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 R be an integral domain and let $r, s \in R$. We say that r *divides* s, written $r\big|s$, if there exists a $t \in R$ such that $s = rt$, i.e. s is a multiple of r. Thus, for example, every $r \in R$ divides 0, but r is divisible by $0 \iff r = 0$.

By definition, r is a unit $\iff r\big|1$. We claim that r is a unit $\iff r\big|s$ for all $s \in R \iff r\big|1$. (Proof: if r is a unit, then, for all $s \in R$, $s = r(r^{-1}s)$ and hence $r\big|s$. Next, $r\big|s$ for all $s \in R \implies r\big|1$, and finally $r\big|1 \implies r$ is a unit.) We will usually ignore units when we discuss factorization because they contribute what are essentially trivial factors.

In case $R = F[x]$, 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 divides every $g \in F[x] \iff f$ divides 1 $\iff f \in F^*$ is a nonzero constant polynomial. Note that, if $c \in F^*$ is a unit, then $f\big|g \iff cf\big|g \iff f\big|cg$.

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

$$g = fq + r.$$

Proof. First we prove existence. The proposition is clearly true if $g = 0$, since then we can take $q = r = 0$. Otherwise, we argue by induction on $\deg g$. If $\deg g = 0$ and $\deg f = 0$, then $f = c \in F^*$ is a nonzero constant, and then $g = c(c^{-1}g) + 0$, so we can take $q = c^{-1}g$ and $r = 0$. If $\deg g = 0$ and $\deg f > 0$, or more generally if $n = \deg g < \deg f = d$, then we can take

$q = 0$ and $r = g$. Now assume that, for a fixed f, the existence of q and r has been proved for all polynomials of degree $< n$, and suppose that g is a polynomial of degree n. As above, we can assume that $n \geq d = \deg f$. Let $f = \sum_{i=0}^{d} a_i x^i$, with $a_d \neq 0$, and let $g = \sum_{i=0}^{n} b_i x^i$. In this case, $g - b_n a_d^{-1} x^{n-d} f$ is a polynomial of degree at most $n - 1$ (or 0). By the inductive hypothesis and the case $g = 0$, there exist polynomials $q_1, r \in F[x]$ with either $r = 0$ or $\deg r < \deg f$, such that

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

Then

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

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

To see uniqueness, suppose that

$$g = f q_1 + r_1 = f q_2 + r_2,$$

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

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

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

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

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

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

a contradiction. Thus $q_1 - q_2 = 0$, hence $r_2 - r_1 = 0$ as well. It follows that $q_1 = q_2$ and $r_2 = r_1$, 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 is **monic**, in other words, $f \neq 0$ and its leading coefficient is 1. The proof is essentially the same.

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

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

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

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

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

Finally, to determine c, we evaluate f at a:

$$f(a) = \mathrm{ev}_a(f) = \mathrm{ev}_a((x - a)g + c) = 0 + c = c.$$

Hence $c = f(a)$. $\square$

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

Corollary 1.6. *Let $f \in F[x]$, $f \neq 0$, and suppose that* $\deg f = d$. *Then there are at most d roots of f 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 $\deg f = d$ and there is no root of f in F, then we are done because $d \geq 0$. Otherwise, let a_1 be a root. Then we can write $f = (x - a_1)g$, where $\deg g = d - 1$. Let a_2 be a root of f 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 (here we must use the fact that F is a field). By induction, g has at most $d - 1$ roots in F (where we allow for the possibility that a_1 is also a root of g). 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 1.7. *Let F be an infinite field. Then the evaluation homomorphism E from $F[x]$ to F^F is injective. In other words, if $f_1, f_2 \in F[x]$ are two polynomials which define the same function, i.e. are such that $f_1(a) = f_2(a)$ for all $a \in F$, then $f_1 = f_2$.*

Proof. It suffices to prove that $\operatorname{Ker} E = \{0\}$, i.e. that if $f \in F[x]$ and $f(a) = 0$ for all $a \in F$, then $f = 0$. This is clear from Corollary 1.6, since a nonzero polynomial can have at most finitely many roots and F was assumed infinite. $\square$

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 1.8 (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 1.9. *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 in Modern Algebra I, 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 $\psi \colon \mathbb{N} \to \mathbb{Z}$ 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. $\square$

Example 1.10. (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 1.11. 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 $f \in I$ be a non-zero polynomial such that $\deg f$ is the minimal possible value among nonnegative integers of the form $\deg g$, where $g \in I$ and $g \neq 0$. More precisely, the set of nonnegative integers

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

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

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

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

1. The polynomial d is a divisor of both f and g: $d \big| f$ and $d \big| g$.
2. If e is a polynomial such that $e \big| f$ and $e \big| g$, then $e \big| d$.

