s

Chapter 17

PID's and CRT's

17.I The Chinese Remainder Theorem, I.

We will state and prove the Chinese Remainder Theorem in 2 versions. First, as an existence (and uniqueness) theorem on solutions of simultaneous congruences. Then, as a structure theorem on quotient rings of a P.I.D.

"*Ring*" still means "commutative ring with 1".

17.I.1 Definition: Let $I_1, I_2, \ldots, I_d$ be ideals of the ring R. The **product ideal** $I_1 \cdot I_2 \cdot \ldots \cdot I_d$ is the ideal consisting of sums of products $m_1 \cdot m_2 \cdot \ldots m_d$ with $m_k \in I_k$

Obviously, the product of a set of ideals is itself an ideal. It is contained in each I_k, hence in their intersection, by the defining properties of ideals. If the ideals I_k are *principal ideals*, $I_k = (m_k)$, then, simply, $I_1 \cdot I_2 \cdot \ldots I_d = (m_1 \cdot m_2 \cdot \ldots \cdot m_d)$

The product may be strictly contained in the intersection. This is illustrated by the simple example $(4) \cdot (6) = (24)$, $(4) \cap (6) = (12)$. Note that $12 = 6 \cdot 4/(6,4)$.

The typical situation in which to expect equality is given by the following definition.

17.I.2 Definition: The ideals $I_1, I_2, \ldots, I_d$ are said to be **pairwise comaximal** if, for each $i, j, i \neq j$,

$$I_i + I_j = R$$

17.I.3 Example: In a P.I.D. R two ideals $(m_1), (m_2)$ are comaximal if, and only if, the ideal generated by m_1 and m_2 contains 1, i.e., iff there are elements $\alpha_1, \alpha_2 \in R$ with

$$\alpha_1 m_1 + \alpha_2 m_2 = 1$$

i.e., iff the m_i have no non-trivial common factor. (Bézout.) ∎

From now on, we will assume that R is a P.I.D. Most of the results, and their proofs, hold for general commutative rings (with identity), mutatis mutandis.

17.I.4 Lemma. *If the ideals I_k are pairwise comaximal, then*

$$I_1 + I_2 \cdot \ldots \cdot I_d = R$$

Proof: By our standing assumption, our ideals are principal, $I_k = (m_k)$

$k = 1, 2, \ldots, d$. By assumption there are elements $\alpha_1, \ldots, \gamma_d$ with

$$\begin{aligned} 1 &= \alpha_1 m_1 + \alpha_2 m_2 \\ 1 &= \beta_2 m_2 + \qquad \beta_3 m_3 \\ &\ldots \\ 1 &= \gamma_1 m_1 + \qquad\qquad \gamma_d m_d \end{aligned}$$

(two terms in each right member) Multiplying left and right members we get

$$\begin{aligned} 1 &= (\text{lots of terms}) \cdot m_1 + \\ &\quad + \alpha_2 \beta_3 \cdots \gamma_d (m_2 m_3 \cdots m_d) \\ &\quad \in (m_1) + (m_2 m_3 \cdots m_d) \end{aligned}$$

∎

Of course! If m_1 has no non-trivial factor in common with the other m_k's then it has none in common with their product. But I wished to avoid divisibility theory.

17.I.5 Theorem. *Same assumptions. Then*

$$I_1 \cap I_2 \cap \ldots \cap I_d = I_1 \cdot I_2 \cdot \ldots \cdot I_d$$

Proof: We start first with the case $d = 2$ and the identity

$$a_1 m_1 + a_2 m_2 = 1$$

Pick any $x = p_1 m_1 = p_2 m_2 \in (m_1) \cap (m_2)$. Then

$$\begin{aligned} x &= x(a_1 m_1 + a_2 m_2) \\ &= x a_1 m_1 + x a_2 m_2 \\ &= p_2 m_2 a_1 m_1 + p_1 m_1 a_2 m_2 \\ &= a_1 p_2 m_1 m_2 + a_2 p_1 m_1 m_2 \\ &= (a_1 p_2 + a_2 p_1) m_1 m_2 \end{aligned}$$

i.e.,

$$(m_1) \cap (m_2) \subset (m_1 m_2)$$

and we have already noted that the reverse inclusion is trivial.

The general case proceeds by induction on the number of ideals. The induction assumption, and then the case $d = 2$, yield

$$\begin{aligned} I_1 \cap (I_2 \cap \ldots \cap I_d) &= I_1 \cap (I_2 \cdot \ldots \cdot I_d) \\ &= I_1 \cdot I_2 \cdot \ldots \cdot I_d \end{aligned}$$

where we also used that I_1 and $I_2 \cdots I_d$ are comaximal, by the lemma. ∎

Remark: The first part may be more transparent in ideal notation: $R = I_1 + I_2 \Rightarrow I_1 \cap I_2 = I_1 \cap I_2(I_1 + I_2) \subset I_2 \cdot I_1 + I_1 \cdot I_2 = I_1 \cdot I_2$ The same goes for the Lemma above.

17.I.6 Theorem. Chinese Remainder Theorem I. *Let $a_i, i = 1, 2, \ldots, d$, and $m_i, i = 1, 2, \ldots, d$ be given elements of R, with the ideals (m_i) pairwise comaximal. Then the system of simultaneous congruences*

$$x \equiv a_1 \ (mod\ m_1)$$
$$x \equiv a_2 \ (mod\ m_2)$$
$$\ldots$$
$$x \equiv a_d \ (mod\ m_d)$$

has a unique solution modulo $m_1 m_2 \cdots m_d$, i.e., the general solution is of the form

$$x = x_0 + r m_1 m_2 \cdots m_d;\ r \in R$$

Proof: *Existence:* It is enough to solve, for each fixed i, the system

$$x_i \equiv \delta_{ij} \ (\text{mod}\ m_j), j = 1, 2, \ldots, d$$

($\delta_{ij} := 1$ if $i = j$, $\delta_{ij} = 0$ otherwise, "Kronecker delta")

The system of the statement above is then satisfied by $x = \sum_{i=1}^{d} a_i x_i$ (superposition, exercise).

For notational convenience we consider only $i = 1$:

$$x_1 \equiv 1 \ (\text{mod}\ m_1)$$
$$x_1 \equiv 0 \ (\text{mod}\ m_j),\ j \neq 1$$

By lemma I.4. we have

$$(m_1) + (m_2 \cdot \ldots \cdot m_d) = (1)$$

so there are elements $b_1, b_2 \in R$ satisfying

$$b_1 m_1 + b_2 (m_2 \cdots m_d) = 1$$

Obviously, $x_1 = b_2 (m_2 \cdots m_d)$ does it!

Uniqueness: If x, x' are two solutions, then we have $x - x' \equiv a_j - a_j = 0 \ (\text{mod}\ m_j)$, $\forall j$, i.e., $x - x'$ belongs to the intersection of the (m_j), hence to their product, by comaximality. ∎

17.I.7 Example: Let $R = k[X]$, k a field, $m_i = (X - x_i), i = 1, 2, \ldots, d$ where the x_i are distinct elements of the field k. Let $y_i, i = 1, 2, \ldots, d$ be given elements of k. We have

$$p(X) \equiv y_i \ (\text{mod}\ (X - x_i))$$

iff, for some $q(X) \in k[X]$, $p(X) = q(X)(X - x_i) + y_i$, i.e., iff $p(x_i) = y_i$. The solution to

$$p_1(X) \equiv 1 \pmod{(X - x_1)}$$
$$p_1(X) \equiv 0 \pmod{(X - x_j)}, j \neq 1$$

is easily found. By the second line above p_1 must be a constant multiple of $(X - x_2)(X - x_3) \cdots (X - x_n)$. By the first we must have $p_1(x_1) = 1$ which gives us the value of the constant. We find

$$p_1(X) = \frac{(X - x_2) \cdots (X - x_d)}{(x_1 - x_2) \cdots (x_1 - x_d)} = \frac{q_1(X)}{q_1(x_1)}$$

where $q_1(X) = q(X)/(X - x_1)$; $q(X) = (X - x_1) \cdots (X - x_d)$

Introducing

$$q_i(X) = \frac{q(X)}{(X - x_i)}, \quad i = 1, 2, \ldots, d$$

we similarly find (on replacing 1 by i) that the system

$$p_i(X) \equiv \delta_{ij} \pmod{(X - x_j)}, j = 1, 2, \ldots, d$$

is satisfied by

$$p_i(X) = \frac{q_i(X)}{q_i(x_i)}$$

so the general system

$$p(X) \equiv y_i \pmod{(X - x_i)}$$

has the solution

$$p(X) = \sum_{i=1}^{d} y_i \frac{q_i(X)}{q_i(x_i)}$$

Since the system is uniquely solvable modulo $q(X)$, by the theorem, and the degree of q is d, this is the unique polynomial of degree $\leq d - 1$, satisfying $p(x_i) = y_i, i = 1, 2, \ldots, d$.

This is Lagrange's Interpolation Formula.

The reader may wish to check that $q_i(x_i) = q'(x_i)$ (formal derivative) and that division of both members by $q(X)$ yields the well-known partial fractions decomposition of $p(X)/q(X)$.

17.I.8 Example: Another interpolation example. What does it mean to solve the system

$$f(X) \equiv px + q \pmod{(X - a)^2}$$
$$f(X) \equiv rx + s \pmod{(X - b)^2}$$

where $a, b \in k, a \neq b$? Writing $px + q = p(X - a) + (q + pa) =: p(X - a) + t$ we see that the first congruence reads

$$f(X) = t + p(X - a) + g(X)(X - a)^2$$

Substituting $X = a$ we get $t = f(a)$. Differentiating (a purely formal operation in an arbitrary field) and substituting $X = a$ again (details left to the reader) we get $f'(a) = p$ which should surprise no one. The system obviously satisifies the assumptions of the C.R.T. since $(X - a)^2$ and $(X - b)^2$ have no non-trivial factor in common. So it has a solution, proving the existence of a polynomial with given values and first derivatives at two given points.

By the uniqueness part of the C.R.T., $f(X)$ is uniquely determined modulo $(X - a)^2(X - b)^2$, so may be chosen of degree ≤ 3, since any solution may be reduced modulo $(X - a)^2(X - b)^2$ using polynomial division. ∎

More generally, given field elements $a_i, i = 1, 2, \ldots k$, no two equal, and polynomials $g_i(X - a_i)$ of degrees $d_i - 1, \sum d_i = d + 1$, we can solve the system

$$f(X) \equiv g_i(X - a_i) \pmod{(X - a_i)^{d_i}} \quad i = 1, 2, \ldots, k$$

to find a polynomial f, of degreee $\leq d$, with prescribed Taylor expansions (of prescribed degrees) at the a_i, as long as the number of interpolation data (number of given values) sum up to the degreee of f plus 1 (note that f has $d+1$ coefficients).

An important use of the C.R.T. is fast polynomial and integer arithmetic. Everything is reduced modulo a suitably large integer (or polynomial) and further reduced modulo its various primary factors (powers of irreducible factors). The C.R.T. then helps us reconstruct the solution from its residues. See, e.g., Lidl-Pilz, Applied Abstract Algebra. (Springer Undergraduate Texts in Mathematics).

17.II . The Chinese Remainder Theorem, II.

First we define a recipe for forming big rings from smaller ones.

