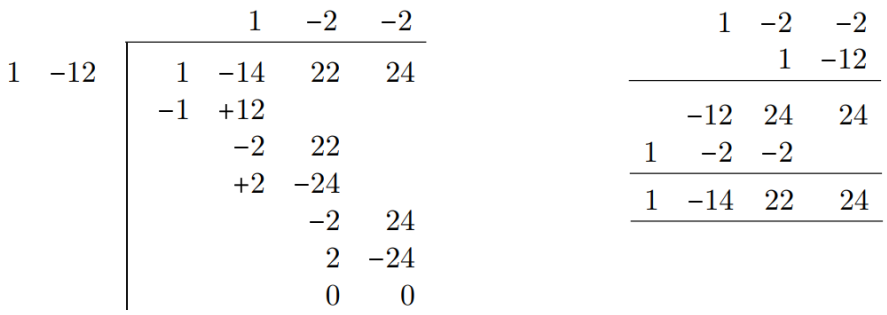


Figure 1: Illustrating division and multiplication of two polynomials by working in columns and manipulating only the coefficients. $(x^3 - 14x^2 + 22x + 24) \div (x - 12)$.

the coefficients are separated, a formal polynomial could be notated in several ways. For instance $P_2(x) = x^3 - 14x^2 + 22x + 24$ could equally well be represented as (1 −14 22 24) or as $|1| - 14|22|24|$. I will use the former of these notations frequently in this article. The examples in Figure 1 show division and multiplication by working with coefficients in columns.

The Euclidean algorithm for the gcd carries over to polynomials, and polynomial division is required. One application is to finding double and multiple factors. If $P(x)$ has a factor raised to some power > 1, then its formal derivative $P'(x)$ will share the same factor. Hence $\gcd(P, P')$ will reveal the common factor. Take the example

$$P(x) = x^6 - 6x^5 + 15x^4 - 25x^3 + 30x^2 - 12x - 8, \qquad P'(x) = 3(2x^5 - 10x^4 + 20x^3 - 25x^2 + 20x - 4).$$

Divide $2P$ by $P'/3$ to obtain quotient $x - 1$ and remainder $r_1 = -5x^3 + 15x^2 - 20 = -5(x^3 - 3x^3 + 4)$. Now divide $P'/3$ by $-r_1/5$ to obtain quotient $2x^2 - 4x + 8$ and remainder $r_2 = -9(x^2 - 4x + 4)$. Finally divide $-r_1/5$ by $-r_2/9$ to obtain quotient $x + 1$ and zero remainder. $\gcd(P, P')$ is the last non-zero remainder so is $x^2 - 4x + 4 = (x - 2)^2$. This, then, is a common factor, so the multiple linear factor in $P(x)$ is $(x - 2)^3$. By polynomial division the other factor is found to be $x^3 + 3x + 1$. With the shift $x \to x - 1$ this becomes $x^3 - 3x^2 + 6x - 3$ and is proved to be irreducible by Eisenstein's criterion with the prime 3. Hence the full factorisation of $P(x)$ over the integers is $(x - 2)^3(x^3 + 3x + 1)$.

One further property of greatest common divisors is worth remembering. If b and c are coprime,

$$\gcd(a, bc) = \gcd(a, b) \,.\, \gcd(a, c). \tag{5}$$

For suppose that $a = k.m.q$, $b = k.r$, $c = m.s$ so that $\gcd(a, b = k)$, $\gcd(a, c) = m$, $\gcd(b, c) = 1$. Then $bc = k.m.r.s$ and $\gcd(a, bc) = km$. Clearly if b and c have a non-trivial common factor the result is not true.

The rest of this article describes through examples some of the theory and techniques which underpin modern fast algorithms for computer factorisation of polynomials.

2 Inefficient algorithms

I open the discussion by looking at some naive factorisation methods which might be tried, but which would be hopelessly inefficient. This will establish a contrast with the efficient methods described in later sections.

2.1 Direct and random search in a finite field

We are given monic polynomial $P(x)$ with degree N and coefficients in field $\mathbb{F}_p$ (see §4 for explanation), and want to find its irreducible polynomial factors $H_1(x)$, $H_2(x)$, ... , $H_n(x)$. N will in general be large – often greater than 10. Would a direct trial and error search for factors be worthwhile?

The answer depends on the probability of hitting upon one factor at each attempt. If the degree of each $H_j(x)$ is h_j, $N = \sum h_j$. The number of ways the h_j could partition N rises steeply with N, being only 5 for $N = 4$, but 11 for $N = 6$, 42 for $N = 10$ and 176 for $N = 15$. Of course many of these will involved more degree-1 factors than modulo p would allow, but nevertheless it is clear that the probability that $P(x)$ has a factor of any particular degree is not large. For example, with $N = 10$, of the 42 partitions only six have degree 6, seven degree 5, and eleven degree 4. So if we are searching for degree-4 factors, there is only about a 25% chance of there being one within $P(x)$. Trial division in search of this degree-4 factor requires a list of irreducible degree-4 polynomials over $\mathbb{F}_p$. The number of reducible and irreducible polynomials $K(x)$ with degree k and coefficients modulo p is given in Table 1 (see also Appendix 3 §A3.3). So, working in $\mathbb{F}_3$ there are 18 irreducibles to try in turn. Assuming we know what these irreducibles are (and building the list requires additional effort), on average about 9 would have be to tried to find one degree-4 factor assuming that it exists, and all 18 even if one does not exist. I will not try to quantify the statistics as it is clear that direct search would be tedious and time consuming.

An alternative 'method' might we be take find the greatest common divisor of $P(x)$ with a set of randomly chosen test polynomials $K(x)$. If the degree k of $K(x)$ is high, it may be built from a large number of factors, one at least of which might happen also to be in $P(x)$. The bottom panel of Table 1 shows that the probability of $K(x)$ being reducible, and hence possibly ripe with candidate factors, is over 80% for $k > 5$. This is a scatter-gun approach, firing products of random factors at $P(x)$ in the hope that one will strike lucky. Since there are more partitions of N of degree-1 and

	Total			
degree	$p = 2$	$p = 3$	$p = 5$	$p = 7$
1	2	3	5	7
2	4	9	25	49
3	8	27	125	343
4	16	81	625	2401
5	32	243	3125	16807
6	64	729	15625	117649

	Irreducible			
	$p = 2$	$p = 3$	$p = 5$	$p = 7$
	2	3	5	7
	1	3	10	21
	2	8	40	112
	3	18	150	588
	6	48	624	3360
	9	116	1380	19544

	% reducible			
	$p = 2$	$p = 3$	$p = 5$	$p = 7$
1	0	0	0	0
2	75	67	60	57
3	75	70	68	67
4	81	78	76	76
5	81	80	80	80
6	86	84	91	83

Table 1: The total number of irreducible polynomials modulo p with degree from 1 to 6. The number of irreducible polynomials. Bottom panel : the percentage of reducible ones.

degree-2, this approach might have some success for such low degree factors.

I have tried this in a limited numerical experiment. Trials were with three polynomials over $\mathbb{F}_3$ which also feature in §6 on distinct degree factorisation.

- The first is $x^6 + x^4 + x^2 + 1$ = (1 0 1 0 1 0 1) in coefficient-only notation, the product of three irreducible quadratics. I tested with 21 different random monic degree-10 polynomials and found quadratic factors with 9 of them plus a 4th-degree factor, the product of two quadratics. This is about 50% success.

- The second $P(x)$ was (1 2 2 2 1 2 0 2 0 1), the product of three degree-3 factors. For this I chose 26 random degree-15 $K(x)$. Only two found a cubic factor, a poor success rate of < 10%.

- The third polynomial had degree 16, being the product of four degree-4 factors. I calculated the $\gcd(P(x), K(x))$ with ten degree-10 then ten degree-12 random $K(x)$. Out of these 20 trials only one found a quadratic factor. In all other trials the gcd was 1.

A further trial aimed to factorise the random degree-10 polynomial $x^{10} + 2x^9 + x^7 + 3x^6 + 2x^4 + 2x^3 + 4x^2 + 1$ = (1 2 0 1 3 0 2 2 4 0 1) modulo 5. Substituting x = 0, 1, 2, 3, 4 shows that it has no linear factors. Most gcd were 1, but the sixth trial found (1 2 0 1). Dividing this out leaves the 7th degree (1 4 2 0 2 2 4 1). Trial number 21 yielded $x^2 + 2$ leaving (1 4 0 2 2 3) to be factored. The requirements for Eisenstein's test are not met even with substitutions $x \to x + 1$ *etc.* Pressing on to 30 trials yielded no further factors, and other test show that $x^5 + 4x^4 + 2x^2 + 2x + 3$ is irreducible. So all factors were found after 21 trials, a success rate is about 10%.

This gcd scatter-gun method has the merit over direct search in that a table of irreducibles does not have to be prepared, but against this is the extra computation of finding the gcd of two possibly large polynomials. We may conclude that search and scatter-gun gcd methods are of almost no use except possibly for polynomials of low degree and small prime field p. However, finding the gcd of $P(x)$ with a carefully chosen polynomial rich in factors is the basis of the powerful Cantor-Zassenhaus equal degree factorisation algorithm described in §9.

2.2 Kronecker's method of factorising a polynomial over $\mathbb{Z}$

Leopold Kronecker proposed this method in the mid 19th century. As a method for factorising a general polynomial over the integers it is almost useless because it produces a long list of trial polynomials which might possibly be factors, every one of which must be tested by division into the given $P(x)$. Nevertheless it is of historic interest and spurred other mathematicians to invent ones that are easier and faster. I will illustrate it with an example, so we will try to factorise $P(x) = x^5 + x^4 - 9x^3 + x^2 - 15x + 15$.

It is always sensible to start by removing any repeated factors by computing the gcd of $P(x)$ and its derivative. Suppose that this has been done and no squared factor has been revealed. It is also sensible to check it modulo 2 to see if it is irreducible in $\mathbb{F}_2$ since that would imply that it was also irreducible over $\mathbb{Q}$. Modulo 2 it is $x^5 + x^4 + x^3 + x^2 + x + 1$. Setting $x = -1$ gives value 0 so $x + 1$ is a factor, so on this evidence we have no reason to think that $P(x)$ is irreducible over $\mathbb{Z}$.

Kronecker's method starts with the observation that since the degree of $P(x)$ is 5, the degrees of any polynomial factors must be in one of the sets {1,4}, {1,1,3}, {1,1,1,2}, {1,1,1,1,1} or {2,3}. So the factorisation would have to involve factors of degree 1 (linear) and/or 2 (quadratic). So we investigate a factorisation $(x^3....)(x^2....) = h(x)g(x)$ where k is a cubic and g a quadratic. Kronecker

then evaluates $P(x)$ at three integer points – three since a quadratic has 3 coefficients. We would prefer the integer values at the chosen integers to be small and not highly composite. I choose

$$P(0) = 15, \qquad P(1) = -6, \qquad P(2) = -35.$$

The argument now goes that if $P(x) = h(x)g(x)$, then $P(x_j) = h(x_j)g(x_j) = y_j$, say, and $g(x_j)|P(x_j)$, that is, $g(x_j)$ divides y_j. At $x = 0$, $y_0 = 15$ which has factors $\pm1, \pm3, \pm5, \pm15$. Any of these could be the value of $g(0)$, but no other integer could be. Similarly at $x = 1$, $y_1 = -6$ which has factors $\pm1, \pm2, \pm3, \pm6$, and y_2 has factors $\pm1, \pm5, \pm7, \pm35$. We do not know which of the 8 possibilities for y_j, $j = 1, 2, 3,$, appertains at each of the three chosen x positions, but we do now have a complete list of possibilities. A unique quadratic curve $a_2x^2 + a_1x + a_0$ can be drawn through any 3 given points and its coefficients can be calculated unambiguously by several methods. Some authors use the Lagrange interpolation formula. I have used matrix inversion. So suppose we guess that the contributions of $g(x)$ to $P(x)$ at 0, 1 and 2 are respectively 3, 1, 1, numbers chosen from the list of factors above. The matrix equation for the coefficients is $\mathtt{Ma} = \mathtt{b}$ with solution $\mathtt{a} = \mathtt{M}^{-1}\mathtt{b}$:

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{pmatrix} \begin{pmatrix} a_2 \\ a_1 \\ a_0 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} \qquad \text{where} \quad \mathtt{M}^{-1} = \tfrac{1}{2} \begin{pmatrix} 1 & -2 & 1 \\ -3 & 4 & -1 \\ 2 & 0 & 0 \end{pmatrix}$$

The result is $g(x) = x^2 - 3x + 3$. Whether this is a factor of $P(x)$ now requires a test division, and in this case the remainder is $-45x + 42$. So $x^2 - 3x + 3$ is not a factor and we must try another selection of three divisors from the lists. Clearly up to $8^3/2 = 256$ quadratics might have to be tried, the 2 because half the polynomials will the negatives of the other half. Some could be quickly ruled out because the quadratic must be monic (so only the top row of $\mathtt{M}^{-1}$ need be calculated) and also the product of the constant terms must be 15. Nevertheless the amount of calculation could be daunting.

3 Congruences

Two integers a and b are said to be congruent with respect of modulus m when they leave the same remainder r upon division by m. Two polynomials are congruent with respect of a modulus polynomial $M(x)$ when they leave the same remainder $R(x)$ upon division by $M(x)$. Two congruent integers therefore have a difference which is by a multiple of m, and two congruent polynomial have a difference which is some polynomial $G(x)$ multiplied by $M(x)$: $a \equiv b \implies a - b = \lambda m$ where λ is an integer or polynomial . Three important facts are:

1. Regarding terminology, $J(x) \equiv R(x) \mod H(x)$ means that $R(x)$ is the remainder after J has been divided by H. If $H(x) = x^4 + 2x^3 + x + 2$, for instance, $R(x)$ is the result of setting $x^4 = -(2x^2 + x + 2)$ in $P(x)$.

2. The equation $ax \equiv b \mod m$ has a solution only if $\gcd(a, m)$ divides b. If this condition is satisfied, there are $d = \gcd(a, m) \mod m$ solutions. For example, if $5x \equiv 1 \mod 3$, $\gcd(5, 3) = 1$ and so there is unique solution modulo 3, namely $x \equiv 2$. On the other hand $3x \equiv 3 \mod 6$ has the two solutions $x \equiv 1,\ 3 \mod 6$ whilst $3x \equiv 2 \mod 6$ has no solutions.

3. If $ka \equiv kb \mod m$, then $a \equiv b \mod \left(\frac{m}{g}\right)$ where $g = \gcd(k, m)$. A special case is where k and m are coprime when the K can be cancelled giving $a \equiv b \mod m$.

Congruences, remainders and residue classes are at the foundation of the theory of finite fields on which main-stream factorisation algorithms are built.

3.1 The Chinese remainder theorem

This is a theorem about an integer x which is known to satisfy simultaneously two or more congruences of the form $x \equiv r_1 \mod m_1$, $x \equiv r_2 \mod m_2$, ..., $x \equiv r_k \mod m_k$. Its history goes back almost 2,000 years. I will illustrate it in much the same way as it is believed to have been used originally. Here is a story:

> At a junior school the head teacher wants to know how many children are on the playground. Instead of trying to count them one by one as they mill around, she asks them to stand in groups of four. If any child cannot join a group of four, he or she is to come forward and be counted. Only one child does so. She then asks all the children to rearrange themselves into groups of five, and again for those who cannot find a group to join to come forward. Two do so. Finally she asks them to form groups of seven, whereupon five children remain over. From this the Head calculates that there are 117 children present. How did she do it?

The problem can be cast as follows. Let x be number of children. We know that

$$x \equiv 1 \mod 4 \text{ so } x = 4\beta_1 + 1, \qquad x \equiv 2 \mod 5 \text{ so } x = 5\beta_2 + 2, \qquad x \equiv 5 \mod 7 \text{ so } x = 7\beta_1 + 5$$

where the β are integers. The Chinese remainder theorem states that there is a solution to this set of congruences modulo the product of the individual moduli.

Modular arithmetic has analogies with other periodic phenomena such as wave motion, where the modulus is equivalent to period or wavelength and the remainder to the phase. Simultaneous congruences have analogies with wave interference, the beating of musical notes which are just out of tune with each other, and with Moiré patterns formed between parallel sets of railings seen one behind the other. In Figure 2 I have plotted the numbers from 70 to 120 which respectively satisfy one of these three congruences. We can see one or other of the three pairs coinciding, but all there come together only at 117.

Figure 2: Numbers x satisfying remainder 1 mod 4 (green), remainder 2 mod 5 (blue), remainder 5 mod 7 (red).

It is clear that any one congruence $x \equiv r \mod m$ advances in steps of m, and that two with moduli m_1, m_2 will synchronise every l, the LCM (least common multiple) of m_1, m_2. If they are coprime, as with the pairs (4, 5), (4, 7) and (5, 7) in the story above, this will be at spacing $l = m_1 m_2$. This is why the (mod 4, mod 5) pair in Figure 2 coincide every 20 units, the (4, 7) pair every 28 units, and the (5, 7) pair every 35. The trio will coincide only every 4.5.7 = 140 units. So if they were to start in phase at 0 (all remainders zero), they would coincide next at 140. In our case the mod 7 condition commenced 5 units previously, which would be equivalent to moving the red markers leftwards by 5 to the positions they occupy in Figure 2, spanning −5 to 135 instead of 0 to 140. Similarly the blue curve is moved back by 2, and the green by 1. The three now coincide at a position L less than 140 where $L = 4\gamma_1 - 1 = 5\gamma_2 - 2 = 7\gamma_3 - 5$ where the γ are integers. On making three short lists {3, 7, 11, 15, 19, 23, 27, ... }, {3, 8, 13, 18, 23, 28 ...} and {2, 9, 16, 23, 30, ...} we spot that the first common number of 23. So the number of children on the playground was 140 − 23 = 117. This is not the only approach to a solution, but does illustrate that lists of suitable integers may be necessary.

The point about the Chinese remainder theorem is that the solution of n simultaneous linear congruences $\mod m_1$, $\mod m_2$, ..., $\mod m_n$ is that it is unique modulo $m_1m_2m_3.....m_n$ provided all pairs m_j, m_k are mutually prime.

There is a formula for the solution of simultaneous congruences. If $x \equiv a_j \mod m_j$, $1 \le j \le n$ and all the m_j are coprime,

$$x \equiv a_1b_1c_1 + a_2b_2c_2 + ... + a_nb_nc_n, \qquad M = \prod_1^n m_j, \quad b_j = \frac{M}{m_j}, \quad b_jc_j \equiv 1 \mod m_j. \qquad (6)$$

The $\gcd(b_j, m_j) = 1$, but $\gcd(b_k, m_j) = 0$, $j \ne k$ because m_j is a factor of every b_k except b_j. Hence taking $x \mod m_j$ picks out only the term in j and then the condition $b_jc_j \equiv 1 \mod m_j$ means that $x \equiv a_j$ as required.

Applying this to the case above, $M = 4.5.7 = 140$ and

- $a_1 = 1$, $m_1 = 4$, $b_1 = 5.7 = 35 \equiv 3 \mod 4$ so $c_1 = 3$,
- $a_2 = 2$, $m_2 = 5$, $b_2 = 4.7 = 28 \equiv 3 \mod 5$ so $c_2 = 2$,
- $a_3 = 5$, $m_3 = 7$, $b_3 = 4.5 = 20 \equiv 6 \mod 7$ so $c_3 = 6$.

