Factorization in Polynomial Rings

Throughout these notes, F denotes a field.

1 Long division with remainder

We begin with some basic definitions.

Definition 1.1. Let $f(x), g(x) \in F[x]$. We say that $f(x)$ *divides* $g(x)$, written $f(x)\big|g(x)$, if there exists an $h(x) \in F[x]$ such that $g(x) = f(x)h(x)$, i.e. $g(x)$ is a multiple of $f(x)$. Thus, for example, every $f(x) \in F[x]$ divides the zero polynomial 0, but $g(x)$ is divisible by 0 $\iff$ $g(x) = 0$.

By definition, $f(x)$ is a unit $\iff$ $f(x)\big|1$. Recall also that the group of units $(F[x])^*$ of the ring $F[x]$ is F^*, the group of units in the field F, and hence the group of nonzero elements of F under multiplication. Thus $f(x)$ divides every $g(x) \in F[x]$ $\iff$ $f(x)$ divides 1 $\iff$ $f(x) \in F^*$ is a nonzero constant polynomial. Finally note that, if $c \in F^*$ is a unit, then $f(x)\big|g(x)$ $\iff$ $cf(x)\big|g(x)$ $\iff$ $f(x)\big|cg(x)$.

Proposition 1.2 (Long division with remainder)**.** *Let $f(x) \in F[x]$, $f(x) \neq 0$, and let $g(x) \in F[x]$. Then there exist unique polynomials $q(x), r(x) \in F[x]$, with either $r(x) = 0$ or $\deg r(x) < \deg f(x)$, such that*

$$g(x) = f(x)q(x) + r(x).$$

Proof. First we prove existence. The proposition is clearly true if $g(x) = 0$, since then we can take $q(x) = r(x) = 0$. Otherwise, we argue by induction on $\deg g(x)$. If $\deg g(x) = 0$ and $\deg f(x) = 0$, then $f(x) = c \in F^*$ is a nonzero constant, and then $g(x) = c(c^{-1}g(x)) + 0$, so we can take $q(x) = c^{-1}g(x)$ and $r(x) = 0$. If $\deg g(x) = 0$ and $\deg f(x) > 0$, or more generally if $n = \deg g(x) < \deg f(x) = d$, then we can take $q(x) = 0$ and $r(x) = g(x)$. Now assume that, for a fixed $f(x)$, the existence of $q(x)$ and $r(x)$ has been proved for all polynomials of degree $< n$, and suppose that $g(x)$ is a polynomial of degree n. As above, we can assume that $n \geq d = \deg f(x)$.

Let $f(x) = \sum_{i=0}^{d} a_i x^i$, with $a_d \neq 0$, and let $g(x) = \sum_{i=0}^{n} b_i x^i$. In this case, $g(x) - b_n a_d^{-1} x^{n-d} f(x)$ is a polynomial of degree at most $n-1$ (or 0). By the inductive hypothesis and the case $g(x) = 0$, there exist polynomials $q_1(x), r(x) \in F[x]$ with either $r(x) = 0$ or $\deg r(x) < \deg f(x)$, such that

$$g(x) - b_n a_d^{-1} x^{n-d} f(x) = f(x) q_1(x) + r(x).$$

Then

$$g(x) = f(x)(b_n a_d^{-1} x^{n-d} + q_1(x)) + r(x) = f(x) q(x) + r(x),$$

where we set $q(x) = b_n a_d^{-1} x^{n-d} + q_1(x)$. This completes the inductive step and hence the existence part of the proof.

To see uniqueness, suppose that

$$g(x) = f(x) q_1(x) + r_1(x) = f(x) q_2(x) + r_2(x),$$

where either $r_1(x) = 0$ or $\deg r_1(x) < \deg f(x)$, and similarly for $r_2(x)$. We have

$$(q_1(x) - q_2(x)) f(x) = r_2(x) - r_1(x),$$

hence either $q_1(x) - q_2(x) = 0$ or $q_1(x) - q_2(x) \neq 0$ and then

$$\deg((q_1(x) - q_2(x)) f(x)) = \deg(q_1(x) - q_2(x)) + \deg f(x) \geq \deg f(x).$$

Moreover, in this case $r_2(x) - r_1(x) \neq 0$. But then

$$\deg(r_2(x) - r_1(x)) \leq \max\{\deg r_1(x), \deg r_2(x)\} < \deg f(x),$$

a contradiction. Thus $q_1(x) - q_2(x) = 0$, hence $r_2(x) - r_1(x) = 0$ as well. It follows that $q_1(x) = q_2(x)$ and $r_2(x) = r_1(x)$, proving uniqueness. $\square$

Remark 1.3. The analogue of Proposition 1.2 holds in an arbitrary ring R (commutative, with unity as always) provided that we assume that $f(x)$ is **monic**.

