1.7 06/24/24 - Ideals, Irreducible Components, Degree II

Today, I want to review some algebra to express our observations from last time in a cleaner way.

1.7.1 Crash Course on Ideals

Definition 1.7.1. Let R be a ring. An ideal of R is an additive subgroup $I \subset R$ such that for all $f \in I$ and $g \in R$, we have $fg \in I$.

The terminology historically comes from thinking of ideals as "ideal numbers". In the 19th century, people came to realize that in some natural rings in number theory, such as $\mathbb{Z}[\sqrt{-5}]$, unique factorization into prime numbers failed. Kummer and Dedekind salvaged this by saying that in these number rings, or in what are now known more generally as Dedekind domains, we do get a unique factorization of numbers into *prime ideal numbers*, i.e. these objects behave the way prime numbers "ideally" would.

If $I \subset R$ is an ideal, we can define an equivalence relation on R called congruence modulo I, by saying $f \sim g$ iff $f - g \in I$. The set of equivalence classes R/I then admits a structure of a ring such that then natural surjection $R \to R/I$ is a ring homomorphism (and this determines the ring structure on R/I completely). This ring R/I is called the quotient of the ring R by the ideal I.

Example 1.7.2.

(a) In any ring R, the set $I = \{0\} \subset R$ is an ideal called the zero ideal. Similarly, $I = R$, i.e. all of R, is also an ideal. We say an ideal $I \subset R$ is a proper ideal if I is a proper subset of R, i.e $I \subsetneq R$.

(b) Given a ring R and an element $f \in R$, we define the principal ideal generated by f to be the ideal $(f) := \{g \in R : f \mid g\}$. An ideal $I \subset R$ is said to be a principal ideal if $I = (f)$ for some $f \in R$; in general, this f is not unique. (E.g. $(2) = (-2)$ in $\mathbb{Z}$.) Note that (0) is the zero ideal, whereas $(1) = R$; more generally, $(u) = R$ iff $u \in R$ is a unit.

(c) More generally, given any subset $S \subset R$, the ideal generated by S is the ideal

$$(S) = \left\{ \sum_{i=1}^{n} a_i s_i : a_i \in R, s_i \in S \right\} \subset R.$$

This is the smallest (with respect to inclusion) ideal containing S, or equivalently the intersection of all ideals containing S.

(d) Any additive subgroup $S \subset \mathbb{Z}$ is of the form (n) for some unique $n \in \mathbb{Z}_{\geq 0}$. In particular, these are all the ideals in $\mathbb{Z}$. (Proof: if $S \cap \mathbb{Z}_{>0} = \emptyset$, then $S = (0)$; else, there is a least $n \in S \cap \mathbb{Z}_{>0}$ by the well-ordering principle, and then $S = (n)$.) A ring R is said to be a principal ideal ring if every ideal of R is principal; a domain R that is a principal ideal ring is called a principal ideal domain, abbreviated PID.

In general, principal ideals don't determine generators (e.g. in $R = \mathbb{Z}/6$, we have $(2) = (4)$); however, in domains[18], principal ideals determine generators up to units.

[18] Fascinatingly, this is not quite a characterization of domains. Other rings, such as local rings, also satisfy this property. I do not know of a complete characterization of rings with this property.

Lemma 1.7.3. If R is a domain and $f, g \in R$, then $(f) = (g)$ iff there is a unit $u \in R^\times$ such that $f = ug$. In other words, a principal ideal in R is determined by, and determines, its generator up to units.

Proof. One direction is clear (which, and why?). For the other direction, by assumption, there are $u, v \in R$ such that $f = ug$ and $g = vf$. Then $f(uv - 1) = 0$, so since R is a domain, one of f and $uv - 1$ is zero. If $f = 0$, then $g = vf = 0$, and $0 = 1 \cdot 0$. Otherwise, $uv = 1$ implies $u \in R^\times$. ■

Proposition/Definition 1.7.4. For a ring R and a proper ideal $P \subset R$, the following are equivalent:

(a) If $f, g \in R$, then $fg \in P$ implies either $f \in P$ or $g \in P$.
(b) The quotient ring R/P is a domain.

A proper ideal $P \subset R$ satisfying these equivalent conditions is called a prime ideal.

Example 1.7.5.

(a) A ring R is a domain iff $(0) \subset R$ is a prime ideal[19]
(b) An ideal $I \subset \mathbb{Z}$ is prime iff either $I = 0$ or $I = (p)$ for some prime integer p.
(c) In general, if R is any ring, then and $0 \neq f \in R$, then f is a prime element iff (f) is a prime ideal.