Therefore $x \equiv 1.35.3 + 2.28.2 + 5.20.6 = 817 \equiv 117 \mod 140$ as found above.

The method carries over to polynomials modulo a polynomial. For example, what is $P(x)$ defined over $\mathbb{F}_3$ if

$$P(x) \equiv x + 1 \mod (x^3 + x^2 + 2) \qquad \text{and also} \qquad \equiv x^2 + 2 \mod (x^3 + 2x + 2) \ ?$$

Here $a_1 = x + 1$, $b_1 = m_2 = x^3 + 2x + 2$. The challenge is to find c_1 such that $b_1c_1 \equiv 1 \mod m_1$. Set $x^3 \equiv -x^2 - 2$ and search through possible values until $c_1 = 2x$ is found. Similarly $a_1 = x^2 + 2$, $b_2 = m_1 = x^3 + x^2 + 2$. Now set $x^3 \equiv -2x - 2$ and search the field until we find that $c_2 = x + 2$. The formula then gives

$$P(x) \equiv (x + 1)(x^3 + 2x + 2).2x + (x^2 + 2)(x^3 + x^2 + 2)(x + 2) \equiv x^6 + 2x^5 + x^2 + 2x + 2.$$

Division by m_1 gives $P(x) = (x^3 + x^2 + 2x + 2)(x^3 + x^2 + 2) + x + 1$, and division by m_2 gives $(x^3 + 2x^2 + x)(x^3 + 2x + 2) + x^2 + 2$. Any multiple of $M(x) = m_1m_2$ may be added to this $P(x)$ and it will remain true.

3.2 Quadratic residues

'Residue' is another name for a remainder upon division. Thinking of integers, a quadratic residue modulo a prime p is a remainder which is the square of itself or another integer. Thus a is a quadratic residue mod p if there is b such that $b^2 \equiv a$. Put another way, quadratic residues are numbers of the field $\mathbb{F}_p$ which have two square roots in the same field. Here are the squares of integer 0 to 6 modulo 7:

$$0^2 = 0, \qquad 1^2 = 1, \qquad 2^2 = 4, \qquad 3^2 = 2, \qquad 4^2 = 2, \qquad 5^2 = 4, \qquad 6^2 = 1,$$

so 0, 1, 2 and 4 are quadratic residues, while 3, 5 and 6 are not. The convention is to omit 0 from the reckoning. Modulo 11 the sets of quadratic residues and non-residues are:

$$\text{Residues : } 1,\ 3,\ 4,\ 5,\ 9. \qquad \text{non-residues : } 2,\ 6,\ 7,\ 8,\ 10.$$

There are as many non-residues as residues: $(p-1)/2$ of each. The negative of each residue a is a non-residue $p-a$. The following fact arises from Fermat's theorem, explained in §5:

if a is a quadratic residue, $a^{(p-1)/2} \equiv 1 \mod p,$, but if a is a non-residue, $a^{(p-1)/2} \equiv -1 \equiv p-1$.

A more general version of this is that if k is an odd integer:

if a is a quadratic residue, $a^{(p^k-1)/2} \equiv 1 \mod p,$, but if a is a non-residue, $a^{(p^k-1)/2} \equiv -1 \equiv p-1$. (7)

If, however, k is even, $k = 2m$, $a^{(p^{2m}-1)/2} \equiv +1 \mod p$ for all a in the field.

Quadratic residues have their close counterparts in polynomials. A knowledge of finite fields and Fermat's little theorem is necessary to understand this, so is deferred to §5. Both for integers and polynomials, the quadratic residues form a subgroup of the non-zero elements of the field, each with order $(q-1)/2$ where q is the number of element in the field.

4 Finite fields

Many algorithms related to factorisation rely on the arithmetic of finite fields. There are excellent articles in Wikipedia and elsewhere on this essential subject, so here are only a few points. The elements of finite fields have properties related to Fermat's little theorem which is the subject of the following section. Appendix 2 gives detailed examples of the formation, isomorphism, and factorisation properties of two sample finite fields.

Recall that a field is a structure such as the rationals, $\mathbb{Q}$, consisting of elements over which two binary operations, addition and multiplication, are defined. Under addition the field elements form an abelian group, meaning that the order of combining any two elements a, b does not affect the result: $a + b = b + a$. There is a unique identity element for addition denoted 0. Similarly under multiplication the elements excluding 0 form an abelian group: $a * b = b * a$. A field has no 'zero divisors' – non-zero numbers which when multiplied together give a product equivalent to zero in the field[1]. There is a unique multiplicative identity, 1, and it is possible to divide any chosen element by any other except 0.

The simplest finite fields, called 'prime fields', are the integers modulo any prime p: $\{0, 1, 2, 3, ..., p-1\}$ where $p \equiv 0$. This is denoted[2] $\mathbb{F}_p$. By $\mathbb{F}_p^*$ is meant the multiplicative group $\{1, 2, 3, ..., p-1\}$. Each number in the field represents an equivalent class of remainders after division by p. Some numbers within the field, called 'primitive elements' will generate the whole multiplicative group $\mathbb{F}_p^*$ by being raised to increasing powers. For instance, if $p = 7$, 3 is primitive since $3^1 \equiv 3$, $3^2 \equiv 2$, $3^3 \equiv 6$, $3^4 \equiv 4$, $3^5 \equiv 5$, $3^6 \equiv 1$. $5 \equiv -3$ is also primitive mod 7. Division is equivalent to solving a congruence. Thus $y = 2/3 \mod 7$ implies $3y \equiv 2 \mod 7$ so $y = 3 + 7k$ is a solution for any integer k, and 3 is the representative of $2/3 \mod 7$.

There are some seemingly strange results when polynomials are solved over a finite field. Whilst in $\mathbb{Z}$ $x^2 - 4$ would have the solutions $x = \pm 2$, in modulo 5 the solutions are $x = 2$, $x = 3$ because $x^2 - 4 \equiv x^2 + 1 = (x+3)(x+2) \mod 5$. Similarly $x^2 - 2 \mod 7$ has the solution $x = 3$, $x = 4$ because $x^2 - 1 \equiv x^2 + 5 = (x+4)(x+3) \equiv (x-3)(x-4) \mod 7$. So what would be $\pm\sqrt{2}$ in $\mathbb{Z}$ has

[1] For example, if we take the integers modulus 6, that is 0, 1, 2, 3, 4, 5, then $2 * 3 = 6 \equiv 0 \mod 6$, so both 2 and 3 are zero divisors. Integers mod p for p prime have no zero divisors and so form a finite field.

[2] It is isomorphic to the set of quotients $\mathbb{Z}/p\mathbb{Z}$, that is the integers modulo p, which form a set of equivalence classes $2 \equiv 2 + p \equiv 2 + 2p \equiv ...$, $3 \equiv 3 + p \equiv 3 + 2p \equiv ...$,.

become 3 and 4 mod 7; $3^2 = 9 \equiv 2$ and also $4^2 = 16 \equiv 2$. Roots which would be complex in $\mathbb{C}$ may be integers in a prime field. One example is $x^2 + 4 = 0$ which is $x = \pm i\sqrt{2}$ in $\mathbb{C}$, but factorises as $(x+4)(x+1) = x^2 + 5x + 4 \equiv x^2 + 4 \mod 5$.

Division of one polynomial by another generally produces a rational function rather than a third polynomial. Therefore polynomials themselves cannot form a field but rather a ring, an algebraic structure with looser properties. Nevertheless a finite field can be created from polynomials by

1. starting with all polynomials of all degrees with coefficients in a field $\mathbb{F}$ and
2. dividing them by a polynomial $K(x)$ which is irreducible over $\mathbb{F}$,
3. identifying the remainders as representative elements of the finite field.

The field, therefore, is polynomials modulo $K(x)$. If $\mathbb{F}$ were $\mathbb{Q}$, an infinite field would be generated by any modulus $K(x)$; it would consist of all polynomials with rational coefficients whose degree was in the range 0 to $k-1$. However by working modulo a prime, with coefficients in $\mathbb{F}_p$, and choosing $K(x)$ to have degree k, a field with p^k elements will be created. Here is how it works; Appendix 2 gives more detail using as examples 2^3 and 3^2.

Take the case $p = 3$, $k = 2$, a field of 9 elements. We work modulo 3, that is in $\mathbb{F}_3$, and choose an irreducible polynomial $K(x)$ of degree $k = 2$. There are 18 quadratics modulo 3. These can be listed by their coefficients, with x^2 in the left-most column: $x^2 = (1\ 0\ 0)$, $x^2 + 1 = (1\ 0\ 1)$, then (1 0 2), (1 1 0), ... (2 2 2). From this list all that are divisible by a degree-1 irreducible are deleted. The linear irreducibles modulo 3 are x, $x+1$, $x+2$, $2x+1$. Quadratics such as $x^2 + x = (1\ 1\ 0)$ that are divisible by x can be deleted. Also $x^2 + 2 \equiv (x+1)(x+2) \mod 3$ and $2x^2 + 1 = (x+1)(2x+1)$ so are also deleted. Eventually the list of irreducible quadratics modulo 3 is whittled down to

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

It does not matter which of these is chosen as $K(x)$ provided it is used consistently. The set of remainders for any $\mathbb{F}_p[x]/K(x)$ is the set of linear irreducibles modulo 3: $\{0, 1, x, x+1, x+2, 2x+1\}$. The field consists of all elements formed by all additions and multiplications of these generating elements reduced both mod 3 and mod $K(x)$. For any $K(x)$ the field elements are

$$\{\ 0,\ \ 1,\ \ 2,\ \ x,\ \ x+1\ \ x+2,\ \ 2x,\ \ 2x+1,\ \ 2x+2\ \}.$$

A philosophical word is appropriate here, to do with the interpretation of x. x is both a place-holder to distinguish the coefficients and also a root of $K(x) = 0$. Since $K(x)$ is irreducible, its roots lie in some extension number field. You might prefer to write the field elements as

$$\{\ 0,\ \ 1,\ \ 2,\ \ \beta,\ \ \beta+1\ \ \beta+2,\ \ 2\beta,\ \ 2\beta+1,\ \ 2\beta+2\ \}$$

thought is makes no practical difference. We do not need to know the actual values of these roots, but to reduce products of elements modulo the chosen $K(x)$. For $K(x) = x^2 + 1$, $x^2 \equiv -1 \equiv p - 1$ and for $x^2 + 2x + 2$, $x^2 \equiv -2x - 2 \equiv x + 1 \mod 3$. In the coefficient-only notation introduced above the field is

$$(0\ 0)\quad (0\ 1)\quad (0\ 2)\quad (1\ 0)\quad (1\ 1)\quad (1\ 2)\quad (2\ 0)\quad (2\ 1)\quad (2\ 2).$$

Indeed this coefficient-only format shows that the field with p^k elements is similar to the direct product of two unnamed quantities.

To illustrate that it is closed under multiplication note that $(2x+2)(x+2) \equiv 2x^2+1 \equiv x \mod (x^2+x+2) \equiv 2x \mod (x^2+2x+2)$. The reciprocal of x is $(x+1) \mod (x^2+x+2)$ because $x(x+1) \equiv 1$. However $1/x = (x+2) \mod (x^2+2x+2)$, showing that the field arithmetic is determined by the defining irreducible $K(x)$, but are isomorphic. The field $\{0,1,2\}$ is called its 'prime subfield'. Using the similarity with a direct product we can write down the elements in the field $\mathbb{F}_8$ using $8 = 2^3$:

$$(0\ 0\ 0)\quad (0\ 0\ 1)\quad (0\ 1\ 0)\quad (0\ 1\ 1)\quad (1\ 0\ 0)\quad (1\ 0\ 1)\quad (1\ 1\ 0)\quad (1\ 1\ 1).$$

We see that the only finite fields have either p or p^k elements, $k = 2, 3, 4, \ldots$ Those with p elements are represented by integers modulo p, that is by polynomials of degree 0. Those with p^k elements are represented by polynomials of degree $k-1$ and less. All fields with the same number of elements are isomorphic, so it is correct to speak of *the* field of, say, 27 elements rather than *a* field of order 27. Appendix 2 gives an example of constructing a finite field with two irreducible polynomials and shows explicitly the isomorphism between them.

One detail will be useful in §8 on Hensel lifting. Suppose that $a \equiv 0 \mod c^2$. This means that $a = jc^2$ for some integer j. If the equation is divided throughout by c we get $a/c = jc$ which is equivalent to 0 modulo c. So

$$a \equiv 0 \mod c^2 \quad \text{implies} \quad \frac{a}{c} \equiv 0 \mod c. \tag{8}$$

The converse is not generally true.

A further significant theorem, a consequence of the Chinese remainder theorem, is that if $K(x)$ creates a finite field of residues $\mathbb{F}_p[x]/K(x)$, and if $K(x) = J_1(x)J_2(x)...J_m(x)$ where J_1 etc. are each irreducible modulo p, then $\mathbb{F}_p[x]/K(x)$ is the direct product of all subfields $\mathbb{F}_p[x]/J_j(x)$:

$$\frac{\mathbb{F}_p[x]}{K(x)} \cong \frac{\mathbb{F}_p[x]}{J_1(x)} \times \frac{\mathbb{F}_p[x]}{J_2(x)} \times \ldots\ldots \times \frac{\mathbb{F}_p[x]}{J_m(x)}. \tag{9}$$

This is really a statement of the Chinese remainder theorem. The $J_j(x)$ are co-prime so

$$\mathbb{F}_p[x] \mod (j_1 j_2 \ldots j_m) = (\mathbb{F}_p[x] \mod j_1) \text{ and } (\mathbb{F}_p[x] \mod j_2) \text{ and } \ldots \text{ and } (\mathbb{F}_p[x] \mod j_m).$$

5 Fermat's Little Theorem

Pierre de Fermat was an older contemporary of Isaac Newton. He made major contributions to number theory including his Little Theorem, so called to distinguish it from the notoriously challenging Last Theorem (that $x^n + y^n = z^n$ has no solution in integers if $n > 2$).

5.1 The theorem for integers

The Little Theorem states that if p is prime and a a positive integer which does not have p as a factor, then

$$a^p \equiv a \mod p \qquad \text{which is equivalent to} \qquad a^{p-1} \equiv 1 \mod p. \tag{10}$$

The logic behind this congruence is as follows. With p prime none of $1, 2, \ldots, p-1$ divides p and so their product $(p-1)!$ is not divisible by p. So let $(p-1)! \equiv q$ where q will be one[3] of the integers $1, 2, \ldots, p-1$. Now multiply each of $1, 2, \ldots, p-1$ by a. The set $\{a, 2a, 3a, \ldots (p-1)a\} \mod p$ will

[3] Actually $q \equiv p-1$: Wilson's theorem.

have the same values as $1, 2, ..., p-1$ though in a different order. To illustrate this look at the two cases $a = 4$, $a = 9$ modulo 7:

m	1	2	3	4	5	6
$4m$	4	8	12	16	20	24
$4m \mod 7$	4	1	5	2	6	3
$9m$	9	18	27	36	45	54
$9m \mod 7$	2	4	6	1	3	5

Multiply the numbers ma together to form $a.2a.3a.....(p-1)a = (p-1)!a^{p-1}$. Then $(p-1)! \equiv q \equiv (p-1)!a^{p-1} \mod p$. Dividing by $(p-1)!$ leaves $a^{p-1} \equiv 1 \mod p$.

Leonard Euler showed that this can be extended to cases where n is not prime but a and n are coprime, that is $\gcd(a, n) = 1$. He introduced his 'totient function' $\phi(n)$ which is the number of integers $< n$ and coprime to n. In this more general version

$$a^{\phi(n)} \equiv 1 \mod n. \tag{11}$$

Some examples will consolidate these simple but powerful results. We use $\phi(4) = 2$ and $\phi(8) = 4$:

$$\begin{array}{llcll} a=3 & n=5 & 3^5 = 243 & \equiv & 3 \mod 5 \\ a=4 & n=7 & 4^7 = 16384 & \equiv & 4 \mod 7 \\ a=3 & n=4 & 3^2 = 9 & \equiv & 1 \mod 4 \\ a=7 & n=8 & 7^4 = 2401 & \equiv & 1 \mod 8 \end{array}$$

The theorem may be used as a short cut to calculating the value of ridiculously large integers modulo a prime. For example $14^{101} \approx 5 \cdot 74 \times 10^{115}$, but $14^{101} \mod 47$ can be found fairly simply using Fermat's theorem. $101 = 2 \times 46 + 9$ so $14^{101} \mod 47 = 14^9 \mod 47$. Next $9 = 2^3 + 1$ and $14^2 \mod 47 \equiv 8$ so $14^4 \equiv 64 \equiv 17$ and $14^8 \equiv 17^2 = 289 \equiv 7$. Finally $14^9 \equiv 7 \times 14 = 98 \equiv 4 \mod 47$.

A slightly different way to look at the theorem is that the equation $x^p - x = 0 \mod p$ has p roots $a_j = 0, 1, 2, 3, \ldots, p-1$ in the finite field $\mathbb{F}_p$. These integers form a field of p congruence classes $\mathbb{Z}/p\mathbb{Z}$. In it $x^p - x \equiv x^p + (p-1)x \mod p$ factorises ('splits') into purely linear factors:

$$x^p - x \;=\; x(x-1)(x-2)....(x-(p-1)) \;\equiv\; (x+(p-1))(x+(p-2))....(x+1)\,x \mod p. \tag{12}$$

Trivially $x = 0$ is a root. The other roots a_j follow from Fermat's theorem because each of $a = 0$ or 1 or 2, 3, ... , $p = 1$ is coprime to p and so $a_j^p \equiv a_j \mod p$ for each field element a_j, $0 \le j < p$. This is an extension to finite fields of Eq 2 of §1.2.

In a finite field all p roots also satisfy $a^M - a \equiv 0$ where $M = p + m(p-1)$, m any integer. The reason is that $a^p \equiv a \mod p$ and $a^{p-1} \equiv 1$ so $a^M = a^p.(a^{p-1})^m \equiv a.1^m \equiv a$. In contrast, for a general exponent j, $a^j - a = 0$ is satisfied by only a few of the integers 0 to $p-1$. The sequence $a^M - a \equiv 0$ where $M = p + m(p-1)$ includes $a^{p^2} - a$ when $M = p$, $a^{p^3} - a$ when $M = p(p+1)$ and $a^{p^4} - a$ when $M = p(p^2 + p + 1)$, and so forth, $a \in \{0, 1, 2, ... p-1\}$. All of the equations

$$x^{p^k} - x = 0 \mod p \text{ have the full set of } p \text{ integer roots } a = 0, 1, 2, ...p-1. \tag{13}$$

The polynomial $x^{p^k} - x \equiv 0 \mod p$ occurs many times in the theory of factorisation; we will meet it in §9 on the Cantor-Zassenhaus algorithm

Fermat's theorem can explain why the number of quadratic residues (see §3.2) equals the number of non-residues modulo a prime p.

$$x^{p-1} - 1 = \left(x^{(p-1)/2} + 1\right)\left(x^{(p-1)/2} - 1\right). \tag{14}$$

Now $x^{p-1} \equiv 1$ has exactly $p-1$ roots and these must be distributed between these two factors. Moreover, neither can have more than $\frac{1}{2}(p-1)$ roots since this is the degree of each polynomial factor. So the roots are distributed equally, giving $\frac{1}{2}(p-1)$ to each.