17.II.1 Definition: Let $R_1, R_2, \ldots, R_d$ be rings. We define their **direct sum**

$$R = R_1 \oplus R_2 \oplus \cdots \oplus R_d = \oplus_{i=1}^{d} R_i$$

to be the set of d-tuples $(r_1, r_2, \ldots, r_d)$, with componentwise addition and multiplication, i.e.,

$$(r_1, \ldots, r_d) + (s_1, \ldots, s_d) = (r_1 + s_1, \ldots, r_d + s_d)$$

$$(r_1, \ldots, r_d) \cdot (s_1, \ldots, s_d) = (r_1 s_1, \ldots, r_d s_d)$$

It is easy to check that this definition turns R into a ring with zero element $(0, 0, \ldots, 0)$ and identity element $1 = (1, 1, \ldots, 1)$. Obviously, the elements $(0, 0, \ldots, r_j, \ldots, 0)$, with zeros in all positions except the jth, form a subring isomorphic to R_j, with identity element $e_j = (0, 0, \ldots, 1, \ldots, 0)$. We will use the notation R_j for this subring, too.

The e_j:s satisfy $e_j^2 = e_j$, i.e., they are *idempotent*. Furthermore $i \neq j \Rightarrow e_i e_j = 0$ and $1 = e_1 + e_2 + \cdots + e_d$. We say that the e_j form a complete system of *orthogonal idempotents* for R.

Each subring R_j is not only a subring but an ideal, as well, since it is obviously closed under multiplication by elements of R:

$$(r_1, \ldots, r_j, \ldots, r_d)(0, \ldots, 0, s_j, 0, \ldots, 0) = (0, \ldots, 0, r_j s_j, 0, \ldots, 0)$$

It is generated, over R, by e_j; $R_j = Re_j$.

The ideal sum of all R_i's but one, R_j say, consists of all elements with a zero in the j:th position.:

$$(r_1, \ldots r_{j-1}, 0, r_{j+1}, \ldots, r_d)$$

We denote this ideal by I_j. It is easy to see that

$$R/I_j \simeq R_j$$

Simply define a surjective homomorphism $R \to R_j$ by $(r_1, \ldots, r_d) \to r_j$ and note that the kernel equals I_j.

So the R_j enter our construction simultaneously as subrings, ideals, and quotient rings.

17.II.2 Theorem. Chinese Remainder Theorem, version II. *R still a P.I.D. Let the ideals $I_j = (m_j), j = 1, 2, \ldots, d$ be pairwise comaximal. Then we have an isomorphism*

$$R/(I_1 \cap \ldots \cap I_d) = R/(I_1 \cdots I_d) \simeq \oplus_{j=1}^{d} R/I_j$$

Proof: Define

$$\varphi : R \to \oplus_{j=1}^{d} R/I_j$$

by

$$\varphi(r) = (r + I_1, r + I_2, \ldots, r + I_d)$$

φ is trivially a homomorphism. Its kernel is, equally trivially, $I := I_1 \cap \ldots \cap I_d$, so φ induces an *injection*

$$R/I \to \oplus_{j=1}^{d} R/I_j =: S$$

So we still have to prove surjectivity. Pick any

$$s = (r_1 + I_1, r_2 + I_2, \ldots, r_d + I_d) \in S$$

We wish to find an $r \in R$ mapping to s under φ. We want

$$\varphi(r) = (r + I_1, r + I_2, \ldots, r + I_d) = (r_1 + I_1, r_2 + I_2, \ldots, r_d + I_d)$$

so we must solve the system of congruences

$$r \equiv r_i \pmod{m_i},\ i = 1, 2, \ldots, d$$

This is possible by the first version of the C.R.T. ∎

We saw above that S possesses a full set of orthogonal idempotents $e'_j, j = 1, 2, \ldots, d$, $1 = e'_1 + \cdots + e'_d$. By the isomorphism just proved, this must hold for R/I too. Let e_j denote the pre-images of the e'_j under φ. Then $1 = e_1 + \cdots + e_d$ and we can view R/I as the direct sum of the subrings (or ideals) $(R/I)e_i$.

The inverse isomorphism

$$\varphi^{-1} : \quad \oplus_j R/I_j \to R/I$$

is given by

$$\varphi^{-1}(r_1, \ldots, r_d) = \sum_j r_j e_j,$$

check this. It is instructive, although not necessary, to check the homomorphism properties of the inverse, just to get the right feeling for the role of idempotents.

17.II.3 Example: let k be field, and $x_i, i = 1, 2, \ldots, d$ distinct elements of k. Let $q(X) = (X - x_1) \cdots (X - x_d)$, and $q_i(X) = q(X)/(X - x_i)$. Then the C.R.T., version II, shows that

$$k[X] \simeq \oplus_{i=1}^{d} k[X]/(X - x_i)$$

by an isomorphism sending a polynomial to its residue classes modulo $(X - x_i)$. By Example I.7., of the previous section, this mapping is essentially an evaluation mapping sending each polynomial to its values at the x_i.

We checked above that the system of congruences (for fixed i):

$$p_i(X) \equiv \delta_{ij} \pmod{(X - x_j)}, \quad j = 1, 2, \ldots, d$$

is satisified by

$$p_i(X) = \frac{q_i(X)}{q_i(x_i)}, \quad i = 1, 2, \ldots d$$

By the proof

$$\varphi(p_i) = (0, 0, \ldots 1, \ldots 0) = e_i'$$

with the "1" in the i:th position. Being preimages of the idempotents e_i', they are a full set of orthogonal idempotents for the ring $k[X]/(q(X))$. I leave it as an exercise to check, directly, that

$$p_1(X) + \cdots + p_d(X) \equiv 1 \pmod{q(X)}$$

$$p_i(X)p_j(X) \equiv \delta_{ij}p_i(X) \pmod{q(X)}$$

You need only check equality of both members at the zeros of $q(X)$ (why?).

The reader is invited to investigate the last example of the previous section in a similar manner.

17.II.4 Example: Let us illustrate the statement

$$\mathbf{Z}/(15) \simeq \mathbf{Z}/(3) \oplus \mathbf{Z}/(5)$$

By the proof of the C.R.T., the isomorphism is given by the mapping

$$\varphi(n + (15)) = (n + (3), n + (5))$$

sending a class modulo 15 to the pair of the corresponding classes modulo 3 and 5. The kernel consists of all classes $n + (15)$ where n belongs to both (3) and (5), hence is divisible by 3 and 5, hence is divisible by 15, i.e., the kernel consists of the zero class, whence φ is injective.

Surjectivity follows directly from this fact, and the fact that both members have 15 elements.

The general theory provides us with a complete orthogonal set of idempotents $e_i, i = 1, 2; 1 = e_1 + e_2$, mapping to the idempotents $(1, 0), (0, 1)$. They are found, on solving the systems

$$x \equiv 1 \pmod 3$$

$$x \equiv 0 \pmod 5$$

and

$$y \equiv 0 \pmod 3$$

$$y \equiv 1 \pmod 5$$

to be

$$e_1 = \overline{10}, e_2 = \overline{6}$$

(classes modulo 15). Check that

$$10^2 \equiv 10 \pmod{15};\ 6^2 \equiv 6 \pmod{15};\ 10 \cdot 6 \equiv 0 \pmod{15}$$

Each element in

$$\mathbf{Z}/(3) \oplus \mathbf{Z}/(5)$$

may be written

$$(m + (3), n + (5))$$

where m and n are uniquely determined modulo 3 and 5, respectively, so we may assume $m = 0, 1, 2$, $n = 0, 1, 2, 3, 4$.

By the C.R.T.- isomorphism the corresponding statement holds in $\mathbf{Z}/(15)$: each element may be written, in a unique manner, as

$$\overline{m}e_1 + \overline{n}e_2 = \overline{m}\overline{10} + \overline{n}\overline{6}$$

with $m = 0, 1, 2$, $n = 0, 1, 2, 3, 4$. In particular, $e_1 + e_2 = \overline{10} + \overline{6} = \overline{16} = \overline{1}$

Note, for instance, that

$$(\overline{m}e_1 + \overline{n}e_2)(\overline{p}e_1 + \overline{q}e_2) =$$

$$\overline{mp}e_1 + \overline{nq}e_2$$

This has an amusing representation in a 3×5-matrix. If we number the rows and columns from 0 to 2, and 0 to 4, respectively, we can represent the class of $m \cdot 10 + n \cdot 6$ as the element in position (m, n).

We immediately find the elements $\overline{k} = k \cdot \overline{1} = k \cdot (e_1 + e_2) = \overline{k \cdot 10 + k \cdot 6}$, $k = 0, 1, 2$, in positions $(0, 0), (1, 1), (2, 2)$:

$$\begin{pmatrix} 0 & ? & ? & ? & ? \\ ? & 1 & ? & ? & ? \\ ? & ? & 2 & ? & ? \end{pmatrix}$$

Next in line is $\overline{3} = \overline{3}(e_1 + e_2) = \overline{0}e_1 + \overline{3}e_2$ in position $(0, 3)$:

$$\begin{pmatrix} 0 & ? & ? & 3 & ? \\ ? & 1 & ? & ? & ? \\ ? & ? & 2 & ? & ? \end{pmatrix}$$

You have guessed the pattern: we proceed along the diagonal until we hit the bottom of a column. We then move to the top of the next column and move along a new diagonal:

$$\begin{pmatrix} 0 & ? & ? & 3 & ? \\ ? & 1 & ? & ? & 4 \\ ? & ? & 2 & ? & ? \end{pmatrix}$$

Similarly, on hitting the right wall of the matrix, we move on to the beginning of the next row obtaining

$$\begin{pmatrix} 0 & ? & ? & 3 & ? \\ ? & 1 & ? & ? & 4 \\ 5 & ? & 2 & ? & ? \end{pmatrix}$$

Once again we hit the bottom of a column and move to the top of the next one. And so on.

We finally wind up with

$$\begin{pmatrix} 0 & \underline{6} & 12 & 3 & 9 \\ \underline{10} & 1 & 7 & 13 & 4 \\ 5 & 11 & 2 & 8 & 14 \end{pmatrix}$$

Reading off the (1,0) and (0,1) elements we find our idempotents!

The first row consists of multiples of 6, modulo 15, the first column of multiples of 10, modulo 15.

Each element in the matrix is the sum of the first element in the same column and the first element in the same row, which is exactly what our direct sum decomposition states, quite concretely.

It *might* be easier to perceive the pattern if you extend the matrix periodically vertically as well as horizontally

$$\begin{pmatrix} 0 & 6 & 12 & 3 & 9 & 0 & 6 & 12 & 3 & 9 & \dots \\ 10 & 1 & 7 & 13 & 4 & 10 & 1 & 7 & 13 & 4 & \dots \\ 5 & 11 & 2 & 8 & 14 & 5 & 11 & 2 & 8 & 14 & \dots \\ 0 & 6 & 12 & 3 & 9 & 0 & 6 & 12 & 3 & 9 & \dots \\ 10 & 1 & 7 & 13 & 4 & 10 & 1 & 7 & 13 & 4 & \dots \\ 5 & 11 & 2 & 8 & 14 & 5 & 11 & 2 & 8 & 14 & \dots \\ \dots & \dots & & & & & & & & & \end{pmatrix}$$

This representation or interpretation of the C.R.T. is at the basis of the so-called Good-Thomas algorithm for Finite Fourier Transforms.

⋆ **17.II.5 Example:** There is another useful bijection between $\mathbf{Z}/(3) \oplus \mathbf{Z}/(5)$ and $\mathbf{Z}/(15)$, this time more conveniently defined in the opposite direction.

Namely,

$$\psi(m + (3), n + (5)) = 5(m + (3)) + 3(n + (5)) + (15) = 5m + 3n + (15)$$

ψ is well-defined and injective. Indeed

$$\overline{5m + 3n} = \overline{0}$$

implies that $5m + 3n$ is divisible by 15, hence a fortiori by 3; hence $5m$, hence also m, is divisible by 3 and, in the same manner, n is divisible by 5. Again, "injective" implies "surjective".

