Contents

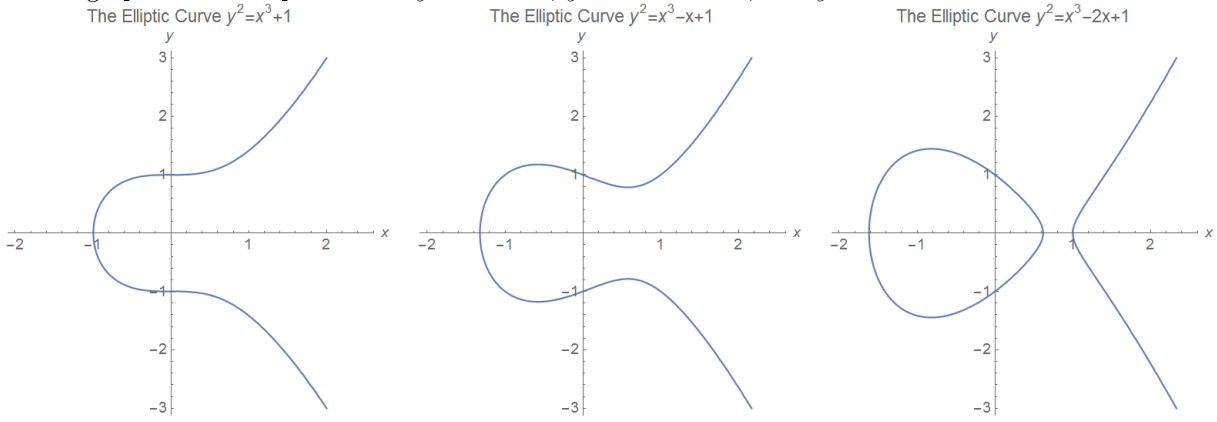
0 Elliptic Curves and Modular Forms

These are lecture notes for the graduate course Math 7359: Elliptic Curves and Modular Forms, taught at Northeastern in Fall 2023.

0.1 (Sep 7) 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 end goals 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.

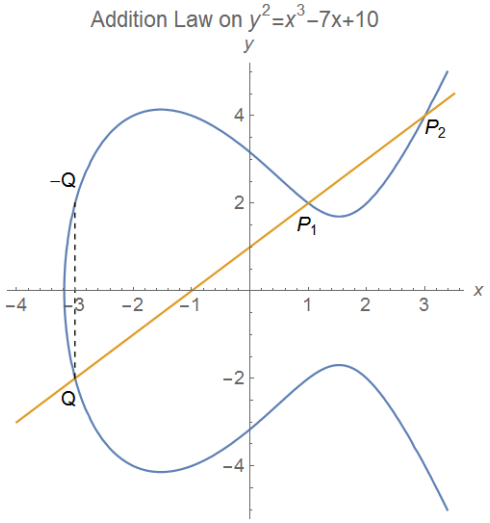
- Once we construct a third point on an elliptic curve this way, we might try to find more points.

  - 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:



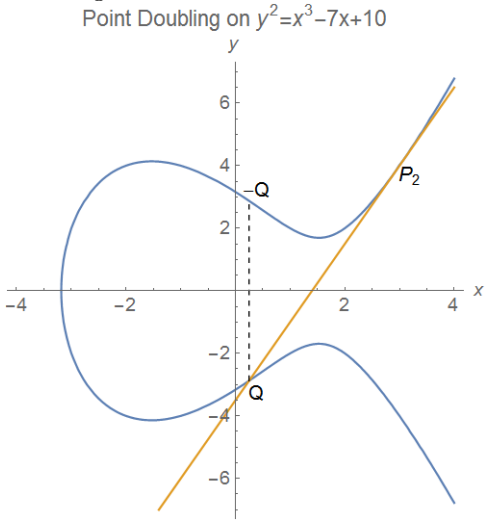
  - The point Q lies on the intersection of $y = x + 1$ and $y^2 = x^3 - 7x + 10$, so $(x + 1)^2 = x^3 - 7x + 10$.
  - This equation is equivalent to $x^3 - x^2 - 9x + 9 = 0$, which factors as $(x - 1)(x - 3)(x + 3) = 0$. Then the x-coordinate of Q is -3 so $Q = (-3, -2)$.
  - Thus, the sum $P_1 + P_2$ 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).

[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).

- 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 $\mathbb{Z}/5\mathbb{Z}$, 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 $P_1 + P_2 = \boxed{(0,3)}$.
  - 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 $P_1 = P_2$.

0.2 (Sep 11) The Group Structure of Elliptic Curves, Nagell-Lutz

- 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.
- 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.

0.3 (Sep 14) More Curves over $\mathbb{Q}$, Mordell's Theorem

- 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: it is $[2]Q = (523620448448041618543485494416 81/434625338111430357812426490000, -351$
  - 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 $\mathbb{Z} \times \mathbb{Z} \times (\mathbb{Z}/2\mathbb{Z})$.

- 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.

[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).

- For any given elliptic curve E, the torsion subgroup $E_{\mathrm{Tor}}(\mathbb{Q})$ can be computed using Nagell-Lutz as we discussed above.
- 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. Elkies has given a construction for an elliptic curve that has rank at least 28 (and it is expected this curve has rank exactly 28)[3]. It has 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 $x^2 + xy + y = x^3 - x^2 - 20067762415575526585033208209338542750930230312178956502x + 3448161179503055646703298569039072037485594435931918036126600829629193944873224342 9$

- Before proceeding with the actual argument, we outline the idea. What we will show is that there exists a homomorphism $\varphi : E(\mathbb{Q})$ with kernel $2E(\mathbb{Q})$ 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 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 $U/2U$, the group of units modulo squares inside R.
  - By the exercise, since R is a direct sum of number fields, the group $U/2U$ 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 18) Affine Space, Affine Algebraic Sets

- 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 $I : [\text{sets}] \to [\text{ideals}]$ and $V : [\text{ideals}] \to [\text{sets}]$.

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 finite, show that the irreducible affine algebraic sets in $\mathbb{A}^n(k)$ are $\emptyset$ and single points.

- Exercise: If k is infinite, 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, $I(V(I)) = \mathrm{rad}(I)$.
    - However, even if I is radical, it is not always true that $I(V(I)) = \mathrm{rad}(I)$: for example, in $\mathbb{A}^1(\mathbb{R})$ we have $V(1 + x^2) = \emptyset$ so that $I(V(1 + x^2)) = \mathbb{R}[x]$.

[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.)

- 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 $I(V(I)) = \mathrm{rad}(I)$.
  - 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 $\mathrm{rad}(I) \subseteq I(V(I))$, one uses the "Rabinowitsch trick" for the other containment.
  - Explicitly, if $g \in I(V(f_1, \ldots, f_r))$, one considers 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 $g \in \mathrm{rad}(I)$.

- 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 $I(V(I)) = \mathrm{rad}(I)$, 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 $I(V(I)) = \mathrm{rad}(I) = I$ 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]

0.5 (Sep 21) Functions on Affine Varieties

- 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 $k = \mathbb{C}$.

- 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 $\mathcal{O}_P(C)$, and thus $m_P(C) = (x,y) = (y)$, as claimed.
- 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.
   - 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 $v(r) \geq 0$.

- 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 some 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.

0.6 (Sep 25) Projective Space and Projective Varieties

- 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.

[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.

- 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.
    - 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 $V(\{a_ix_j - a_jx_i\}_{0 \le i,j \le n}) = \{[a_0 : a_1 : \cdots : a_n]\}$.
    3. $I(\emptyset) = k[x_0, \ldots, x_n]$, $I(\mathbb{P}^n) = 0$ when k is infinite, and $I(\{[a_0 : a_1 : \cdots : a_n]\}) = (\{a_ix_j - a_jx_i\}_{0 \le i,j \le n})$.
    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.

- 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 affine 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 $\Gamma(V) = k[x, y]/(Y^2 + Z^2 - X^2)$, 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 $x_0^{\deg(f)+\deg(g)-\deg(f+g)}(f + g)^* = x_0^{\deg(g)} f^* + x_0^{\deg(f)} g^*$.

- 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.

0.7 (Sep 28) Rational Maps and Morphisms

- 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 isomorphisms $\varphi : \mathbb{A}^n \to \mathbb{A}^n$ are the invertible affine linear transformations, of the form $\varphi(x) = Ax + b$ where A is an invertible $n \times n$ matrix and b is any vector of constants. (Hint: First show that the degree of each coordinate in φ and ψ must be 1.)
- 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), and where $\psi_i(P) \neq 0$ 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 on function fields $\varphi^* : k(W) \to k(V)$ (i.e., a ring homomorphism fixing k) 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 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 $\alpha = \beta$.

0.8 (Oct 2)

- Class cancelled due to instructor injury.

0.9 (Oct 5) 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 $\text{div}(f) \geq 2P_0 - 3P_1$.

- Definition: If D is a divisor on a curve C/k, the Riemann-Roch space associated to D is the set $L(D) = \{\alpha \in k(C)^\times : \text{div}(\alpha) \geq -D\} \cup \{0\}$. 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 $L(D) = \text{span}(1, x^{-1})$, 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 $L(D) = \text{span}(1, x, x^2, x^3)$ 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 $\text{div}(c) = 0$ 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 $L_E(D) = \{\alpha \in E(C)^\times : \text{div}(\alpha) \geq -D\} \cup \{0\}$ 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 $L_\mathbb{R}(D) = \text{span}(1 + x^2)$ 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 $L_\mathbb{C}(P_\infty - P_i) = \text{span}(i - x)$ but $L_\mathbb{C}(2P_\infty - P_i - P_{-i}) = \text{span}(1 + x^2)$; 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.]

0.10 (Oct 12) The Riemann-Roch Theorem + Elliptic Curves (Properly)

- Definition: If D is a divisor on a curve C/k, we define $\ell(D) = \dim_k L(D)$.
  - Examples: By the examples worked out above, for $C = \mathbb{A}^1(\mathbb{C})$ we have $l(P_0) = 2$, $l(3P_\infty) = 4$, and $l(-P_0) = 0$.

- Example: For an arbitrary C, we have $\ell(0) = 1$, since $L(0) = k$.

- Proposition (Properties of $l(D)$): Let C be an algebraic curve over k and let D be a divisor of C.

  1. If $D_1 \leq D_2$, then $\ell(D_1) \leq \ell(D_2)$.
     - Proof: This follows immediately from the definition, since $D_1 \leq D_2$ clearly implies that $L(D_1)$ is a subspace of $L(D_2)$.
  2. If $D_1 \sim D_2$, then $L(D_1) \cong L(D_2)$ and so $\ell(D_1) = \ell(D_2)$.
     - Proof: If $D_1 = D_2 + \operatorname{div}(g)$, then the map from $L(D_1)$ to $L(D_2)$ sending $f \mapsto fg$ is easily seen to be an isomorphism of vector spaces since it has an inverse map $h \mapsto h/g$.
  3. If $\deg(D) \leq 0$, then $L(D) = \{0\}$ and $l(D) = 0$ except when $D = \operatorname{div}(\alpha)$ is principal, in which case $L(D) = \operatorname{span}(\alpha)$ and $l(D) = 1$.
     - Proof: Suppose $f \in L(D)$ and $f \neq 0$. Then $0 = \deg(\operatorname{div}(f)) \geq \deg(-D) = -\deg(D)$.
     - Furthermore, equality can hold only if $D = -\operatorname{div}(f)$ for some $f \in k(C)^\times$, in which case D is principal.
     - If D is principal, then $\ell(D) = \ell(0) = 1$ by (2), and $L(D) = \operatorname{span}(\alpha)$ by the same calculation.
  4. If D_1 and D_2 are divisors with $D_1 \leq D_2$, then $\dim_k(L(D_2)/L(D_1)) \leq \deg(D_2) - \deg(D_1)$.
     - Proof: Induct on the sum of the coefficients of the points in the effective divisor $B - A$. The base case $B - A = 0$ is trivial.
     - For the inductive step, suppose that $D_2 = D_1 + P$ for some point P, and choose $x \in k(C)$ such that $v_P(x) = v_P(D_2) = v_P(D_1) + 1$.
     - Then for any $y \in L(D_2)$, we have $v_P(xy) = v_P(x) + v_P(y) \geq v_P(D_2) - v_P(D_2) \geq 0$, so $xy \in \mathcal{O}_P$, the local ring at P.
     - By composing with the evaluation map at P, we obtain a k-linear transformation $\varphi : L(D_2) \to \mathcal{O}_P/m_P \cong k$ with $\varphi(y) = (xy)(P)$.
     - Then $y \in \ker(\varphi)$ if and only if $(xy)(P) = 0$ if and only if $v_P(xy) \geq 1$ if and only if $v_P(y) \geq 1 - v_P(D_2) = -v_P(D_1)$, and this last statement is equivalent to $y \in L(D_1)$.
     - Thus, by the first isomorphism theorem, we have an injection from $L(D_2)/L(D_1)$ to $\mathcal{O}_P/m_P$. Taking dimensions yields $\dim_k(L(D_2)/L(D_1)) \leq \dim_k(\mathcal{O}_P/m_P) = 1$.
     - This establishes the inductive step, so the general result follows.
  5. For any effective divisor D, we have $\ell(D) \leq \deg(D) + 1$. In fact, this inequality holds for any divisor D of degree ≥ 0.
     - Proof: For effective divisors, this follows immediately by induction on the degree of D using (4), starting with the base case $l(0) = 1$.
     - For general divisors, the result is trivial if $\ell(D) = 0$, so suppose otherwise that $\ell(D) \geq 1$ and let $\alpha \in L(D)$ be nonzero. Then $\operatorname{div}(\alpha) \geq -D$ which is equivalent to $D - \operatorname{div}(\alpha^{-1}) \geq 0$.
     - Then for $D' = D - \operatorname{div}(\alpha^{-1})$, we see that D is equivalent to the effective divisor D', and so by (2) we have $\ell(D) = \ell(D') \leq \deg(D') + 1 = \deg(D) + 1$, as required.
  6. For any divisor D, the quantity $\ell(D)$ is finite.
     - Proof: If $\deg(D) < 0$ then (3) gives $\ell(D) = 0$, while if $\deg(D) \geq 0$ then (5) gives $\ell(D) \leq \deg(D) + 1$.

- What we would like to be able to do now is to calculate the actual dimension $\ell(D)$ for arbitrary divisors D. Rather than delaying the point, we will now get right to our main result:

- Theorem (Riemann-Roch): For any algebraic curve C/k, there exists an integer $g \geq 0$ called the genus of C, and a divisor class $\mathcal{C}$, called the canonical class of C, such that for any divisor $C \in \mathcal{C}$ and any divisor $A \in \operatorname{Div}(K)$, we have $\ell(A) = \deg(A) - g + 1 + \ell(C - A)$.

  - Remark: The divisor class $\mathcal{C}$, as we will explain later in our discussion of differentials, is the divisor class associated with the meromorphic differentials of C.

- Proving the general Riemann-Roch theorem would take us a bit too far our of our way at the moment, so we will instead just derive some of its important consequences in our context.

- Later, when we are discussing elliptic curves over $\mathbb{C}$, we can outline the argument for Riemann-Roch for Riemann surfaces. It contains most of the main ideas but is more accessible since the complex-analytic notion of a differential is quite natural.

- For now, we will run through some consequences of the Riemann-Roch theorem.
- Proposition (Corollaries of Riemann-Roch): Let C/k be an algebraic curve.

1. For any divisor A with $\deg(A) \geq 0$, we have $\deg(A) - g + 1 \leq \ell(A) \leq \deg(A) + 1$.
   - Proof: We showed the upper bound earlier using an inductive argument. The lower bound follows immediately from Riemann-Roch since $\ell(C - A) \geq 0$.
2. For $C \in \mathcal{C}$ we have $\ell(C) = g$ and $\deg(C) = 2g - 2$.
   - Proof: First set $A = 0$ in Riemann-Roch: this yields $\ell(0) = \deg(0) - g + 1 + \ell(C)$, so since $\ell(0) = 1$ and $\deg(0) = 0$, we get $\ell(C) = g$.
   - Now set $A = C$ in Riemann-Roch: this yields $\ell(C) = \deg(C) - g + 1 + \ell(0)$, and so $\deg(C) = \ell(C) + g - 1 - \ell(0) = 2g - 2$.
3. If $\deg(A) \geq 2g - 2$, then $\ell(A) = \deg(A) - g + 1$ except when $A \in \mathcal{C}$ (in which case $\ell(A) = g$).
   - Proof: If $\deg(A) \geq 2g-2$, then $\deg(C-A) \leq 0$, and so $\ell(C-A) = 0$ except when $C-A$ is principal (i.e., when $A \in \mathcal{C}$).
   - When $\ell(C - A) = 0$ Riemann-Roch immediately gives $\ell(A) = \deg(A) - g + 1$, and when $A \in \mathcal{C}$ we have $\ell(A) = g$ by (2).
4. The genus g is unique, as is the equivalence class $\mathcal{C}$.
   - Proof: Pick A of sufficiently large degree: then $\deg(A) - \ell(A) + 1 = g$ by (3), so g is uniquely determined.
   - For $\mathcal{C}$, if $\ell(A) = \deg(A) - g + 1 + \ell(C - A) = \deg(A) - g + 1 + \ell(D - A)$ for some other divisor D, then $\ell(C - A) = \ell(D - A)$ for all A.
   - Setting $A = C$ yields $\ell(D - C) = 1$ and setting $A = D$ yields $\ell(C - D) = 1$, and these are contradictory unless $D - C$ is principal, which is to say, $D \sim C$.

- Our main highlight is that we can use Riemann-Roch to study curves of small genus over an arbitrary field F with algebraic closure k. We start with the simplest genus $g = 0$, so suppose that C is a (smooth projective) curve of genus 0 over the field F, and let $K = F(C)$ be its function field.
  - By Riemann-Roch, we have $\ell(A) = \deg(A) + 1 + \ell(C - A)$ for any divisor A, and also $\deg(C) = -2$.
  - Also, by (3), if $\deg(A) \geq -1$ then $\ell(A) = \deg(A) + 1$. In particular, since $\deg(-C) = 2$, we have $\ell(-C) = 3$.
  - Now, for any point P, we have $\ell(P) \leq \deg(P) + 1$. So, if P is any point with $P \leq C$ (there must be at least one since $\deg(-C)$ is positive), we see $\ell(P) \leq \ell(-C) = 3$. Thus, $\deg(P)$ must be either 1 or 2.
  - First suppose that there is a point P of degree 1. Then $\ell(P) = 2$. Since F is a subspace of $L(P)$, there is a basis of $L(P)$ of the form $\{1, x\}$ for some $x \notin F$.
  - Then since $\deg(\mathrm{div}(x) + P) = 1$ and $\mathrm{div}(x) + P \geq 0$, we must have $\mathrm{div}(x) + P = Q$ for some point Q (necessarily of degree 1). Then $\mathrm{div}(x) = P - Q$, and so $[K : F(x)] = \deg(\mathrm{div}_+(x)) = \deg(P) = 1$, which means $K = F(x)$.
  - Now suppose there is no point P of degree 1: per earlier, we have a point $P \leq C$ of degree 2.
  - Then $\ell(P) = 3$, so again since $L(P)$ contains k, we may take a basis for $L(P)$ of the form $\{1, x, y\}$ for some F-linearly independent $x, y \notin F$.
  - In the same way as above, we see that $\mathrm{div}(x) = P - Q$ and $\mathrm{div}(y) = P - R$ for some (necessarily distinct) points Q and R of degree 2.
  - Then $[K : F(x)] = \deg(\mathrm{div}_+(x)) = 2$ and $[K : F(y)] = \deg(\mathrm{div}_+(y)) = 2$ also. Since $F(x) \neq F(y)$ (by linear independence and the fact that K is a degree-2 extension of both), we see $K = F(x, y)$.
  - Furthermore, Riemann-Roch says that $\ell(2P) = 1 + \deg(2P) = 5$, but we can find six different elements in $L(2P)$, namely $\{1, x, y, x^2, xy, y^2\}$. They must therefore be F-linearly dependent, so we see that x and y satisfy some quadratic relation $ax^2 + bxy + cy^2 + dx + ey = f$.

- Geometrically, this case corresponds to a conic, while the case $K = F(x)$ corresponds to a line (since we can think of $F(x) = F(x, y)$ where y is a linear function of x).

- We can use similar ideas to study curves of genus 1: now suppose that C is a curve of genus 1 over the field F, again with function field K.
  - In this case, for $g = 1$ Riemann-Roch and its corollaries say that $\ell(A) = \deg(A) + \ell(C - A)$, that $\deg(C) = 0$ and $\ell(C) = 1$, and that if $\deg(A) \geq 1$ then $\ell(A) = \deg(A)$.
  - Unlike the case $g = 0$, we are not necessarily guaranteed to have a point of any given degree any more, since we cannot use C to construct a point of small degree – indeed, since $\deg(C) = 0$ and $\ell(C) = 1$, in fact C is principal (and $C \sim 0$).
  - So let us instead merely suppose that we do have a point P of degree 1. Then $\ell(2P) = 2$, so choose a basis $\{1, x\}$ for $L(2P)$, where we necessarily must have $v_P(x) = 2$ since $x \notin L(P)$. Then $\ell(3P) = 3$, so choose a basis $\{1, x, y\}$ for $L(3P)$, where we must necessarily have $v_P(y) = 3$ since $y \notin L(2P)$.
  - Then, as above, $[K : F(x)] = \deg(\mathrm{div}_+(x)) = 2$ and $[K : F(y)] = \deg(\mathrm{div}_+(y)) = 3$, so since 2 and 3 are relatively prime, we see $K = F(x, y)$.
  - Now we would like to identify what kind of algebraic relation x, y must satisfy (they are, after all, algebraically dependent), which we can do by looking at the spaces $L(kP)$ for larger values of k, since the various monomials $x^i y^j$ will all only have poles at P.
  - We have $\ell(4P) = 4$, but we can only identify 4 elements that must lie in this space: $\{1, x, y, x^2\}$. In fact, they are all linearly independent since they all have different valuations at P.
  - Likewise, $\ell(5P) = 5$, but we only have 5 elements in this space: $\{1, x, y, x^2, xy\}$. Again, these elements are all linearly independent since they have different valuations at P.
  - However, $\ell(6P) = 6$, and we can generate 7 elements in this space: $\{1, x, y, x^2, xy, x^3, y^2\}$. We must therefore have a linear dependence among these elements, and in fact since x^3 and y^2 are the only elements with valuation 6 at P, they must both occur with nonzero coefficients.
  - By rescaling x, y appropriately, we obtain an algebraic relation of the form $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ for some $a_1, a_2, a_3, a_4, a_6 \in E$: in other words, C has an equation in Weierstrass form. (Additionally, this also explains where the indices on the coefficients a_i come from: they are giving the "missing" pole valuation at P for the corresponding monomial term.)
  - To make this more precise, we observe that the map $\varphi : C \to \mathbb{P}^2$ defined by $[x : y : 1]$ is a rational map (by definition, because x, y are elements of the function field of C) whose image is a subvariety of the curve $V(Y^2Z + a_1XYZ + a_3YZ^2 - X^3 - a_2X^2Z - a_4XZ^2 - a_6Z^3)$. Thus φ is a morphism, and since it is nonconstant it is surjective.
  - But also, since the function field $F(x, y)$ equals the function field K of the curve, φ has degree 1, so φ is an in fact an isomorphism. Thus, C is isomorphic to a projective curve with a Weierstrass equation, as claimed.

- We therefore see that a curve with genus 1 over F having a point of degree 1 is an elliptic curve in Weierstrass form. We now adopt this as our more highbrow definition of an elliptic curve:

- Definition: Let F be a field. An elliptic curve E over F is a smooth projective curve defined over F with genus 1 that has an F-rational point O.
  - The specific choice of F-rational point O is part of the definition of an elliptic curve: if we take the same projective curve but choose different selections for O, we view the resulting elliptic curves as distinct. (As we will see, however, they will be isomorphic, so the distinction is not of great importance.)

- Now we can establish the existence of the group law on an elliptic curve.
  - In the discussion below, we will denote the divisor of a point P by $[P]$, since we will need to keep separate the notion of P as a divisor and as a point on a curve.

- Theorem (Group Law): Let F be a field and E be an elliptic curve defined over F with an F-rational point O. Then the following hold:

1. If P and Q are F-rational points such that $[P] \sim [Q]$ as divisors, then $P = Q$.
   - Proof: Suppose that $[P] \sim [Q]$, so that $[P] - [Q] = \text{div}(f)$ for some f.
   - Then in particular, $f \in L([Q])$: but Riemann-Roch on E says that $l([Q]) = 1$, so since the constants all lie in $L([Q])$, f must be constant. Then $\text{div}(f) = 0$ and hence $P = Q$, as claimed.
2. For every degree-zero divisor D, there exists a unique point $P \in E$ such that $D \sim [P] - [O]$.
   - Proof: To see that such a P exists, since $\deg(D + [O]) = 1$, our consequences of Riemann-Roch imply that $l(D + [O]) = 1$.
   - Let f span $L(D + [O])$: then $\text{div}(f) \geq -D - [O]$ and $\deg(\text{div}(f)) = 0$, so since $-D - [O]$ has degree -1 we must have $\text{div}(f) = -D - [O] + [P]$ for some degree-1 point P, whence $D \sim [P] - [O]$.
   - Uniqueness of Q then follows immediately from (1), since if $[P] - [O] \sim D \sim [Q] - [O]$ then $P = Q$.
3. If $\sigma : \text{Div}^0(E) \to E$ denotes the map in (2), then σ induces a bijection $\tilde{\sigma} : \text{Pic}^0(E) \to E$.
   - Proof: First observe that $\sigma([P] - [O]) = P$ so σ is certainly surjective from $\text{Div}^0(E)$ to E.
   - Also, by the definition of σ for any divisors D_1 and D_2 we have $\sigma(D_1) - \sigma(D_2) \sim D_1 - D_2$, so $D_1 \sim D_2$ if and only if $\sigma(D_1) = \sigma(D_2)$, which shows that σ descends to a bijection $\tilde{\sigma}$ from $\text{Pic}^0(E)$ to E.
4. With $\tilde{\sigma}$ as in (3), the group operation on E induced from $\text{Pic}^0(E)$ via $\tilde{\sigma}$ is the same as the geometric group law on E. (In other words, if we think of E as a group with the geometric law, then E is isomorphic to $\text{Pic}^0(E)$ via $\tilde{\sigma}$.)
   - Proof: The inverse map of $\tilde{\sigma}$ is $\tau : P \to [P] - [O]$. We want to see that $\tau(P+Q) = \tau(P) + \tau(Q)$, where the addition on the left is the geometric group law, and the addition on the right is the addition of divisor classes in the Picard group.
   - Equivalently, we want to see that $[P+Q] - [P] - [Q] + [O] \sim 0$, where again $P+Q$ represents addition via the geometric group law.
   - Let f be the line through P and Q, let R be the third intersection point of E with this line, and let g be the line through R and O. Then since the line $Z = 0$ intersects E at O with multiplicity 3, we have $\text{div}(f/Z) = [P] + [Q] + [R] - 3[O]$ and $\text{div}(g/Z) = [R] + [P + Q] - 2[O]$.
   - Therefore, $[P + Q] - [P] - [Q] + [O] = \text{div}(f/g) \sim 0$, as required. This means τ is a group homomorphism and thus a group isomorphism, as desired.
5. The group law defines morphisms $+ : E \times E \to E$ mapping $(P, Q) \mapsto P + Q$ and $- : E \to E$ mapping $P \mapsto -P$.
   - Proof: It is enough to show that the maps are rational, since rational maps from a smooth curve to a variety are morphisms. But the addition map and the additive-inverse maps are both rational as we have already seen via the explicit formulas: the only possible exceptions involve adding a point to itself or a point to O. One may check explicitly in these cases that the maps still yield morphisms.
6. For any divisor $D \in \text{Div}(E)$, D is principal if and only if $\deg(D) = 0$ and the formal sum representing D evaluates to O when viewed as a sum of points using the group law.
   - Proof: As we have previously noted, the degree of any principal divisor is 0, so certainly we must have $\deg(D) = 0$.
   - Now if $D \in \text{Div}^0(E)$ is $D = \sum_P n_P[P]$ we have $D \sim 0$ if and only if $\sigma(D) = O$. But $\sigma(D) = \sigma(\sum_P n_P[P]) = \sum_P n_P \sigma([P]) = \sum_P n_P(P - O) = \sum_P n_P P$ by definition of σ and the equivalence of the group operations in (4). So we see $\sigma(D) = O$ if and only if $\sum_P n_P P = O$ when viewed as a sum of points using the group law.

- Exercise: Show that we have an exact sequence $1 \to k^* \to k(E)^* \xrightarrow{\text{div}} \text{Div}^0(E) \xrightarrow{(6)} E \to 0$ where div represents the divisor map $f \mapsto \text{div}(f)$ and (6) represents the map discussed in (6) that takes a divisor $\sum_P n_P[P]$ and evaluates it as a sum of points on E.

0.11 (Oct 16) Differentials on Curves

- We would now like to establish the converse of our theorem above: namely, that every smooth projective curve with a Weierstrass equation $Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3$ is actually an elliptic curve.

- Since $[0:1:0]$ (the affine point at ∞) is always a rational point on this curve, we need only show it has genus 1.
- In order to do this, we need to discuss differentials, since they allow us to calculate the genus.

- Definition: Let C/k be a (smooth projective) curve. The space $\Omega(C)$ of meromorphic differential 1-forms on C is the k-vector space consisting of symbols of the form dx for $x \in k(C)$, subject to the following three relations:

  1. The additivity relation $d(x+y) = dx + dy$ for all $x, y \in k(C)$.
  2. The Leibniz rule $d(xy) = x\,dy + y\,dx$ for all $x, y \in k(C)$.
  3. Derivatives of constants are zero: $da = 0$ for all $a \in k$.

  - We remark also that there is a more general notion of differential form defined using the notion of a derivation from a commutative ring R to an R-module M. (We will not need this added formalism, but we include the definition purely for reference.)
    * A derivation is an additive map $D : R \to M$ such that $D(ab) = aD(b) + bD(a)$ for all $a, b \in R$).
    * If B is a k-algebra, we say D is a k-derivation if k is contained in the kernel of D (i.e., D vanishes on k).
    * Then the module of relative differential forms of B over k is a B-module $\Omega_{B/A}$ together with a k-derivation $d : B \to \Omega_{B/A}$ such that for any other B-module M and any k-derivation $d' : B \to M$ there exists a unique B-module homomorphism $f : \Omega_{B/A} \to M$ such that $d' = f \circ d$.
  - Although the space $\Omega(C)$ contains differentials of the form df for all $f \in k(C)$, and may therefore appear to be very large, in fact the relations impose all of the familiar rules of calculus.
  - Exercise: Show that the relations (1)-(3) also imply the power rule $d(x^n) = nx^{n-1}dx$ and the quotient rule $d(\frac{x}{y}) = \frac{x\,dy - y\,dx}{y^2}$.
  - Exercise: Suppose C/k is a curve and $x_1, x_2, \ldots, x_n \in k(C)$. For any rational function $f \in k(x_1, \ldots, x_n)$, show the "chain rule": that $df = f_{x_1}\,dx_1 + \cdots + f_{x_n}\,dx_n$, where f_{x_i} denotes the usual partial derivative. [Hint: First show the result for polynomials f, then use the quotient rule.]
  - As a corollary of the above exercises, we see immediately that if the function field $k(C)$ is generated (as a field extension) by $x_1, \ldots, x_n$ then $\Omega(C)$ is spanned by dx_1, dx_2, ... , dx_n as a $k(C)$-vector space.
  - Example: For $C = \mathbb{P}^1$, we have $k(C) = k(x)$ for $x = X/Y$. Since x generates the function field by itself we see that $\Omega(C)$ is spanned by dx. In fact, $\{dx\}$ is a basis, since there are no additional relations arising in the definition of $\Omega(C)$.
  - Example: Let p be a prime. For $C = \mathbb{P}^1$ over a field of characteristic not equal to p, by the above we know that $\{dx\}$ is a basis of $\Omega(C)$. Then for $f = x^p$, since $df = px^{p-1}dx$ is a nonzero scalar multiple of dx, we see that $\{df\}$ is also a basis of $\Omega(C)$. On the other hand, over a field of characteristic p, we have $df = px^{p-1}\,dx = 0$, and so $\{df\}$ is not a basis of $\Omega(C)$.
  - Example: For $C = V(Y^2Z - X^3 - XZ^2)$ with $x = X/Z$ and $y = Y/Z$, we have $k(C) = k(x, y)$. By the exercises above we see that $\Omega(C)$ is spanned by dx and dy, but since $y^2 = x^3 + x$ taking differentials yields a linear dependence $2y\,dy = (3x^2 + 1)\,dx$. Thus in fact either dx or dy suffices to span $\Omega(C)$. As with $\mathbb{P}^1$, there is no relation imposed on a single differential by itself, so $\{dx\}$ or $\{dy\}$ is a basis for $\Omega(C)$.
  - More generally, one may show similarly that $\Omega(C)$ is always a 1-dimensional $k(C)$-vector space for any curve C: in general, dx generates $\Omega(C)$ if and only if $k(C)/k(x)$ is a separable extension of finite degree. (The second example above shows that separability is necessary, since if k has characteristic p then $k(x)/k(x^p)$ is not separable, and as noted above in that situation dx^p does not span $\Omega(C)$.)

- Our goal now is to show that we may do calculations with differentials that mirror those for rational functions. First, we will give a well-defined notion of the order of a differential ω at a point P, and then we use it to attach a divisor to a differential.

- Proposition (Properties of Differentials): Let C/k be a curve, let ω be a differential in $\Omega(C)$, and let P be a point of C with a local uniformizer t. Then the following hold:

1. There exists a unique rational function $f \in k(C)$ such that $\omega = f\,dt$. (Since f is unique, we may think of it as the "quotient" ω/dt.)
   - Proof: First, since t is a local uniformizer, the extension $k(C)/k(t)$ has finite degree and is separable. Hence by the discussion above, we see that $\{dt\}$ spans $\Omega(C)$ as a $k(C)$-vector space.
   - This means so there exists a unique rational function $f \in k(C)$ such that $\omega = f\,dt$.
2. If $f \in k(C)$ is defined at P, then df/dt is also defined at P.
   - The most direct proof of this fact follows by working with local Laurent expansions near P. We will not need to do this explicitly, so here is an outline of the idea.
   - One may expand functions in $\mathcal{O}_P$ as infinite formal power series in the formal Laurent series ring of $k((t))$, and the resulting map $D : k(C) \to k((t))$ is a derivation.
   - Elements in the local ring $\mathcal{O}_P$ (i.e., functions f defined at P) have images lying in the formal power series ring $k[[t]]$, and for such elements, one may show that the term-by-term power series derivative f' yields the rational function with $df = f'\,dt$. Since the term-by-term derivative f' lies in $k[[t]]$, it is defined at P.
3. If t' is another local uniformizer at P, then $\operatorname{ord}_P(\omega/dt) = \operatorname{ord}_P(\omega/dt')$. We may therefore define $\operatorname{ord}_P(\omega)$ to be the value $\operatorname{ord}_P(\omega/dt)$ for *any* local uniformizer t.
   - Proof: Taking $f = t'$ in (2) shows that $dt'/dt = g$ is defined at P, and interchanging t and t' shows that $dt'/dt = 1/g$ is also defined at P. Therefore, we have $\operatorname{ord}_P(g) \geq 0$ and $\operatorname{ord}_P(1/g) \geq 0$ whence $\operatorname{ord}_P(g) = 0$.
   - Then we immediately have $\operatorname{ord}_P(\omega/dt) = \operatorname{ord}_P(\omega/dt' \cdot dt'/dt) = \operatorname{ord}_P(\omega/dt') + \operatorname{ord}_P(g) = \operatorname{ord}_P(\omega/dt')$.
4. Let $x \in k(C)^\times$ with $x(P) = 0$. Then $\operatorname{ord}_P(dx) = \operatorname{ord}_P(x) - 1$ except when the characteristic of k divides $\operatorname{ord}_P(x)$, in which case we have $\operatorname{ord}_P(f\,dx) \geq \operatorname{ord}_P(x)$.
   - Intuitively, the idea of this result is the extremely reasonable notion that taking the derivative of a function lowers its order of vanishing by 1, except in situations where the function is something times a pth power in characteristic p.
   - Proof: Since x is not zero we may write $x = ut^n$ for some u of order 0, and $n = \operatorname{ord}_P(x)$. Then $dx = unt^{n-1}\,dt + (du/dt)t^n\,dt$ by the chain rule.
   - From (2) we know that du/dt is defined at P so $\operatorname{ord}_P(du/dt) \geq 0$.
   - If the characteristic of k divides n, then $n = 0$ (in k), so $dx = (du/dt)t^n\,dt$. Then $\operatorname{ord}_P(dx) = \operatorname{ord}_P(dx/dt) = \operatorname{ord}_P(du/dt) + n \geq \operatorname{ord}_P(x)$ as desired.
   - Otherwise, if the characteristic does not divide n, then $n \neq 0$ in k so $\operatorname{ord}_P(unt^{n-1}) = n - 1$ while the order of the second term $(du/dt)t^n$ is at least n (as just calculated above). So since ord_P is a discrete valuation, the order of the sum $unt^{n-1} + (du/dt)t^n$ is $n - 1 = \operatorname{ord}_P(x) - 1$, as desired.
5. For all but finitely many P, we have $\operatorname{ord}_P(\omega) = 0$.
   - Proof: Pick x to be a local uniformizer at an arbitrary point of C: then by (1) we may write $\omega = f\,dx$.
   - Now, f has finitely many zeroes and poles, as noted in our discussion of divisors of functions. Additionally, as we will discuss in more detail later, there are only finitely many points at which $x - x(P)$ fails to be a local uniformizer at P. (These are the points at which x is ramified, when thought of as a map $x : C \to \mathbb{P}^1$.)
   - So there are only finitely many points P where f has a zero or pole, or where $x - x(P)$ fails to be a local uniformizer.
   - Let Q be any other point. Then $x - x(Q)$ is a local uniformizer, so we have $\operatorname{ord}_Q(dx) = \operatorname{ord}_Q(d(x - x(Q)) = 1 - 1 = 0$ by (4).
   - Hence $\operatorname{ord}_Q(\omega) = \operatorname{ord}_Q(f\,dx) = \operatorname{ord}_Q(f) + \operatorname{ord}_Q(dx) = 0 + 0 = 0$ because f is defined and does not vanish at Q. This applies for all but finitely many points Q, so we are done.
6. For any differential ω, its divisor $\operatorname{div}(\omega) = \sum_P \operatorname{ord}_P(\omega)\,P$ is well defined, and for any other differential ω_1 we have $\operatorname{div}(\omega) \sim \operatorname{div}(\omega_1)$. We define the canonical class $\mathcal{C}$ to be the resulting divisor class of $\operatorname{div}(\omega)$ in $\operatorname{Pic}(C)$.
   - Proof: Well-definedness follows immediately from (5), since only finitely many terms in the formal sum are nonzero.

- Now suppose ω_1 is any other differential. By (1) there exists $f \in k(C)$ such that $\omega/\omega_1 = f$: thus $\mathrm{div}(\omega) - \mathrm{div}(\omega_1) = \mathrm{div}(f)$ which means by definition that $\mathrm{div}(\omega) \sim \mathrm{div}(\omega_1)$.
- The well-definedness of the canonical class is then immediate from the equivalence.

7. A differential ω is holomorphic if $\mathrm{div}(\omega) \geq 0$: equivalently, when $\mathrm{ord}_P(\omega) \geq 0$ for all P, which is to say, when ω has no poles. The holomorphic differentials form a finite-dimensional vector space, whose dimension is defined to be g, the genus of C.
   - For completeness, we also say that a differential ω is nonvanishing if $\mathrm{div}(\omega) \leq 0$: equivalently, when $\mathrm{ord}_P \leq 0$ for all P, which is to say, when ω has no zeroes.
   - Proof: Writing $\omega = f\,dt$ we see that ω is holomorphic if and only if $\mathrm{div}(f) \geq -\mathrm{div}(\omega)$. Therefore, the map $\omega \mapsto \omega/dt$ is an isomorphism of the space of holomorphic differentials with the Riemann-Roch space $L(\mathrm{div}(\omega))$, whose dimension $l(\mathrm{div}(\omega)) = l(\mathcal{C})$ is finite, as follows from our properties of Riemann-Roch spaces.

- Of course, the real point of (6) and (7) is to give a proper definition of the canonical class and the genus of a curve that appear in the statement of the Riemann-Roch theorem.
  - We can also give some explanation of why the genus g, defined here as the dimension of the space of holomorphic differentials C, corresponds to the topological genus.
  - The idea is that when we are working over $k = \mathbb{C}$, then viewing C as a (compact, connected) Riemann surface, we may integrate a holomorphic differential along a path inside C.
  - By standard results from complex analysis, if two paths are homotopic then integrating any differential along the two paths yields the same value. Since the set of paths up to homotopy is the first homology group $H_1(C)$, which is a free abelian group of rank g (the topological genus of C), we obtain a pairing between $H_1(C)$ and the space of holomorphic differentials given by $\langle C, \omega \rangle = \int_C \omega$.
  - One then shows that this is a perfect pairing, and so these vector spaces are isomorphic. (Essentially, the idea is that we can obtain independent holomorphic differentials by integrating around independent non-contractible paths on C.)
  - We remark that all of this is just a rephrasing of Poincaré duality applied to the de Rham cohomology groups of C, considered as a 2-dimensional manifold.