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

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

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

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

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

(ii) if d_1 and d_2 are two greatest common divisors of f and g, then by definition $d_1 \big| d_2$ and $d_2 \big| d_1$. Thus there exist $u, v \in F[x]$ such that $d_2 = ud_1$ and $d_1 = vd_2$. Hence $d_1 = uvd_1$. Since a greatest common divisor can never be 0 (it must divide both f and g and at least one of these is non-zero) and $F[x]$ is an integral domain, it follows that $1 = uv$, i.e. both u and v are units in $F[x]$, hence elements of F^*. Thus $d_2 = cd_1$ for some $c \in F^*$.

(iii) To see existence, define

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

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

Remark 2.4. We could specify the gcd of f and g 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 and g, all related by multiplication by a unit of $F[x]$, in other words a nonzero constant polynomial.

Remark 2.5. In fact, we can find the polynomials r, s described in (iii) of the proposition quite explicitly by a variant of the Euclidean algorithm.

Remark 2.6. If R is a general integral domain, then we can define a greatest common divisor of a and b by the obvious analogue of Definition 2.2. However, in a general integral domain, greatest common divisors may not exist, and even when they do always exist, they need not be given as linear combinations $ar + bs$ as in Part (iii) of Proposition 2.3.

Definition 2.7. Let $f, g \in F[x]$. Then f and g are *relatively prime* if 1 is a gcd of f and g. It is easy to see that this definition is equivalent to: there exist $r, s \in F[x]$ such that $1 = rf + sg$. (If 1 is a gcd of f and g, then $1 = rf + sg$ for some $r, s \in F[x]$ by Proposition 2.3. Conversely, if $1 = rf + sg$, then a gcd d of f, g 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 and g.)

Proposition 2.8. *Let $f, g \in F[x]$ be relatively prime, and suppose that $f \big| gh$ for some $h \in F[x]$. Then $f \big| h$.*

Proof. Let $r, s \in F[x]$ be such that $1 = rf + sg$. Then

$$h = rfh + sgh.$$

Clearly $f \big| rfh$, and by assumption $f \big| gh$ and hence $f \big| sgh$. Thus f divides the sum $rfh + sgh = h$. $\square$

Definition 2.9. Let $p \in F[x]$. Then p is *irreducible* if p is neither 0 nor a unit (i.e. p is a non-constant polynomial), and if $p = fg$ for some $f, g \in F[x]$, then either $f = c \in F^*$ and hence $g = c^{-1}p$, or $g = c \in F^*$ and $f = c^{-1}p$. Equivalently, p is not a product fg of two polynomials $f, g \in F[x]$ such that both $\deg f < \deg p$ and $\deg g < \deg p$. 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 2.10. 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 $\mathbb{R}[x]$, and the same is true for $x^3 - 2$. . Likewise $x^2 + 1$ is irreducible in $\mathbb{R}[x]$ but not in $\mathbb{C}[x]$. The polynomial $f = x^2 + x + 1$ is irreducible in $\mathbb{F}_2[x]$ as it does not have a root in $\mathbb{F}_2$. ($f(0) = f(1) = 1$.)

On the other hand, the polynomial $x^4 - 4$ is **not** irreducible in $\mathbb{Q}[x]$, even though it does not have a root in $\mathbb{Q}$.

Proposition 2.11. *Let p be irreducible in $F[x]$.*

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

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

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

(ii) Suppose that $p \big| fg$ but that p does not divide f. By (i), p and f are relatively prime. By Proposition 2.8, since $p \big| fg$ and p and f are relatively prime, $p \big| g$. Thus either $p \big| f$ or $p \big| g$. $\square$

Corollary 2.12. *Let p be irreducible in $F[x]$, let $f_1, \dots, f_n \in F[x]$, and suppose that $p \big| f_1 \cdots f_n$. Then there exists an i such that $p \big| f_i$.*

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

Theorem 2.13 (Unique factorization in polynomial rings)**.** *Let f be a non constant polynomial in $F[x]$, i.e. f is neither 0 nor a unit. Then there exist irreducible polynomials $p_1, \dots, p_k$, not necessarily distinct, such that $f = p_1 \cdots p_k$. In other words, f can be factored into a product of irreducible polynomials (where, in case f is itself irreducible, we let $k = 1$ and view f as a one element "product"). Moreover, the factorization is unique up to multiplying by units, in the sense that, if $q_1, \dots, q_\ell$ are irreducible polynomials such that*

$$f = p_1 \cdots p_k = q_1 \cdots q_\ell,$$

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 $q_i = c_i p_i$.

