Notes on Algebra II

Nilay Kumar

Last updated: April 1, 2013

1 Reducibility

February 25, 2013

Let F be a field, and $F[x]$ be the ring of polynomials over F. Recall we have already shown that every ideal in $F[x]$ is principal, and that there exists a unique gcd of two non-zero polynomials. Additionally, we showed that if f and g are two relatively prime polynomials, then $f|gh \implies f|h$.

Definition 1. A polynomial $p(x) \in F[x]$ is **irreducible** if $\deg p(x) > 0$, i.e. p is not zero and not a unit, and if $p = fg$ implies that one of f, g is a unit and the other is a unit times p. In words, $p(x)$ is irreducible if it does not factor into a product of two polynomials with strictly smaller (non-zero) degree. A polynomial is said to be **reducible** if it is not irreducible.

Example 1. (Reducibility)

(i) Any linear polynomial $x + a$ is obviously irreducible.

(ii) Any quadratic polynomial is clearly reducible if and only if it has two linear factors. This is equivalent to the polynomial having a root, as long division will yield the second factor.

(iii) Similarly, a cubic polynomial is reducible if and only if it has a root.

(iv) For higher degrees, the existence of a root is not equivalent to reducibility, as we will see in the next example.

Example 2. (Simple examples)

- $x^2 - 2$ is irreducible in $\mathbb{Q}[x]$, as it has no roots in $\mathbb{Q}$. It is, however, reducible in $\mathbb{R}[x]$: $x^2 - 2 = (x - \sqrt{2})(x + \sqrt{2})$.
- x^2+1 is irreducible in $\mathbb{R}[x]$ but reducible in $\mathbb{C}[x]$: $x^2+1 = (x-i)(x+i)$.

- $x^3 - 2$ is irreducible in $\mathbb{Q}[x]$, but reducible in $\mathbb{R}[x]$, where we can write it as a product of $x - \sqrt[3]{2}$ and an irreducible quadratic.
- $x^4 - 4 = (x^2 - 2)(x^2 + 2)$ is reducible in $\mathbb{Q}[x]$ but has no roots!

In fact, it is generally a hard problem to determine whether an arbitrary polynomial $f(x) \in \mathbb{Q}[x]$ is irreducible. Note, however, that we can think of irreducibility in analogy to that for natural numbers, as the following dichotomy illustrates.

Remark. If $p(x) \in F[x]$ is irreducible, then for any polynomial $f \in F[x]$, either $p|f$ or p and f are relatively prime.

Proof. Let $d = \gcd(p, f)$. By definition, d divides p. However, as p is irreducible, d must either be a unit or d must be cp for c a unit. In the first case, since the gcd of p and f is a unit, p and f must be relatively prime. In the second case, since $d = cp$ by construction divides f, p must divide f. □

Corollary 1. *If $p \in F[x]$ is irreducible and $p|fg$, then either $p|f$ or $p|g$.*

Proof. By the above remark, either $p|f$ or p and f are relatively prime. If $p|f$, we are done. Otherwise, p is relatively prime to f, and by what we showed last class, $p|g$. □

Theorem 2 (Unique factorization of polynomials)**.** *Let $f(x) \in F[x]$ with $\deg f(x) > 0$. Then there exist k irreducible polynomials in $F[x]$ such that*

$$f(x) = \prod_{i=1}^{k} p_i(x).$$

Additionally, if it is also true that $f(x) = \prod_{i=1}^{l} q_i$, then $k = l$, and after some reordering, there exist nonzero constants such that $q_i = c_i p_i$.

In other words, for any polynomial with degree greater than zero, there always exists a unique factorization into a product of irreducible polynomials.

Proof. Let us first show existence. We proceed by complete induction on the degree of f. If $\deg f = 1$, f is irreducible, and we are done. Otherwise, we assume that the theorem holds for all degrees less than n. Let $\deg f = n$. If f is irreducible, we are done. Otherwise, $f = g_1 g_2$ with $\deg g_1 < n$ and $\deg g_2 < n$. By the inductive hypothesis, g_1 and g_2 are products of irreducible polynomials, and thus f must be as well, and we are done.

The real muscle of this theorem comes in the form of uniqueness. Suppose $f = \prod_{i=1}^{k} p_i = \prod_{j=1}^{l} q_j$, with p_i, q_j reducible. We proceed by induction on k. If $k = 1$, $p_1 = q_1 \cdots q_l$. Clearly, then, $p_1 | q_1 \cdots q_l$, and thus (by induction over the statement at the beginning of lecture), p_1 must divide q_i for some i. But the q_i are irreducible and p_1 is not a constant, so $p_1 = cq_i$ for some unit c. If we now reorder terms, we can assume that $i = 1$ and we can cancel:

$$\begin{aligned} p_1 = cq_1 &= q_1 q_2 \cdots q_l \\ c &= q_2 \cdots q_l. \end{aligned}$$

But this is impossible, as the product of q's has degree greater than zero. Consequently, l must be 1, and thus $p_1 = q_1$ and we have shown that $k = l$. The general case is similar; we write $p_1 \cdots p_k = q_1 \cdots q_l$. Then $p_1 | q_1 \cdots q_l$, and so for some i, $p_1 = cq_i$. After reordering, we can write

$$\begin{aligned} cq_1 p_2 \cdots p_k &= q_1 \cdots q_l \\ cp_2 \cdots p_k &= q_2 \cdots q_l, \end{aligned}$$

and by induction, we know that $k - 1 = l - 1$. Reordering, we can write $p_= c_i q_i$ for $i = 2 \cdots k$, and we are done. □

Note that the irreducible factors need not be distinct.

Theorem 3. *Let F be a field. 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) $I = (p)$, where p is a irreducible polynomial.

Proof. Let us first show that $(i) \implies (ii)$. Say I is maximal. Then, I must be prime. Additionally, I cannot be the zero ideal, as it is not maximal, and so we are done.

Showing $(ii) \implies (iii)$ is a little trickier. Suppose I is a prime ideal with $I \neq \{0\}$. We want to show that the ideal is generated by an irreducible element. Since every ideal in $F[x]$ is principal, $I = (p)$ for some $p \in F[x]$. Let us show that p is irreducible. First note that p cannot be a unit, because otherwise $1 \in (p)$ which implies that $(p) = F[x]$, which is not possible for prime ideals. Furthermore, $p \neq 0$, as I is assumed not to be the zero ideal.

To show that p is irreducible, we need to show that if $p = fg$ then one of f, g is a unit and the other is a unit times p. So take $p = fg$. Then, $fg \in (p) = I$. Since I is prime, either $f \in I$ or $g \in I$. Take the first case, $f \in (p)$. Then, $f = hp$ for some $h \in F[x]$, and so $p = hpg \implies 1 = hg$, i.e. h, g are units, and thus f is a unit times p. Thus, p is irreducible.

Finally, we show that $(iii) \implies (i)$. Let $I = (p)$, with p irreducible. We wish to show that I is maximal, i.e. $(p) \neq F[x]$ and if $(p) \subset J$ then either $J = (p)$ or $J = F[x]$. First note that $(p) \neq F[x]$ because $\deg p > 1$ and so it can't generate constants. Next, since J is necessarily a principal ideal, $J = (f)$, for some $f \in F[x]$. If $(p) \subset (f)$, then $p \in (f)$, so $p = fg$ for some $g \in F[x]$. But p is irreducible, so either f is a unit, in which case $J = (f) = F[x]$, or $f = cp$, for c a unit, in which case $J = (f) = (p)$. Hence, I is maximal. □

This theorem is quite handy in constructing interesting fields, as the following corollary shows.

Corollary 4. *$F[x]/(f)$ is a field if and only if f is irreducible.*

Proof. This follows from above theorem and the fact that $F[x]/(f)$ is a field if and only if (f) is a maximal ideal. □

This allows us to show that certain rings are, in fact, fields – something that may not have been obvious – or, in fact, to find wholly new fields.

Example 3.

- $\mathbb{Q}[x]/(x^2-2)$ is a field, as x^2-2 is irreducible in $\mathbb{Q}[x]$, and its elements, by what we know about long division, are of the form $c + d\alpha$, where $\alpha = x + (x^2 - 2)$. In addition, $\alpha^2 = 2$.
- $\mathbb{R}[x]/(x^2+1)$ is a field, as x^2+1 is irreducible in $\mathbb{R}[x]$, and its elements are of the form $c + d\alpha$ where $\alpha = x + (x^2 + 1)$ satisfies $\alpha^2 = -1$.
- $\mathbb{Q}[x]/(x^3-2)$ is a field, as x^3-2 is irreducible in $\mathbb{Q}[x]$, and its elements are of the form $c + d\alpha + e\alpha^2$, where $\alpha = x + (x^3 - 2)$ satisfies $\alpha^3 = 2$. We often rewrite the elements as $c + d\sqrt[3]{2} + e\sqrt[3]{2}^2$.
- Take the finite field $\mathbb{F}_2$ and the polynomial $x^2 + x + 1 \in \mathbb{F}_2$. Since the only members of $\mathbb{F}_2$ are 0 and 1, it should be clear that this polynomial has no roots, and thus is irreducible in $\mathbb{F}_2[x]$. Consequently, $E = \mathbb{F}_2[x]/(x^2 + x + 1)$ is a field. Its elements are of the form $c + d\alpha$, where

of course $c, d \in \mathbb{F}_2$ and $\alpha = x+(x^2+x+1)$, which satisfies the property that $\alpha^2 = -\alpha - 1 = \alpha + 1$. E has four elements (since c and d can each take 2 values).

2 Field extensions

February 27, 2013

In general, a problem in algebra is to enlarge the domain of discourse, i.e. $\mathbb{R} \to \mathbb{C}$, so that one can solve equations that were hitherto unsolvable. So given a polynomial $f(x) \in F[x]$, we wish to find a root of $f(x)$. Maybe there is not root of $f(x)$ in F, so we wish to enlarge F. A simple idea that we saw earlier was that if R is any ring with $f(x) \in R[x]$, there was a way to enlarge R. We consider $R[x]/(f(x))$, which always has a root $x+(f(x)) = \alpha$. By construction, $f(\alpha) = 0$.

There is a problem with this – we don't know much about the algebraic structure of this new quotient ring, $R[x]/I$. The solution is: if R is a field, and $f(x)$ is irreducible, then $R[x]/I$ is good, i.e. it is a field, as we saw last lecture. But really, even if $f(x)$ is reducible, we should think in analogy to the world of $\mathbb{Z}/n\mathbb{Z}$, where $n > 0$. The full details of this analogy are fleshed out in the file `analogy.pdf`.