- Example: On $C = \mathbb{P}^1$ with $x = X/Y$ as usual, find $\mathrm{div}(dx)$ and then show that there are no nonzero holomorphic differentials.
  - First, since $k(C) = k(x)$, so rather trivially $k(C)/k(x)$ is separable and of finite degree, we see that $\Omega(C)$ is spanned by dx. Thus every differential on C is of the form $\omega = f\,dx$ for some rational function $f \in k(x)$, in which case $\mathrm{div}(\omega) = \mathrm{div}(f) + \mathrm{div}(dx)$.
  - To find $\mathrm{div}(dx)$, first observe that for all $c \in k$ the function $x - c$ is a uniformizer at $[c:1]$, so $\mathrm{ord}_{[c:1]}(dx) = \mathrm{ord}_{[c:1]}(x-c) - 1 = 0$ by our results in (4).
  - Also, at the point at infinity $[1:0]$, the function $1/x$ is a uniformizer, so $\mathrm{ord}_{[1:0]}(x) = -1$ and thus $\mathrm{ord}_{[1:0]}(dx) = \mathrm{ord}_{[1:0]}(x) - 1 = -2$, again by (4).
  - Therefore, $\mathrm{div}(dx) = -2P_{[1:0]}$. We conclude that the canonical class is the image of $-2P_{[1:0]}$ in $\mathrm{Pic}(C)$.
  - In particular, the degree of any differential must be -2. But since the degree of a holomorphic differential is nonnegative, we see immediately that there are no nonzero holomorphic differentials.
  - Hence we see that the genus of $\mathbb{P}^1$ is 0 – as it should be, of course, given the results of our earlier calculations for genus-0 curves using Riemann-Roch.

- Example: On $C = V(Y^2Z - X^3 - XZ^2)$ with $x = X/Z$ and $y = Y/Z$ as usual, show that dx/y is a nonvanishing holomorphic differential, when the characteristic of k is not 2.
  - We have previously shown that $\mathrm{div}(y) = P_{[0:0:1]} + P_{[i:0:1]} + P_{[-i:0:1]} - 3P_{[0:1:0]}$ on this elliptic curve.
  - To find $\mathrm{div}(dx)$ we first compute $\mathrm{div}(x)$: since x is only zero at $[0:0:1]$ and since y is a local uniformizer there, to check the zero order we observe that $x/y^2 = XZ/Y^2 = Z^2/(X^2+Z^2) = 1$ is defined and nonzero, so $\mathrm{ord}_{[0:0:1]}x = 2$. Then since the only pole of x is at $[0:1:0]$ the pole also has order 2, and so $\mathrm{div}(x) = 2P_{[0:0:1]} - 2P_{[0:1:0]}$.

- In the same way we can show that $\operatorname{div}(x-i) = 2P_{[i:0:1]} - 2P_{[0:1:0]}$ and $\operatorname{div}(x+i) = 2P_{[-i:0:1]} - 2P_{[0:1:0]}$.
- Then since $x - x(P)$ is only zero at $x = 0, i, -i$, by property (4) we deduce that the zeroes of dx occur only at $[0:0:1]$, $[-i:0:1]$, and $[i:0:1]$ and the zero order there is $2-1=1$ in each case.
- Likewise, since the only pole of dx is at $[0:1:0]$, by (4) again we see the pole order is $-2-1=-3$. (Here is where we need the fact that the characteristic is not 2.)
- Putting all of this together shows that $\operatorname{div}(dx) = P_{[0:0:1]} + P_{[i:0:1]} + P_{[-i:0:1]} - 3P_{[0:1:0]}$. But this is precisely $\operatorname{div}(y)$, and so that means $\operatorname{div}(dx/y) = 0$ whence dx/y is holomorphic and also nonvanishing.

- Let us now generalize the example above to complete the proof that smooth projective curves of genus 1 having a rational point (per our highbrow definition of elliptic curves) are the same as nonsingular cubic curves in Weierstrass form (per our original definition):

- Proposition (Differentials on Elliptic Curves): Let C/k be a smooth projective curve having an affine Weierstrass equation of the form $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$. Then the following hold:

1. The differential $\omega = \dfrac{dx}{2y + a_1x + a_3} = -\dfrac{dy}{3x^2 + 2a_2x + a_4 - a_1y}$ is holomorphic and nonvanishing on C.
   - Proof: Let $f = y^2 + a_1xy + a_3y - (x^3 + a_2x^2 + a_4x + a_6)$: then by the chain rule we see that $\dfrac{dx}{f_y(x,y)} = -\dfrac{dy}{f_x(x,y)}$, showing that the two expressions are equal.
   - For any finite point $P = (x_0, y_0)$ we also have $\omega = \dfrac{d(x-x_0)}{f_y(x,y)} = -\dfrac{d(y-y_0)}{f_x(x,y)}$ since translating by a constant does not affect differentials.
   - In particular we see that P cannot be a pole of ω since this would require $f_x(P) = f_y(P) = 0$, but that cannot occur because C is smooth at P. So ω could only possibly have a pole at ∞.
   - For zeroes of ω we observe that the map $\varphi : C \to \mathbb{P}^1$ with $[X:Y:Z] \mapsto [X:Z]$ has degree 2, so $\operatorname{ord}_P(x - x_0) \leq 2$ with equality if and only if $f(x_0, y)$ has a double root in y at $y = y_0$, which occurs if and only if $f_y(x_0, y_0) = 0$.
   - Therefore by property (4) we see that $\operatorname{ord}_P(\omega) = \operatorname{ord}_P(dx) - \operatorname{ord}_P(f_y) = \operatorname{ord}_P(x - x_0) - \operatorname{ord}_P(f_y) - 1 = 0$ in both the situation when $\operatorname{ord}_P(x - x_0) = 1$ and in the situation when $\operatorname{ord}_P(x - x_0) = 2$.
   - It remains to check the order at ∞. For this let t be a uniformizer: then because $\operatorname{ord}_\infty(x) = -2$ and $\operatorname{ord}_\infty(y) = -3$ we have $x = t^{-2}u$ and $y = t^{-3}w$ for some $u, w \in k(C)$ that are defined and nonzero at ∞.
   - Then $\omega = \dfrac{dx}{f_y(x,y)} = \dfrac{-2t^{-3}u + t^{-2}(du/dt)}{2t^{-3}w + a_1t^{-2}u + a_3}\,dt = \dfrac{-2u + t(du/dt)}{2w + a_1tu + a_3t^3}\,dt$.
   - When the characteristic of k is not equal to 2, we can then evaluate the function $\dfrac{-2u + t(du/dt)}{2w + a_1tu + a_3t^3}$ at ∞ (note that $t = 0$ at ∞) to obtain $\dfrac{-u(\infty)}{w(\infty)}$ which is defined and nonzero.
   - Exercise: When the characteristic of k is not equal to 3, show that the equivalent formula $\omega = -\dfrac{dy}{f_x(x,y)}$ evaluates to a quantity that is defined and nonzero at ∞.
   - This means $\operatorname{ord}_\infty(\omega) = 0$ as well, so $\operatorname{div}(\omega) = 0$ whence ω is holomorphic and nonvanishing as claimed.
2. The space of holomorphic differentials on C is a 1-dimensional k-vector space, whence C has genus 1.
   - Proof: As shown in (1) above, there exists a holomorphic differential ω such that $\operatorname{div}(\omega) = 0$.
   - From our properties of differentials, any other differential ζ is of the form $f\omega$ for some $f \in k(C)$.
   - But then $\operatorname{div}(\zeta) = \operatorname{div}(f) + \operatorname{div}(\omega) = \operatorname{div}(f)$, so in order for ζ to be holomorphic we must have $\operatorname{div}(f) \geq 0$, meaning that f is a rational function with no poles.
   - But the only such (projective) functions are constants, whence ζ is a k-scalar multiple of ω.
   - This immediately implies that the space of holomorphic differentials on C is a 1-dimensional k-vector space, so C has genus 1 as claimed.
3. Every smooth projective genus-1 curve has a nonsingular Weierstrass equation, and conversely every nonsingular Weierstrass equation gives a smooth projective genus-1 curve.

- Proof: We showed the first part earlier using Riemann-Roch, while the second part is simply (2).

4. The differential ω from (1) is translation-invariant, meaning that for any point Q on E, if $(x,y)+Q=(\tilde{x},\tilde{y})$, then $\omega = \dfrac{d\tilde{x}}{2\tilde{y}+a_1\tilde{x}+a_3}$ as well. We therefore call ω the invariant differential of E.
   - We could in principle show this result just using the point addition formulas, since they give explicit expressions for $\tilde{x}$ and $\tilde{y}$ in terms of x, y, and the coordinates of Q. We will give a less tedious argument.
   - Proof: Let $\tilde{\omega} = \dfrac{d\tilde{x}}{2\tilde{y}+a_1\tilde{x}+a_3}$. Since $\tilde{\omega}$ is obtained by adding Q to all points on C, for any P on C we see that $\mathrm{ord}_P(\tilde{\omega}) = \mathrm{ord}_{P-Q}(\omega) = 0$, and so $\tilde{\omega}$ is also a nonvanishing holomorphic differential.
   - By (2) since the space of holomorphic differentials is 1-dimensional, that means $\tilde{\omega} = c_Q\omega$ for some scalar $c_Q \in k$ that (a priori) depends on Q.
   - Now consider the map $\varphi : E \to \mathbb{P}^1$ sending $Q \mapsto [c_Q : 1]$ for each point Q.
   - This map is necessarily rational (since after all the expressions for $\tilde{x}$ and $\tilde{y}$ are rational functions, so the ratio $\tilde{\omega}/\omega$ is some rational function), but it clearly omits $[1:0]$ since c_Q is defined for all Q.
   - Thus φ is not surjective, meaning that it must be constant since nonconstant rational maps of curves are surjective. Finally, setting Q to be the identity O on E shows $\tilde{\omega}_O = \omega$, so the constant must be 1.
   - We conclude that $\tilde{\omega} = \omega$ for all Q.

0.12 (Oct 19) Riemann-Roch (Redux), Ramification

- To finish this portion of the discussion, we outline the proof of the Riemann-Roch theorem.
  - The main additional definition required is the residue of a rational function $f \in k(C)$ at a point P, which is the general analogue of the residue of a meromorphic function at a point in $\mathbb{C}$. There are various ways to give this definition, but the standard approach is as the coefficient a_{-1} in a local Laurent expansion $f = \sum_{n=-k}^{\infty} a_n t^n$ where t is a local uniformizer. (Nontrivial work is required to make this rigorous, since we may be working in a field that lacks a notion of infinite series.)
  - The residue of a rational function is only nonzero when the function has a pole at P. By the analogue of Cauchy's residue theorem (or Stokes's theorem, depending on one's interpretation), one may also show that the sum of the residues of any rational function over all its poles is zero.
  - If we have an effective divisor $D = P_1 + P_2 + \cdots + P_d$ for distinct points P_i, we obtain a map $\varphi : L(D) \to k^d$ by taking $\varphi(D) = (\mathrm{Res}_{P_1} f, \mathrm{Res}_{P_2} f, \ldots, \mathrm{Res}_{P_d} f)$. The kernel of this map is the set of functions $g \in L(D)$ whose residue is zero at each P_i, which includes all constant functions.
  - Thus, we obtain an exact sequence $0 \to k \to L(D) \xrightarrow{\varphi} k^d$.

- Now we ask the question: how close is the map φ to being surjective? In other words, what conditions are there on the values of the residues of a function in $L(D)$ at the points P_i?
  - We can answer this question by looking at the residues of holomorphic and meromorphic differentials.
  - If ω is holomorphic, we define the residue of ω at P as the residue of the ratio ω/dt at P where dt is a local uniformizer at P.
  - Let D be a divisor on the curve C. Define $\Omega(D)$ to be the space of differentials ζ such that $\mathrm{div}(\zeta) \geq -D$.
  - Exercise: Show $\Omega(D)$ is a vector space, and that $\Omega(D)$ is isomorphic to $L(\mathcal{C} - D)$ where $\mathcal{C}$ is any element of the canonical class of C. [Hint: Fix a differential ω and let $f \in L(\mathcal{C} - D)$ and consider $f \mapsto f\omega$. The proof of (7) above is the special case $D = 0$.]
  - In the same way as for functions, the sum of the residues of any meromorphic differential over all points must be zero: thus, for each holomorphic ω and each $f \in L(D)$, we see that the sum of the residues of $f\omega$ must be zero. This means each differential imposes a linear condition on the possible choices of residues for f.
  - More precisely, we obtain a map $\psi : \Omega(D) \to k^d$ by taking $\psi(D) = (\mathrm{Res}_{P_1}\omega, \mathrm{Res}_{P_2}\omega, \ldots, \mathrm{Res}_{P_d}\omega)$. The kernel of this map is the set of differentials $\omega \in \Omega(D)$ whose residue is zero at each P_i, which includes all of the holomorphic differentials.

- Thus, we obtain another exact sequence $0 \to \Omega(0) \to \Omega(D) \xrightarrow{\psi} k^d$.
- Exercise: If $D \geq 0$, show that Riemann-Roch is equivalent to the statement that $\dim_k[L(D)/L(0)] + \dim_k[\Omega(D)/\Omega(0)] = \deg(D)$.
- The images of the two maps φ and ψ are orthogonal inside k^d by the observation made above: for any $f \in L(D)$ and any $\omega \in \Omega(D)$, the dot product of $\varphi(f)$ and $\psi(\omega)$ is $\sum_{i=1}^d \mathrm{Res}_{P_i}(f)\mathrm{Res}_{P_i}(\omega) = \sum_{i=1}^d \mathrm{Res}_{P_i}(f\omega) = 0$ since this is again the sum of the residues of a differential.
- So, since the images of φ and ψ are orthogonal, we see that $\dim(\mathrm{im}\varphi) + \dim(\mathrm{im}\psi) \leq d = \deg(D)$.
- By the nullity-rank theorem, since $\ker(\varphi) = k$ we get $\dim(\mathrm{im}\varphi) = \dim(L(D)) - 1 = \ell(D) - 1$.
- Likewise, since $\ker(\psi) = \Omega(D)$ we get $\dim(\mathrm{im}\psi) = \dim(\Omega(0)) - \dim(\Omega(D)) = g - \ell(C - D)$.
- Thus, we obtain the inequality $\ell(D) - 1 + g - \ell(C - D) \leq \deg(D)$, which is known as Riemann's inequality.

- The Riemann-Roch theorem is then the statement that we have an actual equality here: i.e., that the images of φ and ψ are actually orthogonal complements.
  - As it stands, we only know that $\ell(D) - 1 + g - \ell(C - D) \leq \deg(D)$ when D is an effective divisor.
  - In the event that $C - D$ is also effective, however, we can extract the desired result: in such a case, we have $\ell(D) - 1 + g - \ell(C - D) \leq \deg(D)$ and also $\ell(C - D) - 1 + g - \ell(D) \leq \deg(C - D) = \deg(C) - \deg(D)$, so adding the two inequalities yields $2g - 2 \leq \deg(C)$. But since $\deg(C) = 2g - 2$ (a calculation we take for granted), we must have equality in both cases.
  - This establishes Riemann-Roch for divisors D where both D and $C - D$ are effective divisors (or equivalent to effective divisors, since as we showed, $\ell(D_1) = \ell(D_2)$ when $D_1 \sim D_2$).
  - In fact, this is nearly enough to get the general result, since as we showed, if $L(D) \neq 0$ then D is equivalent to an effective divisor. In general, one needs to verify that when $\ell(C - D) = 0$, one has $\deg(D) \geq \ell(D) - 1 + g$.
  - Assuming the inequality $\deg(D) \geq \ell(D) - 1 + g$, one obtains the general statement of Riemann-Roch: if both D and $C - D$ are equivalent to effective divisors, the result is as above, and if D is but $C - D$ is not, the result follows from $\deg(D) \geq \ell(D) - 1 + g$, and if $C - D$ is but D is not, the result is equivalent by interchanging D and $C - D$.
  - Finally, if neither D nor $C - D$ is equivalent to an effective divisor (i.e., if $\ell(D) = \ell(C - D) = 0$), then by the inequality above we must have $\deg(D) \geq g - 1$ and $\deg(C - D) \geq g - 1$. But since $\deg(C) = 2g - 2$ this forces $\deg(D) = g - 1$, in which case we do get $\deg(D) = \ell(D) - 1 + g - \ell(C - D)$, as required.

- Our next object of study is how morphisms interact with divisors and differentials. We begin by discussing the notion of ramification.
  - Recall, as we have previously discussed, that if $\varphi : C_1 \to C_2$ is a nonconstant morphism of curves then we obtain a corresponding injection $\varphi^* : k(C_2) \to k(C_1)$ on function fields given by $\varphi^* f = f \circ \varphi$ for $f \in k(C_2)$, and conversely any injection of function fields $k(C_2) \hookrightarrow k(C_1)$ arises from a unique morphism.
  - More generally, this association yields an equivalence of categories, where E is an arbitrary field:
    1. (Objects) Function fields K/E of transcendence degree 1 where $K \cap \overline{E} = E$
       (Morphisms) Field injections fixing 1 (up to isomorphism)
    2. (Objects) Smooth projective curves defined over E
       (Morphisms) Non-constant morphisms defined over E (up to isomorphism)
  - Since both function fields have transcendence degree 1 over k and are finitely generated, the field extension $k(C_1)/\varphi^* k(C_2)$ has finite degree: we define the degree of this extension to be the degree $\deg(\varphi)$. For completeness also we define the degree of constant morphisms to be 0.
  - Additionally, we say φ is separable (or inseparable) when the corresponding field extension is separable (or inseparable) and define the associated separable degree (and inseparable degree) of φ to be the corresponding separable degree (and inseparable degree) of the field extension.

- Example: The degree of the morphism $\varphi : \mathbb{P}^1 \to \mathbb{P}^1$ given by $\varphi[X : Y] = [X^2 : Y^2]$ is 2, since with $x = X/Y$ as usual we have $k(C_1) = k(C_2) = k(x)$ and $\varphi(x) = x^2$, so $\varphi^* k(C_2) = k(x^2)$. Then the corresponding function-field extension is $k(x)/k(x^2)$ which has degree 2. Written affinely, the map is simply $\varphi(x) = x^2$, which we quite reasonably would expect to have degree 2 under any sensible definition. When the field characteristic is not equal to 2, this map is separable, and when the characteristic equals 2, it is (purely) inseparable.

- Definition: Let $\varphi : C_1 \to C_2$ be a nonconstant morphism. For each $P \in C_1$ we define the ramification index $e_\varphi(P)$ to be $\mathrm{ord}_P(\varphi^* t_{\varphi(P)})$, where $t_{\varphi(P)}$ is a local uniformizer at $\varphi(P)$.

  - Intuitively, the ramification index $e_\varphi(P)$ measures by what factor the local order of vanishing changes when we apply φ to move from P to $\varphi(P)$.
  - Note by definition that $(\varphi^* t_{\varphi(P)})(P) = (t_{\varphi(P)} \circ \varphi)(P) = t_{\varphi(P)}(\varphi(P)) = 0$, so the function $\varphi^* t_{\varphi(P)}$ is defined at P and evaluates to zero there. Thus, we have $e_\varphi(P) \geq 1$ with equality if and only if $\varphi^* t_{\varphi(P)}$ is a local uniformizer at P.
  - When $e_\varphi(P) = 1$ we say that P is unramified and otherwise (when $e_\varphi(P) > 1$) we say that P is ramified. We extend this to say a point $Q \in C_2$ is unramified when all its preimages $P \in \varphi^{-1}(Q)$ are unramified.
  - Example: Consider the morphism $\varphi : \mathbb{P}^1(\mathbb{C}) \to \mathbb{P}^1(\mathbb{C})$ given affinely by $\varphi(x) = x^2$. By definition, $\varphi^* f(x) = f(x^2)$. At $P = 2$ we have $\varphi(P) = 4$ and so $t_{\varphi(P)} = x - 4$ is a local uniformizer at $\varphi(P)$. Then $\varphi^* t_{\varphi(P)} = x^2 - 4$, so $\mathrm{ord}_P(\varphi^* t_{\varphi(P)}) = \mathrm{ord}_{x-2}(x^2 - 4) = 1$, so $P = 2$ is unramified. On the other hand, at $Q = 0$ we see that $t_{\varphi(Q)} = x$ so that $\varphi^* t_{\varphi(Q)} = x^2$ and $\mathrm{ord}_Q(\varphi^* t_{\varphi(Q)}) = \mathrm{ord}_x(x^2) = 2$, so $Q = 0$ is ramified. At $R = 0$ we have $t_{\varphi(R)} = 1/x$ so $\varphi^* t_{\varphi(R)} = 1/x^2$ and so $\mathrm{ord}_R(\varphi^* t_{\varphi(R)}) = \mathrm{ord}_{1/x}(1/x^2) = 2$ so $R = \infty$ is ramified. Indeed, one may check that 0 and ∞ are the only ramified points of this morphism.
  - Exercise: Compute the ramification index $e_\varphi(P)$ for all points $P \in \mathbb{P}^1$ for the map $\varphi : \mathbb{P}^1(\mathbb{C}) \to \mathbb{P}^1(\mathbb{C})$ with $\varphi(x) = x^3$.
  - Exercise: Let $f \in k(x)$ be a nonconstant rational function. Show that a finite point $P \in k$ is ramified for the map $f : \mathbb{P}^1(k) \to \mathbb{P}^1(k)$ if and only if $f'(P) = 0$. Deduce that f has only finitely many ramified points. Under what conditions on f will ∞ be ramified?
  - The ramification index defined above is the natural function-field analogue for the ramification index of a prime in a number field.
  - Explicitly, if L/K is an extension of number fields with corresponding rings of integers $\mathcal{O}_L$ and $\mathcal{O}_K$, then each prime ideal R of $\mathcal{O}_L$ lies over a unique prime ideal Q of $\mathcal{O}_K$ with $Q = \mathcal{O}_K \cap R$. If the prime ideal factorization of $Q\mathcal{O}_R$ has its power of R equal to $R^{e(R)}$, then the ramification index of R is $e(R)$. (This quantity is well defined since $\mathcal{O}_L$ is a Dedekind domain and therefore has unique factorization of ideals as a product of prime ideals.)
  - In fact, the ramification index in our situation quite literally *is* the ramification index for the prime ideal m_P associated to the valuation ring $\mathcal{O}_P$ in the field extension $k(C_1)/\varphi^* k(C_2)$.
  - The correspondence is a consequence of the general theorems describing the behavior of prime ideals in finite extensions of Dedekind domains, of which the ring of integers $\mathcal{O}_K$ in a number field and the local ring $\mathcal{O}_P$ for a point on a curve are examples.

- We have various other results following from the general theory, as well:

- Proposition (Properties of Ramification): Let $\varphi : C_1 \to C_2$ be a nonconstant morphism of (smooth projective) curves.

  1. For all $Q \in C_2$, we have $\sum_{P \in \varphi^{-1}(Q)} e_\varphi(P) = \deg \varphi$.

     - Example: For the squaring map $\varphi : \mathbb{P}^1(\mathbb{C}) \to \mathbb{P}^1(\mathbb{C})$ with $\varphi(x) = x^2$ of degree 2, for $Q = 4$ we have $\varphi^{-1}(Q) = \{P_2, P_{-2}\}$ and we may compute $e_\varphi(P_2) = e_\varphi(P_{-2}) = 1$. For $R = 0$ we have $\varphi^{-1}(R) = \{P_0\}$ and as we have already computed, $e_\varphi(P_0) = 2$.
     - This result is the analogue of the so-called "efg" theorem of number fields: if L/K is an extension of number fields and $f_\varphi(R|Q)$ is the relative degree of the prime R of $\mathcal{O}_L$ lying over the prime Q of $\mathcal{O}_K$, then $\sum_{R|Q} e_i(R) f_i(R) = [L : K]$.

- In our situation, the analogous definition of the relative degree would be the vector space dimension $\dim_{\mathcal{O}_P/m_P}(\mathcal{O}_{\varphi^*P}/m_{\varphi^*P})$, but since k is algebraically closes both fields $\mathcal{O}_P/m_P$ and $\mathcal{O}_{\varphi^*P}/m_{\varphi^*P}$ are isomorphic to k, so the relative degree is always 1.
- The proof (in both the number field case and our case) follows from examining the prime ideal factorization in the appropriate extension of Dedekind domains.

2. A point $Q \in C_2$ is unramified if and only if $\#\varphi^{-1}(Q) = \deg \varphi$.
   - Proof: By (1) we have $\sum_{P \in \varphi^{-1}(Q)} e_\varphi(P) = \deg \varphi$.
   - Since there are $\deg \varphi$ terms in the sum and each term is at least 1, the sum is always at least $\#\varphi^{-1}(Q)$, and it equals $\#\varphi^{-1}(Q)$ if and only if $e_\varphi(P) = 1$ for all $P \in \varphi^{-1}(Q)$.
   - So we see $e_\varphi(P) = 1$ for all $P \in \varphi^{-1}(Q)$ if and only if $\#\varphi^{-1}(Q) = \deg \varphi$, as claimed.
3. For all but finitely many $Q \in C_2$, $\#\varphi^{-1}(Q) = \deg_s \varphi$. As a consequence, when φ is separable, there are only finitely many ramified points Q.
   - The idea here is that typically a point $Q \in C_2$ has a total of $\deg_s \varphi$ preimages under φ, with the exceptions occuring when Q is ramified. Ramification corresponds to the situation where these preimages "collide" and yield fewer preimage points than expected (and the number of such collisions is measured by the ramification index).
   - Exercise: Suppose k is a(n algebraically closed) field of characteristic p and let the Frobenius morphism $\text{Frob} : \mathbb{P}^1(k) \to \mathbb{P}^1(k)$ be given by $\text{Frob}(x) = x^p$. Verify that $\#\text{Frob}^{-1}(Q) = 1$ for all $Q \in \mathbb{P}^1$, and show that Frob is ramified at every point. Deduce that the hypothesis that φ be separable in (3) above is necessary to ensure there are only finitely many ramified points.
   - This result is the analogue of the statement that there are only finitely many ramified primes in any extension L/K of number fields, which for number fields is typically proven by examining discriminants.
   - Proof (second part): If φ is separable, the result follows immediately from the first part and (2), since $\deg_s \varphi = \deg \varphi$: so for all but finitely many Q we see that Q is unramified.
4. The ramification index is multiplicative under composition: explicitly, if $\psi : C_2 \to C_3$ is another nonconstant morphism and $P \in C_1$, we have $e_{\psi \circ \varphi}(P) = e_\varphi(P) e_\psi(\varphi(P))$.
   - This result is the analogue of the fact that the ramification index is multiplicative in towers of number fields.
   - Proof (sketch): Applying φ changes the local order of vanishing by a factor of $e_\varphi(P)$, while applying ψ changes the local order of vanishing by a factor of $e_\psi(\varphi(P))$. Thus, the composition changes the local order of vanishing by the product of these two factors.

0.13 (Oct 23) Riemann-Hurwitz, Isogenies

- When we think of $k(C_1)$ as a finite extension of $\varphi^* k(C_2)$, we may use the norm in this extension to construct a map $\varphi_* : k(C_1) \to k(C_2)$.
  - Explicitly, we define $\varphi_* : k(C_1) \to k(C_2)$ via $\varphi_* = (\varphi^*)^{-1} \circ N_{k(C_1)/\varphi^* k(C_2)}$.
  - We will not bother being more explicit here, because our main interest is in the actions of the maps φ^* and φ_* on divisors and differentials, where we can give much nicer formulas.
- Definition: Let $\varphi : C_1 \to C_2$ be a nonconstant map of (smooth projective) curves. We define the inverse image map $\varphi^* : \text{Div}(C_2) \to \text{Div}(C_1)$ on divisor groups by setting $\varphi^*(Q) = \sum_{P \in \varphi^{-1}(Q)} e_\varphi(P) P$ for all $Q \in C_2$ and extending linearly, and we also define the direct image map $\varphi_* : \text{Div}(C_1) \to \text{Div}(C_2)$ by setting $\varphi_*(P) = \varphi(P)$ for all $P \in C_1$ and extending linearly.
  - Notational Remark: In principle we could just refer to φ_* as φ, but we want to keep separate the action of φ as a morphism with its action φ_* on divisors and on the associated function fields, and it will be convenient later to write it as φ_*.
  - Rather vacuously, both φ_* and φ^* are homomorphisms.
  - Example: Let $\varphi : \mathbb{P}^1 \to \mathbb{P}^1$ be the squaring map $\varphi(x) = x^2$. Then for $D = P_4 + 2P_0 - P_\infty$ we have $\varphi^*(D) = P_2 + P_{-2} + 4P_0 - 2P_\infty$ and $\varphi_*(P) = P_{16} + 2P_0 - P_\infty$.

- These actions also extend naturally to differentials:
- Definition: Let $\varphi : C_1 \to C_2$ be a nonconstant map of (smooth projective) curves. We define $\varphi^* : \Omega(C_2) \to \Omega(C_1)$ by setting $\varphi^*(f\,dx) = (\varphi^* f)\,d(\varphi^* x)$ for all $f, x \in k(C_2)$, and we define $\varphi_* : \Omega(C_1) \to \Omega(C_2)$ by setting $\varphi_*(g\,dy) = (\varphi_* g)\,d(\varphi_* y)$ for all $g, y \in k(C_1)$.
  - We will only need the action of φ^*, but the action of φ_* is recorded for completeness.
  - Example: Let $\varphi : \mathbb{P}^1 \to \mathbb{P}^1$ be the squaring map $\varphi(x) = x^2$. Then for $\omega_2 = (x+2)\,dx$ we have $\varphi^*(\omega_2) = (x^2+2)\,d(x^2) = (x^2+2)\,2x dx$.
- Proposition (Properties of φ_* and φ^*): Let $\varphi : C_1 \to C_2$ be a nonconstant map of (smooth projective) curves. Then the following hold:
  1. For any $D \in \mathrm{Div}(C_2)$, we have $\deg(\varphi^* D) = (\deg \varphi)(\deg D)$.
     - Proof: For a single point divisor Q we have $\varphi^* Q = \sum_{P \in \varphi^{-1}(Q)} e_\varphi(P) P$ so $\deg(\varphi^* Q) = \sum_{P \in \varphi^{-1}(Q)} e_\varphi(P) = \deg \varphi$ by property (1) of the ramification index. Now sum over all points in D and apply linearity.
  2. For any $D \in \mathrm{Div}(C_1)$, we have $\deg(\varphi_* D) = \deg D$.
     - Proof: Obvious, since if $D = \sum_{P \in C_1} n_P P$ then $\varphi_* D = \sum_{P \in C_1} n_P \varphi(P)$, whose degree is still $\sum_{P \in C_1} n_P$.
  3. For all $D \in \mathrm{Div}(C_2)$ we have $\varphi_*(\varphi^* D) = (\deg \varphi) D$.
     - Proof: For a single point divisor Q we have $\varphi_*(\varphi^* Q) = \varphi_* \sum_{P \in \varphi^{-1}(Q)} e_\varphi(P) P = \sum_{P \in \varphi^{-1}(Q)} e_\varphi(P) \varphi(P) = [\sum_{P \in \varphi^{-1}(Q)} e_\varphi(P)] Q = (\deg \varphi) Q$. Now sum over all points in D and apply linearity.
  4. If $\psi : C_2 \to C_3$ is another nonconstant map of smooth projective curves, then $(\psi \circ \varphi)^* = \varphi^* \circ \psi^*$ and $(\psi \circ \varphi)_* = \psi_* \circ \varphi_*$ as maps on the appropriate divisor groups.
     - Proof: For a single point divisor $R \in C_3$ we have $(\psi \circ \varphi)^* R = \sum_{P \in (\psi\circ\varphi)^{-1} R} e_{\psi\circ\varphi}(P) P = \sum_{P \in \varphi^{-1}(Q)} [\sum_{Q \in \psi^{-1}(R)} e_\varphi$ $\varphi^* \psi^* R$ using the ramification-in-towers property; now apply linearity.
     - Likewise, for a single point divisor $P \in C_1$ we have $(\psi \circ \varphi)_* P = \psi(\varphi(P)) = (\psi_* \circ \varphi_*)(P)$.
  5. For all nonzero $f \in k(C_2)$ we have $\varphi^*(\mathrm{div}\, f) = \mathrm{div}(\varphi^* f)$.
     - Exercise: For any nonzero $f \in k(C_2)$ and any $P \in C_1$, show that $\mathrm{ord}_P(\varphi^* f) = e_\varphi(P) \mathrm{ord}_{\varphi(P)}(f)$.
     - Proof: By the exercise we see that $\mathrm{div}(\varphi^* f) = \sum_{P \in C_1} \mathrm{ord}_P(\varphi^* f) P = \sum_{P \in C_1} \mathrm{ord}_{\varphi(P)}(f) \cdot [e_\varphi(P) P] = \sum_{Q \in C_2} \mathrm{ord}_Q(f) \cdot [\sum_{P \in \varphi^{-1}(Q)} e_\varphi(P)] P = \sum_{Q \in C_2} \mathrm{ord}_Q(f)\, \varphi^* Q = \varphi^* \sum_{Q \in C_2} \mathrm{ord}_Q(f) Q = \varphi^*(\mathrm{div}\, f)$, as claimed.
  6. For all nonzero $g \in k(C_1)$ we have $\varphi_*(\mathrm{div}\, g) = \mathrm{div}(\varphi_* g)$.
     - This property follows by general facts about the behavior of norms in finite extensions of Dedekind domains.
  7. The map φ is separable if and only if $\varphi^* : \Omega(C_2) \to \Omega(C_1)$ is injective (or equivalently, nonzero).
     - Proof: As noted in our initial discussion of differentials, an element $y \in k(C_2)$ has $\{dy\}$ a basis for $\Omega(C_2)$ if and only if $k(C_2)/k(y)$ is a finite-degree separable extension. Choose such an element y.
     - Applying φ^* shows that $\varphi^* k(C_2)/\varphi^* k(y)$ is also a finite-degree separable extension, and by definition of the action of $\varphi^* y = y \circ \varphi$ we see that $\varphi^* k(y) = k(\varphi^* y)$.
     - Then φ^* is injective $\iff$ $d(\varphi^* y) \neq 0$ $\iff$ $\{d(\varphi^* y)\}$ is a basis for $k(\Omega_1)$ $\iff$ $k(C_1)/k(\varphi^* y)$ is separable $\iff$ $k(C_1)/\varphi^* k(C_2)$ is separable, where the last equivalence follows from the fact that $\varphi^* k(C_2)/\varphi^* k(y)$ is separable and the composition of separable extensions is separable. And this last statement is simply the definition of separability for φ, so we are done.
     - Also, the equivalence of injectivity and nonzeroness for φ^* on differentials follows from the fact that $d(\varphi^* y)$ being nonzero is equivalent to it being a basis.
- We can now establish the fundamental relationship between the genera[6] of curves related by a morphism.

[6] "Genera" is the correct plural of "genus".

- Theorem (Riemann-Hurwitz): Let $\varphi : C_1 \to C_2$ be a nonconstant separable morphism where C_1 and C_2 are smooth projective curves of respective genera g_1 and g_2. Let $\omega \in \Omega(C_2)$ be any nonzero differential and define the ramification divisor $R = \mathrm{div}(\varphi^*\omega) - \varphi^*(\mathrm{div}\omega) \in \mathrm{Div}(C_1)$.

  1. The ramification divisor R is independent of the choice of ω.
     - Proof: Let $\{dx\}$ be any basis for $\Omega(C_2)$ and write $\omega = f\,dx$.
     - Then $\varphi^*\omega = (\varphi^*f)\,d(\varphi^*x)$ so $\mathrm{div}(\varphi^*\omega) = \mathrm{div}(\varphi^*f) + \mathrm{div}[d(\varphi^*x)]$, whereas $\varphi^*(\mathrm{div}\omega) = \varphi^*(\mathrm{div}f) + \varphi^*(\mathrm{div}\,dx)$.
     - Hence $R = \mathrm{div}(\varphi^*\omega) - \varphi^*(\mathrm{div}\omega) = [\mathrm{div}(\varphi^*f) - \varphi^*(\mathrm{div}f)] + \mathrm{div}[d(\varphi^*x)] - \varphi^*(\mathrm{div}\,dx) = \mathrm{div}[d(\varphi^*x)] - \varphi^*(\mathrm{div}\,dx)$ by property (5) above.
     - This last quantity is independent of ω, as desired.
  2. We have $\deg R \geq \sum_{P \in C_1}[e_\varphi(P) - 1]$ with equality if and only if the characteristic of k does not divide $e_\varphi(P)$ for any $P \in C_1$. (In particular, equality holds when the characteristic is zero.)
     - Proof: As shown in (1) we have $R = \mathrm{div}[d(\varphi^*x)] - \varphi^*(\mathrm{div}\,dx)$ for any basis $\{dx\}$ of $\Omega(C_2)$. To compute the order of R at P, we may take $x = t$ where t is a uniformizer at $Q = \varphi(P)$, since as we showed previously, $\{dt\}$ is a basis for $\Omega(C_2)$.
     - By definition, we have $\varphi^*t = us^e$ where s is a uniformizer at P, $e = e_\varphi(P)$ is the ramification index, and $u \in \mathcal{O}_P$ is defined at P with $u(P) \neq 0$.
     - Then $d(\varphi^*t) = [(du/ds)s^e + eus^{e-1}]ds$ so $\mathrm{ord}_P[d(\varphi^*t)] = \mathrm{ord}_P[(du/ds)s^e + eus^{e-1}] = (e-1) + \mathrm{ord}_P[s(du/ds) + eu]$, and we also have $\mathrm{ord}_P[\varphi^*(\mathrm{div}\,dt)] = 0$.
     - Since u is defined at P we see that du/ds is also defined at P, and quite similarly to our calculations with differentials previously, we see that $\mathrm{ord}_P[d(\varphi^*t)] \geq e - 1$ with equality if and only if the characteristic of k does not divide $e = e_\varphi(P)$.
     - Summing over all points $P \in C_1$ yields the result immediately.
  3. We have $2g_1 - 2 = (\deg\varphi)(2g_2 - 2) + \deg R$. As a consequence, $2g_1 - 2 \geq (\deg\varphi)(2g_2 - 2) + \sum_{P \in C_1}[e_\varphi(P) - 1]$ with equality if and only if $\mathrm{char}(k)$ does not divide $e_\varphi(P)$ for any $P \in C_1$.
     - Proof: Taking degrees in the definition of R and rearranging yields $\deg[\varphi^*(\mathrm{div}\omega)] = \deg[\mathrm{div}(\varphi^*\omega)] + \deg R$.
     - By property (1) of φ^*, we have $\deg(\varphi^*\omega) = (\deg\varphi)(\deg\omega) = (\deg\varphi)(2g_2 - 2)$ since ω is a differential on C_2 hence the degree of its divisor is $2g_2 - 2$ as we showed using Riemann-Roch.
     - Likewise, since $\varphi^*(\mathrm{div}\omega)$ is a differential on C_1, its degree is $2g_1 - 2$.
     - Then (2) yields immediately that $2g_1 - 2 \geq (\deg\varphi)(2g_2 - 2) + \sum_{P \in C_1}[e_\varphi(P) - 1]$ with equality if and only if $\mathrm{char}(k)$ does not divide $e_\varphi(P)$ for any $P \in C_1$, as claimed.

- The Riemann-Hurwitz theorem is really a topological result, and we can give some geometric motivation for where it comes from in the situation of Riemann surfaces, where $k = \mathbb{C}$.
  - If we view the curves C_1 and C_2 as surfaces over $\mathbb{R}$, then the morphism φ represents a d-sheeted covering of C_2 by C_1, where each unramified point of C_2 has exactly d preimages in C_1.
  - If φ were unramified everywhere, then (e.g., by considering a triangulation of C_1) we see that the Euler characteristic $\chi_1 = 2 - 2g_1$ of C_1 would be d times the Euler characteristic $\chi_2 = 2 - 2g_2$ of C_2: this is precisely the statement of Riemann-Hurwitz above.
  - At ramified points of φ, the sheets of the covering collide, which introduces an error term into the calculation.
  - Precisely, at a ramified point the ramification index $e_\varphi(P)$ counts the number of sheets that collide at P, and so relative to unramified points (with ramification index 1) the overall characteristic χ_1 is lowered by a total of $e_\varphi(P)$ from what would be expected if the point were unramified.
  - Summing this correction over all of the ramified points yields the general statement of Riemann-Hurwitz: $\chi_1 = (\deg\varphi)\chi_2 - \sum_{P \in C_1}[e_\varphi(P) - 1]$.