5.2 Application to polynomials

Fermat's theorem has important applications to polynomials and their factorisation. The binomial theorem for integers a and b takes on a simple form modulo a prime. Over $\mathbb{Z}$

$$(a+b)^n = {}^nC_0 a^n b^0 + {}^nC_1 a^{n-1} b^1 + {}^nC_2 a^{n-2} b^2 + {}^nC_3 a^{n-3} b^3 + \ldots + {}^nC_{n-1} a^1 b^{n-1} + {}^nC_n a^0 b^n.$$

The coefficients are ${}^nC_r = \frac{n!}{r!(n-r)!}$. If $n = p$, a prime, the p is not cancelled between the factorials in numerator and denominator so the numerator retains p as a factor, unless $r = 0$ or $r - n = 0$ in which cases ${}^nC_0 = {}^nC_n = 1$. Therefore $(a+b)^p \equiv a^p + b^p \mod p$. This deletion of cross terms in the expansion continues to multinomials $(a+b+c)^p = (a+(b+c))^p = a^p + (b+c)^p = a^p + b^p + c^p$ and in general

$$(a+b+c+\ldots+x+y+x)^p \equiv a^p + b^p + c^p + \ldots + x^p + y^p + z^p \mod p. \tag{15}$$

In the USA this equation is called 'the Freshman's dream', I suppose because novices to algebra hopefully suppose that it applies in $\mathbb{Z}$ too.

We can look at this is slightly different way. §4 on finite fields shows that any polynomial $K(x)$ of degree k which is irreducible over F_p defines a field through its quotient $\mathbb{F}_p[x]/K(x)$ with the ring of all polynomials over F_p. The field will have p^k elements because the indeterminates 1, x, x^2, ..., x^{k-1} form a basis for the field and each can be weighted by any integer 0, 1, 2, ... $p-1$ so that any field member is a linear combination of the base vectors; that is, it looks like a polynomial in x of degree $< h$ but can be thought of as a vector. The specific case of $\mathbb{F}_{2^3}$ is examined in Appendix 2. The non-zero members of this field form a cyclic group under multiplication. Some elements will be 'primitive', meaning that their powers generate all the other elements of the group. The order of the primitive elements is $p^k - 1$. Others, members of a cyclic subgroup, will have an order which divides $p^k - 1$ by Lagrange's theorem. Either way, every element except 0 will satisfy

$$x^{p^k-1} \equiv 1 \mod p, \text{ and all elements satisfy } x^{p^k} - x \equiv 0. \tag{16}$$

This is Eq 11 of §5.1. It is important in factorisation algorithms.

It is shown in Appendix 2, §A2.3 that the field-defining polynomial $K(x)$ in $\mathbb{F}_p[x]/K(x)$ divides $x^{p^k} - x$. Indeed, any of the other irreducible polynomials with the same degree as $K(x)$ which could define the same or at least an isomorphic field will also divide into $x^{p^k} - x$. This means that $x^{p^k} - x$ includes in its factors all irreducibles of degree k with coefficients modulo p.

An even stronger statement can be made: the polynomial $x^{p^k} - x$ can be divided by all irreducibles $H(x)$ of lower degree h provided that $h|k$. So if $k = 8$, not only would all irreducible polynomials of degree 8 divide into it, but also those with degrees 4, 2 and 1. This remarkable theorem is the basis of the distinct degree algorithm, §6. A proof is given on §A3.4 of Appendix 3.

It is perhaps not obvious that the symbol x in Eq 12 can replaced by a polynomial $u(t)$ and so provide one factorisation of $u(t)^p - u(t)$ in which each factor steps by 1 from the previous. For example if $u(t) = t^2 + t + 2$ and $p = 3$,

$$(t^2+t+2)^3 - (t^2+t+2) \;=\; (t^2+t+2)(t^2+t+1)(t^2+t) \;=\; t^6+t^3+2t^2+2t \mod 3. \qquad (17)$$

By Eq 18 the term $(t^2+t+2)^3 \mod 3 \equiv t^6+t^3+2$, and $-(t^2+t+2) \equiv 2t^2+2t+1$. By adding these the rightmost expression follows. Note that Fermat's theorem has *not* been applied to reduce t^3 to t, though obviously substituting any of 0, 1, $p-1$ for t will give zero. These principles were developed by Elwyn Berlekamp in the mid 1960s into what became a standard algorithm for machine factorisation of polynomials over a finite field. It is described in §7.

At Eq 12 it was demonstrated for integers that the remarkable polynomial $x^p - x$ is the product of all linear factors in which the roots are all the elements of the field $\mathbb{Z}/p\mathbb{Z}$, namely the integers 0 to $p-1$. This factorisation extends to polynomials in a field $\mathbb{F}_p[x]/K(x)$ defined by some irreducible $K(x)$ of degree k. To be clear, in these field we are working both modulo p and modulo $K(x)$, retaining only the remainders of expressions on division by these two defining quantities. As an example, take $p = 3$ and $K(x) = x^3 + x^2 + 2 = (1\ 1\ 0\ 2)$. If a is a root of $x^3 + x^2 + 2$, modulo $K(x)$ is equivalent to $a^3 = -a^2 - 2$. The product of all 27 linear factors is

$$x(x-1)(x-1)(x-2)(x-a)(x-a-1)(x-a-2)(x-2a)(x-2a-1)(x-2a-2)....$$
$$.....(x-2a^2-2a)(x-2a^2-2a-1)(x-2a^2-2a-2)$$

and this evaluates to $x^{27} - x$, that is to $x(x^{3^3-1} - 1) = x(x^{p^k-1} - 1)$.

In §3.2 we introduced quadratic residues amongst the integers. Using Fermat's little theorem it can be shown that they carry over to polynomials. We have established that if $K(x)$ is irreducible modulo a prime p with degree k, then the residue classes of $\mathbb{F}_p[x]/K(x)$ form a field with p^k elements. There are exactly p^k solutions to the congruence $x^{p^k-1} \equiv 1 \mod p$ and these are the members of this field. Following the approach of §3.2 we look at the square roots $x^{(p^k-1)/2}$ and find that, with 0 not included, half the field give the value +1 and the other half −1. As an example, again take $p = 3$ and $K(x) = x^3 + x^2 + 2 = (1\ 1\ 0\ 2)$. The table below gives the signs of the 26 non-zero elements a under the operation $a^{(3^3-1)/2} = a^{13}$. There are 13 of each polarity. The table shows a similarity

0 0 0		1 0 0	+	2 0 0	−
0 0 1	+	1 0 1	−	2 01	+
0 0 2	−	1 0 2	−	2 0 2	+
0 1 0	+	1 1 0	+	2 1 0	+
0 1 1	+	1 1 1	+	2 1 1	−
0 1 2	−	1 1 2	+	2 1 2	−
0 2 0	−	1 2 0	−	2 2 0	−
0 2 1	+	1 2 1	+	2 2 1	−
0 2 2	−	1 2 2	+	2 2 2	−

to Eq 6 of §3.2. Those polynomials giving +1 are the quadratic residues and those giving −1 the non-residues. The former have two square roots in the field, the non-residues have their square roots in some higher field extension. A few examples are :

$(2\ 0\ 2)^2 = (1\ 0\ 1)^2 = (0\ 1\ 0)$, $\qquad$ $(0\ 1\ 2)^2 = (0\ 2\ 1)^2 = (1\ 1\ 1)$,
$(1\ 1\ 2)^2 = (2\ 2\ 1)^2 = (1\ 2\ 2)$, $\qquad$ $(2\ 0\ 1)^2 = (1\ 0\ 2)^2 = (2\ 1\ 0)$.

Ones such as (1 0 1) and (2 2 1) have no square root modulo 3 and modulo (1 1 0 2). As we might suspect, and as shown again in Appendix 2, A2.2, the product of the all linear terms with the +1 elements of $\mathbb{F}_p[x]/K(x)$, excluding 0, is

$$(x-1)(x-a)(x-a-1)(x-2a)(x-2a-1)(x-a^2)(x-a^2-a)(x-a^2-a-1)\times \\ (x-a^2-a-2)(x-2a^2)(x-2a^2-1)(x-2a^2-2)(x-2a^2-a).$$

This evaluates to $x^{13}-1$. The coset of these roots gives the following product of linear factors:

$$(x-2)(x-a-2)(x-2a)(x-2a-2)(x-a^2-1)(x-a^2-2)(x-a^2-2a)\times \\ (x-2a^2-a-1)(x-2a^2-a-2)(x-2a^2-2a)(x-2a^2-2a-1)(x-2a^2-2a-2)$$

which evaluates to $x^{13}+1$. Clearly the product of these is $x^{26}-1$ as found above.

6 Distinct degree factorisation

'Distinct degree' means that the algorithm will split a given $P(x)$ into blocks of irreducible factors within each of which the factors have the same degree. The degrees in one block will all be h_1, in the next all h_2 and so forth. The algorithm is therefore a coarse filter and so is used as an intermediate stage in several computer factorisation procedures over finite fields. It is built on the fact that $x^{\mathcal{N}}-x$ for $\mathcal{N}=p^k$ in $\mathbb{F}_p$ is the product of every possible monic irreducible polynomial whose degree h divides k, and of no others. This essential theorem is explained and a proof given in Appendix 3, §A3.4. The algorithm takes $P(x)$ and calculates its gcd with $x^{p^h}-x$ for different integers h. For each h all irreducible factors of $P(x)$ whose degree divides h will be picked up in the gcd. It is as simple and yet as clever as that!

I will illustrate the method first with $P(x)=x^{11}+x^9+2x^8+2x^7+x^6+4x^7+4x^4+x^3+x^2+4x+1$ mod 5 or (1 0 1 2 2 1 4 4 1 1 4 1) mod 5. It is relatively easy to find linear factors since there are only five integers to substitute. Thus $x=2$ and 3 are both zeros, making $x+3$ and $x+2$ respectively factors. Nevertheless, if the $\gcd(x^5-x,P(x))$ is calculated, it is found to be (1 0 1) and $x^2+1=(x+3)(x+2)$ mod 5, so both linear factors have been found, admittedly as their product. The calculation of high powers of x is done efficiently by algorithms which partition the exponent and recursively build from x, x^2, x^4, *etc.* by repeated squaring. This is a significant topic in its own right.

It is probably best now to divide $P(x)$ by x^2+1, but if instead the $\gcd(x^{25}-x,P(x))$ is calculated, it is found to be (1 3 0 3 4) = (1 3 4)(1 0 1). The irreducible quadratic factor x^2+3x+4 has therefore also been found. The quotient on dividing by factors found so far is $Q(x)$ =(1 2 0 4 0 3 3 4) and this is now tested for cubic factors by $\gcd(x^{125}-x,Q(x))$. The result is (1 2 0 1). This divides $Q(x)$ with quotient (1 0 0 3 4) which is also irreducible as substitution of 0 to 4 quickly proves. This simple process has given a complete factorisation of P(x):

(1 0 1 2 2 1 4 4 1 1 4 1) = (1 0 0 3 4)(1 2 0 1)(1 3 4)(1 3)(1 2) mod 5.

The second example is more realistic in that $P(x)$ mod 3 has a higher degree. The algorithm should split it into blocks each of which is the product of irreducible factors in $\mathbb{F}_3$ of the same degree. The non-zero coefficients are listed here:

power	31	3	29	28	27	26	25	24	23	22	21	2	19	18	17	16
coeffn	1			1			2	1	1		2	1	2			2

power	15	14	13	12	11	1	9	8	7	6	5	4	3	2	1	0
coeffn			1	1	2	1	1	1	1	2		2	2	1		1

We calculate in turn the greatest common divisor of $P(x)$ and $x^{3^k} - x$ starting at $k = 1$. Factors found at each stage are divided out to leave a reduced polynomial.

1. Since modulo 3 there are only three numbers to test for integer roots, it is easy to show that there are no linear factors. Nevertheless we follow the algorithm and calculate $\gcd(P(x), x^3 - x)$. It is 1. The quotient is $Q_1(x) = P(x)$.

2. $\gcd(Q_1(x), x^9 - x) = x^6 + x^4 + x^2 + 1 = (1\ 0\ 1\ 0\ 1\ 0\ 1)$. This is the product of quadratic factors, so there must be three of them. Dividing this out from $Q_1(x)$ leaves quotient $Q_2(x) =$

power	25	24	23	22	21	20	19	18	17	16	15	14	13
coeffn	1		2	1		2	2	1		2		2	

power	12	11	10	9	8	7	6	5	4	3	2	1	0
coeffn	1					1		1	1	2			1

3. $\gcd(Q_2(x), x^{27} - x) = x^9 + 2x^8 + 2x^7 + 2x^6 + x^5 + 2x^4 + 2x^2 + 1 = (\ 1\ 2\ 2\ 2\ 1\ 1\ 0\ 2\ 0\ 1)$. This is the product of all polynomials of degree 3 so there are three of them. The next quotient is $Q_3(x)$

power	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
coeffn	1	1	1	1	2		2		2	2	2	2		2	1		1

4. The power of 16 in $Q_3(x)$ shows that there can be no polynomials of degree 5, 6 or 7. This, therefore, seems to be the product of four degree-4 polynomials. This is confirmed by $\gcd(Q_3(x), x^{81} - x) = Q_3(x)$.

The three gcds, for $k = 2$, 3 and 4 are reducible, but we have not yet described an algorithm capable of factoring them. The two in widespread use were devised by Berlekamp in the 1960s and by Cantor and Zassenhaus in the 1980s and are described in §7 and §9 respectively.

7 Berlekamp's factoring algorithm

At Eq 17 above we factorised $P(x) = x^6 + x^3 + 2x^2 + 2x \mod 3$ as $(x^2 + x + 2)(x^2 + x + 1)(x^2 + x)$ using the happy circumstance that $P(x) = f(x)^3 - f(x) \mod 3$ where $f(x) = x^2 + x + 2$. For any given $P(x)$ we are not likely to be able to express it precisely as $f(x)^p - f(x) \mod p$ for some $f(x)$ yet to be determined, and hence factorise it using Eq 16. However there may be more scope to find an $f(x)$ such that $P(x)$ merely divides $F(x) = f(x)^p - f(x) \mod p$. That would be progress because $\gcd(P, F) = P(x)$. Furthermore by Eq 12 and Eq 16 $F(x)$ factorises into p coprime factors $g_0 g_1 g_2 ... g_{p-1}$. We can therefore use $\gcd(P, F) = \gcd(P,\ g_0 g_1 g_2 ... g_{p-1}) = \gcd(P, g_0) \cdot \gcd(P, g_1) \cdot \gcd(P, g_2) \cdot$ ....., $\gcd(P, g_{p-1})$. The greatest common divisors can be found with Euclid's algorithm. This is the basis of Berlekamp's algorithm. The challenge is to craft the function $F(x) = f(x)^p - f(x)$ so that the given $P(x)$ will divide it without remainder. If $\partial(..)$ denotes the degree of a polynomial, we want $\partial(F) \geq \partial(P)$ but $\partial(f) < \partial(P)$.

A start can be made by proposing a general form of $f(x)$ from where the challenge moves to finding suitable coefficients. If $\partial(P) = n$, try

$$f(x) \;=\; x^{n-1} + a_{n-2}x^{n-2} + ... + a_1x + a_0, \qquad f^p \;=\; x^{p(n-1)} + a_{n-2}^p x^{p(n-2)} + ... + a_1^p x^p + a_0^p \mod p,$$

$$f(x)^p \;=\; x^{p(n-1)} + a_{n-2}x^{p(n-2)} + ... + a_1x^p + a_0 \mod p \;=\; f(x^p)$$

using Eq 15. On subtracting $f(x)$ the constant terms a_0 cancel.

The next step is to divide $F(x)$ by $P(x)$ and select coefficients which make the remainder zero. This will create a number of simultaneous equations in the a_j. A concrete example should help, so consider again the polynomial $P(x) = x^5 + x^4 - 9x^3 + x^2 - 15x + 15$ which features in §2.2 on Kronecker's method. There a search was made directly in $\mathbb{Z}$. Here we will explore factors modulo 2 and 3. In terms of coefficients $P(x)|_{mod\ 2} = (1\ 1\ 1\ 1\ 1\ 1)$ and in $F_2(x) = (a_4\ 0\ a_3\ 0\ [a_2 - a_4]\ -a_3\ [a_1 - a_2]\ -a_1\ 0)$ where the subscript denotes modulo 2 and $a_j = 0$ or ± 1. The remainder has degree 4 and is $([a_2 + a_4]\ a_3\ [a_1 + a_2 + a_4]\ a_1\ a_3)$ from which $a_1 = a_3 = 0$, $a_2 = a_4 = 1$ making $f(x) = x^4 + x^2$ and $F(x) = x^8 + x^2$ which factors as $f.(f-1) = (x^4 + x^2)(x^4 + x^2 + 1) = g_1.g_2$. We now need $\gcd(P, g_1)$ and $\gcd(P, g_2)$. In fact $g_2|P(x)$ so $\gcd(P, g_2) = g_2$. Euclid's algorithm also gives $\gcd(P, g_1) = x + 1$. The resulting factorisation is

$$P(x)|_{mod\ 2} \;=\; x^5 + x^4 + x^3 + x^2 + x + 1 \;=\; (x^4 + x^2 + 1)(x + 1). \tag{18}$$

The algorithm has worked even though this example has some issues. Observe that $(x^4 + x^2 + 1) = (x^2 + x + 1)^2 \mod 2$. The factorisation above is therefore not into irreducible polynomials. The first step in any factoring algorithm should be to remove squared factors by computing $\gcd(P, P')$. In this case $P'(x)$ divides $P(x) \mod 2$ with quotient $x + 1$, which features in Eq 18. This is best divided out.

I will repeat the process modulo 3 even though a factorisation is obvious: $P(x)|_{mod\ 3} = x^5 + x^4 + x^2 = x^2(x^3 + x^2 + 1)$. $f(x) = (a_4\ a_3\ a_2\ a_1\ a_0)$ again, but $F_3(x)$ has degree 12. The quotient by $P(x)$ therefore has degree 7. The remainder is $([2a_3 + a_2]\ [a_3 + 2a_2 + a_1]\ 0\ 2a_1\ 0)$ which will be zero if $a_1 = 0$, $a_2 = a_3 = 1$ or 2. a_4 does not feature in the quotient but must be 1 for $F_3(x)$ to be a monic polynomial of degree 12. We therefore have two choices for $f(x)$ which imply two factorisations of $F_3(x) = g_1.g_2.g_3$:

1. $f = x^4 + x^3 + x^2$, $F_3(x) = x^{12} + x^9 + x^6 + 2x^4 + 2x^3 + 2x^2 = (x^4 + x^3 + x^2)(x^4 + x^3 + x^2 + 2)(x^4 + x^3 + x^2 + 1)$,
2. $f = x^4 + 2x^3 + 2x^2$, $\quad F_3(x) = x^{12} + 2x^9 + 2x^6 + 2x^4 + x^3 + x^2$
   $= (x^4 + 2x^3 + 2x^2)(x^4 + 2x^3 + 2x^2 + 2)(x^4 + 2x^3 + 2x^2 + 1)$.

What matters are the gcd of $P|_{mod\ 3} = x^5 + x^4 + x^2$ with each of these factors.

