Contents

0 Elliptic Curves and Modular Forms

These are lecture notes for the graduate course Math 7359: Elliptic Curves and Modular Forms, taught at Northeastern in Fall 2026.

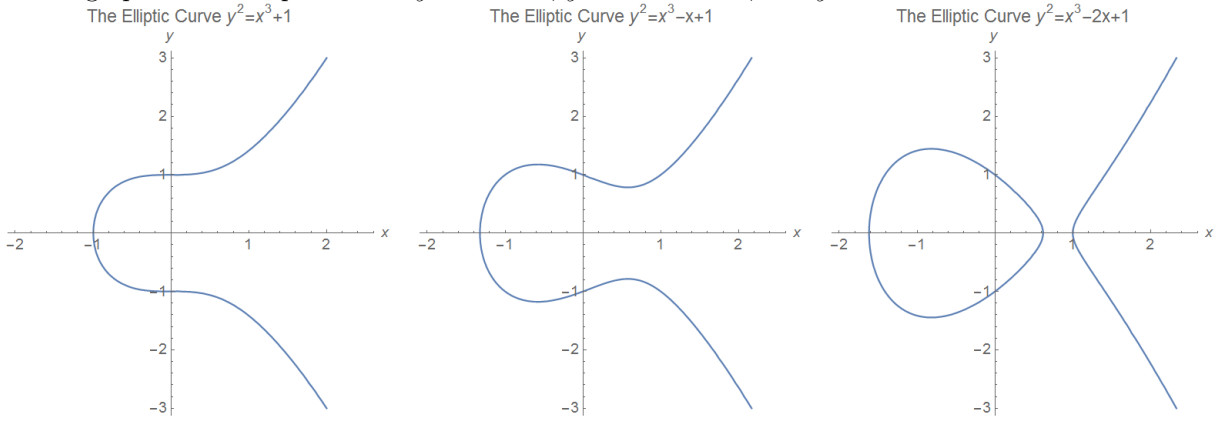
0.1 (Sep 9) Overview + The Group Law

- The goal of this course is to give an overview of elliptic curves and modular forms, highlighting in particular the modern development of modularity, which establishes a very deep connection between these two otherwise very different classes of objects.
    - Elliptic curves are algebraic curves of genus 1 that arise in a wide variety of contexts in mathematics and their study involves techniques from nearly every discipline: algebra, analysis, geometry, topology, and (of course) number theory.
    - Modular forms are analytic functions on the complex upper half-plane satisfying a certain functional equation, and although they are intrinsically analytic objects, they turn out to have surprisingly deep connections to the (seemingly far more) algebraically-flavored elliptic curves.
    - In particular, the connection between elliptic curves and modular forms is central to Wiles's proof of Fermat's Last Theorem, and one of the goals at the end of the course is to elucidate some of the major ideas of this connection.
- In elementary coordinate geometry, one begins by studying the behavior of lines in the plane, which have the general equation $ax+by+c=0$, and then afterwards studies quadratic curves (i.e., the conic sections) having the general equation $ax^2+bxy+cy^2+dx+ey+f=0$.
    - In each case, we often perform simple algebraic manipulations and changes of variable to put the equations into a more standard form.
    - For example, if $b \neq 0$, we can rewrite $ax+by+c=0$ as $y=(-a/b)x+(-c/b)$, which for $m=-a/b$ and $b'=-c/b$ has the more familiar form $y=mx+b'$.
    - Similarly, if $a \neq 0$, we can perform a change of variable $x'=y+(b/(2a))x$ in the equation $ax^2+bxy+cy^2+dx+ey+f=0$ to remove the cross term bxy, and then we can complete the square in x and in y and then rescale the variables to obtain an equation of the form $x^2 \pm y^2 = 1$ or $y=x^2$, depending on which quadratic coefficients are zero.

- Our goal now is to study cubic curves in the plane, which have the general form $ax^3 + bx^2y + cxy^2 + dy^3 + ex^2 + fxy + gy^2 + hx + iy + j = 0$.
    - Like in the case of quadratic curves above, we can perform various changes of variable to reduce the general form to a simpler one.
    - We will not give the full details of the procedure now, as it relies on some facts about cubic curves that we will prove later.
    - Instead, we will summarize matters by saying that as long as the equation is actually cubic (i.e., it is not the case that all of a, b, c, d are zero), then the general equation above can always be transformed using rational changes of variable into one of the form $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$, for appropriate coefficients a_1, a_2, a_3, a_4, a_6.

- Definition: An elliptic curve E over a field K is a curve having an equation of the form $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$, for appropriate coefficients a_1, a_2, a_3, a_4, a_6 in K. This expression is called the Weierstrass form of E.
    - This expression is not the simplest possible one: as long as the characteristic of K is not 2 or 3, we can simplify it further by completing the square in y and completing the cube in x.
    - Explicitly, if we set $y' = y + (a_1/2)x + (a_3/2)$ and $x' = x + (a_2/3)$, we can reduce the Weierstrass equation above to one of the form $(y')^2 = (x')^3 + A(x') + B$.
    - An elliptic curve having an equation of the form $y^2 = x^3 + Ax + B$ is sometimes said to be in "reduced" Weierstrass form.
    - This reduced form is much more amenable for computations, and (in fact) it is nearly unique: the only change of variables that preserves it is one of the form $x = u^2x'$, $y = u^3y'$ for some nonzero u, from which we see that $A = u^4A'$ and $B = u^6B'$.

- Here are graphs of the elliptic curves $y^2 = x^3 + 1$, $y^2 = x^3 - x + 1$, and $y^2 = x^3 - 2x + 1$ over $\mathbb{R}$:



    - In general, we can see that the graph of an elliptic curve $y^2 = x^3 + Ax + B$ will always be symmetric about the x-axis, since if (x, y) satisfies the equation then so does $(x, -y)$. The graph will also have either one or two connected components according to the number of real roots that $x^3 + Ax + B$ has.
    - Exercise: Show that graph of an elliptic curve over $\mathbb{R}$ will have two components when the polynomial $x^3 + Ax + B$ has three distinct real roots, and will have one component otherwise.
    - Notice also that the tangent line at each crossing of the x-axis is vertical for each curve above. Using implicit differentiation, we can compute $y' = \dfrac{3x^2 + A}{2y}$: thus, we see that $y' = \infty$ when y is zero, provided that $3x^2 + A$ is not also zero. This behavior can only occur when $x^3 + Ax + B$ has a root in common with its derivative $3x^2 + A$, which is in turn equivalent to saying that $x^3 + Ax + B$ has a double root.

- Definition: If the polynomial $x^3 + Ax + B$ has a repeated root, we say that the elliptic curve $y^2 = x^3 + Ax + B$ is singular. Otherwise (if the roots are distinct) we say the elliptic curve is nonsingular. A curve is singular if and only if its discriminant $\Delta = -16(4A^3 + 27B^2)$ is zero.

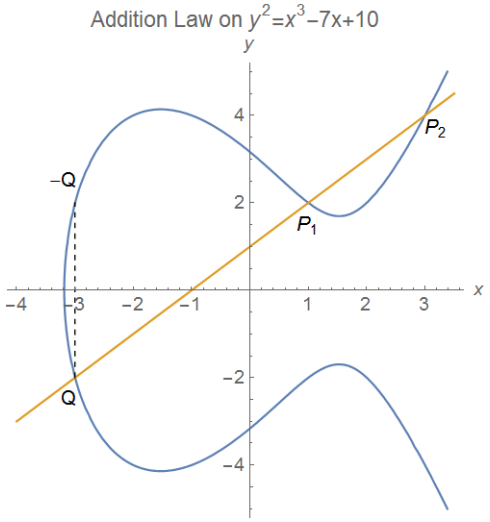
- The second statement follows from the observations above: the polynomial $x^3 + Ax + B$ has a repeated root if and only if it has a root in common with its derivative $3x^2 + A$. This occurs precisely when $x^2 = -A/3$, from which we see that $x(2A/3) + B = 0$ so $x = -3B/(2A)$: then substituting for x yields $\Delta = 0$ almost immediately.
- Remark: The presence of the constant -16 is superfluous here, but there is also a definition of Δ in terms of the original coefficients a_1, a_2, a_3, a_4, a_6 for a general Weierstrass form. To avoid having denominators in that expression, we end up needing an extra factor of -16 in the one we gave above.

- The core property of elliptic curves that makes them so interesting is that if we have two points that lie on the curve, we can use them to construct a third point on the curve.
  - Explicitly, suppose $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ are two distinct points on an elliptic curve E: $y^2 = x^3 + Ax + B$.
  - Draw the line through P_1 and P_2: we claim that this line L must intersect E in a third point Q.
  - To see this, suppose the line through P_1 and P_2 has equation $y = mx + b$. (We are tacitly excluding the possibility that the line is vertical, but we will come back to this case in a moment.)
  - Then the intersection points between L and E are the solutions to the system $y = mx + b$ and $y^2 = x^3 + Ax + B$. Equivalently, we must solve $(mx+b)^2 = x^3 + Ax + B$, or $x^3 + (-m^2)x^2 + (A - 2mb)x + (B - b^2) = 0$.
  - However, we already know that this cubic has the two roots $x = x_1$ and $x = x_2$, so it must have a third root: this gives us the third point Q we wanted.

- Once we construct a third point on an elliptic curve this way, we might try to find more points.
  - If we try this procedure directly using our points P_1, P_2, and Q, however, we will not get anywhere: the line through any of these two points intersects the elliptic curve at the other point.
  - However, we can also exploit the vertical symmetry of the curve to make new points: if $P = (x, y)$ lies on the curve, then the point $-P = (x, -y)$ also lies on the curve.
  - If we combine these two procedures, we can often generate many points on the curve starting from just two.

- Definition (Group Law I): If P_1 and P_2 are two distinct points on the elliptic curve $E : y^2 = x^3 + Ax + B$, let $Q = (x', y')$ be the third intersection point of E with the line L joining P_1 and P_2. We define the sum $P_1 + P_2$ to be the point $-Q = (x', -y')$.
  - It is not immediately clear why we define the sum of two points to be the reflection of Q rather than Q itself. This will become clearer in a moment.
  - Note that if we attempt to add two points which are vertical reflections of one another on the graph of $y^2 = x^3 + Ax + B$, the resulting line will not intersect the curve again.
  - To remedy this, we declare that the curve also includes a point at ∞, which we denote simply as ∞, that we consider as lying on any vertical line. (What we are really doing here is working with the projective model of the curve, rather than the affine one.)

- Example: Given the points $P_1 = (1, 2)$ and $P_2 = (3, 4)$ on the elliptic curve $y^2 = x^3 - 7x + 10$, find the sums $P_1 + P_2$ and $(P_1 + P_2) + P_2$.
  - It is easy to verify that both points lie on the curve. Here is a plot of the curve and the line $y = x + 1$ through the two points:

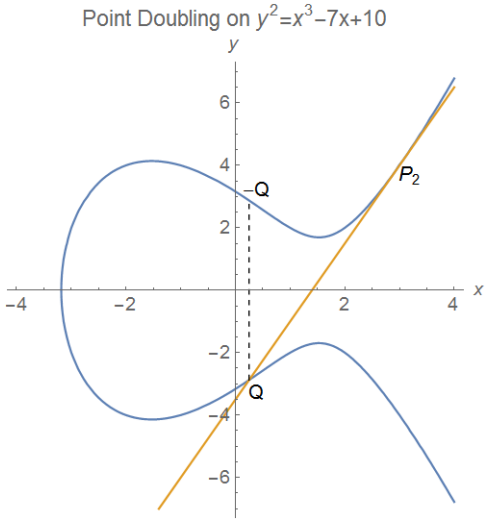


- The point Q lies on the intersection of $y = x + 1$ and $y^2 = x^3 - 7x + 10$, so $(x + 1)^2 = x^3 - 7x + 10$.
- This equation is equivalent to $x^3 - x^2 - 9x + 9 = 0$, which factors as $(x - 1)(x - 3)(x + 3) = 0$. Then the x-coordinate of Q is -3 so $Q = (-3, -2)$.
- Thus, the sum $P_1 + P_2$ is the vertical reflection of Q, which is $\boxed{(-3, 2)}$.
- To find the sum $(P_1 + P_2) + P_2$ we perform a similar procedure: the line through $P_1 + P_2$ and P_2 has equation $y = \frac{1}{3}x + 3$.
- Then we must solve $(\frac{1}{3}x + 3)^2 = x^3 - 7x + 10$, or $x^3 - \frac{1}{9}x^2 - 9x + 1 = 0$.
- Factoring yields $(x - \frac{1}{9})(x + 3)(x - 3) = 0$, so $Q' = (\frac{1}{9}, \frac{82}{27})$, and thus $(P_1 + P_2) + P_2 = \boxed{(\frac{1}{9}, -\frac{82}{27})}$.

- We would also like to be able to add a point to itself.
  - It is straightforward to see from our definition that if P_1 and P_2 are distinct points, then $P_1 + P_2$ is a continuous function of the coordinates of the points.
  - If we are working over $\mathbb{R}$, or another field where limits exist, we could define the addition $P + P$ to be the limit as $P_1 \to P$ of sums $P + P_1$. Geometrically, the lines used in the construction also have a limit as $P \to P_1$: they approach the tangent line to the curve E at the point P.
  - Since we want a definition over any field, we define $P + P$ by letting L be the tangent line to E at P, and then taking Q to be the third point of intersection of L with E.

- Definition (Group Law II): If P is any point on the elliptic curve $E : y^2 = x^3 + Ax + B$, let $Q = (x', y')$ be the third intersection point of E with the tangent line L to E at P. We define the sum $P + P$ to be the point $-Q = (x', -y')$.

- Example: Given the points $P_1 = (1, 2)$ and $P_2 = (3, 4)$ on the elliptic curve $y^2 = x^3 - 7x + 10$, find the sums $P_2 + P_2$ and $(P_1 + P_2) + P_2$.
  - Differentiating implicitly yields $2yy' = 3x^2 - 7$ so that $y' = (3x^2 - 7)/(2y)$. Thus, the tangent line to E at P_2 has slope $\frac{5}{2}$ and its equation is $y = \frac{5}{2}x - \frac{7}{2}$.
  - Here is a plot of the curve and the tangent line at P_2:



- The point Q lies on the intersection of $y = \frac{5}{2}x - \frac{7}{2}$ and $y^2 = x^3 - 7x + 10$, so $(\frac{5}{2}x - \frac{7}{2})^2 = x^3 - 7x + 10$.
- This equation is equivalent to $x^3 - \frac{25}{4}x^2 + \frac{21}{2}x - \frac{9}{4} = 0$, which factors as $(x - \frac{1}{4})(x - 3)(x - 3) = 0$.
- Then the x-coordinate of Q is $1/4$ so $Q = (\frac{1}{4}, -\frac{23}{8})$, and so $P_2 + P_2 = \boxed{(\frac{1}{4}, \frac{23}{8})}$.
- To find the sum $P_1 + (P_2 + P_2)$ we then find the sum of $P_1 = (1, 2)$ with $(\frac{1}{4}, \frac{23}{8})$. The line through these points is $y = -\frac{7}{6}x + \frac{19}{6}$.
- Then we must solve $(-\frac{7}{6}x + \frac{19}{6})^2 = x^3 - 7x + 10$, which has solutions $x = \frac{1}{9}, \frac{1}{4}, 1$.
- Then $Q' = (\frac{1}{9}, \frac{82}{27})$, and thus $P_1 + (P_2 + P_2) = \boxed{(\frac{1}{9}, -\frac{82}{27})}$.

- Note that in the previous two examples, we computed $(P_1 + P_2) + P_2 = (\frac{1}{9}, -\frac{82}{27}) = P_1 + (P_2 + P_2)$, and so we see in this case that the addition law is actually associative. Much more is true:
- Theorem (Group Law): If K is any field and E is any elliptic curve defined over K, then under the addition law defined above, the set of K-valued points on E forms an abelian group with identity ∞ and with the inverse of any point P given by $-P$.
  - We will give arguments for an elliptic curve of the form $y^2 = x^3 + Ax + B$, but the theorem holds in full generality for any elliptic curve.
  - Proof: The addition law is commutative, since the line used in computing $P_1 + P_2$ and $P_2 + P_1$ is the same in each case.
  - To see that ∞ is an identity, consider the sum $P + \infty$. The line passing through P and ∞ is the vertical line through P which also intersects E at the point $-P$. Then by the geometric definition, $P + \infty = -(-P) = P$.
  - To see that $-P$ is an inverse of P, observe that the line passing through P and $-P$ is a vertical line, so the other point on it is ∞. The reflection of ∞ is also ∞, so $P + (-P) = \infty$.
  - Associativity of the addition law is the only nontrivial result in this theorem. It can be done with a tedious algebraic computation using explicit formulas for the addition law (see below).
  - We give another argument using the following consequence of basic linear algebra: if C_1 and C_2 are two distinct plane cubics intersecting in 9 points, then any other cubic D passing through 8 of them must be a linear combination of C_1 and C_2 hence also pass through the 9th point.

- To prove this we merely observe that the vector space of all equations of cubic curves is 10-dimensional and each point imposes one linear condition, so the space of equations passing through 8 given points is 2-dimensional by the nullity-rank theorem (this also requires checking that the linear conditions are independent, which we omit). Since C_1 and C_2 are distinct their equations yield a basis for this space, and so any other cubic D is a linear combination. In particular, then, since the 9th point satisfies the equation for both C_1 and C_2, it also satisfies the equation for any linear combination.
- So suppose P_1, P_2, P_3 are points on an elliptic curve E. Construct the following lines:
  L_1 through P_1, P_2, S $\quad$ L_2 through $-S$, P_3, T $\quad$ L_3 through ∞, U, $-U$
  M_1 through $\infty, S, -S$ $\quad$ M_2 through P_2, P_3, U $\quad$ M_3 through $-U$, P_1, T'
- Then $-T = (P_1 + P_2) + P_3$ and $-T' = P_1 + (P_2 + P_3)$, so we wish to show that $T = T'$.
- Let C_1 be the cubic $L_1L_2L_3$ and C_2 be the cubic $M_1M_2M_3$ (by which we mean, write down the cubic equation $(ax+by+c)(a'x+b'y+c')(a''x+b''y+c'') = 0$ obtained by multiplying together the corresponding equations for the lines).
- Then C_1 and E both pass through the 9 points P_1, P_2, P_3, S, $-S$, ∞, U, $-U$, and T. Since C_2 also passes through the first 8 of these points, it must also pass through the 9th, which is T.
- But since C_2 and E can only intersect in at most 9 points by Bézout's theorem[1], and these 9 points are P_1, P_2, P_3, S, $-S$, ∞, U, $-U$, and T', we must have $T' = T$.

- For convenience in doing numerical computations, we will also write down the general formula for the group law on any reduced Weierstrass curve:

- Proposition (Explicit Group Law): Let $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ be points on the elliptic curve $E : y^2 = x^3 + Ax + B$. Then $P_1 + P_2 = (x_3, y_3)$ where $x_3 = m^2 - x_1 - x_2$ and $y_3 = -m(x_3 - x_1) - y_1$, with $m = \begin{cases} (y_2 - y_1)/(x_2 - x_1) & \text{if } P_1 \neq P_2 \\ (3x_1^2 + A)/(2y_1) & \text{if } P_1 = P_2 \end{cases}$. If m is infinite, then $P_1 + P_2 = \infty$.
  - Observe in particular that the addition formula is rational, in the sense that the result is always a rational function of the inputs. In particular, the sum of two points whose coordinates lie in a field K will also lie in K.
  - Proof: If $P_1 \neq P_2$ then the line joining P_1 and P_2 has equation $y - y_1 = m(x - x_1)$ where $m = \dfrac{y_2 - y_1}{x_2 - x_1}$.
  - We therefore obtain the equation $(mx - mx_1 + y_1)^2 = x^3 + Ax + B$, which has the form $x^3 - m^2x^2 + Cx + D = 0$ for appropriate constants C and D.
  - The polynomial $x^3 - m^2x^2 + Cx + D$ must factor as $(x - x_1)(x - x_2)(x - x_3)$, so upon multiplying out we see that $x_1 + x_2 + x_3 = m^2$. This yields the stated value of x_3, and then $y_3 = m(x_3 - x_1) + y_1$ (where we have multiplied by -1 to account for the vertical reflection).
  - If $P_1 = P_2$ then everything is the same, except instead m is the slope of the tangent line at P_1. By implicit differentiation, we see that $2yy' = 3x^2 + A$ so $m = \dfrac{3x_1^2 + A}{2y_1}$ here, as claimed.