Or we might prove surjectivity first, using the fact that every integer p can be written $p = 5m + 3n$, by Bézout's Identity (5 and 3 being relatively prime).

However, ψ is not a ring homomorphism since it only maps sums to sums, not products to products (check!). So ψ is only an isomorphism of abelian groups (or $\mathbf{Z}$- modules, in the jargon of later sections.) Part of the secret of abstraction is to realize which part of a given structure to ignore and which to retain.

We may at any rate form a matrix similar to the one above. I leave it to the reader to interpret the following:

$$\begin{pmatrix} 0 & \underline{3} & 6 & 9 & 12 \\ \underline{5} & 8 & 11 & 14 & 2 \\ 10 & 13 & 1 & 4 & 7 \end{pmatrix}$$

An excellent introduction to the use of the C.R.T in fast algorithms is given by Chapter 11 of Richard Blahut: "The Theory and Practice of Error Correcting Codes" (Addison-Wesley).

17.II.6 Example: Let us work through a polynomial example. We work over the field $k = \mathbf{Z}/(2)$ and study the ring $k[X]/I$ where I is generated by the polynomial $X^3 + 1 = (X + 1)(X^2 + X + 1)$ the factors of which are irreducible, hence relatively prime.

So we have an isomorphism

$$\overline{R} = k[X]/(X^3 + 1) \simeq k[X]/(X + 1) \oplus k[X]/(X^2 + X + 1)$$

by a mapping sending the class of $f(X)$ to its classes modulo $X+1$ and X^2+X+1.

Again, if $f(X)+(X^3+1)$ is sent to the zero classes, $f(X)$ is divisible by $X+1$ and $X^2 + X + 1$, hence by their product, hence $f(X)$ belongs to the zero class modulo $X^3 + 1$, so our mapping is injective.

Both members may be viewed as vector spaces over k. Since each element in $k[X]/(X^3 + 1)$ has a unique representative of degree < 3 (requiring 3 coefficients) this ring has dimension 3 as a vector space over k. Similarly the two factors in the right member have dimensions 1 and 2, respectively, so the direct sum has dimension 3. By the Dimension Theorem, "injective" (nullspace dimension = 0) implies "surjective" (dimension of range = 3).

(Of course, with a finite ground field we could count elements as well. Both members have $2^3 = 8$ elements.).

From the proof of the theorem it is clear that we find our idempotents by solving Bézout's Identity:

$$1 = p(X)(X + 1) + q(X)(X^2 + X + 1) = e_1 + e_2$$

(The *classes* of e_1, e_2, modulo $X^3 + 1$, are the idempotents).

We could use Euclid's Algorithm for that kind of computation but a partial fractions decomposition may often be more convenient. Decomposing $1/(X^3 + 1)$ we get

$$\frac{1}{X^3 + 1} = \frac{1}{X + 1} + \frac{X}{X^2 + X + 1}$$

Multiplying by X^3+1 we get

$$1 = (X^2+X+1) + (X^2+X)$$

The classes of X^2+X+1 and X^2+X modulo X^3+1 are our idempotents, since they satisfy the systems of congruences of the proof. The reader should check this.

Of course, any element in $\overline{R}$ may be expressed in the idempotents. The coefficients will be uniquely determined modulo $X+1$, and X^2+X+1, respectively, so their representatives may be chosen of degrees <1 and <2. Omitting bars we find

$$\begin{aligned} 1 &= (X^2+X+1) + (X^2+X) \\ X &= (X^2+X+1) + X(X^2+X) \\ X^2 &= (X^2+X+1) + (X+1)(X^2+X) \end{aligned}$$

The remaining 5 elements are suitable $k-$linear combinations of these elements. I invite the reader to ponder over the following matrix

$$\begin{pmatrix} 0 & 0 & 1 & X & 1+X \\ 0 & 0 & \underline{X^2+X} & X^2+1 & 1+X \\ 1 & \underline{X^2+X+1} & 1 & X & X^2 \end{pmatrix}$$

The underlined elements are the idempotents. The first row and first column are the coefficients (replacing row and column indices). All computations are performed modulo X^3+1.

Of course, such a matrix scheme does not exist in the case of an infinite field k.

17.II.7 RSA Ciphers

In cryptology it is common practice to represent words as numbers or classes modulo some large integer. The RSA cipher is a very simple encryption-decryption scheme (at least empirically) with very good properties. It rests on the following simple theorem:

17.II.8 Theorem. . *Let $n = pq$ be a product of two distinct primes p and q. Let e and d be two numbers, relatively prime to $(p-1)(q-1)$. Assume $ed \equiv 1 \pmod{(p-1)(q-1)}$. Then, for every integer w,*

$$w^{ed} \equiv w \pmod{n}$$

Proof: By the C.R.T, we have an isomorphism

$$\mathbf{Z}/(pq) \simeq \mathbf{Z}/(p) \oplus \mathbf{Z}/(q)$$

so, letting w denote the class of w modulo n as well, we may write

$$w = (u, v)$$

where u, v are classes modulo p and q, respectively. Since

$$w^k = (u^k, v^k)$$

it is enough to prove the corresponding result for u and v, say u, i.e., that

$$u^{e\,d} = u$$

This is obvious if $u = 0$, so assume $u \neq 0$. By Fermat's Little Theorem

$$u^{p-1} = 1$$

As

$$ed = j(p-1)(q-1) + 1$$

for some j, we get

$$u^{e\,d} = u \cdot (u^{p-1})^{j(q-1)} = u \cdot 1 = u$$

∎

The RSA scheme is the following. Let n be a product of two very large primes (which are kept secret, of course). Transmit the word $c = w^e$ (reduced modulo n) for some e, relatively prime to $(p-1)(q-1)$. The receiver, who knows d, computes $c^d = w^{ed}$ (modulo n), thereby retrieving the word w. The cipher is safe as long as no one is able to factor the integer n. But no one has really proved that factoring is necessary, or equivalent to breaking the cipher.

The letters R,S,A are the initials of Rivest, Shamir, and Adleman.

17.III Boolean Rings

A showpiece in most courses i Discrete Mathematics is a structure theorem characterizing finite Boolean Algebras. They are isomorphic to the power set of a finite set M, with the usual operations of union, intersection, and complement.

Any Boolean algebra may be turned into a Boolean ring, to be defined below, and vice versa. As an application of the ideas of the preceding sections we prove the structure theorem in terms of rings.

Since we do not wish to assume more than necessary, the word "ring" will be taken in its most general sense. We do not assume multiplication to be commutative, and we do not assume the existence of a unity element. These properties will be *proved* for finite Boolean rings.

The most important examples of non-commutative rings are rings of matrices with elements in a field. All proper ideals of, e.g., $\mathbf{Z}$, may be viewed as rings, without unity. This, however, is not the proper way to view them and natural examples of rings without unity are hard to come by.

We are ready for our first definition

17.III.1 Definition: A **Boolean ring** is a (general) ring R in which every element is idempotent, $a^2 = a \quad \forall a \in R$.

17.III.2 Example:

a) The most obvious example is $\mathbf{Z}_2 = \mathbf{Z}/(2)$. The next most obvious example is a direct sum (product) of several copies of that ring, $R = (\mathbf{Z}_2)^d$.

b) The power set 2^M, M preferrably finite, is a Boolean ring under the operations:

$$+ : M_1 + M_2 = M_1 \cup M_2 \setminus M_1 \cap M_2$$

(symmetric difference, "exclusive or"),

$$\cdot : M_1 \cdot M_2 = M_1 \cap M_2$$

Obviously $N \cdot N = N$ for any subset N.

I leave all verifications to the reader. The empty set serves as zero element, the whole set M as 1. The additive inverse of any subset N is N itself. Note that the union of two subsets is given by

$$M_1 \cup M_2 = M_1 + M_2 + M_1 \cdot M_2$$

The complement of N is $M + N$.

∎

Obviously, subrings and quotient rings of Boolean rings are themselves Boolean.

Before proving general theorems on Boolean rings we show that these two examples are really the same.

17.III.3 Theorem. *Let M be a finite set, with d elements. Then there is an isomorphism*

$$2^M \simeq \mathbf{Z}_2^d$$

of Boolean rings.

Proof: We may assume that M is the set $\{1, 2, \ldots d\}$. The elements of $\mathbf{Z}_2^d$ are strings $(r_1, r_2 \ldots, r_d)$ where each r_i equals 0 or 1. If $N \subset M$ we may define a mapping

$$\varphi : N \to \mathbf{Z}_2^d$$

by $\varphi(N) = (r_1, r_2, \ldots, r_d)$ where $r_j = 1$ if $j \in N$ and $r_j = 0$ otherwise. It easy to check that sums are mapped to sums, and products to products. Details are left to the reader. ∎

Our Structure Theorem is preceded by three Lemmas

17.III.4 Lemma. *A Boolean ring R has characteristic 2, i.e., $a + a = 0$, $\forall a \in R$.*

Proof:

$$2a = a + a = (a + a)^2 = a^2 + 2a^2 + a^2 = 4a^2 = 4a$$

whence $2a = 4a - 2a = 0$. ∎

17.III.5 Lemma. *Boolean rings are commutative*

Proof:

$$a + b = (a + b)^2 = a^2 + ab + ba + b^2 = a + ab - ba + b$$

(we used char.$R = 2$ in the last step), whence

$$ab - ba = 0 \qquad \forall a, b \in R$$

∎

17.III.6 Lemma. *Any finite non-zero Boolean ring has a unity element*

Proof: Among all ideals Re pick one that is maximal under inclusion, i.e., not properly contained in any other ideal of the same kind. We are done if we can prove that $Re = R$ because then, for an arbitrary element $re \in R$, $(re)e = re^2 = re$.

Suppose $Re \neq R$. Pick an element $f \notin Re$ and form the element $e' = e + f + ef$. Then

$$fe' = f(e + f + ef) = fe + f^2 + ef^2 = f + 2ef = f$$

so $f \in Re'$ (we used characteristic 2 again). In the same manner we prove $e \in Re'$, whence $f \in Re'$, $\quad Re \subset Re'$ contradicting the maximality of Re. ∎

Remark: The choice of e' was guided by our set-theoretic example. Think of e and f as subsets of some finite set. We wanted something larger so we picked the element corresponding to their union.

We are ready to prove the Structure Theorem:

17.III.7 Theorem. *Let R be a finite non-zero Boolean ring. Then*

a)

$$R \simeq 2^M$$

for some finite set M

b)

$$R \simeq \oplus_j R_j$$

where each R_j is isomorphic to $\mathbf{Z}_2$

Proof: In view of our previous Theorem it is enough to prove the second statement. The proof proceeds by induction on the cardinality of R. The basis step, $\sharp R = 2$ is obvious, because then $R \simeq \mathbf{Z}_2$.

So assume $\sharp R = n > 2$ and the Theorem proved for all Boolean rings with fewer elements. Pick an element $e \neq 0, 1$. Note that $e(1+e) = e + e^2 = e + e = 0$ so the elements $e, 1+e$ are orthogonal idempotents. They thereby provide us with a direct sum decomposition of the ring R allowing us to apply the induction hypothesis to two smaller rings.

So let us form the subrings (even ideals) $Re, R(1+e)$, Boolean in themselves, neither zero (because they contain e and $1+e$, respectively. They are both different from R, since the first ring is annihilated by $1+e$ and the second by e.