1. $\gcd(P, x^4 + x^3 + x^2) = 2x^3 + x^2 \equiv 2(x^3 + 2x^2)$ (see Appendix 1[4]), $\gcd(P, x^4 + x^3 + x^2 + 2) = x^2 + 2x + 2$, $\gcd(P, x^4 + x^3 + x^2 + 1) = 1$, so $P|_{mod\ 3} = x^2(x + 2)(x^2 + 2x + 2)$.
2. $g_1 = x^4 + 2x^3 + 2x^2$ divides P with quotient $x + 2$ so $\gcd(P, g_1) = g_1$ and the factorisation is $x^2(x + 2)(x^2 + 2x + 2)$ again. The other two gcd are both 1.

The three main stages in this simple but ingenious algorithm are i) dividing F/P, ii) solving for the coefficients of f that make the remainder of F/P zero, and iii) applying Euclid's algorithm several times to find the various $\gcd(P, g_j)$. There can also be the complication that we saw above

[4] By convention the gcd is always taken to be monic.

of having more than one solution for the coefficients of $f(x)$, though all should give the same result. The polished implementations of Berlekamp's algorithm in commercial and open source academic software include numerous refinements to speed computer calculation. Clearly there are advantages in working with a low prime p because of the degrees $\partial(f) = \partial(P) - 1$ and $\partial(F) = p.\partial(f)$ which will create large polynomials.

One device at stages i) and ii) is to recognise that the remainder of F/P has degree $\partial(f) - 1$ or less. Therefore $F(x)$ can be separated into two sections, the first holding terms with degrees of $\partial(f)$ and larger, and the second holding those of lesser degree. Only the first section need be divided by $P(x)$, and this can we done one term at a time as $a_k x^k \div P(x)$. This will produce a remainder for each power of x in $P(x)$, and these can be added to the second section to give the overall remainder. Then linear equations can be set up in a matrix to solve for coefficients which make the remainder the zero polynomial modulo p. An example may explain this, so let us factorise the same $P(x) = x^5 + x^4 - 9x^3 + x^2 - 15x + 15$ modulo 11.

$P(x) = (1\ 1\ 2\ 1\ 7\ 4) \mod 11$. Again take $f = x^4 + a_3x^3 + a_2x^2 + a_1x + a_0$ so the first section of F is $x^{44} + a_3x^{33} + a_2x^{22} + a_1x^{11}$ and the second is $-(x^4 + a_3x^3 + a_2x^2 + a_1x)$. Now find the remainders for each term in the first section. With $a_4 = 1$ and $a_0 = 0$ they are[5]

$$\begin{array}{lll} x^{44} \div P: & r_{44} & = (7\quad 5\quad 3\quad 1\quad 10) \\ x^{33} \div P: & r_{33} & = (10\quad 1\quad 7\quad 7\quad 1).a_3 \\ x^{22} \div P: & r_{22} & = (0\quad 10\quad 10\quad 1\quad 6).a_2 \\ x^{11} \div P: & r_{11} & = (2\quad 9\quad 0\quad 3\quad 6).a_1 \end{array}$$

This gives the remainders on dividing f^{11} by P and the total remainders are obtained by subtracting f. The coefficient of each power of x is required to be 0. The linear relations amongst the coefficients a_3, a_2, a_1 in matrix form are

$$\begin{pmatrix} 7-1 & 10 & 0 & 2 \\ 5 & 1-1 & 10 & 9 \\ 3 & 7 & 10-1 & 0 \\ 1 & 7 & 1 & 3-1 \\ 10 & 1 & 6 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ a_3 \\ a_2 \\ a_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

I have made the subtraction of f explicit. The matrix is not square and so may not have a unique solution[6]. By elementary row operations it can be placed in echelon form:

$$\begin{pmatrix} 1 & 0 & 0 & 9 \\ 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

so $1+9a_1 \equiv 0 \mod 11$ making $9a_1 \equiv 10$, $a_1 \equiv 6$. Similarly, $a_3+8a_1 \equiv 0 \to a_3 \equiv 7$ and $a_2+3a_1 \equiv 0 \to a_2 \equiv 4$. The required polynomial $f(x)$ is therefore $(1\ 7\ 4\ 6\ 0)$ and the factorisation of $P(x)$ is the product of the gcds of $P(x)$ with each of the ten $(1, 7, 4, 6, k)$ for $0 \le k < 11$.

[5] The a_0 cancel between the two section of F. a_0 will later take all values from 0 to $p-1$ when the polynomial factors are created.

[6] If the same $P(x)$ is factored mod 7, seven solutions are found. In each $a_3 = 6$ but a_2 and a_1 are only determined to the extent that $a_2 + 2a_1 \equiv 2$.

We expect most of these gcds to be 1 on account of a theorem which states that the number of zero rows in the echelon form of the matrix above (called its 'nullity') is equal to the number of irreducible polynomial factors of $P(x) \mod p$. The list below gives the 11 versions of (1, 7, 4, 6, k) and their gcd with $P(x) = (1\ 1\ 2\ 1\ 7\ 4) \mod 11$.

$k = 0$ gcd = 1
$k = 1$ gcd = 1
$k = 2$ gcd = 1
$k = 3$ gcd = 1
$k = 4$ gcd = $(7\ \ 10\ \ 1)$
$k = 5$ gcd = 1
$k = 6$ gcd = 1
$k = 7$ gcd = 1
$k = 8$ gcd = 1
$k = 9$ gcd = 1
$k = 10$ gcd = $(7\ \ 8\ \ 0\ \ 9)$

Where the calculated gcd has been a constant, I have replaced it by 1 above. Also $(7\ \ 10\ \ 1) = 7(1\ 3\ 8) \mod 11$, and $(7\ 8\ 0\ 9) = 7(1\ 9\ 0\ 6)$ (see Appendix 1). We arrive at the factorisation

$$P(x) = (1\ 1\ 2\ 1\ 7\ 4) \ = \ (1\ 3\ 8)(1\ 9\ 0\ 6) \mod 11$$

and this is confirmed correct by multiplying the factors $(x^2 + 3x + 8)(x^3 + 9x^2 + 6)$.

Having obtained a factorisation modulo a low prime Hensel lifting (§8) is used to enlarge the field. In this case this is not necessary because in $\mathbb{Z}$ $(1\ 3\ 8)(1\ 9\ 0\ 6) = (1\ \ 12\ \ 35\ \ 78\ \ 18\ \ 48)$ in which all the coefficients exceed those in $P(x) = x^5 + x^4 - 9x^3 + x^2 - 15x + 15$ itself. The problem here is that negative coefficients are not produced by the algorithm so each could be in error by a multiple of the modulus, 11. The constant term of 48 exceeds the true value of 15 by 3×11. The constants in the two factor polynomials clearly should be −3 and −5 respectively. Then the product would be

$$(1\ 3\ \ -3)(1\ 9\ 0\ \ -5) = (1\ \ 12\ \ 24\ \ -32\ \ -15\ \ 15)$$

which is a partial repair. The other coefficients are still too large and there are two options: a) reduce $3x$ in the first factor to $-8x$, and b) reduce the $9x^2$ in the second to $-2x^2$. The latter is the correct choice and we arrive at

$$P(x) \ = \ x^5 + x^4 - 9x^3 + x^2 - 15x + 15 \ = \ (x^2 + 3x - 3)(x^3 - 2x^2 - 5).$$

The adjustments just made to the coefficients involved a short search over a small number of options and could be automated. To emphasise the point here are some factorisations of $P(x)$ modulo other primes, showing how they can be adjusted by adding or subtracting multiples of p to their coefficients to match the given $P(x) \in \mathbb{Z}$.

$$P \mod 13 = (1\ \ 1\ \ 4\ \ 1\ \ 11\ \ 2) \ = \ (1\ \ 3\ \ 10)(1\ \ 11\ \ 0\ \ 8) \ \rightarrow \ (1\ 3\ \ -3)(1\ 9\ 0\ \ -5),$$

$$P \mod 17 = (1\ \ 1\ \ 8\ \ 1\ \ 2\ \ 15) \ = \ (1\ \ 3\ \ 14)(1\ \ 15\ \ 0\ \ 12) \rightarrow (1\ 3\ \ -3)(1\ 9\ 0\ \ -5),$$

$$P \mod 19 = (1\ \ 1\ \ 10\ \ 1\ \ 4\ \ 15) \ = \ (1\ \ 3\ \ 16)(1\ \ 17\ \ 0\ \ 14) \rightarrow (1\ 3\ \ -3)(1\ 9\ 0\ \ -5),$$

$$P \mod 23 = (1\ \ 1\ \ 14\ \ 1\ \ 8\ \ 15) \ = \ (1\ \ 3\ \ 20)(1\ \ 21\ \ 0\ \ 18) \rightarrow (1\ 3\ \ -3)(1\ 9\ 0\ \ -5),$$

8 Hensel's lifting algorithm

This is a clever device for finding a congruence modulo p^{k+1} when we know its value modulo p, p prime. It is a way of obtaining a congruence relation in a much larger finite field from one in a small one. Kurt Hensel was a student of Leopold Kronecker in Berlin and made important contributions to number theory, particularly by introducing the p-adic numbers in 1897. He developed the p-adic numbers by applying ideas about power series and Taylor's theorem to integers. Hensel's lifting lemma states that

> Let $P(x)$ be a given polynomial with integer coefficients and a_k an integer which is a solution of the congruence $P(x) \equiv 0 \mod p^k$: that is $P(a_k) \equiv 0 \mod p^k$. Suppose also that the derivative $P'(a_k) \not\equiv 0 \mod p^k$. Then there exists a unique integer $a_{k+1} = a_k + \beta p^k$ such that $P(a_{k+1}) \equiv 0 \mod p^{k+1}$. The constant β may be determined unambiguously using Bézout's identity.

8.1 The method by example

Suppose we know that $P(a_1) \equiv 0 \mod p^k$ and we ask 'What are the solutions a_2 of $P(a_2) \equiv 0 \mod p^{2k}$, if any?'. Hensel pictured $P(a_2)$ as being a refinement of $P(a_1)$ and related to it by a Taylor series:

$$P(a_2) = P(a_1 + \delta) = P(a_1) + \delta P'(a_1) + \tfrac{1}{2}\delta^2 P''(a_1) + ... \quad . \tag{19a}$$

Modular arithmetic is imposed, requiring that $P(a_2) \equiv 0 \mod p^{2k}$. $P(a_1) \equiv 0 \mod p^k$ also means that $P(a_1 + \beta p^k) \equiv 0 \mod p^k$ for some integer β. Setting $\delta = \beta p^k$ means that the third term in δ^2 in the Taylor series and subsequent terms are all exactly zero modulo p^{2k}. The equation now reads

$$P(a_2) \equiv 0 \equiv P(a_1) + \beta p^k P'(a_1) \mod p^{2k} \tag{19b}$$

and is reminiscent of Newton's numerical method for the roots of a function. $P(a_2)$ will be different from $P(a_1)$ if $P'(a_1) \not\equiv 0 \mod p^k$. From this information a value of β can be found and hence a_2.

The method is best explained by example so let $P(x) = x^2 + 2x \equiv 0 \mod 7$ and we look for a solution to $x^2 + 2x \equiv 0 \mod 49$ and then $\mod 2401 = 7^4$. We first need a solution to the base case $x^2 + 2x \equiv 0 \mod 7$ and a few trials show that $x = a_1 \equiv 5$ is the only one. Of course $P(5) = 35 \not\equiv 0 \mod 7^2$. An addition must be made to 5 to make $P(x) = 0 \mod 49$. This addition will be a multiple of $p = 7$.

Using Eq 19b $P(5) = 35$ and $P'(5) = 12 \not\equiv 0 \mod 7$. The equation for β with $p = 7$, $k = 1$ is

$$P(a_2) = 0 \equiv 35 + 7.12.\beta \mod 49.$$

Following Eq 8 of §4, divide throughout by 7:

$$0 \equiv 5 + 12\beta \mod 7.$$

The solution is $\beta = 6$ and then $a_2 = a_1 + 6.7 = 5 + 42 = 47$. This is correct: $P(47) = 47^2 + 2.47 = 2303 = 47 \times 49 \equiv 0 \mod 49$. Also a numerical search for solutions confirms that this is the only one.

Apply this process again to find a solution modulo 7^4. The process is called 'lifting' as the equation is being lifted into a larger finite field. $P(47) = 2303$ and $P'(47) = 96$. The equation for the new β with $p = 7$, $k = 2$ is

$$P(a_4) = 0 \equiv 2303 + 49.96.\beta \mod 2401.$$

Divide throughout by 49 to get

$$0 \equiv 47 + 96\beta \mod 49.$$

β is readily found using Béout's identity. Recast the last equation as $-2 + 96\beta = 49\gamma$ for some integer γ and use Euclid's algorithm to find the gcd(96,49) then work backwards as in §1. We find that gcd(96,49)= 1 = 24.96 – 47.49. So $\beta = 2.24 = 48$, making $a_4 = 2399$. This is the unique solution to $x^2 + x \equiv 0 \mod 7^4$.

Since we are on a roll, let's find the solution $\mod 7^8 = 5,764,801$. $P(2399) = 5,759,999$ and $P'(2399) = 4800$. The equation is

$$P(a_8) = 0 \equiv 5759999 + 7^4.4800.\beta \mod 7^8.$$

Divide throughout by 7^4 to get

$$0 \equiv 2399 + 4800\beta \equiv -2 + 4800\beta \mod 2401$$

and solve for β from gcd(4800, 2401). Bézout's identity here is $1 = 1200.4800 - 2399.2401$ so $\beta = 2400$ and $a_3 = 2399 + 2400.2401 = 5,764,799$, the unique solution modulo 7^8.

One cannot help noticing that for this equation the solution in each case is $a_k = 7^k - 2$. I have not investigated whether this is a coincidence or an example of a principle.

The method is more general than demonstrated above as it can be used to to find a solution modulo p^{k+1} given a solution at p^k. The reason is that at Eq 19a the term in δ^2 is zero modulo any power of p greater than k. So to solve $x^2 + 2x \equiv 0 \mod 7^3$ we just use $p^k + 1$ for $k = 2$ and build on the solution $a_2 = 47$ for 7^2. Eq 19b becomes

$$P(a_3) \equiv 0 \equiv P(a_2 + \beta p^2) + \beta.p^{2k}P'(a_2) \mod p^{2k+1} \quad \rightarrow \quad 0 \equiv 2303 + 49.96.\beta \mod 7^3.$$

Divide by 7^2 to get $0 \equiv 47 + 96\beta \mod 7$ which is $0 \equiv 5 + 5\beta \mod 7$ so $\beta = -1 \equiv 6$. This makes $a_3 = 47 + 6.49 = 341$. In this way we could go on to calculate that mod 7^5 the solution is $a_5 = 16,805$, mod 7^6 it is $a_6 = 117,647$ and mod 7^7 is $a_7 = 823,541$.

A word of warning. The solution to a congruence mod p^k does depend on the precise coefficients of the starting polynomial, not just on their values mod p. This is because if c is a coefficient and $p^k < c < p^{k+1}$, then $c \mod p^k \neq c \mod p^{k+1}$. This can be clearly seen by examining some variants of $x^2 + 2x$ studied above. These calculations lifted $x^2 + 2x \equiv 0 \mod 7$ and found solutions $5 \mod 7^1$, $47 \mod 7^2$, $341 \mod 7^3$, $2399 \mod 7^4$, etc. If, however, the starting equation were $x^2 + 2x + 7 \equiv 0$ mod 7, the solutions would be $5 \mod 7^1$ as expected, but then both 21 and 26 mod 7^2 so the solution is no longer unique. If $P(x)$ were given as $x^2 + 9x \equiv 0 \mod 7$, the solutions would be $5 \mod 7^1$, 40 mod 7^2, 334 mod 7^3, 2392 mod 7^4, etc., each of which is 7 different from the solutions to $x^2 + 2x$.

8.2 Lifting of polynomials over finite fields

Hensel lifting can also be applied to polynomials to elevate them from a finite field $\mathbb{F}_{p_k}$ to $\mathbb{F}_{p_{k+m}}$. This is perhaps the most technologically important application since it is a crucial stage in many algorithms for factoring polynomials over $\mathbb{Z}$ once a factorisation in a finite field have been found. As a final stage the lifted factors of the given polynomial may need to be combined to give one over $\mathbb{Z}$.

I will illustrate this lifting procedure to factorise $P(x) = x^6 - x^5 - x^4 - 9x^3 + 23x^2 - 7x - 5$ over $\mathbb{Z}$. This is Exercise E7 in Chapter 22 of the book 'A Concrete Introduction to Higher Algebra' by

Lindsay Childs (Springer, 1997). Following his guidance we start by finding a factorisation modulo 2 into two cubics using direct search ro distinct degree factorisation, §6. Taking two general monic cubics and equating coefficients

$$P(x)|_{mod\ 2} = x^6+x^5+x^4+x^3+x^2+x+1 = (x^3+x^2+1)(x^3+x+1), \quad = \quad (1\,1\,0\,1)(1\,0\,1\,1).$$

Let (1 1 0 1)= g_1 and (1 0 1 1)= h_1. The convention is that a remainder is written as a smallest positive integer mod p, so negative signs never appear. The given $P(x)$, however, has several negative coefficients so a way will need to be found to adjust modulo p^k some positive coefficients in the calculated factors to match $P(x)$. I discuss this at the end of these calculations.

We now move up the lifting ladder to modulo 4. The product g_1h_1 will no longer equal $P(x)$ mod 4, but we look for additions to both g_1 and h_1 which will restore equality.

$$P(x)|_{mod\ 4} = (1\,3\,3\,3\,3\,1\,3) \quad \text{whilst} \quad g_1h_1 = (1\,1\,0\,1)(1\,0\,1\,1) = (1\,1\,1\,3\,1\,1\,1).$$

The discrepancy is $P - g_1h_1 = (0\,2\,2\,0\,2\,0\,2) = 2(1\,1\,0\,1\,0\,1) \mod 4$. We want g_2, h_2 such that $P - g_2h_2 \equiv 0 \mod 4$. As with the lifted congruences in §8.1, we add an integer multiple of $p = 2$ and a polynomial such that

$$P(x)|_{mod\ 4} = g_2h_2 = (g_1+2K_g)(h_1+2K_h) = g_1h_1 + 2(g_1K_h + h_1K_g) + 4K_gK_h \mod 4$$

$$P(x)|_{mod\ 4} - g_1h_1 = 2(1\,1\,0\,1\,0\,1) = 2(g_1K_h + h_1K_g) \mod 4,$$

$$(1\,1\,0\,1\,0\,1) = (1\,1\,0\,1)K_h + (1\,0\,1\,1)K_g \mod 2.$$

Here K_g and K_h are polynomials yet to be found. Clearly, since $P - g_1h_1$ has degree 5, neither can have degree greater than 2; they are at most quadratics. Therefore

$$(1\,1\,0\,1\,0\,1) = (1\,1\,0\,1)(b_2\ b_1\ b_0) + (1\,0\,1\,1)(a_2\ a_1\ a_0) \mod 2.$$

This presents us with six simultaneous linear equations for the coefficients a_i and b_j. The solution is $K_g = (0\,1\,0)$, $K_h = (1\,1\,1)$. These make

$$g_2 = (1\,1\,0\,1) + 2(0\,0\,1\,0) = (1\,1\,2\,1), \qquad h_2 = (1\,0\,1\,1) + 2(0\,1\,1\,1) = (1\,2\,3\,3).$$

Multiplying these does indeed give $(x^3+x^2+2x+1)(x^3+2x^2+3x+3) = x^6+3x^5+3x^4+3x^3+3x^2+x+3 = P(x) \mod 4$.