- Example: If $P_1 = (1, 3)$ and $P_2 = (0, 2)$ on the elliptic curve $y^2 = x^3 + 4x + 4$ over $\mathbb{F}_5$ (i.e., with coefficients mod 5), find $P_1 + P_2$ and $P_1 + P_1$.
  - We simply apply the appropriate formulas: adding $Q_1 = (x_1, y_1)$ to $Q_2 = (x_2, y_2)$ produces (x_3, y_3) where $x_3 = m^2 - x_1 - x_2$ and $y_3 = -m(x_3 - x_1) - y_1$, and $m = \begin{cases} (y_2 - y_1)/(x_2 - x_1) & \text{if } Q_1 \neq Q_2 \\ (3x_1^2 + A)/(2y_1) & \text{if } Q_1 = Q_2 \end{cases}$.
  - With $(x_1, y_1) = (1, 3)$ and $(x_2, y_2) = (0, 2)$ we obtain $m = \dfrac{2 - 3}{0 - 1} = 1$, so $x_3 = 0$ and $y_3 = -1(0-1) - 3 = 3$, so $P_1 + P_2 = \boxed{(0, 3)}$.

[1] Bézout's theorem states that two plane curves of degrees m and n not sharing a common component will intersect in mn points over an algebraically closed field, counting multiplicities. Applied when $m = n = 3$, we see that two plane cubics intersect in 9 points (over an algebraically closed field, counting multiplicities).

- Likewise, with $(x_1, y_1) = (x_2, y_2) = (1, 3)$ we obtain $m = \dfrac{3+4}{2 \cdot 3} = 2$, so $x_3 = 2$ and $y_3 = -2(2-1)-3 = 0$, so $P_1 + P_1 = \boxed{(2, 0)}$.

- We will also remark that there are formulas for the addition law on a more general elliptic curve $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$.
  - We use the same geometric construction as for an elliptic curve in reduced Weierstrass form $y^2 = x^3 + Ax + B$, namely, by taking the additive inverse of a point P to be the other point on the (vertical) line joining P to ∞, and by taking the sum $P_1 + P_2$ to be the additive inverse of the other point on the line joining P_1 and P_2 (which is the tangent line when $P_1 = P_2$).

- (Tedious) Exercise: Suppose E is an elliptic curve with a Weierstrass equation $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ with $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ on E. Show that the additive inverse is given by $-P_1 = (x_0, -y_0 - a_1x_0 - a_3)$, and the sum $P_1 + P_2$ is given by ∞ when $x_1 = x_2$ and $y_1 = -y_2 - a_1x_2 - a_3$ and by (x_3, y_3) where $x_3 = m^2 + a_1m - a_2 - x_1 - x_2$ and $y_3 = -(m + a_1)x_3 - b - a_3$ where $y = mx + b$ is the line joining P_1 and P_2 (or the tangent line when $P_1 = P_2$), which explicitly has $m = \dfrac{y_2 - y_1}{x_2 - x_1}$, $b = \dfrac{x_2y_1 - x_1y_2}{x_2 - x_1}$ when $P_1 \neq P_2$ and has $m = \dfrac{3x_1^2 + 2a_2x_1 + a_4 - a_1y_1}{2y_1 + a_1x_1 + a_3}$ and $b = \dfrac{-x_1^3 + a_4x_1 + 2a_6 - a_3y_1}{2y_1 + a_1x_1 + a_3}$ when $P_1 = P_2$.

- Now that we have established the group law, we can now (attempt to) compute the abelian group structure of the set of points on a given elliptic curve over a field K.

- Definition: If E is an elliptic curve with coefficients lying in a field K, we define the set $E(K)$, the K-rational points on E, to be the set of points on E whose entries lie in K, along with the point ∞.
  - When K is finite, clearly $E(K)$ must also be finite, in which case (in principle) we can simply list all of the elements of $E(K)$ and write down the group structure explicitly.

- Example: Find all of the points on the elliptic curve $y^2 = x^3 + 4x + 4$ over $\mathbb{F}_3$ and identify the group structure explicitly.
  - By simply computing $x^3 + 4x + 4$ for each $x \in \mathbb{F}_3$ and testing which are squares we can see that there are 4 points on E: $(0, 1)$, $(0, 2)$, $(1, 0)$, and ∞.
  - The group of points is therefore either cyclic and isomorphic to $\mathbb{Z}/4\mathbb{Z}$, or isomorphic to the Klein 4-group $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$, but since $(0, 1) + (0, 1) = (1, 0)$ is not the identity, in fact the group must be cyclic and generated by $(0, 1)$.

- Exercise: Pick an elliptic curve in Weierstrass form (e.g., $y^2 = x^3 + 4x + 1$) and after checking whether it is nonsingular, find all of its points over $\mathbb{F}_3$, $\mathbb{F}_5$, $\mathbb{F}_7$, $\mathbb{F}_{11}$, and $\mathbb{F}_{13}$, and identify the group structure explicitly in each case.

- We have some additional notation useful when computing orders of elements:

- Definition: If P is a point on an elliptic curve E, we define the multiples of P as $[n]P = \underbrace{P + P + \cdots + P}_{n \text{ terms}}$ for positive integers n, with $[0]P = \infty$ and $[-n]P = -[n]P$. The subgroup of $E(K)$ generated by P is then simply the set $\{[n]P : n \in \mathbb{Z}\}$, and as usual the order of P is the cardinality of this set.
  - Trivially, ∞ is the only point of order 1 on E.
  - The first nontrivial case is to identify the points of order 2: these points satisfy $P+P = \infty$. Geometrically, this means that the tangent line to the graph of E at P passes through $-\infty = \infty$, meaning that the tangent line at P is vertical. From the explicit formula $2yy' = 3x^2 + A$ we see that this is, in turn, equivalent to saying that $y = 0$.
  - Therefore, the points (x, y) of order 2 are those having $y = 0$. Since this requires $x^3 + Ax + B = 0$, we see that there are at most 3 such points.

- For points of order 3, we see that such points P satisfy $P+P+P=\infty$ so that $P+P=-P$, which means that the third intersection point of the tangent line to E at P also goes through P. Equivalently, this says that the point P is an inflection point of the curve.

- We can be more precise if we work with the subgroup of $E(K)$ consisting of all m-torsion points together:

- Definition: The m-torsion subgroup of an elliptic curve E defined over K is the kernel of the multiplication-by-m map (i.e., the points $P \in E(K)$ with $[m]P = \infty$) and is denoted $E_K[m]$.

  - Later, we will also be interested in the full group of m-torsion points on E when we consider E as being defined over the algebraic closure $\overline{K}$: this full m-torsion group is denoted $E[m]$.

- Example: Find the points of order 2 on the elliptic curve $E : y^2 = x^3 + x$ over $\mathbb{Q}$ and over $\mathbb{C}$, and identify the group structure of the 2-torsion group $E[2]$ over each field.

  - From the discussion above, the 2-torsion points are the points with $y = 0$, which requires $x^3 + x = 0$ so that $x = 0$ or $x = \pm i$.
  - Over $\mathbb{Q}$, there is therefore one 2-torsion point $\boxed{(0,0)}$. Then the 2-torsion group $E_{\mathbb{Q}}[2]$ is $\{\infty, (0,0)\}$ and its group structure is isomorphic to $\mathbb{Z}/2\mathbb{Z}$.
  - Over $\mathbb{C}$ we have three 2-torsion points: $\boxed{(0,0),\ (i,0),\ (-i,0)}$. Then the 2-torsion group $E_{\mathbb{Q}}[2]$ is $\{\infty, (0,0), (i,0), (-i,0)\}$ Since all of the nontrivial elements in this group have order 2, the group structure is isomorphic to the Klein 4-group $V_4 \cong (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$.

- For points of higher order, it is more difficult to give nice geometric or algebraic descriptions of $E[m]$ directly from the definition.

  - As we will show later using complex lattices, for any elliptic curve defined over (a subfield of) $\mathbb{C}$, the m-torsion subgroup $E[m]$ is always isomorphic to $(\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/m\mathbb{Z})$, and the same is true more generally in any field of characteristic zero.
  - Exercise: Show that for an elliptic curve $y^2 = x^3 + Ax + B$ over $\mathbb{C}$, the 2-torsion subgroup $E[2]$ is always isomorphic to the Klein 4-group $V_4 \cong (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$.
  - Exercise: Show that for an elliptic curve $y^2 = x^3 + Ax + B$ over $\mathbb{C}$, the 3-torsion subgroup $E[3]$ is always isomorphic to $(\mathbb{Z}/3\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z})$. [Hint: If $P = (x, y)$, write out the x-coordinate of $2P$.]

0.2 (Sep 14) The Nagell-Lutz Theorem and Mazur's Theorem

- The following theorem of Nagell and Lutz provides a convenient way to calculate the torsion points on any elliptic curve over $\mathbb{Q}$:

- Theorem (Nagell-Lutz): Suppose E is an elliptic curve over $\mathbb{Q}$ with Weierstrass equation $y^2 = x^3 + Ax + B$ where A and B are integers, and let $D = -4A^3 - 27B^2$ be the reduced discriminant of E. If $P = (x, y)$ is a rational point of finite order, then x and y are integers. Furthermore, either $y = 0$ or y^2 divides D.

  - Exercise: Show that by making an appropriate change of variables, any rational Weierstrass form can be converted into one with A, B integers. Illustrate by finding a Weierstrass form with integer coefficients for $y^2 = x^3 + \frac{3}{2}x + \frac{2}{5}$.
  - We emphasize here that the Nagell-Lutz theorem is not an if-and-only-if: there can exist points (x, y) with y dividing D that do not have finite order.
  - To prove the theorem we will assemble a few preliminary lemmas.

- Lemma 1: Let $P = (x, y)$ be a point on an elliptic curve E over $\mathbb{Q}$ such that P and $[2]P$ both have integral coordinates. Then either $y = 0$ or y^2 divides the reduced discriminant of E.

  - Proof: If $[2]P = \infty$ then from our discussion of 2-torsion points we see that $y = 0$.

- Otherwise assume $[2]P \neq \infty$. If E has Weierstrass equation $y^2 = x^3 + Ax + B$, then by the explicit group law formula the x-coordinate of $[2]P$ is $\dfrac{(3x^2+A)^2}{4y^2} - 2x$. Since this quantity is an integer by hypothesis, we must have $y|(3x^2 + A)$.
- Now we invoke the identity $D = 27(x^3 + Ax - B)(x^3 + Ax + B) - (3x^2 + 4A)(3x^2 + A)^2$, which can be derived by applying the Euclidean algorithm in $\mathbb{Z}[x]$ to $x^3 + Ax + B$ and the square of its derivative $(3x^2 + A)^2$.
- Since y^2 divides both $x^3 + Ax + B = y^2$ and $(3x^2 + A)^2$, it divides Δ, as claimed.

- We now record some facts about the coordinates of $[m]P$:

- (Tedious) Exercise: Let E be an elliptic curve and $P = (x, y)$ be a point on E. Define the polynomials $\varphi_0 = 0$, $\varphi_1 = 1$, $\varphi_2 = 2y$, $\varphi_3 = 3x^4 + 6Ax^2 + 12Bx - A^2$, $\varphi_4 = 4y(x^6 + 5Ax^4 + 20Bx^3 - 5A^2x^2 - 4ABx - 8B^2 - A^3)$, and in general $\varphi_{2n+1} = \varphi_{n+2} \cdot \varphi_n^3 - \varphi_{n-1}\varphi_{n+1}$ and $\varphi_{2n} = \dfrac{\varphi_n}{2y} \cdot (\varphi_{n+2}\varphi_{n-1}^3 - \varphi_{n-2}\varphi_{n+1}^2)$ for $n \geq 2$.

  1. With $y^2 = x^3 + Ax + B$, show that φ_n can be written as a polynomial in $\mathbb{Z}[x, A, B]$ when n is odd and can be written as y times a polynomial in $\mathbb{Z}[x, A, B]$ when n is even.
  2. Show that φ_n^2 is a polynomial of degree $n^2 - 1$ in x with leading coefficient n^2 while $x\varphi_n^2 - \varphi_{n-1}\varphi_{n+1}$ is a polynomial of degree n^2 in x with leading coefficient 1.
  3. Show that the coordinates of $[n]P$ are (x_n, y_n) where $x_n = \dfrac{x\varphi_n^2 - \varphi_{n-1}\varphi_{n+1}}{\varphi_n^2}$ and $y_n = \dfrac{\varphi_{n+2}\varphi_{n-1}^2 - \varphi_{n-2}\varphi_{n+1}^2}{4y\varphi_n^3}$.

- Lemma 2: If E is an elliptic curve over $\mathbb{Q}$ and P is a point on E such that $[m]P$ has integral coordinates for some $m \geq 1$, then P itself has integral coordinates.

  - Proof: Suppose that $P = (s/t, y)$ where s and t are relatively prime and $t > 0$. By the exercise above, x_m is the quotient of a monic polynomial of degree n^2 in x by a polynomial of degree at most $n^2 - 1$ in x.
  - This means that the value of the numerator polynomial $x\varphi_n^2 - \varphi_{n-1}\varphi_{n+1}$ is of the form a/t^{n^2} with a, t relatively prime, while the denominator polynomial is of the form b/t^{n^2-1} for some b. But then x_m is of the form $a/(tb)$ which cannot be an integer unless $t = 1$. Then the x-coordinate of P is integral, so then since $y^2 = x^3 + Ax + B$ is an integer and y is rational, y is also integral, as required.

- Lemma 3: If $P = (x, y)$ has rational coordinates on $E : y^2 = x^3 + Ax + B$, then $(x, y) = (q/d^2, s/d^3)$ for some positive integer d and some integers p, r relatively prime to d.

  - Proof: Letting $x = q/r$ and $y = s/t$ in lowest terms with $r, t > 0$ and clearing denominators in $y^2 = x^3 + Ax + B$ yields $s^2r^3 = t^2(q^3 + Ar^2p + Br^3)$.
  - Then r is relatively prime to $q^3 + Ar^2p + Br^3$ hence r^3 divides t^2, and likewise t is relatively prime to s^2 hence t^2 divides r^3. Thus $t^2 = r^3$ so letting $d = t/r$ we see $r = d^2$ and $t = d^3$, as required.

- We can now assemble the results for a proof of Nagell-Lutz:

  - Proof (of Nagell-Lutz): Suppose $P = (x, y)$ is a torsion point on E. We first show that P has integer coordinates.
  - Let p be a prime divisor of m and consider $Q = [m/p]P$, which is a torsion point of order p. By Lemma 2 it suffices to show that $Q = (x_Q, y_Q)$ has integer coordinates, since this would imply that P itself has integer coordinates.
  - If $p = 2$ then $[2]Q = \infty$ which requires $y_Q = 0$ so that $x_Q^3 + Ax_Q + B = 0$ so that x_Q is integral by the rational root test; then y_Q is also necessarily integral since it is rational and its square is the integer $x_Q^3 + Ax_Q + B$.
  - Otherwise suppose p is odd. Since $[p]Q = \infty$, the denominator term φ_p of the x-coordinate must vanish when evaluated at x_Q. But for fixed integers A and B and odd p, φ_p is a polynomial in x with integer coefficients of degree $(p^2 - 1)/2$ and leading coefficient p.

- By Lemma 3, in lowest terms we have $x_Q = q/d^2$ for some integer d, so since $\varphi_p(x) = px^{(p^2-1)/2} + O(x^{(p^2-3)/2})$, we see that $d^{p^2-1}\varphi_p(q/d^2) = p \cdot q^{(p^2-1)} + d^2 \cdot k$ for an integer k. Since this quantity must be zero and p is prime, this requires d^2 to divide p, and hence $d = 1$. This means x_Q is an integer, hence so is y_Q by the argument used above. Thus Q is integral and hence P is integral by Lemma 2.
- By repeating the same argument for $[2]P$ we see that $[2]P$ is also integral. Then, finally, by Lemma 1, we conclude that either $y = 0$ or y^2 divides the reduced discriminant of E, as desired.

- The result of the Nagell-Lutz theorem gives us a very effective way to compute all of the torsion points on E: simply find all possible (x, y) on E where $y = 0$ or y^2 divides D, and then test whether these points have finite order.
  - *A priori*, a rational point P could potentially have very large order, but since the torsion points form a subgroup and we have just listed all of the possible elements of this group, we have an upper bound on the possible order of the group and hence on the possible order of P.
  - More efficiently, to test whether P has finite order, we could simply compute the list $\{P, [2]P, [3]P, [4]P, \dots\}$, or even just $\{P, [2]P, [4]P, [8]P, \dots\}$: if any of the multiples of P fail to land on our list, then P cannot have finite order; otherwise, the multiples of P must necessarily repeat since our list is finite, in which case P (and all of its multiples) does have finite order.

- Example: Find the rational torsion points on the elliptic curve $E : y^2 = x^3 - 4x + 3$ and identify their group structure.
  - Here, we have $A = -4$ and $B = 3$, so the discriminant is $D = -4A^3 - 27B^2 = 13$.
  - Since D is squarefree, the only possible y-coordinates are 0 and ± 1.
  - Testing $y = 0$ (so that $x^3 - 4x + 3 = 0$) yields a single rational solution $x = 1$, giving a 2-torsion point $(1, 0)$.
  - Testing $y = \pm 1$ (so that $x^3 - 4x + 3 = \pm 1$) yields no rational solutions in either case, as the resulting cubic is irreducible.
  - Therefore, we see that there are two rational torsion points on E: $\boxed{(1,0) \text{ and } \infty}$. The torsion group has order 2 and is isomorphic to $\mathbb{Z}/2\mathbb{Z}$.

- Example: Find the rational torsion points on the elliptic curve $E : y^2 = x^3 - 351x + 1890$ and identify their group structure.
  - Here, we have $A = -351$ and $B = 1890$, so the discriminant is $D = -4A^3 - 27B^2 = 2^4 3^{14}$.
  - Then the possible y-coordinates are 0 and $\pm 2^a 3^b$ for $a \in \{0, 1, 2\}$ and $b \in \{0, 1, 2, 3, 4, 5, 6, 7\}$.
  - If $y = 0$ then we obtain three 2-torsion points, namely $(-21, 0)$, $(6, 0)$, $(15, 0)$.
  - For the other 24 possible values of y, some computation yields four additional candidate points: $(-3, \pm 54)$ and $(33, \pm 162)$.
  - With $P = (33, 162)$ we can compute $[2]P = (15, 0)$, $[3]P = (33, -162)$, and $[4]P = \infty$, so this point has order 4.
  - Likewise, with $Q = (-3, 54)$ we can compute $[2]Q = (15, 0)$, $[3]Q = (-3, -54)$, and $[4]Q = \infty$, so this point also has order 4.
  - Thus, there are eight rational torsion points on E: $\boxed{(-3, \pm 54),\ (33, \pm 162),\ (-21, 0),\ (6, 0),\ (15, 0), \text{ and } \infty}$. The torsion group has order 8 and is isomorphic to $(\mathbb{Z}/4\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$, where we can take (a, b) mapping to $[a]P + [b](Q - P)$.