Theorem 5. *Let $f(x) \in F[x]$ and suppose $\deg f(x) > 0$, i.e. f is nonconstant. Then, there exists a field E that contains (a subfield isomorphic to) F, and an element $\alpha \in E$ such that $f(\alpha) = 0$.*

Proof. Take $f(x)$ and find an irreducible factor (we know these exist from last time) $p(x)$. We consider $E = F[x]/(p(x))$; E is a field. Additionally, there is a map $F \to E$ that sends $a \in F \mapsto a + (p(x))$. This map is injective (why?), and we identify F with the image subfield. We know that if $\alpha = x + (p(x))$, then $p(\alpha) = 0$. But $p(x)|f(x)$, so $f(\alpha) = 0$ as well, and we are done. □

Corollary 6. *Let $f(x) \in F[x]$, $\deg f(x) > 0$. Then there exists a field E such that, in $E[x]$, $f(x)$ is a product of linear factors.*

Proof. We apply the above theorem to find E_1 and $\alpha_1 \in E_1$ such that $f(\alpha_1) = 0$ ($F \leq E_1$). In $E_1[x]$, $f(x) = (x - \alpha_1)g_1(x)$, where $\deg g_1(x) = \deg f(x) - 1$. Informally speaking, now all we have to do is to keep going! We find E_2 with $E_1 \leq E_2$ such that we can write $g_1(x) = (x - \alpha_2)g_2(x)$ with $\alpha_2 \in E_2$ with $\deg g_2 = \deg f(x) - 2$. We continute until we run out of degrees, and clearly we have factored f into linear factors. □

Let us now switch perspectives. Let's consider the situation where E and F are fields, and $F \leq E$, i.e. F is a subfield of E. We also say that E is an **extension field** of F. Note that we used the machinery of prime and maximal ideals to construct extensions. Now, given an extension, let us use these tools to analyze these fields.

Consider E an extension field of F, and let $\alpha \in E$. Look at $\mathrm{ev}_\alpha : F[x] \to E$. Note that $\operatorname{Im} \mathrm{ev}_\alpha = F[\alpha] \leq E$. At this point, all we know is that $F[\alpha]$ is an integral domain. We claim that there are exactly 2 possibilities.

1. $\ker \mathrm{ev}_\alpha = \{0\}$, i.e. that if $f(x) \in F[x]$, $f(x) \neq 0$, then $f(\alpha) \neq 0$. In this case, we say that α is **transcendental** over F. Additionally, $\mathrm{ev}_\alpha : F[x] \to E$ is injective, which suggests that it extends to an injection $F(x) \to E$, whose image we call $F(\alpha)$. Elements of this image have the form $f(\alpha)/g(\alpha)$ where $f, g \in F[x]$ and $g \neq 0$. This is the smallest subfield of E containing F and α.

   A famous example is that $\pi \in \mathbb{R}$ is transcendental over $\mathbb{Q}$ (Lindemann, 1880). The same holds for e. Note carefully, that $\pi \in \mathbb{R}$ is not transcendental over $\mathbb{R}$, as π is a root of $x - \pi$.c w

2. $\ker \mathrm{ev}_\alpha \neq \{0\}$, i.e. there exists an $f(x) \in F[x]$, $f \neq 0$ such that $f(\alpha) = 0$. In this case, we say that α is **algebraic** over F. What can we say here? An incredible amount, it turns out.

   First note that $\ker \mathrm{ev}_\alpha$ is a principal ideal in $F[x]$: $\ker \mathrm{ev}_\alpha = (p(x))$. Additionally, we know that $F[\alpha] = \operatorname{Im} \mathrm{ev}_\alpha \cong F[x]/\ker \mathrm{ev}_\alpha$. But since $F[x]$ is an integral domain (subring of a field), $(p(x))$ is a prime ideal that is not $\{0\}$. This means that $(p(x))$ is a maximal ideal and $p(x)$ is irreducible (by theorem proved last time). Consequently, $F[x]/(p(x)) = F[x]/\ker \mathrm{ev}_\alpha$ is a field, and $F[\alpha]$ is a field. Recall the example of $\mathbb{Q}[\sqrt[3]{2}]$ being a field. Thus, we now write $F[\alpha] = F(\alpha)$, which is the smallest subfield of E containing F and α.

   In particular, there is a unique monic generator of $(p(x)) = \ker \mathrm{ev}_\alpha$. It is denoted $\mathrm{irr}(\alpha, F, x)$, which is read "the irreducible polynomial for α over F." It satisfies $\mathrm{irr}(\alpha, F, \alpha) = 0$. Let us do a few examples:

   Example 4.

$$\mathrm{irr}(\frac{1}{2}, \mathbb{Q}, x) = x - \frac{1}{2}$$
$$\mathrm{irr}(\sqrt[3]{2}, \mathbb{Q}, x) = x^3 - 2$$
$$\mathrm{irr}(\sqrt[3]{2}, \mathbb{Q}(\sqrt[3]{2}), x) = x - \sqrt[3]{2}$$
$$\mathrm{irr}(\sqrt[3]{2}, \mathbb{Q}(\sqrt{2}), x) = ?$$

Remark. Note that if $f(x) \in F[x]$ is any polynomial such that $f(\alpha) = 0$, then the $\text{irr}(a, F, x) | f(x)$.

Proof. $f(\alpha) = 0 \iff f(x) \in \ker \text{ev}_\alpha = (\text{irr}(\alpha, F, x))$. By definition, then, $\text{irr}(\alpha, F, x) | f(x)$. □

For example, take $f(x) \in \mathbb{R}[x]$. If $f(i) = 0$, then $x^2 + 1 | f(x)$.

Mostly we will be working with the algebraic case, as the transcendental case belongs in a separate course.

Definition 2. Given $F \leq E$ and $\alpha \in E$ algebraic over F, we define the **degree** of α over F as $\deg \text{irr}(\alpha, F, x)$.

For example, $\deg_{\mathbb{Q}} \sqrt[3]{2} = 3$ and $\deg_{\mathbb{R}} \sqrt[3]{2} = 1$ and $\deg_{\mathbb{R}} i = 2$.

Definition 3. Let $F \leq E$. Then we say that E is a **simple extension** of F if $E = F(\alpha)$ for some $\alpha \in E$.

Roughly speaking, this means that we can extend F to E by throwing in only one more element – i.e. we can do this if α is transcendental. Take, for example, $\mathbb{Q}(\sqrt{2})$. $\sqrt{3} \notin \mathbb{Q}(\sqrt{2})$. Then, $\mathbb{Q}(\sqrt{2}) \leq \mathbb{Q}(\sqrt{2})(\sqrt{3}) = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. Then, $\mathbb{Q} \leq \mathbb{Q}(\sqrt{2} + \sqrt{3})$, one can find that $\mathbb{Q}(\sqrt{2} + \sqrt{3}) = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. This shows that simple extensions are not always obviously simple.

3 F-vector spaces

Let us take a detour through vector spaces.

Let F be a field. An **F-vector space** is an abelian group $(V, +)$ and a function $F \times V \to V$ called **scalar multiplication**, which we write as av such that:

1. $a(bv) = (ab)v$
2. $a(v + w) = av + aw$
3. $(a + b)v = av + bv$
4. $1 \cdot v = v$

A very useful example is $V = F^n = \{(a_1 \cdots a_n) : a_i \in F\}$, i.e. the Cartesian product of F with itself n times. We define addition componentwise, and multiply the scalar through each component, as usual. It is easy to

check that F^n is a vector space. The $n = 0$ case is allowed, as it is the zero vector space with only 0.

Another example is the space of functions $X \to F$ on any set X, which we denote by F^X. Functions are added pointwise as usual, and scalar multiplication is done pointwise as well.

The important example for us is actually a bit unexpected. Suppose E is an extension field of F. Then E is an F-vector space. E is already an abelian group and scalar multiplication is defined in the ordinary sense of multiplication. The rest of the axioms follow straightforwardly. Now, this is not as strange as it might look. The complex numbers, for example, are an extension field of the reals, and we are used to going back and forth between numbers/vectors: $a + bi \iff (a, b)$. Similarly, $\mathbb{Q}(\sqrt[3]{2})$ is a $\mathbb{Q}$-vector space as $a + b\sqrt[3]{2} + c\sqrt[3]{2}^2 \iff (a, b, c)$. Furthermore, $\mathbb{R}$ is a $\mathbb{Q}$-vector space (but a really big one).

One might ask, why does one require F to be a field? We can define a similar structure for a ring.

Definition 4. If R is a commutative ring with unity, then an **R-module** M is an abelian group $(M, +)$ with a scalar multiplication $R \times M \to M$ that satisfies the properties defined above for F-vector spaces.

These are, however, more interesting for algebra in general, and not so much for our case, where we will extensively use the field properties.

(March 4, 2013)

What we wish to do with these ideas is to use linear algebraic ideas to understand more deeply field extensions.

Let's talk about a few basic notions.

Definition 5. A **vector subspace** of an F-vector space V is a subgroup W of $(V, +)$ such that for all $a \in F$,$w \in W$, $aw \in W$. W then becomes an F-vector space in its own right.

Definition 6. $F : V_1 \to V_2$ is a **linear map** if

1. f is a homomorphism of abelian groups
2. f preserves scalar multiplication: $a, b \in F, v \in V_1$, $f(av) = af(v)$

A **linear isomorphism** is just a bijective linear map. Its inverse is linear as well.

Definition 7. Let V be an F-vector space and let $v_1, \ldots v_n \in V$. Then a **linear combination** of these vectors is an element of v of the form $a_1v_1 +$

$\ldots + a_n v_n$. We define the **span** of this set of vectors as the set of all such linear combinations. It should be clear that the span of a set of vectors is a vector space.

Definition 8. V is **finite dimensional** if there exists a set of vectors in V whose span equals V.

Note, for example, that F^n is finite dimensional (via the standard basis), but $F[x]$ is not. However, $F[x]$, does have many interesting finite dimensional subspaces. If we define P_n to be the set of polynomials in $F[x]$ with degree n or less, it forms a vector subspace spanned by $1, x, x^2, \ldots, x^n$.

Definition 9. A set of vectors $v_1, \ldots, v_n \in V$ are **linearly independent** if the only linear combination of them that yields zero is where the coefficients in the linear combination are all zero.

Theorem 7. *If V is an F-vector space, and $v_1, \ldots, v_n \in V$ are linearly independent, and $w_1, \ldots, w_m$ span V, then $n \leq m$.*

Definition 10. $v_1, \ldots v_n$ is a **basis** of V if these vectors are linearly independent, and they span V.

