1.6 06/21/24 - Nullstellensatz and Unique Factorization

Last time, we proved that if R is a UFD, then so is $R[t]$. The same circle of ideas allows us to compare irreducibles in $R[t]$ and $K[t]$. Let's prove two results in this direction, and then return to the theory of curves to see their applications.

As before, in what follows we will take R to be a UFD and $K = \operatorname{Frac} R$ to be its fraction field.

Lemma 1.6.1.

(a) If $f \in R[t]$ is irreducible and of positive degree, then f is irreducible in $K[t]$.
(b) If $f \in R[t]$ is primitive and irreducible in $K[t]$, then f is irreducible in $R[t]$.

Proof.

(a) In this case, f is a nonzero nonunit in $K[t]$. If $f = gh$ for $g, h \in K[t]$, then Lemma 1.5.18(a) tells us that $\tilde{f} = \tilde{g}\tilde{h}$, and then $f = (\operatorname{cont}(f) \cdot \tilde{g}) \cdot \tilde{h}$. Since f is irreducible in $R[t]$, either $\operatorname{cont}(f) \cdot \tilde{g}$ is a unit in R, in which case $\tilde{g}$ is a (nonzero) constant and hence $g \in K[t]^\times$ by Lemma 1.5.17(a), or simlarly $\tilde{h}$ is a unit in R, in which case $h \in K[t]^\times$.
(b) This is Lemma 1.5.18(c), given that the terms "prime" and "irreducible" are interchangable in $R[t]$ and $K[t]$ thanks to Proposition 1.5.8 and Theorem 1.5.12.

■

In any UFD S, we say that two elements $f, g \in S$ are relatively prime if there is no prime $p \in S$ such that $p \mid f$ and $p \mid g$.

Lemma 1.6.2. If $f, g \in R[t]$ are relatively prime in $R[t]$, then

(a) they are relatively prime in $K[t]$, and
(b) there are $a, b \in R[t]$ and $0 \neq c \in R$ such that $af + bg = c$.

Proof.

(a) If there is a prime $q \in K[t]$ such that $q \mid f$ and $q \mid g$ in $K[t]$, then by rescaling we can assume without loss of generality that $q \in R[t]$ is primitive (how?), and then Lemma 1.5.18(b) tells us that $q \mid f$ and $q \mid g$ in $R[t]$, and Lemma 1.5.18(c) tells us that q is prime in $R[t]$. This can't happen if $f, g \in R[t]$ are relatively prime in $R[t]$.
(b) This is clear from the Euclidean algorithm and backward substitution if R is a field (make sure you understand this!). In the general case, the first observation and part (a) combine to tell us that there are $a_1, b_1 \in K[t]$ and $0 \neq c_1 \in K$ such that $a_1 f + b_1 g = c_1$. Now we can simply "clear denominators": find a $0 \neq d \in R$ such that $a := a_1 \cdot d$ and $b := b_1 \cdot d$ are in $R[t]$, and $c := c_1 d \in R$.

■

Example 1.6.3. Take $R = \mathbb{Z}$ and $f(t) = t^3 + 1$ and $g(t) = t^2 - 7$. Then we can take $a = -7t + 1$ and $b = 7t^2 - t + 49$ with $c = -342$ via the identity

$$(-7t + 1)(t^3 + 1) + (7t^2 - t + 49)(t^2 - 7) = -342 = -2 \cdot 3^2 \cdot 19.$$

Note that the same polynomial identity holds over any ring R, but something special happens over $R = \mathbb{Z}/2, \mathbb{Z}/3$ and $\mathbb{Z}/19$: the polynomials f and g end up being not relatively prime. In fact, f and g are not relatively prime in $\mathbb{Z}/p$ iff $p \in \{2, 3, 19\}$. This fascinating observation has to do with resultants again–see Remark 1.6.5.