The following is really just a restatement of Proposition 1.2 in more abstract language:

Corollary 1.4. *Let $f(x) \in F[x]$, $f(x) \neq 0$. Then every coset $g(x) + (f(x))$ has a unique representative $r(x)$, where $r(x) = 0$ or $\deg r(x) < \deg f(x)$.*

Proof. By Proposition 1.2, we can write $g(x) = f(x)q(x)+r(x)$ with $r(x) = 0$ or $\deg r(x) < \deg f(x)$. Then $r(x) \in g(x) + (f(x))$ since the difference $g(x)-r(x)$ is a multiple of $f(x)$, hence lies in $(f(x))$. The uniqueness follows as in the proof of uniqueness for Proposition 1.2: if $r_1(x)+(f(x)) = r_2(x)+(f(x))$, with each $r_i(x)$ either 0 of of degree smaller than $\deg f(x)$, then $f(x)\big|r_2(x) - r_1(x)$, and hence $r_2(x) - r_1(x) = 0$, so that $r_1(x) = r_2(x)$. □

Corollary 1.5. *Let $a \in F$. Then every $f(x) \in F[x]$ is of the form $f(x) = (x-a)g(x) + f(a)$. Thus $f(a) = 0 \iff (x-a)\big|f(x)$.*

Proof. Applying long division with remainder to $x-a$ and $f(x)$, we see that $f(x) = (x-a)g(x) + c$, where either $c = 0$ or $\deg c = 0$, hence $c \in F^*$. (This also follows directly, for an arbitrary ring: if $f(x) = \sum_{i=0}^{d} a_i x^i$, write $f(x) = f(x-a+a) = \sum_{i=0}^{d} a_i(x-a+a)^i$. Expanding out each term via the binomial theorem then shows that $f(x) = \sum_{i=0}^{d} b_i(x-a)^i$ for some $b_i \in F$, and then we take $c = b_0$.)

Finally,to determine c, we evaluate $f(x)$ at a: $f(a) = \mathrm{ev}_a(f(x)) = \mathrm{ev}_a((x-a)g(x)+c) = 0 + c = c$. Hence $c = f(a)$. □

Recall that, for a polynomial $f(x) \in F[x]$, a *root* or *zero* of $f(x)$ in F is an $a \in F$ such that $f(a) = \mathrm{ev}_a(f(x)) = 0$.

Corollary 1.6. *Let $f(x) \in F[x]$, $f(x) \neq 0$, and suppose that $\deg f(x) = d$. Then there are at most d roots of $f(x)$ in any field E containing F. In other words, suppose that F is a subfield of a field E. Then*

$$\#\{a \in E : f(a) = 0\} \leq d.$$

Proof. We can clearly assume that $E = F$. Argue by induction on $\deg f$, the case $\deg f = 0$ being obvious. Suppose that the corollary has been proved for all polynomials of degree $d-1$. If $\deg f(x) = d$ and there is no root of $f(x)$ in F, then we are done because $d \geq 0$. Otherwise, let a_1 be a root. Then we can write $f(x) = (x-a_1)g(x)$, where $\deg g(x) = d-1$. Let a_2 be a root of $f(x)$ with $a_2 \neq a_1$. Then

$$0 = f(a_2) = (a_2 - a_1)g(a_2).$$

Since F is a field and $a_2 \neq a_1$, $a_2 - a_1 \neq 0$ and we can cancel it to obtain $g(a_2) = 0$, i.e. a_2 is a root of $g(x)$ (here we must use the fact that F is a field). By induction, $g(x)$ has at most $d-1$ roots in F (where we allow for the possibility that a_1 is also a root of $g(x)$). Then

$$\{a \in F : f(a) = 0\} = \{a_1\} \cup \{a \in F : g(a) = 0\}.$$

Since $\#\{a \in F : g(a) = 0\} \leq d - 1$, it follows that $\#\{a \in F : f(a) = 0\} \leq d$. $\square$