Theorem 8. *If $v_1, \ldots, v_n$ and $w_1, \ldots, w_n$ are two bases of V, then $n = m$. In this case, we define this number $n = \dim_F V$.*

Proof. By the counting theorem above, $n \leq m$ and $m \leq n$, so $n = m$. □

The main example that will be important for us to consider is as follows. Let $f(x) \in F[x]$ with $\deg f(x) = n$ and $f \notin F$. Consider $F \leq F[x]/(f(x))$. In fact, we know that every element in the coset ring is uniquely of the form $g(x) + (f(x))$ where g is zero or $\deg g(x) < n$. This says that the cosets $1 + (f(x)), x + (f(x)), \ldots, x^{n-1} + (f(x))$ are an F-basis for $F[x]/(f(x))$. This implies that $\dim_F F[x]/(f(x)) = n$.

Corollary 9. *Let $F \leq E$ and α algebraic over F. Let $\deg_F \alpha = \deg \operatorname{irr}(\alpha, F, x)$. Then $F(\alpha)$ is a finite dimensional F-vector space: $\dim_F F(\alpha) = \deg_F \alpha = \deg \operatorname{irr}(\alpha, F, x)$.*

We have already seen a few examples: $\dim_{\mathbb{R}} \mathbb{C} = 2$ and $\dim_{\mathbb{Q}} \mathbb{Q}(\sqrt[3]{2}) = 3$. Let's recall some more important linear algebra facts.

Theorem 10. *Let V be a finite-dimensional F-vector space. Then,*

1. *Any set $v_1, \ldots, v_n \in V$ that spans V contains a subset which is a basis.*

2. *Any set $v_1, \ldots, v_n \in V$ which is linearly independent can be completed to a basis.*

3. *If W is a vector subspace of V, then $\dim W \leq \dim V$. If $\dim W = \dim V$, then $W = V$.*

4. *If V_1 and V_2 are two finite-dimensional vector spaces with bases given by v_n, w_m, and $f : V_1 \to V_2$ is a linear map, then*

$$f(v_i) = \sum_{j=1}^{m} a_{ji} w_j$$

where $A = (a_{ij})$ is an $m \times n$ matrix. f determines and is determined by A.

Remark. Suppose F is a finite field, with q elements. Suppose V is a finite-dimensional F-vector space of dimension n, i.e. there exists a basis $v_1, \ldots, v_n$ of V where every vector can be written uniquely in terms of this basis. It should be clear that the number of elements in V is q^n. In fact, any finite dimensional vector space of dimension n is isomorphic to F^n.

In particular, if F itself is finite, its characteristic must be a prime p, and $\mathbb{F}_p \leq F$. Thus, any finite field is an $\mathbb{F}_p$-vector space. Since F is also finite-dimensional, the number of elements in F is simply p^k.

Definition 11. Let E be an extension field of F. Then E is a **finite extension** of F if E is a finite-dimensional F-vector space. In this case, we define $\dim_F E = [E : F]$, the **degree of E over F**.

Example 5. $\mathbb{C}$ is a finite extension of $\mathbb{R}$, as $[\mathbb{C} : \mathbb{R}] = 2$. Similarly, $\mathbb{Q}(\sqrt{2})$ is a finite extension of $\mathbb{Q}$ with $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$. Again, $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$. However, $\mathbb{R}$ is not a finite extension of $\mathbb{Q}$, as $\mathbb{R}$ is an infinite-dimensional $\mathbb{Q}$-vector space.

Furthermore, consider F any field, a subring of $F[x]$. Then, $F \leq F(x)$, but $F(x)$ is not a finite extension of F since $F \subset F[x] \subset F(x)$.

Theorem 11. *Let $E = F(\alpha)$ be a simple extension of F. Then, E is a finite extension of F if and only if α is algebraic over F. In this case, $[E : F] = \deg_F \alpha = \deg \operatorname{irr}(\alpha, F, x)$.*

Note that if α is transcendental, $F[\alpha] \cong F[x]$ (because there is no kernel), which is not a field, but $F[\alpha] \subset F(\alpha) \cong F(x)$. In particular, $F(\alpha)$ (quotients) is the smallest subfield of E containing F and α. Since $F(\alpha) \cong F(x)$, this is not a finite extension of F.

If α is algebraic over F, $\ker \mathrm{ev}_\alpha$ is a maximal ideal and thus $F[\alpha] = \mathrm{Im}\ \mathrm{ev}_\alpha$ which is already a field, that we denote by $F[\alpha] = F(\alpha)$. Again, it is the smallest subfield of E containing F and α. In this case $F(\alpha) \cong F[x]/\ker \mathrm{ev}_\alpha \cong F[x]/(\mathrm{irr}(\alpha, F, x))$. This forms a finite-dimensional F-vector space with basis $1, \alpha, \ldots, \alpha^{d-1}$, where $d = \deg \mathrm{irr}(\alpha, F, x)$. In other words, every element of $F(\alpha)$ is uniquely written $c_0 + c_1\alpha + \ldots + c_{d-1}\alpha^{d-1}$, $c_i \in F$.

One piece of notation: if $F \leq E$, we take $\alpha_1, \ldots, \alpha_n \in E$. We can define $F(\alpha_1, \ldots, \alpha_n)$ as the smallest subfield of E containing F and $\alpha_1, \ldots, \alpha_n$. It is easy to check that there is an inductive definition $F(\alpha_1, \ldots, \alpha_n) = F(\alpha_1, \ldots \alpha_{n-1})(\alpha_n)$. For example, we could look at $\mathbb{Q}(\sqrt{2}, \sqrt{3})$. This is, in fact, the same thing is as $\mathbb{Q}(\sqrt{2})(\sqrt{3})$. If $\alpha = \sqrt{2} + \sqrt{3}$, $\mathbb{Q}(\alpha) \leq \mathbb{Q}(\sqrt{2}, \sqrt{3})$. On the other hand, by an explicit check, one can show that $\mathbb{Q}(\alpha) = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. It is enough to show that $\sqrt{2} \in \mathbb{Q}(\alpha)$ and $\sqrt{3} \in \mathbb{Q}(\alpha)$.

Lemma 12. *Let E be a finite extension of F and let V be a finite dimensional E-vector space. Then V is a finite-dimensional F-vector space and $\dim_F V = [E : F] \dim_E V$.*

Proof. There exists a basis for V as a vector space over E, say $w_1 \cdots w_n$, with $n = \dim_E V$. Additionally, there exists a basis for E as a vector space over F, say $\alpha_1 \cdots \alpha_m$, where $m = \dim_F E$. Consider the collection $\alpha_i w_j$, which has mn elements. We claim that $\alpha_i w_j$ is an F-basis for V. Once we prove the claim, we see that V is a finite-dimensional F-vector space, and that $\dim_F V = mn = [E : F] \dim_E V$.

Let us first show that $\alpha_i w_j$ spans V over F. We know that w_j span V over E, i.e. for all $v \in V$, there exist $\beta_j \in E$ such that $v = \sum_{j=1}^n \beta_j w_j$. But we also know that α_i span E over F, i.e. $\beta_j = \sum b_{ij}\alpha_i$, with $b_{ij} \in F$. Putting these together, we find

$$v = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} b_{ij}\alpha_i \right) w_j = \sum_{i,j=1}^{m,n} b_{ij}(\alpha_i w_j),$$

as desired.

Next we show that $\alpha_i w_j$ are linearly independent. Suppose $\sum_{ij} b_{ij}\alpha_i w_j = 0$ – we wish to show that $b_{ij} = 0$ for all i, j. We simply rewrite the sum as

$$\sum_{j=1}^{n} \left(\sum_{i=1}^{m} b_{ij}\alpha_i \right) w_j = 0,$$

but by linear independence of w_j, the inner sums must be zero, but if the inner sum is zero, by linear independence of α_i, $b_{ij} = 0$ for all i, j, and we are done. □

Theorem 13. *Suppose $F \leq E \leq K$, all fields. Then K is a finite extension of F if and only if K is a finite extension of E and E is a finite extension of F. Furthermore, $[K : F] = [K : E][E : F]$.*

Proof. Let us start with the forwards direction. If K is a finite extension of F, then, K is a finite dimensional F-vector space. In particular, E is a vector subspace of K. But this implies that $\dim_F E$ is also finite, which in turn implies that E is a finite extension of F. Also, K is spanned over F by a finite number of elements $\alpha_1, \ldots, \alpha_n \in K$. But notice that the span of the α_i with E-coefficients includes the span with F-coefficients. This implies that α_i span K over E. In other words, for all $\alpha \in K$, $\alpha = \sum f_i \alpha_i$ with $f_i \in F$. Since $F \subset E$, α is in the E-span of $\alpha_1 \ldots \alpha_n$, which implies that K is spanned over E by $\alpha_1, \ldots, \alpha_n$.

The other direction is trivial via the above lemma; take $V = K$ – then $\dim_F K = [K : F]$, and we are done. □

Corollary 14. *Both $[K : E]$ and $[E : F]$ divide $[K : F]$.*

Proof. Obvious. □

Example 6. Is $\sqrt{2} \in \mathbb{Q}(\sqrt[3]{2})$? No. Why? Because if there were, we'd have $\mathbb{Q} \leq \mathbb{Q}(\sqrt{2}) \leq \mathbb{Q}(\sqrt[3]{2})$. But $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$ and $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$, but two does not divide three, so we are done by contradiction.

Note that the above proof shows that if $\alpha_1 \cdots \alpha_m$ is a basis for E over F and $\beta_1 \cdots \beta_n$ is a basis for K over E, $\alpha_i \beta_j$ is a basis for K over F.

Example 7. $\mathbb{Q} \leq \mathbb{Q}(\sqrt{2}) \leq \mathbb{Q}(\sqrt{2}, \sqrt{3})$ and $\sqrt{3} \neq \mathbb{Q}(\sqrt{2})$. We have $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}] = [\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}(\sqrt{2})][\mathbb{Q}(\sqrt{2}) : \mathbb{Q}]$ where the second term is 2. But we have $\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(\sqrt{2})(\sqrt{3})$ and $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}(\sqrt{2})]$ is the degree of $\mathrm{irr}(\sqrt{3}, \mathbb{Q}(\sqrt{3}), x)$. What is this? It's $x^2 - 3$, but this is irreducible in $\mathbb{Q}(\sqrt{2})[x]$ since it has no root in $\mathbb{Q}(\sqrt{2})$. Thus this degree is 2, and $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}] = 4$.

Note $1, \sqrt{2}$ is a basis for $\mathbb{Q}(\sqrt{2})$ over $\mathbb{Q}$ and the basis for $\mathbb{Q}(\sqrt{2})(\sqrt{3})$ over $\mathbb{Q}(\sqrt{2})$ is $1, \sqrt{3}$ (because in general, $[F(\alpha) : F] = 1, \alpha, \cdots, \alpha^{d-1}$). Then the basis for $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ over $\mathbb{Q}$ is $1, \sqrt{2}, \sqrt{3}, \sqrt{6}$.