(actually the two rings are isomorphic to $R/(1+e)$ and $R/(e)$, respectively)

So both of them are, by induction, isomorphic to the direct sum of a number of $\mathbf{Z}_2$:s. We exhibit the ring R as the direct sum of these rings, thereby finishing the inductive step. The isomorphism is the mapping

$$\psi : Re \oplus R(1+e) \to R$$

given by

$$\psi(re, s(1+e)) = re + s(1+e)$$

I leave it to the reader to check the multiplicative property of this mapping, additivity being obvious.

It is surjective, since $(re, r(1+e))$ maps to r. It is injective, since

$$\psi(re, s(1+e)) = re + s(1+e) = 0$$

yields, on multiplication by e,

$$re^2 + s(1+e)e = re = 0$$

and, similarly, $s(1+e) = 0$. By bijectivity, and the homomorphism property, there is no need to check that the unity element $(e, 1+e)$ maps to the unity of R, but that, too, is easy.

∎

Remark 1: The quickest proof is to appeal to the C.R.T. directly. As $e+1+e=1$, the ideals (e) and $(1+e)$, both of them proper and non-zero, are comaximal. Their intersection equals their product which is zero, since $e(1+e)=0$.

So we then have the C.R.T. isomorphism

$$R \to R/(1+e) \oplus R/(e)$$

and may apply the inductive hypothesis to the right member.

The connection with our proof is the following.

Consider the mapping

$$\alpha : R \to Re; \qquad \alpha(r) = re$$

It is a surjective homomorphism of rings. Additivity is obvious. Multiplicativity follows from:

$$\alpha(rs) = rse = rse^2 = re \cdot se = \alpha(r) \cdot \alpha(s)$$

The kernel is easy to determine: $\alpha(r) = re = 0 \Leftrightarrow re + r = r; r = r(1+e)$, so α induces an isomorphism

$$R/(1+e) \simeq Re$$

and, similarly, we have

$$R/(e) \simeq R(1+e)$$

whence isomorphisms

$$R \to R/(1+e) \oplus R/(e) \to Re \oplus R(1+e)$$

Backtracking the reader may easily convince himself that the mapping of our proof is the inverse of this composition. I chose the route of the proof in order to clarify the structure of R as *internal* direct sum of the two subrings Re and $R(1+e)$

The idea behind the inductive argument is that idempotents split the ring into direct factors. If one at least of the two idempotents e, or $1+e$, can be decomposed into two orthogonal idempotents the ring may be further decomposed. Finally we wind up with a full orthogonal set of indecomposable idempotents, *primitive* idempotents, or *atoms*, of which every ring element is a $\mathbf{Z}_2$-linear combination. The atoms correspond to the singleton subsets (or elements) of a finite set M, and any sum of these corresponds to the subset having these elements.

Remark 2: The reader may easily give the dictionary transforming a Boolean algebra into a Boolean ring, or vice versa. Granted this, our result painlessly translates into the corresponding structure theorem for Boolean algebras.

17.IV Phi and Mu

This section is devoted to the extremely useful Euler Phi and Moebius Mu functions. It will culminate in the Moebius Inversion Formula. The C.R.T plays a minor role in the theory.

17.IV.1 Lemma. *Suppose the (commutative) ring (with 1), R, is the direct sum of rings $R_i, i = 1, 2, \ldots, d$:*

$$R = R_1 \oplus R_2 \oplus \ldots \oplus R_d$$

Then $u = (r_1, \ldots, r_d) \in R$ is a unit in R if and only if all the r_i are units in their respective R_i.

Proof:

$$r_i s_i = 1 \forall i \iff (r_1, \ldots, r_d)(s_1, \ldots, s_d) = (1, 1, \ldots, 1)$$

∎

17.IV.2 Definition: We define the **Euler Phi function** for integers $n > 1$ by $\varphi(n)$ = the number of invertible classes modulo n = the number of integers m, $1 \leq m \leq n-1$, with $(m, n) = 1$.

By convention, $\varphi(1) = 1$.

17.IV.3 Example:

$$\varphi(6) = 6 - 3 - 2 + 1 = 2$$

We start with the six classes and subtract the three classes divisible by 2 and the two classes divisible by 3. But then we subtract the zero class twice so it must be put back again.

17.IV.4 Lemma. *The Phi function is* multiplicative, *i.e.*

$$(m, n) = 1 \Rightarrow \varphi(mn) = \varphi(m)\varphi(n)$$

Proof: Follows from the isomorphism

$$\mathbf{Z}/(mn) \simeq \mathbf{Z}/(m) \oplus \mathbf{Z}/(n)$$

and the first Lemma.

∎

17.IV.5 Example:

$$\varphi(6) = \varphi(2) \cdot \varphi(3) = 1 \cdot 2 = 2$$

From this we see that the Phi function is fully determined by its values for prime powers p^k. In fact we have the following Theorem.

17.IV.6 Theorem. *For p a prime number it holds that*

$$\varphi(p) = p - 1$$

$$\varphi(p^k) = p^k - p^{k-1} = p^k(1 - \frac{1}{p})$$

For an arbitrary positive integer n we have

$$\varphi(n) = n \prod (1 - \frac{1}{p})$$

the product extending over all prime factors of n

Proof: For the proof of the second statement, just note that $(m, p^k) = 1$ if and only if m is not divisible by p. So from the p^k classes modulo p^k delete the p^{k-1} ones corresponding to integers divisible by p.

∎

17.IV.7 Lemma.

$$\sum_{d|n} \varphi(d) = \sum_{d|n} \varphi(\frac{n}{d}) = n$$

Proof: The first equality comes from the fact that an arbitrary divisor of n may also be written in the form n/d.

For the second, let $1 \leq m \leq n-1$, $(m, n) = d$, (so that, in fact, $1 \leq m \leq n - d$) Then $(m/d, n/d) = 1$ and $1 \leq m/d \leq n/d - 1$. This sets up a bijection between those m, $1 \leq m \leq n-1$ with $(m, n) = d$ and the invertible classes modulo n/d.

Summing over all possible d we get the result. ∎

17.IV.8 Definition: Let ϕ, ψ be functions from the positive integers to **Z**. Their **Dirichlet product** (or, **multiplicative convolution**) is defined by

$$\phi * \psi(n) = \sum_{d|n} \phi(d)\psi(\frac{n}{d}) = \sum_{de=n} \phi(d)\psi(e)$$

Here ϕ and ψ may be interchanged, so the Dirichlet product is commutative. It is also associative, in fact:

$$\phi * (\psi * \chi)(n) = (\phi * \psi) * \chi(n) = \sum_{def=n} \phi(d)\psi(e)\chi(f)$$

There is an identity for this product, namely $\iota, \iota(1) = 1, \iota(n) = 0, n > 1$. "Identity" means that $\iota * \phi = \phi$, check this. If $\phi * \psi = \iota$, then we say, of course, that ϕ and ψ are (Dirichlet) inverses of one another.

If two functions from $\mathbf{Z}_+$ to $\mathbf{Z}$ are multiplicative, then so is their Dirichlet product. Exercise (needed below). Every such function has an inverse. If the given function is multiplicative, then so is its inverse. We will not need that observation.

Another useful multiplicative function is the constant function assigning the value 1 to each positive integer. It will be denoted by [1]. Still another example is the identical function I, $I(n) = n$. Do not confuse the identical function I with the identity ι defined above!

Obviously,

$$\sum_{d|n} \phi(d) = [1] * \phi(n)$$

Our result above on the Euler function could have been proved using the fact that both members are mutliplicative functions of n. It is easy to prove for prime powers.

17.IV.9 Moebius

We now wish to find the inverse of the constant function [1]. It will be denoted by μ.

Let us try low prime powers:

$$[1] * \mu(1) = 1 \cdot \mu(1) = 1$$

whence $\mu(1) = 1$; then

$$[1] * \mu(p) = 1 \cdot \mu(p) + 1 \cdot \mu(1) = 0$$

whence $\mu(p) = -1$; and then

$$[1] * \mu(p^2) = 1 \cdot \mu(1) + 1 \cdot \mu(p) + 1 \cdot \mu(p^2) = 0$$

whence $\mu(p^2) = 0$.

From this a pattern emerges. We must have, for any prime p, $\mu(p) = -1$, $\mu(p^k) = 0, k > 1$.

If μ is to be multiplicative we are led to the following definition

17.IV.10 Definition: The **Moebius Mu Function** $\mu : \mathbf{Z}_+ \to \mathbf{Z}$ is defined by

$$\mu(n) = \mu(p_1 p_2 \cdots p_k) = (-1)^k$$

if n is the product of distinct (simple) prime factors p_i, and

$$\mu(n) = 0 \text{ otherwise}$$

It is obvious that μ is multiplicative. We can now state and prove the desired result

17.IV.11 Theorem. *The functions μ and $[1]$ are Dirichlet inverses of one another.*

Proof: Obviously $[1] * \mu(1) = 1$. Also, it is easy to see that

$$[1] * \mu(p^k) = 1 \cdot \mu(p) + 1 \cdot \mu(1) = -1 + 1 = 0$$

all other terms of the sum defining the Dirichlet product being zero.
The identity

$$[1] * \mu(n) \equiv \iota(n)$$

now follows because both members are multiplicative and agree on prime powers.
∎

We now arrive at the beautiful Moebius Inversion Formula.

17.IV.12 Theorem. *If ϕ is a function from $\mathbf{Z}_+$ to $\mathbf{Z}$ and*

$$\psi(n) = [1] * \phi(n) = \sum_{d|n} \phi(d)$$

then

$$\phi(n) = \mu * \psi(n) = \sum_{d|n} \mu(\frac{n}{d})\psi(d)$$

Proof:

$$\phi = \iota * \phi = (\mu * [1]) * \phi = \mu * ([1] * \phi) = \mu * \psi$$

∎

We note the following Corollary

17.IV.13 Corollary.

$$\varphi(n) = \sum_{d|n} \mu(\frac{n}{d})d = \mu * I(n)$$

Proof: This is an immediate consequence of the above Theorem and the identity

$$\sum_{d|n} \varphi(d) = n$$

∎

We leave it to the reader to prove the following multiplicative version:

17.IV.14 Theorem. . *If ϕ is a function from $\mathbf{Z}_+$ to an abelian (multiplicative) group G, and*

$$\psi(n) = \prod_{d|n} \phi(d),$$

then

$$\phi(n) = \prod_{d|n} \psi(\frac{n}{d})^{\mu(d)}$$

∎

17.IV.15 Example:

$$\varphi(6) = \mu(1) \cdot 6 + \mu(2) \cdot 3 + \mu(3) \cdot 2 + \mu(6) \cdot 1 =$$

$$= 6 - 3 - 2 + 1 = 2$$

for the third time.

17.IV.16 An Application

As an application of the Euler function we prove that the multiplicative group of a finite field F is cyclic.

Let us recall that a finite field has $p^n =: q$ elements, where p is a prime number. The non-zero elements are characterized as the distinct roots of the polynomial $X^{q-1} - 1$. The possible orders d of the multiplicative group are divisors of $q - 1$, $q - 1 = de$. It is easy to see, for such d, that the polynomial $X^d - 1$ divides $X^{q-1} - 1$:

$$X^{de} - 1 = (X^d - 1)(X^{(d-1)e} + X^{(d-2)e} + \ldots + 1)$$

So, for each d, $d|q - 1$, the polynomial $X^d - 1$ has exactly d roots in F.

We now let $\Psi(c)$ denote the number of elements of (exact) multiplicative order c in F. (If c does not divide $q - 1$, then $\Psi(c) = 0$) If c divides d, and d divides $q - 1$, then any element of order c is a root of $X^d - 1$, whence

$$\sum_{c|d} \Psi(c) = d = \sum_{c|d} \varphi(c)$$

The Moebius Inversion Formula then yields, for $d|q-1$,

$$\Psi(d) = \sum_{c|d} \mu(c)\frac{d}{c} = \varphi(d)$$

In particular,

$$\Psi(q-1) = \varphi(q-1) > 0$$

(unless $q = 2$, of course) which establishes the existence of at least one element of order $q-1$. ∎

17.IV.17 Another Application.