- We can now apply these general results to the situation of morphisms of elliptic curves. Since we defined an elliptic curve as a smooth projective curve of genus 1 together with a marked rational point O, we require the maps also to preserve the marked point:

- Definition: Let (E_1, O_1) and (E_2, O_2) be two elliptic curves. An isogeny $\varphi : E_1 \to E_2$ is a morphism from E_1 to E_2 such that $\varphi(O_1) = O_2$. If E_1 and E_2 are elliptic curves such that there exists a nonzero isogeny between them, we say they are isogenous.
  - Since the identity is uniquely specified for an elliptic curve, we will usually just write it as O.
  - As we will show later, being isogenous is an equivalence relation on elliptic curves. (It is self-evidently reflexive and transitive, since the identity morphism is an isogeny and the composition of two isogenies is an isogeny.)
  - Since nonconstant morphisms of curves are surjective, and the only constant isogeny is the zero map, nonzero isogenies are surjective.
  - When φ is nonzero, recall that we define the degree of φ to be degree of the function-field extension $k(C_2)/\varphi^* k(C_1)$. We also set $\deg(0) = 0$ for convenience.
  - Exercise: Show that the degree map is multiplicative on isogenies: $\deg(\varphi \circ \psi) = (\deg \varphi)(\deg \psi)$.
- Since E_1 and E_2 are groups, the collection of all isogenies from E_1 to E_2 forms an abelian group, and since compositions of isogenies are isogenies, the set of isogenies from E to E forms a ring.
  - Exercise: Let E_1 and E_2 be elliptic curves and define $\mathrm{Hom}(E_1, E_2)$ to be the collection of all isogenies from E_1 to E_2. Show that $\mathrm{Hom}(E_1, E_2)$ is an abelian group under the addition operation $(\varphi + \psi)P = \varphi(P) + \psi(P)$ for all $P \in E_1$ (where the addition on the right is the sum under the group law on E_2) for $\varphi, \psi \in \mathrm{Hom}(E_1, E_2)$.
  - Exercise: Let E be an elliptic curve and define $\mathrm{End}(E) = \mathrm{Hom}(E, E)$ to be the collection of all isogenies from E to itself. Show that E is a ring with 1 having no zero divisors, with addition given as in the exercise above and multiplication given by composition. [Hint: For the lack of zero divisors, consider degrees.]
- Our most basic example of an isogeny is the multiplication-by-m map:
  - Example: For an integer m, the multiplication-by-m map $[m] : E \to E$ is an isogeny, since as we have previously discussed it is a morphism, and it clearly preserves the group identity O.
  - We showed much earlier during our discussion of Mordell's theorem that the multiplication-by-m map has degree m^2, since as a rational map it is defined by a quotient of polynomials of degree m^2 (we will later give a far nicer and minimally computational proof of this fact).
  - In particular, $[m] \neq 0$ for $m \neq 0$, so the endomorphism ring $\mathrm{End}(E)$ always contains the subring $\mathbb{Z}$ generated by the identity map [1].
  - Additionally, if $\varphi : E_1 \to E_2$ is any isogeny, we see that $\deg(m\varphi) = \deg([m] \circ \varphi) = \deg([m]) \deg(\varphi) = m^2 \deg(\varphi)$. Thus, if φ is a torsion element of $\mathrm{Hom}(E_1, E_2)$ so that $m\varphi = 0$, then this calculation shows $\deg(\varphi) = 0$ whence $\varphi = 0$: this means $\mathrm{Hom}(E_1, E_2)$ is torsion-free.
- As we will see, for many elliptic curves the multiplication-by-m maps are the only endomorphisms! So it requires some nontrivial effort to give other examples.
- Example: Consider the map $i : E \to E$ with $i(x, y) = (-x, iy)$ on the elliptic curve $E : y^2 = x^3 - x$, where $i^2 = -1$ inside the underlying field k (where we assume $\mathrm{char}(k) \neq 2$ to avoid trivialities).
  - This map is a morphism from E to E since $(-x, iy)$ is also a point of E and it is described by rational functions that are defined everywhere, and since it maps $O = \infty$ to itself, it is an isogeny of E.
  - We can easily see that $[i] \circ [i]$ maps $(x, y) \mapsto (x, -y)$, so $[i] \circ [i] = [-1]$.
  - Taking $b[i]$ to be the b-fold sum of $[i]$ with itself, we see that the endomorphism ring $\mathrm{End}(E)$ contains the elements $[a] + b[i]$ for all $a, b \in \mathbb{Z}$. The ring of such elements is isomorphic to the Gaussian integer ring $\mathbb{Z}[i]$ via the obvious map $[a] + b[i] \mapsto a + bi$, and in fact, these are all of the endomorphisms of this elliptic curve.
  - This curve E is an example of an elliptic curve with complex multiplication, as it possesses an endomorphism that behaves like multiplication by a complex number (in this case, $i = \sqrt{-1}$).

- Example: Let $E = V(f)$ be an elliptic curve and let $E^{(p)} = V(f^{(p)})$, where $f^{(p)}$ is obtained by raising all of the coefficients of f to the pth power[7]. Then the Frobenius map $\text{Frob} : E \to E^{(p)}$ with $\text{Frob}(x, y) = (x^p, y^p)$ is an isogeny from E to $E^{(p)}$ since it is clearly a morphism and it preserves the point at ∞.
    - If E is defined over the field $\mathbb{F}_p$, the Frobenius map fixes all of the coefficients (indeed, $\mathbb{F}_p$ is precisely the field fixed by the Frobenius map): then $E^{(p)} = E$ and so Frob is an endomorphism of E.
    - More generally, if E is defined over $\mathbb{F}_q$ for some prime power q, then the qth-power Frobenius map $\text{Frob}(x, y) = (x^q, y^q)$ is an endomorphism of E.

0.14 (Oct 26) Properties of Isogenies

- We obtain various fundamental properties of isogenies that flow (more or less immediately) from the general properties of morphisms we have already shown:
- Theorem (Properties of Isogenies): Let $\varphi : E_1 \to E_2$ be a nonzero isogeny. Then the following hold:

1. The map φ is a group homomorphism from E_1 to E_2: in other words, $\varphi(P + Q) = \varphi(P) + \varphi(Q)$ for all $P, Q \in E_1$.
    - Since isogenies are the natural maps in the category of elliptic curves, and elliptic curves carry a natural group structure (which as we have discussed can be described purely in terms of the divisor group), the fact that isogenies are group homomorphisms is quite reasonable. Indeed, the reason we impose the additional condition that isogenies map the identity of E_1 to the identity of E_2 is precisely to ensure that isogenies are group homomorphisms.
    - Proof: Let P, Q be points of C_1 and O be the identity of C_1.
    - Then by our earlier results, $[P + Q] - [P] - [Q] + [O]$ is a principal divisor on E_1 as it has degree 0 and the underlying sum of points resolves to the identity on E_1.
    - For $\text{div}(f) = [P + Q] - [P] - [Q] + [O]$, we then have $\text{div}(\varphi^* f) = \varphi^* \text{div}(f) = [\varphi(P + Q)] - [\varphi(P)] - [\varphi(Q)] + [\varphi(O)]$, so this latter divisor is principal on E_2.
    - But that implies the resulting sum of points $\varphi(P+Q) - \varphi(P) - \varphi(Q) + \varphi(O)$ resolves to the identity on E_2, so since $\varphi(O)$ is the identity on E_2, we conclude immediately that $\varphi(P + Q) = \varphi(P) + \varphi(Q)$ as claimed.
    - Remark: Another way of making this argument is to observe that we have constructed group isomorphisms $\tau_1 : E_1 \to \text{Pic}^0(E_1)$ and $\tau_2 : E_2 \to \text{Pic}^0(E_2)$ with $\tau_i(P) = [P] - [O]$ as divisor classes. Then $\varphi_* \circ \tau_1 = \tau_2 \circ \varphi$ essentially by definition and the fact that $\varphi(O) = O$, so since φ_* is a homomorphism on the Picard groups, $\varphi = \tau_2^{-1} \circ \varphi_* \circ \tau_1$ is a composition of homomorphisms and thus also a homomorphism.
2. For all $Q \in E_2$, $\#\varphi^{-1}(Q) = \deg_s \varphi$. In particular, $\ker \varphi = \varphi^{-1}(O)$ is a finite subgroup of E_1.
    - Exercise: Suppose that $\varphi : G \to H$ is a surjective group homomorphism. Show that for any $h \in H$ there is a bijection between $\varphi^{-1}(h)$ and $\ker \varphi$.
    - Proof: By our results on ramification we know that $\#\varphi^{-1}(Q) = \deg_s \varphi$ for all but finitely many $Q \in E_2$. Since φ is a group homomorphism by (1) and surjective since it is a nonzero morphism, applying the exercise above yields both results immediately.
3. For all $P \in E_1$, the ramification index $e_\varphi(P) = \deg_i \varphi$, the inseparable degree of φ.
    - Proof: First, let $Q = \varphi(P)$ and take P' to be another point in $\varphi^{-1}(Q)$, and also define $R = P' - P$.
    - Since the translation morphism $\tau_R : E \to E$ defined by $\varphi(A) = A + R$ is an isomorphism and hence unramified, we have $\varphi(R) = O$ and so $\varphi \circ \tau_R = \varphi$.
    - Then $e_\varphi(P) = e_{\varphi \circ \tau_R}(P) = e_\varphi(\tau_R(P)) e_{\tau_R}(P) = e_\varphi(P')$ by the ramification composition formula. This means all points in $\varphi^{-1}(P)$ have the same ramification index.
    - Then $\deg_s \varphi \deg_i \varphi = \deg \varphi = \sum_{P \in \varphi^{-1}(Q)} e_\varphi(P) = \#\varphi^{-1}(Q) \cdot e_\varphi(P) = \deg_s \varphi \cdot e_\varphi(P)$, so we must have $e_\varphi(P) = \deg_i \varphi$ as claimed.

[7] Since the discriminant is a polynomial function of the coefficients of the Weierstrass equation, since the Frobenius map is a field automorphism, the discriminant of $f^{(p)}$ is the pth power of the discriminant of E, so $E^{(p)}$ is also nonsingular when E is nonsingular.

4. If φ is separable then φ is everywhere unramified and $\#\ker\varphi = \deg\varphi$.
   - Proof: By (3) we see immediately that if φ is separable, then $e_\varphi(P) = \deg_i\varphi = 1$ for all P, so φ is unramified. The cardinality of the kernel is immediate from (2).
   - Exercise: Use Riemann-Hurwitz to prove directly that if $\varphi : E_1 \to E_2$ is a nonconstant separable morphism of elliptic curves then φ is everywhere unramified.
5. The kernel $\ker\varphi$ is isomorphic to the automorphism group of the extension $k(E_1)/\varphi^*k(E_2)$ via the map Ξ sending $R \mapsto \tau_R^*$ where τ_R is the translation-by-R morphism.
   - Proof: First, in the same way as noted in the proof of (3), for any $R \in \ker\varphi$ we have $\varphi \circ \tau_R = \varphi$, since $\varphi(x+R) = \varphi(x) + \varphi(R) = \varphi(x)$ since φ is a homomorphism.
   - Then for any $f \in k(E_2)$ we have $\tau_R^*(\varphi^* f) = (\varphi\circ\tau_R)^* f = \varphi^* f$, and so τ_R^* fixes the field $k(E_2)$. Therefore τ_R^* is an automorphism of the extension $k(E_1)/\varphi^*k(E_2)$ so Ξ is well defined.
   - Next, for any $R, S \in \ker\varphi$ since rather obviously $\tau_{R+S} = \tau_S \circ \tau_R$, we have $\tau_{R+S}^* = (\tau_S\circ\tau_R)^* = \tau_R^*\tau_S^*$ so Ξ is a homomorphism.
   - Third, if τ_R^* fixes $k(E_1)$, then for any $f \in k(E_1)$ we have $f \circ \tau_R = f$. Taking f to be a function with poles only at O (which exist by Riemann-Roch since $l(2O) = 2$) we see that $f\circ\tau_R$ has poles only at $-R$, so $R = O$. Thus $\ker\Xi = \{O\}$ so Ξ is injective.
   - Finally, by basic facts about field automorphisms, the cardinality of $\mathrm{Aut}[k(E_1)/\varphi^*k(E_2)]$ is at most the separable degree of the extension $\deg_s(\varphi)$, so by (2) and the fact that Ξ is an injective homomorphism, we must have equality and Ξ is an isomorphism.
6. If φ is separable then the extension $k(E_1)/\varphi^*k(E_2)$ is a Galois extension of degree $\#\ker\varphi$.
   - Proof: By basic Galois theory, the cardinality of $\mathrm{Aut}[k(E_1)/\varphi^*k(E_2)]$ equals the degree of the extension if and only if the extension is Galois. By (4) and (5) combined, this occurs, and the degree equals $\deg\varphi = \#\ker\varphi$.
7. Suppose that $\varphi : E_1 \to E_2$ and $\psi : E_1 \to E_3$ are nonconstant isogenies and that φ is separable. If $\ker\varphi \subseteq \ker\psi$ then there exists a unique isogeny $\gamma : E_2 \to E_3$ such that $\psi = \gamma\circ\varphi$.
   - This result gives us a very convenient universal property of isogenies, and is essentially just an application of the fundamental theorem of Galois theory to the appropriate field extensions rephrased in terms of the corresponding curves.
   - Proof: Since φ is separable, by (6) we know that $k(E_1)/\varphi^*k(E_2)$ is Galois of degree $\#\ker\varphi$. Let the Galois group be G. Since $\ker\varphi \subseteq \ker\psi$, every element of G fixes $\psi^*k(E_3)$, so $\varphi^*k(E_2)$ is a field extension of $\psi^*k(E_3)$.
   - Since field extensions of function fields correspond to morphisms of curves (per the equivalence of categories discussed earlier), there exists a unique morphism $\gamma : E_2 \to E_3$ such that $\varphi^*(\gamma^*k(E_3)) = \psi^*k(E_3)$ which on the level of morphisms is equivalent to saying that $\gamma\circ\varphi = \psi$.
   - Finally, we have $\gamma(O) = \gamma(\varphi(O)) = \psi(O) = O$ since φ and ψ are isogenies, and so γ is an isogeny as well.
8. Suppose that Φ is a finite subgroup of the elliptic curve E. Then there exists a unique elliptic curve E' and a separable isogeny $\varphi : E \to E'$ such that $\ker\varphi = \Phi$.
   - Since φ is a surjective group homomorphism, the first isomorphism theorem immediately implies that the group structure of E' is that of the quotient group E/Φ, so since E' is unique here we often simply write $E' = E/\Phi$.
   - Of course, we can certainly construct the quotient group as a group by itself, but it is not immediately obvious why this quotient should also carry the structure of an algebraic variety (let alone why it should be another elliptic curve).
   - But in fact, one can show that the quotient of any smooth projective curve by a finite group of automorphisms also carries the structure of a variety.
   - Proof: As noted in (5), for each $R \in \Phi$ the translation-by-R map $\tau_R(x) = x + R$ yields an automorphism τ_R^* of $k(E)$; note it is an automorphism since it has an inverse map τ_{-R}^*.
   - By the fundamental theorem of Galois theory, if K is the fixed field of the automorphism group $\Phi^* = \{\tau_R^* : R \in \Phi\}$, then $k(E)/K$ is a Galois extension of degree $\#\Phi^* = \#\Phi$. In particular, K has transcendence degree 1 over k, so by our equivalence of categories, there exists a unique (up to isomorphism) smooth projective curve C/k and a unique finite-degree morphism $\varphi : E \to C$ such that $\varphi^*k(C) = K$.

- Now, since $k(E)/K$ is Galois hence separable, φ is separable. It remains to show that C is an elliptic curve (i.e., that it has genus 1 and that it has a rational point O).
- To do this we first show that φ is unramified, and then we apply Riemann-Hurwitz.
- For any $P \in E$ and $R \in \Phi$ and $f \in k(C)$, we have $f(\varphi(P+R)) = f(\varphi(\tau_R(P)) = (\varphi \circ \tau_R)^* f(P) = \tau_R^* \varphi^* f(P) = \varphi^* f(P) = f(\varphi(P))$ because τ_R^* fixes $\varphi^* f$.
- Since this equality holds for all functions f, by choosing f to be a function with poles only at one point (as in the argument in (5) above) we see that $\varphi(P+R) = \varphi(P)$.
- Therefore, for any $Q = \varphi(P)$ on C, the set $\varphi^{-1}(Q)$ contains the $\#\Phi$ translates $\{Q + R : R \in \Phi\}$. But by our properties of ramification, $\#\varphi^{-1}(Q) \leq \deg \varphi = \#\Phi$ with equality if and only if Q is unramified. Since this holds for all $Q \in C$, this means φ is everywhere unramified.
- Now, since φ is separable and unramified everywhere, by Riemann-Hurwitz we have $2g_E - 2 = (\deg \varphi)(2g_C - 2) + 0$ and so since $\deg \varphi$ is positive and $g_E = 1$, we must have $g_C = 1$ also.
- Finally, if we define $O_C = \varphi(O_E)$, then φ is an isogeny, and then as calculated above $\ker \varphi$ equals $\{O_E + R : R \in \Phi\} = \Phi$ since φ is unramified.

9. Suppose that $\mathrm{char}(k) = p$ and φ is not separable. For $q = \deg_i \varphi$ and $\mathrm{Frob}_q : E_1 \to E_1^{(q)}$ the qth-power Frobenius map, we may decompose $\varphi = \alpha \circ \mathrm{Frob}_q$ for a separable isogeny $\alpha : E_1^{(q)} \to E_2$.
   - This result extends the discussion above about the separable case and allows us to reduce the study of inseparable isogenies to the specific situation of the Frobenius map.
   - Proof: Let K be the separable closure of $\varphi^* k(E_2)$ in $k(E_1)$. Then by standard results about (in)separable extensions over perfect fields, $k(E_1)/K$ is purely inseparable of degree $q = \deg_i \varphi = p^d$ for some d with $K = k(E_1)^q$, while $K/\varphi^* k(E_2)$ is separable of degree $\deg_s \varphi$.
   - Then by definition, we see $K = \varphi^* k(E_1^{(q)})$, so since we have the tower of extensions $k(E_1)/\varphi^* k(E_1^{(q)})/\varphi^* k(E_2)$, converting to a statement about morphisms we see that $\varphi = \alpha \circ \mathrm{Frob}_q$ where $\alpha : E_1^{(q)} \to E_2$ is the morphism corresponding to the separable extension $K/\varphi^* k(E_2)$ and $\mathrm{Frob}_q : E_1 \to E_1^{(q)}$ is the qth-power Frobenius morphism corresponding to $k(E_1)/\varphi^* k(E_1^{(q)})$.
   - Remark: As also follows from the field degree calculations above, the degree of the qth-power Frobenius map Frob is q. This can also be calculated directly.
10. For any $Q \in E_1$, if $\tau_Q : E_1 \to E_1$ is the translation-by-Q map then $\tau_Q^* \omega = \omega$.
   - Proof: We showed this earlier in our discussion of differentials (it is why ω is called the invariant differential).
11. We have $[-1]^* \omega = -\omega$.
   - Proof: Suppose E_1 has general Weierstrass equation $y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6$.
   - Now because $[-1](x,y) = (x, -y - a_1 x - a_3)$, we see that $d[-1]^* x = dx$ and so $[-1]^* \omega = [-1]^* \dfrac{dx}{2y + a_1 x + a_3} = \dfrac{dx}{2(-y - a_1 x - a_3) + a_1 x + a_3} = -\omega$.
12. If ω is the invariant differential on E_1 and φ, ψ are two isogenies from $E_1 \to E_2$, then $(\varphi + \psi)^* \omega = \varphi^* \omega + \psi^* \omega$.
   - Proof: If φ or ψ is the zero isogeny the result is trivial.
   - If $\varphi + \psi = 0$ so that $\psi = [-1] \circ \varphi$ and thus $\psi^* = \varphi^* \circ [-1]^*$, we deduce $\varphi^* \omega + \psi^* \omega = \varphi^* \omega + \varphi^*(-\omega) = 0$ by linearity and (11).
   - Now assume that none of $\varphi, \psi, \varphi + \psi$ is zero. Take independent coordinates (x_1, y_1) and (x_2, y_2) and let $(x_3, y_3) = (x_1, y_1) + (x_2, y_2)$ via the group law, so that x_3 and y_3 are rational functions of x_1, y_1, x_2, y_2.
   - Now let $\omega_i(x_i, y_i)$ be the associated invariant differential $\dfrac{dx_i}{2y_i + a_1 x_i + a_3}$ for each $i = 1, 2, 3$.
   - Writing $x_3 = f(x_1, y_1, x_2, y_2)$,[8] the chain rule for differentials yields $dx_3 = f_{x_1} dx_1 + f_{y_1} dy_1 + f_{x_2} dx_2 + f_{y_2} dy_2$. But since dx_1 and dy_1 are $k(x_1, y_1)$-multiples of $\omega(x_1, y_1)$ and dx_2, dy_2 are $k(x_2, y_2)$-multiples of $\omega(x_2, y_2)$, we see that $\omega(x_3, y_3)$ is a $k(x_1, y_1, x_2, y_2)$-linear combination of the differentials $\omega(x_1, y_1)$ and $\omega(x_2, y_2)$.

[8] Explicitly, $f(x_1, y_1, x_2, y_2) = (\dfrac{3x_1^2 + 2a_2 x_1 + a_4 - a_1 y_1}{2y_1 + a_1 x_1 + a_3})^2 + a_1(\dfrac{3x_1^2 + 2a_2 x_1 + a_4 - a_1 y_1}{2y_1 + a_1 x_1 + a_3}) - a_2 - x_1 - x_2$.

- So now write $\omega_3(x_3, y_3) = g(x_1, y_1, x_2, y_2)\,\omega_1(x_1, y_1) + h(x_1, y_1, x_2, y_2)\,\omega_2(x_2, y_2)$ for some $g, h \in k(x_1, y_1, x_2, y_2)$.
- By working through the rather horrendous calculations explicitly, one may show that in fact the coefficients g and h are both just 1.
- To see this in a more pleasant manner, observe that if we choose any $P \in E_1$ and evaluate our expressions at $x_1 = x_1(P)$ and $y_1 = y_1(P)$, we have $dx_1 = dy_1 = 0$ so $\omega_1 = 0$, while $(x_3, y_3) = P + (x_2, y_2) = \tau_P(x_2, y_2)$.
- Thus $\omega_3 = \tau_P^* \omega_2 = \omega_2$ by translation-invariance from (10), so the linear combination expression reads as $\omega_3 = h(x_1(P), y_1(P), x_2, y_2)\omega_3$ whence $h(x_1(P), y_1(P), x_2, y_2)$ is identically 1 as a rational function in x_2 and y_2. But since this is true for every point $P \in C$, in fact h must be the constant function 1. By a symmetric argument we see $g = 1$ as well.
- So now we know that $\omega_3(x_3, y_3) = \omega_1(x_1, y_1) + \omega_2(x_2, y_2)$.
- Now apply this result to the case where $(x_1, y_1) = \varphi(x, y)$ and $(x_2, y_2) = \psi(x, y)$, so that $(x_3, y_3) = (\varphi + \psi)(x, y)$: it says that $(\omega \circ (\varphi + \psi))(x, y) = (\omega \circ \varphi)(x, y) + (\omega \circ \psi)(x, y)$ whence $(\varphi + \psi)^* \omega = \varphi^* \omega + \psi^* \omega$, as desired.

13. For any integer m we have $[m]^* \omega = m\omega$. In particular, $[m]$ is a finite, separable morphism whenever $\mathrm{char}(k)$ does not divide m.
    - Proof: For $m \geq 0$, induct on m. The base case $m = 0$ is trivial. For the inductive step observe that $[m+1]^* \omega = [m]^* \omega + [1]^* \omega = (m+1)\omega$ using (12) for additivity and the obvious $[1]^* \omega = \omega$.
    - For negative m note that $[m] = [-1] \circ [-m]$ and apply the result for positive m and (11).
    - The last statement follows immediately from the fact discussed earlier that a morphism is separable whenever it is nonzero on differentials.

0.15 (Oct 30) Dual Isogenies and Applications to the Hasse Bound

- Our goal now is to show that "being isogenous" is an equivalence relation on elliptic curves.
  - Since being isogenous is reflexive and transitive as we have already noted, it remains to show that every nonzero isogeny $\varphi : E_1 \to E_2$ induces some other nonzero isogeny $\hat{\varphi} : E_2 \to E_1$.
  - To see that this "dual isogeny" exists, we exploit the contravariant nature of the map $\varphi^* : \mathrm{Div}(E_2) \to \mathrm{Div}(E_1)$.
  - Specifically, because φ^* scales degrees by $\deg \varphi$, as we showed earlier, it maps $\mathrm{Div}^0(E_2)$ into $\mathrm{Div}^0(E_1)$, and therefore it descends onto a well-defined map $\varphi^* : \mathrm{Pic}^0(E_2) \to \mathrm{Pic}^0(E_1)$.
  - But as we also showed, the group operation in $\mathrm{Pic}^0(E)$ is isomorphic to the group law on E (namely, via the map sending a point $P \in E$ to the divisor class $[P] - [O]$), and so by composing these isomorphisms appropriately, we obtain a group homomorphism $\hat{\varphi} : E_2 \to E_1$.
- Of course, it is not at all obvious that this group homomorphism $\hat{\varphi}$ is actually an isogeny, since there are very many possible homomorphisms between the point groups, most of which will not be defined by rational functions.
  - Let us examine what precisely this map would look like when applied to a point $Q \in E_2$: first we map Q to the divisor class $[Q] - [O]$, then we apply φ^* to obtain $\deg_i \varphi \left(\sum_{P \in \varphi^{-1}(Q)} [P] - \sum_{R \in \varphi^{-1}(O)} [R] \right)$, and finally we must resolve this sum to write it in the form $[S] - [O]$.
  - By our results on the group law, we have $S = \deg_i \varphi \left(\sum_{P \in \varphi^{-1}(Q)} P - \sum_{R \in \varphi^{-1}(O)} R \right)$, evaluated as a sum on E_1. But because $\varphi^{-1}(Q) = \{P + R : R \in \varphi^{-1}(O)\}$ for any fixed $P \in \varphi^{-1}(Q)$, the difference between the two sums simply resolves to $\deg_i \varphi \cdot \#\varphi^{-1}(Q)$ times P.
  - But since $\#\varphi^{-1}(Q) = \deg_s \varphi$, the sum simplifies to $[\deg_i \varphi \deg_s \varphi] P = [\deg \varphi] P$.
  - So, to summarize, this map $\hat{\varphi} : E_2 \to E_1$ maps a point $Q \in E_2$ to $[\deg \varphi] P$ where P is any point in $\varphi^{-1}(Q)$.
  - Note that this description of $\hat{\varphi}$ is well posed: regardless of which representative $P \in \varphi^{-1}(Q)$ is chosen, since the difference between any of these representatives lies in $\varphi^{-1}(O) = \ker \varphi$.

- Equivalently, this says $\hat{\varphi}(\varphi(P)) = [\deg\varphi]P$ for all $P \in E_1$, meaning that the composition $\hat{\varphi}\circ\varphi$ is simply the multiplication-by-$[\deg\varphi]$ map on E_1.

- When φ is separable, we may use this observation along with the universal property of isogenies from earlier to show that this map $\hat{\varphi}$ actually is an isogeny, and we may then address the inseparable case by analyzing the Frobenius map:

- Theorem (Dual Isogenies): Let $\varphi : E_1 \to E_2$ be a nonconstant isogeny.

  1. If φ is separable, then there exists a unique isogeny $\hat{\varphi} : E_2 \to E_1$ such that $\hat{\varphi}\circ\varphi$ is multiplication by $\deg\varphi$ on E_1.
     - Proof: Let $\psi = [\deg\varphi]$ be the multiplication-by-$\deg\varphi$ map on E_1 and $E_3 = E_1$. Then since $\#\ker\varphi = \deg\varphi$, by Lagrange's theorem we see that $\ker\varphi \subseteq \ker\psi$.
     - Now by the universal property (7) of separable isogenies, there exists a unique isogeny $\hat{\varphi} : E_2 \to E_1$ such that $\hat{\varphi}\circ\varphi = \psi = [\deg\varphi]$, as claimed.
  2. If $\mathrm{char}(k) = p > 0$ and Frob_p is the pth-power Frobenius morphism $\mathrm{Frob}_p : E \to E^{(p)}$, then there exists a unique isogeny $\widehat{\mathrm{Frob}_p} : E^{(p)} \to E$ such that $\widehat{\mathrm{Frob}_p}\circ\mathrm{Frob}_p$ is multiplication by $p = \deg(\mathrm{Frob}_p)$ on E.
     - Proof: Let ω be the invariant differential on E. By property (13) of isogenies we see that $[p]^*\omega = p\omega = 0$, which means $[p]$ is not separable since it is not injective on differentials.
     - Hence by property (9) of isogenies, we may factor $[p]$ as $[p] = \alpha\circ\mathrm{Frob}_q$ where $q = \deg_i[p] = p^d$ for some integer $d \geq 1$ (note $d \geq 1$ because $[p]$ is not separable).
     - Then since[9] $\mathrm{Frob}_q = (\mathrm{Frob}_p)^d$ we see that $[p] = \alpha\circ(\mathrm{Frob}_p)^{d-1}\circ\mathrm{Frob}_p$: thus, the choice $\hat{\varphi} = \alpha\circ(\mathrm{Frob}_p)^{d-1}$ has the property that $\hat{\varphi}\circ\varphi = [p]$, as desired.
  3. There exists a unique isogeny $\hat{\varphi} : E_2 \to E_1$ such that $\hat{\varphi}\circ\varphi = [\deg\varphi]$ on E_1. This isogeny is called the dual isogeny of φ.
     - We emphasize here that this statement is equivalent to the one we worked out earlier as motivation for the construction for $\hat{\varphi}$: namely, for any $P \in C_1$, with $Q = \varphi(P)$ we have $\hat{\varphi}(Q) = [\deg\varphi]P$.
     - Proof: By property (9) of isogenies, we may decompose $\varphi = \alpha\circ\mathrm{Frob}_q = \alpha\circ(\mathrm{Frob}_p)^d$ where α is separable.
     - By (1) there exists an isogeny $\hat{\alpha}$ with $\hat{\alpha}\circ\alpha = [\deg\alpha]$ and by (2) there exists an isogeny $\widehat{\mathrm{Frob}_p}$ with $\widehat{\mathrm{Frob}_p}\circ\mathrm{Frob}_p = [\deg\mathrm{Frob}_p]$.
     - Then for $\hat{\varphi} = (\widehat{\mathrm{Frob}_p})^d\circ\hat{\alpha}$ we have $\hat{\varphi}\circ\varphi = (\widehat{\mathrm{Frob}_p})^d\circ\hat{\alpha}\circ\alpha\circ(\mathrm{Frob}_p)^d = (\widehat{\mathrm{Frob}_p})^d\circ[\deg\alpha]\circ(\mathrm{Frob}_p)^d = [\deg\alpha]\circ(\widehat{\mathrm{Frob}_p})^d\circ(\mathrm{Frob}_p)^d = [\deg\alpha][\deg\mathrm{Frob}_p]^d = [\deg\varphi]$, where the middle equality follows because multiplication by $\deg\alpha$ commutes with other isogenies since they are group homomorphisms, and the last equality follows from multiplicativity of degrees. Hence $\hat{\varphi}$ exists.
     - For uniqueness, suppose $\tilde{\varphi}\circ\varphi = [\deg\varphi] = \hat{\varphi}\circ\varphi$. Then $(\tilde{\varphi} - \hat{\varphi})\circ\varphi = 0$ so taking degrees yields $\deg(\tilde{\varphi} - \hat{\varphi})\deg\varphi = 0$ so since $\deg\varphi \neq 0$ that means $\deg(\tilde{\varphi} - \hat{\varphi}) = 0$ whence $\tilde{\varphi} = \hat{\varphi}$.
  4. We have $\varphi\circ\hat{\varphi} = [\deg\varphi]$ on E_2.
     - Although this result is very similar to the statement of (3), the multiplication-by-$\deg\varphi$ maps are taking place on different curves. Indeed, we have been slightly abusing terminology this entire time by referring to all of the maps $[m]$ on different curves E using the same name. Luckily, no issues will arise because multiplication by m "commutes" with all isogenies:
     - Exercise: Show that for any integer m and any isogeny $\varphi : E_1 \to E_2$, we have $[m]_{E_2}\circ\varphi = \varphi\circ[m]_{E_1}$.
     - Proof: Notice that $\hat{\varphi}\circ\varphi\circ\hat{\varphi} = [\deg\varphi]\circ\hat{\varphi} = \hat{\varphi}\circ[\deg\varphi]$ by (3) and the fact that the multiplication-by-m maps commute with all isogenies per the exercise above.
     - Thus, $\hat{\varphi}\circ(\varphi\circ\hat{\varphi} - [\deg\varphi]) = 0$, so by taking degrees as usual we see that since $\hat{\varphi} \neq 0$ we have $\varphi\circ\hat{\varphi} = [\deg\varphi]$.
  5. For any isogenies $\varphi : E_1 \to E_2$ and $\psi : E_2 \to E_3$ we have $\widehat{\psi\circ\varphi} = \hat{\varphi}\circ\hat{\psi}$.

[9] Technically, what is actually true is that Frob_q is the composition $F_{d-1}\circ\cdots\circ F_2\circ F_1$ where $F_i : E^{(p^{i-1})} \to E^{(p^i)}$ is the Frobenius map from $E^{(p^i)}$ to $E^{(p^{i+1})}$. By mild abuse of notation, we refer to all of these maps as simply Frob_p, since on the level of coordinates they are all just the pth power map $(x, y) \mapsto (x^p, y^p)$.

- Proof: Observe that $(\hat{\varphi} \circ \hat{\psi}) \circ (\psi \circ \varphi) = \hat{\varphi} \circ [\hat{\psi} \circ \psi] \circ \varphi = \hat{\varphi} \circ [\deg \psi] \circ \varphi = [\deg \psi][\deg \varphi] = [\deg \psi \circ \varphi]$.
- Since the dual isogeny is unique by (3), we must have $\widehat{\psi \circ \varphi} = \hat{\varphi} \circ \hat{\psi}$.

6. For any isogenies $\varphi, \psi : E_1 \to E_2$ we have $\widehat{\psi + \varphi} = \hat{\varphi} + \hat{\psi}$.
   - Proof: If φ, ψ, or $\varphi + \psi$ is zero, the result is trivial, so assume all of them are nonzero.
   - Let (x_1, y_1) and (x_2, y_2) be coordinates on E_1. Then because φ, ψ, and $\varphi + \psi$ are all morphisms, $\varphi(x_1, y_1)$, $\psi(x_1, y_1)$, and $(\varphi + \psi)(x_1, y_1)$ are all elements of the function field $E_2(k(x_1, y_1))$ of E_2 over the field $k(x_1, y_1)$.
   - Let D be the divisor $[(\varphi + \psi)(x_1, y_1)] - [\varphi(x_1, y_1)] - [\psi(x_1, y_1)] + [O]$ on E_2 over $k(x_1, y_1)$ – in other words, in $\mathrm{Div}_{k(x_1,y_1)}(E_2)$ – since it has degree 0 and the point sum resolves to the identity, it is the divisor of some function $f \in k(x_1, y_1)(E_2) = k(x_1, y_1, x_2, y_2)$.
   - Now switch coordinates and consider $\mathrm{div}(f)$ inside the divisor group $\mathrm{Div}_{k(x_2,y_2)}(E_2)$ – i.e., with x_2, y_2 constant and x_1, y_1 the variables. Let us compute the zeroes and poles (and their orders) of f.
   - If $P \in E_1(\overline{k(x_2, y_2)})$ is a point with $\varphi(P) = (x_2, y_2)$, then since D has the term $-[\varphi(x_1, y_1)]$ in it, D has a pole at P of order $e_\varphi(P)$ by the definition of the ramification index. In the same way, if Q has $\psi(Q) = (x_2, y_2)$ then because of the term $-[\psi(x_1, y_1)]$ we see that D has a pole at Q of order $e_\psi(Q)$, and if R has $(\varphi + \psi)(R) = (x_2, y_2)$ then the term $[(\varphi + \psi)(x_1, y_1)]$ contributes a zero of order $e_{\varphi+\psi}(R)$.
   - So that means the divisor of f inside $\mathrm{Div}_{k(x_2,y_2)}(E_2)$ has the form $(\varphi + \psi)^*[(x_2, y_2)] - \varphi^*[(x_2, y_2)] - \psi^*[(x_2, y_2)] + \sum n_i P_i$ for some "constants" $P_i \in E_1(k)$.
   - Since this is the divisor of a function, the sum of all the points resolves to the identity. Since $\sum n_i P_i$ is constant and does not depend on (x_2, y_2) this means the sum $(\widehat{\varphi + \psi})(x_2, y_2) - \hat{\varphi}(x_2, y_2) - \hat{\psi}(x_2, y_2)$ is a constant. Since it is the identity when $(x_2, y_2) = O$, it is always the identity.

7. For any nonzero integer m we have $\widehat{[m]} = [m]$ and $\deg[m] = m^2$.
   - We gave a computational argument much earlier to argue that $\deg[m] = m^2$, but now we can give a conceptually cleaner argument by exploiting dual isogenies.
   - Proof: We clearly have $\widehat{[1]} = [1]$. Then $\widehat{[m]} = [m]$ for positive m follows by a trivial induction from (6), and for negative m it follows by noting that $\widehat{[-1]} = [-1]$ and using (5).
   - For the degree of $[m]$ we note that by definition of the dual isogeny we have $[\deg[m]] = \widehat{[m]} \circ [m] = [m] \circ [m] = [m^2]$, and so $\deg[m] = m^2$.

8. We have $\deg \hat{\varphi} = \deg \varphi$ and $\hat{\hat{\varphi}} = \varphi$.
   - Proof: For the first, taking degrees in $[\deg \varphi] = \hat{\varphi} \circ \varphi$ and using (7) yields $(\deg \varphi)^2 = (\deg \hat{\varphi})(\deg \varphi)$. Cancelling yields the desired $\deg \hat{\varphi} = \deg \varphi$.
   - For the second, observe by definition that $\hat{\hat{\varphi}} \circ \hat{\varphi} = [\deg \hat{\varphi}] = [\deg \varphi] = \varphi \circ \hat{\varphi}$ on E_1. So since $\hat{\varphi}$ is nonzero, the usual degree argument shows that $\hat{\hat{\varphi}} = \varphi$.

9. For any nonzero integer m, if $\mathrm{char}(k)$ does not divide m then the m-torsion subgroup $E[m]$ is isomorphic to $(\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/m\mathbb{Z})$.
   - Proof: By (7), the degree of $[m]$ is m^2 and as we have previously noted using the action on differentials, $[m]$ is separable whenever $\mathrm{char}(k)$ does not divide m.
   - Therefore, by our properties of isogenies, we see that $\#E[m] = \#\ker[m] = \deg[m] = m^2$.
   - In particular, for each prime $p|m$, the group $E[p]$ is an elementary abelian p-group of order p^2, hence is isomorphic to $(\mathbb{Z}/p\mathbb{Z}) \times (\mathbb{Z}/p\mathbb{Z})$.
   - Then for each prime power $p^d|m$, the group $E[p^d]$ has at most two components in its decomposition each of which has order at most p^d, but since $E[p^d]$ has order p^{2d}, that means $E[p^d]$ is isomorphic to $(\mathbb{Z}/p^d\mathbb{Z}) \times (\mathbb{Z}/p^d\mathbb{Z})$.
   - The result for m then follows immediately from the Chinese remainder theorem, or from the structure theorem for finite abelian groups.
   - Exercise: Show that when $\mathrm{char}(k) = 0$, the group E_{tor} of all torsion points on E is isomorphic to $(\mathbb{Q}/\mathbb{Z}) \times (\mathbb{Q}/\mathbb{Z})$. [Hint: Note that E_{tor} is the direct limit of $E[n!]$ as $n \to \infty$.]

10. For any nonzero integer m and if $\mathrm{char}(k) = p$, then either $E[p^d] = \{O\}$ for all $d \geq 1$, or $E[p^d]$ is isomorphic to $\mathbb{Z}/p^d\mathbb{Z}$ for all $d \geq 1$.