But note that we know if $\alpha = \sqrt{2} + \sqrt{3}$, $\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(\alpha)$. We know that α is a root of $x^4 - 10x^2 + 1$ and thus $\mathrm{irr}(\alpha, \mathbb{Q}, x)$ divides this polynomial. But $\deg \mathrm{irr}(\alpha, \mathbb{Q}, x) = [\mathbb{Q}(\alpha) : \mathbb{Q}] = [\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}] = 4$. Consequently, $x^4 - 10x^2 + 1$ is irreducible in $\mathbb{Q}[x]$ and it is the $\mathrm{irr}(\alpha, \mathbb{Q}, x)$. Then, $1, \alpha, \alpha^2, \alpha^3$ is another $\mathbb{Q}$-basis for $\mathbb{Q}(\sqrt{2}, \sqrt{3})$.

Definition 12. Let $F \leq E$. We say that E is an **algebraic extension** of F if for all $\alpha \in E$, α is algebraic over F.

Example 8. $\mathbb{R}$ is not an algebraic extension of $\mathbb{Q}$.

Theorem 15. *If E is a finite extension of F, then E is an algebraic extension of F.*

Proof. If $\alpha \in E$ is not algebraic over F, then we have $F \leq F(\alpha) \leq E$. But we've seen that $F(\alpha)$ is not finite dimensional as an F-vector space. But this contradicts that E is finite-dimensional. Thus every α is algebraic. □

Theorem 16. *Suppose $F \leq E$ and $\alpha, \beta \in E$ are both algebraic over F. Then, $a \pm \beta, \alpha\beta, \alpha/\beta$ are also algebraic over F (where $\beta \neq 0$ for division).*

Proof. Note that $F \leq F(\alpha) \leq F(\alpha, \beta) = F(\alpha)(\beta)$. We know that α is algebraic over F, which implies that $F(\alpha)$ is a finite extension of F. If β is algebraic over F, it is clearly algebraic over $F(\alpha)$. This implies that $F(\alpha)(\beta)$ is a finite extension of $F(\alpha)$. Thus, $F(\alpha, \beta)$ is a finite extension of F. Consequently, every element of $F(\alpha, \beta)$ is algebraic over F. By closure of the field operations, then, and the previous theorem, we are done. □

Corollary 17. *If $F \leq E$ then $\{a \in E : \alpha$ is algebraic over $F\}$ is a subfield of E called the **algebraic closure** of F in E.*

Example 9. The algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$ is $\mathbb{Q}^{\text{alg}}$, which is an algebraic extension of $\mathbb{Q}$, but is not finite. What is the algebraic closure of F in $F(t)$? It is simply F.

(March 11, 2013)
Let's do a quick recap – we've defined 3 types of extensions:

1. E is a simple extension of F if $E = F(\alpha)$. If α is algebraic over F, then $F(\alpha) = F[\alpha]$.

2. E is a finite extension of F if $\dim_F E = [E : F] < \infty$. E is a finite-dimensional F-vector space.

3. E is an algebraic extension of F if for all $\alpha \in E$, α is algebraic over F.

What is the relationship between these three concepts? First of all, we know that if α is algebraic over F, then the simple extension $F(\alpha)$ is a finite extension of F, and in fact, $[F(\alpha) : F] = \deg_F \alpha = \deg \operatorname{irr}(\alpha, F, x)$. Second, we know that if E is a finite extension of F, then it is an algebraic extension of F. What about the converses of these two statements? First, note

that there are algebraic extensions that are not finite. Take, for example, $\mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5}, \ldots)$. This is clearly algebraic but not finite (we've joined all prime numbers). Second, finite doesn't always have to be simple, but in fact, it is *almost always*. We will talk about this later, but recall the example of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ and $\mathbb{Q}(\sqrt{2} + \sqrt{3})$. Of course, simple extensions can sometimes be harder to work with, so this is not always to be taken advantage of.

What else do we know? If $F \leq E$, with $\alpha, \beta \in E$ algebraic over F, then $\alpha \pm \beta, \alpha\beta, \alpha/\beta$ are all algebraic over F as well (if the quotient is defined). This is what motivated us to define above the algebraic closure of F.

Lemma 18. *Let E be an extension of F. Then E is a finite extension of F if and only if there exist $\alpha_1, \cdots, \alpha_n \in E$ with α_i algebraic over F for all i such that $E = F(\alpha_1, \cdots, \alpha_n)$.*

Proof. Let us take the backwards direction. If $E = F(\alpha_1, \cdots, \alpha_n)$, where α_i are algebraic over F, we simply consider the sequence of extensions $F \leq F(\alpha_1) \leq F(\alpha_1, \alpha_2) \leq \cdots \leq F(\alpha_1, \cdots, \alpha_n) = E$. We claim by induction that $F(\alpha_1, \cdots, \alpha_i)$ is a finite extension of F for $i = 1, \cdots, n$. For the case $i = 1$, we have a simple extension, and by hypothesis, α_1 is algebraic over F, we know $F(\alpha)$ is a finite extension of F. Now assume that $F \leq F(\alpha_1, \cdots, \alpha_i) \leq F(\alpha_1, \cdots, \alpha_{i+1})$. We wish to show that this last extension is finite over F. We know that $F(\alpha_1, \cdots, \alpha_{i+1}) = F(\alpha_1, \cdots, \alpha_i)(\alpha_{i+1})$, so this is a simple extension. In particular, α_{i+1} is algebraic over F, and hence algebraic over any bigger field. Then we clearly have a finite extension of $F(\alpha_1, \cdots, \alpha_i)$, and by the theorem proved last time, we have a finite extension of F.

Now we show the forwards direction. Suppose E be a finite extension of F. Then we know that E is an algebraic extension of F, i.e. for all $\alpha \in E$, α is algebraic over F. Now we argue by complete induction on $[E : F]$. If $[E : F] = 1$, $E = F$ and we are done. Suppose that the result is true for all finite extensions of fields of degree less than some $d > 1$. Given E an extension of F with $[E : F] = d$. Clearly $E \neq F$ and so there exists some $\alpha_1 \in E$, $\alpha_1 \notin F$, which is algebraic over F (E is finite). Then, $F \leq F(\alpha_1) \leq E$ and $[E : F] = [E : F(\alpha_1)][F(\alpha_1) : F]$. The second term is bigger than 1, and so $[E : F(\alpha_1)] < d$. By induction, then, there exist $\alpha_2, \cdots, \alpha_n \in E$ that are algebraic over $F(\alpha_1)$ such that $E = F(\alpha_1, \alpha_2, \cdots, \alpha_n)$. In fact, α_i are algebraic over F, again since E is a finite extension, and we are done. □

Thus, we have shown that every finite extension can be broken up into a sequence of simple extensions.

Example 10. We looked at the case $\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \sqrt{2}\sqrt{3}$ by direct computation because $\deg \operatorname{irr}(\sqrt{3}, \mathbb{Q}(\sqrt{2}), x) = 2$. In general, however, such a

statement about polynomials is not always obvious.

Lemma 19. *Take $F \leq E \leq K$. Suppose that E is algebraic over F. Suppose $\alpha \in K$. Then α is algebraic over F if and only if α is algebraic over E.*

Proof. The forward direction is trivial, as if α is the root of $f(x) \in F[x]$ non-zero, we just view that as a member of $E[x]$. The difficult direction is the backwards one. Suppose α is algebraic over E. This means that there exists an $f(x) = x^d + a_{d-1}x^{d-1} + \cdots + a_0 \in E[x]$ such that $f(\alpha) = 0$. Consider $F \leq F(a_0, \cdots, a_{d-1}) \leq F(a_0, \cdots, a_{d-1}, \alpha)$. By the above lemma, $F(a_0, \cdots, a_{d-1})$ is a finite extension of F. Additionally, α is algebraic over $F(a_0, \cdots, a_{d-1})$, and so $F(a_0, \cdots, a_{d-1}, \alpha)$ is a finite extension of F. But this implies that it is an algebraic extension of F, and since $\alpha \in F(a_0, \cdots, a_{d-1}, \alpha)$, α must be algebraic over F. □

Corollary 20. *Suppose $F \leq E \leq K$. Then K is algebraic over F if and only if K is algebraic over E and E is algebraic over F.*

Proof. If K is algebraic over F, then every element of K is algebraic over F. In particular, given $\alpha \in E \leq K$, α is algebraic over F. And given $\alpha \in K$, α is algebraic over F, then α is algebraic over E.

Now assume that K is algebraic over E and E is algebraic over F. Then, we are done by the above lemma. □

Definition 13. A field K is **algebraically closed** if for all $f(x) \in K[x]$ with $\deg f(x) \geq 1$, then $f(x)$ has a root in K.

Theorem 21. *For a field K, the following are equivalent:*

1. *K is algebraically closed.*
2. *Every $f(x) \in K[x]$ non-constant is a product of linear factors, i.e. $f(x)$ is irreducible in $K[x]$ if and only if $\deg f(x) = 1$*
3. *If E is an algebraic extension of K, then $E = K$.*

Proof. These are fairly straightforward, so they are left as an exercise. [Hint: the second is via induction] □

Definition 14. Let F be a field. Then an **algebraic closure** of F is an extension field K such that K is an algebraic extension of F and K is algebraically closed.

Remark. Notice that $\mathbb{C}$ is not an algebraic closure of Q, as $\mathbb{C}$ is not an algebraic extension (it contains transcendental elements over $\mathbb{Q}$).

Let's pause and look at the definitions:

1. If $F \leq E$, we've defined the algebraic closure of F in E.
2. K is algebraically closed.
3. K is an algebraic closure of F.

Theorem 22. *Let $F \leq K$ with K be algebraically closed. Then the algebraic closure of F in K, F^{alg}, is an algebraic closure of F.*

Example 11. Before we prove this proposition, note the following intuitive example. We have some relationship of the form $\mathbb{Q} \leq \mathbb{Q}^{\text{alg}} \leq \mathbb{C}$, and the proposition in this case states says that $\mathbb{Q}^{\text{alg}}$ is an algebraic closure of $\mathbb{Q}$. Note, however, that $\mathbb{C}$ is not an algebraic closure of $\mathbb{Q}$, as $\mathbb{C}$ has transcendental numbers, i.e. $i \in \mathbb{Q}^{\text{alg}}$ but $\pi i \in \mathbb{C} \notin \mathbb{Q}^{\text{alg}}$.