Corollary 1.6 has the following surprising consequence concerning the structure of finite fields, or more generally finite subgroups of the group F^* under multiplication:

Theorem 1.7 (Existence of a primitive root)**.** *Let F be a field and let G be a finite subgroup of the multiplicative group $(F^*, \cdot)$. Then G is cyclic. In particular, if F is a finite field, then the group $(F^*, \cdot)$ is cyclic.*

Proof. Let $n = \#(G)$ be the order of G. First we claim that, for each $d \big| n$, the set $\{a \in G : a^d = 1\}$ has at most d elements. In fact, clearly $\{a \in G : a^d = 1\} \subseteq \{a \in F : a^d = 1\}$. But the set $\{a \in F : a^d = 1\}$ is the set of roots of the polynomial $x^d - 1$ in F. Since the degree of $x^d - 1$ is d, by Corollary 1.6, $\#\{a \in F : a^d = 1\} \leq d$. Hence $\#\{a \in G : a^d = 1\} \leq d$ as well. The theorem now follows from the following purely group-theoretic result, whose proof we include for completeness. $\square$

Proposition 1.8. *Let G be a finite group of order n, written multiplicatively. Suppose that, for each $d \big| n$, the set $\{g \in G : g^d = 1\}$ has at most n elements. Then G is cyclic.*

Proof. Let φ be the Euler φ-function. The key point of the proof is the identity (proved last semester, or in courses in elementary number theory)

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

Now, given a finite group G as in the statement of the proposition, define a new function $\psi \colon \mathbb{N} \to \mathbb{N}$ via: $\psi(d)$ is the number of elements of G of order exactly d. By Lagrange's theorem, if $\psi(d) \neq 0$, then $d \big| n$. Since every element of G has some well-defined finite order, adding up all of values of $\psi(d)$ is the same as counting all of the elements of G. Hence

$$\#(G) = n = \sum_{d \in \mathbb{N}} \psi(d) = \sum_{d|n} \psi(d).$$

Next we claim that, for all $d \big| n$, $\psi(d) \leq \varphi(d)$; more precisely,

$$\psi(d) = \begin{cases} 0, & \text{if there is no element of } G \text{ of order } d; \\ \varphi(d), & \text{if there is an element of } G \text{ of order } d. \end{cases}$$

Clearly, if there is no element of G of order d, then $\psi(d) = 0$. Conversely, suppose that there is an element a of G of order d. Then $\#(\langle a\rangle) = d$, and every element $g \in \langle a\rangle$ has order dividing d, hence $g^d = 1$ for all $g \in \langle a\rangle$. But since there at most d elements g in G such that $g^d = 1$, the set of all such elements must be exactly $\langle a\rangle$. In particular, an element g of order **exactly** d must both lie in $\langle a\rangle$ and be a generator of $\langle a\rangle$. Since the number of generators of $\langle a\rangle$ is the same as the number of generators of any cyclic group of order d, namely $\varphi(d)$, the number of elements of G of order d is then $\varphi(d)$. Thus, if there is an element of G of order d, then by definition $\psi(d) = \varphi(d)$.

Now compare the two expressions

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

Since, for each value of $d\big|n$, $\psi(d) \leq \varphi(d)$, and the sums are the same, we must have $\psi(d) = \varphi(d)$ for all $d\big|n$. In particular, taking $d = n$, we see that $\psi(n) = \varphi(n) \neq 0$. It follows that there exists an element of G of order $n = \#(G)$, and hence G is cyclic. □

Example 1.9. (1) In case p is a prime and $F = \mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$, then a generator for $(\mathbb{Z}/p\mathbb{Z})^*$ is called a *primitive root*.

(2) For $F = \mathbb{C}$, the finite multiplicative subgroups of $\mathbb{C}^*$ are the groups μ_n of n^{th} roots of unity. A generator of μ_n, in other words a complex number whose order in the group $(\mathbb{C}^*, \cdot)$ is exactly n, is called a *primitive n^{th} root of unity*. The standard such generator is $e^{2\pi i/n}$.

Remark 1.10. If on the other hand G is an infinite subgroup of F^*, then G is not in general cyclic. For example, $\mathbb{Q}^*$ is not a cyclic group. The situation for $\mathbb{R}^*$ is even more drastic: $\mathbb{R}^*$ is uncountable, but every cyclic group is either finite or isomorphic to $\mathbb{Z}$, hence countable.

