Unique Factorization in Z[X]

We have already used Eisenstein's criterion, without proof, to produce irreducible polynomials over Z. We will fill the gap in our logic by proving that result now. The essential point is Gauss' theory of the content of a polynomial, and of primitive polynomials. These concepts allow us to compare factorization in Z[X] with that in Q[X], and to deduce that a monic polynomial which is irreducible in Z[X] remains irreducible in Q[X]. The contrapositive statement that an integral polynomial which is reducible in Q[X] is also reducible in Z[X] allows us to obtain (unique) factorization of polynomials in Z[X], and more generally also in $k[X_1,....,X_n]$.

Theorem (Gauss): If R is a ufd, then R[X] is a ufd also.

This is the most general statement we shall prove, in the next section, but we shall proceed to the proof in stages, first proving that Z[X] is a ufd. This proof contains all the essential ideas. It is simple in principle, but the details are tedious to do completely. We will attempt to make the main ideas clear, and we will also try to present essentially all the details. First of all, think back over your own experience, factoring things like $x^2+5x+6 = (x+2)(x+3)$. Notice that when the coefficients of the original polynomial are integers, then the coefficients of the factors are also integers. To be sure, you *can* factor $x^2-2 = (x-2^{1/2})(x+2^{1/2})$ with *irrational* numbers. But if an integral polynomial factors with rational numbers, then it already factors with integers. This is one of the first results we shall prove using Gauss' idea of "content".

Very briefly then, to factor an integral polynomial f over Z[X], for example $f = 6x^2-30x+36$, just remove the gcd of the coefficients (this is the "content"), here $f = 6(x^2+5x+6)$, then factor separately the content and the remaining polynomial $f = (2)(3)(x+2)(x+3)$, and these are the irreducible factors of f over Z[x]. Note there are four irreducible factors here since 2, 3 are not units, but primes in Z[x].

The most important concept is the following one:

Definition: The "content" of a polynomial f in Z[X] is the gcd of the coefficients of f. If f=0, the content is 0. We denote content(f) = c_f.

Thus c_f is a well defined non negative integer which is zero iff f =0.

Definition: A polynomial f in Z[X] is "primitive" iff c_f =1, iff the coefficients of f have no common prime integral factor.

Lemma: Let f,g,h, be non zero polynomials in Z[X].

(i) c in Z^+ is the content of f iff f = cg where g is primitive.

(ii) If c,d are in Z^+, g,h in Z[X] are primitive, and cg = dh, then c=d and g=h.

(iii) Every non zero f in Z[X] has a unique associated primitive polynomial f_0 such that $f = c_f(f_0)$. [Or if f=0, take f_0=1.]

(iv) If $f \neq 0$ in #[X], and $f = cg$ where c is in Z^+ and g is primitive, then $c = c_f$ and $g = f_0$.

proof: Exercise. **QED.**

The main property of the content is that it is multiplicative. We prove this in the following steps.

Lemma: If g,h are primitive in Z[X] and f = gh, then f is primitive.

proof: If p is any prime integer, it suffices to prove that some coefficient of f is not divisible by p. Since this is true for both g and h, among the coefficients of g which are not divisible by p there is a highest one say a_r, and similarly a highest one among the coefficient of h not divisible by p, say b_s. Then the coefficient c of X^{r+s} in f is a sum of terms, of which one is $a_r b_s$ and the others are of form $a_k b_l$ where k+l = r+s. Hence except when k = r, l = s, we must have k > r or l > s. In these cases, either p divides a_k or b_l and thus their product. Hence p divides every term but one in the coefficient c of X^{r+s}, and hence p does not divide c. **QED.**

Lemma: If f,g,h are in #[X] and f = gh, then $c_f = (c_g)(c_h)$.

proof: In the notation introduced above we have $g = c_g(g_0)$, and $h = c_h(h_0)$, whence $f = gh = c_g c_h(g_0 h_0)$, where $g_0 h_0$ is primitive. Thus by the properties given above for content, $c_f = c_g c_h$. **QED.**

Lemma: If a, b, c, d are in Z^+, and g, h are primitive (in Z[X]), and if (a/b)g = (c/d)h, then a/b = c/d, and g = h.

proof: Multiplying by bd, we conclude that adg = bch, whence the properties above of content imply ad = bc, hence a/b = c/d. Dividing through by a/b = c/d, then g =h. **QED.**

Remark: If f is non zero in Q[X], there exist a,b in Z^+ and a primitive g in Z[X] such that f = (a/b)g, since we may take b as a positive common multiple of the denominators of the coefficients of f, and a = content(bf), where bf is in #[X]. By the previous lemma, a/b and g are unique.

Definition: For any non zero f in Q[X] the content is the unique positive element c_f of Q such that $f = c_f(f_0)$ where f_0 is primitive in Z[X]. The unique such f_0 is called the "primitive form" of f.

Multiplicativity holds also for rational contents.