Proof. The theorem contains both an existence and a uniqueness statement. To prove existence, we argue by complete induction on the degree $\deg f$ of f. If $\deg f = 1$, then f is irreducible and we can just take $k = 1$ and $p_1 = f$. Now suppose that existence has been shown for all polynomials of degree less than n, where $n > 1$, and let f be a polynomial of degree n. If f is irreducible, then as in the case $n = 1$ we take $k = 1$ and $p_1 = f$. Otherwise $f = gh$, where both g and h are nonconstant polynomials of degrees less than n. By the inductive hypothesis, both g and h factor into products of irreducible polynomials. Hence the same is the true of the product $gh = f$. 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 $f = p_1 \cdots p_k = q_1 \cdots q_\ell$ where the p_i and q_j are irreducible. The proof is by induction on the number k of factors in the first product. If $k = 1$, then $f = p_1$ and p_1 divides the product $q_1 \cdots q_\ell$. By Corollary 2.12, there exists an i such that $p_1 \big| q_i$. After relabeling the q_i, we can assume that $i = 1$. Since q_1 is irreducible and p_1 is not a unit, there exists a $c \in F^*$ such that $q_1 = cp_1$. We claim that $\ell = 1$ and hence that $q_1 = f = p_1$. To see this, suppose that $\ell \ge 2$. Then

$$p_1 = cp_1 q_2 \cdots q_\ell.$$

Since $p_1 \ne 0$, we can cancel it to obtain $1 = cq_2 \cdots q_\ell$. Thus q_i is a unit for $i \ge 2$, contradicting the fact that q_i 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 $f = p_1 \cdots p_k = q_1 \cdots q_\ell$ where the p_i and q_j are irreducible as above. As before, $p_1 \big| q_1 \cdots q_\ell$ hence, there exists an i such that $p_1 \big| q_i$. After relabeling the q_i, we can assume that $i = 1$ and that there exists a $c_1 \in F^*$ such that $q_1 = c_1 p_1$. Thus

$$p_1 \cdots p_k = c_1 p_1 q_2 \cdots q_\ell,$$

and so canceling we obtain $p_2 \cdots p_k = (c_1 q_2) \cdot \cdots \cdot q_\ell$. 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 $q_i = c_i p_i$ if $i > 2$, and $c_1 q_2 = c_2 p_2$. After renaming $c_1^{-1} c_2$ by c_2, we see that $q_i = c_i p_i$ 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 such that $I = (p)$.*

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 $I = (p)$ for some polynomial p, and must show that p is irreducible. Note that $p \neq 0$, since $I \neq \{0\}$, and p is not a unit, since $I \neq F[x]$ is not the whole ring. Now suppose that $p = fg$. Then $fg = p \in (p)$, and hence either $f \in (p)$ or $g \in (p)$. Say for example that $f \in (p)$. Then $f = hp$ for some $h \in F[x]$ and hence

$$p = fg = hgp.$$

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

(iii) $\implies$ (i): Suppose that $I = (p)$ for an irreducible polynomial p. Since p is not a unit, no multiple of p 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 $J = (f)$ for some $f \in F[x]$. Since $p \in (p) = I \subseteq J = (f)$, we know that $f \big| p$. As p is irreducible, either f is a unit or $f = cp$ for some $c \in F^*$. In the first case, $J = (f) = R$, and in the second case $f \in (p)$, hence $J = (f) \subseteq (p) = I$. Since by assumption $I \subseteq J$, $I = J$. Thus I is maximal. $\square$

Corollary 3.2. *Let $f \in F[x]$. Then $F[x]/(f)$ is a field $\iff$ f 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 is irreducible, and given a nonzero coset $g + (f) \in F[x]/(f)$, we must find a multiplicative inverse for $g + (f)$. Now, assuming that f is irreducible, $g + (f)$ is not the zero coset $\iff$ f does not divide g $\iff$ f and g are relatively prime, by Proposition 2.11 $\iff$ there exist $r, s \in F[x]$ such that $1 = rf + sg$. In this case, the coset $s + (f)$ is a multiplicative inverse for the coset $g + (f)$, since then

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

Thus, the Euclidean algorithm for polynomials gives an effective way to find inverses.

Given a field F and a nonconstant polynomial $f \in F[x]$, we now use the above to construct a possibly larger field E containing a subfield isomorphic to F such that f 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 $\rho(F)[x] \leq E[x]$.

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

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

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

Proof. The proof is by induction on $n = \deg f$ 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 $\deg f = n$, by Corollary 3.4 there exists a field E_1 containing a subfield isomorphic to F and a root α of f in E_1. Thus, in $E_1[x]$, $f = (x-\alpha)g$, where $g \in E_1[x]$ and $\deg g = n-1$. By the inductive hypothesis applied to the field E_1 and the polynomial $g \in E_1[x]$, there exists a field E containing a subfield isomorphic to E_1 such that g 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 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. $\square$