2 Factorization and principal ideals

The outline of the discussion of factorization in $F[x]$ is very similar to that for factorization in $\mathbb{Z}$. We begin with:

Proposition 2.1. *Every ideal in $F[x]$ is a principal ideal.*

Proof. Let I be an ideal in $F[x]$. If $I = \{0\}$, then clearly $I = (0)$ as well, and so I is principal. Thus we may assume that $I \neq \{0\}$. Let $f(x) \in I$ be a nonzero polynomial such that $\deg f(x)$ is the minimal possible value among nonnegative integers of the form $\deg g(x)$, where $g(x) \in I$ and $g(x) \neq 0$. More precisely, the set of nonnegative integers

$$\{\deg g(x) : g(x) \in I \text{ and } g(x) \neq 0\}$$

is a nonempty subset of $\mathbb{N} \cup \{0\}$ and hence by the well-ordering principle has a smallest element, necessarily of the form $\deg f(x)$ for some non-zero polynomial $f(x) \in I$. We claim that $f(x)$ is a generator of I, i.e. that $I = (f(x))$.

Clearly, as $f(x) \in I$, $(f(x)) \subseteq I$. To see the opposite inclusion, let $g(x) \in I$. Then we can apply long division with remainder to $f(x)$ and $g(x)$: there exist $q(x), r(x) \in F[x]$, with either $r(x) = 0$ or $\deg r(x) < \deg f(x)$, such that $g(x) = f(x)q(x) + r(x)$. Since $g(x) \in I$ and $(f(x)) \subseteq I$, $r(x) = g(x) - f(x)q(x) \in I$. But, if $r(x) \neq 0$, then $\deg r(x) < \deg f(x)$, contradicting the choice of $f(x)$. Hence $r(x) = 0$, so that $g(x) = f(x)q(x) \in (f(x))$. Since $g(x)$ was an arbitrary element of I, it follows that $I \subseteq (f(x))$ and hence that $I = (f(x))$. Hence I is principal. □

Definition 2.2. Let $f(x), g(x) \in F[x]$, where not both of $f(x), g(x)$ are zero. A *greatest common divisor* of $f(x)$ and $g(x)$, written $\gcd(f(x), g(x))$, is a polynomial $d(x)$ such that

1. The polynomial $d(x)$ is a divisor of both $f(x)$ and $g(x)$: $d(x)\big|f(x)$ and $d(x)\big|g(x)$.

2. If $e(x)$ is a polynomial such that $e(x)\big|f(x)$ and $e(x)\big|g(x)$, then $e(x)\big|d(x)$.

Proposition 2.3. *Let $f(x), g(x) \in F[x]$, not both 0.*

(i) *If $d(x)$ is a greatest common divisor of $f(x)$ and $g(x)$, then so is $cd(x)$ for every $c \in F^*$.*

(ii) *If $d_1(x)$ and $d_2(x)$ are two greatest common divisors of $f(x)$ and $g(x)$, then there exists a $c \in F^*$ such that $d_2(x) = cd_1(x)$.*

(iii) *A greatest common divisor $d(x)$ of $f(x)$ and $g(x)$ exists and is of the form $d(x) = r(x)f(x) + s(x)g(x)$ for some $r(x), s(x) \in F[x]$.*

Proof. (i) This is clear from the definition.

(ii) if $d_1(x)$ and $d_2(x)$ are two greatest common divisors of $f(x)$ and $g(x)$, then by definition $d_1(x)\big|d_2(x)$ and $d_2(x)\big|d_1(x)$. Thus there exist $u(x), v(x) \in F[x]$ such that $d_2(x) = u(x)d_1(x)$ and $d_1(x) = v(x)d_2(x)$. Hence $d_1(x) = u(x)v(x)d_1(x)$. Since a greatest common divisor can never be 0 (it must divide both $f(x)$ and $g(x)$ and at least one of these is non-zero) and $F[x]$ is an integral domain, it follows that $1 = u(x)v(x)$, i.e. both $u(x)$ and $v(x)$ are units in $F[x]$, hence elements of F^*. Thus $d_2(x) = cd_1(x)$ for some $c \in F^*$.