- We can also use the Nagell-Lutz theorem to establish that a given point has infinite order on E.
  - Most obviously, if a rational point does not have integral coordinates, then it is not a torsion point. Even if its coordinates are integral, if its y-coordinate is nonzero and its square does not divide D, then the point cannot be a torsion point.

- Furthermore, even if all of these conditions are satisfied, if we compute $[2]P, [3]P, [4]P, \ldots$ and any of these points have non-integral coordinates or have a nonzero y-coordinate with y^2 not dividing D, then P must have infinite order.

- Example: Show that the elliptic curve $E : y^2 = x^3 + 2$ has infinitely many rational points.
  - Testing small values of x reveals two integral points: $(x, y) = (-1, \pm 1)$.
  - If we take $P = (-1, -1)$, then P could be a torsion point, since its y-coordinate -1 has its square dividing the discriminant $D = -108$.
  - However, we can calculate $[2]P = (17/4, 71/8)$, and so since $[2]P$ does not have integral coordinates, it is not a torsion point, and thus neither is P.
  - This means that P has infinite order, which is to say, all of the points $P, [2]P, [3]P, [4]P, \ldots$ are distinct. Since these all have rational coordinates, we see that E has infinitely many rational points.
  - Indeed (though this is much harder to prove) the group of rational points on E is generated by P.

- It follows from the Nagell-Lutz theorem that the group of rational torsion points on an elliptic curve is always finite, since there are only finitely many points with $y = 0$ or y^2 dividing D.
  - Although it may seem that the group could potentially be arbitrarily large, in fact, it cannot have order greater than 16.
  - The following quite deep theorem of Mazur establishes that there is a fairly small list of possible torsion groups:

- Theorem (Mazur): If E is an elliptic curve, then the number of rational torsion points can be any integer from 1 to 12 inclusive, excluding 11, or 16. More explicitly, there are 15 possible group structures for the rational torsion points: the trivial group (order 1), $\mathbb{Z}/2\mathbb{Z}$ (order 2), $\mathbb{Z}/3\mathbb{Z}$ (order 3), $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$ or $\mathbb{Z}/4\mathbb{Z}$ (order 4), $\mathbb{Z}/5\mathbb{Z}$ (order 5), $\mathbb{Z}/6\mathbb{Z}$ (order 6), $\mathbb{Z}/7\mathbb{Z}$ (order 7), $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/4\mathbb{Z})$ or $\mathbb{Z}/8\mathbb{Z}$ (order 8), $\mathbb{Z}/9\mathbb{Z}$ (order 9), $\mathbb{Z}/10\mathbb{Z}$ (order 10), $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/6\mathbb{Z})$ or $\mathbb{Z}/12\mathbb{Z}$ (order 12), or $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/8\mathbb{Z})$ (order 16).
  - The proof of this theorem involves quite advanced methods: the idea is to study the points on various modular curves and use a (tremendous!) amount of case analysis to eliminate all of the other possible torsion orders and other possible group structures.
  - There also exist infinite families of elliptic curves having each of the groups listed as its torsion group.

- Over finite fields, all points are torsion points since $E(K)$ is finite, so the question of computing $E(K)$ reduces to that of computing the the torsion points. Over infinite fields, however, $E(K)$ can have many linearly independent points of infinite order, which makes the group structure quite a lot more challenging to determine.
  - Indeed, even over $K = \mathbb{Q}$, it can often be quite computationally intensive to compute generators and relations for $E(\mathbb{Q})$, let alone over larger fields K.
  - Modern software packages such as Sage have functionality to compute generators for $E(K)$ when K is a number field.
  - We do not have the tools currently to prove many of these results, but we will give some examples for illustration.

- Example: Consider the elliptic curve $E : y^2 = x^3 + 4x + 1$ over $\mathbb{Q}$.
  - Quite obviously, $P = (0, 1)$ is a rational point on E.
  - We can then compute $[2]P = (4, -9)$, $[3]P = (9/4, 37/8)$, $[4]P = (28/81, -1135/729)$, $[5]P = (2664/49, 137593/343)$, and so forth.
  - Computing larger multiples of P will yield increasingly complicated rational points on E. Indeed, P has infinite order since $[3]P$ has non-integral coordinates, and so these integer multiples of P yield infinitely many distinct rational points on E.
  - In fact, P is actually a generator for the group $E(\mathbb{Q})$, although this is quite a lot harder to prove. As a consequence, the group $E(\mathbb{Q})$ is isomorphic to $\mathbb{Z}$.

- Example: Consider the elliptic curve $E : y^2 = x^3 - 2023x$ over $\mathbb{Q}$.
  - Quite obviously, $P = (0, 0)$ is a rational point on E. Here, however, since $[2]P = \infty$, we see that P is a torsion point.
  - Searching for other small rational points will reveal none, but in fact there are infinitely many points on E as well.
  - As one may verify with a computer, the point $Q = (84676804/180625, -775601419158/76765625)$ also lies on E, and it necessarily has infinite order since its coordinates are non-integral. Indeed, even just evaluating $[2]Q$ is messy by itself: $[2]Q = (523620448448041618543485494416 81/434625338111430357812426490000, -351504781800873988198159892309132758495471097671/2865316246850633237015114603098197911070000000)$.
  - In fact, P and Q generate the group of rational points, so since they are necessarily linearly independent, the group of rational points on E is isomorphic to $\mathbb{Z} \times (\mathbb{Z}/2\mathbb{Z})$.
- Example: Consider the elliptic curve $E : y^2 = x^3 - 43x + 166$ over $\mathbb{Q}$.
  - Searching for small integer points will eventually reveal that $P = (3, 8)$ lies on E.
  - Computing multiples of P yields $[2]P = (-5, -16)$, $[3]P = (11, -32)$, and $[4]P = (11, 32) = -[3]P$: thus P must have order 7.
  - In fact, the multiples of P turn out to be the only rational points on E, meaning that $E(\mathbb{Q})$ is isomorphic to $\mathbb{Z}/7\mathbb{Z}$.
- Example: Consider the elliptic curve $E : y^2 = x^3 - 2x$ over $\mathbb{Q}(i)$.
  - Clearly, $P = (0, 0)$ has order 2, while the points $Q = (i-1, -2)$ and $R = (-1, 1)$ both have infinite order: for R this follows from Nagell-Lutz since $[2]R = (9/4, -21/8)$, but it is more work to show the result for Q since its entries are not rational (one approach is to show that the powers of 2 in the denominators of $[2^n]Q$ grow as n grows).
  - In fact the points Q and R are linearly independent, although this is even more work to show (since for example there is no reason *a priori* that there couldn't exist a relation like $[7]Q + [8]P = \infty$).
  - But since Q and R actually are linearly independent, each of the points $[a]Q + [b]R$ for integers a, b are distinct. As an example, we have $[2]Q + R = (\frac{13+84i}{25}, \frac{-462+709i}{125})$.
  - With even more effort, one may show that P, Q, and R generate $E(\mathbb{Q}(i))$, which is isomorphic to $\mathbb{Z} \times \mathbb{Z} \times (\mathbb{Z}/2\mathbb{Z})$.

0.3 (Sep 17) Mordell's Theorem

- In all of the examples above, the group $E(K)$ was finitely generated. In fact, the group $E(K)$ is always finitely generated whenever K is a number field[2], as shown for $K = \mathbb{Q}$ by Mordell and then extended to all number fields K by Weil:
- Theorem (Mordell-Weil): If K is a number field and E is an elliptic curve defined over K, then the group $E(K)$ is finitely generated.
  - We will not prove this theorem now since it requires substantially more number-theoretic background; even proving Mordell's theorem for the case $K = \mathbb{Q}$, which we will do below, is not at all trivial.
  - By the structure theorem for finitely generated abelian groups, this says $E(\mathbb{Q}) \cong \mathbb{Z}^r \oplus E_{\text{Tor}}(\mathbb{Q})$ where $E_{\text{Tor}}(\mathbb{Q})$ is the set of $\mathbb{Q}$-torsion points of E (i.e., the set of $\mathbb{Q}$-rational points of E having finite order), which is a finite abelian group and thus is a direct sum of cyclic groups.
  - For any given elliptic curve E, the torsion subgroup $E_{\text{Tor}}(\mathbb{Q})$ can be computed using Nagell-Lutz as we discussed above.

[2]Recall that a number field is a finite-degree field extension of $\mathbb{Q}$, meaning that K is finite-dimensional when considered as a vector space over $\mathbb{Q}$. Number fields can all be written as $K = \mathbb{Q}(\alpha)$ for some algebraic number α (i.e., a root of a nonzero polynomial with rational coefficients).

- The quantity r is called the rank of the elliptic curve, and is equal to the number of linearly-independent points one may construct on E. The rank is much more difficult to compute, and there is no known direct algorithm that is guaranteed to compute it (though in practice the rank of most curves can be computed).
- It is not currently known whether elliptic curves over $\mathbb{Q}$ can have an arbitrarily large rank, and the historical consensus has switched back and forth between thinking ranks can be arbitrarily large and thinking that ranks are uniformly bounded above. In 2006 Elkies gave a construction for an elliptic curve that has rank at least 28 (and it is expected this curve has rank exactly 28)[3], and Elkies and Klagsbrun in 2024 constructed an elliptic curve of rank 29. In 2026 similar curves have been discovered of ranks 30 and 31.
- It has also been shown by Bhargava and Shankar in 2015 that the average rank (suitably defined) of an elliptic curve is at most 7/6: the actual average is expected to be 1/2 (with 50% of elliptic curves having rank 0 and 50% having rank 1, asymptotically).

- Here is the structure of the proof of Mordell's theorem (and its generalization by Weil):
  - First, one proves the so-called "weak Mordell-Weil theorem": that for any positive integer m and any number field K, the group $E(K)/mE(K)$ is finitely generated.
  - Of course, the weak Mordell-Weil theorem does not imply the full Mordell-Weil theorem directly, because there are many non-finitely-generated groups G such that G/mG is finitely generated (for example, $\mathbb{Q}$ and $\mathbb{R}$ both have $G/mG = 0$ for all m).
  - The difficulty is that knowing G/mG is finitely generated does not imply G is finitely generated, because G could contain many elements that are divisible by m.
  - The task then is to eliminate this possibility, which can be done using the theory of heights: one defines a "height function", measuring roughly the complexity of a point on the curve, and then shows that the height of large multiples of a point tends to be larger than the height of the original point.
  - One such height function on points $(x, y) = (p_x/q_x, p_y/q_y)$ over $\mathbb{Q}$ is $\max(\log p_x, \log q_x)$: essentially, the maximum number of digits appearing in the numerator or denominator of the x-coordinate.
  - Next, one shows that there are a bounded number of points in $E(K)$ of height less than any fixed bound: thus, any point that is a multiple of m has to be "large" for large m.
  - By fine-tuning the details of this argument, we can deduce that a finite number of generators will suffice to generate the group $E(K)$: the idea is to show that for any point P on E, we may subtract appropriate multiples of the coset representatives of the finite group $E(K)/mE(K)$ to obtain a new point whose height is bounded independently of P. Since there are then only finitely many such points, adding them to our list will yield a finite generating set for $E(K)$.
  - A structurally similar argument works over arbitrary number fields K, but the details are more complicated.

- Before going into the details of Mordell's theorem for $K = \mathbb{Q}$ we make some additional remarks.
  - The proof of weak Mordell(-Weil) is not effective, meaning that it does not yield an actual algorithm guaranteed to compute generators for $E(K)/mE(K)$, even for specific values of m (typically one uses $m = 2$).
  - Various practical computational methods (e.g., those implemented in Sage) have been developed that can provably compute generators for $E(K)$, but they are not always guaranteed to terminate.
  - In order to obtain Mordell(-Weil) it is only necessary to prove the finiteness of $E(K)/mE(K)$ for a single value of m, typically $m = 2$, which we will do in our argument.

- Theorem (Weak Mordell's Theorem): If E is any elliptic curve defined over $\mathbb{Q}$, the group $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite.

[3]The equation of Elkies' curve is $y^2 + xy + y = x^3 - x^2 - 20067762415575526585033208209338542750930230312178956502x + 3448161179503055646703298569039072037485594435931918036126600829629193944873224342 9$

- Before proceeding with the actual argument, we outline the idea. What we will show is that there exists a homomorphism φ with kernel $2E(\mathbb{Q})$ from $E(\mathbb{Q})$ to a finite direct sum of groups of the form $(K^*)/(K^*)^2$ where K is a number field. We then show the image of $E(\mathbb{Q})$ inside each component is finite, which by the first isomorphism theorem implies that $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite.
- To motivate the existence of this homomorphism, suppose that E has a rational 2-torsion point. By translation we may move this point to the origin, in which case E has a Weierstrass equation $y^2 = x^3 + Cx^2 + Dx$.
- If P, Q, and R are three collinear points on E none of which equals ∞, then these three points lie on some line $y = mx+b$. Hence the x-coordinates of P, Q, and $-(P+Q)$ are the three roots of the equation $(mx + b)^2 = x^3 + Cx^2 + Dx$, or equivalently $x^3 + (C - m^2)x^2 + (D - 2mb)x - b^2$, so by the usual root formulas, the product of these three roots is b^2: the square of a rational number.
- Therefore, except at the 2-torsion points, the x-coordinate function $x : E(\mathbb{Q}) \to \mathbb{Q}^*$ satisfies the relation $x(P)x(Q)x(-P - Q) \in (\mathbb{Q}^*)^2$.
- Since the x-coordinate of $-P - Q$ is the same as that of $P + Q$, and all of the quantities are rational numbers, this observation equivalently says that $x(P)x(Q)$ differs by a square factor from $x(P + Q)$, which is a convoluted way of saying that when we descend instead to the group $(\mathbb{Q}^*)/(\mathbb{Q}^*)^2$, the images of $x(P)x(Q)$ and $x(P + Q)$ are equal.
- In other words, the image of the x-coordinate map satisfies the group homomorphism property, as a map from $E(\mathbb{Q})$ to $(\mathbb{Q}^*)/(\mathbb{Q}^*)^2$.
- This does not give a complete description of the map, because we must still handle the situation of ∞ (which we clearly should map to the identity) and the origin $(0, 0)$, which we map to the value of the derivative of the cubic $x^3 + Cx^2 + Dx$ at that point: namely, D. One may then check that these conditions preserve the homomorphism property.
- Now, $2E(\mathbb{Q})$ is certainly contained in the kernel of this homomorphism, since $x(2P) = x(P + P) = x(P)^2 = 1$ inside $(\mathbb{Q}^*)/(\mathbb{Q}^*)^2$ by the homomorphism property, since $x(P)^2$ is a square. However, the kernel will usually be much larger than $2E(\mathbb{Q})$.
- In order to deal with this, we need to exploit the other points of order 2: that requires us to work with order-2 points located other places than the origin, and these points may not even lie in $E(\mathbb{Q})$. In general, if (α, β) has order 2, we instead want to work with the modified map $x^*(P) = x(P) - \alpha$ having image in $(K^*)/(K^*)^2$, where $K = \mathbb{Q}(\alpha)$, and we take the homomorphism to be the image of P under all three of these maps at once.
- A convenient way to package these calculations is instead to consider the polynomial quotient ring $\mathbb{Q}[x]/(x^3 + Ax + B)$, which automatically keeps track of the x-coordinates of the points of order 2, so we will start with this approach.

- Step 1 (Construction of φ): Suppose E has a Weierstrass equation $y^2 = x^3 + Ax + B$ with A, B integers and let $f(x) = x^3 + Ax + B$.
  - Consider the polynomial quotient ring $R = \mathbb{Q}[x]/(f(x))$, which is a $\mathbb{Q}$-algebra of dimension 3.
  - By the Chinese remainder theorem, if the factorization of $f(x)$ over $\mathbb{Q}$ is a product of 3 linear factors then $R \cong \mathbb{Q} \oplus \mathbb{Q} \oplus \mathbb{Q}$, if $f(x)$ is a product of a linear and quadratic term then $R \cong \mathbb{Q} \oplus K$ where $K/\mathbb{Q}$ is a field extension of degree 2 (generated by a root of the quadratic), and if $f(x)$ is irreducible then $R \cong L$ where $L/\mathbb{Q}$ is a field extension of degree 3 (generated by any root of f).
  - Let U be the group of units of the ring R, which are the residue classes in R represented by the polynomials relatively prime to $x^3 + Ax + B$ in $\mathbb{Q}[x]$.
  - We now construct a group homomorphism $\varphi : E(\mathbb{Q}) \to U/U^2$ with kernel $2E(\mathbb{Q})$. Clearly we must take $\varphi(\infty)$ to be the identity element of U/U^2 if this map is to be a homomorphism.
  - Next suppose $P = (\alpha, \beta)$ is a rational point on E with $y \neq 0$. Then the polynomial $x - \alpha$ is relatively prime to $x^3 + Ax + B$ (as α is not a root of this polynomial because $y \neq 0$), so the residue class of $x - \alpha$ lies in U. We define $\varphi(P)$ to be the residue class $\overline{\alpha - x} + U^2 \in U/U^2$.
  - It remains to define φ on the points $(\alpha, 0)$ of order 2. Since $x - \alpha$ divides $x^3 + Ax + B$, we may write $x^3 + Ax + B = (x - \alpha)g(x)$ for a quadratic $g(x)$; then because α is rational and $x^3 + Ax + B$ has no repeated roots, by the Chinese remainder theorem we have $\mathbb{Q}[x]/(f(x)) \cong \mathbb{Q}[x]/(x - \alpha) \oplus \mathbb{Q}[x]/(g(x))$.

Since the first factor $\mathbb{Q}[x]/(x-\alpha)$ is isomorphic to $\mathbb{Q}$ (by the evaluation map $x \mapsto \alpha$), the element $(f'(\alpha) \bmod x-\alpha, \alpha - x \bmod g(x))$ corresponds to a unique unit in U/U^2.