Let $q = p^n$ still denote the number of elements of a finite field F.

A well-known result states that the polynomial

$$X^q - X,$$

is the product of all monic polynomials, irreducible over $k = \mathbf{Z}_p$, the degree of which divides n.

We wish to find the product P_n of those polynomials whose degree equals n, i.e., those polynomials, the roots a of which generate F, in the sense $F = K[a]$. Please do not confuse this with roots generating the multiplicative group of the field. These *primitive* roots, and their corresponding primitive polynomials, are fewer.

Obviously,

$$X^q - X = \prod_{d|n} P_d$$

Applying the multiplicative version of the Moebius formula to the group of non-zero rational functions over k, we immediately obtain:

$$P_n(x) = \prod_{d|n} (X^{q^d} - X)^{\mu(n/d)}$$

The reader is invited to work out a few examples of his own, and also to determine the number of irreducible polynomials of given degree, and the corresponding formulae for primitive polynomials of given degree (their number and product).

17.V Cyclic codes (assai presto).

In Coding Theory binary words $(a_0, a_1, \ldots a_{n-1}), a_i \in \mathbf{Z}_2$ are often conceived of as polynomials $a_0 + a_1X + \ldots + a_{n-1}X^{n-1}$. The words to be transmitted are mapped into code words possessing greater redundance. They should be "farther apart", the distance between two words simply being the number of differing digits.

If errors occur, the transmitted code word is interpreted as its nearest neighbor, w.r.t. to this distance, if there be one. If the minimum distance between any two code words is $\geq 2t + 1$, this procedure will correct t errors, i.e., if no more than t errors occur there is a unique nearest codeword to the word actually received.

For purposes of encoding and decoding highly structured codes are desirable, for economy and expedience. The first obvious simple requirement is linearity, i.e., that the code words form a vector space over $\mathbf{Z}/(2)$. Of course, one then wants the encoding mapping to be linear.

A very useful additional requirement is *cyclicity*, i.e., if the word $a_0 + a_1X + \ldots a_{n-1}X^{n-1}$ belongs to the code then so does $a_{n-1} + a_0X + a_1X^2 + \ldots a_{n-2}X^{n-1}$. The reader can easily convince himself that this cyclic shift is the same as multiplication by X, modulo $X^n - 1 (= X^n + 1)$. (multiply, and replace X^n by 1.) We say that the code is invariant under multiplication by (the class of) X (modulo $X^n - 1$).

Of course, if the code is invariant under multiplication by X it is invariant under multiplication by any power of X (several cyclic shifts) and, by linearity, under multiplication by (the class of) any polynomial. So linear cyclic codes may be conceived of as ideals of the quotient ring $R = k[X]/(X^n - 1), k = \mathbf{Z}/(2)$.

By a general correspondence theorem there is a bijection between overideals of $(X^n - 1)$ in $k[X]$ and ideals of the quotient ring. The overideals in question are generated by the factors of $X^n - 1$ so to there is a bijection between cyclic codes and monic factors $g(X)$ of $X^n - 1$. $g(X)$ is the *generator polynomial* of the code. The polynomial $h(X) = (X^n - 1)/g(X)$ is the *parity check polynomial* of the code. The class $\overline{r}(X)$ belongs to the code iff $\overline{r(X)h(X)} = \overline{0}$.

⋆ 17.V.1 Generating idempotents

If n is even, $n = 2m$, then $X^n - 1 = (X^m - 1)^2$, by "Freshman's Dream", which is wasteful. From now on we assume n odd. In that case $\frac{d}{dx}(X^n - 1) = nX^{n-1} = X^{n-1}$. Since $(X^{n-1}, X^n - 1) = 1$, $X^n - 1$ has no multiple factors, i.e. it factorizes as $X^n - 1 = m_1(X)m_2(X) \cdot \ldots \cdot m_k(X)$ where the $m_i(X)$ are distinct *irreducible* monic polynomials over k, hence pairwise relatively prime. So the C.R.T yields

$$\overline{R} = k[X] \simeq \oplus_{i=1}^{k} k[X]/(m_i(X)) = \oplus_i F_i$$

where the F_i are *fields*.

Let I be an ideal of $\overline{R}$ and I' its image under the isomorhism. If $(r_1, r_2, \ldots, r_k)$ is a non-zero element of I' with, say $r_1 \neq 0$, then $(1, 0, \ldots 0)(r_1, r_2, \ldots r_k) =$

$(r_1, 0, 0 \ldots, 0) \in I'$. Since F_1 is a field any $(a, 0, \ldots, 0), a \in K$ belongs to I' (simply use $a = ar_1 r_1^{-1}$ and multiply by $(ar_1^{-1}, 0, \ldots 0)$). From this it follows easily that I' is the direct sum of some of the F_i and that a generator is given by the sum of the corresponding idempotents e_i'

So the number of possible subcodes (omitting the trivial zero code) is $2^k - 1$.

Going back to $\overline{R}$ we see that I is generated by a sum of some of the e_i. Now the sum of any number of orthogonal idempotents is itself an idempotent, e.g., $(e_1 + e_2)^2 = e_1^2 + 2e_1e_2 + e_2^2 = e_1 + 0 + e_2$ so I has an idempotent generator $w(X)$, say.

(Actually, the same computation shows that in characteristic 2 the sum of *any* two idempotents is an idempotent, so the idempotents form a ring, as well as a $\mathbf{Z}/(2)$-vector space).

$w(X)$ differs from the $g(X)$ of the previous section by a factor $r(X)$ which is invertible modulo $X^n - 1$, i.e., which has no non-trivial factors in common with $X^n - 1$ (i.e., no common zeros in a splitting field of $(X^n - 1)$.)

The roots of $X^n - 1$, in some suitable splitting field, form a group with n elements, since $a^n = 1, b^n = 1, \Rightarrow (a^{-1})^n = 1, (ab)^n = 1$. By a general theorem in field theory this finite group is cyclic. We denote a generator by α and write the roots as $\alpha^i, i = 0, 1, 2, \ldots n - 1$.

From $w(X)^2 \equiv w(X) \pmod{X^n - 1}$ we see $w(\alpha^i)^2 = w(\alpha^i)$, i.e., $w(\alpha^i) = 0$ or 1 (since the equation $X^2 = X$ has only these two roots in a field). From the equation $w(X) \equiv r(X)g(X) \pmod{X^n - 1}$, and $r(\alpha^i) \neq 0, i = 0, 1, \ldots, n - 1$, we see that $w(\alpha^i)$ equals zero whenever $g(\alpha^i)$ does and equals 1 otherwise.

Of course, $w(X)$ may be determined directly from Bézout's Identity. Write $X^n - 1 = g(X)h(X), (g(X), h(X)) = 1$ and solve

$$1 = p(X)g(X) + q(X)h(X); \quad w(X) = p(X)g(X)$$

as usual.

$1 - w(X) = q(X)h(X)$ is a parity check idempotent. Check(!) this.

The results of this section are proved in Blahut using spectral techniques (Discrete Fourier Transforms).

17.V.2 BCH Codes

If the generator polynomial $g(X) \in k[X]$, hence every member of the code, is chosen to have the roots $\alpha^b, \alpha^{b+1}, \ldots, \alpha^{b+d-2}$, $d - 1$ in number, at least d coefficients of any non-zero member are non-zero. We will prove this below. Since the difference of any two codewords is again a codeword (the code being linear) this means that the distance between any two codewords is at least d (recall that the distance is the number of differing coefficients).

To prove this we introduce the "Discrete Fourier Transform (DFT)" polynomial of

a codeword $q(X)$. It is defined to be

$$Q(X) = \sum_{i=0}^{n-1} q(\alpha^{-i})X^i$$

This actually depends only on the class of $q(X)$ modulo $X^n - 1$, since the roots of the latter are involved. $Q(X)$ is a polynomial with coefficients in the splitting field of $X^n - 1$.

There is an *inversion formula*:

$$q(X) = \sum_{j=0}^{n-1} Q(\alpha^j)X^j$$

It is enough to prove equality of the two members for $X = \alpha^{-k}, k = 0, 1, \ldots, n-1$. Indeed, the difference of the two members is then a polynomal of degree $\leq n-1$, with n zeros, hence equal to the zero polynomial.

Let $R(X)$ denote the right member.

Plugging the first expression into the right member we obtain

$$R(X) = \sum_{i,j} X^j \alpha^{ij} q(\alpha^{-i})$$

and

$$R(\alpha^{-k}) = \sum_{i,j} \alpha^{(i-k)j} q(\alpha^{-i})$$

Summing, w.r.t j, over all terms with $i = k$, we get $q(\alpha^{-k})$ times n ones, adding up to $q(\alpha^{-k})$. Recall that n is odd, $\equiv 1 \pmod 2$.

Summing over terms with a fixed $i \neq k$ we get the geometric series

$$q(\alpha^{-i}) \sum_{j=0}^{n-1} \alpha^{j(i-k)} =$$

$$= q(\alpha^{-i}) \frac{1 - \alpha^{n(i-k)}}{1 - \alpha^{i-k}} = 0$$

So

$$R(\alpha^{-k}) = q(\alpha^{-k}), \quad k = 0, 1, \ldots n-1$$

as claimed.

Remark: In arbitrary characteristic p, if n is relatively prime to p, the inversion formula reads

$$q(X) = \frac{1}{n} \sum_{j=0}^{n-1} Q(\alpha^j)X^j$$

the only difference being the extra factor $1/n$. This is obvious from the proof above. In some texts the DFT (or Mattson-Solomon) polynomial is defined without the minus signs. In that case the signs must be moved to the inversion formula.

Now, if exactly t of the coefficients of $q(X)$ are non-zero, then from the inversion formula we see that $Q(X)$ has exactly $n-t$ roots that are powers of α.

From the first expression we see that if $q(X)$ has the $d-1$ roots prescribed above (consecutive α-powers) then $Q(X)$ will have $d-1$ consecutive coefficients equal to zero.

Multiplying by X will obviously not affect the number of α-power roots of $Q(X)$. Neither will (subsequent) reduction modulo X^n-1 (since α-powers are roots of X^n-1):

$$XQ(X) = k(X^n-1) + r(X)$$

yields

$$Q(\alpha^i) = 0 \iff r(\alpha^i) = 0$$

These two operations amount to a cyclic shift of the coefficients. So, on performing a number of shifts, if necessary, we may assume that the *first* $d-1$ of them are zero, i.e., $Q(X)$ can be replaced by a polynomial of degree $n-1-(d-1) = n-d$ without changing the number of α-power roots.

So their number, $n-t$, is $\leq n-d$, i.e., $t \geq d$.

The letters B, C, H are the initials of Bose, Ray-Chauduri, and Hocquenghem.

17.V.3 Example: There is no reason to expect that all non-zero roots of a $Q(X)$ are $\alpha-$powers. So the true minimum distance may actually be larger than the "designed distance" d.

By way of example, let us consider $n = 7$. It is easy to check that

$$X^7 - 1 = X^7 + 1 = (X+1)(X^3+X+1)(X^3+X^2+1)$$

and that the factors are irreducible.

Consider the code generated by $g(X) = X^3 + X^2 + 1$. Since 7 is a prime number any root $\neq 1$ will generate the group of roots of $X^7 - 1$.

