Notes on splitting fields

Math 361, UMass Boston

May 9, 2022

Our last class ended with *Kronecker's Theorem:* If $f \in F[x]$ is non-constant, then there exists a field extension $F \to E$ in which f has a root. One constructs this extension by first factoring f as a product of irreducibles in $F[x]$, say $f = p_1 \dots p_k$, then putting $E = F[x]/\langle p_1 \rangle$. This will always be a field precisely because p_1 is irreducible. Also, the standard generator $\alpha = x + \langle p_1 \rangle$ satisfies $p_1(\alpha) = 0$ so $f(\alpha) = 0$.

As the next example shows, the extension field E is *not* unique; at the very least it can depend on which irreducible factor of f we use to contruct it.

Example 1. Take $F = \mathbb{Q}$ and $f = x^4 - 5x^2 + 6$. Over $\mathbb{Q}$ we have the factorization $f = (x^2 - 2)(x^2 - 3)$. If we put $E_1 = \mathbb{Q}[x]/\left\langle x^2 - 2 \right\rangle$ then over E_1 we have the further factorization $f = (x - \alpha)(x + \alpha)(x^2 - 3)$.

We would like to know whether the quadratic $x^2 - 3$ has any further factorization over E_1. Since the degree is low, this is equivalent to finding out whether $x^2 - 3$ has any roots in E_1. If $a\alpha + b$ were such a root, we would have $(a\alpha + b)^2 = a^2\alpha^2 + 2ab\alpha + b^2 = 2ab\alpha + (2a^2 + b^2) = 0\alpha + 3$ implying $2ab = 0$ and $2a^2 + b^2 = 3$. If $a = 0$ then $b^2 = 3$, contradicting the fact that b is rational, while if $b = 0$ then $a^2 = 3/2$, contradicting the fact that a is rational. Therefore there is no such root, and $f = (x - \alpha)(x + \alpha)(x^2 - 3)$ is a factorization into irreducibles over E_1.

Alternatively, we could have put $E_2 = \mathbb{Q}[x]/\left\langle x^2 - 3 \right\rangle$. Denoting the standard generator $x + \left\langle x^2 - 3 \right\rangle$ by β we have the factorization $f = (x^2 - 2)(x - \beta)(x + \beta)$ over E_2. By an argument similar to that given above, the factor $x^2 - 2$ has no roots in E_2 so this is a complete factorization over E_2.

Observe that E_1 and E_2 cannot be isomorphic extensions, since E_1 contains an element that squares to 2, while E_2 does not.

This example shows that the extension constructed in the proof of Kronecker's Theorem is unsatifying in at least two ways: it is not unique, even up to isomorphism, and it may not be large enough to contain "all" the roots of f.

Definition 2. Suppose $f \in F[x]$ is non-constant, and $F \to E$ is a field extension. We say that f *splits over* E if f can be written as a product of linear factors in $E[x]$.

Example 3. The polynomial $f = x^2 - 4$ splits over $\mathbb{Q}$, since we can write $f = (x - 2)(x + 2)$. By contrast, $g = x^2 - 5$ does *not* split over $\mathbb{Q}$.

Example 4. Let f and E_1 be as in example 1. Then f does *not* split over E_1 even though it has a root there.

Example 5. Let f be as in example 1. Then f splits over $\mathbb{R}$, since over the reals, f can be factored as $f = (x - \sqrt{2})(x + \sqrt{2})(x - \sqrt{3})(x + \sqrt{3})$.

Definition 6. Suppose $f \in F[x]$ is not constant, and $F \to E$ is a field extension. We say that $F \to E$ is a *splitting field* for f if

1. The polynomial f splits over E, and
2. The extension $F \to E$ is generated by roots of f, i.e. the smallest subfield of E containing (the image of) F and all of the roots of f is E itself.

Example 7. Let f and E_1 be as in example 1. Then $\mathbb{Q} \to E_1$ is *not* a splitting field for f because f does not split over E_1. Also, $\mathbb{Q} \to \mathbb{C}$ is *not* a splitting field for f: even though f does split over $\mathbb{C}$, there are many proper subfields (e.g. $\mathbb{R}$) that contain $\mathbb{Q}$ and all the roots of f.

In the previous example, E_1 is "too small" to be a splitting field because f does not actually split over it. By contrast, $\mathbb{C}$ is "too large" to be a splitting field because f can still be split over proper subfields. (We shall see that $\mathbb{R}$ is also too large to be a splitting field for f). It seems reasonable to expect that there should be some field extension that is "just right" for splitting any given f. This turns out to be so; our main result for today is that splitting fields always exist, and in fact they are unique up to a suitable notion of isomorphism.