- Step 2: The map φ is a homomorphism.
  - Proof: First, because the definition of φ is independent of the y-coordinate of a point P, we have $\varphi(P) = \varphi(-P)$ for all P.
  - It then suffices to show that if P, Q, R are collinear on E, then $\varphi(P)\varphi(Q)\varphi(R) = 1$ in U/U^2: this will imply that $\varphi(P+Q)\varphi(P)\varphi(Q) = \varphi(P+Q)\varphi(-P)\varphi(-Q) = 1$, and since the square of every element in U/U^2 is 1, this would imply the homomorphism condition $\varphi(P+Q) = \varphi(P)\varphi(Q)$.
  - Now suppose that P, Q, R are distinct and not 2-torsion. If $Q = -P$ then $R = \infty$ in which case $\varphi(P)\varphi(Q)\varphi(R) = \varphi(P)^2 = 1$ in U/U^2.
  - Otherwise, the three points lie on a line $y = mx + b$ in which the three x-coordinates x_P, x_Q, x_R of P, Q, R are the roots of $x^3 + Ax + B - (mx+b)^2 = (x - x_P)(x - x_P)(x - x_R)$: then $\varphi(P)\varphi(Q)\varphi(R)$ is the residue class $\overline{x_P - x} \cdot \overline{x_Q - x} \cdot \overline{x_R - x} = \overline{(mx+b)^2} = (\overline{mx+b})^2 = 1$ because this element is the square of a residue class in U/U^2.
  - If one of the points has order 2 (say P) but the others do not, we again check that the image of the product is a square in the various components of $\mathbb{Q}[x]/(f(x))$. For example, if P has order 2 but Q, R do not, then the same argument as above works in the factor $\mathbb{Q}[x]/(g(x))$, while in the factor $\mathbb{Q}[x]/(x - x_P)$ we have $f'(\alpha) = (x_P - x_Q)(x_P - x_R)$ so the product $\varphi(P)\varphi(Q)\varphi(R) = f'(\alpha)^2$ in the first component. Thus since the product is a square in both components, by the Chinese remainder theorem it is a square in U/U^2 hence equals 1.
  - Finally, if all of the points have order 2, a similar argument works to show that the resulting product is a square in each of the three resulting components of $\mathbb{Q}[x]/(f(x)) \cong \mathbb{Q}[x]/(x - x_P) \oplus \mathbb{Q}[x]/(x - x_Q) \oplus \mathbb{Q}[x]/(x - x_R)$.
- Step 3: The kernel of φ is $2E(\mathbb{Q})$.
  - Proof: First observe that for any $P \in E(\mathbb{Q})$, we have $\varphi(2P) = \varphi(P)^2 = 1$ since squares in U/U^2 are 1. Therefore $2E(\mathbb{Q})$ is contained in $\ker(\varphi)$.
  - For the other containment suppose $\varphi(P) = 1$ for $P = (\alpha, \beta)$. Then $\varphi(P)$ is the residue class of $\alpha - x$ in U/U^2, meaning that $\alpha - x$ is a unit and a square in $\mathbb{Q}[x]/(x^3 + Ax + B)$.
  - Suppose $\alpha - x \equiv (c_1x^2 + c_2x + c_3)^2$ modulo $x^3 + Ax + B$ for some $c_1, c_2, c_3 \in \mathbb{Q}$. Note c_1 must be nonzero (otherwise the congruence would have a linear polynomial congruent to one of degree 0 or 2).
  - One may check that $(-c_1x + c_2)(c_1x^2 + c_2x + c_3) \equiv dx + e$ modulo $x^3 + Ax + B$ for $d = Ac_1^2 + c_2^2 - c_1c_3$ and $e = Bc_1^2 + c_2c_3$. Thus the polynomial $(-c_1x + c_2)^2 - (\alpha - x)(dx + e)^2$ is zero modulo $x^3 + Ax + B$: but since this polynomial is a monic cubic, it must equal $x^3 + Ax + B$.
  - Therefore, $x^3 + Ax + B = (-c_1x + c_2)^2 - (\alpha - x)(dx + e)^2$. Geometrically, this means that the line $y = -c_1x + c_2$ intersects $y^2 = x^3 + Ax + B$ at one point with x-coordinate α (namely, at P or at $-P$) and a double intersection at some other point Q. This means $\pm P = 2Q$, and in either case $P \in 2E(\mathbb{Q})$, as desired.
- Step 4: Proof of the weak Mordell theorem.
  - Using the homomorphism φ we can now finish the proof: by the first isomorphism theorem, $E(\mathbb{Q})/2E(\mathbb{Q})$ is isomorphic to the image of φ inside U/U^2, the group of units modulo squares inside R.
  - Since R is a direct sum of number fields, by the Chinese remainder theorem, the group U/U^2 is a direct sum of groups of the form $(K^*)/(K^*)^2$ where K is a number field. It therefore suffices to show that the projection of the image of φ inside each of these groups $(K^*)/(K^*)^2$ is finite, for then the image of φ itself is finite.
  - We will show the result in the situation where E has rational 2-torsion (i.e., when the roots of $f(x)$ are all rational), in which case $U/2U$ is the direct sum of three copies of $(\mathbb{Q}^*)/(\mathbb{Q}^*)^2$.

- Suppose we are considering the component associated to a 2-torsion point P. By translating P to the origin, we may equivalently work with E having a Weierstrass form $y^2 = x^3 + Cx^2 + Dx$, where (as we showed above during the motivation for the argument) the desired map is simply the x-coordinate map $x(\alpha, \beta) = \alpha$.
- We claim that the image of $(\mathbb{Q}^*)/(\mathbb{Q}^*)^2$ lies inside the subgroup generated by -1 and the prime divisors of D. To see this, suppose $Q = (q/d^2, r/d^3)$ is a rational point on E (recall that we showed all rational points have this form in our proof of Nagell-Lutz; the proof works equally well for arbitrary Weierstrass forms). By rescaling d, we may also assume that q is squarefree, in which case $q = \varphi(Q)$ inside $(\mathbb{Q}^*)/(\mathbb{Q}^*)^2$.
- By clearing denominators we see that $r^2 = q^3 + Cq^2d + Dqd^2$. Let p be a prime dividing q (which necessarily does not divide d): then p divides the right-hand side, so p divides r^2 and hence it divides r. But then $Dqd^2 = r^2 - q^3 - Cq^2d$ is divisible by p^2, and since q is squarefree, it is not divisible by p^2, and d^2 is also not divisible by p, so D must be divisible by p.
- Therefore, all prime divisors of $\varphi(Q)$ divide D, and so $\varphi(Q)$ lies inside the finite subgroup of $(\mathbb{Q}^*)/(\mathbb{Q}^*)^2$ generated by -1 and the prime divisors of D. Since this holds for all Q, we see that the image of φ lies inside this finite subgroup, so it is finite.
- In the other cases, where some of the 2-torsion points of E do not lie in $\mathbb{Q}$, one may adapt this argument inside $(K^*)/(K^*)^2$ to see that there are only finitely many possible values for $\varphi(Q)$ when it is a principal ideal. Then because the ideal class group of K is finite, one can extend this argument to show that there are only finitely many possible values for $\varphi(Q)$ in general. (We omit the details since they require some nontrivial calculations with ideal classes.)

- Now that we have finished the weak Mordell's theorem, we can use the result to prove the full version. To do this we will use the following descent theorem:

- Theorem (Descent Theorem): Suppose that G is an abelian group with a "height function" $h : G \to [0, \infty)$ such that (i) for all nonnegative M, the number of elements $g \in G$ with $h(g) \leq M$ is finite, (ii) for all $g_0 \in G$ there is a constant c_g with $h(g+g_0) \leq 2h(g)+c_g$ for all $g \in G$, (iii) there is a constant d such that $h(2g) \geq 4h(g)-d$ for all $g \in G$, and (iv) $G/2G$ is finite. Then G is finitely generated.

  - Proof: Let S be a set of coset representatives for $G/2G$ and let $P_0 \in G$. Then P_0 lies in one of these cosets, say the coset represented by Q_0, meaning that $P_0 - Q_0 \in 2G$. This means $P_0 - Q_0 = 2P_1$ for some $P_1 \in G$.
  - In the same way, P_1 lies in some coset, say represented by Q_1, so that $P_1 - Q_1 = 2P_2$ for some $P_2 \in G$. By iterating this procedure we obtain a sequence of coset representatives $Q_0, Q_1, Q_2, \ldots$ and elements $P_1, P_2, P_3, \ldots$ such that $P_i - Q_i = 2P_{i+1}$ for each $i \geq 0$.
  - By substituting these equations into one another, we see that $P = Q_0+2Q_1+4Q_2+\cdots+2^nQ_n+2^{n+1}P_{n+1}$ for each $n \geq 0$.
  - Now, by (ii) applied with $g_0 = -Q_i$ we have $h(g - Q_i) \leq 2h(g) + c_i$ for some c_i and all $g \in G$.
  - Since there are only finitely many possible Q_i (namely, the elements in the set S of coset representatives), letting c be the maximum of the corresponding c_i shows that $h(g - Q_i) \leq 2h(g) + c$ for all $g \in G$ and all $i \geq 0$.
  - Then by (iii) applied with $g = P_{i+1}$ we have $4h(P_{i+1}) \leq h(2P_{i+1}) + d = h(P_i - Q_i) + d \leq 2h(P_i) + c + d$ where the last step follows from what we just did above.
  - In particular, when $h(P_i) \geq c + d$, we have $h(P_{i+1}) \leq \frac{3}{4}(c + d) \leq \frac{3}{4}h(P_i)$. In other words, if the height of P_i is large enough, then the height of P_{i+1} necessarily decreases exponentially.
  - We therefore see that there must exist some n with $h(P_{n+1}) \leq c + d$. Let T be the set of g with $h(g) \leq c + d$, which is finite by (i).
  - Then $P = Q_0 + 2Q_1 + 4Q_2 + \cdots + 2^nQ_n + 2^{n+1}P_{n+1}$ is a linear combination of the Q_i (which all lie in S) and P_{n+1} (which lies in T). Since this holds for any point P, we deduce that $S \cup T$ generates G: since S and T are both finite, this means G is finitely generated.

- It remains to show that there exists a height function h on the elliptic curve E that satisfies all of the hypotheses of the descent theorem. The final step is to prove that the height function $h(q/r, s/t) = \log \max(|q|, |r|)$ satisfies all of the requirements.

- For (i), clearly there are only finitely many rational numbers with height $\leq M$, since the numerator and denominator must both be at most e^M in absolute value.
- For (ii), suppose $P = (q/d^2, r/d^3)$ with d relatively prime to q, r and E has a Weierstrass form $y^2 = x^3 + Ax + B$. If P has height h, then $|q| \leq h$ and $d^2 \leq h$, and also $r^2 = \left|q^3 + Aqd^2 + Bd^3\right| \leq Ch^3$ where $C = 1 + |A| + |B|$ is a fixed constant.
- Then from the addition formula, if $P_0 = (x, y)$, the x-coordinate of $P + P_0$ is $(\frac{y - y_0}{x - x_0})^2 - x_0 - x$, which can eventually be simplified to the form $\frac{c_1 y + c_2 x^2 + c_3 x + c_4}{c_5 x^2 + c_6 x + c_7}$ for some integers $c_1, \ldots, c_7$. Setting $x = q/d^2$ and $y = r/d^3$ then yields an expression for the x-coordinate that is a ratio of two integers. Applying the triangle inequality and the bounds above to the numerator and denominator then yield an inequality of the desired form.
- For (iii), we must estimate the height of $2P$ in terms of the height of P. If $P = (x, y)$ then the x-coordinate of $2P$ is $\frac{(3x^2 + A)^2}{4(x^3 + Ax + B)} - 2x = \frac{x^4 - 2Ax^2 - 8Bx + A^2}{4(x^3 + Ax + B)}$.
- The desired height estimate then follows from the following general fact about rational functions given by quotients of relatively prime polynomials: if $f(x)$ and $g(x)$ are relatively prime polynomials in $\mathbb{Z}[x]$ with $d = \max(\deg f, \deg g)$, then there exist a constant C such that $\left|h(\frac{f(x)}{g(x)}) - d\,h(x)\right| \leq C$ for all x.
- Showing this estimate is quite nontrivial, and we will omit the details. Intuitively, however, the idea is to show that there can only be a bounded amount of factor cancellation between $f(x)$ and $g(x)$ for each rational x, and so since $f(x)$ and $g(x)$ are relatively prime, the height of $\frac{f(x)}{g(x)}$ up to some bounded amount is the same as the height of $\frac{x^d}{g(x)}$, which is just $h(x^d) = dh(x)$.
- Finally, (iv) is the weak Mordell's theorem, which we proved above.

- Putting all of these facts together, at last, completes the proof of Mordell's theorem that $E(\mathbb{Q})$ is finitely generated.

0.4 (Sep 21) $\mathbb{A}^n$, Affine Varieties

- The explicit calculations resulting in the group law that we worked out during the last few lectures have a rather *ad hoc* feel to them. Our goal now is to give a more coherent approach to the group law on an elliptic curve, in a way that will make more clear that the existence of the group law is not just some mere computational accident, but rather something that is forced to exist by the structural properties of the curve.
  - To do this we will review some basic facts about the algebraic geometry of plane curves.

- Definition: For a field k, we define affine n-space $\mathbb{A}^n(k) = \{(x_1, x_2, \ldots, x_n) : x_i \in k\}$ to be the set of n-tuples of elements of k. The elements of $\mathbb{A}^n(k)$ are called points.
  - Definition: For $f \in k[x_1, \ldots, x_n]$, we define the vanishing locus of f to be $V(f) = \{P \in \mathbb{A}^n(k) : f(P) = 0\}$, the set of points $P \in \mathbb{A}^n(k)$ where f vanishes. We extend this definition to subsets $T \subseteq k[x_1, \ldots, x_n]$ by setting $V(T) = \cap_{f \in T} V(f) = \{P \in \mathbb{A}^n(k) : f(P) = 0 \text{ for all } f \in T\}$.
  - Exercise: Draw $V(x)$, $V(x^2)$, $V(y - x)$, $V(y - x^2)$, $V(xy)$, $V(x, y)$, and $V(y^2 - x^3 - x)$ in $\mathbb{A}^2(\mathbb{R})$.
  - Definition: For a subset $S \subseteq \mathbb{A}^n(k)$, we define the ideal of functions vanishing on S to be $I(S) = \{f \in k[x_1, \ldots, x_n] : f(P) = 0 \text{ for all } P \in S\}$. It is easy to see that $I(S)$ is an ideal of $k[x_1, \ldots, x_n]$ for any set S.
  - Exercise: Identify $I(S)$ in $\mathbb{R}[x, y]$ for $S = \{(t, 0) : t \in \mathbb{R}\}$, $\{(t^2, t) : t \in \mathbb{R}\}$, $\{(1, 1)\}$, $\{(0, 0), (1, 1)\}$, $\{(\cos t, \sin t) : t \in \mathbb{R}\}$, and $\{(t, \sin t) : t \in \mathbb{R}\}$.

- We have various properties of the maps V and I:

1. If I is the ideal generated by $T \subseteq k[x_1, \ldots, x_n]$, then $V(T) = V(I)$. Thus, we need only consider the behavior of V on ideals, meaning that we will only consider I and V as maps I : [sets] $\to$ [ideals] and V : [ideals] $\to$ [sets].
2. $V(0) = \mathbb{A}^n(k)$, $V(1) = \emptyset$, and $V(x_1 - a_1, \ldots, x_n - a_n) = \{(a_1, \ldots, a_n)\}$.
3. $I(\emptyset) = k[x_1, \ldots, x_n]$, $I(\mathbb{A}^n) = 0$ when k is infinite, and $I(\{(a_1, \ldots, a_n)\}) = (x_1 - a_1, \ldots, x_n - a_n)$.
4. $V(\cup_i I_i) = \cap_i V(I_i)$ and $V(IJ) = V(I) \cup V(J)$.
5. For ideals I and J, if $I \subseteq J$ then $V(I) \supseteq V(J)$, and for sets X and Y, if $X \subseteq Y$ then $I(X) \supseteq I(Y)$. (Thus, both I and V are inclusion-reversing.)
6. For any subset S of $k[x_1, \ldots, x_n]$, $S \subseteq I(V(S))$ and $V(S) = V(I(V(S)))$.
7. For any subset X of $\mathbb{A}^n(k)$, $X \subseteq V(I(X))$ and $I(X) = I(V(I(X)))$. Furthermore, $I(X)$ is a radical[4] ideal.
    - Proofs: Exercises.

- Definition: For a field k, an affine algebraic set in $\mathbb{A}^n(k)$ is a subset of $\mathbb{A}^n(k)$ of the form $V(I)$ for some ideal I.
    - Examples: Single points $\{(a_1, \ldots, a_n)\} = V(x_1 - a_1, \ldots, x_n - a_n)$ are affine algebraic sets by (2) above. The sets $\{(t, 0) : t \in k\} = V(y)$ and $\{(t^2, t^3) : t \in k\} = V(y^2 - x^3)$ are affine algebraic sets.
    - By (4), we see that affine algebraic sets are closed under finite unions and arbitrary intersections, and (3) shows that $\mathbb{A}^n$ and $\emptyset$ are affine algebraic sets.
    - Thus, if we consider affine algebraic sets to be closed (with the open sets therefore being their complements), we obtain a topology on $\mathbb{A}^n(k)$. This topology is known as the Zariski topology.
    - By Hilbert's basis theorem, every ideal of $k[x_1, \ldots, x_n]$ is finitely generated, so by (4) above, we see that every affine algebraic set is of the form $V(f_1) \cap V(f_2) \cap \cdots \cap V(f_i)$ for some polynomials $f_1, \ldots, f_i$. (Equivalently, the complements of the sets $V(f_i)$ form a base for the Zariski topology.)
    - It is natural to seek "minimal" elements under the Zariski topology.

- Definition: An affine algebraic set V is reducible if it can be written as $V = V_1 \cup V_2$ where $V_1, V_2 \neq V$, and it is irreducible otherwise.

- We have a few more properties:

8. V is irreducible if and only if $I(V)$ is a prime ideal of $k[x_1, \ldots, x_n]$.
    - Proof: If $V = V_1 \cup V_2$ with $V_1, V_2 \neq V$, then $I(V_1)$ and $I(V_2)$ both properly contain V: if $f \in I(V_1) \backslash V$ and $g \in I(V_2) \backslash V$ then $fg \in I(V_1) \cap I(V_2) = I(V)$, meaning that $I(V)$ is not prime.
    - Conversely, if $fg \in I(V)$ with $f, g \notin I(V)$, we can take $V_1 = V \cap V(f)$ and $V_2 = V \cap V(g)$: then $V_1 \cup V_2 = V$ and $V_1, V_2 \neq V$ so V is reducible.

9. Any affine algebraic set V can be written uniquely as a union of irreducible affine algebraic sets $V_1 \cup V_2 \cup \cdots \cup V_n$ such that $V_i \not\subseteq V_j$ for any $i \neq j$. (These sets V_i are the irreducible components of V.)
    - Proof: Exercise. This result is the geometric version of primary decomposition (generalizing the notion of prime factorization of elements).

- Exercise: If k is algebraically closed, show that the irreducible affine algebraic sets in $\mathbb{A}^2(k)$ are $\emptyset$, $\mathbb{A}^2(k)$, single points, and curves of the form $V(f)$ for a monic irreducible polynomial $f \in k[x, y]$. [Hint: Show that if $f, g \in k[x, y]$ are relatively prime, then (f, g) contains a nonzero polynomial in $k[x]$ and a nonzero polynomial in $k[y]$.]

- Although it may appear that I and V should behave like inverses, they are not quite.
    - For example, even in $\mathbb{A}^1(k)$, we have $V(x^2) = \{0\}$ so that $I(V(x^2)) = (x)$. The point here is that $I = (x^2)$ is not a radical ideal, and in this case, $I(V(I)) = \mathrm{rad}(I)$.

[4]Recall that if I is an ideal of a commutative ring R, then the radical $\mathrm{rad}(I) = \{r \in R : r^n \in I \text{ for some } n \geq 1\}$, and I is a radical ideal if $I = \mathrm{rad}(I)$. (Note that $\mathrm{rad}(I)$ is an ideal, as is easily seen via an application of the binomial theorem.)

- However, even if I is radical, it is not always true that $I(V(I)) = \text{rad}(I)$: for example, in $\mathbb{A}^1(\mathbb{R})$ we have $V(1+x^2) = \emptyset$ so that $I(V(1+x^2)) = \mathbb{R}[x]$.
- Indeed, there is no subset S of $\mathbb{A}^1(\mathbb{R})$ with $I(S) = (1+x^2)$ since the only set S with $I(S) \supseteq (1+x^2)$ is the empty set. The issue here is that $\mathbb{R}$ is not algebraically closed: if instead we work in $\mathbb{C}$, then $S = \{i, -i\}$ does have $I(S) = (1+x^2)$.