Now repeat the process to lift to modulo 8. $P(x)|_{mod\ 8} = (1\,7\,7\,7\,7\,1\,3)$ and the deficit $P - g_2h_2 = (1\,7\,7\,7\,7\,1\,3) - (1\,3\,7\,3\,3\,1\,3) = (4\,0\,4\,4\,0\,0) \mod 8$. We therefore divide by 2 and solve for new K_g, K_h satisfying

$$(2\,0\,2\,2\,0\,0) = (1\,1\,2\,1)K_h + (1\,2\,3\,3)K_g \mod 4.$$

The solution of the implied simultaneous equations for the coefficients is $K_g = (2\,0\,0)$, $K_h = (0\,0\,0)$. Adding 2 times these to g_2 and h_2 gives $g_3 = (1\,5\,2\,1)$, $h_3 = (1\,2\,3\,3)$ and indeed

$$(x^3+5x^2+2x+1)(x^3+2x^2+3x+3) = x^6+7x^5+7x^4+7x^3+7x^2+x+3 \mod 8.$$

We press on, lifting to modulo 16 and further until the factorisation is stable and the modulus is so large that any combination of coefficients in the given $P(x)$ could not exceed any in the lifted product of factors. Modulo 16 $P(x) = (1\ 15\ 15\ 7\ 7\ 9\ 11)$ and $g_3h_3 = (1\ 7\ 15\ 7\ 7\ 9\ 3)$ making the deficit $(0\,8\,0\,0\,0\,0\,8) \mod 16$. Again divide by 2 and solve for a new K_g, K_h satisfying

$$(4\,0\,0\,0\,0\,4) = (1\,5\,2\,1)K_h + (1\,2\,3\,3)K_g \mod 8.$$

The solution is $K_g = 4x^2$, $K_h = 4$ and the factorisation is

$$(x^3 + 13x^2 + 2x + 1)(x^3 + 2x^2 + 3x + 11) \;=\; x^6 + 15x^5 + 15x^4 + 7x^3 + 7x^2 + 9x + 11 \mod 16.$$

Let us remind ourselves that the original polynomial is $P(x) = x^6 - x^5 - x^4 - 9x^3 + 23x^2 - 7x - 5$ which is $x^6 + 15x^5 + 15x^4 + 7x^3 + 7x^2 + 9x + 11 \mod 16$ so in a sense we have almost arrived at the required factorisation. Over $\mathbb{Z}$, however, $(x^3 + 13x^2 + 2x + 1)(x^3 + 2x^2 + 3x + 11) = x^6 + 15x^5 + 31x^4 + 55x^3 + 151x^2 + 25x + 11$ which looks distinctly different from $P(x)$.

If we press on to modulo 32, $P(x)|_{mod\ 32} = x^6+31x^5+31x^4+23x^3+23x^2+25x+27$. The product of current lifted factors is $(x^3+13x^2+2x+1)(x^3+2x^2+3x+11) = x^6+15x^5+31x^4+23x^3+23x^2+25x+11 \mod 32$ giving a discrepancy $16(0\ 1\ 0\ 0\ 0\ 0\ 1)$. Divide by 2 and solve

$$(8\,0\,0\,0\,0\,8) \;=\; (1\ 13\ 2\ 1)K_h + (1\ 2\ 3\ 11)K_g \mod 16.$$

The solution is $K_g = 8x^2$, $K_h = 8$ and so $P(x)|_{mod\ 32} = (x^3 + 29x^2 + 2x + 1)(x^3 + 2x^2 + 3x + 27) \mod 32$. If this is multiplied out in $\mathbb{Z}$ the expansion is $x^6 + 31x^5 + 63x^4 + 119x^3 + 791x^2 + 57x + 27$.

Unless we change our approach it is clear that the coefficients of these cubic factors will go on increasing modulo 2^k as k increases. As commented above, the method we are following will always produce polynomial factors which have strictly positive coefficients and so can be correct only modulo the current prime power. How can the factorisation modulo 32 be converted to the true value $P(x) = x^6 - x^5 - x^4 - 9x^3 + 23x^2 - 7x - 5$? There is no reason why the values of $2K_g$ and $2K_h$ at each stage should be added to g_j, h_j. It is equally valid to subtract them and test whether the resulting factorisation, when expanded in $\mathbb{Z}$, does equal the give $P(x)$. In our case, if $2K_g = 16x^2$, $2K_h = 16$ were subtracted respectively from $(1\ 13\ 2\ 1)$, $(1\ 2\ 3\ 11)$ instead of being added, the factorisation would read $(x^3 - 3x^2 + 2x + 1)(x^3 + 2x^2 + 3x - 5) \mod 32$, and in $\mathbb{Z}$ this expands to $x^6 - x^5 - x^4 - 9x^3 + 23x^2 - 7x - 5$ which is the given $P(x)$. These two cubics are irreducible in $\mathbb{Z}$. The conclusion is that over $\mathbb{Z}$

$$P(x) \;=\; x^6 - x^5 - x^4 - 9x^3 + 23x^2 - 7x - 5 \;=\; (x^3 - 3x^2 + 2x + 1)(x^3 + 2x^2 + 3x - 5).$$

The learning point is that the option of subtracting the K_g, K_h correction terms should be tested as well as the assumed course of adding them.

8.3 Poor choice of polynomial factors mod p

In his example solved above Lindsay Child's steered students to a factorisation modulo 2 into two cubics. Suppose, however, that we did not have this guidance. In this subsection I wish to follow the trail modulo 5 where the factorisation will not be into cubics. I choose 5 because the constant term immediately disappears and x is a factor. Then a search by substituting the numbers 0 to 4 shows that $4 \equiv -1$ is a zero, so $x + 1$ is another factor. Dividing these two out leaves the quartic $(1\ 3\ 1\ 0\ 3)$ and we test whether this is the product of two quadratics by seeing whether the obvious $(1\ a\ 3)(1\ b\ 1)$ can be made to fit. $a = 2$, $b = 1$ and give

$$P(x)|_{mod\ 5} \;=\; x(x + 1)(x^2 + 2x + 3)(x^2 + x + 1).$$

Can this be lifted? The discrepancy modulo 25 is

$$P(x) - g_1h_1 \;=\; (1\ 24\ 24\ 16\ 23\ 18\ 20) - (1\ 4\ 9\ 11\ 8\ 3\ 0) \;=\; (0\ 20\ 15\ 5\ 15\ 15\ 20) \mod 25.$$

If the degrees of each factor are retained, the only way to make up this deficit and retain a monic polynomial on expansion is to add 5 times a constant to each linear factor, and 5 times a linear polynomial to each quadratic.

$$P(x)|_{mod\ 25} = (x+5a)(x+1+5b)(x^2+2x+3+5(cx+d))(x^2+x+1+5(ex+f)) \mod 25$$

Dividing be 5 and solving simultaneously modulo 5 as before, the values of the correction terms are $a = 3$, $b = 4$, $c = 2$, $d = 4$, $e = 0$, $f = 1$ making the factorisation

$$P(x) = (x+15)(x+21)(x^2+12x+23)(x^2+x+6) \mod 25.$$

This looks plausible, so how will we come to see that it is a false trail, not leading to a valid factorisation over $\mathbb{Z}$? One clue comes from multiplying out the constant term in $\mathbb{Z}$. The value is large, $> 20,000$, far from the -5 of $P(x)$ whether the corrections terms are added or subtracted.

Factorisation over a small finite field is initially done with the distinct degree algorithm and then either Berlekamp's or the Cantor-Zassenhaus method of the next section. The Hensel lifting operation has the ability to convert a factorisation modulo a low prime p into one mod p^k where p^k exceeds all coefficients in $P(x)$. It is therefore an essential final stage in moving from a finite field to factorisation over the integers. One aspect which needs attention is to identify the correct degrees of the polynomial factors over $\mathbb{Z}$, as in §8.3 above, and do this at an early stage.

9 The Cantor-Zassenhaus algorithm

In §2.1 I introduced the notion of searching for factors of a given $P(x)$ by calculating its greatest common divisor with a set of random polynomials $K(x)$ over the same prime field $\mathbb{F}_p$. This scatter-gun approach was little better than trial division from a list of known irreducible factors, but it may have given David Cantor and Hans Zassenhaus the germ of an idea. They published their powerful algorithm for factoring a general $P(x)$ over $\mathbb{F}_p$, $p \geq 3$ in 1981, commenting that it builds upon Berlekampe's. They require three preparatory stages:

1. convert $P(x)$ to a monic equivalent if its leading coefficient is not already 1 (see Appendix 1),
2. differentiate to check for multiple factors and divide them out if any are found (see end of §1),
3. carry out distinct degree factorisation, §6, to present $P(x)$ as the product $F_1F_2...F_n$ where each F_j is a product of irreducible polynomials all of the same degree. These polynomials themselves will be found in the next stage, but it is guaranteed by the distinct degree algorithm that the degree h_j of each polynomial factor of F_j is the same, and different from that of F_i, $i \neq j$.

Cantor and Zassenhaus call the next stage 'equal degree factorisation' because it operates on each F_j in turn, splitting it into a number of irreducibles all with the degree h_j which characterises F_j. It differs from Berlekamp's algorithm in that Berlekamp's is mechanistic and will eventually factorise any polynomials, whereas equal degree is probabilistic, using a randomly chosen test polynomial to operate on F_j. Therefore it is not guaranteed to factorise F_j at any one try, but has a fair probability of doing so, and is often faster than Berlekamp's.

I will illustrate the basis of this equal degree algorithm in the simplest case where, after the preliminaries above, $F_1(x)$, say, is the product of only two irreducible factors, H_1, H_2 each of degree h. We do now know what they are, but are sure they have the same degree. In principle they could be separated by making one do something noticeably different from the other under some arithmetic

operations. For instance, H_1 could be shown to be a divisor of some further polynomial $T(x)$ while $\gcd(H_2, T) = 1$. What form should $T(x)$ take?

Before explaining how and why the algorithm works, here is the recipe:

1. Start with $F_1(x)$ (I'll now drop the suffix $_1$) which has been prepared by the stages above to consist of the product of two or more irreducible polynomials all of the same degree h. If the degree $\partial(F) = n$, there will be n/h factors, all to be found. The coefficients of $F(x)$ are in $\mathbb{F}_p$.
2. Calculate the exponent $M = \frac{1}{2}(p^h - 1)$.
3. Choose a polynomial $K(x)$ with random coefficients and degree k, $h < k < n$. Check that it is coprime to F by confirming that $\gcd(K, F) = 1$. Equivalently, choose an irreducible polynomial modulo p.
4. Calculate $V = K^M \mod F$, that is, the remainder after $K(x)^M$ has been divided by $F(x)$.
5. Calculate $\gcd(V + 1, F)$ and $\gcd(V - 1, F)$. These could be 1 or one of the required degree-h factors of $F(x)$, or the product of a few of its factors.
6. If a factor or product of factors is found, divide it out from F and continue from step 3 to find other factors. If gcd is 1, return to step 3 and try with another random $K(x)$.

9.1 Example 1

To make this concrete we will apply it first to the distinct degree polynomials found in §6, where a degree-31 polynomial over $\mathbb{F}_3$ was split into $F_2(x)$ containing three degree-2 factors, $F_3(x)$ containing three of degree-3, and $F_4(x)$ containing four degree-4 ones: $n = h_2 + h_3 + h_4 = 31 = 6 + 9 + 16$.

Degree-2 factors : $F_2 = (1\ 0\ 1\ 0\ 1\ 0\ 1)$. $M = (3^2 - 1) = 4$. There is much scope to choose a random $K(x)$ so plump first for $x^5 + 2x + 1$ which is irreducible. $V = K^4 \mod F_2 = 2(1\ 2\ 0\ 2\ 1)$. Next $\gcd(V + 1, F_2) = (1\ 1\ 2)$, one of the quadratic factors. The same factor is revealed with $K = (1\ 0\ 1\ 0\ 1\ 2)$, but using $\gcd(V - 1, F_2)$. However, $K = (1\ 1\ 0\ 0\ 0\ 2)$ gives $V = K^M \mod F_2 = = 2(1\ 1\ 0\ 1\ 1)$ and $\gcd(V + 1, F_2) = (1\ 2\ 2)$, a second quadratic factor. So with some random polynomials the + sign in $\gcd(V + 1, F_2)$ yields a factor and with others it is the –1 sign. The third factor is found by division; it is $x^2 + 1$.

Degree-3 factors : $F_3 = (1\ 2\ 2\ 2\ 1\ 1\ 0\ 2\ 0\ 1)$. $M = (3^3 - 1) = 13$. We can try the same degree-5 random $K(x) = x^5 + 2x + 1$ as above since $3 < 5 < 9$. This gives $V = K^{13} \mod F_3 = 2(1\ 1\ 1\ 2\ 1\ 1\ 1\ 1\ 1)$ from which

$$\gcd(V - 1, F_3) = (1\ 0\ 2\ 2) \quad \text{whilst } \gcd(V + 1, F_3) = (1\ 2\ 0\ 2\ 0\ 1\ 2).$$

The –1 choice has unearthed one degree-3 factor whilst the +1 choice seems to have produced the product of two such. Next try $K = (1\ 0\ 1\ 0\ 1\ 2)$ to obtain $V = K^{13} \mod F_3 = x^2(1\ 1\ 0\ 1\ 2\ 1\ 2)$, a degree-8 remainder. Then

$$\gcd(V - 1, F_3) = (1\ 1\ 0\ 2) \quad \text{whilst } \gcd(V + 1, F_3) = (1\ 1\ 1\ 2\ 0\ 0\ 2).$$

The three factors of F_3 are therefore $(1\ 0\ 2\ 2)$, $(1\ 1\ 0\ 2)$ and, by division, $(1\ 1\ 2\ 1)$.

Degree-4 factors : F_4 = (1 1 1 1 2 0 2 0 2 2 2 2 0 2 1 0 1) of degree 16. This time $M = (3^4 - 1) = 40$. Using the same degree-5 $K(x)$ polynomials as above, with $K(x) = x^5 + 2x + 1$ we obtain $V = K^{40}$ mod F_4 = 2(1 1 1 1 2 1 0 2 1 0 0 2 1 1 0).

$$\gcd(V-1, F_4) = (1\ 0\ 1\ 2\ 1\ 2\ 2\ 1\ 1) \quad \text{whilst } \gcd(V+1, F_4) = (1\ 1\ 0\ 1\ 2\ 2\ 0\ 2\ 1)$$

two degree-8 polynomials each of which must be the product of two degree-4 irreducible factors. So we try the next $K(x)$ = (1 0 1 0 1 2). This gives degree-15 remainder $V = K^{40}$ mod F_4 =(1 2 0 1 2 1 0 0 2 0 0 0 0 1 1 0).

$$\gcd(V-1, F_4) = (1\ 0\ 1\ 0\ 2) \quad \text{whilst } \gcd(V+1, F_4) = (1\ 1\ 0\ 0\ 0\ 1\ 2\ 2\ 0\ 1\ 1\ 0\ 2).$$

So here a genuine degree-4 irreducible has been found. Dividing the above degree-12 polynomial by one of the degree-8 ones from the previous choice of $K(x)$ gives a second factor (1 0 0 2 2). I won't go through the calculation with a further random $K(x)$ as by now the procedure should be quite clear. The other two degree-4 factors are (1 1 0 0 2) and (1 0 0 1 2).

9.2 Example 2

The algorithm is so important that I give a further example. $P(x)$ has been constructed as the product of four degree-3 irreducibles modulo 7:

$$H_1H_2H_3H_4 = (1\ 1\ 2\ 4)(1\ 6\ 0\ 4)(1\ 0\ 3\ 5)(1\ 4\ 1\ 3) = (1\ 4\ 5\ 2\ 2\ 3\ 3\ 3\ 1\ 6\ 4\ 1\ 2) \bmod 7.$$

$M = \frac{1}{2}(7^3 - 1) = 171$. In this example the random test polynomials $K(x)$ are not necessarily irreducible in $\mathbb{F}_7$, but their gcd with $P(x)$ is has been checked to be 1.

- Try $K_1 = x^5 + 2x + 1$. Remainder mod $P(x)$ is V_1 = 6(1 4 3 0 6 3 3 2 1 2 2 1). $\gcd(V_1 - 1, P)$ = (1 4 4 6 2 0 1) = f_1, a partial factorisation. $\gcd(V_1 + 1, P)$ = (1 0 1 6 0 1 2) = f_2. This has split the degree-12 P into two of degree 6. Note that $\gcd(V_1 + b, P) = 1$ for all b other than 1 and $p - 1$.

- Try $K_2 = x^7 + x + 1$. The remainder K_2^M mod $P(x)$ is 1 so this has no useful outcome.

- Try $K_3 = x^7 + x^3 + 1$. Remainder K_3^{171} mod $P(x)$ is V_3 = 3(1 1 1 6 0 4 4 0 4 1)). $\gcd(V_3 - 1, P)$ = (1 1 2 4) = H_1, the first degree-3 factor to be found. $\gcd(V_3 + 1, P)$ = (1 3 0 6 5 0 4 0 0 4) which must be the product of three irreducible factors. Dividing this by f_1 above reveals a second degree-3 factor of H_2 = (1 6 0 4). The product of the two found is (1 1 2 4)(1 6 0 4) = f_2.

- There is only f_1 to split so we can try a test polynomial of lower degree. Try $K_4 = x^4 + x + 2$ which has gcd of 1 with f_1. Remainder K_4^{171} mod $f_1(x)$ is V_4 = 3(1 4 1 3 0 2). $\gcd(V_4 - 1, f_1) = H_3$ = (1 0 3 5) and $\gcd(V_4 + 1, f_1) = H_4$ = (1 4 1 3).

Except for K_2 all the random test polynomials have yielded a contribution to the factorisation, even if in about half the cases only a partial factorisation has resulted. The $\gcd(V-1, P)$ and $\gcd(V+1, P)$ have both been useful.

9.3 Why it works

Let the given polynomial $P(x) = H_1(x)H_2(x)...H_L(x)$, that is, it has L irreducible factors $H_j(x)$ over $\mathbb{F}_p$ which, following application of the distinct degree algorithm, are all of degree h. The degree of P is therefore Lh. Start by considering only a single factor H_1, say. By quotient with $\mathbb{F}_p[x]$, H_1 defines a field of p^h elements, represented by all polynomials mod p with degree less than h. By Eq 16 of §5.2 regarding Fermat's little theorem, every element $K_1(x)$ of this field satisfies $K_1^{p^h-1} \equiv 1$

mod H_1, mod p. The degree of K_1 will be $h-1$ of less. We also know from §3.2 on quadratic residues that this factorises into $(K_1^M-1)(K_1^M+1) \equiv 0 \mod H_1$, mod p where $M = \frac{1}{2}(p^h-1)$. Since $K \nmid H$, either $K_1^M-1 \equiv 0$ or $K_1^M+1 \equiv 0$. K_1^M will have higher degree that H_1 so either of these is a condition for H_1 to divide K_1^M+1 or K_1^M-1. If $K_1^M+1 \equiv 0$, $K_1^M-1 \equiv -2 \mod p$, and if $K_1^M-1 \equiv 0$, $K_1^M+1 \equiv +2 \mod p$. Because H_1 is irreducible, if H_1 divides K_1^M+1, $\gcd(K_1^M+1, H_1) = H_1$ and correspondingly for the negative case K_1^M-1. If the given polynomial $P(x)$ consisted of only the one factor H_1, this factor would appear as $\gcd(K_1^M+1, P) = H_1$.