Let α denote one root of $g(X)$ in a suitable extension field (actually the field with 8 elements). By general principles (Freshman's Dream, Frobenius automorphism) the remaining roots are α^2, α^4. These $\alpha-$powers are then roots of all codeword polynomials. Since we have only two consecutive $\alpha-$ power roots, α and α^2, our theorem predicts that the minimum distance code is ≥ 2. But, actually, it is $= 3$.

As a vector space over $\mathbf{Z}/(2)$ our code is spanned by

$$g(X) = X^3 + X^2 + 1$$
$$Xg(X) = X^4 + X^3 + X$$
$$X^2g(X) = X^5 + X^4 + X^2$$
$$X^3g(X) = X^6 + X^5 + X^3$$

By a tedious, but trivial, computation the reader could check that any of the 16 linear combinations (i.e., any sum) of these polynomials has at least 3 non-zero coefficients. However, this is unnecessary. Simply remark that a polynomial of the form x^k or $x^m + x^n, 0 \leq m < n < 7$ couldn't possibly have the root α. The first is obvious, the latter because we would then have $1 + \alpha^{n-m} = 0$ in conflict with the true order 7.

The code of our example is the so-called Hamming (7,4) code. The 7 refers to the *blocklength*, the 4 (= 7 minus degree of generator) to the *dimension* of the code.

s

17.VI Linear Recursion

17.VI.1 The E operator

Let k be field. We will be concerned with sequences

$$\mathbf{s} = (s_d)_{d=0}^{\infty}$$

of elements in k. Such sequences may be added, componentwise, and multiplied with scalars (elements of k), again componentwise, so they constitute a vector space over k, albeit of infinite dimension.

We will sometimes write

$$s_d = \mathbf{s}(d)$$

so we can attach some indices to the letter **s**!

Let $a_0 = 1, a_1, a_2, \ldots, a_n \neq 0$ be given elements of k. A *linear recurring sequence* is a sequence **s** satisfying

$$s_{d+n} + a_1 s_{d+n-1} + a_2 s_{d+n-2} + \ldots + a_n s_d = 0 \qquad (*)$$

for all $d \geq 0$. It is clear that such a sequence is uniquely determined by its *inital data* (or *state vector*)

$$\begin{pmatrix} s_0 \\ s_1 \\ \vdots \\ s_{n-1} \end{pmatrix}$$

and that each such state vector really gives rise to a sequence satisfying (*) (since each element can be expressed in the n elements preceding it). Further the correspondence between state vectors (elements of k^n) and solutions is not only bijective, but linear as well, so:

17.VI.2 Theorem. *The solutions of (*) form a vector space, of dimension n.*

This last statement does not refer specifically to one given field k and so is valid for any field K containing the coefficients. For instance, if the equation has real coefficients, we may consider solutions in **C** alongside solutions in **R**. The complex dimension of the space of **C**-solutions then equals the real dimension of the **R**-solutions. A basis for the real space will be one for the complex space, as well.

Piling all equations (*) on top of each other we see that (*) expresses a linear relation between the *sequences*

$$(s_{d+n})_{d=0}^{\infty}, (s_{d+n-1})_{d=0}^{\infty}, \ldots, (s_d)_{d=0}^{\infty}$$

Here each sequence is obtained from the next one by deleting the first term. We denote this linear mapping, *the shift operator*, by E. So if

$$\mathbf{s} = s_0, s_1, \ldots s_d, \ldots$$

then

$$E\mathbf{s} = s_1, s_2, \ldots, s_{d+1}, \ldots$$

or

$$(E\mathbf{s})(d) = \mathbf{s}(d+1) = s_{d+1}$$

Equation (*) therefore reads

$$E^n\mathbf{s} + a_1E^{n-1}\mathbf{s} + \ldots a_n\mathbf{s} = \mathbf{0}$$

or

$$(E^n + a_1E^{n-1} + \ldots a_nI)\mathbf{s} = 0$$

$$p(E)\mathbf{s} = 0$$

where $p(X) = X^n + a_1X^{n-1} + \ldots a_n$ is the *generating*, or *characteristic, polynomial* of equation (*), and I is the identity operator.

Note that $E\mathbf{s}$ is a solution if $\mathbf{s}$ is:

$$p(E)E\mathbf{s} = Ep(E)\mathbf{s} = E0 = 0.$$

17.VI.3 Matrix notation

It is often useful to regard a single recursion as a system in the sequences $\mathbf{s}, E\mathbf{s}, \ldots, E^{n-1}\mathbf{s}$. Obviously

$$\begin{pmatrix} s_{j+1} \\ s_{j+2} \\ \vdots \\ s_{j+n} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \ldots & & & & 0 \\ -a_n & -a_{n-1} & -a_{n-2} & \ldots & -a_1 \end{pmatrix} \begin{pmatrix} s_j \\ s_{j+1} \\ \vdots \\ s_{j+n-1} \end{pmatrix}$$

Denoting the n/n-matrix by A, setting $j = 0, 1, \ldots, d-1$, and iterating, we find

$$\mathbf{s}^{(d)} := \begin{pmatrix} s_d \\ s_{d+1} \\ \vdots \\ s_{d+n-1} \end{pmatrix} = A^d \begin{pmatrix} s_0 \\ s_1 \\ \ldots \\ s_{n-1} \end{pmatrix} = A^d\mathbf{s}^{(0)}$$

from which s_d may be found on multiplication by the row matrix $(1, 0, \ldots, 0.)$.

More important, $\det A = \pm a_n \neq 0$, by assumption. Note that the only admissible product without zero factors comes from the ones above the diagonal and the far left element in the bottom line.

So A, hence also A^d, is invertible. This means that the initial data $\mathbf{s}^{(0)}$ may be reconstructed from any segment $\mathbf{s}^{(d)}$ of the sequence.

In a finite field, since there are only a finite number of possible segment vectors, we must get a repeat:

$$A^{n+p}\mathbf{s}^{(0)} = A^n\mathbf{s}^{(0)} \qquad (***)$$

for some $n > 0, p \geq 0$. Since A is invertible we get, on multiplying by A^{m-n},

$$A^{m+p}\mathbf{s}^{(0)} = A^m\mathbf{s}^{(0)}$$

for *all* m. In other words, the sequence is *periodic*.

17.VI.4 The equation $(E - I)^n\mathbf{s} = \mathbf{0}$

We will ultimately reduce everything to the equation $(E - I)^n\mathbf{s} = 0$. We proceed to describe the solution space of this equation.

In the case of characteristic 0, it is fairly easy to prove, by induction, that the sequences $\mathbf{s}_k(d) = d^k, k = 0, 1, \ldots n-1$ constitute a system of n linearly independent solutions and we are done in this case; cf. the remark below.

However, in positive characteristic p, this is no longer true for $n > p$. (The impatient reader may skip the following discussion) For instance, in characteristic 2, the equation

$$(E - I)^4\mathbf{s} = (E^4 - I)\mathbf{s} = 0; \quad s_{k+4} = s_k$$

is satisfied by all sequences of period 4 and thus cannot be expressed as a function of d modulo 2. Here is where binomial coefficients come to the rescue.

To avoid confusion we will reserve the notation $\binom{d}{k}$ for the expression

$$\binom{d}{k} := \frac{d!}{k!(d-k)!}$$

For the true binomial coefficients, i.e., the coefficients of

$$(a + b)^d$$

we write $C(d; k)$:

$$(a + b)^d = \sum_{k=0}^{d} C(d; k)a^k b^{d-k}$$

They are, of course, our ordinary binomial coefficients, reduced (mod p) (if the characteristic $p > 0$), but that is a wasteful way of computing them. We want to determine them in terms of entities modulo p.

It is still true (and obvious) that

$$C(d;k) = C(d;d-k)$$

Looking at the way terms group together on multiplication by $a+b$ we still get

$$C(d+1;k) = C(d;k) + C(d;k-1) \qquad (**)$$

which is Pascal's Triangle. In characteristic 2 the first few lines are

$$\begin{pmatrix}
 & & & & 1 & & 0 & & 0 & & 0 & & 0 \\
 & & & 1 & & 1 & & 0 & & 0 & & 0 & \\
 & & 1 & & 0 & & 1 & & 0 & & 0 & & 0 \\
 & 1 & & 1 & & 1 & & 1 & & 0 & & 0 & \\
1 & & 0 & & 0 & & 0 & & 1 & & 0 & & 0 \\
 & 1 & & 0 & & 0 & & 1 & & 1 & & 0 & \\
1 & & 1 & & 0 & & 0 & & 1 & & 1 & & 0 \\
 & & & & & & & & & & & &
\end{pmatrix}$$

Here each digit is the sum of its two neighbors in the row above it. The reason for adding zeros to the right of the triangle will presently be revealed.

Introducing the sequences $\mathbf{s}_k, \mathbf{s}_k(d) = C(d;k), k = 0, 1, \ldots$ (the k-th diagonal of the triangle above) starting with $k = 0$, we may rewrite equation $(**)$ as

$$\mathbf{s}_k(d+1) - \mathbf{s}_k(d) = \mathbf{s}_{k-1}(d)$$

i.e.,

$$(E - I)\mathbf{s}_k = \mathbf{s}_{k-1} \quad k > 0$$

and

$$(E - I)^n \mathbf{s}_{d-1} = \mathbf{0}$$

$$(E - I)^{n-1} \mathbf{s}_{d-1} = \mathbf{s}_0 \neq \mathbf{0}$$

(since $\mathbf{s}_0$ is all ones).

We will use a lemma from Linear Algebra

17.VI.5 Lemma. *Let $F : V \to V$ be a linear mapping on the k-vector space V. Let $\mathbf{v} \in V$ be a vector such that $F^n(\mathbf{v}) = \mathbf{0}, F^{n-1}(\mathbf{v}) \neq \mathbf{0}$. Then the vectors*

$$\mathbf{v}, F(\mathbf{v}), \ldots F^{n-1}(\mathbf{v})$$

are linearly independent.

Proof: Suppose

$$\lambda_0 \mathbf{v} + \lambda_1 F(\mathbf{v}) + \ldots + \lambda_{n-1} F^{n-1}(\mathbf{v}) = \mathbf{0}$$

Let F^{n-1} act on both members. We get

$$\lambda_0 F^{n-1}(\mathbf{v}) = \mathbf{0}$$

hence $\lambda_0 = 0$, since $F^{n-1}(\mathbf{v}) \neq \mathbf{0}$.

We are left with a shorter relation involving $\mathbf{u} = F(\mathbf{v})$, $F^{n-1}(\mathbf{u}) = \mathbf{0}, F^{n-2}(\mathbf{u}) \neq \mathbf{0}$. So the rest is induction on n, the basis step $n = 1$ being obvious. ∎

Applying the Lemma to $E - I$ and $V =$ solution space of $(E - I)^n \mathbf{s} = \mathbf{0}$ we see that the sequences $\mathbf{s}_j$ above are indeed a basis for V.

Remark: Actually, it easy to check directly that these sequences are independent. Simply note that $\mathbf{s}_0$ begins with a 1, $\mathbf{s}_1$ with 0,1,...; ... $\mathbf{s}_d$ with $d - 1$ zeros and a 1, and so on.

In characteristic 0 each of the $\mathbf{s}_k(d)$ is a polynomial in d of degree k (exactly):

$$\binom{d}{k} = \frac{d(d-1)(d-2)\cdot \ldots \cdot (d-k+1)}{k!}$$

so they constitute a basis for the space of polynomials in d, of degree $\leq k$. So in this case the solution space consists of all such polynomials in d. We thereby justify the claim in the second paragraph of this subsection.

It remains to determine the $C(d;k)$ in positive characteristic p.