(iii) To see existence, define

$$(f(x), g(x)) = (f(x)) + (g(x)) = \{r(x)f(x) + s(x)g(x) : r(x), s(x) \in F[x]\}.$$

It is easy to see that $(f(x), g(x))$ is an ideal (it is the ideal sum of the principal ideals $(f(x))$ and $(g(x))$) and that $f(x), g(x) \in (f(x), g(x))$. By Proposition 2.1, there exists a $d(x) \in F[x]$ such that $(f(x), g(x)) = (d(x))$. In particular, $d(x) = r(x)f(x) + s(x)g(x)$ for some $r(x), s(x) \in F[x]$, and, as $f(x), g(x) \in (d(x))$, $d(x)\big|f(x)$ and $d(x)\big|g(x)$. Finally, if $e(x)\big|f(x)$ and $e(x)\big|g(x)$, then it is easy to check that $e(x)$ divides every expression of the form $r(x)f(x)+s(x)g(x)$. Hence $e(x)\big|d(x)$, and so $d(x)$ is a greatest common divisor of $f(x)$ and $g(x)$. $\square$

Remark 2.4. We could specify the gcd of $f(x)$ and $g(x)$ uniquely by requiring that it be monic. However, for more general rings, this choice is not available, and we will allow there to be many different gcds of $f(x)$ and $g(x)$, all related by multiplication by a unit of $F[x]$, in other words a nonzero constant polynomial.

Definition 2.5. Let $f(x), g(x) \in F[x]$. Then $f(x)$ and $g(x)$ are *relatively prime* if 1 is a gcd of $f(x)$ and $g(x)$. It is easy to see that this definition is equivalent to: there exist $r(x), s(x) \in F[x]$ such that $1 = r(x)f(x) + s(x)g(x)$. (If 1 is a gcd of $f(x)$ and $g(x)$, then $1 = r(x)f(x) + s(x)g(x)$ for some $r(x), s(x) \in F[x]$ by Proposition 2.3. Conversely, if $1 = r(x)f(x) + s(x)g(x)$, then a gcd $d(x)$ of $f(x), g(x)$ must divide 1 and hence is a unit c, and hence after multiplying by c^{-1} we see that 1 is a gcd of $f(x)$ and $g(x)$.)

Proposition 2.6. *Let $f(x), g(x) \in F[x]$ be relatively prime, and suppose that $f(x)\big|g(x)h(x)$ for some $h(x) \in F[x]$. Then $f(x)\big|h(x)$.*

Proof. Let $r(x), s(x) \in F[x]$ be such that $1 = r(x)f(x) + s(x)g(x)$. Then

$$h(x) = r(x)f(x)h(x) + s(x)g(x)h(x).$$

Clearly $f(x)\big|r(x)f(x)h(x)$, and by assumption $f(x)\big|g(x)h(x)$ and hence $f(x)\big|s(x)g(x)h(x)$. Thus $f(x)$ divides the sum $r(x)f(x)h(x)+s(x)g(x)h(x) = h(x)$. $\square$

Definition 2.7. Let $p(x) \in F[x]$. Then $p(x)$ is *irreducible* if $p(x)$ is neither 0 nor a unit (i.e. $p(x)$ is a non-constant polynomial), and if $p(x) = f(x)g(x)$ for some $f(x), g(x) \in F[x]$, then either $f(x) = c \in F^*$ and hence $g(x) = c^{-1}p(x)$, or $g(x) = c \in F^*$ and $f(x) = c^{-1}p(x)$. Equivalently, $p(x)$ is not a product $f(x)g(x)$ of two polynomials $f(x), g(x) \in F[x]$ such that both $\deg f(x) < \deg p(x)$ and $\deg g(x) < \deg p(x)$. In other words: an irreducible polynomial is a non-constant polynomial that does not factor into a product of polynomials of strictly smaller degrees. Finally, we say that a polynomial is *reducible* if it is not irreducible.

Example 2.8. A linear polynomial (polynomial of degree one) is irreducible. A quadratic (degree 2) or cubic (degree 3) polynomial is reducible $\iff$ it has a linear factor in $F[x]$ $\iff$ it has a root in F. Thus for example $x^2 - 2$ is irreducible in $\mathbb{Q}[x]$ but not in $\mathbb{R}[x]$.