Now use the Chinese remainder theorem to build up this behaviour from one factor to $P(x)$ with L factors. For each factor H_j of $P(x)$ we associate a polynomial $K_j(x)$ for which $K_j^M \pm 1 \equiv 0 \mod H_j$, the $\pm$ sign meaning either + or -1, not both. First note that the product of these K_j is just another member of the fields H_j when reduced modulo each H_j. $\hat{K}$ is the polynomial which simultaneously satisfies all the congruences $\hat{K} \equiv K_j \mod H_j$. Next, from Eq 4 of §1, if H_1 and H_2 are coprime and $\gcd(A, H_1) = \alpha$, $\gcd(A, H_2) = \beta$, then $\gcd(A, H_1H_2) = \alpha\beta$. It will be random whether $K_j^M \equiv +1$ or $-1 \mod H_k$ for any pair j, k. The chance of either is $\frac{1}{2}$. This carries over to the product of the K_j and means that $\hat{K}^M \mod H_k$ has a 50% change of being +1 or -1 modulo any H_k.

Suppose for illustration that $L = 3$, that $\hat{K}$ has been randomly chosen with degree k, $h < k < Lh$, and that $\hat{K}^M \mod H_j = -1$ for each of $j = 1$, 2 and 3. Then $\hat{K}^M + 1 \equiv 0 \mod H_1H_2H_3$; in other words, $P = H_1H_2H_3$ divides $\hat{K}^M$ and $\gcd(\hat{K}^M+1, P) = P$. Such a result would be true but not helpful in factorising $P(x)$ since it has returned $P(x)$ unchanged. But another random choice of test polynomial $\hat{K}$ might have $\hat{K}^M \mod H_1 = -1$, $\hat{K}^M \mod H_2 = -1$ but $\hat{K}^M \mod H_3 = +1$. Then $\gcd(\hat{K}^M+1, H_1) \equiv H_1$, $\gcd(\hat{K}^M+1, H_2) \equiv H_2$, $\gcd(\hat{K}^M+1, H_3 \equiv -2$, but $\gcd(\hat{K}^M-1, H_3) \equiv H_3$, Then $\gcd(\hat{K}^M+1, P) \equiv H_1H_2$ and $\gcd(\hat{K}^M-1, P) \equiv H_3$ and $P(x)$ has been factorised.

Here is a non-trivial numerical example. Take $p = 5$, $L = 4$ and

$$H_1 = x^3+x+4, \quad H_2 = x^3+2x+1, \quad H_3 = x^3+2x^2+4x+2, \quad H_4 = x^3+3x^2+2,$$

$K_1 = x^2+4x+4 \mod H_1$, $K_2 = 3x^2+x+1 \mod H_2$, $K_3 = 2x+2 \mod H_3$, $K_4 = x^2+2x+4 \mod H_4$.
Therefore $P(x) = H_1H_2H_3H_4 = (1\ 0\ 3\ 1\ 2\ 0\ 3\ 1\ 1\ 0\ 0\ 3\ 1) \mod 5$, a 12th degree polynomial. Using formula Eq 6 §3.1 the value of $\hat{K}$ can be found to satisfy all four congruences. Use $\hat{K} = \sum K_j b_j c_j$ where $b_j = \hat{H}/Hj$, $\hat{H} = H_1H_2H_3H_4$ and c_j is the inverse of b_j modulo H_j, so $b_jc_j \equiv 1 \mod H_j$. Thus

$$\begin{aligned} b_1 &= (1\ 0\ 2\ 2\ 0\ 0\ 0\ 2\ 1\ 4), & c_1 &= (3\ 4\ 1) \\ b_2 &= (1\ 0\ 1\ 0\ 0\ 4\ 3\ 3\ 1\ 1), & c_2 &= (1\ 1\ 2) \\ b_3 &= (1\ 3\ 3\ 1\ 2\ 1\ 1\ 1\ 3\ 3), & c_3 &= (0\ 2\ 4) \\ b_4 &= (1\ 2\ 2\ 3\ 4\ 4\ 0\ 3\ 4\ 3), & c_4 &= (1\ 1\ 1) \end{aligned}$$

making $\hat{K} = 2(1\ 0\ 0\ 3\ 1\ 2\ 0\ 3\ 1\ 1\ 0\ 0\ 3\ 1)$, a 13th degree polynomial. Table 2 lists the remainders and greatest common divisors of $\hat{K}^M \pm 1$. The bottom line gives $\gcd(\hat{K}^M+1, P)$ and $\gcd(\hat{K}^M-1, P)$, showing how one of the irreducible factors of $P(x)$ has been split away. The test polynomial $\hat{K}(x)$ is arbitrary provided $\gcd(\hat{K}, P) = 1$. When applying the algorithm, several values of $\hat{K}$ would be picked at random to see if any would split up $P(x)$.

This and other algorithms call for the calculation of large powers of a polynomial $P(x)$ over finite fields. This must be done efficiently. One device is to reduce by the modulus as soon as possible at each stage of the evaluation. Another device is to partition the exponent so that it can be achieved as the product of the square of $P(x)$ and squares of this square. It is beyond the scope of this article to discuss these essential aspects of efficient algorithms. Much of the literature and the coding of these algorithms is concerned with machine-efficient formulations.

	Rem +	gcd +	Rem −	gcd −
H_1	2	1	0	H_1
H_2	2	1	0	H_2
H_3	0	H_3	3	1
H_4	2	1	0	H_4
P		H_3		b_3

Table 2: Remainders and gcd for $\hat{K}^M \pm 1$ with H_1, *etc.* Rem + means $\hat{K}^M + 1 \mod H_j$ or $\mod P$ according to the row label. gcd − means $\gcd(\hat{K}^M - 1, H_j)$, $M = \frac{1}{2}(p^h - 1) = 62$, $p = 5$, $h = 3$

.

10 Calculating the discriminant Δ

The discriminant of a monic polynomial $F(x)$ with roots $\rho_1, ...\rho_n$ is

$$\Delta(F) \;=\; (-1)^{n(n-1)/2} \prod_{i,j,i\neq j}^{i=n,j=n} (\rho_i - \rho_j) \;=\; \prod_{i,j<i}^{i=n,j=n-1} (\rho_i - \rho_j)^2 \,.$$

It carries information about the roots: i) if $\Delta(F) = 0$, there are repeated roots, ii) if $\Delta > 0$, all roots are real, iii) if $\Delta < 0$, there is at least one pair of complex conjugate roots. There is a related quantity often defined which I will call $\delta(F)$ and which is the signed square root of Δ:

$$\delta(F) \;=\; \prod_{i,j<i}^{i=n,j=n-1} (\rho_i - \rho_j), \qquad \Delta = \delta^2.$$

Since $\delta^2 = \Delta$ and $\Delta \in \mathbb{Q}$, $\delta(F)$ will be either a purely real number or a purely imaginary number.

Since the roots are not usually known, it is essential that Δ can be computed from the coefficients only. We are assured that Δ can be expressed in terms of the coefficients of $F(x)$ because it is clearly a symmetric function of the roots. For all but equations of degree 2 and 3, however, this is not an efficient method. This is largely because the number of terms in Δ increases rapidly with degree n, from 2 terms for a quadratic, 5 for a cubic, 16 for a quartic, 59 for a quintic and 246 for a 6th degree polynomial. A faster method has been found using the so-called Sylvester matrix.

In its general form the Sylvester matrix is associated with the so-called 'resultant' or 'eliminant' of two polynomials, F and G. The resultant $Res(F, G)$ is the determinant of a matrix in the coefficients of both F and G such that if F and G share a roots in common, the $Res(F, G) = 0$. If

$$F \;=\; p_n x^n + p_{n-1}x^{n-1} + ... + p_1 x + p_0, \qquad G \;=\; q_m x^m + p_{m-1}x^{m-1} + ... + q_1 x + q_0,$$

then $Res(F, G)$ is the determinant of the $n + m$ square matrix constructed from successively shifted

rows of the two sets of coefficients:

$$\begin{pmatrix} f_n & f_{n-1} & f_{n-2} & .... & .... & 0 & 0 & 0 \\ 0 & f_n & f_{n-1} & f_{n-2}.... & & .... & 0 & 0 \\ 0 & 0 & f_n & f_{n-1} & f_{n-2}.... & & .... & 0 \\ ... & ... & ... & ... & ..... & & ... & ... \\ 0 & 0 & .... & ... & .... & f_2 & f_1 & f_0 \\ g_m & g_{m-1} & g_{m-2} & .... & .... & 0 & 0 & 0 \\ 0 & g_m & g_{m-1} & g_{m-2}.... & & .... & 0 & 0 \\ 0 & 0 & g_m & g_{m-1} & g_{m-2}.... & & .... & 0 \\ ... & ... & ... & ... & ..... & & ... & ... \\ 0 & 0 & .... & ... & .... & g_2 & g_1 & g_0 \end{pmatrix}$$

There are m rows from F and n from G. The discriminant of $F(x)$ is the resultant of $F(x)$ and its derivative, $F'(x)$. The reason is that if $F(x)$ has a double factor $(x-\rho_k)^2$, its derivative will share the factor $x-\rho_k$. Thus to calculate Δ for the polynomial $F(x) = x^4 - 4x^3 + 4x^2 + 12x + 5$ we have $n = 4$ and $m = 3$ so the 7×7 Sylvester matrix is

$$\begin{pmatrix} 1 & -4 & 4 & 12 & 5 & 0 & 0 \\ 0 & 1 & -4 & 4 & 12 & 5 & 0 \\ 0 & 0 & 1 & -4 & 4 & 12 & 5 \\ 4 & -12 & 8 & 12 & 0 & 0 & 0 \\ 0 & 4 & -12 & 8 & 12 & 0 & 0 \\ 0 & 0 & 4 & -12 & 8 & 12 & 0 \\ 0 & 0 & 0 & 4 & -12 & 8 & 12 \end{pmatrix}.$$

Its determinant is fairly readily evaluated with computer software and has the value 29952, which is not a square.

John Coffey, August 2022

Appendix 1: Non-monic polynomials modulo m

Polynomials over the integers $\mathbb{Z}$ will in general have positive and negative coefficients and the leading coefficient is generally not +1. Much of the theory, however, has been developed for monic polynomials, and also factorisation algorithms work modulo an integer. The question, therefore, is how to deal with non-monic polynomials modulo an integer m. There is no issue if the arithmetic operation is addition of two non-monic polynomials, or their subtraction, multiplication or raising to a power, even if m is not a prime or prime power. Complications arise only with division.

If the coefficient field is $\mathbb{Q}$, the matter is solved simply by dividing by c_n, the leading coefficient, in which case some or even all of the other coefficients will become ordinary fractions. If fractions are not permitted, there are at least two methods which may be tried to find an equivalent monic polynomial. The first is my preferred method.

Fractions modulo m

Modulo integer m the fraction n/d is only defined in the sense that $n/d = r$ where r satisfies $n \equiv dr \mod m$. This always has a solution with r in the range $\{0, 1, \ldots\ p-1\}$ if $m = p$, a prime. Some examples are

$$\tfrac{1}{2} \mod 3 \equiv 2, \qquad \tfrac{3}{4} \mod 5 \equiv 2, \qquad \tfrac{5}{9} \mod 11 \equiv 3.$$

It may also work if m is a prime power:

$$\tfrac{1}{3} \mod 4 \equiv 3, \qquad \tfrac{3}{8} \mod 9 \equiv 6, \qquad \tfrac{3}{5} \mod 8 \equiv 7.$$

However, if m is composite, including being a prime power, a fraction can be defined only if m and d are coprime. Thus none of these is defined:

$$\tfrac{1}{2} \mod 4, \qquad \tfrac{2}{5} \mod 15, \qquad \tfrac{5}{9} \mod 12.$$

The reason is that only a prime modulus can define a simple finite field. General rings have zero divisors. (Finite fields of order p^k require paired numbers to represent their elements.)

Using this approach a polynomial such as $4x^5 + 2x^4 + x^3 + 3x^2 + 5 = (4\ 2\ 1\ 3\ 0\ 5)$ can be expressed modulo 7 or any other prime p by dividing by $c_5 = 4$ and solving the congruences for the respective values of r.

$$4\left(1 \quad \tfrac{2}{4} \quad \tfrac{1}{4} \quad \tfrac{3}{4} \quad \tfrac{0}{4} \quad \tfrac{5}{4}\right) \equiv 4(1\ 4\ 2\ 6\ 0\ 3) \mod 7. \qquad (A1)$$

The original form is recovered by multiplying out the bracket and reducing modulo 7.

Change of variable

An alternative approach, noted in §1, is to change the variable from x to u with $x = u/c_n$. Then

$$c_n x^n + c_{n-1}x^{n-1} + c_{n-2}x^{n-2} + \ldots\ldots + c_1 x + c_0 \to \frac{1}{c_n^{n-1}}\left(u^n + c_{n-1}u^{n-1} + c_{n-2}c_n u^{n-2} \ldots + c_1 c_n^{n-2}u + c_n^{n-1}c_0\right)$$

so one must keep track of the a factor of c_n^{n-1}. As a numerical example the polynomial (4 2 1 3 0 5) above with $x = u/4$ becomes $\frac{1}{256}(1\ \ 2\ \ 4\ \ 48\ \ 0\ \ 1280)$ in u. This has been calculated with coefficients in $\mathbb{Q}$. It can be reduced modulo a prime. Thus $1/256 \mod 7 \equiv 2$ so $\frac{1}{256}(1\ \ 2\ \ 4\ \ 48\ \ 0\ \ 1280) \equiv 2(1\ \ 2\ \ 4\ \ 6\ \ 0\ \ 6) \mod 7$, a polynomial in u. Its equivalence to Eq A1 in x is shown by replacing u by $4x$: $2(4^5\ \ 2.4^4\ \ 4.4^3\ \ 6.4^2\ \ 0\ \ 6) \mod 7 \equiv (4\ \ 2\ \ 1\ \ 3\ \ 0\ \ 5)$ while the right side of Eq A1 is $(4\ \ 16\ \ 8\ \ 24\ \ 0\ \ 12) \equiv (4\ \ 2\ \ 1\ \ 3\ \ 0\ \ 5) \mod 7$, which is where this example began.

Appendix 2: Further properties of finite fields

In §A2.1 this Appendix demonstrates the construction of a finite field $\mathbb{F}_8$ of 8 elements using two different irreducible polynomials, and deduces the map between them. §A2.2 considers the order of the field elements and their relation to the irreducible polynomial whose quotient with the integers creates the field. The final section proves that $x^{p^k} - x \mod p$ is divisible by any and all irreducible polynomials of degree k with coefficients in $\mathbb{F}_p$.

A2.1 Isomorphism

The field of $2^3 = 8$ elements is represented by all polynomials of degree $3-1 = 2$ with integer coefficients modulo 2. There are only two irreducible cubics modulo 2: $K_1 = x^3 + x^2 + 1$ and $K_2 = x^3 + x + 1$. Any randomly chosen polynomial of degree ≥ 3 will have a remainder after division by either K_1 or K_2, and the sets of these remainders will be the same. The table below lists the 8 field elements and gives them letter labels within the respective fields modulo K_1 and K_2.

Element	(0 0 0)	(0 0 1)	(0 1 0)	(0 1 1)	(1 0 0)	(1 0 1)	(1 1 0)	(1 1 1)
in K_1	0	a	b	c	d	e	f	g
in K_2	0	p	q	r	s	t	u	v

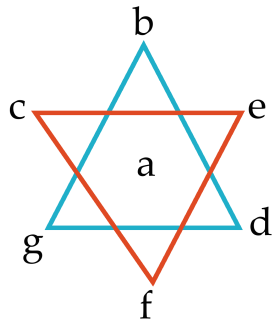
Though the elements are the same in each version, their multiplication tables are not because K_1 and K_2 produce different remainders. The addition tables do agree because there is no increase in degree on addition. The pair of tables below are the multiplication tables under K_1 (left) and K_2 (right). Each table is symmetric and 0 has been omitted. Only a few of the corresponding elements are in the same positions in both tables. The order of each element except 0 and 1 is 7. The elements excluding 0 = (0 0 0) form the cyclic group C_7 under multiplication. Every element other that 1 is a 'primitive element' meaning that its powers generate all other members of the group, including 1.

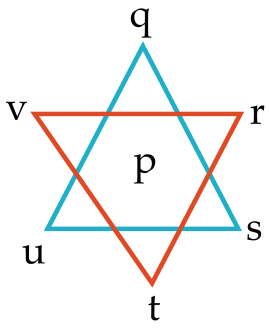
a	b	c	d	e	f	g
	d	f	e	g	a	c
		e	a	b	g	d
			g	c	b	f
				f	d	a
					c	e
						b

p	q	r	s	t	u	v
	s	u	r	p	v	t
		t	v	s	p	q
			u	q	t	p
				v	r	u
					q	s
						r

The next step is to deduce the map which pairs the elements so that these tables translate one to the other. I have approached this by calculating the squares and cubes of each element in both K_1 and K_2. Briefly, in K_1 $b^2 = d$, $d^2 = g$, $g^2 = b$, so forming a cycle. Similarly $c^2 = e$, $e^2 = f$, $f^2 = c$, forming a second cycle. Also $b^3 = e$, $e^3 = d$, $d^3 = f$, $f^3 = g$, $g^3 = c$, $c^3 = b$. These three cycles have been combined into one diagram in the left panel of the figure. e has been placed between b and d because it is intermediate in cube values.

The right panel in the figure is the equivalent for K_2. Corresponding points in the diagram spell out the isomorphism. The same conclusion can be reached by comparing powers of, say, (0 1 0). When it is used to reorder the rows and columns, we obtain a second multiplication table for K_2 which corresponds in all positions with that of K_1. This is presented in the second pair of tables. It shows, for instance, that (1 0 1) behaves in K_2 as does (1 1 0) in K_1.





a	b	c	d	e	f	g
	d	f	e	g	a	c
		e	a	b	g	d
			g	c	b	f
				f	d	a
					c	e
						b

p	q	v	s	r	t	u
	s	t	r	u	p	v
		r	p	q	u	s
			u	v	q	t
				t	s	p
					v	r
						q

A2.2 The order of field elements

The above example also illustrates that in any finite field the multiplicative group (omitting 0) is cyclic and the non-zero elements form a cyclic group under multiplication. If there are $q = p^k$ elements in the field, there are $q-1$ non-zero ones so the cyclic multiplicative group has order $q-1$. If this is prime, as for $q = 8$ above, every element except 1 is 'primitive' meaning that its powers generate all the other group members including 1. This Appendix now considers $\mathbb{F}_{3^2}$ with 9 elements.