In Exercise 2.3.3, you are invited to find all prime ideals of the ring $k[x, y]$ when k is algebraically closed. Finally, we will need one more fact about ideals.

Proposition 1.7.6. Let R be a ring and $I \subset R$ be a proper ideal (i.e. $I \neq R$). Then there is a prime ideal $Q \subset R$ containing I.

Proof. Let $\mathcal{C}$ be the partially ordered set of all proper ideals of R containing I ordered by inclusion; this is nonempty because $I \in \mathcal{C}$. If (I_α) is an ascending chain of ideals in $\mathcal{C}$, then $\bigcup_\alpha I_\alpha \subset R$ is also a proper ideal of R (check!); this proves that every chain in $\mathcal{C}$ has an upper bound, and hence $\mathcal{C}$ has a maximal element Q (this element need not be unique). We claim that Q is prime. Indeed, if it were not, then there would $p, q \in R$ such that $pq \in Q$ but neither $p \in Q$ nor $q \in Q$. Then we claim that $Q + pR$ is a strictly larger ideal in $\mathcal{C}$; that it contains I is clear, that it is strictly larger follows from $p \notin Q$, and that it is proper follows from the following argument. If $Q + pR = R$, then we can write $1 = s + pt$ for some $s \in Q$ and $t \in R$. Multiplying by q yields $q = qs + pqt$, but $qs \in Q$ (because $s \in Q$) and $pqt \in Q$ (because $pq \in Q$) and hence $q \in Q$, which is a contradiction. ■

In fact, this maximal element Q of $\mathcal{C}$ as in the above proof if actually a maximal ideal of R, i.e. an ideal not contained in any other proper ideals of R (almost by definition!), and it is a general fact, which we showed in this proof, that any maximal ideal is prime.

1.7.2 Irreducible Components and Degree II

Let's return to the theory of curves; recall that we are over an algebraically closed field k. The idea here is that if $C \subset \mathbb{A}^2_k$ is a curve, then the vanishing ideal $\mathbb{I}(C)$ of C defined in Definition

[19] Here, and always, we use the convention that domains are nonzero.

1.6.11 is an ideal of the ring $k[x,y]$, and, in fact, by Theorem 1.6.12 a principal ideal.

Definition 1.7.7. Given a curve $C \subset \mathbb{A}^2_k$, a minimal polynomial of C is a generator of the principal ideal $\mathbb{I}(C) \subset k[x,y]$.

Note that any minimal polynomial must necessarily be reduced (why?). By Lemma 1.7.3, any two minimal polynomials of C differ by multiplication by units in $k[x,y]$, i.e. nonzero scalars–this is why we sometimes speak of "the minimal polyomial". If $C = C_f$ for a nonconstant $f \in k[x,y]$ then a minimal polynomial of C can be taken to $\mathrm{rad}(f)$. This gives us a perfect translation between alegbra and geometry. For instance, we can use this to define the degree of curve.

Definition 1.7.8 (Degree)**.** Given a curve $C \subset \mathbb{A}^2_k$, the degree of C is defined to be the degree of any minimal polynomial for C.

You may verify that if $k = \overline{k}$, then this definition agrees with Definition 1.2.2. Similarly, Corollary 1.6.13 can be restated as

Corollary 1.7.9 (Hilbert's Nullstellensatz for Curves, Version III)**.** Over an algebraically closed field k, there is a bijective correspondence

$$\{\text{curves } C \subset \mathbb{A}^2_k\} \dashleftarrow\dashrightarrow \{\text{pr. ideals of } k[x,y] \text{ gen. by nonconst. reduced } f \in k[x,y]\}$$

given by sending a curve C to $\mathbb{I}(C)$ and an ideal I to C_f for any generator f of I. Under this correspondence, the curve C is irreducible iff $\mathbb{I}(C)$ is a prime ideal.

Finally, from unique factorization in $k[x,y]$, we also obtain a decomposition for curves.

Theorem 1.7.10 (Unique Factorization/Irreducible Decomposition for Curves)**.** If $k = \overline{k}$, then given any curve $C \subset \mathbb{A}^2_k$, there is an integer $n \geq 1$ and irreducible curves $C_1, \ldots, C_n \subset \mathbb{A}^2_k$ such that $C_i \neq C_j$ for $i \neq j$ and

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

Further, if $m \geq 1$ is any other integer and $D_1, \ldots, D_m \subset \mathbb{A}^2_k$ irreducible curves such that $D_i \neq D_j$ for $i \neq j$ and

$$C = D_1 \cup D_2 \cup \cdots \cup D_m,$$

then $m = n$ and for all i, we have $C_i = D_{\sigma(i)}$ for some bijection $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$.