- Proof: As in (9) we know that $\deg[p^d] = p^{2d}$, but now since $p | p^d$, the map $[p^d]$ is inseparable.
- If φ is the pth-power Frobenius map, then as we showed in (2), $\hat{\varphi} \circ \varphi = [p]$, so $(\hat{\varphi} \circ \varphi)^d = [p^d]$.
- By our properties of isogenies, we have $\#E[p^d] = \#\ker[p^d] = \deg_s[p^d] == \deg_s(\hat{\varphi} \circ \varphi)^d = \deg_s(\hat{\varphi})^d$ because $\deg_s \varphi = 1$ as φ is purely inseparable.
- Now, since $\deg \hat{\varphi} = \deg \varphi = p$ by (8), and $\deg_s \hat{\varphi} \deg_i \hat{\varphi} = p$, we either have $\deg_s \hat{\varphi} = 1$ or $\deg_s \hat{\varphi} = p$.
- In the first case we see that $\#E[p^d] = 1$ for all d, whence $E[p^d] = \{O\}$. In the second case we see that $\#E[p^d] = p^d$ for all d, so since $\#E[p] = p$ each of the groups must be cyclic hence $E[p^d]$ is isomorphic to $\mathbb{Z}/p^d\mathbb{Z}$ for all $d \geq 1$.

- We can apply these results very fruitfully to establish some structural statements about the group of points on an elliptic curve over a finite field. Let us first establish a few facts about quadratic forms.
  - Recall that if G is an abelian group, a function $d : G \to \mathbb{Z}$ is a quadratic form when $d(-g) = g$ for all $g \in G$ and when the pairing $\langle \cdot, \cdot \rangle : G \times G \to \mathbb{Z}$ with $\langle g, h \rangle = \frac{1}{2}[d(g+h) - d(g) - d(h)]$ is bilinear (i.e., is $\mathbb{Z}$-linear in both g and h).
  - We also say that a quadratic form is positive-definite when $d(g) \geq 0$ for all $g \in G$, with equality if and only if $g = 0$.
  - The usual Cauchy-Schwarz inequality holds for the bilinear pairing associated with a positive-definite quadratic form: $\langle g, h \rangle^2 \leq d(g)\, d(h)$.
  - To prove this note that if $d(g) = 0$ the result is trivial, and for $d(g) > 0$, for all integers a, b we have $a^2 d(g) - 2ab\langle g, h \rangle + b^2 d(h) = \langle ag - bh, ag - bh \rangle = d(ag - bh) \geq 0$ by bilinearity and positive-definiteness.
  - Setting $a = \langle g, h \rangle$ and $b = d(g)$ then yields $d(g)[d(g)^2 d(h) - \langle g, h \rangle^2] \geq 0$, and so upon dividing by $d(g)$ we obtain the desired inequality.

- Now we can apply these facts to the Frobenius map to count the number of points on an elliptic curve over a finite field.

- Theorem (Points on Elliptic Curves over $\mathbb{F}_q$): Let $q = p^d$ be a prime power and let $E/\mathbb{F}_q$ be an elliptic curve defined over $\mathbb{F}_q$.

  1. The degree map $\deg : \mathrm{Hom}(E_1, E_2) \to \mathbb{Z}$ is a positive-definite quadratic form.
     - Proof: First, $\deg(-\varphi) = \deg([-1])\deg(\varphi) = \deg(\varphi)$.
     - Second, the associated pairing $\langle \varphi, \psi \rangle = \deg(\varphi + \psi) - \deg(\varphi) - \deg(\psi)$ is bilinear, because we may write $[\langle \varphi, \psi \rangle] = [\deg(\varphi + \psi)] - [\deg(\varphi)] - [\deg(\psi)] = \widehat{\varphi + \psi} \circ (\varphi + \psi) - \hat{\varphi} \circ \varphi - \hat{\psi} \circ \psi = (\hat{\varphi} + \hat{\psi}) \circ (\varphi + \psi) - \hat{\varphi} \circ \varphi - \hat{\psi} \circ \psi = \hat{\psi} \circ \varphi + \hat{\varphi} + \psi$ using (6).
     - But now this last expression is linear in both φ and ψ by (6), so the pairing is bilinear.
     - Finally, the degree map is clearly positive-definite since $\deg(\varphi) \geq 0$ with equality if and only if $\varphi = 0$.
     - Exercise: On the elliptic curve $y^2 = x^3 - x$ with the isogeny $[i](x, y) = (-x, iy)$ discussed previously, for $\varphi = [a] + [b][i]$ with $a, b \in \mathbb{Z}$, calculate $\hat{\varphi}$. Use the result to find $\deg \varphi$ and compute the associated quadratic form.
  2. The Frobenius map $\varphi = \mathrm{Frob}_q$ has the property that $1 - \varphi$ is separable.
     - By light abuse of notation, we write $1 - \varphi$ instead of $[1] - \varphi$, since $[1] - \varphi$ is much uglier to read.
     - Proof: Let ω be the invariant differential on E. As we have previously shown,
     - By additivity of inverse image maps on differentials, we have $(1 - \varphi)^* \omega = [1]^* \omega - \varphi^* \omega = \omega$ since $\varphi^* \omega = 0$ because φ is inseparable hence is trivial on differentials).
     - But now since $1 - \varphi$ is nontrivial on differentials, it is separable.
     - Exercise: Show more generally that $a + b\varphi$ is separable if and only if the characteristic p does not divide a.
  3. (Hasse Bound) The number of points on $E(\mathbb{F}_q)$ satisfies $|\#E(\mathbb{F}_q) - q - 1| \leq 2\sqrt{q}$.
     - Proof: By basic Galois theory of finite fields, an element $x \in \overline{\mathbb{F}_q}$ lies in $\mathbb{F}_q$ if and only if $x^q = x$, which is to say, if and only if it is fixed by the qth-power Frobenius map φ.

- So now if we choose a Weierstrass equation for E over $\mathbb{F}_q$, since $E^{(q)} = E$ since the coefficients lie in $\mathbb{F}_q$ by hypothesis, we see a point $[X:Y:Z] \in E(\mathbb{F}_q)$ if and only if $\varphi(X:Y:Z) = [X:Y:Z]$, which is equivalent to saying that $[X:Y:Z] \in \ker(1-\varphi)$.
- Hence $\#E(\mathbb{F}_q) = \#\ker(1-\varphi)$. By (2), the map $1-\varphi$ is separable, so $\#\ker(1-\varphi) = \deg(1-\varphi)$ by our results on kernels and degrees.
- By (1), since the degree map is a positive-definite quadratic form, we may apply the Cauchy-Schwarz inequality to see that $\langle 1, -\varphi\rangle^2 \leq \deg[1]\deg(-\varphi) = q$ whence $|\langle 1, -\varphi\rangle| \leq \sqrt{q}$.
- Since $\langle 1, -\varphi\rangle = \frac{1}{2}[\deg(1-\varphi) - \deg(-\varphi) - \deg(1)] = \frac{1}{2}[\deg(1-\varphi) - q - 1]$, applying the results above yields $|\#E(\mathbb{F}_q) - q - 1| \leq 2\sqrt{q}$, as claimed.

0.16 (Nov 2) The Zeta Function, The Weil Conjectures, and The Tate Module

- Exercise: Verify the Hasse bound for $E : y^2 = x^3 + 4x + 1$ over $\mathbb{F}_3$, $\mathbb{F}_5$, $\mathbb{F}_7$, $\mathbb{F}_{11}$, and $\mathbb{F}_{13}$ (optionally, also over $\mathbb{F}_9$, $\mathbb{F}_{25}$, and $\mathbb{F}_{27}$).
- We can give some intuition for why we might expect an inequality like the Hasse bound to hold.
  - Assuming characteristic not equal to 2 for simplicity, consider a Weierstrass equation $y^2 = p(x)$ for E. For each of the q possible finite values of x, there are either 2, 1, or 0 possible values of y, according to whether x is a nonzero square, zero, or a nonsquare. Since the squaring map $x \mapsto x^2$ is a homomorphism with kernel $\{\pm 1\}$ in $\mathbb{F}_q$, there are $(q-1)/2$ nonzero squares and $(q-1)/2$ nonsquares, so the expected number of values of y for any given x is equal to 1.
  - Since there are q possible x, the expected number of finite points (x, y) is q, so together with the point at ∞, this gives an expected $q+1$ points on $E(\mathbb{F}_q)$.
  - We trivially have the inequality $|\#E(\mathbb{F}_q) - q - 1| \leq q$ since the number of points is at least 1 and at most $2q+1$. Hasse's bound is therefore a strengthening of the error term from this "trivial estimate" q to the estimate $2\sqrt{q}$. In fact we can give some statistical motivation for why this estimate on the deviation is somewhat reasonable:
  - Exercise: Suppose X is the sum of q independent random variables each of which takes the values 0 and 2 each with probability 1/2. Show that the standard deviation of X is $\sqrt{q}$.
  - If we approximate the point-count on E as the sum of q independent coin flips each of which yields 0 or 2 points, then by the exercise above, the standard deviation in the total number of points would be $\sqrt{q}$. The Hasse bound thus says our count will always be within 2 standard deviations of the mean. (Of course, this is only a heuristic, since the actual variables themselves are not independent, but it's useful for seeing why the results come out near $\sqrt{q}$.)
- Perhaps surprisingly, the error estimate in the Hasse bound is actually tied to much deeper results related to the Riemann hypothesis for algebraic varieties, via the Weil conjectures. To explain how this works we first define the zeta function of a variety:
- Definition: Let q be a prime power and V be a smooth projective variety defined over the field $\mathbb{F}_q$. For each $n \geq 1$, define $a_n = \#V(\mathbb{F}_{q^n})$ to be the number of points of V that lie in the extension field $\mathbb{F}_{q^n}$. Then the zeta function of V is defined to be the power series $\zeta_V(T) = \exp(\sum_{n=1}^{\infty} a_n \frac{T^n}{n})$.
  - Example: For $V = \mathbb{P}^1$ we have $a_n = q^n + 1$ for each n, and so $\zeta_{\mathbb{P}^1}(T) = \exp(\sum_{n=1}^{\infty} \frac{(qT)^n}{n} + \sum_{n=1}^{\infty} \frac{T^n}{n}) = \exp(-\ln(1-qT) - \ln(1-T)) = \frac{1}{(1-qT)(1-T)}$ using the usual series expansion $-\ln(1-T) = \sum_{n=1}^{\infty} \frac{T^n}{n}$.
  - Exercise: Find $\zeta_V(T)$ for $V = \mathbb{P}^n$ and for $\mathbb{P}^1 \times \mathbb{P}^1$.
- It is not especially clear from this definition why exactly we call this the zeta function of V. We can give some clearer motivation in the situation where $V = C$ is a curve:

- Proposition (Sum Formula for Zeta Function): Suppose C is a smooth projective curve defined over $\mathbb{F}_q$ and let b_n be the number of effective divisors $D \geq 0$ of degree n in the divisor group $\mathrm{Div}_{\mathbb{F}_q}(C)$. Then the zeta function $\zeta_C(T)$ equals $\sum_{n=0}^{\infty} b_n T^n$.
  - Recall that for a point $P \in C(\overline{\mathbb{F}_q})$, the degree of P is defined to be the degree of the field extension $\mathbb{F}_q(P)/\mathbb{F}_q$, and the divisor associated to P is the sum $\mathrm{div}(P) = \sum_{\sigma \in \mathrm{Gal}(\mathbb{F}_q(P)/\mathbb{F}_q)} \sigma(P)$, which has degree $\deg(P)$.
  - Proof: Note that any effective divisor $D \geq 0$ in $\mathrm{Div}_{\mathbb{F}_q}(C)$ is of the form $\sum_{P \in C(\overline{\mathbb{F}_q})} n_P \mathrm{div}(P)$ for nonnegative integers n_P, and the degree of this divisor is $\sum_{P \in C(\overline{\mathbb{F}_q})} n_P \deg(P)$.
  - So by the usual properties of generating functions, we have $\sum_{n=0}^{\infty} b_n T^n = \prod_{P \in C(\overline{\mathbb{F}_q})} (1 + T^{\deg P} + T^{2 \deg P} + \cdots) = \prod_{P \in C(\overline{\mathbb{F}_q})} (1 - T^{\deg P})^{-1}$ as a formal power series. (If one formally multiplies out the middle product, each divisor $\sum_{P \in C(\overline{\mathbb{F}_q})} n_P \mathrm{div}(P)$ of total degree n yields one term T^n.)
  - Then $\ln[\sum_{n=0}^{\infty} b_n T^n] = -\sum_{P \in C(\overline{\mathbb{F}_q})} \ln(1 - T^{\deg P}) = \sum_{P \in C(\overline{\mathbb{F}_q})} \sum_{k=1}^{\infty} \dfrac{T^{k \deg P}}{k}$, whose coefficient of T^n is the sum $\sum_{P \in C(\overline{\mathbb{F}_q}) \,:\, k \deg(P) = n} \dfrac{1}{k} = \sum_{P \in C(\overline{\mathbb{F}_q}) \,:\, \deg(P) | n} \dfrac{\deg(P)}{n}$, which when we "glue" all of the $\deg(P)$ Galois-conjugate points $\sigma(P)$ together, evaluates simply to $\dfrac{1}{n} \#\{P \in C(\overline{\mathbb{F}_q}) \,:\, \deg(P) | n\} = \dfrac{1}{n} \#C(\mathbb{F}_{q^n}) = \dfrac{a_n}{n}$, where the first equality follows from the fact that an element of $\overline{\mathbb{F}_q}$ lies in $\mathbb{F}_{q^n}$ if and only if its degree divides n.
  - So we conclude that $\ln(\sum_{n=0}^{\infty} b_n T^n) = \sum_{n=0}^{\infty} a_n \dfrac{T^n}{n}$, which upon exponentiating yields the desired formula.
- Now we can explain the analogy for why the zeta function is called a zeta function.
  - From the sum formula, we have $\zeta_V(T) = \sum_{n=0}^{\infty} b_n T^n = \sum_{D \geq 0} T^{\deg(D)} = \sum_{D \geq 0} \dfrac{1}{N(D)^s}$ where $N(D) = q^{-\deg(D)}$ and $T = q^{-s}$.
  - This latter expression is the analogue of the Riemann zeta function's definition $\zeta(s) = \sum_{n \geq 1} \dfrac{1}{n^s}$.
  - The idea is that effective divisors on C are the natural analogue of the positive integers, and that the norm function $N(D)$ gives the proper "size" of a divisor. The points in $C(\overline{\mathbb{F}_q})$ are the analogues of the primes showing up in the Euler product $\zeta(s) = \prod_p (1 - p^{-s})^{-1}$, analogous to the Euler product $\zeta_C(T) = \prod_P (1 - T^{\deg P})^{-1}$ worked out in the proposition.
- Now that we have worked out some facts about zeta functions, we can state the Weil conjectures:
- Theorem (Weil Conjectures): Let V be a smooth projective variety of dimension n defined over $\mathbb{F}_q$ with associated zeta function $\zeta_C(T)$. Then the following properties hold:
  1. (Rationality) The zeta function $\zeta_C(T)$ is a rational function of T. More specifically, $\zeta_C(T) = \prod_{i=0}^{2n} p_i(T)^{(-1)^{i+1}} = \dfrac{p_1(T) p_3(T) \cdots p_{2n-1}(T)}{p_0(T) p_2(T) \cdots p_{2n}(T)}$ for appropriate polynomials $p_i(T) \in 1 + T\mathbb{Z}[T]$, where $p_0(T) = 1 - T$, $p_{2n}(T) = 1 - q^n T$, and $p_i(T) = \prod_j (1 - \alpha_{i,j} T)$ for some $\alpha_{i,j} \in \mathbb{C}$.
  2. (Functional Equation / Poincaré Duality) The zeta function has a functional equation $\zeta_C(q^{-n} T^{-1}) = \pm q^{nE/2} T^E \zeta_C(T)$, where $E = 2 - 2g$ is the Euler characteristic of V. In particular, the map $\alpha \mapsto q^n/\alpha$ maps the zeroes of p_i to the zeroes of p_{2n-i}.
  3. (Riemann Hypothesis) For each i, j, the inverse zeroes $\alpha_{i,j}$ of p_i have $|\alpha_{i,j}| = q^{i/2}$. Equivalently, with $T = q^{-s}$, all of the zeroes of $p_k(T)$ lie on the line $\mathrm{Re}(s) = k/2$.
  4. (Betti Numbers) If V is the reduction modulo $\tilde{p} = \mathrm{char}(\mathbb{F}_q)$ of a smooth variety X defined over an algebraic number field, then the degree of p_i is the ith Betti number of the space $X(\mathbb{C})$ of complex points on X.
- The Weil conjectures have a long history. Here is a brief summary of some of it:

- In the early 1800s, Gauss identified some components of these general results in particular examples for certain curves, in the context of counting points on elliptic curves modulo p.
- In 1924, Artin conjectured the general results for curves and Hasse independently proved the results for elliptic curves.
- In 1949, Weil formulated the general statement of the Weil conjectures (he had previously established Artin's conjectured statements in the case of curves).
- Establishing the Weil conjectures in full took the development of about 20 more years' worth of algebraic geometry machinery: Dwork proved (1) in 1960, while Grothendieck proved (1), (2), and (4) in the 1960s, and Deligne finished (3) in 1973.

- In the specific case $n = 1$ (i.e., for curves), the Weil conjectures read as follows:
  1. $\zeta_C(T)$ is a rational function of the form $\zeta_C(T) = \dfrac{L_C(T)}{(1-T)(1-qT)}$ for some polynomial $L_C(T) = \prod_j (1 - \alpha_j T)$.
  2. For $\xi_C(T) = T^{1-g}\zeta_C(T)$, we have $\xi_C(q^{-1}T^{-1}) = \xi_C(T)$.
  3. The roots of L_C all have $|\alpha_j| = q^{-1/2}$.
  4. The degree of L_C is $2g$.
- Exercise: Verify the Weil conjectures for $C = \mathbb{P}^1$.
- Exercise: Show that for elliptic curves, the Weil conjectures are equivalent to the statement that $\zeta_C(T) = \dfrac{(1-\alpha T)(1-\beta T)}{(1-T)(1-qT)}$ where α and β are complex conjugates of absolute value $\sqrt{q}$.
  - Let us unwind precisely what this statement says about the coefficients $a_n = \#E(\mathbb{F}_{q^n})$.
  - Suppose for the moment that we know the Weil conjectures are true, so that $\zeta_C(T) = \dfrac{(1-\alpha T)(1-\beta T)}{(1-T)(1-qT)}$.
  - Then $\ln \zeta_C(T) = -\ln(1-T) - \ln(1-qT) + \ln(1-\alpha T) + \ln(1-\beta T) = \sum_{n=1}^{\infty} \dfrac{1^n + q^n - \alpha^n - \beta^n}{n} T^n$, and so we have $\#E(\mathbb{F}_{q^n}) = 1 + q^n - \alpha^n - \beta^n$ for some complex conjugates α and β of absolute value $\sqrt{q}$.
  - Notice that when $n = 1$, this says $\#E(\mathbb{F}_q) = 1 + q - \alpha - \beta$ where α and β are complex conjugates of absolute value $\sqrt{q}$, meaning that $|\#E(\mathbb{F}_q) - q - 1| = 2|\mathrm{Re}(\alpha)| \leq 2\sqrt{q}$: precisely the statement of the Hasse bound!
- So, how could we try to prove the Weil conjectures? As with the proof of the Hasse bound, we need to convert things to a statement about the qth-power Frobenius map φ.
  - First, we observe that $P \in \overline{\mathbb{F}_q}$ lies in $\mathbb{F}_{q^n}$ if and only if P is fixed by φ^n if and only if $P \in \ker(1 - \varphi^n)$. Thus, $\#E(\mathbb{F}_{q^n}) = \#\ker(1-\varphi^n) = \deg(1-\varphi^n)$ since $1 - \varphi^n$ is separable by the same argument used in (2) of the Hasse bound proof.
  - From properties of duals, we have $[\deg(1-\varphi^n)] = \widehat{(1-\varphi^n)} \circ (1-\varphi^n) = (1-\hat{\varphi}^n) \circ (1-\varphi^n) = [1] - \varphi^n - \hat{\varphi}^n + \hat{\varphi}^n \circ \varphi^n = [1] - \varphi^n - \hat{\varphi}^n + [q^n]$.
  - This is fairly close to the result we want: we would just need to show that $\varphi^n + \hat{\varphi}^n = [\alpha^n + \beta^n]$ where α and β are complex conjugates of absolute value $\sqrt{q}$.
  - Let's now pretend φ is a different kind of object entirely: namely, a linear transformation on a complex vector space, with $\hat{\varphi}$ being its dual transformation (or adjoint, depending on terminology) with respect to an inner product.
  - Then the sum $\varphi^n + \hat{\varphi}^n$ would represent the trace $\mathrm{tr}(\varphi^n)$ of the linear transformation φ^n, which by basic linear algebra equals the sum of the nth powers of the eigenvalues of φ. So we would obtain a statement of the desired form if φ had exactly 2 eigenvalues (i.e., if φ were an operator on a 2-dimensional vector space) that were complex conjugates of absolute value $\sqrt{q}$.
  - Of course, none of these statements are really true: although φ is a linear transformation, it acts on the field $\overline{\mathbb{F}_q}$ of characteristic p. In order to make statements about eigenvalues that are complex numbers, we would need to have an action of φ on something in characteristic 0.

- Let us now discuss how to construct an object in characteristic 0 on which φ has a natural 2-dimensional representation. In fact, with no added difficulty, we can do this for any Galois automorphism (or any endomorphism) of any elliptic curve E.

- So let E be an elliptic curve defined over the field F with algebraic closure k as usual, and let $\sigma \in G = \mathrm{Gal}(k/F)$ be any automorphism in the Galois group. Since E is defined over F, σ maps points of E to other points of E, and indeed σ is a group homomorphism from E to E since the addition law of points in F is defined over F as well.
  - For any $P \in E[m]$, we have $[m]\sigma(P) = \sigma([m]P) = \sigma(O) = O$, and so G acts on $E[m]$.
  - Since for any integer m not divisible by $p = \mathrm{char}(\mathbb{F}_q)$, the m-torsion subgroup $E[m]$ is isomorphic to $(\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/m\mathbb{Z})$, this means G has a group action on $(\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/m\mathbb{Z})$, which is to say, we have a representation $G \to \mathrm{Aut}[(\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/m\mathbb{Z})] \cong GL_2(\mathbb{Z}/m\mathbb{Z})$.
  - This is a 2-dimensional representation of the Galois group, which is at least in the right direction for what we want, but we really need a representation in characteristic 0, not characteristic m. To deal with this, we can exploit the fact that we have representations for all integers m, not just individual ones.
  - Since by the Chinese remainder theorem, the action of the representation is completely determined by the action on the prime-power torsion groups, it's enough to instead study the behavior on the l-power torsion subgroups $E[l^d]$ for $l \neq p$, which are isomorphic to $(\mathbb{Z}/l^d\mathbb{Z}) \times (\mathbb{Z}/l^d\mathbb{Z})$.
  - We may glue the l-power torsion groups $E[l^d]$ together in a natural way using inverse limits using the fairly simple observation that if P is an l^d-torsion point, then $[l]P$ is an l^{d-1}-torsion point.

- To warm up with a simpler example, let us construct the ring $\mathbb{Z}_l$ of l-adic integers using the inverse system $\mathbb{Z}/l\mathbb{Z} \xleftarrow{\pi} \mathbb{Z}/l^2\mathbb{Z} \xleftarrow{\pi} \mathbb{Z}/l^3\mathbb{Z} \xleftarrow{\pi} \mathbb{Z}/l^4\mathbb{Z} \xleftarrow{\pi} \cdots$ of rings with projection maps $\pi : \mathbb{Z}/l^{d+1}\mathbb{Z} \to \mathbb{Z}/l^d\mathbb{Z}$ given by the natural projection (i.e., reduction modulo l^d).
  - The elements of the inverse system are tuples $(b_1, b_2, b_3, b_4, \dots)$ such that $\pi(b_{d+1}) = b_d$ for each d, which is to say, $b_{d+1} \equiv b_d \pmod{l^d}$. If we take the unique representative for each b_i with $0 \leq b_i < l^i$, then we have $b_{d+1} = b_d + a_d l^d$ for some unique integer $a_{d+1} \in \{0, 1, 2, \dots, l-1\}$.
  - Iterating, we see that we have $b_{d+1} = a_0 + a_1 l + \cdots + a_d l^d$ for some sequence of "base-l digits" $a_i \in \{0, 1, \dots, l-1\}$.
  - Therefore, we can equivalently describe the elements of the inverse limit $\varprojlim_d (\mathbb{Z}/l^d\mathbb{Z})$ as infinite base-l expansions $a_0 + a_1 l + a_2 l^2 + a_3 l^3 + \cdots$ for appropriate digits $a_i \in \{0, 1, \dots, l-1\}$.[10]
  - Then $\mathbb{Z}_l$ is a ring via componentwise addition and multiplication since all of the projections are ring homomorphisms, and the resulting ring operations are simply that of base-l arithmetic on the resulting digits. In particular, note that $\mathbb{Z}_l$ has characteristic zero.
  - Indeed, $\mathbb{Z}_l$ also inherits a metric space topology (the l-adic topology) from the natural l-adic valuation $v_l(\sum a_i l^i)$ given by the minimal power i with $a_i \neq 0$. (Intuitively, two points are close together under this topology when their expansions agree for many terms.)
  - We may use a very similar inverse limit construction on the torsion groups $E[l^d]$.
  - Explicitly, consider the inverse system $E[l] \xleftarrow{[l]} E[l^2] \xleftarrow{[l]} E[l^3] \xleftarrow{[l]} E[l^4] \xleftarrow{[l]} \cdots$ of groups whose elements are tuples $(P_1, P_2, P_3, P_4, \dots)$ with $P_d \in E[l^d]$ and where $[l]P_{d+1} = P_d$.
  - One may think of these tuples as being obtained by starting with the identity O and then successively choosing inverse images P_1, P_2, P_3, P_4, ... under the multiplication-by-l map.
  - Since all of the maps are group homomorphisms, the set of such tuples is a group under componentwise addition: it is the inverse limit $\varprojlim_d E[l^d]$.
  - Indeed, since each $E[l^d]$ is a $(\mathbb{Z}/l^d\mathbb{Z})$-module, the inverse limit actually carries a $\mathbb{Z}_l$-module structure, and hence also inherits the l-adic topology.

- Definition: Let E be an elliptic curve and l be a prime. The l-adic Tate module of E is the $\mathbb{Z}_l$-module $T_l(E) = \varprojlim_d E[l^d]$.

[10] This description shows another standard way to construct $\mathbb{Z}_l$: namely, as the completion of $\mathbb{Z}$ under the l-adic metric.

- When $l \neq \text{char}(k)$, when we apply the inverse limit construction starting with generators P and Q of $E[l]$, we obtain topological generators for $T_l(E)$ yielding a group isomorphism $T_l(E) \cong \mathbb{Z}_l \times \mathbb{Z}_l$.
- When $l = \text{char}(k)$ we instead have $T_l(E) \cong \mathbb{Z}_l$ or 0, according to whether $E[l^d] \cong \mathbb{Z}/l^d\mathbb{Z}$ or 0, respectively.
- Now, returning to our discussion, if E is defined over F and $\sigma \in \text{Gal}(k/F)$, then σ acts naturally on the Tate module via $\sigma(P_1, P_2, P_3, \dots) = (\sigma P_1, \sigma P_2, \sigma P_3, \dots)$, and since this action is clearly a group action, it yields a representation of $\text{Gal}(k/F)$ on $\text{Aut}[T_l(E)]$.
- In fact, since the Galois group acts continuously on each component $E[l^d]$ of the inverse limit (rather trivially, since it is a profinite group and they are all discrete groups), the Galois action is also continuous.

- Definition: Let E be an elliptic curve defined over the field F with algebraic closure k, and let $l \neq \text{char}(k)$ be a prime. The l-adic Galois representation associated to E is the map $\rho_l : \text{Gal}(k/F) \to \text{Aut}[T_l(E)]$ defined by $\rho_l(\sigma)(P_1, P_2, P_3, \dots) = (\sigma P_1, \sigma P_2, \sigma P_3, \dots)$.
  - Since $l \neq \text{char}(k)$ we know that $T_l(E)$ is isomorphic to $\mathbb{Z}_l \times \mathbb{Z}_l$, so $\text{Aut}[T_l(E)]$ is isomorphic to $\text{Aut}(\mathbb{Z}_l \times \mathbb{Z}_l) \cong GL_2(\mathbb{Z}_l)$.
  - Now, $\mathbb{Z}_l$ is not a field, but it is an integral domain, so it embeds in its field of fractions $\mathbb{Q}_l = \mathbb{Z}_l[l^{-1}]$, and so by embedding $GL_2(\mathbb{Z}_l)$ inside $GL_2(\mathbb{Q}_l)$, we obtain a 2-dimensional representation of $\text{Gal}(k/L)$ over a field of characteristic zero. (At last, progress!)

0.17 (Nov 6) The Weil Pairing and The Weil Conjectures (again)

- The remaining ingredient for our plan in proving the Weil conjectures is to find an analogue of an inner product structure associated to the action of the Galois group on $\text{Aut}[T_l(E)]$.
  - As with our construction of the Tate module, we will do this by constructing a pairing on the components $E[l^d]$ used in the inverse limit construction of $T_l(E)$. Indeed, for no additional cost, we can construct the pairing on $E[m]$.
  - Here is a simple way to try to do this: by choosing a basis $\{P, Q\}$ of $E[m]$, we have an isomorphism $E[m] \cong (\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/m\mathbb{Z})$, so elements are of the form $aP + bQ$ for $a, b \in \mathbb{Z}/m\mathbb{Z}$.
  - Then a natural pairing with many convenient properties is $\langle aP + bQ, cP + dQ \rangle = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \pmod{m}$. (For instance, the pairing is bilinear, alternating, and nondegenerate, all of which are properties we would want for something analogous to an inner product.)
  - Of course this pairing does not take values in a field unless m is prime, but we can easily deal with this shortcoming by instead taking the pairing to be $\langle aP + bQ, cP + dQ \rangle = \zeta^{ad-bc}$ where $\zeta \in k$ is some primitive mth root of unity.
  - However, this construction relies on several choices (the basis $\{P, Q\}$ and the mth root of unity ζ). In order to take an inverse limit, we want to give a more natural pairing that doesn't depend on particular choices of basis and generator for the group of mth roots of unity.

- So, let us take a different approach for constructing a pairing on the set of m-torsion points $E[m]$ whose values are mth roots of unity.
  - Fix a positive integer $m \geq 2$ not divisible by $p = \text{char}(k)$.
  - For any $Q \in E[m]$, since the divisor $m[Q] - m[O]$ has degree 0 and the sum of points resolves to the identity on E, it is principal: say $m[Q] - m[O] = \text{div}(f_Q)$, for a function $f_Q \in k(C)$ unique up to scaling.
  - We claim that the divisor $[m]^*Q - [m]^*O$ is also principal. To see this choose any $Q' \in [m]^{-1}Q$: then by definition we have $[m]^*Q - [m]^*O = \sum_{R \in E[m]}([Q' + R] - [R])$ which is also principal since it has degree 0 and the underlying sum of points is $\sum_{R \in E[m]} Q' = [m^2]Q' = [m]Q = O$.
  - This means $[m]^*Q - [m]^*O = \text{div}(g_Q)$ for some function g_Q that is unique up to scaling.
  - Now, we have $\text{div}(g_Q^m) = \sum_{R \in E[m]}(m[Q'+R] - m[R])$ and also $\text{div}(f_Q \circ [m]) = \sum_{R \in E[m]}(m[Q'+R] - m[R])$, so g_Q^m and $f_Q \circ [m]$ have the same divisor, meaning that they differ by a nonzero scalar factor (since the divisor of their ratio is zero, hence is constant).

- Hence by rescaling f_Q, we may assume that $f_Q \circ [m] = g_Q^m$.
- Now suppose we have some other point $P \in E[m]$. Then for any $X \in E$, we see that $g_Q(X+P)^m = f_Q([m]X + [m]P) = f_Q([m]X) = g_Q(X)^m$. Thus, as long as $g_Q(X)$ is not zero or ∞, the ratio $g_Q(X+P)/g_Q(X)$ is some mth root of unity.
- Exercise: Suppose $h \in k(E)$ is a rational function that takes only finitely many values on E. Show that h is constant. (Note as always that k is algebraically closed.)
- By the exercise, since the ratio $g_Q(X+Q)/g_Q(X) \in k(E)$ is a rational function that takes only finitely many values (namely, the roots of unity, and potentially some other values at the finitely many zeroes and poles of g_Q), it must in fact be constant, meaning that it is independent of X.
- Furthermore, since g is defined uniquely up to a constant factor, the ratio $g_Q(X+P)/g_Q(X)$ is independent of the specific choice of g.
- Thus, we obtain a well-defined pairing $e_m(P,Q) = g_Q(X+P)/g_Q(X)$ from $E[m] \times E[m]$ to the multiplicative group of mth roots of unity $\mu_m = \{\zeta \in k : \zeta^m = 1\}$ in k. (Note that because m is not divisible by $\mathrm{char}(k)$, the group μ_m is cyclic of order m.)

- This pairing is called the Weil pairing:

- Definition: Let E/k be an elliptic curve and $m \geq 2$ be an integer not divisible by $p = \mathrm{char}(k)$. The Weil pairing $e_m : E[m] \times E[m] \to \mu_m$ is defined as follows: for any $P, Q \in E[m]$, choose any $g_Q \in k(C)$ such that $\mathrm{div}(g_Q) = [m]^*Q - [m]^*O$, and then define $e_m(P,Q) = g_Q(X+P)/g_Q(X)$ for any $X \in E$ such that the ratio is defined.
  - From our discussion above, the definition of $e_m(P,Q)$ is independent from the specific choice of the function g_P and from the choice of the point X where the ratio is evaluated.

- The Weil pairing has various canonical properties:

- Proposition (Properties of the Weil Pairing): Let E be an elliptic curve and $m \geq 2$ be an integer not divisible by $p = \mathrm{char}(k)$, with $e_m : E[m] \times E[m] \to \mu_m$ the Weil pairing on E. Then the following hold:

1. (Bilinearity) We have $e_m(P_1+P_2,Q) = e_m(P_1,Q)e_m(P_2,Q)$ for any $P_1, P_2, Q \in E[m]$, and $e_m(P, Q_1+Q_2) = e_m(P,Q_1)e_m(P,Q_2)$ for any $P, Q_1, Q_2 \in E[m]$.
   - Proof: For linearity in P we have $e_m(P_1+P_2,Q) = \dfrac{g_Q(X+P_1+P_2)}{g_Q(X)} = \dfrac{g_Q(X+P_1+P_2)}{g_Q(X+P_2)} \cdot \dfrac{g_Q(X+P_2)}{g_Q(X)} = e_m(P_1,Q)e_m(P_2,Q)$ since $\dfrac{g_Q(X+P_1+P_2)}{g_Q(X+P_2)} = \dfrac{g_Q(Y+P_1)}{g_Q(Y)}$ for $Y = X+P_2$.
   - For linearity in Q, let $Q_3 = Q_1+Q_2$ and take f_i, g_i with $\mathrm{div}(f_i) = m[Q_i] - m[O]$ and $\mathrm{div}(g_i) = [m]^*Q_i - [m]^*O$ so that $f_i \circ [m] = g_i^m$ for each $i = 1,2,3$. Since the divisor $[Q_3] - [Q_2] - [Q_1] + [O]$ has degree 0 and resolves to the identity, it is $\mathrm{div}(h)$ for some h.
   - Then $\mathrm{div}(f_3) - \mathrm{div}(f_1 f_2) = m\mathrm{div}(h)$, so $f_3 = c f_1 f_2 h^m$ for some scalar c. Composing on the right with $[m]$ then yields $g_3^m = f_3 \circ [m] = (cf_1f_2h^m) \circ [m] = c(f_1 \circ [m])(f_2 \circ [m])(h \circ [m])^m = cg_1^m g_2^m (h \circ [m])^m$ and now extracting mth roots yields $g_3 = c' g_1 g_2 (h \circ [m])$ for some mth root c' of c.
   - Now we have $e_m(P, Q_1+Q_2) = \dfrac{g_3(X+P)}{g_3(X)} = \dfrac{c'g_1(X+P)g_2(X+P)h([m]X+[m]P)}{c'g_1(X)g_2(X)h([m]X)} = \dfrac{g_1(X+P)}{g_1(X)}\dfrac{g_2(X+P)}{g_2(X)} = e_m(P,Q_1)e_m(P,Q_2)$, where $h([m]X+[m]P) = h([m]X)$ since $P \in E[m]$.
2. (Alternating) We have $e_m(P,P) = 1$ for all $P \in E[m]$, or equivalently, $e_m(P,Q) = e_m(Q,P)^{-1}$ for all $P, Q \in E[m]$.
   - Proof: Take f, g with $\mathrm{div}(f) = m[P] - m[O]$ and $\mathrm{div}(g) = [m]^*P - [m]^*O$ with $g^m = f \circ [m]$.
   - Now for each integer k if we let $\tau_{-kP} : E \to E$ be the translation map $\tau_{-kP}(X) = X - kP$ and also take $f_k = f \circ \tau_{-kP}$, then $\mathrm{div}(f \circ \tau_{-kP}) = m[(1+k)P] - m[kP]$ since composing with τ_{-kP} simply translates zeroes and poles by kP.
   - We can see that $\mathrm{div}(f_0 f_1 \cdots f_{m-1}) = 0$ since the divisor sum telescopes, meaning that the product $f_0 f_1 \cdots f_{m-1}$ is constant. Then for $g_k = g \circ \tau_{-kP'}$ for any P' with $[m]P' = P$, we see that $(g_0 g_1 \cdots g_{m-1})^m = (f_0 f_1 \cdots f_{m-1}) \circ [m]$ is constant whence $g_0 g_1 \cdots g_{m-1}$ is constant.

- This means $g(X)g(X+P')\cdots g(X+(m-1)P') = g_0(X)g_1(X)\cdots g_{m-1}(X) = g_0(X+P')g_1(X+P')\cdots g_{m-1}(X+P') = g(X+P')g(X+2P')\cdots g(X+mP')$ and so cancelling the common terms yields $g(X) = g(X+mP') = g(X+P)$, whence $e_m(P,P) = 1$.
- For the second statement we have $1 = e_m(P+Q,P+Q) = e_m(P,P)e_m(P,Q)e_m(Q,P)e_m(Q,Q) = e_m(P,Q)e_m(Q,P)$ using bilinearity.