Working modulo 3 the elements of this field will be the residue classes of an irreducible quadratic. $K = x^2 + 2x + 2 = (1\ 2\ 2)$ will serve to form the quotient $\mathbb{F}_3[x]/K(x)$. The field members are therefore (0 0), (0 1), (0 2), (1 0), (1 1), (1 2), (2 0), (2 1), (2 2). In this case $q-1 = 8$ so there will be subgroups with elements whose orders divide 8, by Lagrange's theorem on groups. The powers are listed in the table below. It shows four primitive elements, two of order 4 and one of order 2. The additive identity 0 incorporated with one subgroup forms the subfield $\mathbb{F}_3 = \{0, 1, 2\}$. However 0 incorporated with the order 4 subgroup $\{0, 1, 2, x+1, 2x+2\}$ does not constitute the field $\mathbb{F}_5$ because x, $2x$ and $x+2$ would be formed under addition but are absent from the set. In the case of $\mathbb{F}_{2^3}$ in §A2.1 the only subfield is $\{0, 1\}$. The order of the subfields in both cases is p so they are called 'prime subfields'.

power	element							
1	1	(0 2)	(1 0)	(1 1)	(1 2)	(2 0)	(2 1)	(2 2)
2		1	(1 1)	(0 2)	(2 2)	(1 1)	(2 2)	(0 2)
3			(2 1)	(2 2)	(0 2)	(1 2)	(1 0)	(1 1)
4			(0 2)	1	(0 2)	(0 2)	(0 2)	1
5			(2 0)		(2 1)	(1 0)	(1 2)	
6			(2 2)		(1 1)	(2 2)	(1 1)	
7			(1 2)		(1 0)	(2 1)	(2 0)	
8			1		1	1	1	

Powers of elements in $\mathbb{F}_9 = \mathbb{F}[x]/(x^2 + 2x + 2) \mod 3$. Patterns repeat to bottom of each column.

If the order of an element a is c, such that $a^c \equiv (0\ 1) = 1 \mod 3$, then clearly $a^{2c} \equiv 1$, $a^{4c} \equiv 1$ and therefore every multiplicative element, 1 included, satisfies $a^q \equiv 1 \mod p$. Element 0 can be included too if the statement is widened to say that in general the equation $x^{p^k} - x = 0 \mod p$ has p^k roots where k is the degree of $K(x)$, and these roots are all the p^k elements of the finite field $\mathbb{F}_{p^k}$. This can also be stated as $\mathbb{F}_p[x]/K(x)$ is the splitting field of $x^{p^k} - x$. This can be made explicit by taking the product of all linear factors. We are working both modulo 3 and modulo $K(x)$. The latter implies that $\beta^2 + 2\beta + 2 \equiv 0$ where β is any root of $K(x) = 0$, and the polynomial (2 1) means $2\beta + 1$. All the terms in β cancel:

$$(x-0)(x-1)(x-2)(x-\beta-0)(x-\beta-1)(x-\beta-2)(x-2\beta-0)(x-2\beta-1)(x-2\beta-2)$$

$$= x^9 + 2(\beta^6+\beta^4+\beta^2+1)x^3 + \beta^2(\beta^4+\beta^2+1)x = x^9+2x \equiv x^{3^2}-x \mod 3 \text{ where } \beta^2 = \beta+1. \quad (A2.1)$$

Since $K(x)$ is irreducible over $\mathbb{F}_p$, β is an irrational or complex number, but its value does not have to be known; we simply use $\beta^2 = -2\beta - 2 \equiv \beta + 1 \mod 3$ to simplify expressions because this substitution is equivalent to taking the remainder after dividing by $K(x)$.

A2.3 The polynomial $x^{p^k} - x$

The reduction to $x^9 - x$ at Eq A2.2 was made with $K(x) = x^2 + 2x + 2$, but there are two other quadratics modulo 3 which could have been used instead to form the quotient with $\mathbb{F}_3$ and so define the field. These are $x^2 + 1$ and $x^2 + x + 2$. Since there is no reason to prefer one irreducible quadratic over another, we might expect to obtain similar cancelling of terms in β, and reduction to $x^9 - x$ with both of these, and indeed this is the case. (Use $\beta^2 = 2$ and $\beta^2 = 2\beta + 1$ respectively; of course the root β is numerically different for each choice of $K(x)$.) The product of these three irreducible quadratics is $x^6 + x^4 + x^2 + 1$ which is very similar to the cyclotomic polynomials in Eq BB of §1, examined further in Appendix 3. In fact $(x^2 - 1)(x^6 + x^4 + x^2 + 1) = x^8 + 2 \equiv x^{3^2-1} - 1$, which is the product of all the non-zero field elements and the essential statement in Eq A2.1. This illustrates an important fact: that each of the equivalent field-defining irreducible polynomials $K(x)$ of degree k divides $x^{p^k} - x$. The other irreducibles appear as factors of the quotient $(x^{p^k} - x)/K(x)$.

A proof can be given along these lines. Write

$$x^{p^k} - x = Q(x)K(x) + R \qquad \text{or} \qquad x^{p^k} - x - Q(x)K(x) = R \qquad (A2.2)$$

where $Q(x)$ is the quotient and R the remainder. We are working both modulo p and modulo $K(x)$. Now β is both a root of $K(x) = 0$ and also the element (1 0) of the field. By Eq CC of §5 on Fermat's theorem $\beta^{p^k} \equiv \beta$. At $x = (1\ 0)$ the value of R is therefore

$$\beta^{p^k} - \beta - Q(\beta)K(\beta) = 0 - 0 = 0. \qquad (A2.2)$$

R is therefore zero, meaning that $K(x)$ does divide $x^{p^k} - x$. This is true for every irreducible polynomials capable of defining the field as the quotient $\mathbb{F}_p[x]/K$, and is consistent with the observation above regarding (1 0 1), (1 1 2), (1 2 2) in modulo 3.

A stronger version of this theorem in given at the end of Appendix 3.

A2.4 Representation by matrices

This topic is not developed in this article, but is of interest. The non-zero elements of a finite field $\mathbb{F}_p[x]/K(x)$ can be represented not just as polynomials but as matrices. The 'companion matrix' is

constructed from the monic polynomial $K(x)$ as follows:

$$K(x) = x^k + c_{k-1}x^{k-1} + c_{k-2}x^{k-2} + ... + c_1x + c_0 \quad \text{translates to} \quad \begin{pmatrix} 0 & 0 & & .... & 0 & -c_0 \\ 1 & 0 & & & 0 & -c_1 \\ 0 & 1 & 0 & & 0 & -c_2 \\ 0 & 0 & 1 & & 0 & \vdots \\ & & & \ddots & 0 & \\ & & & & 0 & \vdots \\ 0 & 0 & 0 & .... & 1 & -c_{k-1} \end{pmatrix}$$

Referring back to §A2.1 and the field $\mathbb{F}_{2^3}$, we used two polynomials $K(x)$ to define the field: $K_1 = x^3 + x^2 + 1$ and $K_2 = x^3 + x + 1$. The companion matrix of K_1 and its powers modulo 2 are :

$$C(K_1) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \qquad C(K_1)^2 = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \qquad C(K_1)^3 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \qquad C(K_1)^4 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix},$$

$$C(K_1)^5 = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad C(K_1)^6 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \qquad C(K_1)^7 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The matrix $C(K_1)$ therefore has degree 7 and corresponds with a generator of the multiplicative group. The equivalent matrices for $K_2(x)$ are

$$C(K_2) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad C(K_2)^2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \qquad C(K_2)^3 = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \qquad C(K_2)^4 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix},$$

$$C(K_2)^5 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}, \qquad C(K_2)^6 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \qquad C(K_2)^7 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Remarkably, except for the identity matrix, not one matrix in the powers of $C(K_2)$ is the same as one for $C(K_1)$. As evidence that each set of matrices constitutes a field, note that the inverse of $C(K_1)^j$ is $C(K_1)^{7-2j}$, and that $C(K_1)^2 + C(K_1)^5 = C(K_1)^4$, $C(K_1)^3 + C(K_1)^6 = C(K_1)^5$, and so on.

Appendix 3: Factors of $x^N - 1$

The purpose of this rather lengthy Appendix is to illustrate and sketch a proof of the theorem which underpins the discrete degree factorisation algorithm in §8: namely

> $P^*(x) = x^{\mathcal{N}} - x \mod p$, for p prime, k an integer and $\mathcal{N} = p^k$, is the product of every irreducible monic polynomial of degree k such that $h|k$, and of no others.

Since the factors x and $x-1$ occur throughout this study, it is convenient to define three polynomials which differ only in that one or both of these linear factors has been divided out:

$$P^*(x) = x^{\mathcal{N}} - x, \qquad P(x) = x^N - 1,\ N = \mathcal{N} - 1, \qquad \bar{P}(x) = \frac{x^N - 1}{x - 1}.$$

The monic irreducible factors of $x^N - 1$ are the cyclotomic polynomials, $\Phi_d(x)$. $x^N - 1$ is the product of all Φ_d for which d divides N:

$$P(x) = x^N - 1 = \prod_{d|N} \Phi_d(x).$$

This Appendix examines this factorisation both in integers and in finite fields before dealing explicitly with the theorem in §A3.4.

A3.1 Factorisation in $\mathbb{Z}$

Table 1 lists some full factorisations as a way of gaining familiarity with the behaviour. $\Phi_1(x) = x - 1$ is always a factor of $P(x) = x^N - 1$ so it is not given explicitly in Table 1 to save space. The Table gives the irreducible factors of $\bar{P}(x) = x^{N-1} + x^{N-2} + ... + x + 1 = (1\ 1\ 1\ 1\ ....\ 1\ 1)$ in coefficient-only notation.

The individual factors in each row of Table 1 are cyclotomic polynomials, Φ_N. They are obtained in sequence as N is increased from 1 as follows. Divide $x^N - 1$ by the product of all lower Φ_j for which j is a divisor of N. The quotient is an irreducible polynomial not already identified and is labelled Φ_N. Thus for $N = 2$, $(x^2 - 1)/\Phi_1 = x + 1$ or $(1\ 1)$ and is labelled Φ_2. Similarly $\Phi_{10} = (x^{10} - 1)/(\Phi_1.\Phi_2.\Phi_5) = (1\ {-1}\ 1\ {-1}\ 1)$. Each Φ_j is the minimal polynomial for a set of roots of unity. Thus the roots of $x^{10} = 1$ are furnished by Φ_1 contributing +1, Φ_2 contributing −1, Φ_5 adding $0{\cdot}31 \pm 0{\cdot}95i$ and $-0{\cdot}81 \pm 0{\cdot}59i$, and finally Φ_{10} adding the mirror image of these four in the imaginary axis to complete the division of the unit circle into 10 equal arcs. Some patterns which result from the way these polynomials are formed include

1. for N prime, p, there is only one irreducible factor of degree $p-1$; it has p terms with coefficients $(1\ 1\ 1\ 1\ .....\ 1\ 1\ 1)$.
2. for N even $(1\ 1)$ is a factor. If $4|N$, $\Phi_4 = (1\ 0\ 1)$ is also a factor and if $8|N$ so is $\Phi_8 = (1\ 0\ 0\ 0\ 1)$ (written as $(1_4...\ 1)$ for $N = 32$. Similarly multiples of $N = 3$ all contain the $(1\ 1\ 1)$ polynomial, and multiples of $N = 5$ contain $(1\ 1\ 1\ 1\ 1\ 1)$.
3. If $N = 2p$, p an odd prime, then Φ_{2p} is like Φ_p except that it has negative coefficients in alternate positions.
4. where N is a power of a prime p, $N = p^k$, the degrees of the irreducible factors are in geometric progression with ratio p and starting at $p - 1$. The factors are obtained by multiplying the factors of p^{k-1} by a new cyclotomic of degree $p^k - p^{k-1}$. Within each factor Φ_j the terms are equally spaced in degree – that is, their degrees are in arithmetic sequence whose common difference is a power of p. Table 2 makes this clear.

N	factors
1	
2	(1 1)
3	(1 1 1)
4	(1 0 1)(1 1)
5	(1 1 1 1 1)
6	(1 −1 1)(1 1 1)(1 1)
7	(1 1 1 1 1 1 1 1)
8	(1 0 0 0 1)(1 0 1)(1 1)
9	(1 0 0 1 0 0 1)(1 1 1)
10	(1 −1 1 −1 1)(1 1 1 1 1)(1 1)
11	(1 1 1 1 1 1 1 1 1 1 1)
12	(1 0 −1 0 1)(1 −1 1)(1 0 1)(1 1 1)(1 1)
14	(1 1 1 1 1 1 1)(1 −1 1 −1 1 −1 1)(1 1)
15	(1 −1 0 1 −1 1 0 −1 1)(1 1 1 1 1)(1 1 1)
16	(1_8 .... 1)(1 0 0 0 1) (1 0 1)(1 1)
18	(1 0 0 1 0 0 1)(10 0 −1 0 0 1)(1 −1 1)(1 1 1)(1 1)
20	(1 0 −1 0 1 0 −1 0 1)(1 −1 1 −1 1)(1 1 1 1 1)(1 0 1)(1 1)
24	(1 0 0 0 −1 0 0 0 1)(1 0 −1 0 1)(1 0 0 0 1)(1 −1 1)(1 0 1)(1 1)
25	(1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 1)(1 1 1 1 1)
26	(1 −1 1 −1 1 −1 1 −1 1 −1 1 −1 1)(1 1 1 1 1 1 1 1 1 1 1 1)(1 1)
27	(1_{18} .... 1_9 .... 1)(1 0 0 1 0 0 1)(1 1 1)
30	(1 1 0 −1 −1 −1 0 1 1)(1 −1 0 1 −1 1 0 −1 1)(1 −1 1 −1 1)(1 1 1 1 1)(1 1 1)(1 −1 1)(1 1)
32	(1_{16} .... 1)(1_8 .... 1)(1_4 ... 1)(1 0 1)(1 1)

Table 3: Factors of $\bar{P}(x) = \frac{x^N-1}{x-1}$ over $\mathbb{Z}$ in coefficient-only notation for selected values of N. The subscripts for $N = 27,\ 32$ denote the power of x where all other coefficients are zero.

A3.2 Factorisation in a finite field

Since most factoring algorithms work modulo a prime, we look at how the above results in $\mathbb{Z}$ are changed in a finite field. Table 3 lists selected factorisations, some of which can be compared with Table 1 in $\mathbb{Z}$. It was shown at Eq 16 §7 that $x^p + q \equiv (x+q)^p \mod p$ so $x^p - 1 = (x-1)^p \mod p$ where $-1 \equiv p-1$, giving p identical linear factors. Unlike Table 1, I have included this $x-1$ factor in Table 3 because it often fits into a sequence such as (1 4)(1 3)(1 2)(1 1) for $x^4 - 1 \mod 5$.

You will see that if N is prime, $N = q$, it may retain its single factor (1 1 1 1 .... 1 1 1) of degree $q-1$, but by no means in all cases. Table 4 gives further information on this. It shows the number ν of irreducible factors in $(x^q-1)/(x-1) \mod p$ for q a prime between 3 and 103, and for $p = 2,\ 3,\ 5$ and 7. About half the q values retain the single factor of the (1 1 1 .... 1 1 1) type. Some, however, such as $x^{31} - 1 \mod 5$ and $x^{71} - 1 \mod 5$ split into about a dozen factors. Note that ν divides $q-1$. Moreover, all ν of the factor polynomials have the same degree. For example, for $x^{31}-1$ mod 5 the 10 factors are all of degree 3, and the 14 factors of $x^{71} - 1 \mod 5$ all have degree 5.

We continue to look at the factorisation of $\Phi_q(x)$, q a prime, as it is the simplest case. Later in §A3.3 we examine a theorem that $\nu = (q-1)/\omega$ for $N = q$, q prime and $p \nmid q$, where ω is the so-called order of $p \mod q$; that is, the smallest power to which p must be raised to be equivalent to 1 modulo q. Moreover, the degree of each polynomial is ω. There is here a double interaction:

prime p	exponent k	degrees of factors
2	1	(1 0)
	2	(2 0) (1 0)
	3	(4 0) (2 0) (1 0)
	4	(8 0) (4 0) (2 0) (1 0)
	5	(16 0) (8 0) (4 0) (2 0) (1 0)
	6	(32 0) (16 0) (8 0) (4 0) (2 0) (1 0)
3	1	(2 1 0)
	2	(6 3 0) (2 1 0)
	3	(18 9 0) (6 3 0) (2 1 0)
	4	(54 27 0) (18 9 0) (6 3 0) (2 1 0)
5	1	(4 3 2 1 0)
	2	(20 15 10 5 0) (4 3 2 1 0)
	3	(100 75 50 25 0) (20 15 10 5 0) (4 3 2 1 0)
7	1	(6 5 4 3 2 1 0)
	2	(42 35 28 21 14 7 0) (6 5 4 3 2 1 0)
	3	(294 245 196 147 98 49 0) (42 35 28 21 14 7 0) (6 5 4 3 2 1 0)

Table 4: Degrees of non-zero factors of $\bar{P}(x) = \dfrac{x^{p^k}-1}{x-1}$. Thus (18 9 0) means $x^{18}+x^9+1$. The degree of the polynomial in each row is p^k-1.

p	N	coefficients mod p
2	3	(1 1 1) (1 1)
	5	(1 1 1 1 1 1) (1 1)
	7	(1 1 0 1) (1 0 1 1) (1 1)
	11	(1 1 1 1 1 1 1 1 1 1 1) (1 1)
	15	(1 1 0 0 1) (1 0 0 1 1) (1 1 1 1 1) (1 1 1)(1 1)
	31	(1 0 1 1 1 1) (1 1 1 1 0 1)(1 1 1 0 1 1)(1 0 0 1 0 1)(1 1 0 1 1)(1 0 1 0 1)(1 1)
3	7	(1 1 1 1 1 1) (1 2)
	8	(1 2 2) (1 1 2)(1 0 1) (1 1) (1 2)
	26	(1 2 2 2) (1 2 1 1) (1 2 0 1) (1 1 2 1) (1 1 1 2) (1 1 0 2) (1 0 2 2) (1 0 2 1) (1 1) (1 2)
	80	(1 2 2 1 2)(1 2 1 2 1)(1 2 1 1 2)....(1 0 1 0 2)(1 0 0 2 2)(1 0 0 1 2)(1 2 2)(1 1 2)(1 0 1)(1 1)(1 2)
5	4	(1 4) (1 3) (1 2) (1 1)
	7	(1 1 1 1 1 1 1) (1 4)
	11	(1 4 4 1 3 4) (1 2 4 1 1 4) (1 4)
	13	(1 3 0 3 1) (1 2 1 2 1) (1 1 4 1 1) (1 4)
	24	(1 4 2) (1 4 1) (1 3 4) (1 3 3) (1 2 4) (1 2 3) (1 1 2) (1 1 1) (1 0 3) (1 0 2) (1 4)(1 3)(1 2)(1 1)
7	6	(1 6) (1 5) (1 4) (1 3) (1 2) (1 1)
	13	(1 1 1 1 1 1 1 1 1 1 1 1 1) (1 6)
	48	(1 6 6)(1 6 4)(1 6 3)(1 5 5)(1 5 3).....(1 1 3)(1 0 4)(1 0 2)(1 0 1)(1 6)(1 5)... (1 1)

Table 5: Coefficients of $P(x) = x^N - 1 \mod p$ for a selection of values of N.

modulo p on the cyclotomic polynomial, and modulo q on p^k. The theorem is illustrated by the data in Table 4.