Definition 8. Suppose (F, E_1, ι_1) and (F, E_2, ι_2) are extensions of the same base field. We say that these extensions are *isomorphic* if there exists a field isomorphism $\phi : E_1 \to E_2$ with $\phi \circ \iota_1 = \iota_2$.

Example 9. Take $F = \mathbb{Q}$. Put $E_1 = \mathbb{Q}[x]/\left\langle x^2 - 2 \right\rangle$ with $\iota(a) = (a + 0x + 0x^2 + \dots) + \left\langle x^2 - 2 \right\rangle$ (i.e. we send the rational number a to the coset of the constant polynomial with value a).

Next, put $E_2 = \{r + s\sqrt{2} \,|\, r, s \in \mathbb{Q}\}$ (here $\sqrt{2}$ denotes the positive real number whose square is 2, so that E_2 is a subset of $\mathbb{R}$). One verifies that E_2 is in fact a subfield of $\mathbb{R}$. (The hardest part of the verification is showing that every non-zero element is a unit. For this one uses the calculation $\frac{1}{r+s\sqrt{2}} = \frac{r-s\sqrt{2}}{r^2-2s^2} = \frac{r}{r^2-2s^2} + \frac{-s}{r^2-2s^2}\sqrt{2}$.) Define $\iota_2 : \mathbb{Q} \to E_2$ by $\iota_2(r) = r + 0\sqrt{2}$.

We claim that $(\mathbb{Q}, E_1, \iota_1)$ and $(\mathbb{Q}, E_2, \iota_2)$ are isomorphic field extensions. To see this, define a map $\psi : \mathbb{Q}[x] \to \mathbb{R}$ by the formula $\psi(a_0 + a_1x + a_2x^2 + a_3x^3 \dots) = a_0 + a_1\sqrt{2} + a_2\sqrt{2}^2 + a_3\sqrt{2}^3 + \cdots = (a_0 + 2a^2 + \dots) + (a_1 + 2a_3 + \dots)\sqrt{2}$. Evidently $\mathrm{im}(\psi) = E_2$. It is straightforward though tedious to verify that ψ is a unital ring homomorphism. Next, $\ker(\psi)$ is an ideal of $\mathbb{Q}[x]$ that contains $x^2 - 2$, so it must be a principal ideal, generated by some divisor of $x^2 - 2$. But $x^2 - 2$ is irreducible over $\mathbb{Q}$, so this divisor must actually be an associate of $x^2 - 2$, from which it follows that $\ker(\psi) = \left\langle x^2 - 2 \right\rangle$. Now by applying the Fundamental Theorem on Homomorphisms, we get a monomorphism $\widehat{\psi} : E_1 \to \mathbb{R}$ whose image

is E_2; restricting the codomain gives an isomorphism from E_1 to E_2. Finally, $\widehat{\psi}(\iota_1(r)) = \widehat{\psi}((r + 0x + \dots) + \langle x^2 + 1 \rangle) = \psi(r + 0x + \dots) = r + 0\sqrt{2} = \iota_2(r)$, so $\widehat{\psi}$ is indeed an isomorphism of field extensions.

Remark 10. In the previous example, the field E_1 is a rather "abstract" object; its elements are cosets of polynomial expressions modulo multiples of a fixed modulus, and no one who has not made a rather extensive study of abstract algebra will be able to comprehend it. On the other hand, E_2 is very "concrete;" it is an actual set of real numbers, and it could be presented to middle-schoolers.

Since E_1 and E_2 are isomorphic, why should we ever think about such a complicated object as E_1? The answer is that E_1 is much better suited to machine-based *symbolic computation*. An element of E_1 can be stored on a machine by storing a pair of rational numbers (the coefficients in the expression $r + sx + \langle x^2 - 2 \rangle$), and this in turn can be done by storing four integers (the numerators and denominators of r and s), which can be done exactly, with *no approximation whatsoever* and hence no problems associated with accumulating roundoff error. By contrast, an element of E_2 is an actual real number, typically an irrational one, and one must inevitably make approximations when manipulating elements of E_2.

The fact that these field extensions are isomorphic means that we can have the best of both worlds. When making calculations, we can work with elements of E_1 and can manipulate them all day long without fear of accumulating errors. Then, when all of our calculations are made, we can push our "final answers" through the isomorphism $\widehat{\psi}$ to obtain decimal approximations suitable for engineering purposes.