3. (Nondegeneracy) If $e_m(P,Q) = 1$ for all $P \in E[m]$, then $Q = O$.
   - Proof: Take f_Q, g_Q with $\mathrm{div}(f_Q) = m[Q] - m[O]$ and $\mathrm{div}(g_Q) = [m]^*Q - [m]^*O$ with $g_Q^m = f_Q \circ [m]$.
   - Suppose $e_m(P,Q) = 1$ for all $P \in E[m]$, meaning that $g_Q(X+P) = g_Q(X)$ for all $P \in E[m]$.
   - This means $g_Q \circ \tau_P = g_Q$ for all translation maps τ_P with $P \in E[m]$. But as we have shown, these translation maps are the elements of the Galois group of the extension $k(E)/[m]^*k(E)$ via the map Ξ sending $P \mapsto \tau_P^*$.
   - Hence g_Q is Galois-invariant, so it is an element of the base field $[m]^*k(E)$, which is to say $g_Q = h \circ [m]$ for some $h \in k(E)$.
   - But now $f_Q \circ [m] = g_Q^m = h^m \circ [m]$ so $f_Q = h^m$. This means $\mathrm{div}(f_Q) = m\mathrm{div}(h)$ so $\mathrm{div}(h) = [Q] - [O]$. But because $[Q] - [O]$ is principal, it must resolve to the identity: thus $Q = O$.
4. (Galois-equivariance) If E is defined over F, then for any $\sigma \in \mathrm{Gal}(k/F)$ we have $e_m(\sigma P, \sigma Q) = \sigma[e_m(P,Q)]$.
   - Proof: Take f_Q, g_Q with $\mathrm{div}(f_Q) = m[Q] - m[O]$ and $\mathrm{div}(g_Q) = [m]^*Q - [m]^*O$ with $g_Q^m = f_Q \circ [m]$.
   - Then $\mathrm{div}(\sigma f_Q) = m[\sigma Q] - m[O]$ and $\mathrm{div}(\sigma g_Q) = [m]^*\sigma Q - [m]^*O$ and $(\sigma g_Q)^m = (\sigma f_Q) \circ [m]$ since the Galois action carries through on divisors and functions, so we have $f_{\sigma Q} = \sigma f_Q$ and $g_{\sigma Q} = \sigma g_Q$.
   - Then $e_m(\sigma P, \sigma Q) = \dfrac{g_{\sigma Q}(X+\sigma P)}{g_{\sigma Q}(X)} = \dfrac{\sigma g_Q(\sigma^{-1}X + P)}{\sigma g_Q(\sigma^{-1}X)} = \sigma\left[\dfrac{g_Q(Y+P)}{g_Q(Y)}\right] = \sigma[e_m(P,Q)]$ where $Y = \sigma^{-1}X$.
5. (Compatibility) For any $P \in E[mm']$ and $Q \in E[m]$ we have $e_{mm'}(P,Q) = e_m([m']P, Q)$.
   - Proof: Take f_Q, g_Q with $\mathrm{div}(f_Q) = m[Q] - m[O]$ and $\mathrm{div}(g_Q) = [m]^*Q - [m]^*O$ with $g_Q^m = f_Q \circ [m]$.
   - Then $\mathrm{div}(f_Q^{m'}) = mm'[Q] - mm'[O]$ and $(g_Q \circ [m'])^{mm'} = (f_Q \circ [m'])^{m'}$.
   - Hence $e_{mm'}(P,Q) = \dfrac{(g \circ [m'])(X+P)}{(g \circ [m'])(X)} = \dfrac{g([m']X + [m']P)}{g([m']X)} = e_m([m']P, Q)$.
6. (Surjectivity) For any mth root of unity ζ_m, there exist $P, Q \in E[m]$ with $e_m(P,Q) = \zeta_m$.
   - Proof: By (1) and (2), we see that the image of $e_m : E[m] \times E[m] \to \mu_m$ is a subgroup of μ_m.
   - Suppose the image has order $d|m$. Then for all P and Q we have $e_m(P,Q)^d = 1$, which by (1) says that $e_m(P, [d]Q) = 1$.
   - By the non-degeneracy property (3), this implies $[d]Q = O$ for all $Q \in E[m]$, which can only happen when $d = m$. Hence e_m is onto, as claimed.
   - Exercise: Suppose E is defined over F and $E[m] \subseteq E(F)$. Show that F contains the mth roots of unity.
   - Exercise: Suppose E is defined over $\mathbb{Q}$ and $p > 2$ is a prime. Show that the p-torsion subgroup of $E(\mathbb{Q})$ is either cyclic or trivial.
7. (Adjoints) For any isogeny $\varphi : E_1 \to E_2$ and any $P \in E_1[m]$ and $Q \in E_2[m]$, we have $e_m^{(1)}(P, \hat{\varphi}(Q)) = e_m^{(2)}(\varphi(P), Q)$ where $e_m^{(i)}$ is the Weil pairing on E_i.
   - Proof: Take f_Q, g_Q with $\mathrm{div}(f_Q) = m[Q] - m[O]$ and $\mathrm{div}(g_Q) = [m]^*Q - [m]^*O$ with $g_Q^m = f_Q \circ [m]$.
   - Observe that $\varphi^*[Q] - \varphi^*[O] - [\hat{\varphi}(Q)] + [O] \in \mathrm{Div}(E_1)$ is principal on E_1 since it has degree 0 and the sum of points resolves to zero, since $\hat{\varphi}(Q)$ is defined to be the sum $\sum_{Q' \in \varphi^{-1}(Q)} Q' - \sum_{R \in \varphi^{-1}(0)} R$ and these are exactly the points in the sum for $\varphi^*[Q]$ and $\varphi^*[O]$ respectively.
   - So choose h with $\mathrm{div}(h) = \varphi^*[Q] - \varphi^*[O] - [\hat{\varphi}(Q)] + [O]$.
   - Now, we have $\mathrm{div}(f_Q \circ \varphi) = \varphi^*\mathrm{div}(f_Q) = m\varphi^*[Q] - m\varphi^*[O]$ by our properties of φ^* acting on divisors, and so $\mathrm{div}\left[\dfrac{f_Q \circ \varphi}{h^m}\right] = m[\hat{\varphi}(Q)] - m[O]$, meaning that we may take $f_{\hat{\varphi}(Q)} = \dfrac{f_Q \circ \varphi}{h^m}$.

- To find a corresponding $g_{\hat{\varphi}(Q)}$ we can observe that $f_{\hat{\varphi}(Q)} \circ [m] = \dfrac{f_Q \circ \varphi}{h^m} \circ [m] = \dfrac{f_Q \circ [m] \circ \varphi}{h^m \circ [m]} = \dfrac{g_Q^m \circ \varphi}{h^m \circ [m]} = \left(\dfrac{g_Q \circ \varphi}{h \circ [m]}\right)^m$, so we may take $g_{\hat{\varphi}(Q)} = \dfrac{g_Q \circ \varphi}{h \circ [m]}$.
- Then $e_m^{(1)}(P, \hat{\varphi}(Q)) = \dfrac{g_{\hat{\varphi}(Q)}(X+P)}{g_{\hat{\varphi}(Q)}(X)} = \dfrac{(g_Q \circ \varphi)(X+P)/(h \circ [m])(X+P)}{(g_Q \circ \varphi)(X)/(h \circ [m])(X)} = \dfrac{g_Q(\varphi(X)+\varphi(P))}{g_Q(\varphi(X))} \cdot \dfrac{h(mX)}{h(mX+mP)} = \dfrac{g_Q(Y+\varphi(P))}{g_Q(Y)} = e_m^{(2)}(\varphi(P), Q)$ where $Y = \varphi(X)$.

- Now that we have given a more natural construction of the Weil pairing on $E[m]$, we can extend this pairing to the Tate module by taking inverse limits.
  - Explicitly, for a prime $l \neq \mathrm{char}(k)$, we have a Weil pairing $e_{l^d} : E[l^d] \times E[l^d] \to \mu_{l^d}$.
  - The Tate module is formed using the inverse system $E[l] \xleftarrow{[l]} E[l^2] \xleftarrow{[l]} E[l^3] \xleftarrow{[l]} E[l^4] \xleftarrow{[l]} \cdots$, and we have the corresponding inverse system on the groups of l-power roots of unity, namely $\mu_l \xleftarrow{l} \mu_{l^2} \xleftarrow{l} \mu_{l^3} \xleftarrow{l} \mu_{l^4} \xleftarrow{l} \cdots$, where the map $l : \mu_{l^{d+1}} \to \mu_{l^d}$ is the lth-power map.
  - Since the groups μ_{l^d} are isomorphic to $\mathbb{Z}/l^d\mathbb{Z}$ by choosing a specific root of unity as generator and making consistent choices the inverse system becomes $\mathbb{Z}/l\mathbb{Z} \xleftarrow{l} \mathbb{Z}/l^2\mathbb{Z} \xleftarrow{l} \mathbb{Z}/l^3\mathbb{Z} \xleftarrow{l} \mathbb{Z}/l^4\mathbb{Z} \xleftarrow{l} \cdots$, which (by using the isomorphism $l\mathbb{Z}/l^{d+1}\mathbb{Z} \cong \mathbb{Z}/l^d\mathbb{Z}$ via dividing representatives by l) is equivalent to our inverse system $\mathbb{Z}/l\mathbb{Z} \xleftarrow{\pi} \mathbb{Z}/l^2\mathbb{Z} \xleftarrow{\pi} \mathbb{Z}/l^3\mathbb{Z} \xleftarrow{\pi} \mathbb{Z}/l^4\mathbb{Z} \xleftarrow{\pi} \cdots$ used to construct $\mathbb{Z}_l$.
  - Hence, by selecting consistent choices of generators for the l^d-power roots of unity (i.e., generators $\zeta_1, \zeta_2, \ldots, \zeta_d, \ldots$ with $\zeta_{d+1}^l = \zeta_d$), which is equivalent to selecting a topological generator of μ_{l^∞}, we may view the Weil pairing as taking its values in $\mathbb{Z}_l$.
  - It remains to show that the inverse-limit structure of $\mathbb{Z}_l$ is consistent with the inverse-limit structure of the Tate module.

- Proposition (Weil Pairing on Tate Module): Let E/k be an elliptic curve and l be a prime with $l \neq \mathrm{char}(k)$. Then the Weil pairings $e_{l^d} : E[l^d] \times E[l^d] \to \mu_{l^d}$ extend to a pairing $e : T_l[E] \times T_l[E] \to \varprojlim_d \mu_{l^d} \cong \mathbb{Z}_l$. This l-adic Weil pairing is bilinear, alternating, nondegenerate, Galois-equivariant, and the dual of an isogeny behaves as an adjoint.
  - Proof: First, the Weil pairings e_{l^d} are compatible with the inverse limit $\varprojlim_d \mu_{l^d}$, since by the compatibility and bilinearity properties we have $e_{l^{d+1}}(P,Q)^l = e_{l^d}([l]P, Q)^l = e_l([l]P, [l]Q)$.
  - The other properties all follow by taking the inverse limit of the properties we showed above for the individual Weil pairings e_{l^d}.

- The l-adic Weil pairing provides the final ingredient for proving the Weil conjectures for elliptic curves:

- Theorem (Weil Conjectures for Elliptic Curves): Let E be an elliptic curve defined over the finite field $\mathbb{F}_q$ of characteristic p and let φ be the qth-power Frobenius map. Then the following hold:

  1. For any prime $l \neq p$, if ψ_l is the image of φ under the l-adic Galois representation $\rho_l : \mathrm{Gal}(k/F) \to \mathrm{Aut}[T_l(E)]$, then $\det(\psi_l) = \deg \varphi$ and $\mathrm{tr}(\psi_l) = 1 + \deg(\varphi) - \deg(1-\varphi)$. In particular, the determinant and trace of ψ_l are integers that are are independent of l, and the characteristic polynomial $\det(T - \psi_l) = T^2 - \mathrm{tr}\psi_l T + \det \psi_l$ has two complex-conjugate roots of absolute value $\sqrt{q}$.
     - Proof: Choose a $\mathbb{Z}_l$-basis $\{v, w\}$ for $T_l(E)$: then the matrix associated to ψ_l with respect to this basis is some 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, meaning that $\psi_l(v) = av + cw$ and $\psi_l(w) = bv + dw$.
     - Using the l-adic Weil pairing we then have $e(v,w)^{\deg \varphi} = e([\deg \varphi]v, w) = e((\hat{\varphi} \circ \varphi)v, w) = e(\varphi v, \varphi w) = e(av+cw, bv+dw) = e(v,w)^{ad-bc} = e(v,w)^{\det \psi_l}$ using the bilinearity, adjoint, and alternating properties. But now since e is nondegenerate, we must have $\deg \varphi = \det \psi_l$.
     - The same calculation also shows that $\deg(1-\varphi) = \det(1-\psi)$. Finally, we have $\mathrm{tr}(\psi_l) = a + d = 1 + (ad - bc) - [(1-a)(1-d) - (-b)(-c)] = 1 + \det(\psi) - \det(1-\psi) = 1 + \deg(\varphi) - \deg(1-\varphi)$, as desired.

- The fact that the determinant and trace are integers and independent of l are immediate, since $\deg\varphi$ and $\deg(1-\varphi)$ are both fixed integers.
- Now, for any rational number m/n, we have $\det(m/n-\psi_l) = \det(m-n\psi_l)/n^2 = \deg(m-n\varphi)/n^2 \geq 0$ since isogenies have nonnegative degree.
- Hence by continuity, the characteristic polynomial $\det(T-\psi_l)$ is nonnegative on $\mathbb{R}$, so it cannot have distinct real roots: thus its roots α and β are complex conjugates (possibly equal), and since their product is $\deg\varphi = q$, each has absolute value $\sqrt{q}$ as claimed.

2. For any $n \geq 1$, $\#E(\mathbb{F}_{q^n}) = q^n + 1 - \alpha^n - \beta^n$ for some complex conjugates α and β of absolute value $\sqrt{q}$.
   - Proof: As we noted in our earlier discussion of the Weil conjectures, $P \in E(\overline{\mathbb{F}_{q^n}})$ if and only if $\varphi^n(P) = P$ if and only if $P \in \ker(1-\varphi^n)$.
   - Then since $(1-\varphi^n)^*\omega = \omega$ the map $1-\varphi^n$ is separable, so $\#E(\mathbb{F}_{q^n}) = \#\ker(1-\varphi^n) = \deg(1-\varphi^n)$.
   - Now since φ^n is the q^nth-power Frobenius map, applying (1) to it yields $\deg(1-\varphi^n) = 1+\deg(\varphi^n) - \mathrm{tr}(\psi_l^n) = 1 + q^n - \alpha^n - \beta^n$ for some complex conjugates α and β of absolute value $\sqrt{q}$.

3. The zeta function $\zeta_C(T) = \dfrac{(1-\alpha T)(1-\beta T)}{(1-T)(1-qT)}$ for some complex conjugates α and β of absolute value $\sqrt{q}$. As an immediate consequence, the Weil conjectures hold for E.
   - Proof: By definition and (2), we have $\ln\zeta_C(T) = \sum_{n=1}^{\infty} \#E(\mathbb{F}_{q^n})\dfrac{T^n}{n} = \sum_{n=1}^{\infty}(1^n+q^n-\alpha^n-\beta^n)\dfrac{T^n}{n} = -\ln(1-T) - \ln(1-qT) + \ln(1-\alpha T) + \ln(1-\beta T)$.
   - Exponentiating immediately yields $\zeta_C(T) = \dfrac{(1-\alpha T)(1-\beta T)}{(1-T)(1-qT)}$.

0.18 (Nov 9) Endomorphism Rings, Part 1

- We now use the Tate module to study isogenies and endomorphisms. To begin, observe that since isogenies also commute with the multiplication-by-l maps, they also act on Tate modules.[11]
  - Explicitly, suppose $\varphi : E_1 \to E_2$ is an isogeny. Then since $\varphi \circ [l] = [l] \circ \varphi$, the action of φ induces a natural map of $\mathbb{Z}_l$-modules $\varphi_l : T_l(E_1) \to T_l(E_2)$ via $\varphi_l(P_1, P_2, P_3, \dots) = (\varphi(P_1), \varphi(P_2), \varphi(P_3), \dots)$.
  - Since the componentwise action is clearly additive in the isogeny φ, we obtain a group homomorphism $\Psi : \mathrm{Hom}(E_1, E_2) \to \mathrm{Hom}(T_l(E_1), T_l(E_2))$.
  - Exercise: Show that when $E_1 = E = E_2$, the action $\Psi : \mathrm{End}(E) \to \mathrm{End}(T_l(E))$ with $\Psi(\varphi)$ mapping $(P_1, P_2, P_3, \dots) \in T_l(E)$ to $(\varphi(P_1), \varphi(P_2), \varphi(P_3), \dots) \in T_l(E)$ is a ring homomorphism.
  - Indeed, when $T_l(E_1) \neq 0$, which we know occurs whenever $l \neq \mathrm{char}(k)$, this homomorphism $\Psi : \mathrm{Hom}(E_1, E_2) \to \mathrm{Hom}(T_l(E_1), T_l(E_2))$ is injective.
  - To see this suppose that $\varphi \in \ker(\Psi)$ so that $\varphi(T_l(E_1)) = 0$, which is equivalent to saying that $E[l^d] \in \ker\varphi$ for all d. In particular, $\ker\varphi$ is infinite: but as we showed, nonzero isogenies have a finite kernel, and so we must have $\varphi = 0$.
  - In fact, a much stronger statement is actually true:

- Proposition (Isogeny Action on Tate Modules): Let E_1 and E_2 be elliptic curves over k and l be a prime not equal to $\mathrm{char}(k)$. Then the natural map $\Psi_l : \mathrm{Hom}(E_1, E_2) \otimes \mathbb{Z}_l \to \mathrm{Hom}(T_l(E_1), T_l(E_2))$ defined by mapping $\varphi \otimes 1 \mapsto \varphi_l$ and then extending $\mathbb{Z}_l$-linearly, is injective, where $\varphi_l(P_1, P_2, P_3, \dots) = (\varphi(P_1), \varphi(P_2), \varphi(P_3), \dots)$.
  - The proof of this result is somewhat involved. Let us motivate the general idea by attempting to give a direct argument first.
  - Suppose $\varphi \in \ker\Psi_l$, so that $\varphi_l = 0$. By definition of the tensor product, φ is a finite sum of simple tensors $\varphi = \varphi_1 \otimes \alpha_1 + \cdots + \varphi_k \otimes \alpha_k$ for some isogenies $\varphi_i : E_1 \to E_2$ and some scalars $\alpha_i \in \mathbb{Z}_i$.
  - Now, the fact that $\varphi_l = 0$ means that φ annihilates $E[l^d]$ for each $d \geq 1$.

[11] In fact, we have really been using this observation already, since we used some facts about $1 - \mathrm{Frob}$ to prove the Weil conjectures for elliptic curves.

- So, if we write $\alpha_i = a_{i,0} + a_{i,1}l + a_{i,2}l^2 + \cdots$, then for $b_{i,d} = a_{i,0} + a_{i,1}l + \cdots + a_{i,d}l^d$ (i.e., the reduction of α_i mod l^d) we see that $b_{1,d}\varphi_1 + b_{2,d}\varphi_2 + \cdots + b_{k,d}\varphi_k$ annihilates $E[l^d]$ for each $d \geq 1$.
- We would now like to apply a similar argument as the one earlier: namely, to observe that we would have an isogeny whose kernel is infinite, hence it must be zero. Unfortunately, this does not work because all of the isogenies $\psi_d = b_{1,d}\varphi_1 + b_{2,d}\varphi_2 + \cdots + b_{k,d}\varphi_k$ are different for different d, so there is nothing to prevent them from having increasingly large kernels.
- Instead, observe that $\ker \psi_d$ contains $E[l^d] = \ker[l^d]$, and so since $[l^d]$ is separable, by our results on composition of isogenies, we know that there exists some other isogeny λ_d such that $\psi_d = [l^d] \circ \lambda_d$.
- What we would like to say is that this forces all of the coefficients $b_{i,d}$ to be multiples of l^d, which since they are obtained by reduction of α_i mod l^d, would mean that they are all zero. This would follow, for instance, if the φ_i were linearly independent and $\lambda_d \in \text{span}(\varphi_i)$.
- We can recover the core of the argument using the following lemma:
- Lemma: Let M be a finitely generated subgroup of $\text{Hom}(E_1, E_2)$ and let $M_{\text{div}} = \{\varphi \in \text{Hom}(E_1, E_2) : [m] \circ \varphi \in M \text{ for some } m \geq 1\}$ be the subgroup of M-divisible elements. Then M_{div} is also finitely generated.
- Proof (of Lemma): Since M is finitely generated, the tensor product $M \otimes \mathbb{R}$ is a finitely generated $\mathbb{R}$-vector space, which we may endow with the natural topology from $\mathbb{R}$. We may then extend the degree map to give a continuous real-valued function on $M \otimes \mathbb{R}$: the degree map with integer coefficients is a quadratic form on M as we previously showed, so we may simply extend scalars to $\mathbb{R}$, in which case the corresponding real-valued quadratic form is certainly continuous.
- By continuity, the set $U = \{\varphi \in M \otimes \mathbb{R} : \deg \varphi < 1\}$ is open, and since M is torsion-free (because $\text{Hom}(E_1, E_2)$ is torsion-free), we see that M_{div} injects into $M \otimes \mathbb{R}$ (indeed, it is a subset of $M \otimes \mathbb{Q}$).
- Since $M_{\text{div}} \cap U = \{0\}$ since all elements of M_{div} are isogenies and thus nonzero elements of M_{div} have degree at least 1, this means M_{div} is a discrete subgroup of a finite-dimensional real vector space, and hence is finitely generated (in fact, the number of generators is at most $\dim_{\mathbb{R}}(M \otimes \mathbb{R}) = \text{rank}_{\mathbb{Z}} M$).
- Now we can prove the original result:
- Proof (of Proposition): Suppose $\varphi \in \text{Hom}(E_1, E_2) \otimes \mathbb{Z}_l$. Then φ is some finite sum of simple tensors. Let M be the subgroup of $\text{Hom}(E_1, E_2)$ spanned by the isogenies in those tensors.
- By the Lemma, the subgroup M_{div} is finitely generated, say with a basis $\varphi_1, \ldots, \varphi_k$. Then since $\varphi \in M_{\text{div}}$ we have $\varphi = \varphi_1 \otimes \alpha_1 + \cdots + \varphi_k \otimes \alpha_k$ for some isogenies $\varphi_i : E_1 \to E_2$ and some scalars $\alpha_i \in \mathbb{Z}_i$.
- As above, writing $\alpha_i = a_{i,0} + a_{i,1}l + a_{i,2}l^2 + \cdots$ and setting $b_{i,d}$ to be the reduction of α_i mod l^d, we see that $\psi_d = b_{1,d}\varphi_1 + b_{2,d}\varphi_2 + \cdots + b_{k,d}\varphi_k$ annihilates $E[l^d]$ for each $d \geq 1$. Thus $\ker \psi_d$ contains $E[l^d] = \ker[l^d]$ and so $\psi_d = [l^d] \circ \lambda_d$ for some isogeny λ_d.
- But now $\lambda_d \in M_{\text{div}}$, so $\lambda_d = c_1\varphi_1 + \cdots + c_k\varphi_k$ for some integers $c_1, \ldots, c_k$, since $\{\varphi_1, \ldots, \varphi_k\}$ is a basis for M_{div}.
- Then the statement $\psi_d = [l^d] \circ \lambda_d$ implies that $l^d c_1\varphi_1 + \cdots + l^d c_k\varphi_k = b_{1,d}\varphi_1 + b_{2,d}\varphi_2 + \cdots + b_{k,d}\varphi_k$ so all the coefficients must agree. But since $b_{i,d}$ is the reduction of α_i mod l^d, this means $b_{i,d} = 0$ for each i.
- Since this holds for each $d \geq 1$, we see that all of the l-adic series coefficients of α_i are zero for each i, meaning each $\alpha_i = 0$ and thus that $\varphi = 0$ as desired.

- Now we can use this result to describe more explicitly the structure of isogeny groups and endomorphism rings:

- Proposition (Isogeny Groups and Endomorphism Rings): Let E_1, E_2, and E be elliptic curves. Then the following hold:

  1. The group of isogenies $\text{Hom}(E_1, E_2)$ is a free abelian group of rank at most 4.
     - Proof: Because $\mathbb{Z}_l$ has characteristic zero and the isogeny group is torsion-free as we showed previously, by basic properties of tensor products we see that $\text{rank}_{\mathbb{Z}}\text{Hom}(E_1, E_2) = \text{rank}_{\mathbb{Z}_l}[\text{Hom}(E_1, E_2) \otimes \mathbb{Z}_l]$.
     - Explicitly, a $\mathbb{Z}$-basis of $\text{Hom}(E_1, E_2)$ is also a $\mathbb{Z}_l$-basis of $\text{Hom}(E_1, E_2) \otimes \mathbb{Z}_l$, and if there is no $\mathbb{Z}$-basis for $\text{Hom}(E_1, E_2)$ because its rank is infinite, an infinite linearly independent subset of $\text{Hom}(E_1, E_2)$ yields an infinite linearly independent subset of $\text{Hom}(E_1, E_2) \otimes \mathbb{Z}_l$.

- By the proposition above, $\mathrm{Hom}(E_1, E_2) \otimes \mathbb{Z}_l$ injects into $\mathrm{Hom}(T_l(E_1), T_l(E_2)) \cong \mathrm{Hom}(\mathbb{Z}_l \times \mathbb{Z}_l, \mathbb{Z}_l \times \mathbb{Z}_l) \cong M_{2\times 2}(\mathbb{Z}_l)$, which has rank 4 as a $\mathbb{Z}_l$-module.
- Hence $\mathrm{rank}_{\mathbb{Z}}\mathrm{Hom}(E_1, E_2) = \mathrm{rank}_{\mathbb{Z}_l}[\mathrm{Hom}(E_1, E_2) \otimes \mathbb{Z}_l] \leq \mathrm{rank}_{\mathbb{Z}_l}[\mathrm{Hom}(T_l(E_1), T_l(E_2))] = 4$, as claimed.

2. The endomorphism ring $\mathrm{End}(E)$ is a $\mathbb{Z}$-algebra of characteristic 0 and rank at most 4 possessing an anti-involution $\varphi \mapsto \hat{\varphi}$ such that $\widehat{\varphi + \psi} = \hat{\varphi} + \hat{\psi}$, $\widehat{\varphi\psi} = \hat{\psi}\hat{\varphi}$, $\hat{\hat{\varphi}} = \varphi$, $\widehat{[m]} = [m]$ for all $m \in \mathbb{Z}$, and for which $\varphi\hat{\varphi} = \hat{\varphi}\varphi$ is a nonnegative integer that equals 0 only when $\varphi = 0$.
   - Proof: The multiplication-by-m maps are a subring of $\mathrm{End}(E)$ isomorphic to $\mathbb{Z}$, and since $[m]$ commutes with all isogenies, $\mathrm{End}(E)$ is a $\mathbb{Z}$-algebra of characteristic 0.
   - The rank statement is simply (1) in the case $E = E_1 = E_2$. The existence of the anti-involution and all of the associated properties all follow from our results on dual isogenies and the fact that $\hat{\varphi} \circ \varphi = [\deg \varphi] \geq 0$.

- We can now give a general classification of the possible endomorphism rings of an elliptic curve based on the properties established in (2) of the proposition above.

  - If F is a field, recall that an associative F-algebra is a ring extension A of F that is also a finite-dimensional F-vector space, where the vector space and ring structures are compatible. (Explicitly, for $r, s \in A$ and $\alpha \in F$, this means $\alpha(rs) = (\alpha r)s = r(\alpha s)$.)
  - Example: Any finite-degree field extension L/F is an associative F-algebra, as is the matrix ring $M_{n\times n}(F)$.
  - Example: The division ring of real quaternions $\mathbb{H} = \mathbb{R} + \mathbb{R}i + \mathbb{R}j + \mathbb{R}k$ with $i^2 = j^2 = k^2 = ijk = -1$ (so that $ij = -ji$, etc.) is an associative $\mathbb{R}$-algebra. (In fact, other than $\mathbb{R}$ and $\mathbb{C}$, the quaternions are the only other associative $\mathbb{R}$-algebra having no zero divisors: this is a theorem of Frobenius.)
  - If A is an associative F-algebra, an order of A is a finitely-generated subring R of A such that $R \otimes F = A$. (We emphasize here that part of this definition also includes the fact that A and R have the same multiplicative identity.)
  - Equivalently, and more concretely, an order is a subring R (with 1) such that as a $\mathbb{Z}$-module, we have $\mathrm{rank}_{\mathbb{Z}} R = \dim_{\mathbb{Q}} A$.

- As we will show, there are three different types of $\mathbb{Z}$-algebras of characteristic 0 with rank at most 4 having an anti-involution and a positive-definite quadratic form defined on them, which are as follows:

  1. The ring $\mathbb{Z}$.
  2. Orders in imaginary quadratic fields. Recall that an imaginary quadratic field is a field extension $K/\mathbb{Q}$ of degree 2 which is not a subfield of $\mathbb{R}$. Explicitly, $K = \mathbb{Q}(\sqrt{-D})$ for some squarefree positive integer $D > 0$, meaning that $K = \mathbb{Q} + \mathbb{Q}\sqrt{-D}$. Equivalently, an imaginary quadratic field is an associative $\mathbb{Q}$-algebra of the form $A = \mathbb{Q} + \mathbb{Q}\alpha$ with multiplication rules $\alpha^2 \in \mathbb{Q}$ and $\alpha^2 < 0$.
  3. Orders in definite quaternion algebras. Recall that a definite quaternion algebra is the quaternionic analogue of an imaginary quadratic field extension of $\mathbb{Q}$: it is an associative $\mathbb{Q}$-algebra of the form $A = \mathbb{Q} + \mathbb{Q}\alpha + \mathbb{Q}\beta + \mathbb{Q}\alpha\beta$ with multiplication rules $\alpha^2, \beta^2 \in \mathbb{Q}$ with $\alpha^2 < 0$, $\beta^2 < 0$, and $\alpha\beta = -\beta\alpha$.

  - Exercise: If $K = \mathbb{Q}(\sqrt{-D})$ is an imaginary quadratic field, its ring of integers $\mathcal{O}_K$ is $\mathbb{Z}[\alpha]$ where $\alpha = \begin{cases} \sqrt{-D} & \text{when } -D \equiv 2, 3 \pmod 4 \\ (1 + \sqrt{-D})/2 & \text{when } -D \equiv 1 \pmod 4 \end{cases}$. Show that the orders of K are the rings of the form $R = \mathbb{Z} + f\mathcal{O}_K$ for a positive integer f, the conductor of R. [Hint: Show $[\mathcal{O}_K : R] = f$ is finite.]
  - Exercise: Show that both the ring of naive integral quaternions $R = \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}k$ and the ring of Hurwitz quarternions $H = \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}\dfrac{1 + i + j + k}{2}$ are orders in the algebra $A = \mathbb{Q} + \mathbb{Q}i + \mathbb{Q}j + \mathbb{Q}k$ of rational quaternions, where $i^2 = j^2 = k^2 = ijk = -1$ as usual.
  - Exercise: Show that if S is a Noetherian integrally closed domain with fraction field F, then $R = M_{n\times n}(S)$ is an order in $A = M_{n\times n}(F)$.

- Let us now show the claimed result:

- Theorem (Endomorphism Rings): Let R be a ring with 1, of characteristic 0, having no zero divisors, which has rank at most 4 as a $\mathbb{Z}$-module, and has an anti-involution $r \mapsto \hat{r}$ such that $\widehat{r+s} = \hat{r} + \hat{s}$, $\widehat{rs} = \hat{s}\hat{r}$, $\hat{\hat{r}} = r$, $\hat{m} = m$ for all $m \in \mathbb{Z}$, and for which $r\hat{r} = \hat{r}r$ is a nonnegative integer that equals 0 only when $r = 0$. Then R is either isomorphic to $\mathbb{Z}$, to an order in an imaginary quadratic field, or to an order in a definite quaternion algebra.
  - Proof: Since R is finitely generated (as it is torsion-free and has rank at most 4 as an additive group), it suffices to prove that $K = R \otimes \mathbb{Q}$ is either $\mathbb{Q}$, an imaginary quadratic field, or a definite quaternion algebra.
  - We can extend the anti-involution to K in the natural way by taking $\widehat{r \otimes (p/q)} = (p/q)\hat{r}$ and extending linearly. It is immediate that the resulting anti-involution on K retains all of the same properties except that $r\hat{r}$ is now a nonnegative rational number.
  - Define norm and trace functions on K via $\mathrm{nm}(\alpha) = \alpha\hat{\alpha}$ and $\mathrm{tr}(\alpha) = \alpha + \hat{\alpha}$ for $\alpha \in K$.
  - Clearly the norm is always rational and multiplicative, and since $1 + \mathrm{nm}(\alpha) - \mathrm{nm}(1-\alpha) = 1 + \alpha\hat{\alpha} - (1-\hat{\alpha})(1-\alpha) = \alpha + \hat{\alpha} = \mathrm{tr}(\alpha)$ we see that the trace is rational and additive. Additionally, for rational numbers x we have $\mathrm{tr}(x) = x + \hat{x} = 2x$.
  - Observation: if $\mathrm{tr}(\alpha) = 0$, then $0 = (\alpha - \alpha)(\alpha - \hat{\alpha}) = \alpha^2 - \mathrm{tr}(\alpha)\alpha + \mathrm{nm}(\alpha) = \alpha^2 + \mathrm{nm}(\alpha)$, so $\alpha^2 \in \mathbb{Q}$ and $\alpha^2 \leq 0$ with equality only when $\alpha = 0$.
  - If $K = \mathbb{Q}$ we are trivially done so suppose $K \neq \mathbb{Q}$ and pick $\alpha \notin \mathbb{Q}$. Then the element $\beta = \alpha - \frac{1}{2}\mathrm{tr}(\alpha)$ has $\mathrm{tr}(\beta) = \mathrm{tr}(\alpha) - \mathrm{tr}(\frac{1}{2}\mathrm{tr}(\alpha)) = \mathrm{tr}(\alpha) - \mathrm{tr}(\alpha) = 0$, and β is nonzero.
  - Hence by the Observation, $\beta^2 \in \mathbb{Q}$ and $\beta^2 < 0$, so if $K = \mathbb{Q}(\beta)$ we are done since K is an imaginary quadratic field. Otherwise, suppose $K \neq \mathbb{Q}(\beta)$ and select $\gamma \notin \mathbb{Q}(\beta)$.
  - Define the element $\delta = \gamma - \frac{1}{2}\mathrm{tr}(\gamma) - \frac{1}{2} \cdot \frac{\mathrm{tr}(\beta\gamma)}{\beta^2}\beta$ and observe that $\frac{\mathrm{tr}(\beta\gamma)}{\beta^2}$ is rational since β^2 is a negative rational.
  - Then $\mathrm{tr}(\delta) = \mathrm{tr}(\gamma) - \mathrm{tr}(\gamma) - \frac{1}{2} \cdot \frac{\mathrm{tr}(\beta\gamma)}{\beta^2}\mathrm{tr}(\beta) = 0$ since $\mathrm{tr}(\beta) = 0$ as calculated earlier, and hence $\delta^2 < 0$.
  - We also have $\mathrm{tr}(\beta\delta) = \mathrm{tr}(\beta\gamma) - \frac{1}{2}\mathrm{tr}(\gamma)\mathrm{tr}(\beta) - \frac{1}{2}\mathrm{tr}(\beta\gamma) = 0$ as well.
  - So now since $\mathrm{tr}(\beta) = \mathrm{tr}(\delta) = \mathrm{tr}(\beta\delta) = 0$ we have $\beta = -\hat{\beta}$, $\delta = -\hat{\delta}$, and so $\beta\delta = -\widehat{\beta\delta} = -\hat{\delta}\hat{\beta} = -\delta\beta$.
  - This means $\mathbb{Q}[\beta, \delta] = \mathbb{Q} + \mathbb{Q}\beta + \mathbb{Q}\delta + \mathbb{Q}\beta\delta$ has $\beta^2 < 0$, $\delta^2 < 0$, and $\beta\delta = -\delta\beta$, so it is a quaternion algebra.
  - Now we claim that $\{1, \beta, \delta, \beta\delta\}$ is $\mathbb{Q}$-linearly independent, so suppose we had a linear dependence $w + x\beta + y\delta + z\beta\delta = 0$.
  - Taking the trace yields $2w = 0$ so $w = 0$. Now multiplying $x\beta + y\delta + z\beta\delta = 0$ on the left by β and on the right by δ yields $(x\beta^2)\delta + (y\delta^2)\beta + z\beta^2\delta^2 = 0$, where we used $\beta^2, \delta^2 \in \mathbb{Q}$ to move terms.
  - But this is a $\mathbb{Q}$-linear dependence between β, δ, and 1, but because $\beta \notin \mathbb{Q}$ and $\delta \notin \mathbb{Q}(\beta)$ all coefficients must be zero, and so $\{1, \beta, \delta, \beta\delta\}$ is linearly independent.
  - This means $\mathbb{Q}[\beta, \delta]$ is a 4-dimensional $\mathbb{Q}$-vector space, hence it must equal K since the dimension of K is at most 4 (since the rank of R is at most 4) by assumption.
- Interestingly, the characteristic-0 and characteristic-p behaviors for the endomorphism ring structure are quite different.
  - In characteristic zero, as we will show later in a more concrete way using analytic methods, the endomorphism ring cannot be a quaternion order, and indeed it can only be an order in an imaginary quadratic field in rather special situations: typically it is just $\mathbb{Z}$.
  - In positive characteristic, any of the three possible endomorphism ring types can occur, but there is a rather interesting separation of the behaviors.
  - For an elliptic curve defined over $\overline{\mathbb{F}_p}$, the endomorphism ring is always bigger than $\mathbb{Z}$, and we can characterize fairly precisely when the endomorphism ring is a definite quaternion order or an imaginary quadratic order.
  - Over other fields of positive characteristic, when the elliptic curve is not defined over $\overline{\mathbb{F}_p}$, the endomorphism ring is always $\mathbb{Z}$.

0.19 (Nov 13) Endomorphism Rings, Part 2

- For now, let us continue our study of the endomorphism ring by using some general tools from the study of central simple algebras and the Brauer group.
  - For a field k, a central simple k-algebra A is a (finite-dimensional) associative k-algebra which is simple and whose center is precisely k.
  - Exercise: If k is a field and D is a division ring with center k, show that the matrix algebra $M_{n\times n}(D)$ is a central simple k-algebra.
  - In fact, by Wedderburn's theorem, *every* central simple k-algebra is of the form $M_{n\times n}(D)$ for some (unique up to isomorphism) division ring D with center k, and some (unique) n.
  - When two central simple k-algebras have the same (or to be pedantic, isomorphic) division ring D, we say $A \sim B$. This relation is clearly an equivalence relation, and its equivalence classes are denoted $[A]$.
  - The classes of central simple k-algebras form an abelian group $\mathrm{Br}(k)$, called the Brauer group of k, with multiplication $[A][B] = [A \otimes_k B]$ given by tensor product, identity $[k]$, and inverses given by $[A]^{-1} = [A^{\mathrm{opp}}]$, where A^{opp} denotes the opposite ring of A (the ring with the same elements and addition, but where multiplication is reversed: $r^{\mathrm{opp}} + s^{\mathrm{opp}} = r + s$ and $r^{\mathrm{opp}} s^{\mathrm{opp}} = sr$).
  - Restricting now to the case where k is a global field (either an algebraic number field or a function field over $\mathbb{F}_q$ of transcendence degree 1), a place $\mathfrak{p}$ of k is either a finite prime ideal of the ring of integers of k (which has an associated $\mathfrak{p}$-adic metric), or an infinite place associated to one of the embeddings of k into its algebraic closure. To each place we have an associated metric (either the $\mathfrak{p}$-adic metric for a finite place $\mathfrak{p}$ or an archimedean metric for an infinite place) yielding a corresponding completion $k_{\mathfrak{p}}$ under that metric.
  - When $k = \mathbb{Q}$, the places are simply the integer primes p, whose associated metric is the p-adic metric yielding the completion $\mathbb{Q}_p$ (the fraction field of the p-adic integer ring $\mathbb{Z}_p$), along with a single infinite place ∞ whose metric is the usual absolute value and whose completion is $\mathbb{R}$.
  - For each place $\mathfrak{p}$ of k, we obtain a homomorphism $\mathrm{Br}(k) \to \mathrm{Br}(k_{\mathfrak{p}})$ defined by $[A] \mapsto [A \otimes_k k_{\mathfrak{p}}]$. One may prove that $[A \otimes_k k_{\mathfrak{p}}] = 1$ for all but finitely many $\mathfrak{p}$ (this is the analogue of the fact that any integer has only finitely many prime factors), and thus there is a well-defined homomorphism $\mathrm{Br}(k) \to \bigoplus_{\mathfrak{p}} \mathrm{Br}(k_{\mathfrak{p}})$.
  - In fact, this map is injective (this is a result known as the Albert-Brauer-Hasse-Noether theorem), and a stronger result, due to Hasse, fits this map into an exact sequence $1 \to \mathrm{Br}(k) \to \bigoplus_{\mathfrak{p}} \mathrm{Br}(k_{\mathfrak{p}}) \xrightarrow{\sum \mathrm{inv}_p} \mathbb{Q}/\mathbb{Z} \to 0$, where inv_p denotes the Hasse invariant map, which we will not describe in general. (The exactness of this sequence is a fundamental result connecting local and global class field theory.)
  - Furthermore, the sum of all of the Hasse invariants (over all places) is an integer, and it follows from the Grunwald-Wang theorem that the order of $[A]$ in the Brauer group is equal to $\sqrt{[A:k]}$ where k is the center of A.
  - We will also mention that there is a very nice way to construct central simple algebras using cohomology classes (known as the crossed product construction). Through this construction one may prove that the Brauer group $\mathrm{Br}(k)$ is canonically isomorphic to the second cohomology group $H^2(G_{\overline{k}/k}, \overline{k}^*)$.
- Now we can apply these results to the algebra $A = \mathrm{End}(E) \otimes \mathbb{Q}$, which is a central simple k-algebra for $k = \mathbb{Q}$ (if $\mathrm{End}(E)$ is $\mathbb{Z}$ or a quaternion order) or an imaginary quadratic field (if $\mathrm{End}(E)$ is an imaginary quadratic order).
  - In the specific case $k = \mathbb{Q}$, the Hasse invariant inv_p is calculated as follows: when $A \otimes \mathbb{Q}_p \cong M_{n\times n}(\mathbb{Q}_p)$ (when $[A_p]$ is the identity class in which case we say A splits at p) we have $\mathrm{inv}_p = 0$, and otherwise (when $[A_p]$ is not the identity class in which case we say A ramifies at p) we have $\mathrm{inv}_p = 1/2$.
  - Additionally, from the Brauer-group exact sequence, the sum $\sum_p \mathrm{inv}_p$ is always an integer.
  - Also, the order of $[A]$ in the Brauer group is equal to $\sqrt{[A:k]}$ where k is the center of A, so when $\mathrm{End}(E)$ is a quaternion order, the order of $[A]$ equals $\sqrt{4} = 2$.
- Proposition (Endomorphisms in Characteristic Zero, I): Suppose E is an elliptic curve in characteristic zero. Then $\mathrm{End}(E)$ cannot be an order in a quaternion algebra.