Proof. We know, by definition, that F^{alg} is an algebraic extension of F, because it is the set of all things in K that are algebraic over F. We wish to show that F^{alg} is algebraically closed. In other words, given some $f(x) \in F^{\text{alg}}[x]$ non-constant, we must show that there exists an α, a root of $f(x)$ in F^{alg}. We know that $f(x) \in F^{\text{alg}} \leq K[x]$. Since K is algebraically closed, there exists an α in K such that $f(\alpha) = 0$. We want to show that $\alpha \in F^{\text{alg}}$. But this is equivalent to saying that α is algebraic over F. But we know that α is algebraic over F^{alg} and that F^{alg} is algebraic over F. By the lemma we proved earlier, then, we have that K must be algebraic over F, and thus α must be in F^{alg}. □

Remark. It is a fact that for any field F, there exists an algebraic closure of F and any two of these algebraic closures of F are isomorphic.

Remark. Suppose one can construct a length $\alpha \in \mathbb{R}$. It can be shown that α can be constructed via compass/straightedge if and only if $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 2^k$. Then, one can show that trisecting a 60 degree angle allows one to construct $\alpha = \cos 20$, but $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 3$, as one can show, so an angle cannot be trisected. Likewise, one can show that it is impossible to double the cube, i.e. construct the cube root of two. Finally, it can be show that it is impossible to construct a square whose side length is $\sqrt{\pi}$, which is, of course, impossible, as π and $\sqrt{\pi}$ are transcendental.

4 Multiple roots and derivatives

Suppose F is a field and $f(x) \in F[x]$, with $\deg f(x) \geq 1$. α is a root of $f(x)$ if and only if $(x - a)|f(x)$. In general, given any $f(x) \in F[x]$ non-constant

and any $\alpha \in F$, there exists an integer $m \geq 0$ such that $(x-\alpha)^m | f(x)$ but $(x-\alpha)^{m+1}$ does not divide $f(x)$. Notice that $m = 0$ if and only if α is not a root. This number m is called the **multiplicity** of the root α.

Remark. We might have $F \leq E$ and we might be looking at $\alpha \in E$ – we talk about divisibility in $E[x]$, as $x - \alpha$ is not necessarily in the smaller field F.

If the multiplicity of α is 1, we say that α is a simple root, and if the multiplicity of α is greater than 1, we say that α is a **multiple** or **repeated** root. To see if α is a repeated root, we write $f(x) = (x-\alpha)g(x)$ if and only if $f(\alpha) = 0$. Notice also that α is a repeated root if and only if $f(\alpha) = f'(\alpha) = 0$. But what is a derivative in any field? Forget about limits – we define derivatives purely formally.

Definition 15. Given $f(x) = \sum_{i=0}^{n} a_i x^i \in F[x]$, we define the **derivative** of f in $F[x]$ purely formally as

$$Df(x) = \sum_{i=1}^{n} i a_i x^{i-1}.$$

Note that $\deg Df(x) \leq \deg f(x) - 1$. In particular, we may define the derivative generally as a function $D : F[x] \to F[x]$. In the context of an extension $F \leq E$ we can define compatibly $D : E[x] \to E[x]$ as the derivatve does not particularly care about where the coefficients live. D is determined by $D(x^i) = ix^{i-1}$ and is F-linear. It should be obvious that D is not a ring homomorphism, as it instead follows the product rule:

$$D(x^a x^b) = D(x^{a+b}) = (a+b)x^{a+b-1} = ax^{a-1}x^b + x^a b x^{b-1} = (Dx^a)x^b + x^a Dx^b$$

and by linearity, this holds for all polynomials.

There is a corollary of the product rule that states

$$D((f(x))^n) = nf(x)^{n-1} \cdot Df(x),$$

proved via induction.

We can, of course, inquire into the kernel of D. It is not just constant functions, however, as if F has characteristic p, $Dx^p = px^{p-1} = 0$! In fact, it is easy to see that if the characteristic of F is zero, the $\ker D = F \leq F[x]$, but if the characteristic is p, then $\ker D = F[x^p] = \operatorname{Im} \text{Frob} \leq F[x]$, i.e. the polynomials in x^p.

Now what is the connection of the derivative to multiple roots? Let us work in $F \leq E$ with $\alpha \in E$. Let m be the multiplicity of α in $f(x)$, i.e.

$(x-\alpha)^m$ divides $f(x)$ but higher powers do not. In other words, we can write $f(x) = (x-\alpha)^m g(x)$ with $g(\alpha) \neq 0$. If $m = 0$, $f(\alpha) \neq 0$. Let's assume that $m \geq 1$. Then we write $f(x) = (x-\alpha)^m g(x)$. Taking the derivative, we find

$$\begin{aligned} Df(x) &= D\left((x-a)^m\right) g(x) + (x-\alpha)^m Dg(x) \\ &= m(x-a)^{m-1} g(x) + (x-\alpha)^m Dg(x) \end{aligned}$$

Now, if $m = 1$, $Df(x) = f(x) + (x-a)Dg(x)$ and $Df(\alpha) = g(\alpha) \neq 0$. If $m \geq 2$, on the other hand, we have at least one overall factor of $x-\alpha$. Here, unlike before, $Df(\alpha) = f(\alpha) = 0$.

Theorem 23. *Any $\alpha \in E$ is a multiple root of $f(x)$ if and only if $f(\alpha) = Df(\alpha) = 0$.*

This is not exactly what we want, however, as we want a condition that $f(x)$ has a multiple root in some extension field without knowing what the root or the extension field are.

Lemma 24. *Let $F \leq E$ and $f(x), g(x) \in F[x]$ not both 0, $f(x)$ not constant.*

1. *$f(x)|g(x)$ in $F[x]$ if and only if $f(x)|g(x)$ in $E[x]$.*
2. *$d(x) \in F[x]$ is a gcd of f, g in $F[x]$ if and only if it is a gcd in $E[x]$.*
3. *f, g are relatively prime in $F[x]$ if and only if they are relatively prime in $E[x]$.*

Proof. 1. The forward direction is obvious. Suppose $g(x) = f(x)h(x)$ with $h(x) \in E[x]$. We wish to show that $h(x) \in F[x]$. We can apply long division with remainder in $F[x]$: $g(x) = q(x)f(x) + r(x)$ with $q, r \in F[x]$ and either $r = 0$ or $\deg r < \deg f$. We now have two different expression for g, but we know that in $E[x]$ long division with remainder is unique! Thus, they must be the same. This implies that $h(x) = q(x)$ and $r(x) = 0$, i.e. that $h(x) \in F[x]$.

2. Say $d(x)$ is a gcd of f, g in $F[x]$. In particular, we know that $d|f, d|g$ and $d = af + bg$. To show that d is a gcd of f, g in $E[x]$, we need to show that anything that divides f, g divides d. Suppose $e(x)$ in $E[x]$ divides f, g. Then $e|af + bg = d$ so $e|d$ and d is a gcd in $E[x]$. Conversely suppose $d(x) \in F[x]$ is a gcd of f, g in $E[x]$ so $d|f, d|g$ in $E[x]$. But by the first statement of this lemma, we know that $d|f, d|g$ in $F[x]$. Then if $e \in F[x]$ with $e|f, e|g$, then $e|f, e|g$ in $E[x]$ implies $e|d$ in $E[x]$ and thus $e|d$ in $F[x]$. This shows that e is a gcd in $F[x]$.

3. f, g are relatively prime in $F[x]$ if and only if 1 is a gcd of f, g in $F[x]$, which by statement 2 of the lemma, implies that 1 is a gcd of f, g in $E[x]$. Then f, g are relatively prime in $E[x]$.

□

Corollary 25. *Let $f(x) \in F[x]$ non-constant. Then there exists an extension field E of F and an $\alpha \in E$ which is a multiple root of $f(x)$ if and only if $f(x)$ and $Df(x)$ are not relatively prime in $F[x]$.*

Proof. Suppose that $f(x)$ has a multiple root α in some extension E. This implies that $x - \alpha | f(x)$ and $x - \alpha | Df(x)$. This implies that f, Df are not relatively prime in $E[x]$. But if they are not relatively prime in $E[x]$, they are not relatively prime in $F[x]$.

If $f(x), Df(x)$ are not relatively prime in $F[x]$, then there exists an irreducible $p(x)$ in $F[x]$ such that $p(x)|f(x)$ and $p(x)|Df(x)$. But we know that there exists an extension field E of F and a root α of $p(x) \in E$. In E, then, α is a root of $f(x)$ and $Df(x)$, which implies that α is a multiple root. □

Corollary 26. *Let $f(x)$ be an irreducible polynomial in $F[x]$. $f(x)$ has a multiple root in some extension field E of F if (and only if) $Df(x) = 0$.*

Proof. If $f(x)$ has a multiple root in some extension field, then $f(x), Df(x)$ are not relatively prime, by the above corollary. But we know that $f(x)$, by hypothesis, is irreducible, and $Df(x)$, if non-zero, has smaller degree (not necessarily $n - 1$ in positive characteristic fields). Since $f(x)$ is irreducible, we know that either the gcd of $f(x), Df(x)$ is one or $f(x)|Df(x)$. However, the first is not true due to the multiple root, and the second is not true due to the degrees of the polynomials. Thus we reach a contradiction, and $Df(x) = 0$. □

Corollary 27. *If the characteristic of F is zero, an irreducible $f(x) \in F[x]$ never has a multiple root in an extension field.*

Proof. $Df(x) \neq 0$. □

In fact, this does happen if the characteristic of F is p, but not if F is finite.

Example 12. Let $F = \mathbb{F}_p(t)$, an infinite field of rational functions with characteristic p. Now consider $f(x) = x^p - t$. A zero α of $f(x)$ is equal to

a pth root of t. Thus $f(x)$ has no root in $\mathbb{F}_p(t)$ just because the polynomial would have to have power $1/p$. Note that one can check that $f(x)$ is irreducible.

Let α be a root of $f(x)$ in some extension field: $\alpha^p = t$. In $E[x]$, $f(x) = x^p - t = x^p - \alpha^p = (x - \alpha)^p$. Thus, α is a root of multiplicity p. Notice that in this case $D(x^p - t) = px^{p-1} = 0$.

When we come back from break, we will move on to using this to classify finite fields.

5 Classification of finite fields

(March 25, 2013)

Although we will not work much with finite fields in this course, they are often useful in coding theory and number theory.

Every finite field $\mathbb{F}$ has prime characteristic. Consequently, $\mathbb{F}_p \leq \mathbb{F}$. Furthermore, if $\dim_{\mathbb{F}_p} \mathbb{F} = n$, then $\mathbb{F}$ is a finite-dimensional $\mathbb{F}_p$-vector space and the number of elements in $\mathbb{F}$ is $q = p^n$. We know that $\mathbb{F}^* = \langle \beta \rangle$ is cyclic. Thus, $\mathbb{F} = \mathbb{F}_p(\beta)$ and $\mathbb{F}$ is a simple extension of $\mathbb{F}_p$ (or of any subfield $\mathbb{F}'$).