Example 1.6.4. Consider the polynomials $f(x,y) = x^3 - 12x - y^2$ and $g(x,y) = x^2 - xy - y^2 + 5$ in $k[x,y]$ for some field k (e.g. $k = \mathbb{C}$). Applying the above procedure to $R = k[y]$ with variable $t = x$ yields

$$\begin{aligned} a_y &= (-2y^2 + 17)x + y(3y^2 - y - 22), \\ b_y &= (2y^2 - 17)x^2 + y(-y^2 + y + 5)x + (y^4 + y^3 - 46y^2 + 289), \text{ and} \\ c_y &= -y^6 - 4y^5 + 52y^4 + 27y^3 - 519y^2 + 1445. \end{aligned}$$

On the other hand, applying the above procedure to $R = k[x]$ with variable $t = y$ yields

$$\begin{aligned} a_x &= (-x)y + (x^3 - 2x^2 - 12x - 5), \\ b_x &= xy + (-x^3 + x^2 + 12x + 5), \text{ and} \\ c_x &= x^6 - 3x^5 - 23x^4 + 26x^3 + 154x^2 + 120x + 25. \end{aligned}$$

Remark 1.6.5 (Resultants)**.** If we fix integers $m, n \geq 1$, and take $R = \mathbb{Z}[a_0, \ldots, a_m, b_0, \ldots, b_n]$ with $f(t) = a_0 t^m + \cdots + a_m$ and $g(t) = b_0 t^n + \cdots + b_n$, then Lemma 1.6.2 gives us $a, b \in R[t]$ and $0 \neq c \in R$ such that $af + bg = c$.[16] The c of least such degree is (up to a negative sign perhaps) none other than the resultant $\mathrm{Res}_t(f,g)$ of f and g with respect to t, essentially because it is the "universal" polynomial which in the coefficients which tests the coprimality of f and g. This is not a hard result, but we won't need it directly, so I won't give a proof; you are invited to prove it (perhaps using the definition from Exercise 2.2.4) if you'd like. Lemma 1.6.2 then gives us the important consequence that the resultant of two polynomials can be written as a polynomial-linear combination of them with coefficients in the ring generated by *their* coefficients.

1.6.1 Finite Intersection of Curves, Nullstellensatz, and Irreducibility II

Let's now return to the theory of curves. One important consequence of Lemma 1.6.2, evident already from Example 1.6.4 is

Theorem 1.6.6 (Finite Intersection)**.** If k is any field and $f, g \in k[x,y]$ are nonconstant relatively prime polynomials, then the intersection $C_f \cap C_g$ is finite.

Proof. Applying Lemma 1.6.2 to $R = k[y]$ with variable $t = x$ yields $a, b \in k[x,y]$ and $0 \neq c \in k[y]$ such that $af + bg = c$. Therefore, if $(p,q) \in C_f \cap C_g$, then $c(q) = 0$, so q is one of the finitely many roots of c, and hence can only take on finitely many values. Reversing the roles of x and y, we conclude that p can only take on finitely many values as well, and hence $C_f \cap C_g$ is finite. ∎

This result generalizes the one from Exercise 2.1.7 (how?). Geometrically, what is happening is this: the roots of the polynomial c are (or at least include) the projections of the points in $C_f \cap C_g$ to the y-axis, and similarly for the corresponding polynomial in x. This yields a finite grid of horizontal and vertical lines, the finitely many intersection points of which contain $C_f \cap C_g$. See Figure 1.7 for an illustration of this phenomenon for the polynomials f and g of Example 1.6.4. We have now arrived at one of the most important results in this theory.

[16] Technically, you have to check that $f(t)$ and $g(t)$ are relatively prime in $R[t]$, but this follows because they are the "universal" polynomials–if they were not, then every pair of polynomials over any ring would have a common factor, which is absurd.