Let $d = \sum_i d_i p^i$ be the unique p-ary representation of the number d. Here each $0 \leq d_i \leq p - 1$. We may write

$$(a+b)^d = \prod_i (a+b)^{d_i p^i} = \prod_i (a^{p^i} + b^{p^i})^{d_i}$$

by Freshman's Dream. Here each factor may be computed by the regular Binomial Theorem (using the $\binom{d_i}{k_i}$ since each $d_i < p$.)

Now let

$$k = \sum_i k_i p^i$$

The only way to form a $a^k b^{d-k}$-term is to take the $a^{p^i k_i} b^{p^i (d_i - k_i)}$-term of the i:th factor and multiply these contributions. If, for one i at least, $k_i > d_i$, there will be no contribution at all, otherwise exactly one. This is because the p-ary representation of any number is unique.

You had better check this for yourself in a simple example.

Setting

$$\binom{d}{k} = 0 \text{ if } \ k > d$$

we get

$$C(d;k) = \prod_i \binom{d_i}{k_i}$$

In characteristic 2 this amounts to 1 if $d_i = 1 \Rightarrow n_i = 1$; 0 otherwise. For instance,

$$C(5;2) = C(2^2 + 2^0; 2^1) = 0 \quad C(6;2) = C(2^2 + \underline{2^1}; \underline{2^1}) = 1$$

$$C(15;k) = C(2^3 + 2^2 + 2^1 + 2^0; k_3 2^3 + k_2 2^2 + k_1 2 + k_0 2^0) = 1$$

for $k \leq 15$, since every non-zero digit of k is a digit of 15.

The first two examples are reflected in the full expansions below:

$$(a+b)^5 = (a+b)^4(a+b) = (a^4 + b^4)(a+b) = a^5 + a^4 b + ab^4 + b^5$$

(no a^2b^3–term),

$$(a+b)^6 = (a+b)^4(a+b)^2 = (a^4 + b^4)(a^2 + b^2) = a^6 + a^4 b^2 + a^2 b^4 + b^6$$

with non-zero a^2b^4–term. You had better check the last example to get the idea.

Remark: As an easy corollary (in the case $p = 2$) we get that that a line in Pascal's Triangle is all ones if and only if d is a power of 2 minus 1. In other words all the regular binomial coefficients are odd if, and only if, d is of the form $2^m - 1$. This was the only problem that stumped me in my first exam at the University. I had never heard of Freshman's Dream back then and I wasn't fresh enough to dream up such a frivolous rule!

17.VI.6 The equation $(E - \lambda I)^n \mathbf{s} = \mathbf{0}$

The solution K- space for $(E - \lambda I)^n \mathbf{s} = \mathbf{0}$, where $\lambda \neq 0$ is an element of some extension field K of k, is easily determined from the previous case.

Simply set

$$s_d = \lambda^d t_d$$

with the t_d yet to be determined.

The $d-$th term of $(E - \lambda I)\mathbf{s}$ is easily found to be

$$\lambda^{d+1} t_{d+1} - \lambda^{d+1} t_d = \lambda^{d+1}(t_{d+1} - t_d)$$

Repeating we obtain the d-th term of $(E - \lambda)^n \mathbf{s}$. It is

$$\lambda^{d+n}((E - I)^n \mathbf{t})(d) = 0$$

So $\mathbf{t}$ satisfies the equation of the preceding paragraph, and we have the following

17.VI.7 Theorem. *The solution space of*

$$(E - \lambda I)^n \mathbf{s} = \mathbf{0}$$

has the following basis:

$$\mathbf{t}_0 = (\lambda^d)_{d=0}^{\infty}, \mathbf{t}_1 = (\lambda^d C(d;1)), \mathbf{t}_2 = (\lambda^d C(d;2)), \ldots, \mathbf{t}_{n-1} = (\lambda^d C(d;n-1))$$

In characteristic 0, or if $n < p$, the solutions are

$$t_d = \lambda^d p(d)$$

where p is a polynomial of degree $\leq n-1$

17.VI.8 The general case, $k = \mathbf{C}$ or $\mathbf{R}$.

We return to the general equation

$$p(E)\mathbf{s} = \mathbf{0}$$

and assume for the time being that $k = \mathbf{C}$.

Now let $\lambda_i, i = 1, 2, \ldots r$, with multiplicities e_i, be the roots of $p(\lambda)$ as an ordinary complex polynomial,

$$p(\lambda) = \prod_i (\lambda - \lambda_i)^{e_i} = \prod_i p_i(\lambda)$$

The polynomials

$$q_i(\lambda) = p(\lambda)/p_i(\lambda)$$

have no factor in common. The factor $(\lambda - \lambda_i)^{e_i}$ enters all the q_j except q_i, which does not have the root λ_i at all. Therefore the polynomials q_i generate the unit ideal in $k[X]$, i.e., there are polynomials $A_i(\lambda)$ satisfying

$$1 = \sum_i A_i(\lambda) q_i(\lambda) = \sum_i A_i(\lambda) p(\lambda)/p_i(\lambda)$$

Substituting the operator E for λ we obtain:

$$I = \sum_i A_i(E) q_i(E)$$

Letting both members act on the solution sequence $\mathbf{s}$ we get

$$\mathbf{s} = \sum_i A_i(E) q_i(E)\mathbf{s} = \sum_i \mathbf{s}_i$$

i.e., every solution $\mathbf{s}$ has an (acutally unique) expression as a sum

$$\mathbf{s} = \mathbf{s_1} + \mathbf{s_2} + \ldots + \mathbf{s_r}$$

with each $\mathbf{s}_i$ satisfying

$$(E - \lambda_i I)^{e_i}\mathbf{s_i} = p_i(E)A_i(E)q_i(E)\mathbf{s_i} = A_i(E)p(E)\mathbf{s_i} = 0$$

So we get the general solution:

$$\mathbf{s}(d) = \sum_i p_i(d)\lambda_i^d$$

where each p_i is a polynomial of degree $\leq e_i - 1$.

Uniqueness now follows from the fact that the solutions $d^k\lambda_i^k, k \leq e_i - 1$, n in number, span the solution space, and so are linearly independent.

This concludes the complex case.

17.VI.9 Real case, conjugacy constraints.

Now let us turn to the case where p has real coefficients. Letting λ_i still denote the complex roots we ask for the *real* solutions to the equation.

Now real numbers are characterized as those complex numbers λ satisfying $\lambda' = \lambda$, the $'$ denoting conjugation. We also know that the non-real roots of p appear in conjugate pairs and that the multiplicities of two conjugate roots are the same.

So if λ and λ' are a non-real conjugate pair of roots one real solution $\mathbf{s}$ might look like this:

$$\mathbf{s}(d) = p(d)\lambda^d + q(d)(\lambda')^d + \ldots$$

Conjugating this, and noting that that the left member equals its conjugate we also have

$$\mathbf{s}(d) = p(d)'(\lambda')^d + q(d)'(\lambda)^d + \ldots$$

By the uniqueness of the representation we see, on comparing the two expressions, that $p(d)' = q(d)$, i.e. we have a *conjugacy constraint*; coefficients appearing in front of conjugate powers must be conjugate to one another.

17.VI.10 Example: An easy example is given by the equation

$$(E^2 + I)^2\mathbf{s} = \mathbf{0}$$

i.e., written out in full:

$$(E^4 + 2E^2 + I)\mathbf{s} = \mathbf{0}$$

The roots are $\lambda = \pm i$, both of multiplicity 2. So the complex solutions are

$$\mathbf{s}(d) = (Ad + B)i^d + (Cd + E)(-i)^d =$$

$$(Ad+B)(\cos(d\pi/2)+i\sin(d\pi/2))+(Cd+E)(\cos(d\pi/2)-i\sin(d\pi/2))$$

For real solutions the conjugacy constraint gives us $A=C'$; $B=E'$ whence, after some regrouping

$$\mathbf{s}(d)=(Pd+Q)\cos(d\pi/2)+(Rd+S)\sin(d\pi/2)$$

with $P,Q,R,S,\in \mathbf{R}$. The cosines and sines constitute periodic sequences, of period 4, starting with $1,0,-1,0$ and $0,1,0,-1$ respectively. ∎

17.VI.11 Finite Fields

Now let k be a finite field and let K denote a splitting field for the generating polynomial $p(\lambda)$. Let again $\lambda_i \in K, i=1,2,\ldots,r$ denote the roots of p in K, of multiplicities e_i. The basic result is the same as above, except that polynomials must be replaced by binomial coefficients.:

17.VI.12 Theorem. *The general solution of the equation $p(E)\mathbf{s}=\mathbf{0}$ (over K) is given by*

$$\mathbf{s}(d)=\sum_i C_i(d)\lambda_i^d$$

where the $C_i(d)$ are $K-$linear combinations of binomial coefficients

$$C(d;j), j=0,\ldots,e_i-1.$$

In case all roots are simple, i.e., $e_i=1,\forall i$, the C_i are constants, i.e., elements of K, independent of d.

17.VI.13 Conjugacy Constraints

For simplicity of notation we henceforth assume that all e_i are $=1$. We again ask for a description of solutions lying in the field k. Again there is a conjugacy condition. If k has $q=p^m$ elements, then , for $\alpha \in K$:

$$\alpha \in k \iff \alpha^q=\alpha,$$

a standard result in the theory of finite fields.

We also know that if $r(\lambda)$ is an irreducible factor of $p(\lambda)$, of degree t, then for $\lambda_0 \in K$,

$$n(\lambda_0)=0 \iff n(\lambda_0^q)=0$$

and the roots of $r(\lambda)$ are given by

$$\lambda_0(=\lambda_0^{q^t}),\lambda_0^q,\ldots,\lambda_0^{q^{t-1}}$$

Now suppose

$$\mathbf{s}(d) = \sum_i C_i \lambda_i^d$$

is a solution in k. Applying the Frobenius $\alpha \to \alpha^q$ to both members, and using Freshman's Dream, we get

$$\mathbf{s}(d)^q = \mathbf{s}(d) = \sum_i C_i^q \lambda_i^{dq}$$

Again, as in the real case, uniqueness of the representation shows that if $r(\lambda)$ is an irreducible factor of degree t, and if λ_0 is a root of $r(\lambda)$, then the contribution from that root, and its conjugates, must be of the form

$$C\lambda_0^d + C^q \lambda_0^{dq} + \ldots + C^{q^{t-1}} \lambda_0^{dq^{t-1}}$$

17.VI.14 Example:

Let $k = \mathbf{Z}_3$, the field with three elements. Let further $p(\lambda) = \lambda^3 + 2\lambda + 2$. This polynomial is irreducible over k. As it is of degree 3 it is enough to check that it has no roots in k.

This polynomial splits in the extension field

$$K = k[\lambda]/(p(\lambda))$$

One root is the class of λ which we denote by α. By Frobenius the other roots are $\alpha^3 = \alpha + 1$ and $\alpha^9 = (\alpha+1)^3 = \alpha^3 + 1 = \alpha + 2$. Of course, $\alpha^{27} = \alpha + 3 = \alpha$.

We study the recursion

$$p(E)\mathbf{s} = 0; \qquad s_0 = 2, s_1 = 0, s_2 = 0$$

The K–solutions of this recursion are of the form

$$s_d = a_0\alpha^d + a_1(\alpha+1)^d + a_2(\alpha+2)^d, \qquad a_i \in K, i = 0, 1, 2$$

The k–solutions are characterized by the conjugacy constraints

$$a_1 = a_0^3, \quad a_2 = a_1^3$$

Of course, then, $a_0 = a_2^3 = a_0^{27}$, by standard field theory.