If the number of elements in $\mathbb{F}$ is q, the number of elements in $\mathbb{F}^*$ is clearly $q - 1$. If $\alpha \in F^*$ then $\alpha^{q-1} = 1$, by Lagrange's theorem. Multiplying by α, we can write that $\alpha^q = \alpha$. This statement now holds for all α (including zero). This is a generalization of Fermat's little theorem. Consequently, every $\alpha \in \mathbb{F}$ is a root of $x^q - z \in \mathbb{F}_p[x]$.

Recall the Frobenius homomorphism, $\sigma_p : \mathbb{F} \to \mathbb{F}$ given by $\sigma_p(\alpha) = \alpha^p$ with the nice properties that $(\alpha + \beta)^p = \alpha^p + \beta^p$ and $(\alpha\beta)^p = \alpha^p\beta^p$. The kernel $\ker \sigma_p = \{\alpha : \alpha^p = 0\} = \{0\}$, as we are in an integral domain. Thus, σ_p is injective. But because $\mathbb{F}$ is finite, σ_p must be surjective, as well, and thus in the case of finite fields, the Frobenius homomorphism is actually an isomorphism. This implies that every element of $\mathbb{F}$ is a pth power (we say $\mathbb{F}$ is **perfect**).

We can take

$$(\sigma_p)^2(\alpha) = \sigma_p(\sigma_p(\alpha)) = (\alpha^p)^p = \alpha^{p^2},$$

and, in general,

$$\sigma_p^k = \sigma_{p^k}.$$

In particular, $\sigma_q(\alpha) = \alpha^q = \alpha$ and so $\sigma_q = \mathrm{Id}$. We should be careful to note, however, that although every element of $\mathbb{F}$ satisfies $\alpha^q = \alpha$, $x^q - x$ is *not* the zero polynomial in $\mathbb{F}[x]$ or $\mathbb{F}_p[x]$.

Theorem 28 (Classification of finite fields)**.** *Let p be a prime throughout.*

(i) If $q = p^n$, $n \in \mathbb{N}$, then there exists a finite field $\mathbb{F}$ with q elements.

(ii) If $\mathbb{F}_1$ and $\mathbb{F}_2$ are two finite fields, $\mathbb{F}_1 \cong \mathbb{F}_2$ if an only if they have the same number of elements.

(iii) Given $q = p^n$, $q' = p^m$, and $\mathbb{F}$ a field with q elements and $\mathbb{F}'$ a field with q' elements, then $\mathbb{F}'$ is isomorphic to a subfield of $\mathbb{F}$ if and only if $q = (q')^d$ for some $d \in \mathbb{N}$, i.e. m divides n.

Proof. Let us begin with (i). Consider $x^q - x \in \mathbb{F}_p[x]$. We know that there exists a field E, $\mathbb{F}_p \leq E$, such that in $E[x]$, $x^q - x = \prod_i^q (x - \alpha_i)$. Then $\alpha \in E$ satisfies $\alpha^p = \alpha$ if and only if $\alpha = \alpha_i$ for some i. We claim that $x^q - x$ has no multiple root, and that all α_i are distinct. We have $D(x^q - x) = qx^{q-1} - 1 = -1$, a unit. By definition, then, $x^q - x, D(x^q - x)$ are relatively prime, i.e. there are no multiple roots.

We now define

$$\mathbb{F}_q \subset E = \{\alpha_1, \ldots, \alpha_q\} = \{\alpha \in E : \alpha^q = \alpha\} = \{\alpha \in E : \sigma_q(\alpha) = \alpha\}.$$

This is called a **fixed field** of σ_q. We know that the number of elements in $\mathbb{F}_q$ is q. We claim that $\mathbb{F}_q$ is a subfield of E. Say $\alpha, \beta \in \mathbb{F}_q$. They satisfy $\alpha^q = \alpha, \beta^q = \beta$. Note that

$$(\alpha \pm \beta)^q = \sigma_q(\alpha \pm \beta) = \sigma_q(\alpha) \pm \sigma_q(\beta) = \alpha \pm \beta$$

and similarly $(\alpha\beta)^q = \alpha\beta$ and $(\alpha/\beta)^q = \alpha/\beta$ (for $\beta \neq 0$). Thus we have a field, since these operations are closed.

Next, we wish to prove the forward implication in (iii). We wish to show that given $\mathbb{F}, \mathbb{F}'$ with q, q' elements respectively, if $\mathbb{F}'$ is isomorphic to a subfield of $\mathbb{F}$, then $q = (q')^d$ for some d. We can identify $\mathbb{F}'$ with a subfield of $\mathbb{F}$. This implies that $\mathbb{F}$ is an $\mathbb{F}'$-vector space that is finite-dimensional. Let $d = \dim_{\mathbb{F}'} \mathbb{F}$. We then have d elements in a basis, so the number of elements of $\mathbb{F}$ is equal to the number of elements of $\mathbb{F}'$ to the dth power, i.e. $q = (q')^d$, and we are done.

Next we show the backwards implication for (iii). Let us show this first for the field we constructed in the proof for (i), i.e. $\mathbb{F} = \mathbb{F}_q$, the set of roots of $x^q - x$ in some E. $\mathbb{F}'$ is any field whose order is q' and $q = (q')^d$. We wish to show that $\mathbb{F}'$ is isomorphic to some subfield of $\mathbb{F}_q$. We know that $\mathbb{F}' = \mathbb{F}_p(\beta)$, where $\langle\beta\rangle = (\mathbb{F}')^*$. We have the $\operatorname{irr}(\beta, \mathbb{F}_p, x) \in \mathbb{F}_p[x]$. We claim that $\beta^q = \beta$. We know that $\beta^{q'} = \beta$, i.e. $\sigma_{q'}(\beta) = \beta$ so $\sigma_{q'}^d(\beta) = \beta$ and thus

$\sigma_q(\beta) = b$. Consequently, β is a root of $x^p - x \in \mathbb{F}_p[x]$. This implies that $\text{irr}(\beta, \mathbb{F}_p, x)$ divides $x^q - x$. We can write in $\mathbb{F}_p[x]$

$$x^q - x = \text{irr}(\beta, \mathbb{F}_p, x) \cdot p(x),$$

for some polynomial p. But we know that in $\mathbb{F}_q[x]$ that $x^q - x$ factors into a product of linear polynomials. We thus have two different factorizations of $x^q - x$. Thus, by unique factorization, it must be that for some i, $x - \alpha_i$ divides $\text{irr}(\beta, \mathbb{F}_p, x)$ in $\mathbb{F}_q[x]$. Thus there exists an i such that α_i is a root of $\text{irr}(\beta, \mathbb{F}_p, x)$. This implies that $\text{irr}(\alpha_i, \mathbb{F}_p, x) = \text{irr}(\beta, \mathbb{F}_p, x)$. But recall that $\mathbb{F}_p(\beta) \cong \mathbb{F}_p[x]/(\text{irr}(\beta, \mathbb{F}_p[x], x)))$, where the isomorphism is given by ev_β. There is also a homomorphism ev_{α_i} from this quotient to $\mathbb{F}_q$. This yields an injective homomorphism $\text{ev}_{\alpha_i} \circ (\text{ev}_\beta)^{-1} : \mathbb{F}' \to \mathbb{F}_q$ (because homomorphisms between fields are injective). The image of this homomorphisms is a subfield of $\mathbb{F}_q$, automatically isomorphic to $\mathbb{F}'$, and we are done.

Let's prove (ii). Suppose $\mathbb{F}_1$ and $\mathbb{F}_2$ are two finite fields with the same number of elements, q. By what we've shown so far in (iii), since $q = q^1$ there exists an injective homomorphism from $\mathbb{F}_1 \to \mathbb{F}_q$. Since $\mathbb{F}_1, \mathbb{F}_q$ both that the same number of elements, the injection is also an isomorphism, and thus $\mathbb{F}_1 \cong \mathbb{F}_q$. Likewise, $\mathbb{F}_2 \cong \mathbb{F}_q$, and thus $\mathbb{F}_1 \cong \mathbb{F}_2$.

Finally, we must finish (iii): if $\mathbb{F}$ is any field with q elements and $\mathbb{F}'$ a field with q' elements with $(q')^d = q$, we've seen that $\mathbb{F}'$ is isomorphic to a subfield of $\mathbb{F}_q \cong \mathbb{F}$. But the composition of these isomorphisms is an isomorphism between $\mathbb{F}'$ and a subfield of $\mathbb{F}$. □

Note that from an isomorphism $\sigma_q : E \to E$ of fields, we constructed a fixed field $\{\alpha \in E : \sigma_q(\alpha) = \alpha\}$. Furthermore, we used $\mathbb{F}_p(\beta) \cong \mathbb{F}_p[x]/(\text{irr}(\beta, \mathbb{F}_p, x))$ to construct a homomorphism. These were, in some sense, the two main tricks of the proof – we will come back to these techniques later.

Note also that we usually call $\mathbb{F}_q$ *the* field with q elements. This field is defined by the roots of the polynomial $x^q - x$. This polynomial is not irreducible in $\mathbb{F}_p[x]$: $0, 1, \ldots, p-1$ are roots.

Definition 16. Let $N_(k)$ be the number of irreducible monic polynomials in $\mathbb{F}_p[x]$ of degree k. For example, $N_p(1) = p$.

In fact, there exists a formula, with $q = p^n$,

$$\sum_{d|n} dN_p(d) = p^n$$

6 Factorization in integral domains

Throughout this section, R will denote an integral domain. We will have two general questions. First, what about $F[x]$ can be generalized? Second, what are the criteria for when a polynomial (even with rational coefficients) is irreducible.

Definition 17. The **units** R^* are $\{u \in R : (\exists r \in R)ur = 1\}$ and given $r, s \in R$ we say that r **divides** s, $r|s$ if there exists $t \in R$ such that $s = rt$. We say $r, s \in R$ are **associates** if there exists a $u \in R^*$ such that $s = ur, r = u^{-1}s$. Note that the associate property is an equivalence relation. Two associates behave the same way with respect to factorization.