- Working over an algebraically closed field resolves all of these difficulties: this is the content of Hilbert's Nullstellensatz, which has various forms:

- Theorem (Affine Nullstellensatz): Suppose k is an algebraically closed field.
  - (Weak) If I is a proper ideal of $k[x_1, \ldots, x_n]$, then $V(I) \neq 0$.
  - (Strong) If I is any ideal of $k[x_1, \ldots, x_n]$, then $I(V(I)) = \text{rad}(I)$.
  - We outline the arguments, leaving the full details as an exercise in commutative algebra or looking up the appropriate references.
  - For the weak Nullstellensatz, it suffices to show the result for maximal ideals, and then since $k[x_1, \ldots, x_n]$ is Noetherian, it is in turn sufficient to show that the finitely many generators of a maximal ideal I must have a common zero somewhere.
  - This in turn follows from showing that the quotient ring $k[x_1, \ldots, x_n]/I$, which is a field extension of k because I is maximal, is a finite-degree extension of k: then since k is algebraically closed, k has no finite-degree field extensions, so the extension simply equals k itself. Then if $\varphi : k[x_1, \ldots, x_n]/I \to k$ is the associated isomorphism, all elements of I vanish at the point $\varphi(x_1, \ldots, x_n)$.
  - For the full Nullstellensatz, after noting that $\text{rad}(I) \subseteq I(V(I))$, one uses the "Rabinowitsch trick" for the other containment: if $g \in I(V(f_1, \ldots, f_r))$, consider the ideal $J = (f_1, \ldots, f_r, x_{n+1}g - 1)$ of $k[x_1, \ldots, x_n, x_{n+1}]$, which has empty vanishing locus hence cannot be proper (by the weak Nullstellensatz), so it contains 1. By writing 1 as an appropriate linear combination of the generators of J and then setting $x_{n+1} = 1/g$ and clearing denominators appropriately, one obtains g^N as a linear combination of the f_i for some N, so $g \in \text{rad}(I)$.

- Definition: If k is algebraically closed, an irreducible affine algebraic set in $\mathbb{A}^n(k)$ is called an affine variety.

- Per the Nullstellensatz we see that I and V give nice bijections between various sets in $\mathbb{A}^n(k)$ and ideals of $k[x_1, \ldots, x_n]$.
  - By the full Nullstellensatz, since $I(V(I)) = \text{rad}(I)$, we obtain a correspondence between radical ideals and affine algebraic sets.
  - Furthermore, by the weak Nullstellensatz, if I is a proper ideal then $V(I)$ must contain some point $(a_1, \ldots, a_n)$, whence I is contained in $I(\{(a_1, \ldots, a_n)\}) = (x_1 - a_1, \ldots, x_n - a_n)$. But since the quotient of $k[x_1, \ldots, x_n]$ by $(x_1 - a_1, \ldots, x_n - a_n)$ is isomorphic to k via the evaluation map $p \mapsto p(a_1, \ldots, a_n)$, the latter ideal is maximal. Thus, the maximal ideals of $k[x_1, \ldots, x_n]$ correspond precisely with points $(a_1, \ldots, a_n)$.
  - Also, by the full Nullstellensatz, if I is a prime ideal, then $I(V(I)) = \text{rad}(I) = I$ since prime ideals are radical, and so by property (8) earlier, we see that $V(I)$ is irreducible. Thus, the prime ideals of $k[x_1, \ldots, x_n]$ correspond with irreducible affine algebraic sets (i.e., affine varieties).
  - To summarize, we have the following correspondences:

    [Affine algebraic sets] $\underset{V}{\overset{I}{\rightleftarrows}}$ [Radical Ideals]

    [Affine varieties] $\underset{V}{\overset{I}{\rightleftarrows}}$ [Prime Ideals]

    [Points of $\mathbb{A}^n(k)$] $\underset{V}{\overset{I}{\rightleftarrows}}$ [Maximal Ideals]

- Next, we bring rational functions into the discussion:

- Definition: If $V = V(I)$ is an affine variety, the coordinate ring of V is the ring $\Gamma(V) = k[x_1, \ldots, x_n]/I(V)$, and its associated field of rational functions (or function field) $k(V)$ is the field of fractions of $\Gamma(V)$.

- Recall that if R is an integral domain, the field of fractions of R consists of the equivalence classes of elements of the form a/b with $a, b \in R$ and b nonzero under the usual equivalence $a/b \sim c/d$ if and only if $ad = bc$. (The field of fractions is also the localization of R at $R\backslash\{0\}$.)
- Exercise: Let $\mathcal{F}(V, k)$ be the ring of k-valued functions on V. We say $f \in \mathcal{F}(V, k)$ is a polynomial function if there exists $g \in k[x_1, \ldots, x_n]$ such that $f(P) = g(P)$ for all $P \in V$. Show that $\Gamma(V)$ is the set of equivalence classes of polynomial functions under the relation $g_1 \sim g_2$ if $g_1(P) = g_2(P)$ for all $P \in V$.
- By the exercise above, the coordinate ring of V can be thought of as the collection of distinct polynomial functions on V, and thus the field of rational functions is, quite explicitly, the collection of rational functions on V.
- Examples: For the affine variety $V = V(y - x^2)$ in $\mathbb{A}^2(\mathbb{C})$, some examples of functions in the coordinate ring are x, $3x - 7$, y, and x^2. For this variety, the functions y and x^2 are the same, since they represent the same coset in the quotient ring. Some examples of rational functions are y/x, which also equals x, and $(x^2 + 5)/(3x - 7)$.

- Rational functions can have poles, which are points $P \in V$ where the function is not defined.
- Definition: If V is an affine variety, we say $f \in k(V)$ is defined at a point P if $f = a/b$ for some $a, b \in \Gamma(V)$ and $b(P) \neq 0$. If f is defined at P, its value $f(P)$ is the ratio $a(P)/b(P) \in k$. The local ring of V at P, denoted $\mathcal{O}_P(V)$, is the set of rational functions $f \in k(V)$ that are defined at P. The points P for which f is not defined are the poles of f, since they are necessarily zeroes of its denominator.
  - Exercise: Show that $\Gamma(V) = \cap_{P \in V} \mathcal{O}_P(V)$: in other words, that a function with no poles is a polynomial. (Note of course that k is still assumed to be algebraically closed.)
  - Examples: For the affine variety $V = V(y - x^2)$ in $\mathbb{A}^2(\mathbb{C})$, the rational function $f(x, y) = y/(x - 2)$ is defined everywhere on V except at the point $(1, 1)$, which is a pole of f.
  - The local ring $\mathcal{O}_P(V)$ has a unique maximal ideal $m_P(V)$ given by the polynomials f that vanish at P (i.e., with $f(P) = 0$.).
  - Exercise: Show that the evaluation-at-P map $\varphi_P : \mathcal{O}_P(V) \to k$ is a surjective ring homomorphism with kernel $m_P(V)$. Deduce that $m_P(V)$ is maximal, and show also that if $P = (a_1, \ldots, a_n)$ then $m_P(V)$ is generated by the polynomials $x_i - a_i$ for $1 \leq i \leq n$.
  - We remark also that the local ring $\mathcal{O}_P(V)$ is simply the localization of the function field $k(V)$ at the ideal $m_P(V)$; this is why $\mathcal{O}_P(V)$ is called the "local ring" of V at P.
  - When the variety V is clear from context we will often just write $\mathcal{O}_P$ and m_P.

- We will emphasize here that there may be numerous ways to write $\alpha = f/g$ as a quotient of polynomials inside the function field $k(V)$, and it may be necessary to work with different "equivalent" formulas in order to verify that α is defined at a particular point P.
- Example: Consider the affine variety $V = V(y^2 - x^2 + 1)$ in $\mathbb{A}^2(k)$ for $k = \mathbb{C}$ and the rational function $\alpha = \dfrac{x - 1}{y} \in k(V)$.
  - It is clear from the expression $\alpha = \dfrac{x - 1}{y}$ that α is defined at all points $P = (x, y) \in V$ where $y \neq 0$.
  - However, because $\Gamma(V) = k[x, y]/(y^2 - x^2 + 1)$, we see that $y^2 = x^2 - 1$ in $\Gamma(V)$, so by factoring and rearranging we see that $\dfrac{x - 1}{y} = \dfrac{y}{x + 1}$ inside $k(V)$. Therefore, α is also equal to $\dfrac{y}{x + 1}$, and this latter expression shows that f is also defined at the point $(1, 0)$.
  - On the other hand, there is no way to rewrite $\alpha = \dfrac{x - 1}{y}$ in such a way that it is defined at $(-1, 0)$: if $\dfrac{x - 1}{y} = \dfrac{p}{q}$ then $(x - 1)q = yp$ but then evaluating both sides at $P = (-1, 0)$ produces $-2q(P) = 0$, which is a contradiction.
  - Remark: More generally, the same argument shows that if the expression for $\alpha(P)$ is of the form $a/0$ for $a \neq 0$, then α is not defined at P. (If, of course, we obtain an expression $0/0$, then f could possibly be defined at P.)

- Definition: If V is an affine variety with function field $k(V)$, its dimension is defined to be the transcendence degree of $k(V)$ over k. An affine curve is an affine variety of dimension 1.
  - Examples:$V(y-x)$ and $V(y^2+a_1xy+a_3y-x^3-a_2x^2-a_4x-a_6)$ are affine curves in $\mathbb{A}^2(k)$.
  - If we think of $V=V(I)$ as being cut out from $\mathbb{A}^n(k)$ by the generators of I, then the dimension (as defined above) agrees with the intuitive topological sense of the dimension of $V(I)$ as a (hyper)surface, when $k=\mathbb{C}$.

0.5 (Sep 24) Affine Curves, $\mathbb{P}^n$

- We outline some additional facts about affine curves in $\mathbb{A}^2(k)$:

1. Via the correspondence $C \mapsto V(f)$, an affine plane curve C is the same as a nonconstant monic irreducible polynomial $f \in k[x,y]$. We define the degree of C to be the degree of the corresponding polynomial f.
   - As noted in an exercise earlier, the irreducible affine sets in $\mathbb{A}^2(k)$ are $\emptyset$ (dimension 0), single points (dimension 0), $\mathbb{A}^2(k)$ (dimension 2), and the sets of the form $V(f)$ where f is a monic irreducible polynomial (these are the only sets of dimension 1, so they are the only curves).
2. If P is a point of the affine curve $C=V(f)$, we say P is a singular point if $f_x(P)=f_y(P)=0$, and otherwise we say P is a nonsingular point (or smooth point or simple point). We say that C itself is smooth if all points of C are smooth points.
   - The main idea here is that a point P is singular if and only if C does not have a well-defined tangent line at P.
   - To find the tangent line(s) to a curve at a point P, we simply expand the defining polynomial f as a local Taylor series centered at $P=(x_0,y_0)$, i.e., as $f=a_{0,0}+a_{1,0}(x-x_0)+a_{0,1}(y-y_0)+a_{2,0}(x-x_0)^2+a_{1,1}(x-x_0)(y-y_0)+a_{0,2}(y-y_0)^2+\cdots$. Then the tangent lines are obtained by factoring the lowest-degree homogeneous component appearing in the factorization.
   - In particular, since $a_{0,0}=f(P)=0$, $a_{1,0}=f_x(P)$, and $a_{0,1}=f_y(P)$ by the usual Taylor expansion, we see that there is a unique tangent line precisely when the linear term does not vanish (i.e., P has multiplicity 1), which is to say, precisely when $f_x(P)$ and $f_y(P)$ are not both zero.
   - Example: The point $(0,0)$ lies on the variety $V(x+x^3-2y-y^5)$. Writing the curve locally near $(0,0)$ yields $f=(x-2y)+x^3-y^5$, and the lowest-degree homogeneous component is $x-2y$. Here, the curve has a unique tangent line at $(0,0)$ given by $x-2y=0$ (which one may check explicitly using calculus).
   - Example: The elliptic curve $V(y^2-x^2-x^3)$ has a singular point at $(0,0)$. Writing the curve locally near $(0,0)$ yields $f=-x^2+y^2-x^3$, and the lowest-degree homogeneous component is $(-x^2+y^2)=(-x+y)(-x-y)$. Here, the curve has two different tangent lines, $y=x$ and $y=-x$, yielding a node at $(0,0)$.
   - Example: The elliptic curve $V(y^2-x^3)$ has a singular point at $(0,0)$. Writing the curve locally near $(0,0)$ yields $f=y^2-x^3$, and the lowest-degree homogeneous component is y^2. Here, the curve has a double tangent line $y=0$, yielding a cusp at $(0,0)$.
   - The degree of the lowest term with a nonzero coefficient in the local expansion of f at P is called the multiplicity of P. One may show that for sufficiently large n, the multiplicity of C at P is equal to $\dim_k(m_P^n/m_P^{n+1})$, where m_P is the maximal ideal of the local ring $\mathcal{O}_P$ at P.
3. If P is a smooth point of the curve C, then the maximal ideal $m_P(C)$ of the local ring $\mathcal{O}_P(C)$ is principal. Any generator for this maximal ideal is called a uniformizer at P.
   - Exercise: Let R be a commutative ring with 1 having a maximal ideal M. Show that M^n/M^{n+1} is a vector space over the field $k=R/M$ for each positive integer n.
   - Exercise: Let R be a local ring (a commutative ring with 1 having a unique maximal ideal M). Show that every element of R is either a unit or an element of M.
   - The principality of m_P follows from the more general statement that if P is a smooth point of a variety, then $\dim_k(m_P/m_P^2)=\dim(V)$. Since curves have dimension 1, this yields $\dim_k(m_P/m_P^2)=1$. If t generates the vector space m_P/m_P^2, then in fact one may show that $m_P=(t)$, though this takes some effort.

- Example: Consider the smooth elliptic curve $C : y^2 = x^3 + x$ over $\mathbb{C}$, whose function field is $k(C) = \mathbb{C}[x,y]/(y^2 - x^3 - x)$. At the point $P = (0,0)$, the associated maximal ideal of the local ring is $m_P = (x,y)$ having $m_P^2 = (x^2, xy, y^2)$. A priori we can see that m_P/m_P^2 is spanned as a vector space by x and y, but in fact since $x = y^2 - x^3 \equiv 0 \bmod m_P^2$, m_P/m_P^2 is generated by y by itself. (In terms of the local expansion near $(0,0)$, this is the same as saying that the lowest-degree homogeneous component has degree 1, which is in turn simply saying that C is smooth at P.) As elements of the local ring $\mathcal{O}_P(C)$ we can observe that $x = \dfrac{y^2}{x^2+1}$ and since $\dfrac{1}{x^2+1} \in \mathcal{O}_P(C)$ does not vanish at P, this means $x \in (y^2)$ as an ideal of $\mathcal{O}_P(C)$, so again $m_P(C) = (x,y) = (y)$.
- Exercise: Show that for any elliptic curve in reduced Weierstrass form $y^2 = x^3 + Ax + B$ and any point $P = (a,b)$ on C, then the corresponding maximal ideal $m_P = (x-a, y-b)$ of the local ring is principal and generated by either $y - b$ (when $y'(P) \neq 0$) or $x - a$ (when $y'(P) = 0$).

4. If P is a smooth point of the curve C and $g \in \mathcal{O}_P(C)$, we define the order of vanishing of $v_P(g)$ at P to be the maximum n for which $g \in m_P(C)^n$. We extend this map to rational functions $\alpha = f/g \in k(C)$ by setting $v_P(f/g) = v_P(f) - v_P(g)$.
   - We will often also write $\mathrm{ord}_P(\alpha)$ interchangeably with $v_P(\alpha)$.
   - If $g(P) \neq 0$ then the order of vanishing is zero, while if $g(P) = 0$ then the order of vanishing is 1 or larger. Since $\bigcap_{n=0}^{\infty} m_P(C)^n = 0$, any nonzero g has a finite order of vanishing (as usual we take the convention that $v_P(0) = \infty$).
   - Although the definition is somewhat complicated, the point is that this order-of-vanishing map is simply the familiar notion of the multiplicity of a zero or pole of a rational function (or of a convergent power series, in analytic contexts).
   - Example: On $C = \mathbb{A}^1(\mathbb{C})$, consider $P = 0$. Then $\mathcal{O}_P$ is the set of rational functions $\dfrac{f(x)}{g(x)}$ with $g(0) \neq 0$ (i.e., rational functions defined at 0) and m_P is the set of rational functions vanishing at 0. It is easy to see that $m_P = (x)$, and so in general if we write a nonzero rational function in the form $x^a \dfrac{f(x)}{g(x)}$ where $f(0), g(0) \neq 0$, then $v_P(x^a \dfrac{f(x)}{g(x)}) = a$. For example, $v_P(x^2) = 2$ while $v_P(\dfrac{1}{x+1}) = 0$ and $v_P(\dfrac{x-1}{x^2+x}) = -1$. In each case we are simply computing the order of the zero or pole of the rational function at $x = 0$.
   - Example: On $C : y^2 = x^3 + x$ over $\mathbb{C}$, consider $P = (0,0)$. As we saw above, $m_P = (y)$ and $x = \dfrac{y^2}{x^2+1}$, so for example we have $v_P(y) = 1$, $v_P(x) = 2$, and $v_P(\dfrac{1}{y^2 - x}) = v_P(\dfrac{1}{x^3}) = -6$.

5. If C is a smooth curve, then for any nonzero rational function $\alpha \in k(C)$, there are only finitely many points P such that $v_P(\alpha) \neq 0$. When $v_P(\alpha) = d > 0$ we say that α has a zero of order d at P, and when $v_P(\alpha) = -d < 0$ we say that α has a pole of order d at P.
   - Exercise: Suppose C is a plane curve and $f(x,y)$ is a polynomial that is not identically zero on C. Show that there are only finitely many $P \in C$ for which $f(P) = 0$.
   - The idea is that for $\alpha = f/g$, any point with $v_P(\alpha) > 0$ requires $f(P) = 0$ and any point with $v_P(\alpha) < 0$ requires $g(P) = 0$. By the exercise above, for any fixed nonzero polynomials f and g on C, there are only finitely many such $P \in C$ with $f(P) = 0$ or $g(P) = 0$.
   - Example: On $C = \mathbb{A}^1(\mathbb{C})$, the rational function $\alpha = \dfrac{x^3}{x^2 - 1}$ has a zero of order 3 at $P = 0$ and poles of order 1 at $P = -1$ and $P = 1$, with no other zeroes or poles.

6. The order-of-vanishing map at a point P is in fact a discrete valuation on the function field $k(C)$.
   - Recall that a discrete valuation on a field F is a surjective function $v : F^\times \to \mathbb{Z}$ such that $v(ab) = v(a) + v(b)$ for all $a, b \in F^\times$ and $v(a+b) \geq \min(v(a), v(b))$ for all $a, b \in F^\times$ with $a + b \neq 0$. By convention we also take $v(0) = \infty$, in which case the statements hold for all a, b. The valuation ring R is the set of elements $r \in F$ with $v(r) \geq 0$.
   - The valuation properties $v(ab) = v(a) + v(b)$ and $v(a+b) \geq \min(v(a), v(b))$ are essentially immediate, and the only result that needs to be checked is that there is an element of the maximal ideal $m_P(C)$ whose valuation equals 1. This follows by generalizing the observation made in the exercise following (3) that for $P = (a,b)$, either $x - a$ or $y - b$ must have valuation 1, for if not, the point P would be singular.

- Exercise (Properties of DVRs): Let F be a field with a discrete valuation v and valuation ring R. Also let $t \in R$ be a uniformizer (i.e., an element with $v(t) = 1$). Show that
  (a) For any $r \in F^\times$, either r or $1/r$ is in R.
  (b) An element $u \in R$ is a unit of R if and only if $v(u) = 0$. In particular, if $\zeta \in F$ is any root of unity, then $v(\zeta) = 0$.
  (c) If $r \in R$ is nonzero and $v(r) = n$, then r can be written uniquely in the form $r = ut^n$ for a unit $u \in R$.
  (d) Every nonzero ideal of R is of the form (t^n) for some $n \geq 0$.
  (e) The ring R is a Euclidean domain (hence also a PID and a UFD) and also a local ring.
  (f) The ring S is a DVR if and only if it is a PID and a local ring but not a field.