Proposition 2.9. *Let $p(x)$ be irreducible in $F[x]$.*

(i) *For every $f(x) \in F[x]$, either $p(x) \big| f(x)$ or $p(x)$ and $f(x)$ are relatively prime.*

(ii) *For all $f(x), g(x) \in F[x]$, if $p(x) \big| f(x)g(x)$, then either $p(x) \big| f(x)$ or $p(x) \big| g(x)$.*

Proof. (i) Let $d(x) = \gcd(p(x), f(x))$. Then $d(x) \big| p(x)$, so $d(x)$ is either a unit or a unit times $p(x)$, hence we can take for $d(x)$ either 1 or $p(x)$. If 1 is a gcd of $p(x)$ and $f(x)$, then $p(x)$ and $f(x)$ are relatively prime. If $p(x)$ is a gcd of $p(x)$ and $f(x)$, then $p(x) \big| f(x)$.

(ii) Suppose that $p(x) \big| f(x)g(x)$ but that $p(x)$ does not divide $f(x)$. By (i), $p(x)$ and $f(x)$ are relatively prime. By Proposition 2.6, since $p(x) \big| f(x)g(x)$ and $p(x)$ and $f(x)$ are relatively prime, $p(x) \big| g(x)$. Thus either $p(x) \big| f(x)$ or $p(x) \big| g(x)$. □

Corollary 2.10. *Let $p(x)$ be irreducible in $F[x]$, let $f_1(x), \dots, f_n(x) \in F[x]$, and suppose that $p(x) \big| f_1(x) \cdots f_n(x)$. Then there exists an i such that $p(x) \big| f_i(x)$.*

Proof. This is a straightforward inductive argument starting with the case $n = 2$ above. □

Theorem 2.11 (Unique factorization in polynomial rings)**.** *Let $f(x)$ be a non constant polynomial in $F[x]$, i.e. $f(x)$ is neither 0 nor a unit. Then there exist irreducible polynomials $p_1(x), \dots, p_k(x)$, not necessarily distinct, such* *that $f(x) = p_1(x)\cdots p_k(x)$. In other words, $f(x)$ can be factored into a product of irreducible polynomials (where, in case $f(x)$ is itself irreducible, we let $k = 1$ and view $f(x)$ as a one element "product"). Moreover, the factorization is unique up to multiplying by units, in the sense that, if $q_1(x), \dots, q_\ell(x)$ are irreducible polynomials such that*

$$f(x) = p_1(x)\cdots p_k(x) = q_1(x)\cdots q_\ell(x),$$

then $k = \ell$, and, possibly after reordering the q_i, for every i, $1 \le i \le k$, there exists a $c_i \in F^*$ such that $q_i(x) = c_i p_i(x)$.

Proof. The theorem contains both an existence and a uniqueness statement. To prove existence, we argue by complete induction on the degree $\deg f(x)$ of $f(x)$. If $\deg f(x) = 1$, then $f(x)$ is irreducible and we can just take $k = 1$ and $p_1(x) = f(x)$. Now suppose that existence has been shown for all polynomials of degree less than n, where $n > 1$, and let $f(x)$ be a polynomial of degree n. If $f(x)$ is irreducible, then as in the case $n = 1$ we take $k = 1$ and $p_1(x) = f(x)$. Otherwise $f(x) = g(x)h(x)$, where both $g(x)$ and $h(x)$ are nonconstant polynomials of degrees less than n. By the inductive hypothesis, both $g(x)$ and $h(x)$ factor into products of irreducible polynomials. Hence the same is the true of the product $g(x)h(x) = f(x)$. Thus every polynomial of degree n can be factored into a product of irreducible polynomials, completing the inductive step and hence the proof of existence.

To prove the uniqueness part, suppose that $f(x) = p_1(x)\cdots p_k(x) = q_1(x)\cdots q_\ell(x)$ where the $p_i(x)$ and $q_j(x)$ are irreducible. The proof is by induction on the number k of factors in the first product. If $k = 1$, then $f(x) = p_1(x)$ and $p_1(x)$ divides the product $q_1(x)\cdots q_\ell(x)$. By Corollary 2.10, there exists an i such that $p_1(x)\big|q_i(x)$. After relabeling the q_i, we can assume that $i = 1$. Since $q_1(x)$ is irreducible and $p_1(x)$ is not a unit, there exists a $c \in F^*$ such that $q_1(x) = cp_1(x)$. We claim that $\ell = 1$ and hence that $q_1(x) = f(x) = p_1(x)$. To see this, suppose that $\ell \ge 2$. Then

$$p_1(x) = cp_1(x)q_2(x)\cdots q_\ell(x).$$

Since $p_1(x) \neq 0$, we can cancel it to obtain $1 = cq_2(x)\cdots q_\ell(x)$. Thus $q_i(x)$ is a unit for $i \ge 2$, contradicting the fact that $q_i(x)$ is irreducible. This proves uniqueness when $k = 1$.