We are nearly ready to prove that splitting fields exist and are unique up to isomorphism, but we need one additional technical tool for the proof.

Definition 11. Suppose $f \in F[x]$ is not constant. The *non-split part* of f over F, denoted $NS_F(f)$, is the product of the non-linear irreducible factors of f in $F[x]$. (If there are no such factors, then by convention we take the *empty product* to be 1.)

Example 12. We shall compute $NS_{\mathbb{Q}}(x^5 - 3x^3 + x)$. The prime factorization over $\mathbb{Q}$ is $x^5 - 3x^3 + x = x(x-2)(x+2)(x^2+1)$. Discarding the linear factors gives $NS_{\mathbb{Q}}(x^5 - 3x^3 + x) = x^2 + 1$.

Remark 13. The non-split part is actually defined only up to unit multiples, since we are always free to move units around in the factorization. For instance, in the example above, we could have factored as $x^5 - 3x^3 + x = \left(\frac{1}{2}x\right)(x - 2)(x + 2)(2x^2 + 2)$ which would have given $NS_{\mathbb{Q}}(x^5 - 3x^3 + x) = 2x + 2$. Thus, technically speaking, we should think of $NS_F(f)$ not as a single polynomial but as an *associate class* of polynomials. However, all polynomials in an associate class have the same degree, so $\deg(NS_F(f))$ is an unambiguous integer; this will become important below.

Example 14. We now compute $NS_{\mathbb{C}}(x^5 - 3x^3 + x)$. The prime factorization over $\mathbb{C}$ is $x^5 - 3x^3 + x = x(x-2)(x+2)(x+i)(x-i)$. All factors are linear, giving $NS_{\mathbb{C}}(x^5 - 3x^3 + x) = 1$.

Example 15. Let f and E_1 be as in example 1. Then, making use of the factorization given there, we obtain $NS_{E_1}(f) = x^2 - 3$.

We are now ready to prove our central result:

Theorem 16. *Suppose F is a field and $f \in F[x]$ is a non-constant polynomial with coefficients in F. Then f has a splitting field. Moreover, splitting fields are unique up to isomorphism of field extensions.*

Proof. The proof is by induction on $\deg(NS_F(f))$.

In the base case, $\deg(NS_F(f)) = 0$, implying that all of the irreducible factors of f are linear, i.e. f already splits over F. In this case we take $E = F$ and we take ι to be the identity map. Certainly f splits over E. Also the smallest subfield of E containing F (and all the roots of f) is evidently E itself, since $E = F$. This proves existence.

For uniqueness, suppose that (F, E', ι') is a second splitting field. Since f already splits over F, all of the roots of f in E' already lie in the image $\iota'[F]$. Since E' is generated by the roots of f, this forces $E' = \iota'[F]$ so that in fact ι' is a field isomorphism. Define $\phi : E \to E'$ by the simple formula $\phi = \iota'$ (recall that $E = F$). Since ι is the identity map, we do indeed have $\phi \circ \iota = \iota'$ and thus ϕ is an isomorphism of extensions. This concludes the proof in the base case.

Now we turn to the inductive step. Since we are not in the base case we have $\deg(NS_F(f)) > 0$. In particular, f must have some non-linear irreducible factor p_1. Put $E_1 = F[x]/\langle p_1 \rangle$ and $\iota_1(a) = (a + 0x + 0x^2 + \dots) + \langle p_1 \rangle$. Since p_1 has a root in E_1, when factoring f over E_1 it will be possible to separate at least one additional linear factor in $E_1[x]$, beyond those in $F[x]$. Thus, $\deg(NS_{E_1}(f)) < \deg(NS_F(f))$.

Temporarily regarding E_1 as a new base field, we can invoke the inductive hypothesis and assume that there is a splitting field (E_1, E, ι_2) over E_1. The composite map $\iota = \iota_2 \circ \iota_1$ takes F to E, so we can regard E as an extension of F. To complete the existence proof, we need to show that (i) f splits over E, and (ii) E is generated by $\iota[F]$ and the roots of f. But (i) is automatic because (E_1, E, ι_2) is a splitting field and thus f does indeed split over E. For (ii), note that $E_1 = \{a_0 + a_1\alpha + \dots + a_{\deg(p_1)-1}\alpha^{\deg(p_1)-1} \mid a_i \in F\}$ and α is a root of f, so any subfield of E containing $\iota[F]$ and the roots of f will contain all of $\iota_2[E_1]$ and then, by induction, it will be all of E. This completes the existence proof.