- The main issue with affine space is that it is missing points in a way that creates many unpleasant special cases and exceptions to various fundamental results. In order to rectify these issues we now enlarge our perspective to work in projective space:

- Definition: For a field k, we define projective n-space $\mathbb{P}^n(k) = \{[x_0 : x_1 : \cdots : x_n] : x_i \in k \text{ not all zero}\}/\sim$, where $P \sim Q$ if $P = \lambda Q$ for some nonzero $\lambda \in k$. Equivalently, $\mathbb{P}^n(k)$ is the set of lines through the origin in $\mathbb{A}^{n+1}(k)$.

  - We use the notation $[x_0 : x_1 : \cdots : x_n]$ to evoke the idea of considering only the ratios between the coordinates, since (for example) in $\mathbb{P}^1(k)$ the points $[1 : 1]$ and $[2 : 2]$ are the same. The coordinates x_i of a point $P \in \mathbb{P}^n(k)$ are not well-defined, but since the equivalence is only up to scaling by a nonzero constant, the statement "$x_i = 0$" is still well-defined, as are the ratios x_i/x_j.
  - For the set $U_i = \{[x_0 : x_1 : \cdots : x_n] : x_i = 1\}$, we can see that U_i looks exactly like $\mathbb{A}^n(k)$ (if we just delete the coordinate $x_i = 1$), and $\mathbb{P}^n(k) = \cup_{i=0}^n U_i$.
  - The complement of the set U_i is the hyperplane $x_i = 0$, and it looks exactly like $\mathbb{P}^{n-1}$ (if we just delete the coordinate $x_i = 0$).
  - Thus, somewhat informally, we have $\mathbb{P}^n(k) = \mathbb{A}^n(k) \cup \mathbb{P}^{n-1}(k)$, where we can think of $\mathbb{A}^n(k)$ as being the points with $x_n = 1$ and $\mathbb{P}^{n-1}(k)$ as being the points with $x_n = 0$.
  - Example: We have $\mathbb{P}^1(k) = \{[x : 1] : x \in k\} \cup \{[1 : 0]\}$, which looks like $\mathbb{A}^1$ along with a point at ∞.

- Evaluating an arbitrary polynomial on a projective point is not well defined, since projective points have various equivalent representatives, and the resulting polynomial value is not well-defined even up to scaling. But we are only interested in vanishing sets, which can be sensibly defined.

  - A natural but somewhat ill-advised option would be to say that $P \in \mathbb{P}^n(k)$ is in the vanishing set of $f \in k[x_0, \ldots, x_n]$ if $f(P) = 0$ for *all* choices of coordinates for P.
  - Exercise: Suppose k is an infinite field, $P \in \mathbb{A}^{n+1} \backslash \{0\}$, and $f \in k[x_0, \ldots, x_n]$. If we write $f = f_0 + f_1 + \cdots + f_d$ for homogeneous[5] polynomials f_i of degree i, show that $f(\lambda P) = 0$ for all $\lambda \in k^\times$ if and only if $f_i(P) = 0$ for all i. [Hint: Use linear algebra and the fact that Vandermonde determinants are nonvanishing.]
  - Per the exercise above, we see that when k is an infinite field, requiring $f(P) = 0$ for all choices of coordinates for P is equivalent to requiring that all of the homogeneous components of f vanish.
  - For consistency with finite fields (which have nonzero polynomials that vanish everywhere, causing issues with the argument above), we instead define the vanishing of a polynomial f on a projective point P in terms of homogeneous components.

- Definition: If $f \in k[x_0, \ldots, x_n]$ is a polynomial with $f = f_0 + f_1 + \cdots + f_d$ for homogeneous polynomials f_i of degree i, we say that f vanishes at $P \in \mathbb{P}^n(k)$, and write $f(P) = 0$, if $f_i(P) = 0$ for each i.

  - Note that $f_i(\lambda P) = \lambda^i f(P)$ so the vanishing condition on f_i does not depend on which equivalent coordinates are used for P.

[5] Recall that a polynomial is homogeneous of degree d if all of its monomial terms have total degree d. For example, $x^2y - 3x^3 + xyz$ is homogeneous of degree 3.

- Example: The polynomial $f(x,y) = x^2 - y^2$ vanishes at the projective point $[1:1]$ since its only nonzero homogeneous component $x^2 - y^2$ vanishes at P, but the polynomial $g(x,y) = x - y^2$ does not since its homogeneous components are x and $-y^2$ and these do not vanish at $[1:1]$.
- The main theme is that when we want to work with polynomials in projective space, we want to consider only homogeneous polynomials.

- Now that we have given a reasonable definition of vanishing for projective points, we can define the projective versions of the operators V and I:

- Definition: If S is any set of polynomials in $k[x_0, \ldots, x_n]$, we define the vanishing locus $V(S) = \{P \in \mathbb{P}^n(k) : f(P) = 0 \text{ for all } f \in S\}$. Conversely, if X is any set of points in $\mathbb{P}^n(k)$, we define the ideal of functions vanishing on X as $I(X) = \{f \in k[x_0, \ldots, x_n] : f(P) = 0 \text{ for all } P \in X\}$.

  - Exercise: Identify $V(x_0)$, $V(x_0^2)$, $V(x_1 - x_0)$, $V(x_1 - x_0^2)$, $V(x_1^2 - x_0^2)$, $V(x_0, x_1)$, $V(x_0, x_1, x_2)$, and $V(x_0x_1 - x_2^2)$ in $\mathbb{P}^2(k)$.

- All of the basic properties of the affine operators I and V also hold for the projective I and V (suitably modified):

  1. If I is the ideal generated by $T \subseteq k[x_0, \ldots, x_n]$, then $V(T) = V(I)$.
  2. $V(0) = \mathbb{P}^n(k)$, $V(1) = \emptyset$, and $V(\{a_ix_j - a_jx_i\}_{0 \leq i,j \leq n}) = \{[a_0 : a_1 : \cdots : a_n]\}$.
  3. $I(\emptyset) = k[x_0, \ldots, x_n]$, $I(\mathbb{P}^n) = 0$ when k is infinite, and $I(\{[a_0 : a_1 : \cdots : a_n]\}) = (\{a_ix_j - a_jx_i\}_{0 \leq i,j \leq n})$.
  4. $V(\cup_i I_i) = \cap_i V(I_i)$ and $V(IJ) = V(I) \cup V(J)$.
  5. For ideals I and J, if $I \subseteq J$ then $V(I) \supseteq V(J)$, and for sets X and Y, if $X \subseteq Y$ then $I(X) \supseteq I(Y)$.
  6. For any subset S of $k[x_0, \ldots, x_n]$, $S \subseteq I(V(S))$ and $V(S) = V(I(V(S)))$.
  7. For any subset X of $\mathbb{P}^n(k)$, $X \subseteq V(I(X))$ and $I(X) = I(V(I(X)))$. Furthermore, $I(X)$ is a radical ideal.

- Owing to our definition of vanishing in terms of homogeneous components, the ideals of sets in $\mathbb{P}^n(k)$ have an additional property:

- Definition: An ideal I of $k[x_0, \ldots, x_n]$ is homogeneous if, for any $f \in I$ with homogeneous decomposition $f = f_0 + f_1 + \cdots + f_d$, it is true that each component $f_i \in I$.

  - It is easy to see that $I(X)$ is homogeneous, since for any $f = f_0 + f_1 + \cdots + f_d \in I(X)$, by definition of vanishing we see that for any $P \in X$ we have $f_i(P) = 0$ and so $f_i \in I(X)$.
  - Exercise: Show that an ideal I of $k[x_0, \ldots, x_n]$ is homogeneous if and only if I is generated by finitely many homogeneous polynomials.

- We also have a projective version of the Nullstellensatz, which is essentially the same as the affine version except that we must account for the fact that the vanishing locus of the ideal $(x_0, x_1, \ldots, x_n)$ in $\mathbb{P}^n$ is empty since $[0 : 0 : \cdots : 0]$ is not a point of $\mathbb{P}^n$:

- Theorem (Projective Nullstellensatz): Let k be an algebraically closed field and I be a homogeneous ideal of $k[x_0, \ldots, x_n]$. Then the following hold:

  1. (Weak) $V(I) = \emptyset$ if and only if I contains all monomials of sufficiently large degree, if and only if $\text{rad}(I)$ contains $(x_0, \ldots, x_n)$.
  2. (Strong) If $V(I) \neq \emptyset$, then $I(V(I)) = \text{rad}(I)$.
     - The proofs are similar to those of the affine Nullstellensatz, and are left as exercises.
     - Owing to the fact that its vanishing locus is trivial, and thus can essentially be ignored when doing computations, the ideal $(x_0, x_1, \ldots, x_n)$ in $k[x_0, \ldots, x_n]$ is called the irrelevant ideal.

0.6 (Sep 28) Projective Varieties, Rational Maps, Morphisms

- Next, we define algebraic sets, varieties, and coordinate rings in $\mathbb{P}^n$. The ideas proceed essentially the same way:
- Definition: A projective algebraic set is a set in $\mathbb{P}^n(k)$ of the form $V(I)$ for some ideal I of $k[x_0, \ldots, x_n]$. A projective algebraic set V is reducible if it can be written as $V = V_1 \cup V_2$ where $V_1, V_2 \neq V$, and it is irreducible otherwise. A projective variety is an irreducible projective algebraic set.
    - As in the affine case, V is irreducible if and only if $I(V)$ is a prime ideal of $k[x_0, \ldots, x_n]$, and any projective algebraic set can be written uniquely as a union of irreducible components $V_1 \cup V_2 \cup \cdots \cup V_n$ such that $V_i \not\subseteq V_j$ for any $i \neq j$.
- Definition: If V is a projective variety, then its (homogeneous) coordinate ring is the integral domain $\Gamma(V) = k[x_0, \ldots, x_n]/I(V)$.
    - As before, we may decompose the polynomials $f \in \Gamma(V)$ as $f = f_0 + f_1 + \cdots + f_d$ where f_i is homogeneous of degree i.
    - Since $I(V)$ is prime, the coordinate ring is an integral domain, so its fraction field is well defined. Unlike in the affine case, however, the elements of this fraction field do not generally determine functions on V, because a ratio of polynomials need not be a function on V.
    - The first obvious issue is that for a ratio $\dfrac{f}{g} = \dfrac{f_0 + f_1 + \cdots + f_d}{g_0 + g_1 + \cdots + g_d}$, the various homogeneous terms in the numerator and denominator will not transform the same way if we choose a different representative for the projective point $P \in V$ at which we are attempting to evaluate f/g. (For example: what is the value of $\dfrac{x + y^2}{x + y}$ at the projective point $[1 : 1]$?)
    - To handle this issue, we must only have a single homogeneous component in the numerator and denominator. But even here, in order for the ratio to be well-defined, the degrees of the numerator and denominator must be equal.
    - When we restrict to rational functions of this form, however, we do obtain well-defined functions on projective points: if f, g are both homogeneous of degree d, then $\dfrac{f(\lambda P)}{g(\lambda P)} = \dfrac{\lambda^d f(P)}{\lambda^d g(P)} = \dfrac{f(P)}{g(P)}$, so the ratio f/g is well defined regardless of the representative of P we use.
- Definition: If V is a projective variety, its function field $k(V)$ is the set of elements z in the fraction field of $\Gamma(V)$ such that z can be written in the form $z = \dfrac{f}{g}$ for some homogeneous polynomials $f, g \in k[x_0, \ldots, x_n]$ of the same degree. We say z is defined at a point $P \in V$ if $z = f/g$ for some g with $g(P) \neq 0$. The local ring of V at P is $\mathcal{O}_P(V) = \{z \in k(V) : z \text{ is defined at } P\}$ with maximal ideal $m_P(V) = \{z \in \mathcal{O}_P(V) : z(P) = 0\}$.
    - As in the affine case, we may require different expressions $z = f/g$ at different points P.
- Example: Consider the variety $V = V(Y^2 + Z^2 - X^2)$ in $\mathbb{P}^2(\mathbb{C})$ and the rational function $f = \dfrac{X - Z}{Y} \in k(V)$.
    - It is clear from the expression $f = \dfrac{X - Z}{Y}$ that f is defined at all points $P = [X : Y : Z] \in V$ where $Y \neq 0$, which is to say, at all points of the form $[X : 1 : Z]$ after rescaling. The only points of V with $Y = 0$ are those with $X^2 = Z^2$, which gives two points: $[1 : 0 : 1]$ and $[1 : 0 : -1]$.
    - However, because $\Gamma(V) = k[X, Y, Z]/(Y^2 + Z^2 - X^2)$, we see that $Y^2 = X^2 - Z^2$ in $\Gamma(V)$, by factoring and rearranging we see that $\dfrac{X - Z}{Y} = \dfrac{Y}{X + Z}$ inside $k(V)$. Therefore, f is also equal to $\dfrac{Y}{X + Z}$, and this latter expression shows that f is also defined at the point $[1 : 0 : 1]$ (and in fact it vanishes there).
    - On the other hand, there is no way to rewrite $f = \dfrac{X - Z}{Y}$ in such a way that it is defined at $[1 : 0 : -1]$: if $\dfrac{X - Z}{Y} = \dfrac{p}{q}$ then $(X - Z)q = Yp$ but then evaluating both sides (as polynomials in X, Y, Z) at $X = 1$, $Y = 0$, $Z = -1$ produces $-2q(1, 0, -1) = 0$, which is a contradiction since this means $q(P) = 0$.

- Remark: Note that this is just the projective version of the example we did earlier for the affine variety $V = V(y^2 + 1 - x^2)$ in $\mathbb{A}^2$.

- As clearly indicated by the similarity of the calculations in the example above and the nearly-identical affine example from earlier, there is quite a lot of interplay between projective and affine spaces.
  - One such correspondence is obtained by viewing $\mathbb{P}^n$ as the lines through the origin in $\mathbb{A}^{n+1}$, so for any set S in $\mathbb{P}^n$ we may write down the set of its corresponding points in $\mathbb{A}^{n+1}$ by converting the point $[x_0 : x_1 : \cdots : x_n]$ to the point $(x_0, x_1, \ldots, x_n)$.
  - Explicitly, if $S \subseteq \mathbb{P}^n$, the cone $C(S)$ of S in $\mathbb{A}^{n+1}$ is the set $\{(x_0, x_1, \ldots, x_n) : [x_0 : x_1 : \cdots : x_n] \in S\} \cup \{(0, 0, \ldots, 0)\}$.
  - Exercise: When V is a nonempty projective algebraic variety, show that $I_{\text{affine}}(C(V)) = I_{\text{projective}}(V)$, and when I is a homogeneous ideal with $V_{\text{projective}}(I) \neq \emptyset$, show that $C(V_{\text{projective}}(I)) = V_{\text{affine}}(I)$.

- Although the cone of a variety shares the same underlying ideal, and thus has the same coordinate ring and function field, its dimension is different.
  - We would like instead to think of $\mathbb{P}^n$ as being $\mathbb{A}^n$ plus a hyperplane at ∞, and so an affine variety in $\mathbb{A}^n$ should give rise to one that looks essentially the same in $\mathbb{P}^n$, except for having some additional points in the hyperplane at ∞.
  - The main idea, as exemplified by comparing the example above to its affine version, is that of homogenization and dehomogenization.

- Definition: If $F \in k[x_0, x_1, \ldots, x_n]$ is a polynomial, its dehomogenization with respect to x_0 is $F_* = F(1, x_1, \ldots, x_n)$. Inversely, if $f \in k[x_1, \ldots, x_n]$ is a polynomial, its homogenization with respect to x_0 is $f^* = x_0^{\deg(f)} f(x_1/x_0, x_2/x_0, \ldots, x_n/x_0)$.
  - More explicitly, if $f \in k[x_1, \ldots, x_n]$ has homogeneous decomposition $f = f_0 + f_1 + \cdots + f_d$, then $f^* = x_0^d f_0 + x_0^{d-1} f_1 + \cdots + f_d$.
  - Example: The homogenizations of $x_1^2 + x_2$, $4 + x_1 x_3 - 3x_4^5$, and 1 are $x_1^2 + x_0 x_2$, $4x_0^5 - x_0^3 x_1 x_3 - 3x_4^5$, and 1 respectively.
  - Example: The dehomogenizations of $x_0^2 + 3x_0 x_1 + x_1 x_2$, $x_0^3 + 4x_0 x_2^2 + x_3^3$, and x_0^2 are $1 + 3x_1 + x_1 x_2$, $1 + 4x_2^2 + x_3^3$, and 1 respectively.
  - The main idea is that dehomogenizing removes the variable x_0 by setting it equal to 1 (thereby usually creating a non-homogeneous polynomial in the remaining variables $x_1, \ldots, x_n$) while homogenizing takes a non-homogeneous polynomial in $x_1, \ldots, x_n$ and makes it homogeneous in $x_0, x_1, \ldots, x_n$ by using the extra variable x_0 to make all of the terms have the same degree.
  - Homogenization and dehomogenization are essentially inverses of one another, aside from occasionally losing powers of x_0.
  - Exercise: Show that $(FG)_* = F_* G_*$, $(fg)^* = f^* g^*$, $(f^*)_* = f$, $(F_*)^* = F / x_0^{v_{x_0}(f)}$, $(F + G)_* = F_* + G_*$, and $x_0^{\deg(f)+\deg(g)-\deg(f+g)}(f + g)^* = x_0^{\deg(g)} f^* + x_0^{\deg(f)} g^*$.
  - It is also possible to homogenize and dehomogenize with respect to the other variables, but we will not need to do this in any substantial way.

- The point is that homogenizing an affine equation creates a projective one, and dehomogenizing a projective equation yields an affine one, thereby giving a correspondence between affine varieties and projective varieties. Since our interest is specifically in plane curves, we will use affine variables $x = x_1$ and $y = x_2$ (written in lowercase), and projective variables X, Y, and Z, with homogenizations performed with respect to Z and dehomogenizations defined via $x = X/Z$, $y = Y/Z$.
  - Motivating Example: Homogenizing the affine equation $x + y = 1$ yields the projective equation $X + Y = Z$. An affine point (x, y) satisfying $x + y = 1$ yields a projective point $[x : y : 1] = [X : Y : Z]$ satisfying $X + Y = Z$. If we compare the affine points to the projective ones, we see that the projective variety consists of the points $[x : y : 1]$, which all correspond to affine points, along with one additional point $[1 : -1 : 0]$ which we think of as the point at ∞ on this line.

- Motivating Example: Dehomogenizing the projective equation $Y^2Z = X^3 + XZ^2$ yields the affine equation $y^2 = x^3 + x$. A projective point $[X : Y : Z]$ satisfying $Y^2Z = X^3 + XZ^2$ yields an affine point $(x, y) = (X/Z, Y/Z)$ satisfying $y^2 = x^3 + x$, as long as $Z \neq 0$. When we dehomogenize, the projective points $[X : Y : Z]$ with $Z = 0$ "disappear" from the affine curve: note here that there is only one such point, namely $[0 : 1 : 0]$, which represents the point at ∞ on this elliptic curve.
- Exercise: Show that there exists a unique projective point $[X : Y : Z]$ with $Z = 0$ on any elliptic curve with a Weierstrass equation $Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3$.

- We can extend the notion of homogenization to ideals and then to algebraic sets in fairly natural ways:
  - Definition: If $I = (f_1, \ldots, f_k)$ is an ideal of $k[x_1, \ldots, x_n]$, the homogenization of I is the ideal $I^* = (f_1^*, \ldots, f_k^*)$ generated by the homogenizations of the generators of I. Conversely, if J is an ideal of $k[x_0, x_1, \ldots, x_n]$, the dehomogenization of J is the ideal $J_* = \{g_* : g \in J\}$ of dehomogenizations of the elements of J (and is generated by the dehomogenizations of the generators of J).
  - We are only interested in the situation of plane curves, in which the ideals are principal, so the homogenization and dehomogenization are quite easy to handle.
  - Definition: If V is an affine algebraic set, then for $I = I_{\text{affine}}(V)$ we define the homogenization of V to be the projective algebraic set $V^* = V_{\text{projective}}(I^*)$. Conversely, if W is a projective algebraic set, then for $J = I_{\text{projective}}(W)$ we define the dehomogenization of W to be the affine algebraic set $W_* = V_{\text{affine}}(J_*)$.