Setting $d = 0, 1, 2$, and using $s_0 = 2, s_1 = 0, s_2 = 0$, we can determine the a_i from a linear system of three equations in three unknowns. By the usual elimination procedure we get ...

$$a_2 = \alpha^2 + \alpha = (\alpha+1)\alpha$$

You may want to check this computation.

There is no need to back-substitute. Using the Frobenius we get

$$a_0 = a_2^3 = (\alpha + 1 + 1)(\alpha + 1) = (\alpha + 2)(\alpha + 1)$$

$$a_1 = a_0^3 = (\alpha + 1 + 2)(\alpha + 1 + 1) = \alpha \cdot (\alpha + 2)$$

since the Frobenius, in this special example, replaces α by $\alpha + 1$.

So

$$s_d = (\alpha + 2)(\alpha + 1)\alpha^d + \alpha \cdot (\alpha + 2)(\alpha + 1)^d + (\alpha + 1)\alpha \cdot (\alpha + 2)^d,$$

that is,

$$s_d = \alpha(\alpha + 1)(\alpha + 2)[(\alpha^{d-1} + (\alpha + 1)^{d-1} + (\alpha + 2)^{d-1}]$$

The first factor is readily seen to be 1.

The reader is invited to check the initial data.

17.VI.15 Maximal Period Sequences

First of all we need a general Lemma.

17.VI.16 Lemma. *The solution space V of $p(E)\mathbf{s} = \mathbf{0}$ is* cyclic, *i.e., there is one solution $\mathbf{s}_0$ such that*

$$\mathbf{s}_0, E\mathbf{s}_0, \ldots E^{n-1}\mathbf{s}_0$$

are a $k-$ basis of V.

Proof: Let $\mathbf{s}_0$ be the *impulse response solution* characterized by the initial data

$$\begin{pmatrix} s_0 \\ s_1 \\ \vdots \\ s_{n-2} \\ s_{n-1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

Then it is easy to see that $\mathbf{s}_0, E\mathbf{s}_0, E^2\mathbf{s}_0, \ldots$ are given by the initial state vectors

$$\begin{pmatrix} \vdots \\ \vdots \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} \vdots \\ \vdots \\ 0 \\ 1 \\ ? \end{pmatrix}, \begin{pmatrix} \vdots \\ \vdots \\ 1 \\ ? \\ ?? \end{pmatrix}, \ldots$$

They are obviously linearly independent, and n in number. ∎

We henceforth assume that $p(\lambda)$ is irreducible over k. So its roots are $\lambda_0, \lambda_1 = \lambda_0^q, \ldots, \lambda_{n-1} = \lambda_0^{q^{n-1}} \in K$.

Since the non-zero elements of K form a group of order $q^n - 1$ we must have

$$\lambda_i^{d+q^n-1} = \lambda_i^d, \quad i = 0, 1, \ldots, n-1$$

for all d, so every solution of period at most $= q^n - 1$. If one λ_i (hence all) generates a group of order less than $q^n - 1$ the period must be strictly shorter, of course. In this case we say that the polynomial p is *imprimitive* over k.

In the opposite case of primitive p, i.e., when one of its roots (hence all) generates the group, there is at least one solution of maximal possible period. This is the content of our next Theorem.

17.VI.17 Theorem. *Let $p(E)\mathbf{s} = \mathbf{0}$, with p irreducible and primitive over k, be given. Let $\mathbf{s}_0$ denote the impulse response solution. Then $\mathbf{s}_0$ is periodic of period $q^n - 1$.*

Proof: Suppose not. Suppose the true period is $r < q^n - 1$. Then $\mathbf{s}_1 := E\mathbf{s}_0, \mathbf{s}_2 := E^2\mathbf{s}_0 \ldots, \mathbf{s}_{n-1} := E^{n-1}\mathbf{s}_0$ also will have period r, hence every $K-$linear combination of them will have period r. But the $\mathbf{s}_i$ span the solution K-space and the sequence

$$s_d = \lambda_0^d$$

is one solution, of period $q^n - 1$, by assumption. This contradiction concludes the proof of the Theorem. ∎

.

Maximal period sequences have some random-like properties making them ueseful in Computer Science. One can prove that all strings of reasonable length (as compared to the period) appear with approximately the same frequency.

For this, and other deeper properties of linear recurring sequences, I refer to Lidl-Niederreiter's book on Finite Fields (Textbook version, Cambridge).

17.VI.18 Example:

In the example above the polynomial $\lambda^3 + 2\lambda + 2$ is not primitive. The true order of α is not $27 - 1 = 26$ but 13. You may want to check this, starting with $\alpha^9 = \alpha + 2$, $\alpha^3 = \alpha + 1$.

The period of any solution is then at most 13. However, the least period of any solution must divide 13 (this is the ususal division argument), and 13 is a prime. Therefore all solutions, except the zero solution, are of period 13.

17.VI.19 Inhomogenous Equations, Non-resonant Case

We next turn to the equation

$$p(E)\mathbf{s} = \mathbf{t}$$

where $\mathbf{t}$ is some given sequence. If $\mathbf{s}_P$ (P as in "particular") is one solution, and $\mathbf{s}$ is an arbitrary solution, then for $\mathbf{s}_H := \mathbf{s} - \mathbf{s}_H$ we obviously have

$$p(E)\mathbf{s}_H = p(E)\mathbf{s} - p(E)\mathbf{s}_P = \mathbf{t} - \mathbf{t} = \mathbf{0}$$

so we may write

$$\mathbf{s} = \mathbf{s}_P + \mathbf{s}_H$$

i.e., the general solution is the sum of one particular solution and the general solution of the *homogeneous* equation

$$p(E)\mathbf{s} = \mathbf{0}$$

So we can concentrate our efforts on finding *one* solution. We will do this in case $\mathbf{t}$ is itself a linear reccuring sequence, of generating polynomial $q(\lambda)$.

If $g(\lambda) = \prod_i (\lambda - \lambda_i)^{e_i}$ then, as we have seen, $\mathbf{t}$ is a k-linear combination of solutions to the simpler equations

$$(E - \lambda_i)^{e_i}\mathbf{t} = \mathbf{0}$$

So in practice one may solve the equation for right members satisfying this equation and then combine the results linearly.

We first deal with the case

$$(q, p) = 1$$

In that case, by Bézout, there are polynomials Q, P satisfying

$$Qq + Pp = 1; \;\; Q(E)q(E) + P(E)p(E) = I$$

From this we get

$$\mathbf{s} = (Q(E)q(E) + P(E)p(E))\mathbf{s} = Q(E)q(E)\mathbf{s} + P(E)\mathbf{t}$$

Assuming $q(E)\mathbf{s} = \mathbf{0}$ we see that the unique possibility is

$$\mathbf{s} = P(E)\mathbf{t}$$

It is easily checked:

$$p(E)P(E)\mathbf{t} = (I - Q(E)q(E))\mathbf{t} = \mathbf{t} - \mathbf{0}$$

We have proved

17.VI.20 Theorem. *Let $p, q \in k[\lambda]$, $(q,p) = 1$, be given polynomials. Suppose* $\mathbf{t}$ *is a reccuring sequence satisfying* $q(E)\mathbf{t} = \mathbf{0}$. *Then the equation*

$$p(E)\mathbf{s} = \mathbf{t}$$

has a unique solution s_H *satisfying* $q(E)\mathbf{s} = \mathbf{0}$ *, i.e., one of the same form as* $\mathbf{t}$.

In particular, if $q(\lambda) = (\lambda - \lambda_0)^e$ *and* λ_0 *is not a root of* $p(\lambda)$ *the displayed equation will have exactly one solution satisfying*

$$(E - \lambda_0 I)^e \mathbf{s} = \mathbf{0}$$

Note that we actually proved:

17.VI.21 Theorem. *Let* $N(q(E))$ *denote the solution space of* $q(E)\mathbf{s} = \mathbf{0}$ *and suppose* p *is a polynomial with* $(q,p) = 1$. *Then the mapping* $p(E) : N(q(E)) \to N(q(E))$ *is bijective.*

Proof: Its inverse is $P(E)$.

∎

17.VI.22 Resonance

We next turn to the case $(p,q) = d \neq 1$. Let $p = p_0 d', q = q_0 d, (p_0, q) = 1$, $d | d'$. Here d' is the product of all prime factors in p that appear in d. Note that p/d could very well have non-trivial factors in common with q.

17.VI.23 Theorem. *The equation*

$$d(E)\mathbf{s} = \mathbf{u}; \;\; q_0(E)d(E)\mathbf{u} = \mathbf{0}; (q_0, d) = 1$$

has a solution $\mathbf{s}$ *satisfying*

$$q(E)d'(E)\mathbf{s} = q_0(E)d(E)d'(E)\mathbf{s} = \mathbf{0}$$

Proof: It is easy to see that $\mathbf{r} := d'(E)\mathbf{s}$ satisfies the equation $q(E)\mathbf{r} = \mathbf{0}$ if, and only if, $\mathbf{s}$ satisfies $q(E)d'(E)\mathbf{s} = \mathbf{0}$. In other words, the mapping

$$d'(E) : N(qd'(E)) \to N(q(E))$$

is *surjective*. From the case above, since $(p_0, q) = 1$, we also have that

$$p_0(E) : N(q(E) \to N(q(E))$$

is bijective. Composing we see that

$$p(E) = p_0(E)d'(E) : N(qd'(E)) \to N(q(E))$$

is surjective, which is exactly what the Theorem states.

∎

Let us give a more explicit statement in case $q(\lambda) = (\lambda - \lambda_0)^e$. I give the Theorem for $k = \mathbf{C}$ leaving to the reader the task of stating and proving the corresponding result in characteristic $p > 0$.

17.VI.24 Theorem. *Suppose λ_0 is a root of $p(\lambda)$, of multiplicity k. Further, let $m(d)$ be a polynomial in d, of degree $e - 1$. Then the equation*

$$p(E)\mathbf{s}(d) = m(d)\lambda_0^d$$

has a particular solution $\mathbf{s}$ of the form

$$\mathbf{s}(d) = d^k m'(d)\lambda_0^d$$

where $m'(d)$ is a polynomial of degree $\leq e - 1$

Proof: . Here $q(\lambda) = (\lambda - \lambda_0)^e$, and $d'(\lambda) = (\lambda - \lambda_0)^k$. So there is solution $\mathbf{s}$ satisfying

$$q(E)d'(E)\mathbf{s} = (E - \lambda_0 I)^{e+k}\mathbf{s} = \mathbf{0}$$

i.e.,

$$\mathbf{s}(d) = M(d)\lambda_0^d$$

with M of degree $\leq e + k - 1$. Here we may delete all terms of degree $< k$ since they satisfy $p(E)\mathbf{s} = \mathbf{0}$, hence contribute nothing. This proves the theorem.

Remark. We couldn't do any better than that. The reader may check for himself that the operator $F = E - \lambda_0$ acting on " polynomial $\cdot \lambda^d$ " lowers the degree of the polynomial factor exactly one step (if it is > 0), hence d' steps when applied d' times.

∎

17.VI.25 Example: An easy example is the equation

$$(E^2 - \mu^2 I)\mathbf{s}(d) = \lambda_0^d$$

In case $\lambda_0 \neq \pm\mu$, the Theorem states that there is one solution of the form $C\lambda_0^d$. Plugging this expression into the equation we immediately get $C = 1/(\lambda_0^2 - \mu^2)$.

Let us work out the case $\lambda_0 = \mu$ as well. Here the Theorem tells us to look for a solution of the form $Bd\mu^d$ (no constant is needed). Trying this we obtain

$$\mathbf{s}(d) = \frac{1}{2} d\mu^d$$

∎