For uniqueness, suppose (F, E', ι') is another splitting field for f. Then unique factorization implies that p_1 splits over E'; in particular, p_1 has some root $\beta \in E'$. Define a ring homomorphism $\psi : F[x] \to E'$ by the formula $\psi(a_0 + a_1x + a_2x^2 + \dots) = \iota'(a_0) + \iota'(a_1)\beta + \iota'(a_2)\beta^2 + \dots$. By construction, $\psi(p_1) = 0$ so $\ker(\psi)$ is generated by some divisor of p_1. But p_1 is irreducible so in fact $\ker(\psi) = \langle p_1 \rangle$ and now the Fundamental Theorem of Homomorphisms gives a monomorphism $\widehat{\psi} : E_1 \to E'$. Thus, we can regard E as an extension of E_1.

Once again we temporarily regard E_1 as the base field. Since f splits over E and is generated (over F) by roots of f, it is also generated (over E_1) by roots of f and thus $(E_1, E', \widehat{\psi})$ is a splitting field for f. By induction, there must be a field isomorphism $\phi : E \to E'$ with $\phi \circ \iota_2 = \widehat{\psi}$. But then $\phi \circ \iota = \phi \circ \iota_2 \circ \iota_1 = \widehat{\psi} \circ \iota_1 = \iota'$ and so, shifting back to the perspective in which F is the base field, ϕ is still an isomorphism of field extensions. This completes the uniqueness proof. □

Example 17. Returning to the setup of example 1, we shall compute the splitting field of $f = x^4 - 5x^2 + 6$ over $\mathbb{Q}$. As in that example, we put $E_1 = \mathbb{Q}[x]/\langle x^2 - 2\rangle$ and we denote the standard generator $x + \langle x^2 - 2\rangle$ by α. Factoring f over E_1 gives $f = (x - \alpha)(x + \alpha)(x^2 - 3)$, so E_1 is not yet the splitting field.

Now put $E = E_1[x]/\langle x^2 - 3\rangle$ and denote the standard generator $x + \langle x^2 - 3\rangle$ by β. Factoring f over E gives $f = (x - \alpha)(x + \alpha)(x - \beta)(x + \beta)$, so indeed f splits over E.

To see that E is generated by the roots of f, it is useful to make a more explicit description of the elements of E. By construction, we have

$$E = \{b_0 + b_1\beta \,|\, b_0, b_1 \in E_1\}$$

with "rule of arithmetic" $\beta^2 = 3$ but also

$$E_1 = \{a_0 + a_1\alpha \,|\, a_0, a_1 \in \mathbb{Q}\}$$

with "rule of arithmetic" $\alpha^2 = 2$. Combining these descriptions gives

$$\begin{aligned} E &= \{(a + b\alpha) + (c + d\alpha)\beta \,|\, a, b, c, d \in \mathbb{Q}\} \\ &= \{a + b\alpha + c\beta + d\alpha\beta \,|\, a, b, c, d \in \mathbb{Q}\} \end{aligned}$$

with "rules of arithmetic" $\alpha^2 = 2$ and $\beta^2 = 3$. This description makes it clear that the smallest subfield of E containing $\mathbb{Q}$ and both α and β is E itself. Since α and β are both roots of f, this shows that E is the splitting field of F.

Remark 18. As in example 9, it is useful to consider alternative (but isomorphic) splitting fields for f. Thus, consider the set of real numbers

$$E' = \{a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6} \,|\, a, b, c, d \in \mathbb{Q}\}.$$

It is straightforward to show that E' is a unital subring of $\mathbb{R}$, and that it is in fact isomorphic to E, via the isomorphism $a + b\alpha + c\beta + d\alpha\beta \mapsto a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}$.

Then, as in example 9, it becomes reasonable to ask why anyone would work with the highly abstract object E rather than the highly concrete set of real numbers E'. As before, E is more suitable for machine computation—its elements can be stored without approximation and without risk of accumulating error. In addition, it is not so easy to recognize that E' is a field, except by observing that it is isomorphic to E. Specifically, there is no obvious method to compute multiplicative inverses in E'. However, in E, we can use the extended Euclidean algorithm to invert elements. Thus, by porting the inversion problem

across the isomorphism $E \simeq E'$, we do obtain an inversion method for E', but it would be quite difficult to discover this method if we had no knowledge of the "abstract" object E.