- Proposition (Equivalence of Function Fields): If V is an affine variety with projective closure V^*, then the function fields $k(V)$ and $k(V^*)$ are isomorphic. Furthermore, if P is any point on V with corresponding point P^* on V^*, then the isomorphism of $k(V)$ and $k(V^*)$ also yields an isomorphism of $\mathcal{O}_P(V)$ with $\mathcal{O}_{P^*}(V^*)$ and of $m_P(V)$ with $m_{P^*}(V^*)$.
  - The point here is that locally (near a given point P) the affine variety V and its projective closure V^* look equivalent, and their function fields are also the same.
  - As such, we may also immediately import all of our other constructions defined in terms of the local ring and its maximal ideal from the affine case: most importantly, the notion of the order of vanishing of a rational function at a point.

- So far, we have mostly been assuming that the constant field k is algebraically closed. In particular, since we are interested in elliptic curves over $\mathbb{Q}$ and $\mathbb{F}_q$, we will need to remove this assumption.
  - Explicitly, suppose V is a variety over k and E is a subfield of k. We would naively like to define the set of E-points of V as $V \cap \mathbb{A}^n(E)$ if V is affine, and as $V \cap \mathbb{P}^n(E)$ if V is projective.
  - We may make this more precise using Galois actions: specifically, assuming that $k = \overline{E}$, then the Galois group of k/E acts naturally on the k-points of V.

- Definition: Let E be a field with algebraic closure k, and let $G = \text{Gal}(k/E)$. If V is a variety over k, we define the E-points of V to be the set of points of V over k that are fixed by G.
  - Explicitly, P is an E-point of V if and only if $\sigma(P) = P$ for all $\sigma \in \text{Gal}(k/E)$.
  - The set of E-points of V is precisely $V \cap \mathbb{A}^n(E)$ if V is affine, and is $V \cap \mathbb{P}^n(E)$ if V is projective, since the given condition is equivalent to saying that all of the coordinates of the point lie in E. (Note that for projective points, we only need one representative to have all its coordinates in E.)
  - Example: For $E = \mathbb{F}_5$ and $V = V(y^2 - x^2 - 1)$ in $\mathbb{A}^2$, the set of E-points of V is $(x, y) = (0, 1)$, $(0, 4)$, $(2, 0)$, and $(3, 0)$.
  - Example: For $E = \mathbb{F}_3$ and $V = V(Y^2Z^2 - XZ^3 - X^4)$ in $\mathbb{P}^2$, the set of E-points of V is $[X : Y : Z] = [0 : 0 : 1]$, $[0 : 1 : 0]$, $[1 : 0 : 2]$.
  - We can also define the elements of the coordinate ring and function field of V over E, namely, as the elements of $\Gamma(V)$ and $k(V)$ fixed by $\text{Gal}(k/E)$, respectively.

- Definition: If E is a field with algebraic closure k, we say that a variety V is defined over E if $I(V)$ can be generated by polynomials with coefficients in E.

- We will think of all varieties as implicitly being defined over an algebraically closed field, even if it they are actually defined over a subfield.
- Thus, we may meaningfully speak of the points of V on arbitrary algebraic extensions of E.
- If P is a point on $k(V)$, we define the degree of P over E to be the degree of the smallest field extension L/E such that $P \in L(V)$.

- We now discuss maps between varieties. The most natural starting point is to consider maps defined by polynomials:

- Definition: If V is an affine variety in $\mathbb{A}^n(k)$ and W is an affine variety in $\mathbb{A}^m(k)$, a map $\varphi : V \to W$ is called a polynomial map from V to W if there exist polynomials $T_1, \ldots, T_m \in k[x_1, \ldots, x_n]$ such that $\varphi(a_1, \ldots, a_n) = (T_1(a_1, \ldots, a_n), T_2(a_1, \ldots, a_n), \ldots, T_m(a_1, \ldots, a_n))$.
  - Example: The map $\varphi : \mathbb{A}^1 \to \mathbb{A}^1$ with $\varphi(a) = a^2 + a$ is a polynomial map, as is the map $\varphi : \mathbb{A}^1 \to \mathbb{A}^3$ with $\varphi(a) = (a, a^2, a^3)$.
  - Example: The map $\varphi : V(x^2 + y^2 - 1) \to \mathbb{A}^1$ with $\varphi(x, y) = x$ is a polynomial map.
  - Example: The map $\varphi : \mathbb{A}^2 \to V(x^2 + y^2 - z^2)$ with $\varphi(a, b) = (2ab, a^2 - b^2, a^2 + b^2)$ is a polynomial map. Note that this map is well-defined because $(2ab)^2 + (a^2 - b^2)^2 - (a^2 + b^2)^2$ is indeed zero for all $(a, b) \in \mathbb{A}^2$, so $(2ab, a^2 - b^2, a^2 + b^2) \in V(x^2 + y^2 - z^2)$.
  - Example: The map $\varphi : V(y - x^2) \to V(z - xy)$ with $\varphi(x, y) = (x, y, x^3)$ is a polynomial map. Note that this map is well-defined because for all $(x, y) \in V(y - x^2)$ we have $y = x^2$, and then $(x, y, x^3) \in V(z - xy)$.

- Polynomial maps are equivalent to homomorphisms of coordinate rings:

- Proposition (Polynomial Maps and Coordinate Rings): If V and W are affine varieties, then any polynomial map $\varphi : V \to W$ induces a homomorphism $\varphi^* : \Gamma(W) \to \Gamma(V)$ on coordinate rings via "plugging in": $\varphi^*(f) = f \circ \varphi$. Conversely, any homomorphism $\varphi^* : \Gamma(W) \to \Gamma(V)$ is induced by a unique polynomial map $\varphi : V \to W$ with $\varphi^*(f) = f \circ \varphi$.
  - Proof: First suppose $\varphi : V \to W$ is a polynomial map. For any $f \in k[x_1, \ldots, x_n]$, define $\psi(f) = f \circ \varphi$. Clearly, ψ is a ring homomorphism (since it is just polynomial evaluation). Furthermore, this map ψ descends to a well-defined map $\tilde{\varphi} : \Gamma(W) \to \Gamma(V)$: this follows by noting that if $f \in \Gamma(W)$ is the $I(W)$-residue of a polynomial $G(x_1, \ldots, x_n)$, then $\tilde{\varphi}(f) = f \circ \varphi$ is the $I(V)$-residue of the polynomial $G(T_1, \ldots, T_m)$.
  - For the converse, we can simply reconstruct the map φ from its action on each variable x_i. Explicitly, suppose that $\tilde{\varphi} : \Gamma(W) \to \Gamma(V)$ is a homomorphism. Then $\tilde{\varphi}$ maps $x_i + I(W)$ to some polynomial $T_i + I(V)$ for each $1 \leq i \leq m$. Then the map $\varphi(a_1, \ldots, a_n) = (T_1(a_1, \ldots, a_n), \ldots, T_m(a_1, \ldots, a_n))$ is a polynomial map from $\mathbb{A}^n$ to $\mathbb{A}^m$, and it induces a map $\hat{\varphi} : \Gamma(\mathbb{A}^m) \to \Gamma(\mathbb{A}^n)$. From the information given we know that $\hat{\varphi}(I(W)) \subseteq I(V)$, so $\varphi(V) \subseteq W$. Thus, $\varphi|_V$ is a polynomial map from V to W, and $\tilde{\varphi}(f) = f \circ \varphi$ as required.

- Definition: If V and W are affine varieties, a polynomial map $\varphi : V \to W$ is an isomorphism if it possesses an inverse polynomial map $\psi : W \to V$ (i.e., with $\varphi \circ \psi = \mathrm{id}_W$ and $\psi \circ \varphi = \mathrm{id}_V$).
  - By the above, we see that V and W are isomorphic if and only if their coordinate rings are isomorphic as k-algebras (i.e., if their coordinate rings are isomorphic as rings where the isomorphism also fixes k).
  - Example: The map $\varphi : V(x - y) \to V(x - 2y)$ with $\varphi(x, y) = (2x, y)$ is an isomorphism with inverse $\psi(x, y) = (x/2, y)$.
  - Exercise: Show that the affine linear maps $\varphi : \mathbb{A}^n \to \mathbb{A}^n$ of the form $\varphi(x) = Ax + b$ where A is an invertible $n \times n$ matrix and b is any vector of constants, are isomorphisms. Find another isomorphism $\varphi : \mathbb{A}^n \to \mathbb{A}^n$ not of this form.

- We would like to write down a similar definition for projective varieties, which we can do at the cost of a bit of added complexity.
  - The most immediate issue is that we need to insist that all of the polynomials T_i be homogeneous of the same degree, in order to ensure that "plugging in" to a polynomial map is well defined.

- However, this is not the only obstruction; difficulties also arise in the event that all of the polynomials T_i vanish simultaneously, since then the resulting value does not yield a well-defined point in $\mathbb{P}^1$.

- Definition: If V and W are projective varieties, a rational map from V to W is a map of the form $\varphi = [\varphi_0 : \varphi_1 : \cdots : \varphi_m]$ where the $\varphi_i \in k[x_0, \ldots, x_n]$ are homogeneous polynomials of the same degree, and such that for all $f \in I(W)$, we have $f \circ \varphi = f(\varphi_0(x_0, \ldots, x_n), \ldots, \varphi_m(x_0, \ldots, x_n)) \in I(V)$.

  - If φ is a rational map, then for $P \in V$ we can evaluate $\varphi(P) = [\varphi_0(P) : \varphi_1(P) : \cdots : \varphi_m(P)] \in W$ as long as not all of the values $\varphi_i(P)$ are zero. We can see that this value $\varphi(P)$ is well defined because the φ_i are homogeneous of the same degree, and $\varphi(P) \in W$ precisely because $f \circ \varphi \in I(V)$ for any $f \in I(W)$.
  - To illustrate, consider the map $\varphi : V(X^2 + Y^2 - Z^2) \to \mathbb{P}^1$ given by $\varphi[X : Y : Z] = [X + Z : Y]$. On its face, this would appear to be a perfectly well-defined function, since for any equivalent representative $[\lambda X : \lambda Y : \lambda Z]$ we have $\varphi[\lambda X : \lambda Y : \lambda Z] = [\lambda X + \lambda Z : \lambda Y] = [X + Z : Y] = \varphi[X : Y : Z]$.
  - However, for the point $P = [1 : 0 : -1]$ in $V(X^2 + Y^2 - Z^2)$, the definition states $\varphi(P) = [0 : 0]$, which is not a point of $\mathbb{P}^1$.
  - Notice, though, that if we work inside $\Gamma(V)$, we see that $[X + Z : Y] = [(X + Z)(X - Z) : Y(X - Z)] = [-Y^2 : Y(X - Z)] = [-Y : X - Z]$ and this latter expression *is* defined at $[1 : 0 : -1]$ since it evaluates to $[0 : 2]$.
  - We would like to extend our interpretation of the value of $\varphi(P)$ in a way that allows us to make these kinds of manipulations.

- Definition: If $\varphi : V \to W$ is a rational map, we say that $\varphi = [\varphi_0 : \cdots : \varphi_m]$ is defined at P if there exist homogeneous polynomials $\psi_0, \ldots, \psi_n$ of the same degree such that $\varphi_i \psi_j \equiv \varphi_j \psi_i \pmod{I(V)}$ for all pairs (i, j) with $\psi_i(P) \neq 0$ for some i, and we write $\varphi(P) = [\psi_0(P) : \cdots : \psi_m(P)]$.

  - The idea here is that, inside $\Gamma(V)$, we view the homogeneous coordinates $[\varphi_0 : \cdots : \varphi_m]$ and $[\psi_0 : \cdots : \psi_m]$ as being projectively equivalent.
  - We call these "rational maps" because if we work affinely, they arise from rational functions.

- Definition: If V and W are varieties, a morphism from V to W is a rational map that is defined at all points of V. An isomorphism is a morphism possessing an inverse morphism.

  - If $\varphi : V \to W$ is a morphism, then φ induces an injective homomorphism on function fields $\varphi^* : k(W) \to k(V)$ via composition: $\varphi^*(f) = f \circ \varphi$.
  - As in the affine case for polynomial maps, the converse is true as well: any injective k-algebra homomorphism (i.e., a ring homomorphism fixing k) on function fields $\varphi^* : k(W) \to k(V)$ yields a morphism $\varphi : V \to W$.
  - Example: The map $\varphi : V(Y^2Z - X^3 - XZ^2) \to \mathbb{P}^1$ given by $\varphi[X : Y : Z] = [Y : Z]$ is a morphism. (Note that there are no points of $V(Y^2Z - X^3 - XZ^2)$ where φ is undefined, since if $Y = Z = 0$ then X would also be zero.)
  - Example: The map $\varphi : V(X^2 + Y^2 - Z^2) \to \mathbb{P}^1$ given by $\varphi[X : Y : Z] = [X + Z : Y]$ is a morphism, since it is defined at all points of $V(X^2 + Y^2 - Z^2)$ as shown earlier.
  - Example: The map $\psi : \mathbb{P}^1 \to V(X^2 + Y^2 - Z^2)$ given by $\psi[S : T] = [S^2 - T^2 : 2ST : S^2 + T^2]$ is a morphism. In fact, it is the inverse of the previous morphism, since we have $(\varphi \circ \psi)[S : T] = \varphi[S^2 - T^2 : 2ST : S^2 + T^2] = [2S^2 : 2ST] = [S : T]$ and $(\psi \circ \varphi)[X : Y : Z] = \psi[X + Z : Y] = [(X + Z)^2 - Y^2 : 2Y(X + Z) : (X + Z)^2 + Y^2] = [2X(X + Z) : 2Y(X + Z) : 2Z(X + Z)] = [X : Y : Z]$.
  - Example: The map $\psi : V(Y^2Z - X^3 - XZ^2) \to V(Y^2Z - X^3 - XZ^2)$ given by $\psi[X : Y : Z] = [X : -Y : Z]$ is a morphism. In fact, it is an isomorphism, since it is its own inverse. (This is the additive inverse map on this elliptic curve.)
  - Example: The map $\psi : V(Y^2Z - X^3 - Z^3) \to V(Y^2Z - X^3 - Z^3)$ given by $\psi[X : Y : Z] = [2XY(Y^2 - 9Z^2) : Y^4 + 18Y^2Z^2 - 27Z^4 : 8Y^3Z]$ is a morphism. (Actually checking that it is well-defined using the definition is rather unpleasant, but it does work out!) This morphism does not magically arise from nowhere, of course: in fact it is simply the doubling map on this elliptic curve.

- Example: More generally, if E is any elliptic curve in projective Weierstrass form, the multiplication-by-m map is a morphism from E to E, as follows from the fact that it has a formula as a rational function (as we saw semi-explicitly in our proof of Nagell-Lutz) and is defined at every point on E.
- Example: If k has characteristic q and V is defined over $\mathbb{F}_q$, the map $\varphi : V \to V$ given by $\varphi[X_0 : X_1 : \cdots : X_n] = [X_0^q : X_1^q : \cdots : X_n^q]$ is a morphism called the Frobenius morphism.
- Example: The map $\varphi : \mathbb{P}^1 \to \mathbb{P}^2$ given by $\varphi[X : Y] = [X^2 : XY : Y^2]$ is a morphism giving an embedding of $\mathbb{P}^1$ into $\mathbb{P}^2$ (it is an example of the general family of d-uple embeddings). The image of φ is the variety $V(XZ - Y^2)$.

- Restricting now to the case of projective curves, we have the following facts:

1. If C_1 is a smooth projective curve, then any rational map $\varphi : C_1 \to C_2$ is automatically a morphism.
   - The idea here is that if P is any point on C_1, then since C_1 is smooth at P (meaning that the local ring $\mathcal{O}_P(V)$ is a DVR), we may choose a local uniformizer t at P (i.e., a generator for the maximal ideal $m_P(V)$).
   - Then we can rescale the components of $\varphi = [\varphi_0 : \varphi_1 : \cdots : \varphi_m]$ by an appropriate power of t in order to make the minimum valuation among the φ_i equal to zero, at which point we see that φ is defined at P.
2. If $\varphi : C_1 \to C_2$ is a nonconstant morphism of projective curves, then φ is surjective, and $k(C_1)$ is a finite-degree extension of $\varphi^*(k(C_2))$. The degree of this field extension $[k(C_1) : \varphi^*(k(C_2))]$ is the degree of the map φ.
   - The first statement follows from the result that the image of a morphism of a projective variety is itself a projective variety (this is usually phrased as saying that projective varieties are complete). Thus, the image $\varphi(C_1)$ is a subvariety of C_2: if its dimension is 1 then since C_2 is irreducible this means $\varphi(C_1) = C_2$, and otherwise if its dimension is 0 then $\varphi(C_1)$ would be a single point and φ would be constant, which we assumed it was not.
   - The fact that $k(C_1)$ is an extension of $k(C_2)$ follows from the fact that φ is surjective, and the fact that the extension has finite degree follows because both $k(C_1)$ and $k(C_2)$ have transcendence degree 1 over k.
3. If $\iota : k(C_2) \to k(C_1)$ is an injection fixing k, then there is a unique nonconstant morphism $\varphi : C_1 \to C_2$ such that $\varphi^* = \iota$.
   - Example: Let C be any smooth projective curve and $\alpha \in k(C)$ be any rational function on C. Then α defines a rational map $\alpha : C \to \mathbb{P}^1$ via $P \mapsto \alpha(P)$, where we take $\alpha(P) = \infty$ when α is not defined at P. By (1) above, this map α is a morphism.
   - In fact, aside from constant maps, the rational maps described above are all possible morphisms from C to $\mathbb{P}^1$: if $\alpha : C \to \mathbb{P}^1$ is a rational map, say $\alpha = [f : g]$, then either g is identically zero in which case α is constant, or the function $\beta = f/g \in k(C)$ has $\beta(P) = \alpha(P)$ for all P so that $\alpha = \beta$.

0.7 (Oct 1) Divisors on Curves

- Our next task is to study divisors on curves, which will be central to our analysis of elliptic curves using the tools from algebraic geometry we have discussed so far.
  - In all of our discussion, C will be a smooth projective curve defined over the algebraically closed field k.
  - We emphasize here that when we we say "points of C", we are implicitly thinking of C as being defined over an algebraically closed field, and the points have coordinates in this algebraically closed field.
  - For notational convenience, we will also usually give examples written in affine form rather than projective form, because the notation is less cumbersome: the point $x \in \mathbb{A}^1$ will be written P_x, while the point $(x, y) \in V(f(x, y))$ inside $\mathbb{A}^2$ will be written $P_{(x,y)}$, and the point at ∞ will be written P_∞.

- Definition: Let C be a smooth curve. The divisor group of C, written $\mathrm{Div}(C)$, is the additive free abelian group generated by the k-points of C. The degree of a divisor $D = \sum_{P \in C} n_P P$ is $\deg(D) = \sum_{P \in C} n_P$. The degree map is a homomorphism from D_C to $\mathbb{Z}$; its kernel is the set of degree-0 divisors $\mathrm{Div}^0(C)$.

- The elements of $\mathrm{Div}(C)$ are of the form $D = \sum_{P \in C} n_P P$ for $n_P \in \mathbb{Z}$, where all but finitely many of the n_P are zero. We will write $\mathrm{ord}_P(D) = n_P$.
- Note that the degree map is well defined since all but finitely many n_P are zero.