Proof. If f is a minimal polynomial of C, and we write $f = f_1 \cdots f_n$ for some $n \geq 1$ and distinct irreducible $f_1, \ldots, f_n \in k[x,y]$, then taking $C_i = C_{f_i}$ for $1 \leq i \leq n$ gives us the indicated decomposition, where we are using both that f is reduced and Corollary 1.6.8 to conclude that $C_i \neq C_j$ for $i \neq j$ (how?). If we have a decomposition $C = D_1 \cup \cdots \cup D_m$, and for each j with $1 \leq j \leq m$, we take a minimal polynomial $g_j \in k[x,y]$ for D_j, then each g_j is irreducible by Corollary 1.6.13(a), and for $i \neq j$, the polynomials g_i and g_j are not scalar multiples of each other by the hypothesis that $D_i \neq D_j$. Then the reduced polynomials f and $g := g_1 \cdots g_m$ define the same curve C, and hence by Corollary 1.6.13(c) are related by nonzero scalars; then we are done by unique factorization in $k[x,y]$, which is Corollary 1.5.14 (how?). ∎

The curves $C_1, \ldots, C_n \subset C$ occuring in such a decomposition are called the irreducible components of C, and they correspond to the irreducible factors of any minimal polynomial of C. Finally, we can upgrade Theorem 1.6.6 slightly to get

Theorem 1.7.11 (Finite Intersection Revisited)**.** If $C, D \subset \mathbb{A}^2_k$ are two curves that don't share any common irreducible components, then the intersection $C \cap D$ is finite.

Proof. Decompose $C = C_1 \cup \cdots \cup C_n$ and $D = D_1 \cup \cdots \cup D_m$ into irreducible components as in Theorem 1.7.10. For each pair (i, j) with $1 \leq i \leq n$ and $1 \leq j \leq m$, if we take minimal polynomials f_i and g_j of C_i and D_j respectively, then f_i and g_j are irreducible (by Corollary 1.6.13(a)) and $C_i \neq D_j$ implies that f_i and g_j are not scalar multiples of each other and hence relatively prime. It follows from Theorem 1.6.6 that each $C_i \cap D_j$ is finite, and hence so is

$$C \cap D = \bigcup_{1 \leq i \leq n} \bigcup_{1 \leq j \leq m} C_i \cap D_j.$$

■

1.7.3 A Few Examples of Irreducible Curves

That's enough general theory. Let's work out a few specific examples.

Example 1.7.12. For any field k, the linearpolynomial $\ell = x + y + 1 \in k[x, y]$ is irreducible: indeed, applying Lemma 1.6.1 to $R = k[x]$ with $t = y$, it suffices to show that ℓ is irreducible in $K[y]$ where $K = k(x)$, but that is true simply because $\ell \in K[y]$ is a linear polynomial![20] Therefore, the line $C_\ell \subset \mathbb{A}^2_k$ is irreducible. The same argument shows that any line in $\mathbb{A}^2_k$ is irreducible, or many generally, that if $f(x, y) \in k[x, y]$ is any polynomial that is linear in either x or y, then $f(x, y)$ is irreducible. For instance, the polynomial $f(x, y) = y - x^2 \in k[x, y]$ is irreducible, so that the parabola $C = \{(t, t^2) : t \in k\} \subset \mathbb{A}^2_k$ is as well.

Example 1.7.13. For any field k, the polynomial $f(x, y) := xy - 1 \in k[x, y]$ is irreducible thanks to Lemma 1.6.1 applied to $R = k[x]$ with $t = y$–note that although $f(x, y)$ is not monic in y, it is still primitive. Over $k = \mathbb{R}$, the polynomial $f(x, y) = xy - 1 \in \mathbb{R}[x, y]$ defines the rectangular hyperbola C with two components. Why does this not contradict irreducibility? Well, firstly: the connection between (topological) irreducibility of curves and polynomials only works over algebraically closed fields such as $k = \mathbb{C}$: over $k = \mathbb{C}$, the "hyperbola" defined by f is a topologically a sphere punctured at two points, which is connected. Secondly, the rectangular hyperbola $C \subset \mathbb{A}^2_{\mathbb{R}}$ is still algebraically irreducible:

Lemma 1.7.14. If $g(x, y) \in \mathbb{R}[x, y]$ is a polynomial that vanishes on one branch of the hyperbola C, (or, in fact, any infinite subset of C) then $f \mid g$ in $\mathbb{R}[x, y]$, so that g must also vanish on the second branch.