- Proof: Suppose by way of contradiction that $\mathrm{End}(E)$ is an order in a quaternion algebra. Let $A = \mathrm{End}(E) \otimes \mathbb{Q}$ and let l be a finite prime.
- By our results on isogenies and the Tate module we know that $\mathrm{End}(E) \otimes \mathbb{Z}_l$ injects into $\mathrm{Aut}(T_l(E)) \cong M_{2\times 2}(\mathbb{Z}_l)$.
- Tensoring with $\mathbb{Q}_l$ shows that $A \otimes \mathbb{Q}_l$ injects into $M_{2\times 2}(\mathbb{Z}_l) \otimes \mathbb{Q}_l \cong M_{2\times 2}(\mathbb{Q}_l)$.
- When $\mathrm{End}(E)$ is a quaternion order, its rank is 4, and so $A \otimes \mathbb{Q}_l$ is (isomorphic to) a 4-dimensional subspace of $M_{2\times 2}(\mathbb{Q}_l)$, but this subspace must therefore be all of $M_{2\times 2}(\mathbb{Q}_l)$.
- Thus, $A \otimes \mathbb{Q}_l \cong M_{2\times 2}(\mathbb{Q}_l)$ and thus $\mathrm{inv}_l[A] = 0$ for all finite places l.
- But since there is a unique infinite place and the sum of the invariants is an integer, we must also have $\mathrm{inv}_\infty[A] = 0$.
- Now, because the map $\mathrm{Br}(\mathbb{Q}) \to \bigoplus_p \mathrm{Br}(\mathbb{Q}_p)$ is injective, this means $[A]$ is the identity element of the Brauer group. But this is, at last, a contradiction, because $\mathrm{End}(E) \otimes \mathbb{Q}$ is not a matrix group since it has no zero divisors. (Alternatively, we could get a contradiction from observing that the order of $[A]$ is 1, whereas $\sqrt{[A : \mathbb{Q}]} = 2$.)

- Exercise: Show that the endomorphism ring of $y^2 = x^3 - x$ with the isogeny $[i](x, y) = (-x, iy)$ discussed previously, is isomorphic to $\mathbb{Z}[i]$.

- On the other hand, in positive characteristic, we have several different classes based on the isomorphism type of the elliptic curve.
  - It turns out that over $\mathbb{F}_q$, there are 2 possibilities: either $\mathrm{End}(E)$ is an order in an imaginary quadratic field or $\mathrm{End}(E)$ is an order in a quaternion algebra.
  - We have previously seen that over $\mathbb{F}_q$ there are also 2 different possibilities for the p-power torsion: either all of the groups are cyclic, or all of the groups are trivial.
  - In fact, perhaps rather surprisingly, it turns out that these two situations align completely: the p-power torsion groups are trivial if and only if the endomorphism ring is an order in a quaternion algebra!

- In order to phrase this result properly we first require a quantity that allows us to characterize elliptic curves up to isomorphism.

- Definition: Let E be an elliptic curve defined over a field F. If E has a Weierstrass equation $y^2 = x^3 + Ax + B$, the j-invariant of E is defined to be the quantity $j = -1728\dfrac{(4A)^3}{\Delta}$ where $\Delta = -16(4A^3 + 27B^2)$ is the discriminant of E.
  - We note that the j-invariant is well-defined because the discriminant Δ is nonzero because elliptic curves are defined to be nonsingular. Additionally, even though the Weierstrass equation for E is not unique because we may perform a rescaling $A' = u^4A$, $B' = u^4B$, the discriminant transforms as $\Delta' = u^{12}\Delta$ and so we see $j' = j$ is invariant. (This is why it is called the j-invariant.)
  - For an arbitrary Weierstrass form $y^2+a_1xy+a_3y = x^3+a_2x^2+a_4x+a_6$ these definitions are more involved. Setting $b_2 = a_1^2 + 4a_2$, $b_4 = 2a_4 + a_1a_3$, $b_6 = a_3^2 + 4a_6$, and then $b_8 = a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2$, we have $\Delta = -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6$ and $j = (-b_2^3 + 36b_2b_4 - 216b_6)^3/\Delta$.
  - For simplicity we will give the discussion in the case $\mathrm{char}(k) \neq 2, 3$ since then we may use the much more understandable formula $j = -1728(4A)^3/\Delta$.

- The j-invariant characterizes an elliptic curve up to isomorphism, in the sense that E_1 and E_2 are isomorphic (over the algebraic closure k) if and only if they have the same j-invariant:

- Theorem (j-Invariants and Isomorphism): Let k be an algebraically closed field.

1. If E_1 and E_2 are elliptic curves over the algebraically closed field k, then E_1 and E_2 are isomorphic over k if and only if $j(E_1) = j(E_2)$.
   - Recall that the only transformations preserving a reduced Weierstrass form (in characteristic not 2 or 3) are $(x', y') = (u^{-2}x, u^{-3}y)$ yielding $(A', B') = (u^4A, u^6B)$.

- Proof (for $\mathrm{char}(k) \neq 2,3$): Our calculations above show that if E_1 and E_2 are isomorphic then $j(E_1) = j(E_2)$ since any transformation of the Weierstrass form of E_1 leaves the j-invariant unchanged.
- Conversely, suppose $j(E_1) = j(E_2)$ with $E_1 : y^2 = x^3 + A_1x + B_1$ and $E_2 : y^2 = x^3 + A_2x + B_2$. Then $j(E_1) = j(E_2)$ implies $A_1^3/(27A_1^3+4B_1^2) = A_2^3/(27A_2^3+4B_2^2)$ which upon clearing denominators becomes $27A_1^3A_2^3 + 4A_1^3B_2^2 = 27A_1^3A_2^3 + 4A_2^3B_1^2$ whence $A_1^3B_2^2 = A_2^3B_1^2$.
- If $A_1 = 0$ then since $\Delta \neq 0$ we must have $B_1 \neq 0$ and so $A_2 = 0$. Then $(A_2, B_2) = (u^4A_1, u^6B_1)$ for $u = (B_2/B_1)^{1/6}$.
- If $B_1 = 0$ then since $\Delta \neq 0$ we must have $A_1 \neq 0$ and so $B_2 = 0$. Then $(A_2, B_2) = (u^4A_1, u^6B_1)$ for $u = (A_2/A_1)^{1/4}$.
- If $A_1, B_1 \neq 0$ then $A_2, B_2 \neq 0$ also. Then $(A_2, B_2) = (u^4A_1, u^6B_1)$ for $u = (A_2/A_1)^{1/4} = (B_2/B_1)^{1/6}$; these are equal because $A_1^3/A_2^3 = B_1^2/B_2^2$.
- Remark: The proof in characteristics 2 and 3 is essentially the same, just with more subcases based on the various possibilities for the additional possibilities for reduced 2-parameter Weierstrass equations (e.g., $y^2 + xy = x^3 + a_4x + a_6$ in characteristic 2, or $y^2 = x^3 + a_2x^2 + a_6$ in characteristic 3).

2. If F is any subfield of k and $j_0 \in F$, then there exists an elliptic curve defined over F with j-invariant j_0.
   - Proof: If $j_0 \neq 0, 1728$ then it is not hard to check that the Tate curve $E : y^2 + xy = x^3 - \dfrac{36}{j_0 - 1728} - \dfrac{1}{j_0 - 1728}$ has discriminant $\Delta = j_0^2/(j_0 - 1728)^3$ and j-invariant $j = j_0$.
   - For the remaining cases, we can similarly check that the curve $y^2 + y = x^3$ has $\Delta = -27$ and $j = 0$, and the curve $y^2 = x^3 + x$ has $\Delta = -64$ and $j = 1728$. These yield nonsingular curves in characteristic not 2 or 3.
   - When $\mathrm{char}(k) = 2, 3$ we have $1728 = 0$ and so there is only one remaining j-invariant to check. In characteristic 2 the first curve is nonsingular while in characteristic 3 the second curve is nonsingular, so this accounts for all cases.

- Now we can give various equivalent conditions for the endomorphism ring structure of an elliptic curve over $\mathbb{F}_q$:

- Theorem (Endomorphism Rings Over $\mathbb{F}_q$): Let F be a perfect field of characteristic p (e.g., $\mathbb{F}_q$) and E be an elliptic curve defined over F. For each integer $r \geq 1$ define $\varphi_r : E \to E^{(p^r)}$ to be the p^r-power Frobenius map and denote $\varphi_1 = \varphi$. Then the following are equivalent:

  1. The p-torsion group $E[p]$ is trivial.
  2. The p-power torsion groups $E[p^r]$ are all trivial.
  3. The dual Frobenius isogeny $\hat{\varphi}$ is purely inseparable.
  4. The dual Frobenius isogenies $\hat{\varphi}_r$ are all purely inseparable.
  5. When F is finite, $\mathrm{tr}(\varphi) \equiv 0 \bmod p$ (where the trace is computed in $\mathrm{End}(T_l(E))$ for any $l \neq p$).
  6. The multiplication-by-p map $[p]$ is purely inseparable and $j(E) \in \mathbb{F}_{p^2}$.
  7. The endomorphism ring $\mathrm{End}(E)$ is an order in a quaternion algebra.
  8. The sum of Hasse invariants $\sum_p \mathrm{inv}_p E$ is equal to 1.
  9. The endomorphism algebra $\mathrm{End}(E) \otimes \mathbb{Q}$ is the unique (up to isomorphism) definite quaternion algebra ramified at p and ∞.
  10. When $p > 2$ and E has a Weierstrass equation $y^2 = f(x)$, the coefficient of x^{p-1} in $f(x)^{(p-1)/2}$ is zero.

  An elliptic curve satisfying any of these equivalent conditions is called supersingular.

  - Proof: (1) $\Leftrightarrow$ (2): Since $E[p]$ is the p-torsion subgroup of $E[p^r]$, one being trivial is equivalent to the other being trivial.
  - (1) $\Leftrightarrow$ (3): We have $\#E[p] = \#\ker[p] = \deg_s[p] = \deg_s \hat{\varphi}$, so $E[p]$ is trivial if and only if $\hat{\varphi}$ is purely inseparable.

- (3) $\Leftrightarrow$ (4): Since $\hat{\varphi}_r = \hat{\varphi}^r$ the inseparability of all of the $\hat{\varphi}_r$ is equivalent to the inseparability of $\hat{\varphi}$.
- (3) $\Leftrightarrow$ (5): Over a finite field during our discussion of the Weil conjectures we established that $\hat{\varphi} = [\mathrm{tr}\varphi] - \varphi$. So we have $\hat{\varphi}^*\omega = ([\mathrm{tr}\varphi] - \varphi)^*\omega = (\mathrm{tr}\varphi)\omega$, whence $\hat{\varphi}$ is inseparable precisely when $\mathrm{tr}(\varphi) \equiv 0 \bmod p$.
- (3) $\Rightarrow$ (6): We have $[p] = \hat{\varphi} \circ \varphi$ and φ is purely inseparable, so if $\hat{\varphi}$ is purely inseparable then so is $[p]$.
- Furthermore, by our results on inseparable isogenies, $\hat{\varphi} : E^{(p)} \to E$ factors as $\hat{\varphi} = \psi \circ \varphi'$ where $\varphi' : E^{(p)} \to E^{(p^2)}$ is the Frobenius map on $E^{(p)}$ and $\psi : E^{(p^2)} \to E$ is some other isogeny. But since $\deg \hat{\varphi} = p = \deg \varphi'$ we must have $\deg \psi = 1$ and so ψ is an isomorphism. This means E and $E^{(p^2)}$ are isomorphic, so they have the same j-invariant.
- But the j-invariant is a rational function of the coefficients of a Weierstrass form, hence $\varphi^2 j(E) = j(E^{(p^2)}) = j(E)$: thus φ^2 fixes $j(E)$ whence $j(E) \in \mathbb{F}_{p^2}$, the fixed field of φ^2.
- (6) $\Rightarrow$ (7): Suppose $[p]$ is inseparable and $j(E) \in \mathbb{F}_{q^2}$. Suppose $\psi : E \to E'$ is some isogeny: then since $\psi \circ [p]_E = [p]_{E'} \circ \psi$, taking inseparability degrees yields $\deg_i \psi \deg_i [p]_E = \deg_i [p]_{E'} \deg_i \psi$ so since $\deg_i \psi > 0$ we see $\deg_i [p]_{E'} = \deg_i [p]_E = p^2$ meaning $[p]$ is also purely inseparable on E', and so $\#E'[p] = \deg_s [p]_{E'} = 1$.
- Applying (1) $\Rightarrow$ (6) on E' shows that $j(E')$ is also an element of $\mathbb{F}_{q^2}$, and so by our results on the j-invariant, there are only finitely many possible E' up to isomorphism.
- Now suppose $\mathrm{End}(E)$ is not an order in a quaternion algebra, so that $K = \mathrm{End}(E) \otimes \mathbb{Q}$ is a number field (either $\mathbb{Q}$ or an imaginary quadratic field).
- Exercise: Show that if E and E' are isogenous then $\mathrm{End}(E) \otimes \mathbb{Q} \cong \mathrm{End}(E') \otimes \mathbb{Q}$. [Hint: Let $\varphi : E \to E'$ be an isogeny. Show that the map sending $f \in \mathrm{End}(E)$ to $\frac{1}{\deg \varphi} \varphi \circ f \circ \hat{\varphi} \in \mathrm{End}(E') \otimes \mathbb{Q}$ is an injective ring homomorphism.]
- As noted above there are only finitely many E' isogenous to E, and by the exercise above, each of their endomorphism rings is an order in $K = \mathrm{End}(E) \otimes \mathbb{Q}$. Now let $l \neq p$ be an inert prime[12] in K not dividing the conductor of any of the orders $\mathrm{End}(E')$ for any E' isogenous to E: then (l) remains prime in each ring $\mathrm{End}(E')$.
- Now choose some k larger than the number of possible isogenous curves E' and let P be a point of order l^k on E. Let Φ_i be the cyclic subgroup of order l^i generated by $[l^{k-i}]P$: then we have $\Phi_1 \subset \Phi_2 \subset \Phi_3 \subset \cdots \subset \Phi_{k-1} \subset \Phi_k$.
- By our results, each of the quotients E/Φ_i is an elliptic curve isogenous to E, and so by the pigeonhole principle some pair of them must be isogenous: say $E/\Phi_a \cong E/\Phi_{a+b}$. Since we also have a natural projection map $\pi : E/\Phi_{a+b} \to E/\Phi_a$ with cyclic kernel isomorphic to $\mathbb{Z}/l^b\mathbb{Z}$, composing the projection with the isomorphism yields an endomorphism λ of E_a whose kernel is also cyclic of order l^b.
- But now λ is a factor of $[l^k]$ since $\ker \lambda \subseteq \ker[l^k] = E[l^k]$, so $\lambda | l^k$ in $\mathrm{End}(E)$. Since l is a prime element of $\mathrm{End}(E')$ by the above discussion, this means $\lambda = ul^c$ for some exponent c and some unit $u \in \mathrm{End}(E)$ (i.e., an automorphism, which since it is invertible must have degree 1).
- Taking degrees shows that $l^b = \#\ker(\lambda) = \deg_s(\lambda) = \deg(\lambda) = \deg(u)\deg[l^c] = l^{2c}$, so $c = b/2$. But then we would have $\ker(\lambda) = \ker[l^c] = \ker[l^{b/2}]$ but this is a contradiction because the kernel of λ is cyclic of order l^b while the kernel of $[l^{b/2}]$ is a product of two cyclic groups of order $l^{b/2}$ and is not cyclic.
- $\neg(3) \Rightarrow \neg(7)$. Suppose $\hat{\varphi}$ is separable: we will show that $\mathrm{End}(E)$ is commutative. Since $\hat{\varphi}$ is separable, we know that $E[p^d] \cong \mathbb{Z}/p^d\mathbb{Z}$ and thus the Tate module $T_p(E) \cong \mathbb{Z}_p$.
- As we have already seen, any endomorphism of E naturally acts as an endomorphism of the Tate module, and this action must be injective as follows from our argument earlier on the Tate module $T_l(E)$: specifically, if $\psi \in \mathrm{End}(E)$ acts as zero on the Tate module, then $E[p^d] \in \ker \psi$ for all d. Then since $[p^d] = \varphi_d \circ \hat{\varphi}_d$ and φ_d is onto, we see that $\varphi_d(\ker \psi) \supseteq \ker \hat{\varphi}_d$ so $\#\ker \psi \geq \#\ker \hat{\varphi}_d = \deg_s \hat{\varphi}_d = (\deg_s \hat{\varphi})^d = p^d$ since $\hat{\varphi}$ is separable of degree d. Then $\#\ker \psi$ is infinite so $\psi = 0$.

[12] There are three types of behavior for a prime l in a quadratic extension $K/\mathbb{Q}$: the ideal (l) can remain prime (l is inert), the ideal $(l) = L_1 L_2$ can factor as the product of two distinct prime ideals (l splits), or the ideal $(l) = L^2$ can factor as the square of a prime ideal (l ramifies). If the discriminant of the field is Δ, then the splitting behavior is determined by the Legendre symbol $\left(\frac{\Delta}{l}\right)$: when the symbol is $+1$ the prime l splits, when the symbol is 0 the prime l ramifies, and when the symbol is -1 the prime l is inert. As such, the ramified primes are those dividing the discriminant, and by Dirichlet's theorem on primes in arithmetic progressions, there are infinitely many split primes and infinitely many inert primes (more precisely, asymptotically half of primes exhibit each behavior).

- But now the endomorphism ring $\mathrm{End}(E)$ injects into $\mathrm{Aut}(T_p(E)) \cong \mathrm{Aut}(\mathbb{Z}_p) \cong \mathbb{Z}_p$, which is commutative, so $\mathrm{End}(E)$ is commutative.
- (7) $\Rightarrow$ (8), (9): Let $l \neq p$ be a prime and let $A = \mathrm{End}(E) \otimes \mathbb{Q}$. Then by the same argument as given above in the characteristic-zero case, $A \otimes \mathbb{Q}_l \cong M_{2\times 2}(\mathbb{Q}_l)$ and so $\mathrm{inv}_l(A) = 0$ for all finite primes $l \neq p$. However, since the element $[A]$ in the Brauer group $\mathrm{Br}(\mathbb{Q})$ has order $\sqrt{[A:\mathbb{Q}]} = 2$, $[A] \neq 1$ and thus A must be ramified at at least 2 places.
- Since the only remaining places are p and ∞, A must be ramified at both: thus the Hasse invariant $\sum_p \mathrm{inv}_p(E) = 1$, and by the injectivity of $\mathrm{Br}(\mathbb{Q}) \to \bigoplus_p \mathrm{Br}(\mathbb{Q}_p)$, this information uniquely determines the class $[A]$ in $\mathrm{Br}(\mathbb{Q})$. But since $\dim_{\mathbb{Q}} A = 4$, the class $[A]$ characterizes A itself up to isomorphism.
- (8) $\Rightarrow$ (7), (9) $\Rightarrow$ (7): Obvious. (The sum $\sum_p \mathrm{inv}_p[A]$ is zero when A is a field.)
- (5) $\Leftrightarrow$ (10): Since (2) implies (3) we know that $j(E) \in \mathbb{F}_{p^2}$ and so E is defined over a finite field $F = \mathbb{F}_q$. Now let χ be the nontrivial quadratic character on $\mathbb{F}_q$ (+1 on unit squares, 0 on zero, -1 on nonsquares): as we have previously noted, $\#E(\mathbb{F}_q) = q + 1 + \sum_{x\in\mathbb{F}_q} \chi(f(x))$.
- Exercise: When q is odd, for any $a \in \mathbb{F}_q$ show that $\chi(a) = a^{(q-1)/2}$. [Hint: $\mathbb{F}_q^\times$ is cyclic.]
- Exercise: For a positive integer k, show that $\sum_{x\in\mathbb{F}_q} x^k$ is 1 when $(q-1)|k$ and is 0 when $(q-1) \nmid k$.
- Let $f(x)^{(q-1)/2} = \sum_{k=0}^{3(q-1)/2} c_k x^k$. By the first exercise, we have $\#E(\mathbb{F}_q) = q + 1 + \sum_{x\in\mathbb{F}_q} f(x)^{(q-1)/2} = q + 1 + \sum_{k=0}^{3(q-1)/2} \sum_{x\in\mathbb{F}_q} c_k x^k = q + 1 + \sum_{k=0}^{3(q-1)/2} c_k (\sum_{x\in\mathbb{F}_q} x^k) = q + 1 - c_{q-1}$ in $\mathbb{F}_q$, where the last step follows from the second exercise.
- But as we have previously shown, $\#E(\mathbb{F}_q) = q + 1 - \mathrm{tr}(\varphi)$, so $\mathrm{tr}(\varphi) = c_{q-1}$ as an element of $\mathbb{F}_q$. Hence one is zero modulo p if and only if the other is zero modulo p.

0.20 (Nov 16) Elliptic Curves over $\mathbb{C}$

- Exercise: Show that $y^2 = x^3 + x + 1$ is supersingular over $\mathbb{F}_{17}$ by computing both $\#E(\mathbb{F}_{17})$ and the coefficient of x^{16} in $(x^3 + x + 1)^8$ mod 17.
- Exercise: Show that for an odd prime p, $y^2 = x^3 + x$ is supersingular over $\mathbb{F}_p$ if and only if $p \equiv 3 \pmod 4$.
- Exercise: Show for a prime $p > 3$, an elliptic curve $E/\mathbb{F}_p$ is supersingular if and only if $\#E(\mathbb{F}_p) = p + 1$. Deduce that the pth-power Frobenius map φ has $\varphi^2 = [-p]$ and that $\hat{\varphi} = -\varphi$.
- Exercise: Show that the elliptic curve $y^2 = x(x-1)(x-\lambda)$ in Legendre form is supersingular over $\mathbb{F}_q$ if and only if λ is a root of the polynomial $H_p(t) = \sum_{k=0}^{(p-1)/2} \binom{(p-1)/2}{k}^2 t^k$. [Remark: One may show that H_p is separable. By using some basic facts about equivalences of Legendre forms, one may give a precise count of the number of supersingular curves over $\mathbb{F}_p$.]
- We can now describe the endomorphism ring structure over arbitrary fields of positive characteristic:
- Theorem (Endomorphism Rings in Positive Characteristic): Let F be a field of positive characteristic p and let E be an elliptic curve defined over F. Then

1. $\mathrm{End}(E) \cong \mathbb{Z}$ precisely when $j(E)$ is transcendental over $\mathbb{F}_p$.
2. $\mathrm{End}(E)$ is an order in an imaginary quadratic field precisely when $j(E)$ is algebraic over $\mathbb{F}_p$ and $[p]$ is not purely inseparable.
3. $\mathrm{End}(E)$ is an order in a definite quaternion algebra precisely when $j(E) \in \mathbb{F}_{p^2}$ and $[p]$ is purely inseparable.
    - Proof: We have already shown (3) in our discussion of supersingular curves earlier. It remains to show that when E is not supersingular, then $\mathrm{End}(E)$ contains some element other than a multiplication-by-m map if and only $j(E)$ is algebraic over $\mathbb{F}_p$.
    - So, first suppose $j(E)$ is algebraic over $\mathbb{F}_p$, so that $j(E) \in \mathbb{F}_{p^d}$ for some $d \geq 1$. By our results on j-invariants this means E is isomorphic to a curve defined over $\mathbb{F}_{p^d}$; since the endomorphism ring is invariant under isomorphism classes, we may therefore replace E with this curve without changing anything.

- Then since E is defined over $\mathbb{F}_{p^r}$, the p^dth-power Frobenius map φ is an endomorphism of E (it fixes all of the coefficients). We claim that φ is not a multiplication-by-m map for any m.
- If we had $\varphi = [m]$ then taking degrees yields $p^d = \deg\varphi = \deg[m] = m^2$ so that $m = \pm p^{d/2}$. But then taking kernels yields $1 = \#\ker\varphi = \#\ker[\pm p^{d/2}] = p^{d/2}$ since by assumption E is not supersingular so its $p^{d/2}$-torsion subgroup is cyclic of order $p^{d/2}$. This is a contradiction since it would give $d = 0$.
- Hence $\mathrm{End}(E)$ is strictly larger than $\mathbb{Z}$, so it must be an order in an imaginary quadratic field.
- Now suppose $j(E)$ is transcendental over $\mathbb{F}_p$ and suppose by way of contradiction that $\mathrm{End}(E)$ is strictly larger than $\mathbb{Z}$: then $K = \mathrm{End}(E)\otimes\mathbb{Q}$ is an imaginary quadratic field.
- Let $l \neq p$ be any prime not dividing N and let $\Phi_1 \subset \Phi_2 \subset \cdots$ be a chain of subgroups with Φ_i cyclic of order l^i (e.g., generated by the terms $(P_1, P_2, \dots)$ of a generator of $T_l(E)$).
- Then $E_i = E/\Phi_i$ is also an elliptic curve and its j-invariant is also transcendental over $\mathbb{F}_p$. Furthermore, we have an associated isogeny $\varphi : E \to E_i$ whose degree is l^i, and by an earlier exercise we have $\mathrm{End}(E_i)\otimes\mathbb{Q} = K$ as well.
- Now let $\alpha_i \in \mathrm{End}(E_i)\otimes\mathbb{Q}$ have $\alpha_i^2 \in \mathbb{Q}$ with $\alpha_i^2 < 0$; by rescaling we may assume $\alpha_i^2 = -N_i$ for some positive integer N_i, and by rescaling further if necessary we may assume $\alpha_i \in \mathrm{End}(E_i)$ with $\alpha_i^2 = -N_i$.
- Then in $\mathrm{End}(E)$ we may observe that $(\hat{\varphi}_i\alpha_i\varphi_i)^2 = \hat{\varphi}_i\alpha_i^2\varphi_i = \alpha_i^2(\deg\varphi_i)^2$, so taking the square root yields $\hat{\varphi}_i\alpha_i\varphi_i = \pm\alpha l^i$, and left-composing with φ_i and cancelling yields $\alpha_i\circ\varphi_i = \pm\varphi_i\circ\alpha$.
- In particular, this means $\alpha(\ker\varphi_i) \subseteq \ker\varphi_i$, so that $\alpha\Phi_i \subseteq \Phi_i$. This holds for all i, so upon taking inverse limits we see that α acts as scalar multiplication on the Tate module $T_l(E)$, meaning that $\alpha(P_1, P_2, \dots) = c(P_1, P_2, \dots)$ for some $c \in \mathbb{Z}_l$.
- But the characteristic polynomial of α on the Tate module has integer coefficients (it is an algebraic integer), so $c \in \mathbb{Q}$ is actually rational. But this is a contradiction, since then $\alpha^2 = c^2$ would be positive, contradiction.

- We now shift focus in our discussion to study elliptic curves over the complex numbers. There are several different threads that will all converge in this discussion, so we will start with the historical motivation in analysis for studying elliptic curves.
  - Consider the problem of calculating the arclength of the ellipse $x^2/a^2 + y^2/b^2 = 1$. Using the natural parametrization $x = a\cos t$, $y = b\sin t$ we see that the arclength equals $s = 4\int_0^{\pi/2}\sqrt{a^2\sin^2 t + b^2\cos^2 t}\,dt = 4b\int_0^{\pi/2}\sqrt{1-k^2\sin^2 t}\,dt$ with $k^2 = 1 - a^2/b^2$.
  - Substituting $x = \sin t$ with $dx = \cos t\,dt$ yields $s = 4b\int_0^1\sqrt{\dfrac{1-k^2x^2}{1-x^2}}\,dx = 4b\int_0^1\dfrac{y\,dx}{1-x^2}$ where $y^2 = (1-x^2)(1-k^2x^2)$.
  - We therefore see that the ellipse arclength is obtained by integrating the differential $\omega = \dfrac{y\,dx}{1-x^2}$ on the curve $y^2 = (1-x^2)(1-k^2x^2)$.
  - In fact, this curve has genus 1, as can most efficiently be seen by verifying that the differential dx/y is holomorphic and nonvanishing, so since it obviously has rational points, it is an elliptic curve.
  - Explicitly, by substituting $x' = a\dfrac{x-1}{x+1}$ and $y' = \dfrac{ey}{(x+1)^2}$ with $a = \dfrac{k+1}{k-1}$ and $e = \dfrac{2i(k+1)}{(k-1)^2}$, one may eventually verify[13] that $y^2 = (1-x^2)(1-k^2x^2)$ becomes the elliptic curve $(y')^2 = x'(x'-1)(x'-\lambda)$ in Legendre form with $\lambda = \dfrac{(k+1)^2}{(k-1)^2}$.
  - Therefore, calculating the arclength of an ellipse (after some amount of torment) eventually becomes a problem of computing an integral of the form $\int_C f(x)\omega$ on an elliptic curve, where ω is the invariant differential and C is some contour. (Since ω is a basis for the space of differentials, all integrals on E are of this form.)
  - Indeed, the fact that computing the arclength of an ellipse eventually leads to computation of an integral on an elliptic curve is the historical reason that elliptic curves are so named!

[13] To find this actual change of coordinates, we need only calculate explicitly the functions used in the Riemann-Roch argument for showing elliptic curves have Weierstrass equations.

- So let us now discuss the general problem of integrating differentials on elliptic curves.
  - The most basic integral of this form would be the integral of the invariant differential $\int_C \omega = \int_C \frac{dx}{y} = \int_C \frac{dx}{\sqrt{x(x-1)(x-\lambda)}}$ when written as an integral in the single variable $x \in \mathbb{C}$.
  - However, an obvious difficulty arises: namely, that the square root function is not single-valued but rather double-valued, so as an explicit contour integral in the complex plane, the integral is not well defined until we choose a specific branch cut of the square root.
  - In general, the square root function $\sqrt{x}$ changes its value by a factor of -1 along a continuous path that circulates with winding number 1 around the singular point $x = 0$.
  - Therefore, the product $\sqrt{x(x-1)}$ can be made single-valued if we make a branch cut from $x = 0$ to $x = 1$ (i.e., meaning that we do not allow any of our contours to cross the branch cut), since then any path that circulates once around the branch cut will enclose both singular points $x = 0$ and $x = 1$, and so the value of the product will be scaled by $(-1)^2 = 1$, which is to say, it will not change.
  - In a similar way, we may make the function $\sqrt{x-\lambda}$ single-valued by making a branch cut from $x = \lambda$ to $x = \infty$ (since the square root function is also singular there).
  - More properly, what we are actually doing is making branch cuts from 0 to 1, and from λ to ∞ in two copies of $\mathbb{P}^1(\mathbb{C})$, and then gluing them together along the branch cuts to form a Riemann surface. (The two different copies correspond to the two possible choices of sign in the square root.)
  - Exercise: Show that the surface obtained by gluing together two spheres along two branch cuts is topologically a torus.
  - Geometrically, we can keep track of which copy of $\mathbb{P}^1(\mathbb{C})$ we are on by introducing another variable y, whose square is equal to the product $x(x-1)(x-\lambda)$: then the sign of y keeps track of the correct location in the Riemann surface.
  - Of course, this is just a convoluted way of saying that the resulting Riemann surface is simply the elliptic curve $E : y^2 = x(x-1)(x-\lambda)$, thought of as a 2-dimensional surface over $\mathbb{R}$ rather than a 1-dimensional curve over $\mathbb{C}$. (This is also exceedingly reasonable since the underlying Riemann surface is a torus, which has genus 1.)
  - So a far more sensible way to do all of this is to view $\int_C \omega$ as an integral where C is a path in $E(\mathbb{C})$.
- We may now try to construct a map from E to $\mathbb{C}$ by sending a point P to the integral $\int_C \omega$, where C is any contour starting at the origin O and ending at P.
  - By standard results in complex analysis, the value of a contour integral is deformation-invariant, meaning that a continuous deformation of the contour does not change the value of the integral (as long as the starting and ending points are the same).
  - However, because of the branch cuts, the value $\int_O^P \omega$ is not uniquely determined, since there are inequivalent paths from O to P (namely, paths winding different numbers of times around the branch cuts).
  - Let α be a path looping around the r_1-r_2 branch cut once, and let β be a path looping around the $r_3 - \infty$ branch cut once. Since α and β generate the first homology group $H_1(T)$ of the torus T, the difference between any two paths between 0 and P on the Riemann sphere $\mathbb{P}^1(\mathbb{C})$ with branch cuts is homotopic to a linear combination of α and β.
  - This means the integral $\int_0^P \omega$ is well-defined up to adding a $\mathbb{Z}$-linear combination of the periods $\omega_1 = \int_\alpha \omega$ and $\omega_2 = \int_\beta \omega$, which is to say, we obtain a well-defined map $E(\mathbb{C}) \to \mathbb{C}/\Lambda$ via $P \mapsto \int_0^P \omega \pmod{\Lambda}$, where $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$.
  - In fact, by a somewhat involved Green's theorem calculation, we can show that the periods are $\mathbb{R}$-linearly independent. Explicitly, taking a particular branch cut of the square root defining y allows us to write $\omega_1 = 2\int_{r_1}^{r_2} \frac{dx}{y}$ and $\omega_2 = 2\int_{r_3}^{\infty} \frac{dx}{y}$.

- Writing $x = u + iv$, we have $\frac{i}{2}\iint_{\mathbb{C}} \omega \wedge \overline{\omega} = \dfrac{(du + i dv) \wedge (du - i\, dv)}{|y^2|} = \iint_{\mathbb{C}} \dfrac{1}{|4x^3 - Ax - B|}\, du\, dv$. This integral converges since it is integrable near the three poles (as can be seen by using polar coordinates), and its value is clearly positive.
- But by applying Green's theorem to the contour bounded by paths from $r_1 \to r_2 \to r_3 \to \infty \to r_1$, and decomposing ω into real and imaginary parts and computing their various signs in the regions around the branch cuts, one may show that $\frac{i}{2}\iint_{\mathbb{C}} \omega \wedge \overline{\omega}$ also equals $\mathrm{Re}(\omega_1)\mathrm{Im}(\omega_2) - \mathrm{Im}(\omega_1)\mathrm{Re}(\omega_2)$. Since this quantity is positive, ω_1 and ω_2 are $\mathbb{R}$-linearly independent.
- So: the two periods are always $\mathbb{R}$-linearly independent, and so the subgroup Λ is discrete, hence since it has rank 2 inside the 2-dimensional real vector space $\mathbb{C}$, it is a lattice.
- The quotient space $\mathbb{C}/\Lambda$ as a group is then isomorphic (and homeomorphic) to $(\mathbb{R}/\mathbb{Z}) \times (\mathbb{R}/\mathbb{Z}) \cong S^1 \times S^1$, which is topologically a torus.
- Indeed, it is easy to see this directly, since the lattice Λ has a fundamental parallelogram with vertices 0, ω_1, ω_2, $\omega_1 + \omega_2$: then the quotient space $\mathbb{C}/\Lambda$ consists of this parallelogram with opposite edges identified with the same orientations, which is a torus.
- Since $E(\mathbb{C})$ and $\mathbb{C}/\Lambda$ are both complex tori and we have an analytic map from one to the other (namely, given by integrating the invariant differential of E), it is reasonable to expect that the map is a complex analytic isomorphism.
- In fact this is true (as we will show): to give a taste of how this works, note that because ω is translation-invariant, we have $\int_0^{P+Q} \omega = \int_0^P \omega + \int_P^{P+Q} \omega = \int_0^P \omega + \int_0^Q \omega$, so the integral map is a group isomorphism.

- Unfortunately, calculating the integrals directly, using a Weierstrass equation of E, is difficult to do for a generic curve.
  - Given a specific equation, of course, we could simply compute numerical approximations of the periods using numerical integration procedures. If we are lucky, we may even find that the period integrals for certain curves can be evaluated exactly.
  - But a general kind of calculation is rather beyond our reach.
  - So what we will do instead is approach this problem from the other side: namely, starting with a lattice $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$, construct the corresponding elliptic curve E whose periods are ω_1 and ω_2. (This is quite a reasonable thing to do because lattices in $\mathbb{C}$ are much easier to understand!)
  - Now, under the assumption that $E(\mathbb{C})$ and $\mathbb{C}/\Lambda$ are complex-analytically isomorphic, the coordinate functions x and y for E will correspond to well-defined meromorphic functions on $\mathbb{C}/\Lambda$, which is to say, meromorphic functions on $\mathbb{C}$ whose values are independent of the specific representative in $\mathbb{C}/\Lambda$ where they are evaluated.
  - We therefore want to study functions with this property.

0.21 (Nov 20) Elliptic Functions, The Weierstrass $\wp$-Function

- Definition: Let $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ be a lattice in $\mathbb{C}$. An elliptic function relative to Λ is a meromorphic function on $\mathbb{C}$ that satisfies $f(z + \omega) = f(z)$ for all $\omega \in \Lambda$ and $z \in \mathbb{C}$. The set of all elliptic functions relative to Λ is denoted $\mathbb{C}(\Lambda)$.
  - Remark: When the lattice Λ is clear from context, or not relevant, we will simply say "elliptic function" without explicitly saying "relative to Λ".
  - Elliptic functions are also commonly called doubly-periodic functions since the general condition above is equivalent to $f(z + \omega_1) = f(z + \omega_2) = f(z)$: in other words, saying that f has two different periods ω_1 and ω_2.
  - Obviously, constant functions are elliptic functions. Keeping in mind the general principle that elliptic functions will correspond to rational functions on the associated elliptic curve E, we should expect it to be somewhat challenging to construct elliptic functions, since most functions on E will not be rational. We will therefore study general properties of elliptic functions first, and then use the results to give constructions of elliptic functions.

- As with any meromorphic function, we may fruitfully discuss the order of vanishing, zeroes, poles, and residues of an elliptic function.
- Explicitly, if f is a nonzero elliptic function on $\mathbb{C}$, then for any $z_0 \in \mathbb{C}$ we have a local Laurent expansion $f(z) = \sum_{n=k}^{\infty} a_n (z - z_0)^n$ at z_0, where we assume the leading coefficient $a_k \neq 0$ (note that when $k \geq 0$ this is a familiar power series, while when $k < 0$ this is a Laurent series).
- For this Laurent expansion, the order of vanishing of f at z_0, denoted $\mathrm{ord}_{z_0}(f)$, is the value k. We say that f has a pole of order $|k|$ at z_0 when $k < 0$ and a zero of order k at z_0 when $k > 0$.
- The residue of f at z_0, denoted $\mathrm{res}_{z_0}(f)$, is the coefficient a_{-1}. Note that the residue can be nonzero only when f has a pole at z_0.

- Let us now collect some basic facts about elliptic functions:

- Proposition (Properties of Elliptic Functions): Let $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ be a lattice in $\mathbb{C}$, let $\mathbb{C}(\Lambda)$ denote the field of elliptic functions with respect to Λ, and let D be a fundamental region for $\mathbb{C}/\Lambda$ (e.g., the parallelogram with vertices 0, ω_1, ω_2, $\omega_1 + \omega_2$ or some $\mathbb{C}$-translate of it). Then the following hold:

  1. A nonzero elliptic function $f \in \mathbb{C}(\Lambda)$ has finitely many zeroes and poles inside of D.
     - Note that this is the analogue of the statement that a nonzero rational function in $k(C)$ has only finitely many zeroes and poles.
     - Proof: Since the fundamental parallelogram D is compact, if f had infinitely many poles they would have an accumulation point, but poles of a meromorphic function are discrete. Hence f has only finitely many poles.
     - Applying the argument to $1/f$ shows that f also has finitely many zeroes, so f has finitely many zeroes and poles.
  2. An elliptic function with no zeroes, or no poles, is constant.
     - Note that this is the analogue of the statement that a rational function in $k(C)$ with no zeroes or no poles is constant.
     - Proof: If f has no poles then f is holomorphic on all of $\mathbb{C}$ (i.e., f is an entire function).
     - Since $\mathbb{C}/\Lambda$ is compact and f is continuous, f is bounded on D, hence on all of $\mathbb{C}$ because f is doubly periodic. But then f is an entire function that is bounded, so by Liouville's theorem, f is constant.
     - If f has no zeroes, then applying the same argument to $1/f$ shows that $1/f$ hence f is constant.
  3. For any $f \in \mathbb{C}(\Lambda)$, we have $\sum_{w \in \mathbb{C}/\Lambda} \mathrm{res}_w(f) = 0$, where the sum is evaluated over any fundamental region D.
     - Note that the sum of residues is well defined by (1), since f has only finitely many poles hence finitely many nonzero residues.
     - Proof: Choose any fundamental region D whose boundary contains no zeroes or poles of f: this is possible since there are only finitely many zeroes and poles by (1), but there are uncountably many inequivalent translations to select for D.
     - Consider the integral $\int_{\partial D} f(z)\, dz$: since f takes the same values on parallel edges of ∂D, the contributions to the integral on opposite sides cancel since they have opposite orientations, so the integral is zero.
     - Then Cauchy's residue theorem immediately yields $\sum_{w \in \mathbb{C}/\Lambda} \mathrm{res}_w(f) = \dfrac{1}{2\pi i} \int_{\partial D} f(z)\, dz = 0$.
  4. For any $f \in \mathbb{C}(\Lambda)$, we have $\sum_{w \in \mathbb{C}/\Lambda} \mathrm{ord}_w(f) = 0$, where the sum is evaluated over any fundamental region D.
     - This result says that f has the same number of zeroes and poles, counted with multiplicity: it is the analogue of the result that $\deg(\mathrm{div}\, f) = 0$ for any nonzero $f \in k(C)$.
     - Proof: As in (3), choose any fundamental region D whose boundary contains no zeroes or poles of f.
     - Since f is doubly-periodic so is its derivative f' hence so too is the ratio f'/f.
     - If the Laurent series for f at w is $a_k(z-w)^k + \cdots$, then the Laurent series for f' is $ka_k(z-w)^{k-1} + \cdots$ and so the Laurent series for the ratio f'/f is $k(z-w)^{-1} + \cdots$, and so $\mathrm{res}_w(f'/f) = k = \mathrm{ord}_w(f)$.