Lemma: For any g,h in $[X], if f=gh then $c_f = c_g c_h$.

proof: The proof is the same as for integral contents. **QED.**

It follows that the "primitive form" is also multiplicative:

Lemma: For any g,h in Q[X], $(gh)_0 = (g_0)(h_0)$.

proof: The lemmas imply that $(gh)_0$ is the unique primitive polynomial P such that gh is a positive rational multiple of P. But $gh = c_g(g_0)c_h(h_0) = c_g c_h(g_0 h_0)$, where c_g, c_h are positive and rational and $(g_0 h_0)$ is primitive. **QED.**

Now we can go through the proof that Z[X] is a ufd, by replacing every polynomial by its primitive form whenever possible. The point is that the primitive polynomials have the same divisibility properties in Q[X] as in Z[X], allowing us to bring unique factorization down from Q[X] to Z[X]. (We emphasize that primitive polynomials are always elements of Z[X], and the only constant primitive polynomials are 1, -1.) More precisely:

Lemma: If f is primitive, then f is reducible in Z[X] iff f is reducible in Q[X]. In fact if f = gh, with g,h non units in Q[X], then also $f = (g_0)(h_0)$, where g_0, h_0 are the primitive forms of g,h.

proof: If f is reducible in Z[X], f = gh, then both g,h have degree _1 since f is primitive, hence g,h are non units in Q[X] and f is reducible in Q[X]. If f = gh, with g,h non units in Q[X], by multiplicativity of primitive forms we have $f = f_0 = (g_0)(h_0)$, so f is reducible in Z[X]. **QED.**

Remark: The previous lemma fails in one direction for non primitive polynomials; eg. 3X+3 is reducible in Z[X] but not in Q[X]. It still holds in the other direction, as the next lemma shows.

Lemma: If f in Z[X] is reducible in Q[X], f is also reducible in Z[X] and can be factored into factors of degree ≥ 1 in Z[X].

proof: If f = gh, with f in Z[X] and g,h of degree ≥ 1 in Q[X], then

$c_f(f_0) = f = gh = c_g c_h(g_0 h_0)$. Thus $c_g c_h = c_f$ is an integer, and $f =$

$c_f(g_0 h_0)$ is a factorization of f over Z[X] with degrees $g_0, h_0 \geq 1$. **QED.**

Remark: Note that since Z is a domain, deg(fg) = deg(f) + deg(g), so constants in Z can have only constant factors, hence prime integers in Z are also irreducible in Z[X].

Now we can prove existence of factorization into irreducibles in Z[X].

Lemma: Every non zero, non unit element of Z[X] can be factored into irreducible elements.

proof: Let f be non zero, non unit in Z[X]. If f is in Z, then the previous remark shows the prime factorization in # gives a factorization into irreducibles in Z[X]. If deg(f)≥1, factor it as $f = c_f(f_0)$. Then $f_0 = \prod g_i$, with g_i irreducible in Q[X]. By the previous lemmas, then $f_0 = \prod (g_i)_0$, where the $(g_i)_0$ are primitive forms of the g_i. Then each $(g_i)_0$ is a non zero rational multiple of g_i, hence still irreducible in Q[X] and also primitive, hence irreducible in Z[X], by our lemma above. Factoring $c_f = \prod p_i$ into primes in Z gives us the factorization of $f = \prod p_i \cdot \prod (g_i)_0$ into irreducibles in Z[X]. **QED.**

Remark: Existence of irreducible factorizations is not really the hard part of the theory in this case, since we already know Z[X] is a noetherian domain, and it can be proved easily that factorization into irreducibles is always possible in any noetherian domain. The uniqueness however is not always true in a noetherian domain. The proof just given of existence of factorizations in Z[X] will also work in R[X] where R is a non noetherian ufd.

We need one more technical property of primitive polynomials.

Lemma: If f in Z[X] is primitive, and g is in Z[X], then f|g in Z[X] iff f|g in Q[X].

proof: If f|g in Z[X] then g = fh, for h in Z[X], hence also h in Q[X], so f|g in Q[X]. And if f|g in Q[X], then g = fh, for h in Q[X]. Then $c_g = c_f c_h = c_h$, so $c_h = c_g$ is an integer. Then $h = c_h (h_0)$ is in Z[X], so f divides g in Z[X]. **QED.**

Remark: Again one direction fails for non primitive polynomials, since 3X+3 divides X+1 in Q[X], but not in Z[X].

The next property is the key to proving uniqueness of factorization.

Lemma: If f,g,h are in Z[X], f is irreducible, and f|gh, then f|g or f|h.

proof: First note that an integer c divides a polynomial F in Z[X] iff c divides all the coefficients of F, iff $c|c_F$. Hence if f is irreducible in Z[X] and an integer, f=p is prime in Z. Then if p divides gh, p divides $c_{gh} = c_g c_h$, so p divides either c_g or c_h by the corresponding lemma in Z. Hence f=p divides either g or h. That settles this case.

If deg(f) ≥ 1, then f irreducible implies f is primitive, hence f is also irreducible in Q[X] and divides gh also in Q[X]. Since the present lemma holds in Q[X], f divides either g or h in Q[X]. Since f is primitive, then f divides either g or h also in Z[X]. **QED.**

Lemma: Factorization into irreducibles is unique in Z[X], up to order of factors and sign.

proof(same proof as in Z): If $\prod g_i = \prod h_j$ where all g_i, h_j are irreducible in Z[X], then g_1 divides the left side, hence also the right, so by the previous lemma g_1 divides some h_j which we may renumber as h_1. Since h_1 is irreducible, and the only units in Z[X] are ±1, then $h_1 = \pm g_1$. Hence we may cancel g_1 from both sides leaving $(g_2)(.....)(g_n) = \pm(h_2)(....)(h_m)$. Continuing with g_2,....., we eventually cancel all terms. I.e. there are the same number of g's and h's and after renumbering the indices, for every i, $g_i = \pm h_i$.

If you want the proof to appear more rigorous, use induction on the number n of factors g_i. If there is only one g_i there can be only one h_j since g_i is irreducible. This proves the result for n=1. Assuming the theorem for n-1 factors g_i, we are done after we cancel g_1 from both sides as above, since then by induction n-1 = m-1, hence n=m and the factors g_2,....,g_n must agree with the factors $\pm h_2$,....,h_n up to order and multiplication by units. **QED.**