- As an aside, we will give a bit of motivation for why divisors are called "divisors", since they seem to have nothing obvious to do with divisibility.
  - If $C_1 = V(f)$ and $C_2 = V(g)$ are two distinct projective plane curves sharing no common component, then their intersection $C_1 \cap C_2 = V(f, g)$ is finite. (Indeed, more is true: Bézout's theorem states that the number of intersection points is at most $\deg(f) \cdot \deg(g)$.)
  - We may then associate a divisor to this intersection $C_1 \cap C_2$ as $\sum_{P \in C_1 \cap C_2} n_P P$, where n_P is the intersection number of $C_1 \cap C_2$ at P given by $n_P = \dim_k \mathcal{O}_P(\mathbb{P}^2)/(f, g)$.
  - For polynomials in one variable, the ideal (f, g) is principal and generated by the gcd of f and g. (One may check that the intersection number at a point P, under the definition above, is the power of $x - P$ that divides their gcd.)
  - For polynomials in two variables (f, g) will no longer be principal, but it still carries the natural sense of being a "common divisor". Thus, we can think (roughly) of the divisor $\sum_{P \in C_1 \cap C_2} n_P P$ as describing the precise way in which the curves C_1 and C_2 intersect.
  - It is not particularly obvious that this value $\dim_k \mathcal{O}_P(\mathbb{P}^2)/(f, g)$ is really the right definition. It is not hard to see that the value is invariant under linear changes of coordinates, and that the intersection number is 1 whenever P is a simple point of C_1 and C_2 where C_1 and C_2 meet transversally (i.e., their tangent lines at P are different). It is also additive when we take unions of curves.
  - We will not really use this particular formulation of divisors; it is merely some motivation for how divisors arise in a fairly natural way in the context of curves.

- If E is a subfield of the algebraically closed field k over which C is defined, the Galois group $\mathrm{Gal}(k/E)$ acts on the k-rational points of C, and thus it also acts on divisors pointwise.

- Definition: Suppose C is a smooth curve defined over the algebraically closed field k, and E is a subfield of k with $\overline{E} = k$. If $\sigma \in \mathrm{Gal}(k/E)$ is an element of the Galois group and $D = \sum_{P \in C} n_P P$ is a divisor, we define the action of σ on D via $\sigma(D) = \sum_{P \in C} n_p \sigma(P)$. We then say a divisor D is defined over E when $\sigma(D) = D$ for all $\sigma \in \mathrm{Gal}(k/E)$, and we denote the subgroup of divisors defined over E as $\mathrm{Div}_E(C)$.
  - If all of the points with nonzero coefficients in D are defined over E then certainly D is defined over E, but this is not necessary. All that is required is for Galois-conjugate points to have the same coefficients, as is seen immediately by comparing $\sigma(D) = \sum_{P \in C} n_p \sigma(P)$ to the reindexed sum $D = \sum_{P \in C} n_{\sigma(P)} \sigma(P)$: one requires $n_{\sigma(P)} = n_P$ for all $P \in C$ and all $\sigma \in \mathrm{Gal}(k/E)$.
  - For example, for the curve $C = \mathbb{A}^1(\mathbb{C})$, with $P = i$ and $Q = -i$, the divisor $2P + Q$ is defined over $\mathbb{Q}(i)$ (any element of the Galois group $\mathbb{C}/\mathbb{Q}(i)$ fixes i and $-i$, hence sends P to P and Q to Q) while the divisor $P + Q$ is defined over $\mathbb{Q}$ (any element of the Galois group $\mathbb{C}/\mathbb{Q}$ either fixes i or maps it to $-i$, and these operations map $P + Q$ to $P + Q$ or $Q + P$ respectively).

- We can attach a divisor to a rational function on C using its zeroes and poles:

- Definition: Let C be a smooth curve and $\alpha \in k(C)$ be a nonzero rational function on C. We define the divisor of α, denoted $\mathrm{div}(\alpha)$, as $\mathrm{div}(\alpha) = \sum_{P \in C} v_P(\alpha) P$. The divisors of the form $\mathrm{div}(\alpha)$ for some $\alpha \in k(C)^\times$ are called principal divisors.
  - Remark: In many sources, the divisor of α is often written (α). In our context, this can lead to ambiguities, since the same notation is also used for the ideal generated by α. As such, we will only ever write $\mathrm{div}(\alpha)$ for the divisor of α.
  - As we have already shown, for any nonzero α, $v_P(\alpha)$ is nonzero only for finitely many $P \in C$, so $\mathrm{div}(\alpha)$ is well defined.
  - Now because $\mathrm{ord}_P(\alpha/\beta) = \mathrm{ord}_P(\alpha) - \mathrm{ord}_P(\beta)$, summing over all primes shows that $\mathrm{div}(\alpha/\beta) = \mathrm{div}(\alpha) - \mathrm{div}(\beta)$, so the principal divisors are a subgroup of the divisor group $\mathrm{Div}(C)$.

- When computing the divisor of a function on a smooth curve C, remember that all of our curves are (implicitly) projective curves. (Thus, even when we are working with an affine equation, we must also include the point at ∞.)
- Example: For $C = \mathbb{P}^1(\mathbb{C})$, with points denoted $[X : Y]$, for the rational function $\alpha = \dfrac{X}{Y}$ we have $\mathrm{div}(\alpha) = P_{[0:1]} - P_{[1:0]}$ where $P_{[0:1]}$ represents the point $[0 : 1]$ (i.e., the point where $X = 0$) and $P_{[1:0]}$ represents the point $[1 : 0]$ (i.e., the point where $Y = 0$).
- Example: For $C = \mathbb{P}^1(\mathbb{C})$, for the rational function $\beta = \dfrac{X^3 - XY^2}{Y^3}$ we have $\mathrm{div}(\beta) = P_{[0:1]} + P_{[1:1]} + P_{[-1:1]} - 3P_{[1:0]}$, using the same notation as above. Note that β has three single zeroes at $[0 : 1]$, $[1 : 1]$, and $[-1 : 1]$ and a triple pole at $[1 : 0]$.
- If we dehomogenize the two examples above, so as to work instead with the affine line $\mathbb{A}^1$, the corresponding rational functions are $\alpha_* = x$ and $\beta_* = x^3 - x$, with associated divisors $\mathrm{div}(\alpha_*) = P_0 - P_\infty$ and $\mathrm{div}(\beta_*) = P_0 + P_1 + P_{-1} - 3P_\infty$.
- Exercise: On $C = \mathbb{A}^1(\mathbb{C})$, suppose $\alpha = u\dfrac{(x-p_1)^{a_1}\cdots(x-p_l)^{a_l}}{(x-q_1)^{b_1}\cdots(x-q_m)^{b_m}}$ for $u \in k^\times$ and take distinct elements $p_1, \ldots, p_l, q_1, \ldots, q_m \in k$ having associated points $P_1, \ldots, P_k, Q_1, \ldots, Q_l$ respectively. Show that $\mathrm{div}(a) = a_1P_1 + \cdots + a_lP_l - b_1Q_1 - \cdots - b_mQ_m + [\sum_j b_j - \sum_i a_i]\infty$. [Hint: This is just a generalization of the examples above.]
- Exercise: Show that for any $C = \mathbb{A}^1(\mathbb{C})$ and any nonzero rational function $\alpha \in \mathbb{C}(C)$ we have $\deg(\mathrm{div}(\alpha)) = 0$.
- Example: For $C = V(Y^2Z - X^3 - XZ^2)$ consider the rational function $\gamma = \dfrac{Y}{Z}$. The zeroes for γ can only occur when $Y = 0$ yielding the points $[0 : 0 : 1]$, $[i : 0 : 1]$, $[-i : 0 : 1]$, while the poles for γ can only occur when $Z = 0$ yielding the point $[0 : 1 : 0]$. To compute the order of vanishing γ at each point we may compute a local uniformizer (for the three zeroes, $\gamma = Y/Z$ is itself a local uniformizer, while for the pole, Z/X is a local uniformizer). One obtains $\mathrm{ord}_{[0:0:1]}\gamma = \mathrm{ord}_{[i:0:1]}\gamma = \mathrm{ord}_{[-i:0:1]}\gamma = 1$ and also $\mathrm{ord}_{[0:1:0]}\gamma = -3$, so $\mathrm{div}(\gamma) = P_{[0:0:1]} + P_{[i:0:1]} + P_{[-i:0:1]} - 3P_{[0:1:0]}$.
- If we dehomogenize the example above, so as to work instead with the affine model $y^2 = x^3 + x$, the corresponding rational function is $\gamma_* = y$ with associated divisor $\mathrm{div}(\gamma_*) = P_{(0,0)} + P_{(i,0)} + P_{(-i,0)} - 3P_\infty$.

- Motivated by the calculations for $K = \mathbb{C}(t)$, we can also pick out the zeroes (respectively, poles) of an element by extracting only the portion of its divisor with positive (respectively, negative) coefficients:
- Definition: If $\alpha \in k(C)^\times$ has divisor $\mathrm{div}(\alpha) = \sum_P n_P P$, we define the "zero divisor" $\mathrm{div}_+(\alpha) = \sum_P \max(0, n_P)P = \sum_{P:n_P>0} n_P P$ and the "pole divisor" $\mathrm{div}_-(a) = \sum_P \min(0, n_P)P = \sum_{P:n_P<0} n_P P$.
  - Notice that $\mathrm{div}(\alpha) = \mathrm{div}_+(\alpha) - \mathrm{div}_-(\alpha)$ for any element $\alpha \in k(C)^\times$.
  - Remark: There are various other notations for these quantities that are often used, such as $(a)_0$ for div_+ and $(a)_\infty$ for div_-, which are intended to evoke the idea of picking out the zeroes and poles of a.
  - Exercise: For any field k, if $f(t), g(t) \in k[t]$ are relatively prime, show that $[k(t) : k(\frac{f(t)}{g(t)})] = \max(\deg f, \deg g)$. [Hint: Use Gauss's lemma to show that $q(y) = f(y) - \frac{f(t)}{g(t)}g(y) \in k(\frac{f(t)}{g(t)})[y]$ is the minimal polynomial of t over $k(\frac{f(t)}{g(t)})$.]
  - In the example above, we can also compute that $\deg(\mathrm{div}_+(\alpha)) = \deg(\mathrm{div}_-(\alpha)) = \deg(f) = \deg(g)$, and by the exercise above, this quantity is equal to the extension degree $[k(C) : k(\alpha)]$. In fact, this result is true in general:

- Theorem (Divisor Degrees): For any nonconstant $\alpha \in k(C)^\times$ on a curve C/k, we have $\deg(\mathrm{div}_+(\alpha)) = \deg(\mathrm{div}_-(\alpha)) = [k(C) : k(\alpha)]$. As a consequence, $\deg(\mathrm{div}(\alpha)) = 0$ for all such α.
  - We will defer the proof of this result temporarily, since it would otherwise require developing a lot of additional material out of order.
  - Our main observation here is that the divisor of an element $a \in k(C)^\times$ always has degree 0, which is to say, the principal divisors are actually a subgroup of the group of degree-0 divisors.

- Definition: On a curve C/k, we say two divisors D_1 and D_2 are linearly equivalent (and write $D_1 \sim D_2$) if $D_1 - D_2$ is principal. The resulting equivalence classes (i.e., divisors modulo principal divisors) form a group called the class group, or the Picard group, of C.
    - Exercise: Verify that this relation is an equivalence relation and that the equivalence classes are the elements in the quotient group of divisors modulo principal divisors.
    - Some notation for all of these various groups: $\mathrm{Div}(C) = D_C$ is the group of all divisors on C, $\mathrm{Div}^0(C)$ is the group of degree-0 divisors on C, $\mathrm{Cl}(C) = \mathrm{Pic}(C) = \mathrm{Div}(C)/[\text{principal divisors}]$ is the class group of C.
    - Since principal divisors all have degree zero, we can also form the reduced Picard group $\mathrm{Pic}^0(C) = \mathrm{Div}^0(C)/[\text{principal divisors}]$.
- For $\mathbb{P}^1$, the reduced Picard group is trivial:
- Proposition (Reduced Picard Group of $\mathbb{P}^1$): If $C = \mathbb{P}^1$, then $\mathrm{Pic}^0(C) = \mathrm{Div}(C)/[\text{principal divisors}]$ is the trivial group, and $\mathrm{Pic}(C) \cong \mathbb{Z}$.
    - Proof: The result is equivalent to showing that every divisor of degree 0 is principal, so suppose $D = \sum_P b_P P$ has degree 0.
    - For a point $P = [a : b] \in \mathbb{P}^1$ let f_P be the rational function $f_P = \dfrac{bX - aY}{Y}$, whose divisor is easily seen to be $\mathrm{div}(f_P) = P - P_\infty$.
    - Now consider the rational function $\alpha = \prod_P f_P^{b_P}$: by the calculation above we have $\mathrm{ord}_P(\alpha) = b_P$ for each point $P \neq \infty$, but since $\sum_P b_P \deg(P) = 0$ by the assumption on D, and $\deg(\mathrm{div}(\alpha)) = 0$ as well, we must have $\mathrm{ord}_\infty(a) = b_\infty$ also.
    - Then $\mathrm{ord}_P(\alpha) = b_p$ for all $P \in \mathbb{P}^1$, so $\mathrm{div}(\alpha) = D$ and so D is principal as claimed.
    - The statement that $\mathrm{Pic}(K) \cong \mathbb{Z}$ follows immediately from $\mathrm{Div}(K)/\mathrm{Div}^0(K) \cong \mathbb{Z}$.
    - Remark: It can be shown that the case $C = \mathbb{P}^1$ is essentially the only situation where the reduced Picard group is trivial. So do not be misled by the convenience of this particular result!
- We have a fundamental analogy between divisors on curves and ideals of algebraic number fields.
    - If $K/\mathbb{Q}$ is an algebraic number field, we have an exact sequence $1 \to [\text{units of } \mathcal{O}_K] \to K^* \to [\text{fractional ideals of } \mathcal{O}_K] \to [\text{ideal class group of } K] \to 1$.
    - If C is an algebraic curve defined over k, the analogous exact sequence is $1 \to k^* \to k(C)^* \to \mathrm{Div}^0(C) \to \mathrm{Pic}^0(C) \to 1$.
    - The constant field k plays the role of the units of an algebraic number field, the group of degree-0 divisors plays the role of the fractional ideals in the ring of integers, and the reduced Picard group plays the role of the ideal class group.
- We now put a partial ordering on divisors that is motivated by the idea of divisibility for integers and rational functions.
    - The idea is that if we look at p-adic valuations of elements of $\mathbb{Q}$, we can identify the elements of $\mathbb{Z}$ as those whose valuations are nonnegative at every finite prime p.
    - The same principle holds for considering valuations of a rational function at points on an algebraic curve C: we can identify polynomial functions as those having no poles except at the points at ∞.
- Definition: If a divisor $D = \sum_P n_P P$ on a curve C/k has $n_P \geq 0$ at all points P, we say D is effective and we write $D \geq 0$. We extend this notion to a partial ordering on divisors by writing $D_1 \leq D_2$ if and only if $D_2 - D_1$ is effective.
    - Exercise (easy): Check that the relation $D_1 \leq D_2$ is a partial ordering on divisors.
    - The partial ordering on divisors allows us to specify the order of zeroes and poles: to illustrate, for $\mathbb{A}^1$, saying that f has a pole of order at most 2 at $x = 0$ and a zero of order at least 3 at $x = 1$ is equivalent to saying $\mathrm{div}(f) \geq 2P_0 - 3P_1$.

- Definition: If D is a divisor on a curve C/k, the Riemann-Roch space associated to D is the set $L(D) = \{\alpha \in k(C)^\times : \operatorname{div}(\alpha) \geq -D\} \cup \{0\}$. Equivalently, an element $\alpha \in k(C)^\times$ is in $L(D)$ if and only if $v_P(a) \geq -v_P(D)$ at all points $P \in C$.
  - When D is an effective divisor, $L(D)$ represents all rational functions whose poles are "no worse" than D.
  - More generally, if $D = \sum_P n_P P - \sum_Q m_Q Q$ with $n_i, m_i > 0$, then $L(D)$ consists of all $\alpha \in k(C)^\times$ such that α has a zero of order at least m_Q at each point Q, and may have poles only at the points P, of order at most n_P at P.
  - It is not hard to see that $L(D)$ is a k-vector space: if $\alpha, \beta \in L(D)$, then $\alpha+\beta \in L(D)$ because $v_P(\alpha+\beta) \geq \min(v_P(\alpha), v_P(\beta))$ for each point P, and $c\alpha \in L(D)$ for all $c \in k$ since $v_P(c\alpha) = v_P(c) + v_P(\alpha) = v_P(\alpha)$ since $v_P(c) = 0$ at all points P.
  - Example: For $C = \mathbb{A}^1$ and $D = P_0$, we can see that $L(D) = \operatorname{span}(1, x^{-1})$, since the only possible poles of an element $f/g \in L(D)$ function occur at $x = 0$ (of order 1) and the function must also have $\deg g \geq \deg f$ since there is no pole at the point at infinity P_∞.
  - Example: For $C = \mathbb{A}^1$ and $D = 3P_\infty$, we can see that $L(D) = \operatorname{span}(1, x, x^2, x^3)$ since the function f/g has no poles except a pole of order at most 3 at P_∞ (meaning that $\deg g \leq \deg f + 3$), which is to say, f/g is a polynomial of degree at most 3.
  - Example: For $C = \mathbb{A}^1$ and $D = -P_0$, we can see that $L(D) = \{0\}$, since any nonzero element $f/g \in L(D)$ would need to be zero at $x = 0$ and defined at all other points, but this cannot occur because g would have to be constant, but then $\deg f > \deg g$ would force f/g to have a pole at P_∞.
  - Example: For arbitrary C/k, we have $L(0) = k$, since $\operatorname{div}(c) = 0$ for all $c \in k^\times$, but any element $x \in k(C)^\times \backslash k$ necessarily has at least one pole (its degree as a rational function must be positive, and then any zero of the denominator yields a pole).
  - Exercise: Determine $L(D)$ when $C = \mathbb{A}^1(\mathbb{C})$ for $D = P_0 - P_\infty$, $P_0 + P_\infty$, and $P_0 + P_1$.
  - We can also consider Riemann-Roch spaces over non-algebraically-closed fields. The only alteration to considering $L_E(D)$ for some subfield E of its algebraic closure k is that $L_E(D) = \{\alpha \in E(C)^\times : \operatorname{div}(\alpha) \geq -D\} \cup \{0\}$ consists only of the elements of the function field that are defined over E.
  - Example: For $C = \mathbb{A}^1$ and $D = P_\infty - P_i$, we have $L_{\mathbb{R}}(D) = \{0\}$ because any such rational function f/g would necessarily be a polynomial in $\mathbb{R}[x]$ of degree at most 1 (since it could only have a pole of order 1 at ∞) and would have to be zero at $x = i$, but any such polynomial would also be zero at $x = -i$ meaning that its degree is too large.
  - Example: For $C = \mathbb{A}^1$ and $D = 2P_\infty - P_i - P_{-i}$, we have $L_{\mathbb{R}}(D) = \operatorname{span}(1 + x^2)$ because as above any element would be a real polynomial of degree at most 2 that is zero at both $x = i$ and $x = -i$, hence is a multiple of $1 + x^2$.
  - Note that the field of definition affects the dimension in the first example above but not the second, since over $\mathbb{C}$ we have $L_{\mathbb{C}}(P_\infty - P_i) = \operatorname{span}(i - x)$ but $L_{\mathbb{C}}(2P_\infty - P_i - P_{-i}) = \operatorname{span}(1 + x^2)$; that difference arises merely because the divisor D is actually defined over $\mathbb{R}$ while the divisor from the first example is not.
  - Exercise: Suppose E is a subfield of k and D is a divisor of k that is defined over E. Show that $\dim_k[L_k(D)] = \dim_E[L_E(D)]$. [Hint: Show that a basis for L_E remains a basis over L_k.]

Well, you're at the end of my handout. Hope it was helpful.