Example 13. For example, if $R = \mathbb{Z}$, then $R^* = \{\pm 1\}$. If $R = F[x]$, then $R^* = F^*$. If $R = \mathbb{Z}[i]$, then the units are $R^* = \{\pm 1, \pm i\}$, i.e. $3 + 4i, -4 + 3i, 4 - 3i, -3 - 4i$ are all associates. If $R = \mathbb{Z}[\sqrt{2}]$, $1 + \sqrt{2}$ is a unit because its inverse is $-(1 - \sqrt{2})$. In fact, $\mathbb{Z}[\sqrt{w}]^* \cong \mathbb{Z} \times (\mathbb{Z}/2\mathbb{Z})$. Finally, if $R = \mathbb{Z}[\sqrt{-2}] = \mathbb{Z}[i\sqrt{2}]$, then $R^* = \{\pm 1\}$.

Definition 18. An element $r \in R$ is **irreducible**, or **an irreducible**, if $r \neq 0, r \notin R^*$ and if $r = st$, then either s is a unit and t is an associate of r, or s is an associate of r and t is a unit.

Example 14. In $\mathbb{Z}$, irreducibles are $\pm p$, with p prime. In $F[x]$, irreducibles are irreducible polynomials.

(March 27, 2013)

Definition 19. An integral domain R is a **unique factorization domain (UFD)** if

1. for any $r \in R$, $r \neq 0$ and $r \notin R^*$, there exist irreducibles $p_1, \ldots, p_n \in R$ such that $r = p_1 \cdots p_n$.
2. if $p_1 \cdots p_n = q_1 \cdots q_n$, p_i, q_j are irreducibles, then $n = m$ and, after reordering, p_i and q_i are associates.

Note that $\mathbb{Z}$ and $F[x]$ are both UFDs.

Remark. Instead of writing $r = p_1 \cdots p_n$ we might write $r = up_1^{q_1} \cdots p_r^{q_r}$, p_i irreducible, $q_i \in \mathbb{N}$ and if $i \neq j$ p_i, p_j are not associates, and u is a unit. For example, in $\mathbb{Z}$, $-4 = (-1)2^2$.

Definition 20. An integral domain R is a **principal ideal domain** (PID) if every ideal I in R is a principal ideal.

Note that we have shown that $\mathbb{Z}$ and $F[x]$ are PIDs.

Theorem 29. *Every PID is a UFD.*

Proof. We will not do the full proof here. The first idea is to show the existence of factorization into irreducibles, i.e. if $r \in R$ non-zero and not a unit, then there exist $p_1, \ldots, p_n \in R$ irreducible such that r is the product of these p_i. This is actually true in a very general class of rings which have some finiteness property (these are called **Noetherian rings**). We'll define a special class of PIDS shortly where we can prove this directly.

Next we need to show uniqueness of this factoring. The main point here is to show that if p is irreducible and $p|rs$, then $p|r$ or $p|s$. This is easy to do using the same arguments we used for $F[x]$. We then use induction: if $p_1 \cdots p_n = q_1 \cdots q_m$, then $p_1|q_1 \cdots q_m$, and thus $p_1|q_i$ for some i. Since q_i is an irreducible, p_1, q_i are associates. We can relabel so that $i \mapsto 1$, cancel off, and continue. □

What are some examples of UFDs that are not PIDs? Well, take $F[x, y]$. The ideal (x, y) is clearly not a principal ideal. But, we will show later that $F[x, y]$ is, in fact, a UFD. Indeed, $F[x_1, \ldots, x_n]$ is in general a UFD, but not a PID if $n \geq 2$. Additionally, $\mathbb{Z}[x_1, \ldots, x_n]$ is a UFD but not a PID if $n \geq 1$. Note that all of these examples are polynomial rings – in fact, we will show that if R is a UFD, then $R[x]$ is a UFD. In general, though, most integral domains are not UFDs.

Recall that in any integral domain, $r|s$ if $s = rt$. We can also define a gcd of r, s as a $d \in R$ such that $d|r, d|s$ and if $e \in R$, and $e|r, e|s$, then $e|d$ (as long as r, s are not both zero). It is easy to see that if d_1, d_2 are 2 gcds of r and s, then d_1, d_2 are associates, because they must divide each other, and since neither is zero, if $d_1 = ud_2, d_2 = vd_1$, then $uv = 1$.

Theorem 30. *If R is a UFD with $r, s \in R$ not both zero, then a gcd of r and s exists.*

Proof. This proof is rather ugly, so let us just sketch it. We can assume r, s are not both 0 or units (as these are the easy cases), and therefore that both can be factored into irreducibles: $r = up_1^{a_1} \cdots p_r^{a_r}, s = vp_1^{b_1} \cdots v_r^{b_r}$ with u, v units and $a_i, b_i \geq 0$. Additionally, we assume that if $i \neq j$ then p_i, p_j are not associates. We must show that this is possible, but we shall not go into this here. But once we get here, we are done, as we just define $c_i = \min\ \{a_i, b_i\}$ and let $d = p_1^{c_1} \cdots p_r^{c_r}$. It should be clear that $d|r, d|s$ and that if $e|r, e|s$ one can factorize e and show that the exponents will be smaller, and thus $d|e$. □

Theorem 31. *If R is a PID and $r, s \in R$, not both zero, then the gcd of r, s exists and $d = ar + bs$ for some $a, b \in R$.*

Proof. Consider the ideal $I = (r, s) = (r) + (s) = \{ar + bs : a, b \in R\}$. Since I is an ideal, it is principal, since R is a PID. Consequently, $I = (d)$ and by definition, $d = ar + bs$ for some $a, b \in R$. But $r, s \in I$ and thus $d|r, d|s$. If $e|r, e|s$, then $e|ar + bs$ so $e|d$. □

Note that in a general UFD, gcds will *not* be linear combinations. Take, for example, $F[x, y]$, a UFD. What is the $\gcd(x, y)$? Well both x, y are irreducible and not associates so $\gcd(x, y) = 1$. Here it is obvious that we cannot write 1 as $f(x, y)x + g(x, y)y$, because if we insert $x = y = 0$, we get $0 = 1$.

Definition 21. $r, s \in R$ are **relatively prime** if $\gcd(r, s) = 1$.

Theorem 32. *If R is a UFD, and $r, s \in R$ are relatively prime, and $r|st$, then $r|t$. Hence if p is an irreducible, and $p|rs$, either $p|r$ or $p|s$.*

Proof. Straightforward but messy. □

These statements can be proved cleanly in a PID by exactly the same arguments that we used for $F[x]$. The above dichotomy about divisibility can be extended to ideals.

Theorem 33. *If R is a UFD then $p \in R$ non-zero is irreducible if and only if (p) is a prime ideal.*

Proof. This is just saying that $rs \in (p)$ if and only if $r \in (p)$ or $s \in (p)$. □

We'll see examples of integral domains R such that there exist p irreducible with $r, s \in R$ with $p|rs$ but p doesn't divide r or s and hence (p) cannot be prime.

Theorem 34. *Let R be a PID, not a field. Let I be an ideal in R. Then the following are equivalent:*

- *I is a maximal ideal.*
- *I is a prime ideal and $I \neq \{0\}$.*
- *there exists $p \in R$ such that $I = (p)$*

Proof. Again, the proof is similar to what we've done before for $F[x]$. □

Definition 22. An integral domain R is a **Euclidean domain** if there exists $N : R - \{0\} \to \mathbb{Z}$ called the **Euclidean norm** such that:

- $N(r) \geq 0$ for all $r \in R - \{0\}$
- For all $a, b \in R$ with $a \neq 0$, there exist $q, r \in R$ such that $b = aq + r$ and either $r = 0$ or $N(r) < N(q)$.

Note that this roughly says that one can do long division with remainder. However, it does not require the remainder to be unique. A simple example is $R = \mathbb{Z}$, where $N(k) = |k|$ (this is defined even for 0). In the case of $R = F[x]$ we have the degree of a polynomial as the norm (not defined for 0).

Definition 23. A Euclidean norm N is **submultiplicative** if for all $r, s \in R - \{0\}$, $N(r) \leq N(rs)$. N is **multiplicative** if for all $a, b \in R - \{0\}$, $N(ab) = N(a)N(b)$.

Note: Robert Friedman made the word "submultiplicative" up.

Theorem 35. *Let R be a Euclidean domain. Then R is a PID.*

Proof. Let I be an ideal in R. If $I = \{0\}$, then $I = (0)$. Assume $I \neq \{0\}$. We choose $d \in I$ non-zero such that $N(d)$ is the smallest possible. We claim that $I = (d)$. We know that $(d) \subset I$. If $b \in I$, we can write $b = dq + r$ with either $r = 0$ or $N(r) < N(d)$. But $r = b - dq \in I$ so $r \neq 0$. However, this implies that $N(r) < N(d)$, which contradicts the choice of d. Thus $r = 0$ and $b = dq$,so $b \in (d)$ and $I \subset (d)$. Consequently, $I = (d)$. □

Note that there are PIDs that are *not* Euclidean.

Lemma 36. *Suppose R is Euclidean and N is a submultiplicative norm. Then, given $r, s \in R - \{0\}$, either s is a unit and $N(r) = N(rs)$ or s is not a unit and $N(r) < N(rs)$.*

Proof. Say s is a unit. $N(r) \leq N(rs)$ and $N(rs) \leq N(rss^{-1}) = N(r)$. Thus, $N(rs) = N(r)$. Conversely, supposed $N(r) = N(rs)$. We must show that s is a unit (hence if s is not a unit, $N(r) < N(rs)$). The idea is to long divide rs into r. We can write $r = (rs)q + t$ where either $t = 0$ or $N(t) < N(rs) = N(r)$. We claim that $N(t) < N(r)$ is impossible because $t = r - rsq = r(1 - sq)$ implies $N(t) = N(r(1 - sq)) \geq N(r)$. Hence, $t = 0$, which implies that $r = rsq$ or $sq = 1$, i.e s is a unit. □

This has various consequences; if R is Euclidean, N submultiplicative

- $c \in R^*$ if and only if $N(r) = N(1)$
- If $r \in R$, non-zero, not a unit, then r can be factored into a product of irreducibles $r = p_1 \cdots p_n$

The first is left for homework but the second can be proved as follows. If r is irreducible, then we are done. Otherwise, $r = st$, neither a unit, and $N(r) = N(st)$ with $N(s) < N(r)$ and $N(t) < N(r)$. One can keep going and argue by complete induction on $N(r)$ to further reduce s, t.

Remark. In any Euclidean domain, we can implement the Euclidean algorithm to find a gcd. Look up the Euclidean algorithm if you are unfamiliar with it.

Example 15. Take the **Gaussian integers** $\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}$. We take $N : \mathbb{Z}[i] \to \mathbb{Z}$ such that $N(a + bi) = a^2 + b^2$. Consequently, if we write $\alpha = a+bi$, we have that $N(\alpha) = \alpha\bar{\alpha}$. It is easy to see that $N(\alpha) \geq 0$, $N(\alpha) = 0$ if and only if $\alpha = 0$, and that N is multiplicative (in fact, N extends to a function from $\mathbb{Q}(i) \to \mathbb{Q}$, which in fact becomes a homomorphism). Less obvious is the fact that we can do long division with remainder.