Case 1 : $x^{17} - 1 \mod 3$ **:** $q - 1 = 16$. The powers of 3 mod 17 are 3, 9, 10, 13, 5, 15, 11, 16, 14, 8, 7, 4, 12, 2, 6, 1, a seemingly random sequence of the integers 1 to 16. The order ω is 16 since it is

q	Mod 2	Mod 3	Mod 5	Mod 7	q	Mod 2	Mod 3	Mod 5	Mod 7
3	1		1	2	47	2	2	1	2
5	1	1		1	53	1	1	1	2
7	2	1	1		59	1	2	2	2
11	1	2	2	1	61	1	6	2	1
13	1	4	3	1	67	1	3	3	1
17	2	1	1	1	71	2	2	14	1
19	1	1	2	6	73	8	6	1	3
23	2	2	1	1	79	2	1	2	1
29	1	1	2	4	83	1	2	1	2
31	6	1	10	2	89	4	1	2	1
37	1	2	1	4	97	2	2	1	1
41	2	4	2	1	101	1	1	4	1
43	3	1	1	7	103	2	3	1	2

Table 6: Number ν of irreducible factors in the factorisation of $\Phi_q(x) = \dfrac{x^q-1}{x-1} \mod p$ for q prime. For a given p and q all factors have the same degree given by $(q-1)/\nu$.

not until 3^{16} that 1 is reached. According to the theorem the number of irreducible factors is 16/16 = 1, as Table 4 records.

Case 2 : $x^{13}-1 \mod 5$ **:** $q-1 = 12$. $3^k \mod 13$ generates the sequence 5, 12, 8, 1, 5, 12, 8, 1, ... This is a short recurring sequence with $\omega = 4$. Hence the theorem predicts $12/4 = 3$ irreducible factors or order 4, as is indeed the case.

Case 3 : $x^{73}-1 \mod 2$ **:** $q-1 = 72$. The sequence $2^k \mod 73$ runs 2, 4, 8, 16, 32, 64, 55, 37, 1, 2, 4, 8, ... , another recurring sequence. $\omega = 9$ and so there are $72/9 = 8$ irreducible factors, as observed.

Where N in x^N-1 is not a prime q, Table 1 has shown that it will have two or more irreducible factors over $\mathbb{Z}$, and in $\mathbb{F}_p$ each of these can split. The theorem just mentioned states that the numerator $N-1$ is replaced by $\phi(d)$, the Euler totient function for d where d is the degree of a cyclotomic factor of x^N-1. This means that each cyclotomic factor of degree m over $\mathbb{Z}$ will split into $\phi(m)/\omega_m$ sub-factors over $\mathbb{F}_p$. These examples illustrate the theorem:

Case 4 : $x^{15}-1 \mod 2$ **:** In $\mathbb{Z}$ (Table 1) this factorises as $\Phi_{15}\,\Phi_5\,\Phi_3\,\Phi_1$ where Φ_{15} is (1 –1 0 1 –1 1 0 –1 1). In Table 3 Φ_1, Φ_3 and Φ_5 are unchanged, but Φ_{15} has split into (1 1 0 0 1) (1 0 0 1 1) neither of which is a cyclotomic polynomial. Φ_{15} is reducible modulo 2 and x^4+x^3+1, x^4+x+1 are its irreducible factors. Now $\phi(15) = 8$ and $2^4 \equiv 1 \mod 15$ so $\omega = 4$ and the theorem predicts 8/4 =2 irreducible factors of degree 4. This is what we observe.

Case 5 : $x^{26}-1 \mod 3$ **:** In $\mathbb{Z}$ (Table 1) this factorises as $\Phi_{26}\,\Phi_{13}\,\Phi_1$, the first two being degree 12 cyclotomics. Table 3 shows that each of these has split into four irreducibles, making ten factors in all. The totient functions are $\phi(13) = \phi(26) = 12$; also $3^3 \equiv 1 \mod 26$ so $\omega = 3$. Therefore ν for each of Φ_{26} and Φ_{13} is $12/3 = 4$, each of degree 3.
A similar process is going on with other paired values of N and p which have a large number of irreducible factors of low degree.

A3.3 Counting factors of $x^{p^k} - x$

Case 5 had $N = p^3 - 1 = 26$. The row for $p = 3$, $N = 80 = p^4 - 1$ in Table 3 shows a string of factors of degree 4 with seemingly every permutation of the numbers 0, 1, 2 amongst their coefficients. Such cases appear particularly to split into many factors of low degree, so from the many pairs N, p which could be studied over finite fields, we pick out these. Referring to Table 3, since the factors are each irreducible and the sum of their degrees is N, they must be the full set of irreducible factors of $P(x) = x^N - 1 \mod p$. However, not every permutation of coefficient values is present so we ask whether those that are missing either represent reducible polynomials or ones which do not divide $x^N - 1$. Here are a few cases:

Case A : $p = 2$, $N = 7$. There are only 8 monic polynomials modulo 2 with degree 3. Two are listed in Table 3. The missing six others are (1 1 1 1), (1 0 0 1), (1 1 1 0), (1 1 0 0), (1 0 1 0), (1 0 0 0) all factorise modulo 2. For instance (1 0 0 1) = (1 1 1)(1 1). There are only two irreducible monic polynomials of degree 3 and these are in the row of Table 3. We might ask why there are no monic irreducible factors with degree 4 or 5. For instance (1 1 1 1 1) is irreducible, but it does not divide $x^7 - 1 \mod 2$, while (1 0 1 1 1) does divide it but is reducible.

Case B : $p = 3$, $N = 80$. $x^{80} - 1 \mod 3$ has two factors of degree 1, three of degree 2 and 18 of degree 4. (1 2 1 1 1) is missing but splits as (1 2 2)(1 1)(1 2), all of which are irreducible factors listed in Table 3. There are no factors of degree 3. Some like (1 1 2 1) do not divide $x^{80} - 1$; others like (1 1 1 1) do divide it, but factor into ones of lower degree, all of which appear in the row in Table 3. There are $3^4 = 81$ monic 4th degree polynomials modulo 3 – are the 18 listed in Table 3 the only irreducible ones? I have checked the 54 which have a non-zero constant term. Those with zero constant clearly have x as a linear factor. There are indeed exactly 18 irreducible degree-4 monic polynomials modulo 3. Here is the list:

(1 0 0 1 2)	(1 0 0 2 2)	(1 0 1 0 2)	(1 0 1 1 1)	(1 0 1 2 1)	(1 0 2 0 2)
(1 1 0 0 2)	(1 1 0 2 1)	(1 1 1 0 1)	(1 1 1 1 1)	(1 1 1 2 2)	(1 1 2 2 2)
(1 2 0 0 2)	(1 2 0 1 1)	(1 2 1 0 1)	(1 2 1 1 2)	(1 2 1 2 1)	(1 2 2 1 2).

Case C : $p = 5$, $N = 24$. There are 20 monic quadratic polynomials modulo 5 which have a non-zero constant term. I have checked all 20 and confirm that only 10 are irreducible, and these are the 10 listed in Table 3. This is the second example of $x^N - 1$, $N = p^k - 1$ factorising into a complete set of irreducible polynomials all of the same degree, here 2. The other four factors are linear, (1 1) to (1 4) and are the complete set of irreducible monic degree-1 polynomials.

Case D : $p = 5$, $N = 71$. This case is different in that N is not of the form $5^k - 1$. It does, however, have 15 irreducible factors, and we ask whether this 14 of degree 5 plus one of degree 1 is the complete set of irreducibles. There are $5^5.4 = 2500$ possible monic polynomials of degree 5 with non-zero constant term. (1 4 0 4 2 2) is only one monic irreducible modulo 5 which does not divide $x^{71} - 1$. We can conclude that this pair of p, N does *not* furnish a full set of irreducible polynomials.

The number of irreducible polynomials factors of $P^*(x) = x^{p^k} - x$ was worked out in the early 19th century by August Möbius and others. The way the pattern develops is fairly clear from the tables above where factorisation of a high power of x contains the factorisations of lower powers. Table 6 lists the number of irreducible factors of degrees from 1 to 8 in $x^{p^k} - x$. There are always p linear factors $x(x+1)(x+2)...(x+p-1)$. For $p = 2$, (1 1 1) is the only irreducible quadratic and it first appears in $x^4 - x$. There are only two irreducible cubics modulo 2, (1 0 1 1) and (1 1 0 1), and these first appear in $x^{16} - x$.

Table 6 gives evidence that only polynomials whose degree k divides k can be factors of $x^{p^k} - x$. With this knowledge the numbers of factors can be calculated for any p and k. For example, if k is prime, its only divisors are 1 and k so after the p linear factors have been stated, the degree remaining must be divided by k. With $x^{128} - x$, 128 – 2 = 126 and 126/7 = 18. Similarly, with $p = 3$, $k = 6$ the divisors of k are 1, 2, 3, 6. From 729 subtract the degrees of the three linear factors $x(x+1)(x+2)$, the three quadratic ones which are (1 0 1), (1 1 2), (1 2 2), and the eight cubics which first appear in $x^{27} - x$. This leaves a degree of 729 – 3 – 6 – 24 = 696 to be shared equally amongst factors of degree 6. There are 696/6 = 116 of them.

						degree, k					
p	k	p^k	1	2	3	4	5	6	7	8	next terms
2	2	4	2	1							
	3	8	2		2						
	4	16	2	1		3					
	5	32	2				6				
	6	64	2	1	2			9			
	7	128	2						18		
	8	256	2	1		3				30	56, 99, 186
3	2	9	3	3							
	3	27	3		8						
	4	81	3	3		18					
	5	243	3				48				
	6	729	3	3	8			116			312, 810
5	2	25	5	10							
	3	125	5		40						
	4	625	5	10		150					624, 2580
7	2	49	7	21							
	3	343	7		112						

Table 7: The number of irreducible polynomials of various degrees in the factorisation of $x^{p^k} - x$. The column on the right shows how the sequences continue.

Now that these examples have given a fair picture of what is happening, some algebraic analysis is appropriate. The formula in the literature for the number of irreducible polynomials modulo a prime uses Möbius's arithmetic μ function. This is defined for an integer N as:

- $\mu(N) = 0$ if N contains a square factor, or higher power. Therefore $\mu(24) = \mu(2^3.3) = 0$.
- $\mu(N) = (-1)^j$ if N is the product of j distinct primes. Therefore $\mu(15) = \mu(3.5) = +1$ and $\mu(30) = \mu(2.3.5) = -1$.
- $\mu(1) = 1$ for consistency.

Möbius proved his 'inversion formula' (perhaps 'reversion' would be better) which relates two general arithmetic functions of the integers $f(n)$, $g(n)$. It states that

$$\text{if } \; g(N) = \sum_{d|N} f(d), \quad \text{then } f(N) \; = \; \sum_{d|N} \mu\left(\frac{N}{d}\right).g(d) \; = \; \sum_{d|N} \mu(d).g\left(\frac{N}{d}\right).$$

Denote the number of irreducible factors of $P^*(x) = x^{p^k} - x$ with degree k by ν_h. Since the product of all such factors over all values of k is $P^*(x)$, the sum of their degrees is p^k. In other words, the

sum of all degrees for all irreducible factors whose degree divides k is

$$\sum_{h|k} h.\nu_h = p^k.$$

Applying Mobius' inversion formula

$$h\nu_h \;=\; \sum_{j|k} \mu\left(\frac{k}{j}\right).p^j \;=\; \sum_{j|k} \mu(j).p^{k/j}.$$

The number ν_m of irreducible polynomials with any degree m in $\mathbb{F}_p$ is obtained from this by a small change of notation.

$$\nu_m \;=\; \frac{1}{m}\sum_{j|m} \mu\left(\frac{m}{j}\right).p^j \;=\; \sum_{j|m} \mu(j).p^{m/j}.$$

A couple of numerical examples from above will illustrate this.
Case A : $p = 2,\ N = 7$ in $P(x),\ m = 3$. The number of irreducible polynomials with degree 3 is $\nu_3 = \frac{1}{3}[\mu(1)2^{3/1} + \mu(3)2^{3/3}] = (8-2)/3 = 2$ as observed. The total degree is 8 but 2 has been contributed by the two linear factors.
Case B : $p = 3,\ N = 80,\ m = 4$. The number of monic irreducibles with degree 4 is $\nu_4 = \frac{1}{4}[\mu(1)3^{4/1} + \mu(2)3^{4/2} + \mu(4)3^{4/4}] = (81 - 9 + 0)/4 = 72/4 = 18$ as observed.
Case of : $p = 3,\ N = 729$ in $P^*(x),\ m = 6$. $\nu_6 = \frac{1}{6}[\mu(1)3^{6/1} + \mu(2)3^{6/2} + \mu(3)3^{6/3} + \mu(6)3^{6/6}] = (729 - 27 - 9 + 3)/6 = 696/6 = 116$. Degrees $27 - 3$ are contributed by the cubic factors and 9 jointly by the quadratic and linear ones.

The Möbius formula has parallels with the Inclusion-Exclusion Principle for sets. In Case B the total degree is 81, the first term in the Möbius series for ν_4. The 9 subtracted accounts for the aggregate of degrees contributed by factors degree 2 and 1. The remainder is shared amongst the fourth degree monic polynomials.

A3.4 The Distinct Degree factor theorem

The essential theorem which underlies the distinct degree factorisation algorithm is the one quoted at the beginning of this Appendix, namely that in the finite field $\mathbb{F}_p$ and p prime

$$P^*(x) \;=\; x^{p^k} - x \;=\; \prod_{\substack{f(x)\in Irr(x)\\ h|k}} f(x) \mod p.$$

That is, $x^{p^k} - x$ is the product of all irreducible polynomials $H(x)$ modulo p whose degree h divides k. Earlier in this article two special cases of $h = k$ have been proved. Below is a summary of these followed by a sketch proof of the general case $h|k,\ h \neq k$.

Case $h = k = 1$ This is just Fermat's Little Theorem. The field consists only of the polynomials of constant terms, that is the integers 0, 1, 2, ... $p-1$. $x^{p^k} - x$ is merely $x^p - x \mod p$ which factors as at Eq 16, §5, into the product of all degree 1 polynomials modulo p. All linear polynomials are irreducible, so the theorem gives a correct statement.

Case $h = k \neq 1$ This was proved in Appendix 2, §A2.3. To run over it again, $K(x)$ is an irreducible polynomial of degree k over $\mathbb{F}_p$. It forms a field with p^k elements, these elements being represented by the remainders of $\mathbb{F}[x]_p/K(x)$; that is by integers modulo p and polynomials of all degrees $h < k$. Every member of the multiplicative group (*i.e.* all excluding 0) has an order which divides $p^k - 1$ by

Lagrange's theorem. If element 0 is included, the statement must be modified to $P^*(x) = x^{p^k} - x \equiv 0 \mod p$. $K(x)$ is the minimal polynomial of some algebraic number β which satisfies $K(\beta) = 0$. β is also the field element (1 0) in coefficient-only notation. Division of $P^*(x)$ by $K(x)$ can be written in terms of quotient $Q(x)$ and remainder $R(x)$ as $P^*(x) = QK + R$. Substituting the particular value β for the place-holder x gives $\beta^{p^k} - \beta - Q(\beta)(K(\beta) = R(\beta) = 0$. Hence $K(\beta)$ divides $P^*(\beta)$. With a change of notation this is $K(x) \,|\, P^*(x)$ as the theorem states. There are in general several irreducibles of degree k which could act as $K(x)$, and the result is equally true for these. Therefore every irreducible of order k and modulo p divides $x^{p^k} - x$.

Case $1 < h \leq k$ To address the general case we consider the two fields $\mathbb{F}_H = \mathbb{F}[x]_p/H(x)$ and $\mathbb{F}_K = \mathbb{F}[x]_p/K(x)$. H has degree h and K degree k. The elements of $\mathbb{F}_H$ are all polynomials with coefficients in $\mathbb{F}_p$ and with degrees $\leq h-1$. Similarly the elements of $\mathbb{F}_K$ are all polynomials with coefficients modulo p $\mathbb{F}_p$ and with degrees $\leq k-1$. If we look at the situation from the point of view of algebraic field extensions, as is done in Galois theory, $H(x)$, being irreducible, is the minimal polynomial of a root α of $H(x) = 0$. So $\mathbb{F}_H = \mathbb{F}_p(\alpha)$, a field extension of degree h. By Fermat's little theorem in $\mathbb{F}_H$ $\alpha^p \equiv \alpha \mod p$ and also, as shown in Case $h = k$ above, $\alpha^{p^h} \equiv \alpha \mod p$. If $h \,|\, k$, $k = mh$ for some integer m,

$$\alpha^{p^k} = \alpha^{p^{mh}} = (....(((\alpha^{p^h})^{p^h})^{p^h})....)^{p^h}, \quad \text{recursive raising } \alpha \text{ to power of } p^h \text{ through } m \text{ stages.}$$

$$= (...(((\alpha)^{p^h})^{p^h})....)^{p^h} = ((\alpha^{p^h})^{p^h}) = ... = \alpha.$$

Hence $\alpha^{p^k} \equiv \alpha$ meaning that α is a root of $K(x) = 0$ as well as of $H(x) = 0$. Finally, invoke division similar to the case above:

$$R(x) = (x^{p^k} - x) - Q(x)H(x) \implies R(\alpha) = (\alpha^{p^k} - \alpha) - Q(\alpha)H(\alpha) = 0 - 0.$$

Therefore, with zero remainder, $H(x)$ divides $x^{p^k} - x$.

Continuing in terms of field extensions, to build $\mathbb{F}_K$ one or more roots β_j of $K(x) = 0$ must be added; $\mathbb{F}_K = \mathbb{F}_p(\alpha, \beta_1, \beta_2, ..)$. The tower looks like

$$\mathbb{F}_p \subset \mathbb{F}_p(\alpha) \subset ..... \subset \mathbb{F}_p(\alpha, \beta_1, \beta_2, ..) = \mathbb{F}_K.$$

By the tower rule the order of the largest field extension is the product of the extensions of all the intermediate stages : $|\mathbb{F}_K : \mathbb{F}_p| = h.j_1.j_2.....j_m = k$. This means that h is a factor of k and gives the condition for $\mathbb{F}_H$ to be a subfield of $\mathbb{F}_K$.

Some texts which sketch a proof for this theorem make use of the following auxiliary theorem: for any positive integers a, k and $h < k$,

$$\gcd(a^k - 1, a^h - 1) = a^g - 1 \quad \text{where} \quad g = \gcd(h, k). \tag{A3.1}$$

In words, the gcd of the functions mirrors the gcd of their exponents. To see this note that since $k > h$

$$a^k - 1 = (a^h - 1)(a^{k-h} + a^{k-2h} + a^{k-3h} + ... + a^r) + (a^r - 1).$$

This has the form $P_k = P_h Q + R$. The sequence of successively subtracting h from the index of a in Q must stop when to subtract another h would make the power of a negative. Therefore $r < h$. This, of course, is the first stage of Euclid's algorithm for the gcd of k and h: $k = qh + r$ for some quotient q. Clearly if $r = 0$ the terms in the larger bracket stop at $a^0 = 1$ and V_h divides V_k. If the algorithm

is followed through to the last non-zero remainder, we would determine $g = \gcd(h, k)$ and conclude that $\gcd(a^k - 1, a^h - 1) = a^g - 1$.

From this theorem, if $h|k$, $\gcd(h, k) = h$ and so $p^h - 1 \,|\, p^k - 1$. These are the orders of the multiplicative groups in the fields with p^h, p^k elements defined by $\mathbb{F}[x]_p/H(x)$ and $\mathbb{F}[x]_p/K(x)$ respectively. It shows that two conditions must be fulfilled for $\mathbb{F}_H$ to be a subfield of $\mathbb{F}_K$. First $\mathbb{F}_H$ with p^h elements has to fit into a tower of field extensions ending in $\mathbb{F}_K$ with p^k elements. Second, the multiplicative group $\mathbb{F}_H \backslash \{0\}$ with order $p^h - 1$ must be a subgroup of $\mathbb{F}_K \backslash \{0\}$ with order $p^k - 1$. Both are attained in $h \,|\, k$.