- As in (3), the integral $\int_{\partial D} \frac{f'(z)}{f(z)}\,dz$ is zero, so then Cauchy's residue theorem yields $\sum_{w\in\mathbb{C}/\Lambda} \mathrm{ord}_w(f) = \sum_{w\in\mathbb{C}/\Lambda} \mathrm{res}_w(f'/f) = \frac{1}{2\pi i}\int_{\partial D} \frac{f'(z)}{f(z)}\,dz = 0$.

5. For any $f \in \mathbb{C}(\Lambda)$, we have $\sum_{w\in\mathbb{C}/\Lambda} \mathrm{ord}_w(f)w \in \Lambda$, where the sum is evaluated over any fundamental region D.
   - Note that choosing a different fundamental region D will potentially shift points w in the sum by an element of Λ, so unlike the sums in (3) and (4) which are independent of the choice of D, this sum is only well-defined modulo Λ.
   - Proof: As in (4) we choose a fundamental region D whose boundary contains no zeroes or poles of f: say with vertices a, $a+\omega_1$, $a+\omega_1+\omega_2$, $a+\omega_2$ in counterclockwise order.
   - By Cauchy's residue theorem we have $\sum_{w\in\mathbb{C}/\Lambda} \mathrm{ord}_w(f)w = \sum_{w\in\mathbb{C}/\Lambda} \mathrm{res}_w(zf'/f) = \frac{1}{2\pi i}\int_{\partial D} z\frac{f'(z)}{f(z)}\,dz$.
   - Decomposing the integral into components along the four sides of D, and then applying ellipticity of f'/f yields

$$\begin{aligned}\int_{\partial D} z\frac{f'(z)}{f(z)}\,dz &= \int_a^{a+\omega_1} z\frac{f'(z)}{f(z)}\,dz + \int_{a+\omega_1}^{a+\omega_1+\omega_2} z\frac{f'(z)}{f(z)}\,dz + \int_{a+\omega_1+\omega_2}^{a+\omega_2} z\frac{f'(z)}{f(z)}\,dz + \int_{a+\omega_2}^{a} z\frac{f'(z)}{f(z)}\,dz \\ &= \int_a^{a+\omega_1} z\frac{f'(z)}{f(z)}\,dz + \int_a^{a+\omega_2} (z+\omega_1)\frac{f'(z)}{f(z)}\,dz - \int_a^{a+\omega_1} (z+\omega_2)\frac{f'(z)}{f(z)}\,dz - \int_a^{a+\omega_2} z\frac{f'(z)}{f(z)}\,dz \\ &= -\omega_2\int_a^{a+\omega_1} \frac{f'(z)}{f(z)}\,dz + \omega_1\int_a^{a+\omega_2} \frac{f'(z)}{f(z)}\,dz\end{aligned}$$

   - But now since f'/f is elliptic, we have $(f'/f)(a) = (f'/f)(a+\omega_1)$, so $\int_a^{a+\omega_j} \frac{f'(z)}{f(z)}\,dz$ equals $2\pi i$ times the winding number $W_{\gamma_j}(0)$ around 0 of the curve $\gamma_j : [0,1] \to \mathbb{C}$ with $\gamma(t) = f(a+t\omega_j)$.
   - Hence we obtain $\sum_{w\in\mathbb{C}/\Lambda} \mathrm{ord}_w(f)w = \frac{1}{2\pi i}\int_{\partial D} z\frac{f'(z)}{f(z)}\,dz = -\omega_2 W_{\gamma_1}(0) + \omega_1 W_{\gamma_2}(0)$, which is an element of Λ because the winding numbers are both integers.

6. An elliptic function with at most one pole, with pole order at most 1 there, is constant.
   - Proof: Suppose f were elliptic and had a single simple pole. Then by (3), since the sum of the residues of f is 0, the residue at that pole would be zero, but then f would be holomorphic hence constant.

- So far we have established some properties of elliptic functions without actually describing any such functions aside from constants. Let us use these properties to (try to) give a construction of an elliptic function.
  - From (2) we know that any nonconstant elliptic function must have at least one pole, and from (6) we see that the total pole order must be at least 2.
  - Taking motivation from the x-coordinate function on an elliptic curve (which has one pole, of order 2, at ∞), let us try to construct an elliptic function $f(z)$ with a double pole.
  - By translation we may place this pole anywhere, so let us put it at 0. Then the Laurent expansion of $f(z)$ at $z=0$ is $c_{-2}z^{-2} + O(z^{-1})$ for some $c \neq 0$, and so by rescaling we may assume $c_{-2} = 1$.
  - Now, by (3), since f has only one pole (up to periodicity), the residue at that pole must be zero, so the z^{-1} coefficient in the Laurent expansion at $z=0$ must be zero.
  - So in fact, the Laurent expansion for $f(z)$ is of the form $f(z) = z^{-2} + c_0 + c_1 z + c_2 z^2 + \cdots$ for some power series $c_0 + c_1 z + c_2 z^2 + \cdots$ that is necessarily holomorphic in a neighborhood of 0. In other words, the difference $f(z) - z^{-2}$ is holomorphic near 0.
  - But $f(z)$ is actually doubly periodic, so $f(z)$ also has a double pole at each point $a\omega_1 + b\omega_2$ of the lattice Λ; by the above argument, we see that $f(z) - (z-\omega)^{-2}$ will be holomorphic near an arbitrary $\omega \in \Lambda$.
  - So now, we ask: what happens if we subtract all of these "pole contributions" $(z-\omega)^{-2}$ for all $\omega \in \Lambda$ from $f(z)$? The resulting function would then have no poles at all, hence be entire, hence (under the assumption it is elliptic) constant. By shifting so that this constant is zero, we would obtain a formula for $f(z)$: namely, $f(z) = \sum_{\omega\in\Lambda} (z-\omega)^{-2}$.

- We now turn around and try to use this series as our construction of an elliptic function: namely, $f(z) = \sum_{\omega\in\Lambda}(z-\omega)^{-2}$ where the sum ranges over all elements $\omega = a\omega_1 + b\omega_2 \in \Lambda$.
  - Unfortunately, this construction does not quite work, for a critical reason: the series $\sum_{\omega\in\Lambda}(z-\omega)^{-2}$ does not converge absolutely!
  - Exercise: Let $\omega = a\omega_1 + b\omega_2$. Show that $|\omega|^2 = xa^2 + yab + zb^2$ is a positive-definite quadratic form in (a,b), where $x = |\omega_1|^2$, $y = 2\mathrm{Re}(\omega_1\overline{\omega_2})$, $z = |\omega_2|^2$.
  - Exercise: Show that if $Q(a,b)$ is a positive-definite real quadratic form, then $\sum_{(0,0)\neq(a,b)\in\mathbb{Z}\times\mathbb{Z}} \frac{1}{Q(a,b)^k}$ diverges for $k \leq 1$ and converges absolutely for $k > 1$. [Hint: Compare to the corresponding integral, diagonalize the quadratic form, and use polar coordinates.]
  - Exercise: Let $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ be a lattice. Show that $\sum_{0\neq\omega\in\Lambda} |\omega|^{-k}$ diverges for $k \leq 2$ and converges absolutely for $k > 2$.
  - Now, letting $\sum_{\omega\in\Lambda*}$ denote a sum over nonzero elements in Λ, for z bounded (e.g., in a fundamental region) the absolute value series is $\sum_{\omega\in\Lambda} |z-\omega|^{-2} = \frac{1}{z^2} + \sum_{\omega\in\Lambda*} \left|\frac{1}{\omega^2} + \frac{2z}{\omega^3} + \frac{3z^2}{\omega^4} + \cdots\right| = \sum_{\omega\in\Lambda*} \left|\omega^{-2} + O(\omega^{-3})\right|$ is on the order of $\sum_{\omega\in\Lambda*} |\omega|^{-2}$ which diverges by the exercises above.
  - So all of this says that our attempted construction does not work because the associated series does not converge. But notice that it barely fails to converge: indeed, if we were able to get rid of the ω^{-2} term, then the remaining series would be $\sum_{\omega\in\Lambda,\omega\neq0} \left|\frac{2z}{\omega^3} + \frac{3z^2}{\omega^4} + \cdots\right| = \sum_{\omega\in\Lambda,\omega\neq0} 2z\,|\omega|^{-3}$, which does converge absolutely.
  - In fact, it is not at all hard to remove that term: simply subtract ω^{-2} from each term of the series where $\omega \neq 0$.
  - So our new construction is the function $f(z) = \frac{1}{z^2} + \sum_{\omega\in\Lambda*}\left[\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2}\right]$. By the calculations above, this series does converge absolutely and uniformly on compact subsets of $\mathbb{C}$ to a meromorphic function having a double pole at each element of Λ.
- The function we have just constructed is called the Weierstrass $\wp$-function. Since the convergence of the various series $\sum_{\omega\in\Lambda^*}\omega^{-k}$ will also be important, we also define them now:
- Definition: Let ω_1, ω_2 are $\mathbb{R}$-linearly independent complex numbers and $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ be the associated complex lattice. The Weierstrass $\wp$-function (with respect to Λ) is defined to be $\wp(z;\Lambda) = \frac{1}{z^2} + \sum_{\omega\in\Lambda*}\left[\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2}\right]$, and the Eisenstein series of weight $2k$ (with respect to Λ) is $G_{2k}(\Lambda) = \sum_{\omega\in\Lambda*}\frac{1}{\omega^{2k}}$ where the sums are over all nonzero $\omega \in \Lambda$.
  - When Λ is clear from context, we will just write $\wp(z)$ in place of $\wp(z;\Lambda)$ and G_{2k} in place of $G_{2k}(\Lambda)$.
  - We index the Eisenstein series as G_{2k} because the odd-indexed sums $\sum_{\omega\in\Lambda^*}\frac{1}{\omega^{2k+1}}$ are all zero, as follows trivially by substituting $\omega \mapsto -\omega$.
- Let us now establish the key properties of this function:
- Theorem (Properties of the $\wp$-Function): Let $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ be a complex lattice with associated Weierstrass $\wp$-function $\wp(z;\Lambda) = \frac{1}{z^2} + \sum_{\omega\in\Lambda*}\left[\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2}\right]$ and Eisenstein series $G_{2k}(\Lambda) = \sum_{\omega\in\Lambda*}\frac{1}{\omega^{2k}}$. Then the following hold:

  1. The Eisenstein series $G_{2k}(\Lambda)$ is absolutely convergent for $k > 1$ but not for $k \leq 1$.
     - This result follows from the discussion above, but we will give a separate self-contained argument.

- Proof: By standard geometric results about lattices[14], if the fundamental parallelogram for Λ has area Δ, then the number of $\omega \in \Lambda$ with $|\omega| \leq R$ is $\frac{\pi}{\Delta}R^2 + O(R)$ as $R \to \infty$.
- Then for arbitrary R and sufficiently large d, the number n_R of $\omega \in \Lambda$ with $R \leq |\omega| < R+d$ is $\Theta(R)$.
- Hence by grouping ω together into the annuli $R \leq |\omega| < R + d$, by the comparison test we see that $\sum_{\omega \in \Lambda^*} |\omega|^{-2k}$ has the same behavior as the series $\sum_{R=1}^{\infty} \frac{\#\{\omega \in \Lambda : Rd \leq |\omega| < Rd + d\}}{(Rd)^k} \sim \sum_{R=1}^{\infty} \frac{R}{R^{2k}}$ which as a p-series is convergent for $k > 1$ and divergent for $k \leq 1$.

2. The series defining $\wp(z)$ converges absolutely and uniformly on compact subsets of $\mathbb{C}\backslash\Lambda$.
   - Proof: For $|\omega| > 2|z|$, we have $\left|\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2}\right| = \frac{|z|\,|2\omega - z|}{|\omega|^2\,|\omega - z|^2} \leq \frac{10|z|}{|\omega|^3}$.
   - Hence the tail of the series $\frac{1}{z^2} + \sum_{\omega \in \Lambda*} \left[\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2}\right]$ with $|\omega| > 2|z|$ is bounded in absolute value by $\sum_{\omega \in \Lambda^*} \frac{10|z|}{|\omega|^3}$ which converges absolutely by (a).
   - Hence by the Weierstrass M-test, the series defining $\wp(z)$ converges absolutely and uniformly on compact subsets of $\mathbb{C}\backslash\Lambda$.

3. The $\wp$-function is meromorphic on $\mathbb{C}$ with a double pole with residue 0 at each point of Λ (and no other poles).
   - Proof: For $\omega \in \Lambda$ the local expansion of $\wp(z)$ at ω is $(z-\omega)^2 + O((z-\omega)^0)$ so there is a double pole with residue 0 at Λ. Since the series for $\wp$ is absolutely convergent on $\mathbb{C}\backslash\Lambda$ by (2), $\wp$ has no other poles.

4. The $\wp$-function is an even function: $\wp(-z) = \wp(z)$.
   - Proof: We have $\wp(-z) = \frac{1}{(-z)^2} + \sum_{\omega \in \Lambda*} \left[\frac{1}{(-z-\omega)^2} - \frac{1}{\omega^2}\right] = \frac{1}{z^2} + \sum_{\omega \in \Lambda*} \left[\frac{1}{(z+\omega)^2} - \frac{1}{\omega^2}\right] = \wp(z)$ by substituting $\omega \mapsto -\omega$ in the sum.

5. The derivative $\wp'(z) = -2\sum_{\omega \in \Lambda} \frac{1}{(z-\omega)^3}$ is an odd function with a triple pole at each point of Λ (and no other poles).
   - Proof: Since the series for $\wp$ converges uniformly on compact subsets of $\mathbb{C}\backslash\Lambda$ its derivative is obtained by differentiating the series term by term, immediately yielding the given sum.
   - Then $\wp'$ is odd since derivatives of even functions are odd, and $\wp'$ has a triple pole at each point of Λ since differentiating a pole creates a pole of one higher order but does not otherwise create new poles.

6. The $\wp$-function and its derivative are elliptic functions with respect to Λ.
   - Proof: First, $\wp'(z)$ is elliptic since the series expression in (5) is clearly invariant under translation by elements of Λ.
   - For $\wp(z)$, taking the antiderivative of $\wp'(z+\omega) = \wp'(z)$ yields $\wp(z+\omega) = \wp(z) + C_\omega$ for some constant C_ω depending only on ω and not on z. Setting $z = -\omega/2$ and using evenness of $\wp$ immediately yields $C_\omega = 0$, and so $\wp$ is also elliptic.

7. The field of even elliptic functions $\mathbb{C}(\Lambda)$ is equal to $\mathbb{C}(\wp(z))$.
   - Proof: Suppose that f is an even elliptic function, with $f(-z) = f(z) = f(z+\omega)$ for all $\omega \in \Lambda$.
   - Our goal is to construct an elliptic function having the same zeroes and poles as f using only expressions of the form $\wp(z) - c$ for constants c: then the ratio of f to this function is elliptic and has no zeroes nor poles hence is constant.

[14] Explicitly, imagine tiling the interior of the disc $|z| \leq R$ with copies of the fundamental parallelogram of Λ. Each copy contains exactly one point of Λ. If the area were perfectly covered, the number of parallelograms would equal the area of the circle πR^2 divided by the area of the parallelogram Δ. The amount of over/undercounting produced by this tiling procedure (i.e., comparing the tiling that fits the maximum number of copies of the parallelogram strictly inside the circle, to the one that uses the smallest number of copies to cover the circle) is on the order of the perimeter of the region, which is $O(R)$.

- Let D be a fundamental parallelogram for Λ and let H be a fundamental domain for $(\mathbb{C}/\Lambda)/\{\pm 1\}$ (i.e., half of the fundamental parallelogram, consisting of a unique representative chosen among the two points $\{\zeta, \omega_1 + \omega_2 - \zeta\}$ for each $\zeta \in D$).
- Now, since f is even, for each $\zeta \in D$ we have $\text{ord}_\zeta(f) = \text{ord}_{\omega_1+\omega_2-\zeta}(f)$, and also for the half-lattice points ζ with $2\zeta \in \Lambda$, we see that $\text{ord}_\zeta(f)$ is even because $f^{(i)}(z) = (-1)^{i-1} f^{(i)}(-z)$ hence $f^{(i)}(\zeta) = 0$ since $\zeta \equiv -\zeta \bmod \Lambda$.
- Now list all of the zeroes $\{a_1, \ldots, a_k\}$ and poles $\{b_1, \ldots, b_k\}$ of f inside H, including appropriate multiplicities, where we list any zero or pole ζ with $2\zeta \in \Lambda$ with half multiplicity.
- We claim that the function $g(z) = \prod_{i=1}^k \dfrac{\wp(z) - \wp(a_i)}{\wp(z) - \wp(b_i)}$ has the same zero and pole orders as f.
- To see this, observe that $\wp(z) - \wp(a_i)$ has a zero at a_i and a zero at $-a_i$ (if $a_i = -a_i$ this is a double zero) and a double pole at 0.
- Hence by construction, $g(z)$ has the same zero and pole order as f does at all points except possibly at 0.
- But because f and g are both elliptic, the sum of both of their orders over all points is 0, and so they must have the same order at 0 as well. Hence the ratio $f(z)/g(z)$ is elliptic with no zeroes or poles, so it is constant. We conclude that $f(z) \in \mathbb{C}(\wp(z))$ as claimed.

8. The field of elliptic functions $\mathbb{C}(\Lambda)$ is equal to $\mathbb{C}(\wp(z), \wp'(z))$.
   - Proof: If $f(z)$ is elliptic, then both of the functions $\dfrac{f(z) + f(-z)}{2}$ and $\dfrac{f(z) - f(-z)}{2\wp'(z)}$ are even and elliptic, hence by (7) they are both rational functions of $\wp(z)$.
   - Then if $g(\wp(z)) = \dfrac{f(z) + f(-z)}{2}$ and $h(\wp(z)) = \dfrac{f(z) - f(-z)}{2\wp'(z)}$, we have $f(z) = g(\wp(z)) + \wp'(z) \cdot h(\wp(z)) \in \mathbb{C}(\wp(z), \wp'(z))$.
   - Remark: In fact, this shows every elliptic function is a rational function in $\wp(z)$ plus $\wp'(z)$ times another rational function in $\wp(z)$.

0.22 (Nov 27) Elliptic Curves via the Weierstrass $\wp$-Function

- The goal of this entire discussion of elliptic functions was to find the analogues of the coordinate functions x and y on $\mathbb{C}/\Lambda$.
  - Since $\wp(z)$ has a double pole at 0 and $\wp'(z)$ has a triple pole at 0, these two functions are natural candidates for x and y, following the Riemann-Roch analogy (in which x was constructed as an element of $L(2P)$ not in $L(P)$ and y was constructed as an element of $L(3P)$ not in $L(2P)$).
  - We therefore can hope that there exists a relation of the form $\wp'(z)^2 = \wp(z)^3 + A\wp(z) + B$ for some constants A and B (which necessarily will depend on the lattice).
  - Indeed, we know there must be some algebraic relation of this general form, because $\wp'(z)^2$ is an even elliptic function, hence by (7) in the proposition above it must be a rational function of $\wp(z)$.
  - We can use (7) to compute the precise relation, which requires only understanding the zeroes and poles of $\wp'(z)$. This will give us one form of the cubic expression we seek.
  - Alternatively, we could simply calculate the Laurent expansions of each of the terms near $z = 0$ and compute an appropriate linear combination that is holomorphic: then it will be a holomorphic elliptic function hence constant. This will give us a second form of the cubic expression.
- Now we carry out these two different calculations, which will allow us to establish the precise nature of the map associating $\mathbb{C}/\Lambda$ with the complex points of an elliptic curve $E(\mathbb{C})$:
- Theorem (Elliptic Curves and $\wp$-Functions): Let $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ be a complex lattice with associated Weierstrass $\wp$-function $\wp(z) = \dfrac{1}{z^2} + \sum_{\omega \in \Lambda^*} \left[\dfrac{1}{(z-\omega)^2} - \dfrac{1}{\omega^2} \right]$ and Eisenstein series $G_{2k} = \sum_{\omega \in \Lambda^*} \dfrac{1}{\omega^{2k}}$. Then the following hold:

1. The derivative $\wp'(z)$ has three single zeroes, located at the nonzero half-lattice points $\omega_1/2$, $\omega_2/2$, $(\omega_1+\omega_2)/2$.
   - Proof: We have already shown that $\wp'(z)$ has a triple pole at 0, and so it must also have three zeroes.
   - From the fact that $\wp'$ is both elliptic and odd, we can see that $\wp'(\omega_1/2) = \wp'(\omega_1/2-\omega_1) = \wp'(-\omega_1/2) = -\wp'(\omega_1/2)$, and so $\wp'(\omega_1/2)=0$.
   - Likewise we also have $\wp'(\omega_2/2)=0$ and $\wp'((\omega_1+\omega_2)/2)=0$, and so $\wp'$ has zeroes at the nonzero half-lattice points $\omega_1/2$, $\omega_2/2$, $(\omega_1+\omega_2)/2$. Since $\wp'$ only has three zeroes, these are all of the zeroes.
2. We have $\wp'(z)^2 = 4(\wp(z)-e_1)(\wp(z)-e_2)(\wp(z)-e_3)$ for $e_1=\wp(\omega_1/2)$, $e_2=\wp(\omega_2/2)$, and $e_3=\wp((\omega_1+\omega_2)/2)$.
   - Proof: Applying the proof of (7), we see that for $e_1=\wp(\omega_1/2)$, $e_2=\wp(\omega_2/2)$, and $e_3=\wp((\omega_1+\omega_2)/2)$, the function $(\wp(z)-e_1)(\wp(z)-e_2)(\wp(z)-e_3)$ has the same zeroes and same zero multiplicities as $\wp'(z)^2$, and both functions also have a pole of order 6 at 0, so they are equal up to a constant factor.
   - To find this constant factor we can simply observe that $\wp(z) = z^{-2}+O(z^{-1})$ near $z=0$ while $\wp'(z) = -2z^{-3}+O(z^{-2})$ near $z=0$, so $(\wp(z)-e_1)(\wp(z)-e_2)(\wp(z)-e_3) = z^{-6}+O(z^{-5})$ while $\wp'(z)^2 = 4z^{-6}+O(z^{-5})$.
   - Hence the constant factor is the ratio, between the leading coefficients, which is 4. We conclude that $\wp'(z)^2 = 4(\wp(z)-e_1)(\wp(z)-e_2)(\wp(z)-e_3)$, as claimed.
3. The Laurent series for $\wp(z)$ around $z=0$ is given by $\wp(z) = z^{-2}+\sum_{k=1}^{\infty}(2k+1)G_{2k+2}z^{2k}$.
   - Proof: For z closer to 0 than the nearest nonzero $\omega\in\Lambda^*$, we have
$$\begin{aligned}\wp(z) &= z^{-2}+\sum_{\omega\in\Lambda*}\left[\frac{1}{(z-\omega)^2}-\frac{1}{\omega^2}\right] = z^{-2}+\sum_{\omega\in\Lambda*}\left[\frac{1}{\omega^2}\cdot\frac{1}{(1-z/\omega)^2}-\frac{1}{\omega^2}\right]\\ &= z^{-2}+\sum_{\omega\in\Lambda*}\left[\sum_{n=1}^{\infty}(n+1)\frac{z^n}{\omega^{n+2}}\right] = z^{-2}+\sum_{n=1}^{\infty}(n+1)z^n\sum_{\omega\in\Lambda*}\left[\frac{1}{\omega^{n+2}}\right]\\ &= z^{-2}+\sum_{n=1}^{\infty}(n+1)G_{n+2}z^n\end{aligned}$$
     where the change in summation order is allowed since the series converges absolutely.
   - The given formula follows immediately upon noting that G_n is zero for odd n.
4. We have $\wp'(z)^2 = 4\wp(z)^3-60G_4\wp(z)-140G_6$.
   - Proof: Using (3) and basic series manipulations, we can work out the first few terms of various Laurent expansions:
$$\begin{aligned}\wp(z) &= z^{-2}+3G_4z^2+5G_6z^4+\cdots\\ \wp(z)^2 &= z^{-4}+6G_4+10G_6z^2+\cdots\\ \wp(z)^3 &= z^{-6}+9G_4z^{-2}+15G_6+\cdots\\ \wp'(z) &= -2z^{-3}+6G_4z+20G_6z^3+\cdots\\ \wp'(z)^2 &= 4z^{-6}-24G_4z^{-2}-40G_6+\cdots\end{aligned}$$
     and so $\wp'(z)^2-4\wp(z)^3-60G_4\wp(z) = 140G_6+\cdots$.
   - Hence the difference is an elliptic function with no pole at 0 hence no poles anywhere, since 0 is the only pole of $\wp$ and $\wp'$. It is therefore constant, hence equals $140G_6$, its value at 0.
5. For $g_2=60G_4$ and $g_3=140G_6$, the polynomial $f(x)=4x^3-g_2x-g_3$ has distinct roots, so $y^2=f(x)$ is an elliptic curve.
   - Proof: From (2) the roots of $f(x)$ are the values $e_1=\wp(\omega_1/2)$, $e_2=\wp(\omega_2/2)$, and $e_3=\wp((\omega_1+\omega_2)/2)$, so we need only see they are distinct.
   - For this we observe that $\wp(z)-\wp(\omega_i/2)$ is even hence has a double zero at ω_i, but since its total pole order is 2, we see it only vanishes at ω_i. In particular, it does not vanish at the other two half-lattice points, and so e_1,e_2,e_3 are distinct.

- The proposition above establishes an explicit correspondence between complex tori $\mathbb{C}/\Lambda$ and complex elliptic curves E, via the Weierstrass $\wp$-function and its derivative. In fact, this correspondence is natural, in both the category of Riemann surfaces and in the category of groups:

- Theorem (Elliptic Curves and $\wp$-Functions): Let Λ be a complex lattice with $g_2 = 60G_4$ and $g_3 = 140G_6$ and let E be the elliptic curve $y^2 = 4x^3 - g_2x - g_3$. Define the map $\Phi : \mathbb{C}/\Lambda \to E(\mathbb{C})$ via $\Phi(z) = (\wp(z), \wp'(z))$, with $\Phi(0) = \infty$.

  1. The map Φ is a bijection.
     - Proof: By (5) we see that the image of Φ is a subset of $E(\mathbb{C})$.
     - To show Φ is onto, choose a finite point $(x, y) \in E(\mathbb{C})$: then $\wp(z) - x$ is a nonconstant elliptic function hence has a zero, say at $z = a$.
     - Then $\wp'(a)^2 = 4a^3 - g_2a - g_3 = y^2$ so (swapping a for $-a$ if needed) we have $\wp'(a) = y$: then $\Phi(a) = (x, y)$.
     - To show Φ is one-to-one, if $\Phi(z_1) = \Phi(z_2)$ then $\wp(z) - \wp(z_1)$ is an elliptic function vanishing at z_1, $-z_1$, and z_2. Since it only has order 2, two of these points must be equivalent modulo Λ.
     - If $2z_1 \notin \Lambda$ then we see $z_2 \equiv \pm z_1 \pmod{\Lambda}$, in which case $\wp'(z_1) = \wp'(z_2) = \wp'(\pm z_1) = \pm\wp'(z_1)$ so we must have the plus sign, and so $z_2 = z_1$ in $\mathbb{C}/\Lambda$.
     - If $2z_1 \in \Lambda$ then as noted in (5), $\wp(z) - \wp(z_1)$ has a double zero at z_1, so since it vanishes also at z_2, we again have $z_2 = z_1$ in $\mathbb{C}/\Lambda$.
  2. The map Φ is a globally analytic isomorphism of Riemann surfaces.
     - Proof: To show Φ is an analytic isomorphism, observe that $\Phi^*(\frac{dx}{y}) = \frac{d\wp(z)}{\wp'(z)} = \frac{\wp'(z)\,dz}{\wp'(z)} = dz$, so Φ^* maps the invariant differential of $E(\mathbb{C})$ to the invariant differential dz of $\mathbb{C}/\Lambda$.
     - This means Φ is locally an analytic isomorphism, and since Φ is a bijection from (1), it is a global isomorphism.
  3. The map Φ is a group isomorphism.
     - Proof: By (1) we need only show that Φ is a homomorphism. Let $z_1, z_2 \in \mathbb{C}$: per the geometric group law, this requires showing that $\Phi(z_1)$, $\Phi(z_2)$, $\Phi(-z_1 - z_2)$ are the three intersection points of a line with E.
     - If $z_1 = 0$ or $z_2 = 0$ then the result follows by noting $\Phi(-z) = (\wp(-z), \wp'(-z)) = (\wp(z), -\wp'(z)) = -\Phi(z)$, and the case $z_1 = -z_2$ follows in the same way.
     - Otherwise, if the line through $\Phi(z_1)$ and $\Phi(z_2)$ is $y = mx+b$ then the elliptic function $\wp'(z) - m\wp(z) - b$ has a triple pole at 0 hence has exactly three zeroes, two of which are z_1 and z_2. (If $z_1 = z_2$ this argument is still valid, as long as we use the tangent line and count with multiplicity.)
     - But by property (5) of elliptic functions, summing the coordinates of all zeroes and poles yields an element of Λ: hence the remaining zero must be $-z_1 - z_2$ modulo Λ, so the third point is indeed $\Phi(-z_1 - z_2)$ as required.

- Let us give a bit of context to the (very nice!) result of this theorem.
  - This theorem provides an explicit parametrization of the points on the elliptic curve $E/\mathbb{C}$ with Weierstrass equation $y^2 = 4x^3 - g_2x - g_3$, namely as $(x, y) = (\wp(z), \wp'(z))$ for $z \in \mathbb{C}/\Lambda$.
  - This is essentially the nicest possible form of a parametrization for the points on $E/\mathbb{C}$, as the parameter functions are meromorphic. (The only thing nicer would be if they were actually rational functions, but a rational parametrization would give an isomorphism with $\mathbb{P}^1(\mathbb{C})$ hence is only possible in genus 0.)
  - Indeed, this development nicely parallels the genus-0 case for the circle $x^2 + y^2 = 1$, which has a parametrization $x = \cos z$, $y = \sin z$ for $x \in \mathbb{C}/2\pi i\mathbb{Z}$.
  - In the genus-0 case, the parameter functions are also obtained by inverting the integrals of the differential $\omega = \frac{dx}{y}$: here this yields $\int_C \frac{dx}{y} = \int_C \frac{dx}{\sqrt{1-x^2}}$ which is the well-understood inverse sine integral that can be made well-defined using a branch cut from -1 to 1. (Then, up to sign, the other parameter function is obtained as the derivative of the first.)

- In our genus-1 case, the parameter function $\wp$ (up to a minus sign) is obtained instead by inverting the elliptic integral $\int_C \frac{dx}{y} = \int_C \frac{dx}{\sqrt{4x^3 - g_2 x - g_3}}$.

- Our next task is to bring isogenies into the discussion.
  - The theorem above indicates that we have a very robust correspondence between $\mathbb{C}$ modulo lattices and elliptic curves over $\mathbb{C}$, so we should expect that the natural morphisms in the category of elliptic curves (namely, isogenies) should have an equally natural counterpart for lattices.
  - Since the correspondence respects the group structures, we are seeking an analytic mapping that sends a lattice Λ_1 into another lattice Λ_2.
  - The only obvious analytic maps with this property are linear functions of the form $\varphi(z) = \alpha z + \beta$ for some $\alpha, \beta \in \mathbb{C}$, as any nonlinear function would distort the lattice structure.
  - Since our lattices all contain 0, we must have $\beta = \varphi(0) \in \Lambda_2$. But since we only care about the maps modulo the lattice, we may simply apply a translation to the image to move β to 0, and thereby put φ into the form $\varphi(z) = \alpha z$.
  - In order for this to be a well-defined map from $\mathbb{C}/\Lambda_1$ to $\mathbb{C}/\Lambda_2$, we would require the image of Λ_1 to be contained in Λ_2, meaning that $\alpha\Lambda_1 \subseteq \Lambda_2$.
  - We now show that indeed, these complex scalings on lattices correspond to isogenies of elliptic curves.

- Theorem (Isogenies and Lattices): Let Λ_1 and Λ_2 be complex lattices.

  1. The only holomorphic functions $\varphi : \mathbb{C}/\Lambda_1 \to \mathbb{C}/\Lambda_2$ with $\varphi(0) = 0$ are the scalings $\varphi_\alpha(z) = \alpha z$ such that $\alpha\Lambda_1 \subseteq \Lambda_2$, and conversely each such scaling is a holomorphic function from $\mathbb{C}/\Lambda_1$ to $\mathbb{C}/\Lambda_2$.
     - Proof: The fact that each of these maps is a holomorphic function from $\mathbb{C}/\Lambda_1$ to $\mathbb{C}/\Lambda_2$ is obvious (the maps are well-defined by the requirement $\alpha\Lambda_1 \subseteq \Lambda_2$, and the maps are clearly holomorphic).
     - Now suppose that $\varphi : \mathbb{C}/\Lambda_1 \to \mathbb{C}/\Lambda_2$ is holomorphic with $\varphi(0) = 0$. Since $\mathbb{C}$ is simply connected, by the lifting property of universal covers, we may lift φ to a holomorphic function $f : \mathbb{C} \to \mathbb{C}$ such that $f(0) = 0$ and $\varphi \circ \pi_1 = \pi_2 \circ f$, where $\pi_i : \mathbb{C} \to \mathbb{C}/\Lambda_i$ is the natural projection map.
     - Then for any $\omega_1 \in \Lambda_1$ we have $f(z + \omega_1) \equiv f(z) \pmod{\Lambda_2}$, so that $f(z + \omega_1) - f(z) \in \Lambda_2$. Fixing ω_1 and letting z vary continuously yields a continuous function $f(z + \omega_1) - f(z)$ from $\mathbb{C}$ to the discrete subset Λ_2, so this difference must be constant, hence independent of z.
     - Differentiating then yields $f'(z + \omega_1) = f'(z)$ and so f' is a holomorphic elliptic function hence constant. Hence f is linear, and then since $f(0) = 0$ we see that $f(z) = \alpha z$ for some α. Finally, since $f(\Lambda_1) \subseteq \Lambda_2$ per the lifting property, we have $\alpha\Lambda_1 \subseteq \Lambda_2$.
  2. If $\varphi : E_1 \to E_2$ is an isogeny of elliptic curves defined over $\mathbb{C}$ with associated lattices Λ_1 and Λ_2, then φ corresponds to a unique scaling map $\varphi_\alpha(z) = \alpha z$ for some α with $\alpha\Lambda_1 \subseteq \Lambda_2$, and conversely each such scaling corresponds to an isogeny.
     - Proof: Because isogenies are rational functions that are defined everywhere, under the correspondence Φ described earlier, each isogeny is a holomorphic function from $\mathbb{C}/\Lambda_1$ to $\mathbb{C}/\Lambda_2$. By (1), such a map must be a scaling.
     - Conversely, suppose we have a scaling map $\varphi_\alpha(z) = \alpha z$ for some α with $\alpha\Lambda_1 \subseteq \Lambda_2$. Then the map φ is given explicitly by $\varphi(\wp(z;\Lambda_1), \wp'(z;\Lambda_2)) = (\wp(\alpha z;\Lambda_2), \wp'(\alpha z;\Lambda_2))$.
     - Now because $\alpha\Lambda_1 \subseteq \Lambda_2$, for any $\omega_1 \in \Lambda_1$ we have $\wp(\alpha(z + \omega_1);\Lambda_2) = \wp(\alpha z + \alpha\omega_1;\Lambda_2) = \wp(\alpha z;\Lambda_2)$ where the last equality follows because $\alpha\omega_1 \in \Lambda_2$.
     - But this means $f(z) = \wp(\alpha z;\Lambda_2)$ is an elliptic function with respect to Λ_1, hence by our results it is some rational function in $x = \wp(z;\Lambda_1)$ and $y = \wp'(z;\Lambda_2)$. Differentiating shows that $\wp'(\alpha z;\Lambda_2)$ is also a rational function in $x = \wp(z;\Lambda_1)$ and $y = \wp'(z;\Lambda_2)$.
     - So, writing $x = \wp(z;\Lambda_1)$ and $y = \wp'(z;\Lambda_2)$, we see that $\varphi(x, y)$ is a rational function of x and y that is defined everywhere, so it is a morphism. Since it (trivially) maps 0 to 0, it is an isogeny.
  3. Two complex tori $\mathbb{C}/\Lambda_1$ and $\mathbb{C}/\Lambda_2$ are isomorphic if and only if there exists some nonzero α such that $\alpha\Lambda_1 = \Lambda_2$. (We say the lattices are homothetic.)

- Proof: By (2), the existence of an isomorphism is equivalent to saying that there are scalings $\varphi_\alpha(z) : \Lambda_1 \to \Lambda_2$ and $\varphi_\beta(z) : \Lambda_2 \to \Lambda_1$ such that $\varphi_\alpha \circ \varphi_\beta$ is the identity map, and that $\alpha\Lambda_1 \subseteq \Lambda_2$ and $\beta\Lambda_2 \subseteq \Lambda_1$.
- Obviously in that case we must have $\alpha\beta = 1$, and then $\Lambda_1 = \beta\alpha\Lambda_1 \subseteq \beta\Lambda_2 \subseteq \Lambda_1$ requires that we have equality everywhere, so $\alpha\Lambda_1 = \Lambda_2$. Conversely, if $\alpha\Lambda_1 = \Lambda_2$ then clearly the map $\varphi_{1/\alpha}$ is an inverse of φ_α; it is well defined because $(1/\alpha)\Lambda_2 = \Lambda_1$.
- Exercise: If the associated Weierstrass equation for Λ is $E : y^2 = 4x^3 + Ax + B$ for $A = -60G_4(\Lambda)$ and $B = -140G_6(\Lambda)$ and $\alpha \neq 0$, calculate the associated Weierstrass equation for $\alpha\Lambda$ and verify directly that the resulting elliptic curve is isomorphic to E.

4. (Uniformization): For any $A, B \in \mathbb{C}$ with $A^3 - 27B^2 \neq 0$, there exists a unique complex lattice Λ such that $g_2(\Lambda) = A$ and $g_3(\Lambda) = B$: specifically, it is the period lattice for the elliptic curve $E : y^2 = 4x^3 - Ax - B$.
   - We will temporarily defer the proof of this result, since it requires establishing some properties of the j-invariant as a function on lattices.
5. For an elliptic curve E with associated period lattice Λ, the two functions $\varphi : \mathbb{C}/\Lambda \to E(\mathbb{C})$ with $\varphi(z) = (\wp(z), \wp'(z))$ and $F : E(\mathbb{C}) \to \mathbb{C}/\Lambda$ with $F(P) = \int_O^P \frac{dx}{y}$ are analytic isomorphisms that are inverses.
   - Proof: We have previously shown that φ is an analytic isomorphism and that F is analytic.
   - Now, $(F \circ \varphi)(z) = \int_O^{\varphi(z)} \frac{dx}{y} = \int_O^{(\wp(z), \wp'(z))} \frac{dx}{y}$. In particular, rather trivially, $(F \circ \varphi)(0) = 0$.
   - This means $F \circ \varphi$ is an analytic map on $\mathbb{C}/\Lambda$ sending 0 to 0, so by (1) it is a complex scaling: say $(F \circ \varphi)(z) = \alpha z$.
   - Now, we have $(F \circ \varphi)^*(dz) = \varphi^* \circ F^*(dz) = \varphi^*(\frac{dx}{y}) = \frac{d\wp(z)}{\wp'(z)} = dz$, but since $(\alpha z)^*(dz) = \alpha\, dz$, we must have $\alpha = 1$.
   - Hence $F \circ \varphi$ is the identity function, so since φ is an analytic isomorphism, F is its inverse (and also an analytic isomorphism).

- We can now completely transfer back and forth between elliptic curves and lattices. More precisely, the theorem implies that we have the following equivalence of categories:

1. (Objects) Elliptic curves defined over $\mathbb{C}$ (up to isomorphism)
   (Morphisms) Isogenies $\varphi : E_1 \to E_2$.
2. (Objects) Complex lattices Λ (up to homothety)
   (Morphisms) Complex scalings $\varphi_\alpha(z) = \alpha z$ with $\alpha\Lambda_1 \subseteq \Lambda_2$.

- With this lattice perspective, we can give a much more concrete analysis of the endomorphism ring of E:

- Theorem (Endomorphism Rings in Characteristic Zero, Again): Let E be an elliptic curve defined over $\mathbb{C}$ with associated lattice $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$, and also let τ be the nonreal complex number ω_2/ω_1.