More generally, we could work with $\mathbb{Z}[\sqrt{-d}] \subset \mathbb{Q}(\sqrt{-d})$ with $d \in \mathbb{N}$ – these are called **imaginary quadratic fields**. Since $\mathbb{Z}[\sqrt{-d}] = \left\{a + b\sqrt{-d} : a, b \in \mathbb{Z}\right\}$, we define a norm $N(a + b\sqrt{-d}) = a^2 + db^2$. In other words, there's nothing special here about $d = 1$. In general, though, for $d \geq 3$, this is never a Euclidean norm, and in fact, $\mathbb{Z}[-\sqrt{d}]$ is never a UFD (for $d \geq 3$).

We could even look at $\mathbb{Z}[\sqrt{d}]$ where $d \in \mathbb{N}$ is not a perfect square. This would be a subring of $\mathbb{Q}(\sqrt{d}) \subset \mathbb{R}$. These are called **real quadratic fields**. Here, we could have a norm given by $N(a+b\sqrt{d}) = |(a+b\sqrt{d})(a-b\sqrt{d})| = |a^2 - bd^2|$. This norm is multiplicative, and sometimes Euclidean. It is in fact an unsolved problem about whether there are finitely many or infinitely many d such that $\mathbb{Z}[\sqrt{d}]$ is a UFD!

Theorem 37. *For the Gaussian integers, $\mathbb{Z}[i]$, the norm $N : \mathbb{Z}[i] \to \mathbb{Z}$ such that $N(a + bi) = a^2 + b^2$ is a Euclidean norm, i.e. for all $\alpha, \beta \in \mathbb{Z}[i]$, for $\alpha \neq 0$, there exists ξ, ρ such that $\beta = \alpha\xi + \rho$ with either $\rho = 0$ or $N(\rho) < N(\alpha)$.*

Proof. Given $\alpha, \beta \in \mathbb{Z}[i]$, $\alpha \neq 0$, we want to find $\xi \in \mathbb{Z}[i]$ such that $\beta/\alpha - \xi = \gamma$ with $N(\gamma) < 1$, where $\gamma \in \mathbb{Q}(i)$. Once we've found this, we set $\rho = \alpha\gamma = \beta - \alpha\xi$. Since β, α, ξ are Gaussian integers, ρ is as well. But $N(\rho) = N(\alpha\gamma) = N(\alpha)N(\gamma) < N(\alpha)$. Note that this allows for $\rho = 0$.

We write $\beta/\alpha \in \mathbb{Q}(i)$. We know $\beta/\alpha = q_1 + q_2 i$, $q_1, q_2 \in \mathbb{Q}$. If $\xi = n + mi$, then $\beta/\alpha - \xi = (q_1 - n) + (q_2 - m)i$. Then $N(\beta/\alpha - \xi) = (q_1 - n)^2 + (q_2 - m)^2$.

But for any rational number q_1, there exists an $n \in \mathbb{Z}$ such that $|q_1 - n| \leq 1/2$. We choose n such that $|q_1 - n| \leq 1/2$ and m such that $|q_2 - m| \leq 1/2$. Then, $N(\gamma) = N(\beta/\alpha - \xi) \leq 1/4 + 1/4 < 1$. Finally, we take $\rho = \beta - \alpha\xi$, and by the above paragraph, we are done. □

Corollary 38. *The Gaussian integers, $\mathbb{Z}[i]$, are both a PID and a UFD.*

Proof. Follows from the previous theorem and our remarks about the norm chosen – $\mathbb{Z}[i]$ is a Euclidean domain and thus a PID and a UFD. □

Let's now examine factorization in $\mathbb{Z}[i]$. First of all, what are the units? We've seen that they are $\{\pm 1, \pm i\}$. In fact, we claim that $\alpha \in \mathbb{Z}[i]$ is a unit if and only if $N(\alpha) = 1$: If $\alpha\beta = 1$, we can write $1^2 = 1 = N(1) = N(\alpha)N(\beta)$ and since $N(\alpha) \in \mathbb{Z}$, $N(\alpha)|1$ and thus $N(\alpha) = 1$. Conversely, if $N(\alpha) = 1 = \alpha\bar{\alpha}$, then $\bar{\alpha}$ is an inverse in $\mathbb{Z}[i]$, and thus α is a unit.

A harder question to ask is: what are the irreducibles in $\mathbb{Z}[i]$? Note that if $n \in N$, $n = N(\alpha)$ for some $\alpha = a + bi \in \mathbb{Z}[i]$, we know $n = a^2 + b^2$. Thus n is a sum of 2 integer squares (and the converse). This shows that factorization in $\mathbb{Z}[i]$ is in fact connected to a classical problem in number theory. We will approach the problem slightly differently.

Lemma 39. *If p is a prime number and $\alpha \in \mathbb{Z}[i]$ with $N(\alpha) = p$, then α is irreducible in $\mathbb{Z}[i]$.*

Proof. If $\alpha = \beta\gamma$, then $N(\alpha) = N(\beta)N(\gamma) = p$, but since p is prime, either $N(\beta) = 1$ or $N(\gamma) = 1$. But by what we noticed above, this requires either of β, γ to be a unit, i.e. that α is irreducible. □

Example 16. $N(1+i) = 2$ and thus $1+i$ is irreducible. Similarly $N(2+i) = 5$ and thus $2 + i$ is irreducible. $N(3 + i) = 10$, and in fact, $3 + i$ is not irreducible.

Lemma 40. *p is a prime number, and p is not irreducible in $\mathbb{Z}[i]$ if any only if there exists an $\alpha \in \mathbb{Z}[i]$ such that $p = N(\alpha)$, which again simply means that p is a sum of two integer squares. Note that if $p = N(\alpha)$, then α is irreducible.*

Proof. If p is not irreducible, then it factors $p = \alpha\beta$, where neither α, β is a unit. Then we take norms, $N(p) = N(\alpha\beta) = N(\alpha)N(\beta) = p^2$, and since we know that $N(\alpha) \neq 1$ and $N(\beta) \neq 1$, we must have $N(\alpha) = N(\beta) = p$. This implies that α is irreducible and that $p = \alpha\bar{\alpha}$, a sum of squares. Conversely, if $p = \alpha\bar{\alpha}$, p is not irreducible since $N(\alpha) = N(\bar{\alpha}) = p$, $\alpha, \bar{\alpha}$ not units. □

Consequently, if p is a prime, either p is irreducible in $\mathbb{Z}[i]$ or $p = N(\alpha) = \alpha\bar{\alpha}$ where α is irreducible in $\mathbb{Z}[i]$.

Lemma 41. *If $\pi \in \mathbb{Z}[i]$ is irreducible then there exists a prime number p such that $\pi|p$ in $\mathbb{Z}[i]$. The only possibilities are that either π is p or an associate, or $N(\pi) = \pi\bar{\pi} = p$.*

Proof. Given π, consider $\pi\bar{\pi} = N(\pi) \in \mathbb{N}$ an $N(\pi) > 1$ as π is not a unit. We factor: $\pi\bar{\pi} = p_1 \cdots p_k$, p_i primes. Then $\pi|p_1 \cdots p_k$. Since $\mathbb{Z}[i]$ is a UFD, $\pi|p_i = p$ for some i. If p is itself irreducible in $\mathbb{Z}[i]$ we have one irreducible dividing another, and thus π must be equal to p up to associate. Otherwise, $p = \alpha\bar{\alpha}$ where $\alpha, \bar{\alpha}$ are irreducible. Then $\pi|\alpha\bar{\alpha}$ which means that either $\pi|\alpha$ or $\pi|\bar{\alpha}$. Thus π is an associate of α or $\bar{\alpha}$. Then $N(\pi) = \pi\bar{\pi} = N(\alpha) = p$. □

Lemma 42. *If $p \equiv \mod 4$ then there exists $k \in \mathbb{Z}$ such that $k^2 \equiv -1 \mod p$, i.e there exists an $a \in \mathbb{Z}/p\mathbb{Z}$ such that $a^2 = -1$.*

Proof. $p \equiv 1 \mod 4$ if and only if $4|(p-1)$. The order of $(\mathbb{Z}/p\mathbb{Z})^*$ is $p-1$. We know that this is a cyclic group, and thus that there exists a cyclic subgroup $\langle a \rangle$ of order 4, i.e. $a \in (\mathbb{Z}/p\mathbb{Z})^*$, $a^4 = 1, a^2 \neq 1$. We know $(a^2)^2 = 1$ so a^2 is a root of $x^2 - 1$ in $\mathbb{Z}/p\mathbb{Z}[x]$. But we can factor this to $(x-1)(x+1)$, and thus a^2 is either 1 or -1. It must be -1, since its order is 4, and thus we are done. □

Theorem 43. *The irreducibles in $\mathbb{Z}[i]$ are*

1. *$1 + i$ and its associates. Note that $N(1+i) = 2$*
2. *$p \in \mathbb{Z}$ a prime number such that $p \equiv 3 \mod 4$ and associates*
3. *$\pi \in \mathbb{Z}[i]$ such that $N(\pi) = p$ a prime, $p \equiv 1 \mod 4$, and moreover, if $p \equiv 1 \mod 4$ then $p = N(\pi)$ for some irreducible π.*

Proof.

1. $2 = 1 \times 1 = N(1+i)$ and we are done.
2. If $p \equiv 3 \mod 4$, then p can't be written as the sum of two integer squares, which implies that p is irreducible in $\mathbb{Z}[i]$. Why? Because if $p = a^2 + b^2$, $a, b \in \mathbb{Z}$ with p odd, they can't both be even or odd. Thus one is even and the other is odd. If we take $a = 2k+1, b = 2k$, we have that $a^2 + b^2 \equiv 1 \mod 4$.
3. Must show that if $p \equiv 1 \mod 4$, then $p = N(\pi) = a^2 + b^2$. By the lemma above, there exists k such that $k^2 \equiv -1 \mod p$, which implies that $p|(k^2+1) = (k+i)(k-i)$. We claim that this means p is not

irreducible, because if it were, then since $p|(k+i)(k-i)$, we would have to have $p|(k \pm i)$. But $k/p \pm 1/p$ is not a Gaussian integer, and thus p doesn't divide either factor, and consequently, p is not irreducible. But if a prime is not irreducible, it is the norm of something that is irreducible, and we are done.

□

Corollary 44. *A prime number p is a sum of 2 integer squares if and only if $p = 2$ or $p \equiv 1 \mod 4$.*