For the inductive step, suppose that uniqueness has been proved for all polynomials which are a product of $k - 1$ irreducible polynomials, and let $f(x) = p_1(x)\cdots p_k(x) = q_1(x)\cdots q_\ell(x)$ where the $p_i(x)$ and $q_j(x)$ are irreducible as above. As before, $p_1(x)\big|q_1(x)\cdots q_\ell(x)$ hence, there exists an i

such that $p_1(x)\big|q_i(x)$. After relabeling the $q_i(x)$, we can assume that $i = 1$ and that there exists a $c_1 \in F^*$ such that $q_1(x) = c_1p_1(x)$. Thus

$$p_1(x)\cdots p_k(x) = c_1p_1(x)q_2(x)\cdots q_\ell(x),$$

and so canceling we obtain $p_2(x)\cdots p_k(x) = (c_1q_2(x))\cdot\cdots\cdot q_\ell(x)$. Then, since the product on the left hand side involves $k-1$ factors, by induction $k-1 = \ell-1$ and hence $k = \ell$. Moreover there exist $c_i \in F^*$ such that $q_i(x) = c_ip_i(x)$ if $i > 2$, and $c_1q_2(x) = c_2p_2(x)$. After renaming $c_1^{-1}c_2$ by c_2, we see that $q_i(x) = c_ip_i(x)$ for all $i \geq 1$. This completes the inductive step and hence the proof of uniqueness. □

3 Prime and maximal ideals in $F[x]$

Theorem 3.1. *Let I be an ideal in $F[x]$. Then the following are equivalent:*

(i) *I is a maximal ideal.*

(ii) *I is a prime ideal and $I \neq \{0\}$.*

(iii) *There exists an irreducible polynomial $p(x)$ such that $I = (p(x))$.*

Proof. (i) $\implies$ (ii): We know that if an ideal I (in any ring R) is maximal, then it is prime. Also, the ideal $\{0\}$ is not a maximal ideal in $F[x]$, since there are other proper ideals which contain it, for example (x); alternatively, $F[x]/\{0\} \cong F[x]$ is not a field. Hence if I is a maximal ideal in $F[x]$, then I is a prime ideal and $I \neq \{0\}$.

(ii) $\implies$ (iii): Since every ideal in $F[x]$ is principal by Proposition 2.1, we know that $I = (p(x))$ for some polynomial $p(x)$, and must show that $p(x)$ is irreducible. Note that $p(x) \neq 0$, since $I \neq \{0\}$, and $p(x)$ is not a unit, since $I \neq F[x]$ is not the whole ring. Now suppose that $p(x) = f(x)g(x)$. Then $f(x)g(x) = p(x) \in (p(x))$, and hence either $f(x) \in (p(x))$ or $g(x) \in (p(x))$. Say for example that $f(x) \in (p(x))$. Then $f(x) = h(x)p(x)$ for some $h(x) \in F[x]$ and hence

$$p(x) = f(x)g(x) = h(x)g(x)p(x).$$

Canceling the factors $p(x)$, which is possible since $p(x) \neq 0$, we see that $h(x)g(x) = 1$. Hence $g(x)$ is a unit, say $g(x) = c \in F^*$, and thus $f(x) = c^{-1}p(x)$. It follows that $p(x)$ is irreducible.

(iii) $\implies$ (i): Suppose that $I = (p(x))$ for an irreducible polynomial $p(x)$. Since $p(x)$ is not a unit, no multiple of $p(x)$ is equal to 1, and hence $I \neq R$.