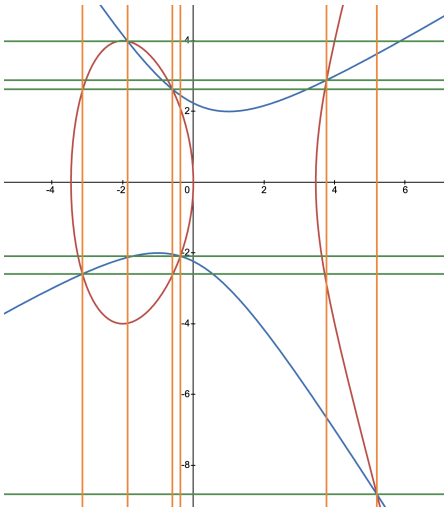

Figure 1.7: An illustration of Theorem 1.6.6 for the f and g in Example 1.6.4. The red curve is C_f, the blue curve is C_g, the green lines correspond to the roots of c_y, and the orange lines correspond to the roots of c_x. The intersection $C_f \cap C_g$ is contained in the finitely many points of the green-orange grid. Picture made with Desmos.

Theorem 1.6.7 (Hilbert's Nullstellensatz for Curves)**.** If k is an algebraically closed field, and $f, g \in k[x, y]$ are nonconstant polynomials, then $C_g \subset C_f$ iff there is some integer $n \geq 1$ such that $g \mid f^n$.

Proof. One direction is clear (which?). For the other direction, it suffices to show that if $q \in k[x, y]$ is a prime factor of g, then $q \mid f$. If there were a prime factor q for which this were not the case, then q and f would be relatively prime in $k[x, y]$, and so by Theorem 1.6.6, the intersection $C_q \cap C_f$ would be finite. But now, $C_q \subset C_g \subset C_f$ implies that $C_q \cap C_f = C_q$, which is infinite by the fact that q is nonconstant and Lemma 1.5.1.[17] ■

Note that the Nullstellensatz–German for "the theorem on the location of zeroes"–uses crucially that k is algebraically closed. We will henceforth return to our convention that k is an algebraically closed field. One important corollary we can extract is

Corollary 1.6.8. If $f, g \in k[x, y]$ are nonconstant polynomials with f irreducible, then $C_g \subset C_f$ implies $C_g = C_f$.

Proof. By Theorem 1.6.7, there is some $n \geq 1$ such that $g \mid f^n$. Then primality of f (using Corollary 1.5.14 and Proposition 1.5.8) tells us that $f \mid g$, so the easy direction of Theorem 1.6.7 implies that $C_f \subset C_g$ as needed. ■

We are now ready to prove Theorem 1.5.6, which we restate here.

[17] This is the only step where we use that k is algebraically closed.

Theorem 1.5.6. If an $f \in k[x,y]$ is irreducible, then C_f is irreducible, and conversely if $C \subset \mathbb{A}^2_k$ is an irreducible curve, then there is an irreducible $f \in k[x,y]$ such that $C = C_f$.

Proof. If f is irreducible and $C_f = C_g \cup C_h$ for nonconstant $g, h \in k[x,y]$, then Corollary 1.6.8 gives us that $C_f = C_g = C_h$, showing irreducibility of C_f. Conversely, if $C = C_{f_0} \subset \mathbb{A}^2_k$ is an irreducible curve for some $f_0 \in k[x,y]$, then we claim that there is an irreducible $f \in k[x,y]$ and an integer $n \geq 1$ such that $f_0 = f^n$. If this were not the case, we would be able to write $f_0 = gh$ for nonconstant relatively prime g, h, from which it would follow that $C = C_g \cup C_h$. Then irreducibility of C would tell us that either $C = C_g$ or $C = C_h$; suppose, without loss of generality, that $C = C_g$. Then Theorem 1.6.7 applied to the containment $C \subset C_g$ would imply that there is some $n \geq 1$ such that $f_0 \mid g^n$, which is a contradiction to the factorization $f_0 = gh$ in the UFD $k[x,y]$, since g and h are relatively prime. ■