Proof. Either g is zero and we are done, or g is nonconstant, in which case we may consider $C_g(\mathbb{C}) \subset \mathbb{A}^2_{\mathbb{C}}$. By hypothesis, $C_g(\mathbb{C})$ and $C_f(\mathbb{C})$ intersect in infinitely many points, so it follows from Theorem 1.6.6 that $f, g \in \mathbb{C}[x, y]$ are not relatively prime. Since $f \in \mathbb{C}[x, y]$ is irreducible by Example 1.7.13, this can only happen if $f \mid g$ in $\mathbb{C}[x, y]$, so that $g/f \in \mathbb{C}[x, y] \cap \mathbb{R}(x, y) = \mathbb{R}[x, y]$. ■

[20] This uses that we understand irreducibility in the polynomial ring $K[y]$ in one variable y over a field K really well.

In other words, just one branch of the hyperbola C is not an algebraic curve by itself. This proposition illustrates that sometimes we can prove results over non algebraically closed fields by using Theorem 1.4.5, and also that curves are incredibly rigid: any polynomial vanishing on any collection of infinite points of one curve must vanish on all of it. This is a manifestation of the coarseness of the Zariski topology.

Example 1.7.15. For any field k, the polynomial $f(x,y) := y^2 - x^3 + x \in k[x,y]$ is irreducible as well. There are a few ways to prove this. One way is sketched in Exercise 2.3.1. Another way to invoke Lemma 1.6.1 again to reduce the problem to showing that $y^2 - x^3 + x \in K[y]$ is irreducible where $K = k(x)$. If it were not irreducible, then it would split into two linear factors; we can assume without loss of generality that these factors of the form $y \pm p(x)$ for some $p(x) \in K$ (why?). Then $p^2 = x^3 - x \in K$, and there are many ways to see why this can't happen. One possible approach is to note that although $x^3 - x$ is not squarefree in general (when $\operatorname{ch} K = 2$), the power of x dividing $x^3 - x$ is still exactly one, and in particular odd. Therefore, if we use that $k[x]$ is a UFD to write $p = r/s$ for some coprime $r, s \in k[x]$ with $s \neq 0$, then $r^2 = x(x^2-1)s^2$ leads to a contradiction to unique factorization.

Again, over $k = \mathbb{R}$, the curve C_f of Example 1.7.15 has two components. Again, however, $C_f(\mathbb{C})$ is a punctured torus (hence connected, even irreducible) and the two components visible in $C_f(\mathbb{R})$ are vestiges of slicing this torus and the fact that $\mathbb{R}$ is not algebraically closed. Finally, an argument identical to that in the proof of Lemma 1.7.14 shows that neither of the pieces of $C_f(\mathbb{R})$ are algebraic curves by themselves.

1.7.4 A Sneak Peek at Curve Intersections

Given two curves $C, D \subset \mathbb{A}^2_k$, in how many points do C and D intersect? Well, they could share a component and have infinitely many points in common, but at least when they don't share a component this intersection is finite (this was Theorem 1.7.11). A little experimenting seems to suggest that if C and D are curves of degree m and n respectively, then C and D usually intersect in mn points, but this is not always true. For instance:

(a) When $k = \mathbb{R}$, the parabola C_f defined by $f(x,y) = y - x^2$ and the line C_ℓ defined by $\ell(x,y) = y - x + 1$ do not intersect at all, since $x^2 - x + 1 \in \mathbb{R}[x]$ has no real roots. However, this problem doesn't really appear over algebraically closed fields such as $k = \mathbb{C}$.

(b) Even over fields such as $k = \mathbb{C}$, we have to account for tangency. For instance, if we take $f(x,y) = y - x^2$ again and $\ell(x,y) := y - 2x + 1$, then the polynomial $x^2 - 2x + 1 = (x-1)^2 \in \mathbb{C}[x]$ still has only one root over $\mathbb{C}$. This is because this line C_ℓ is tangent to the parabola, and should really count as having "intersection multiplicity" two.

(c) Finally, even if we account for intersection multiplicites, we can have asymptotes or parallel lines. For instance, the lines defined by $\ell_1(x,y) := y - x$ and $\ell_2(x,y) = y - x + 1$ never intersect in $\mathbb{A}^2_k$ for any field k because they are "parallel". To rectify this situation, we need to account for intersections "at infinity".

As it turns out, these are the only four problems. Our eventual goal is to show the theorem of Bézout (Theorem 1.14.1) which says that if k is an algebraically closed field, then any two projective plane curves $C, D \subset \mathbb{P}^2_k$ of degrees $m, n \geq 1$ respectively that do not share a common component intersect in exactly mn points, when counted with multiplicity. Over the next few lectures, we'll develop tools to prove this theorem, starting with smoothness and intersection multiplicity.