Suppose that J is an ideal of R and that $I \subseteq J$. We must show that $J = I$ or that $J = R$. In any case, we know by Proposition 2.1 that $J = (f(x))$ for some $f(x) \in F[x]$. Since $p(x) \in (p(x)) = I \subseteq J = (f(x))$, we know that $f(x) \big| p(x)$. As $p(x)$ is irreducible, either $f(x)$ is a unit or $f(x) = cp(x)$ for some $c \in F^*$. In the first case, $J = (f(x)) = R$, and in the second case $f(x) \in (p(x))$, hence $J = (f(x)) \subseteq (p(x)) = I$. Since by assumption $I \subseteq J$, $I = J$. Thus I is maximal. $\square$

Corollary 3.2. *Let $f(x) \in F[x]$. Then $F[x]/(f(x))$ is a field $\iff$ $f(x)$ is irreducible.* $\square$

Remark 3.3. While the above corollary may seem very surprising, one way to think about it is as follows: if $f(x)$ is irreducible, and given a nonzero coset $g(x) + (f(x)) \in F[x]/(f(x))$, we must find a multiplicative inverse for $g(x) + (f(x))$. Now, assuming that $f(x)$ is irreducible, $g(x) + (f(x))$ is not the zero coset $\iff$ $f(x)$ does not divide $g(x)$ $\iff$ $f(x)$ and $g(x)$ are relatively prime, by Proposition 2.9 $\iff$ there exist $r(x), s(x) \in F[x]$ such that $1 = r(x)f(x) + s(x)g(x)$. In this case, the coset $s(x) + (f(x))$ is a multiplicative inverse for the coset $g(x) + (f(x))$, since then

$$\begin{aligned}(s(x) + (f(x)))(g(x) + (f(x))) &= s(x)g(x) + (f(x)) \\ &= 1 - r(x)f(x) + (f(x)) = 1 + (f(x)).\end{aligned}$$

In fact, we can find the polynomials $r(x), s(x)$ quite explicitly by a variant of the Euclidean algorithm. We will discuss this later.

Given a field F and a nonconstant polynomial $f(x) \in F[x]$, we now use the above to construct a possibly larger field E containing a subfield isomorphic to F such that $f(x)$ has a root in E. Here, and in the following discussion, if $\rho\colon F \to E$ is an isomorphism from F to a subfield $\rho(F)$ of E, we use ρ to identify $F[x]$ with $\rho(F)[x] \le E[x]$.

Theorem 3.4. *Let $f(x) \in F[x]$ be a nonconstant polynomial. Then there exists a field E containing a subfield isomorphic to F such that $f(x)$ has a root in E.*

Proof. Let $p(x)$ be an irreducible factor of $f(x)$. It suffices to find a field E containing a subfield isomorphic to F such that $p(x)$ has a root α in E, for then $f(x) = p(x)g(x)$ for some $g(x) \in F[x]$ and $f(\alpha) = p(\alpha)g(\alpha) = 0$. The quotient ring $E = F[x]/(p(x))$ is a field by Corollary 3.2, the homomorphism $\rho(a) = a + (p(x))$ is an injective homomorphism from F to E, and the coset $\alpha = x + (p(x))$ is a root of $f(x)$ in E. $\square$

Corollary 3.5. *Let $f(x) \in F[x]$ be a nonconstant polynomial. Then there exists a field E containing a subfield isomorphic to F such that $f(x)$ factors into linear factors in $E[x]$. In other words, every irreducible factor of $f(x)$ in $E[x]$ is linear.*

Proof. The proof is by induction on $n = \deg f(x)$ and the case $n = 1$ is obvious. Suppose that the corollary has been proved for all fields F and for all polynomials in $F[x]$ of degree $n-1$. If $\deg f(x) = n$, by Corollary 3.4 there exists a field E_1 containing a subfield isomorphic to F and a root α of $f(x)$ in E_1. Thus, in $E_1[x]$, $f(x) = (x-\alpha)g(x)$, where $g(x) \in E_1[x]$ and $\deg g(x) = n-1$. By the inductive hypothesis applied to the field E_1 and the polynomial $g(x) \in E_1[x]$, there exists a field E containing a subfield isomorphic to E_1 such that $g(x)$ factors into linear factors in $E[x]$. Since E contains a subfield isomorphic to E_1 and E_1 contains a subfield isomorphic to F, the composition of the two isomorphisms gives an isomorphism from F to a subfield of E. Then, in $E[x]$, $f(x)$ is a product of $x-\alpha$ and a product of linear factors, and is thus a product of linear factors. This completes the inductive step. $\square$