1.6.2 Unique Factorization II

Here's the picture that we are building to: there is a parallel between the unique factorization in $k[x,y]$ and of curves in $\mathbb{A}^2_k$, namely each curve $C \subset \mathbb{A}^2_k$ can be decomposed as a finite union of irreducible curves

$$C = C_1 \cup C_2 \cup \cdots \cup C_n,$$

and these are determined uniquely upto ordering the factors. For this, the first question we can ask is:

Question 1.6.9. To what extent does a curve $C \subset \mathbb{A}^2_k$ determine a defining polynomial $f \in k[x,y]$, i.e. a polynomial f such that $C = C_f$?

The answer here is: almost, the only problem being multiplicity. Specifically, consider

Definition 1.6.10. Let R be a UFD.

(a) If a nonzero $f \in R$ is decomposed as

$$f = c f_1^{m_1} \cdots f_n^{m_n}$$

where $c \in R^\times$ is a unit, $n \geq 1$ an integer, $f_1, \ldots, f_n \in R$ irreducibles and $m_1, \ldots, m_n \geq 1$, then we define the radical of f by

$$\mathrm{rad}(f) := f_1 \cdots f_n.$$

Note that this is well-defined up to units in R.

(b) We say that a nonzero $f \in R$ is reduced if $f = \mathrm{rad}(f)$ (up to units).

Taking $R = k[x,y]$ in this definition and given any nonconstant $f \in k[x,y]$, the radical $\mathrm{rad}(f)$ is again nonconstant, and we have that

$$C_f = C_{\mathrm{rad}(f)}.$$

Therefore, a curve C cannot distinguish a polynomial from its radical. The Nullstellensatz tells us, however, that the radical can however be recovered from the curve.

Definition 1.6.11. Given a curve $C \subset \mathbb{A}^2_k$, consider the subset

$$\mathbb{I}(C) := \{g \in k[x,y] \text{ nonconstant} : C \subset C_g\} \cup \{0\} \subset k[x,y].$$

This is called the (vanishing) ideal of C. (We will define the term "ideal" properly next time.) The key claim here is then

Theorem 1.6.12. If k is algebraically closed, and $f \in k[x,y]$ is a nonconstant polynomial, then a polynomial $g \in k[x,y]$ is in $\mathbb{I}(C_f)$ iff $\text{rad}(f) \mid g$. In particular, $\text{rad}(f)$ is uniquely determined (up to nonzero scalars) by the curve C.

Proof. If g is nonconstant, then $C_f \subset C_g$ implies by Theorem 1.6.7 that for some $n \geq 1$, we have $f^n \mid g$. Since $\text{rad}(f) \mid f^n$, we are done. Finally, $\text{rad}(f)$ is simply the nonzero polnyomial of least degree in $\mathbb{I}(C)$ (up to nonzero scalars). ∎

We say that $\text{rad}(f)$ is a generator $\mathbb{I}(C)$, and call it the minimal polynomial of C.

Corollary 1.6.13 (Hilbert's Nullstellensatz for Curves, Version II)**.** Over an algebraically closed field k, there is a bijective correspondence

$$\{\text{curves } C \subset \mathbb{A}^2_k\} \dashleftarrow\dashrightarrow \{\text{nonconstant reduced } f \in k[x,y]\}/(\text{nonzero scalars})$$

given by sending an f to C_f and a curve C to its minimal polynomial. In particular, two nonconstant reduced polynomials define the same curve iff they are nonzero scalar multiples of each other.

Under this correspondence,

(a) the curve C is irreducible iff its minimal polynomial its, and
(b) the union of curves corresponds to taking the product of the minimal polynomials (and then the radical).

This result is one of the earliest manifestations of the systematization of the parallels between algebra and geometry, which is the heart and soul of algebraic geometry. We will discuss more consequences of this bijective correspondence next time.