1. If $K = \mathbb{Q}(\tau)$ is not an imaginary quadratic field, then $\mathrm{End}(E) \cong \mathbb{Z}$.
   - Proof: Per our analysis above, if $E/\mathbb{C}$ has associated lattice Λ, then the endomorphisms of $E/\mathbb{C}$ correspond to complex scalings $\varphi_\alpha(z) = \alpha z$ such that $\alpha\Lambda \subseteq \Lambda$.
   - Since we only care about the lattice Λ up to homothety, we may replace $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ with its rescaling $\Lambda = \mathbb{Z} + \mathbb{Z}\tau$ by $1/\omega_1$.
   - Then $\alpha\Lambda \subseteq \Lambda \iff \alpha, \alpha\tau \in \Lambda \iff \alpha = a + b\tau$ and $\alpha\tau = c + d\tau$ for some $a, b, c, d \in \mathbb{Z}$.
   - If $b = 0$ then we have $\alpha = a$ (so that α is multiplication by an integer) and then $\alpha\tau = d\tau$ yields no condition on τ since we can just take $d = \alpha = a$.
   - Otherwise, if $b \neq 0$, then the conditions imply that $\tau(a + b\tau) = c + d\tau$, so that τ is a root of the quadratic polynomial $b\tau^2 + (a - d)\tau - c = 0$.
   - So since τ is nonreal, this means τ generates an imaginary quadratic extension $K = \mathbb{Q}(\tau)$. Thus, when $\mathbb{Q}(\tau)$ is not an imaginary quadratic field, the endomorphism ring cannot contain any α other than multiplication by an integer.

2. If $K = \mathbb{Q}(\tau)$ is an imaginary quadratic field, then $\text{End}(E)$ is an order of K.
   - Proof: Continuing the notation and argument from (1), if $K = \mathbb{Q}(\tau)$ is an imaginary quadratic field, then the possible scalar multiples α are the ones for which there exist integers a, b, c, d with $\alpha = a + b\tau$ and $\alpha\tau = c + d\tau$.
   - These equations yield $(a - \alpha)\tau \cdot (d - \alpha) = b\tau \cdot c$ so cancelling τ and rearranging yields $\alpha^2 - (a + d)\alpha + (ad - bc) = 0$. This means α is the root of a monic polynomial with integer coefficients, so it is the element of the ring of integers $\mathcal{O}_K$ of the field K.
   - This means the endomorphism ring $\text{End}(E)$ is the subring of $\mathcal{O}_K$ containing all elements $\alpha \in \mathcal{O}_K$ that can be written in the form $a + b\tau$ for $a, b \in \mathbb{Z}$.
   - Since $\tau = \dfrac{p + q\sqrt{-D}}{r}$ for some integers p, q, r, D with $q \neq 0$ and $D > 0$, we have $r\tau = p + q\sqrt{-D} \in \mathcal{O}_K$, so there is a nonreal element $\alpha \in \text{End}(E)$.
   - Then $\text{End}(E)$ contains the $\mathbb{Q}$-linearly independent set $\{1, \alpha\}$, whence $\text{End}(E) \otimes \mathbb{Q} = \mathbb{Q}(\tau)$ and so $\text{End}(E)$ is an order in the imaginary quadratic field $\mathbb{Q}(\tau)$, as claimed.
3. If E' is any elliptic curve defined over a field F of characteristic zero, then $\text{End}(E)$ is isomorphic either to $\mathbb{Z}$ or to an order in an imaginary quadratic field.
   - The general principle here, that algebraic geometry over an arbitrary algebraically closed field of characteristic zero is the same as algebraic geometry over $\mathbb{C}$, is known as the Lefschetz principle.
   - The idea is simply that all of our results deal with finite sets of varieties, morphisms, and points (or, at worst, countably many of them), which can each be written in terms of finitely many polynomials or rational functions.
   - Then, taking the field F' to be generated by all the coefficients of these equations (over $\mathbb{Q}$), we see that any question over F about these objects deals exclusively with calculations inside the algebraic closure $\overline{F'}$. But since F' has countable transcendence degree over $\mathbb{Q}$, it is isomorphic to a subfield of $\mathbb{C}$ by a Zorn's lemma argument, and then $\overline{F'}$ is also isomorphic to a subfield of $\mathbb{C}$ by the uniqueness of algebraic closures.
   - Proof: If F is a subfield of $\mathbb{C}$, then the result follows immediately from (2).
   - Otherwise, since $\text{End}(E)$ is finitely generated (as we have previously proven, it has rank at most 4), the field F' generated by the coefficients of E and all endomorphisms of E has finite degree over $\mathbb{Q}$, so it is isomorphic to a subfield of $\mathbb{C}$. Then the result follows immediately as above.

0.23 (Nov 30) Complex Multiplication, The Modular Group

- Exercise: Under the correspondence of $E(\mathbb{C})$ with C/Λ, show that the m-torsion points on $E(\mathbb{C})$ correspond to the m-division points $\frac{1}{m}\Lambda = \{z : mz \in \Lambda\}$ in $\mathbb{C}/\Lambda$. Deduce that $E[m] \cong (\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/m\mathbb{Z})$.
- Exercise: Continuing the exercise above, let e_m be the Weil pairing on $E[m]$ with $E[m]$ viewed as $\frac{1}{m}\Lambda$. For $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$, show that $e_m(\frac{a\omega_1 + b\omega_2}{m}, \frac{c\omega_1 + d\omega_2}{m}) = e^{2\pi i(ad-bc)/m}$.
- Exercise: Let φ be an endomorphism of an elliptic curve $E/\mathbb{C}$ corresponding to a scaling α on the associated lattice Λ. Show that $\deg \varphi = \#\ker \varphi = \#(\Lambda/\alpha\Lambda) = \alpha\overline{\alpha} = V$, where V is the ratio of the area of a fundamental parallelogram for $\alpha\Lambda$ to the area of a fundamental parallelogram for Λ.
- Let us now discuss the interesting situation of elliptic curves having an endomorphism ring larger than $\mathbb{Z}$.
- Definition: If E is an elliptic curve over $\mathbb{C}$ whose endomorphism ring is an imaginary quadratic order, we say E has complex multiplication (often shortened to "E has CM").
  - The lattice perspective explains the terminology: the endomorphisms of E correspond to complex scalings of the associated lattice, and so when E has endomorphisms other than the usual multiplication-by-m maps $[m]$, these "extra endomorphisms" correspond to multiplications by a nonreal complex number on the associated lattice.
  - Example: For the lattice $\Lambda = \mathbb{Z} + \mathbb{Z}i$, the endomorphism ring of the associated elliptic curve is $\mathbb{Z}[i]$, since the set of scalars $\alpha \in \mathbb{C}$ with $\alpha\Lambda \subseteq \Lambda$ is the Gaussian integer ring $\mathbb{Z}[i]$, which is an order in the field $\mathbb{Q}(i)$.

- Exercise: For the lattices $\Lambda = \mathbb{Z}+\mathbb{Z}\rho$ where $\rho = e^{2\pi i/3}$ is a nonreal cube root of unity, and $\Lambda = \mathbb{Z}+\mathbb{Z}\sqrt{-5}$, find the endomorphism rings of the associated elliptic curves.

- There is very much to say about elliptic curves with complex multiplication, so we will only give a very minimal overview of some of their properties.
  - For convenience we will restrict attention to elliptic curves with complex multiplication by the full ring of integers $\mathcal{O}_K$ of an imaginary quadratic field $K = \mathbb{Q}(\sqrt{-D})$.
  - So suppose we have a lattice $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ with corresponding elliptic curve E.
  - From our discussion above, we have $\mathrm{End}(E) \cong \{\alpha \in K : \alpha\Lambda \subseteq \Lambda\}$, so in particular, if Λ is an ideal of the ring $\mathcal{O}_K$, then all $\alpha \in \mathcal{O}_K$ will have $\alpha\Lambda \subseteq \Lambda$, and so the endomorphism ring will be isomorphic to $\mathcal{O}_K$. (Indeed, this is even true when Λ is any K-scalar multiple of an ideal of $\mathcal{O}_K$, which is to say, when Λ is a fractional ideal[15] of K.)
  - However, since we are only interested in lattices up to homothety, scaling a lattice by an element of $\mathcal{O}_K$ will yield an isomorphic elliptic curve, so we should identify ideals that differ by a principal factor. In other words, we only want representatives from the group of nonzero ideals modulo principal ideals, which is simply the ideal class group of K.
  - In fact, these yield all of the elliptic curves with complex multiplication by $\mathcal{O}_K$:

- Theorem (Complex Multiplication and Class Groups): Let K be an imaginary quadratic field with ring of integers $\mathcal{O}_K$, let $\mathfrak{a}$ be a nonzero fractional ideal of $\mathcal{O}_K$, and let Λ be a lattice whose associated elliptic curve E_Λ has complex multiplication by $\mathcal{O}_K$. Then the following hold:

  1. The product $\mathfrak{a}\lambda = \{\alpha_1\lambda_1 + \cdots + \alpha_r\lambda_r : \alpha_i \in \mathfrak{a}_i,\ \lambda_i \in \Lambda\}$ is a lattice in $\mathbb{C}$.
     - Proof: If $\mathfrak{a} = \frac{1}{d}I$ for an integral ideal I of $\mathcal{O}_K$, then $\mathfrak{a}$ is clearly a discrete subgroup of $\mathbb{C}$ (as it is contained in $\frac{1}{d}\mathcal{O}_K$) and it has rank 2 since I has rank 2 (as it is a nonzero ideal hence contains $r\mathcal{O}_K$ for any nonzero $r \in I$).
  2. The elliptic curve $E_{\mathfrak{a}\Lambda}$ has complex multiplication by $\mathcal{O}_K$.
     - Proof: For any $\beta \in \mathbb{C}$ we have $\beta \in \mathrm{End}(E_{\mathfrak{a}\Lambda}) \iff \beta\mathfrak{a}\Lambda \subseteq \mathfrak{a}\Lambda \iff \mathfrak{a}^{-1}\beta\mathfrak{a}\Lambda \subseteq \mathfrak{a}^{-1}\mathfrak{a}\Lambda \iff \beta\Lambda \subseteq \Lambda \iff \beta \in \mathrm{End}(E_\Lambda)$, where $\mathfrak{a}^{-1}$ is the fractional ideal inverse of $\mathfrak{a}$ in $\mathcal{O}_K$.
  3. For any other fractional ideal $\mathfrak{b}$, the curves $E_{\mathfrak{a}\Lambda}$ and $E_{\mathfrak{b}\Lambda}$ are isomorphic if and only if $\mathfrak{a}$ and $\mathfrak{b}$ have the same ideal class in the class group of K.
     - Proof: As we have shown, two curves are isomorphic if and only if their lattices are homothetic, which in this case is equivalent to saying there exists some $c \in \mathbb{C}$ with $c(\mathfrak{a}\Lambda) = \mathfrak{b}\Lambda$.
     - Then both $\mathfrak{a}\mathfrak{b}^{-1}c$ and $(\mathfrak{a}\mathfrak{b}^{-1}c)^{-1}$ scale Λ to itself, so they are both ideals of $\mathcal{O}_K$, but the only invertible ideal of $\mathcal{O}_K$ is $\mathcal{O}_K$ itself.
     - This means $\mathfrak{a}\mathfrak{b}^{-1}c = \mathcal{O}_K$ and so $\mathfrak{a}c = \mathfrak{b}$, which means $\mathfrak{a}$ and $\mathfrak{b}$ have the same ideal class in the class group of K.
  4. The class group $\mathrm{Cl}(K)$ has a simply transitive action on the isomorphism classes of elliptic curves with complex multiplication by $\mathcal{O}_K$, via $\mathfrak{a} \cdot E_\Lambda = E_{\mathfrak{a}\Lambda}$.
     - Proof: The scaling operation of fractional ideals on lattices is clearly a group action since lattice scaling clearly satisfies $\mathfrak{a}(\mathfrak{b}\Lambda) = (\mathfrak{a}\mathfrak{b})\Lambda$.
     - Working instead with the associated elliptic curves yields a group action of fractional ideals on curves. This action then descends to an action of the class group on isomorphism classes of elliptic curves by (3).
     - To see that the action is transitive suppose Λ_1 and Λ_2 are lattices whose associated elliptic curves have endomorphism ring $\mathcal{O}_K$.

[15] Recall that a fractional ideal of an integral domain R is an R-submodule of the field of fractions of R having the form $d^{-1}I$ for some nonzero $d \in R$ and some ideal I of R. A fractional ideal $\mathfrak{a}$ is invertible when there exists another fractional ideal $\mathfrak{b}$ with $\mathfrak{a}\mathfrak{b} = R$; since $\mathcal{O}_K$ is a Dedekind domain, every nonzero fractional ideal is invertible (indeed, this property actually characterizes Dedekind domains). In the quadratic integer ring $\mathcal{O}_K$, one has $(d^{-1}I)^{-1} = (d/N(I))\overline{I}$ where $\overline{I}$ is the conjugate ideal of I and $N(I) = \#\mathcal{O}_K/I$ is the ideal norm of I.

- Choose any nonzero $\lambda_1 \in \Lambda_1$ and let $\mathfrak{a}_1 = \frac{1}{\lambda_1}\Lambda_1$: then $\mathfrak{a}_1$ is a fractional ideal of K since it is a finitely generated subring of K. Similarly choose any nonzero $\lambda_2 \in \Lambda_2$ and let $\mathfrak{a}_2 = \frac{1}{\lambda_2}\Lambda_2$.
- Then $\frac{\lambda_2}{\lambda_1}\mathfrak{a}_2\mathfrak{a}_1^{-1}\Lambda_1 = \Lambda_2$ so for $\mathfrak{b} = \mathfrak{a}_2\mathfrak{a}_1^{-1}$ we have $\mathfrak{b} \cdot E_{\Lambda_1} = E_{\mathfrak{b}\Lambda_1} = E_{(\lambda_1/\lambda_2)\Lambda_2} \cong E_{\Lambda_2}$.
- So there exists a fractional ideal sending E_{Λ_1} to E_{Λ_2}, meaning the action is transitive. Finally, the action is simply transitive, since if $\mathfrak{a} \cdot E_\Lambda = \mathfrak{b} \cdot E_\Lambda$ then by (3), $\mathfrak{a}$ and $\mathfrak{b}$ represent the same ideal class.

5. Up to isomorphism, the number of elliptic curves over $\mathbb{C}$ with complex multiplication by $\mathcal{O}_K$ is equal to the ideal class number $h(K)$.
   - Exercise: Let G be a group acting simply transitively on a set S, meaning that for any $a, b \in S$ there exists a unique $g \in G$ with $g \cdot a = b$. Prove that $\#G = \#S$.
   - Proof: Apply the exercise to (4).
6. If $E/\mathbb{C}$ is any elliptic curve with complex multiplication by $\mathcal{O}_K$, then the j-invariant $j(E)$ is algebraic over $\mathbb{Q}$, and in fact $[\mathbb{Q}(j(E)) : \mathbb{Q}] \leq h(K)$. As a consequence, E is isomorphic to a curve defined over an algebraic extension of $\mathbb{Q}$.
   - Proof: Let σ be any field automorphism of $\mathbb{C}$. Then $\mathrm{End}(\sigma(E)) \cong \mathrm{End}(E)$ via the map $\varphi \mapsto \sigma\varphi\sigma^{-1}$.
   - Hence if E has complex multiplication by $\mathcal{O}_K$, so does $\sigma(E)$ for any automorphism σ of $\mathbb{C}$.
   - Additionally, since $j(\sigma(E)) = \sigma(j(E))$ since the j-invariant is a rational function of the Weierstrass coefficients of E, by (5) there are at most $h(K)$ possible values for $j(\sigma(E))$ since $\sigma(E)$ has complex multiplication by $\mathcal{O}_K$.
   - But this means there are at most $h(K)$ possible values for $\sigma(j(E))$ as σ ranges over automorphisms of $\mathbb{C}$, so $j(E)$ is algebraic (as any transcendental element of $\mathbb{C}$ can be mapped to any other transcendental element of $\mathbb{C}$ via a Zorn's lemma argument) and then it has algebraic degree at most $h(K)$ over $\mathbb{Q}$ since it has at most $h(K)$ possible Galois conjugates over $\mathbb{Q}$.
   - The last part now follows immediately, because E is isomorphic to a curve defined over $\mathbb{Q}(j(E))$ as we showed previously in our analysis of j-invariants.

- We will remark that (6) in the theorem above can be quite substantially strengthened, as the inequality is always an equality: when E has complex multiplication by the ring of integers $\mathcal{O}_K$, the extension degree $[\mathbb{Q}(j(E)) : \mathbb{Q}]$ always equals the class number $h(K)$.
  - In fact, the extension field $K(j(E))$ is the Hilbert class field of K, the maximal unramified abelian extension of K.
  - As a consequence, the j-invariant $j(E)$ is rational if and only if the class number $h(K)$ equals 1.
  - There are exactly 9 imaginary quadratic fields of class number 1, as proven by Baker using linear forms in logarithms (in 1966), and Heegner (in 1952) and Stark (in 1967) using modular functions. As such, there are 9 rational j-invariants yielding elliptic curves with complex multiplication by a full ring of integers $\mathcal{O}_K$. Here is a table:

Field	$\mathbb{Q}(i)$	$\mathbb{Q}(\sqrt{-2})$	$\mathbb{Q}(\sqrt{-3})$	$\mathbb{Q}(\sqrt{-7})$	$\mathbb{Q}(\sqrt{-11})$
$j(E)$	1728	8000	0	-3375	-32768
Field	$\mathbb{Q}(\sqrt{-19})$	$\mathbb{Q}(\sqrt{-43})$	$\mathbb{Q}(\sqrt{-67})$	$\mathbb{Q}(\sqrt{-163})$	
$j(E)$	-884736	-884736000	-294395904000	-262537412640768000	

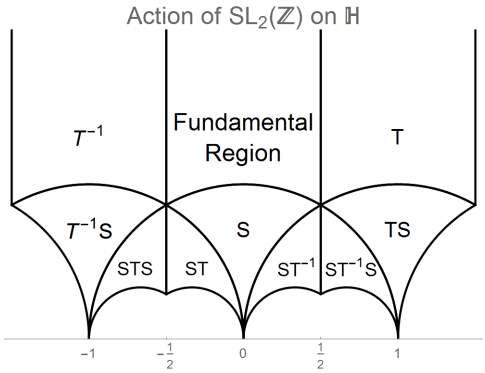
  - Exercise: Compute the endomorphism rings of $y^2 = x^3 + x$, $y^2 = x^3 - 35x + 98$, and $y^2 = x^3 + 4x^2 + 2x$.
  - Exercise: Find an elliptic curve having complex multiplication by $\mathcal{O}_{\sqrt{-11}} = \mathbb{Z}[\frac{1+\sqrt{-11}}{2}]$.
  - Exercise: For the lattice $\Lambda = \mathbb{Z} + \mathbb{Z}i$ show that $G_6(\Lambda) = 0$. Deduce that the associated elliptic curve is of the form $y^2 = x^3 + Ax$. What is the j-invariant of this curve? What is its endomorphism ring?
  - Exercise: For the lattice $\Lambda = \mathbb{Z} + \mathbb{Z}e^{2\pi i/3}$ show that $G_4(\Lambda) = 0$. Deduce that the associated elliptic curve is of the form $y^2 = x^3 + B$. What is the j-invariant of this curve? What is its endomorphism ring?

- We now shift our focus back to lattices and functions on lattices, to study the question of classifying all complex lattices up to homothety.

- From our equivalence of categories, this is equivalent to classifying elliptic curves up to isomorphism, which (as we have already seen) is completely solved by calculating the j-invariant. So, we can keep in mind the fact that the answer will turn out to be classified by the j-invariant as a function on complex lattices.
- For now, let us work directly with lattices up to homothety. Since we may rescale the lattice $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ arbitrarily, rescaling by $1/\omega_1$ allows us to work instead with a lattice of the form $\Lambda = \mathbb{Z} + \mathbb{Z}\tau$ for $\tau = \omega_2/\omega_1$, which (by sending $\tau \mapsto -\tau$ if needed) we may assume is a complex number in the upper half-plane $\mathbb{H} = \{\tau \in \mathbb{C} : \text{im}(\tau) > 0\}$.
- However, many such τ yield equivalent lattices, since there are many changes of basis (e.g., $\tau \mapsto \tau + 1$) that preserve the lattice structure or send Λ to a homothetic lattice (e.g., $\tau \mapsto -\tau^{-1}$).
- We first observe that for a lattice $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$, the set $\{\omega_3, \omega_4\}$ is a basis of Λ if and only if there exists some matrix $M \in GL_2(\mathbb{Z})$ such that $M(\omega_1, \omega_2) = (\omega_3, \omega_4)$, since the inverse change of basis M^{-1} must also have integer entries.
- If we then impose the additional restrictions that $\text{im}(\omega_2/\omega_1) > 0$ and $\text{im}(\omega_4/\omega_3) > 0$, then the matrix M preserves basis orientations hence must have positive determinant, and so $M \in SL_2(\mathbb{Z})$.
- Now suppose that $\Lambda = \mathbb{Z} + \mathbb{Z}\tau$ is homothetic to $\Lambda' = \mathbb{Z} + \mathbb{Z}\tau'$ for some $\tau, \tau' \in \mathbb{H}$, say with $\Lambda' = \alpha\Lambda$.
- This is equivalent to requiring that $\{1, \tau'\}$ is a change of basis from $\{\alpha, \alpha\tau\}$, which by our observations above is equivalent to saying that $(\alpha, \alpha\tau') = M(1, \tau)$ for some $M \in SL_2(\mathbb{Z})$.
- Writing $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ explicitly, this says $\tau' = a\alpha\tau + b\alpha$ and $1 = c\alpha\tau + d\alpha$, so by taking the ratio we see $\tau' = \dfrac{a\tau + b}{c\tau + d}$: in other words, τ' is the image of τ under some fractional linear transformation $\gamma(z) = \dfrac{az + b}{cz + d}$ where $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SL_2(\mathbb{Z})$.
- Exercise: For $a, b, c, d \in \mathbb{R}$ with $ad - bc > 0$ and $\tau \in \mathbb{H}$ show $\text{im}\left(\dfrac{a\tau + b}{c\tau + d}\right) = \dfrac{(ad - bc)\text{im}(\tau)}{|c\tau + d|^2} > 0$.
- Conversely, if $\tau' = \dfrac{a\tau + b}{c\tau + d}$ for some $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SL_2(\mathbb{Z})$, then for $\alpha = c\tau + d$ we see that $\alpha\Lambda' = \mathbb{Z}(a\tau + b) + \mathbb{Z}(c\tau + d) = \Lambda$ since $\{a\tau + b, c\tau + d\}$ is an integral change of basis from $\{1, \tau\}$, and by the exercise, the basis has the correct orientation.
- So we see that in order to obtain a unique value of τ for each lattice up to homothety, we want to take representatives for the group action of $SL_2(\mathbb{Z})$ on $\mathbb{H}$ acting via $\gamma\tau = \dfrac{a\tau + b}{c\tau + d}$ for each $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SL_2(\mathbb{Z})$.
- Remark: One may check with some tedious algebra that this action is a group action, but it follows much more naturally by observing that the group of fractional linear transformations $\gamma(z) = \dfrac{az + b}{cz + d}$ with $ad - bc \neq 0$ on $\mathbb{C} = \mathbb{A}^1(\mathbb{C})$ are simply the invertible linear transformations $[z_0 : z_1] \mapsto [az_0 + bz_1 : cz_0 + dz_1]$ on $\mathbb{P}^1(\mathbb{C})$; the nonzero-determinant condition is required to ensure that the map is well-defined, and then the composition being the same as matrix multiplication is obvious.
- Notice also that $-I = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \in SL_2(\mathbb{Z})$ acts trivially on $\mathbb{H}$, so in fact we really have an action of $SL_2(\mathbb{Z})/\{\pm I\}$ on $\mathbb{H}$.

- Definition: The group $\Gamma(1) = SL_2(\mathbb{Z})/\{\pm I\} = PSL_2(\mathbb{Z})$ is called the modular group.
- Although $\Gamma(1)$ is a quotient group, we will simply write its elements as matrices, taking as implicit the equivalence of any matrix with its negative.
- Exercise: Show that the action of the modular group on $\mathbb{H}$ is faithful (i.e., that the identity is the only element acting trivially on all of $\mathbb{H}$).
- The modular group has a convenient pair of generators, given by the matrices $S = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ and $T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ corresponding to the maps $z \mapsto -z^{-1}$ and $z \mapsto z + 1$ respectively.

- Exercise: For $S = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ and $T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ in $\Gamma(1)$, show that S has order 2 and ST has order 3.
- We can also describe a fundamental domain for the action of $\Gamma(1)$ on $\mathbb{H}$: specifically, we take the region D with $|\mathrm{Re}(z)| \le \frac{1}{2}$ and $|z| \ge 1$. Then S and T map D to various other regions, as illustrated below:



- Since S and T are fractional linear transformations, they map generalized circles[16] to generalized circles, and so since the boundary of D consists of two half-lines and a circular arc, each of the images of D under a word in S and T is a generalized triangular region (whose "sides" are arcs of generalized circles), some of which are in the diagram above.

- Let us now prove that D is a fundamental domain for the action of $\Gamma(1)$ on $\mathbb{H}$:

- Proposition (Fundamental Domain for $\Gamma(1)$): Let D be the region consisting of all $z \in \mathbb{H}$ with $|\mathrm{Re}(z)| \le \frac{1}{2}$ and $|z| \ge 1$, and let $\Gamma(1)$ be the modular group with its usual action on $\mathbb{H}$ with elements $S = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ and $T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

  1. For any $z \in \mathbb{H}$ there exists some $\gamma \in \langle S, T \rangle$ with $\gamma z \in D$.
     - Proof: Let $\Lambda = \mathbb{Z} + \mathbb{Z}z$. Since Λ is discrete, there are only finitely many pairs (c, d) with $|cz + d|$ within a given radius of the origin.
     - Now take $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \langle S, T \rangle$ such that $|cz + d|$ is minimized (this is possible by the observation above). Equivalently, this means $\mathrm{Im}(\gamma z) = \dfrac{\mathrm{Im}(z)}{|cz + d|^2}$ is maximized among all $\gamma \in \langle S, T \rangle$.
     - Now select an integer n such that $|\mathrm{Re}(T^n \gamma z)| \le 1/2$; this is possible since applying T shifts the real part by 1.
     - We claim that $z' = T^n \gamma z$ lies in D: for this, simply note that if $|z'| < 1$ then $\mathrm{Im}(Sz') = \dfrac{\mathrm{Im}(z')}{|z'|^2} > \mathrm{Im}(z') = \mathrm{Im}(\gamma z)$, but this would contradict the maximality above because then the element $ST^n\gamma$ would have $ST^n \gamma z = Sz'$ with larger imaginary part than γz.
  2. If $z \in D$ and $\gamma \in \Gamma(1)$ are such that $\gamma \ne 1$ and $\gamma z \in D$, then z and γz must lie on the boundary of D. More precisely, either $\mathrm{Re}(z) = \pm 1/2$ and $\gamma z = z \mp 1$, or $|z| = 1$ and $\gamma z = -1/z$.
     - Proof: By interchanging z and γz as necessary, we may assume $\mathrm{Im}(z) \le \mathrm{Im}(\gamma z)$.
     - For $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, since $\mathrm{Im}(\gamma z) = \dfrac{\mathrm{Im}(z)}{|cz + d|^2}$ we have $|cz + d| \le 1$. But since $|cz + d| \ge |c| \,\mathrm{Im}(z) \ge |c| \,(\sqrt{3}/2)$, we must have $c = 0$, -1, or 1, and by rescaling γ by -1 we are reduced just to the cases $c = 0$ and $c = 1$.
     - If $c = 0$, then we must have $d = \pm 1$, so rescaling γ by -1 we may assume $d = 1$, and then since $\det \gamma = 1$ we have $a = 1$ also. Then $\gamma z = z + b$, but since γ is not the identity the only possibility is to have $b = \mp 1$ so that $\mathrm{Re}(z) = \pm 1/2$ and $\gamma z = z \mp 1$.

[16] Recall that a generalized circle in $\mathbb{C}$ refers to a circle or a line (which we think of as a circle with infinite radius).

- If $c = 1$, then $|z + d| \leq 1$ implies that $d = 0$ and thus $|z| = 1$, except in the two situations where $z = e^{\pi i/3}$ or $e^{2\pi i/3}$ (in which case respectively $d = -1$ and $d = 1$ are possible), as is easily seen from the diagram of D, but in those cases we still have $|z| = 1$.
- When $d = 0$ since $\det \gamma = 1$ we have $b = -1$ so that $\gamma z = a - 1/z$ and as above the only way this can occur is for $a = 0$ (in which case $\gamma z = -1/z$) or for $z = e^{\pi i/3}$ (where $a = -1$ is also possible) or $e^{2\pi i/3}$ (where $a = 1$ is also possible).
- One may then check the finite number of possible cases for γ to see that when $z = e^{\pi i/3}$ or $e^{2\pi i/3}$ we also have $\gamma z = e^{\pi i/3}$ or $e^{2\pi i/3}$, so the result holds.

3. $\Gamma(1)$ is generated by S and T.
   - Proof: Let $\gamma \in \Gamma(1)$, choose any z_0 in the interior of D such as $z_0 = 3i$, and let $z = \gamma z_0$.
   - By (1), there exists some $\gamma' \in \langle S, T \rangle$ such that $\gamma' z \in D$, which is to say, $(\gamma'\gamma)z_0 \in D$.
   - But by (2), since z_0 and $(\gamma'\gamma)z_0$ are both in D, and they are not both on the boundary, they must be equal, and $\gamma'\gamma = 1$. Hence $\gamma = (\gamma')^{-1} \in \langle S, T \rangle$, as required.
   - Remark: In fact, one may show that $\Gamma(1)$ has a presentation $\langle S, T : S^2 = (ST)^3 = e \rangle$, meaning that $S^2 = e$ and $(ST)^3 = e$ are essentially the only relations between S and T. Equivalently, $\Gamma(1)$ is the free product of the subgroups $\langle S \rangle$ of order 2 and $\langle ST \rangle$ of order 3.

- The quotient space[17] $\Gamma(1)\backslash\mathbb{H}$ classifies lattices up to homothety, and our discussion above shows that we may write down convenient representatives for this quotient space: namely, the region D (with appropriate identifications made on its boundary).
  - More explicitly, we identify the left and right edges $\mathrm{Re}(z) = -1/2$ and $\mathrm{Re}(z) = 1/2$ of the boundary, and we also identify the left and right halves of the arc of $|z| = 1$.
  - Topologically, this quotient space is isomorphic to $\mathbb{C}$.
  - If we look at other $\Gamma(1)$-translates of the fundamental domain D (such as the one obtained by applying S to D) we see that they are all generalized triangles, suggesting that we are actually "missing" the third vertex of the triangle, which corresponds to the point at ∞ inside D.
  - In these other $\Gamma(1)$-translates of D, the "third vertex" is either at ∞ or is a point on the real axis: more specifically, it is a rational point, since $\gamma\infty = a/c$ is always rational for $\gamma \in \Gamma(1)$.
  - This suggests, in order to obtain the proper action, we should adjoin the points of $\mathbb{P}^1(\mathbb{Q})$ to $\mathbb{H}$, and work instead with the action of $\Gamma(1)$ on this slightly larger set.

- Definition: The extended upper half-plane $\mathbb{H}^*$ is the set $\mathbb{H} \cup \mathbb{P}^1(\mathbb{Q}) = \mathbb{H} \cup \mathbb{Q} \cup \{\infty\}$.
  - We have a natural action of $\Gamma(1)$ on $\mathbb{P}^1(\mathbb{Q})$: namely, by taking $\gamma[x : y] = [ax + by : cx + dy]$ for $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.
  - Indeed, this is just the action of (fractional) linear transformations on $\mathbb{P}^1(\mathbb{C})$ we mentioned earlier.
  - This action is transitive: for any $a/c \in \mathbb{Q}$ in lowest terms, by the Euclidean algorithm there exist $b, d \in \mathbb{Q}$ with $ad - bc = 1$; then $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \Gamma(1)$ has $\gamma\infty = a/c$, so ∞ may be moved to any point in $\mathbb{P}^1(\mathbb{Q})$.
  - Exercise: Show that the stabilizer in $\Gamma(1)$ of ∞ is the subgroup $\langle T \rangle$.

- By putting the action of $\Gamma(1)$ on $\mathbb{P}^1(\mathbb{Q})$ together with the action on $\mathbb{H}$, we get an action of $\Gamma(1)$ on $\mathbb{H}^*$. Let us examine the resulting quotient spaces:

- Definition: The modular curves $X(1)$ and $Y(1)$ are defined as $X(1) = \Gamma(1)\backslash\mathbb{H}^*$ and $Y(1) = \Gamma(1)\backslash\mathbb{H}$.
  - The points in $X(1)$ not in $Y(1)$ are called cusps of $X(1)$.
  - The reason we call these quotient spaces "modular curves" is because they have a natural Riemann surface structure that is in fact also algebraic (i.e., they are algebraic curves over $\mathbb{C}$).

[17] We write the quotient space as $\Gamma(1)\backslash\mathbb{H}$ rather than the more traditional $\mathbb{H}/\Gamma(1)$ because $\Gamma(1)$ acts on $\mathbb{H}$ via a left group action. There are other groups that we would like to take quotients by as well, sometimes simultaneously with $\Gamma(1)$, and these will mostly act on the right, so we put $\Gamma(1)$ on the left.

- Geometrically, by identifying the appropriate boundary components of the fundamental domain D, we immediately obtain the local Riemann surface structure on the interior of D (namely, the one from $\mathbb{C}$ itself).
- By applying an appropriate alteration of the fundamental domain, we may make a similar construction for the boundary points, but difficulties arise at the points $z = i$ and $z = e^{i\pi/3}$ because any open neighborhood of either one contains points equivalent under the action of $\Gamma(1)$.
- As such, writing down the Riemann surface structure explicitly (in terms of local coordinates) is quite tedious. Rather than giving the uninspiring details, we will simply say that the idea is a special case of the general notion of an orbifold, a topological space that is locally the quotient of Euclidean space by a finite group (namely, the stabilizer at the given point).
- For $Y(1)$, the local quotient is by the trivial group, except at the points i (where the stabilizer has order 2) and $e^{i\pi/3}$ (where the stabilizer has order 3).
- As a Riemann surface, $Y(1)$ is not compact. Its compactification is $X(1)$, which one may check has genus 0 via the geometric identification of the sides of the fundamental domain.

0.24 (Dec 4) Modular Functions and Modular Forms

- Now that we are viewing $X(1)$ as a Riemann surface, the next natural thing to do is study the meromorphic functions defined on it.
  - Equivalently, these are meromorphic functions defined on $\Gamma(1)\backslash\mathbb{H}^*$: namely, meromorphic functions on the upper half-plane that transform in a natural way under the action of the modular group $\Gamma(1)$.
  - Our entire motivation for discussing this quotient space in the first place was to characterize lattices up to homothety (this was the space $\Gamma(1)\backslash\mathbb{H}$), and so a meromorphic function defined on this space is the same as a meromorphic function that is well-defined on lattices up to homothety.
  - We have already constructed several meromorphic functions on lattices, such as the Eisenstein series $G_{2k}(\Lambda) = \sum_{\omega\in\Lambda^*} \frac{1}{\omega^{2k}}$. These are not well-defined up to homothety, since $G_{2k}(\alpha\Lambda) = \alpha^{-2k}G_{2k}(\Lambda)$, but by taking appropriate combinations of these functions, we can construct functions that are invariant under lattice scalings.
  - Rather than narrow our focus specifically to these functions invariant under scalings, it will be far more valuable to consider the wider class of functions with scaling properties analogous to those of the Eisenstein series.
  - For the Eisenstein series in particular, with $\Lambda = \mathbb{Z} + \mathbb{Z}\tau$, writing $G_{2k}(\tau) = \sum_{m\tau+n\in\Lambda^*} \frac{1}{(m\tau+n)^{2k}}$, then for any $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \Gamma(1)$ we have $G_{2k}(\gamma\tau) = \sum_{m\tau+n\in\Lambda^*} \frac{1}{(m(\gamma\tau)+n)^{2k}} = \sum_{m\tau+n\in\Lambda^*} \frac{(c+d\tau)^{2k}}{(m(a\tau+b)+n(c\tau+d))^{2k}}$ $= (c+d\tau)^{2k} \sum_{u+w\tau\in\Lambda^*} \frac{1}{(u\tau+w)^{2k}} = (c\tau+d)^{2k}G_{2k}(\tau)$, where the middle equality follows from the fact that $\{a\tau+b, c\tau+d\}$ is also a basis of Λ by the assumption $\gamma \in \Gamma(1)$.
  - Therefore, the invariance condition we want is $f(\gamma\tau) = (c\tau+d)^{2k}f(\tau)$ for each $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \Gamma(1)$.
- Definition: Let $k \in \mathbb{Z}$ and $f(\tau)$ be a meromorphic function on $\mathbb{H}$. We say that f is weakly modular of weight $2k$ (for $\Gamma(1)$) when $f(\gamma\tau) = (c\tau+d)^{2k}f(\tau)$ for each $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \Gamma(1)$ and each $\tau \in \mathbb{H}$.
  - Example: For $k \geq 2$, the Eisenstein series $G_{2k}(\tau) = \sum_{(m,n)\neq(0,0)} \frac{1}{(m\tau+n)^{2k}}$ is weakly modular of weight $2k$, as calculated above.
  - Exercise: Show that the space of weakly modular functions of weight $2k$ is a $\mathbb{C}$-vector space, and that the product of weakly modular functions of weights $2k$ and $2l$ yields a weakly modular function of weight $2k+2l$.

- Per the exercise we see that the space of weakly modular forms (of all weights) naturally carries the structure of a graded $\mathbb{C}$-algebra.
- Although the invariance condition may seem somewhat arbitrary, it arises rather naturally from considering differential forms:
- Exercise: Show that $d(\gamma z) = (cz+d)^{-2}dz$. Deduce that f is weakly modular of weight $2k$ if and only if the differential k-form $f(z)\,dz^k$ is invariant under the action of $\Gamma(1)$.
- The definition can also be generalized in a number of ways, such as by using various subgroups of $\Gamma(1)$ rather than $\Gamma(1)$ itself: we will do this later.
- In our definition above, we can restrict attention to the case where the weight is even, because if $f(\gamma\tau) = (c\tau+d)^k f(\tau)$ for an odd k, then taking $\gamma = -I$ yields $f(\tau) = -f(\tau)$ so f is identically zero.
- One may construct a more interesting theory of modular functions of odd weight by introducing a character $\chi(d)$ to the definition: namely, requiring $f(\gamma\tau) = (c\tau+d)^k\chi(d)f(\tau)$ for some multiplicative character χ. (We will not pursue this topic further, but it leads to many rich and interesting results.)

- Since $\Gamma(1)$ is generated by S and T, f is weakly modular if and only if $f(\tau+1) = f(\tau)$ and $f(-1/\tau) = \tau^{2k}f(\tau)$ for all $\tau \in \mathbb{H}$.
  - In particular, f is periodic with period 1, so it has a Fourier expansion in terms of the variable $q = e^{2\pi i\tau}$, say as $\tilde{f}(q) = \sum_{n=-\infty}^{\infty} a_n q^n$.
  - Then since f is meromorphic on $\mathbb{H}$, $\tilde{f}$ is meromorphic in the open unit disc $0 < |q| < 1$ with the origin removed.

- Definition: Let f be a weakly modular function with Fourier expansion $\tilde{f}(q) = \sum_{n=-\infty}^{\infty} a_n q^n$. If the expansion is actually a Laurent expansion (i.e., of the form $\sum_{n=-k}^{\infty} a_n q^n$ for some k) then we say f is meromorphic at ∞, and if the expansion is a power series (i.e., of the form $\sum_{n=0}^{\infty} a_n q^n$) we say f is holomorphic at ∞. A weakly modular function that is holomorphic at ∞ is called a modular form, and if in addition $f(\infty) = 0$, it is called a cusp form. The $\mathbb{C}$-vector space of all modular forms of weight $2k$ is denoted $\mathcal{M}_{2k}$, while the $\mathbb{C}$-vector space of all cusp forms of weight $2k$ is denoted $\mathcal{S}_{2k}$.
  - As we will show below, the Eisenstein series give examples of modular forms, and all modular forms can be written in terms of them.
  - Since the holomorphicity condition is additive and multiplicative, we see immediately that we can fit the spaces $\mathcal{M}_{2k}$ of modular forms of weight $2k$ together into a natural $\mathbb{C}$-algebra $\mathcal{M} = \bigoplus_{k=0}^{\infty} M_{2k}$.

- Proposition (Basics of Modular Forms): We have the following:

  1. Let $k \geq 2$ be an integer. The Eisenstein series $G_{2k}(\tau) = \sum_{(m,n)\neq(0,0)} \dfrac{1}{(m\tau+n)^{2k}}$ is a modular form of weight $2k$, and its value at ∞ is $G_{2k}(\infty) = 2\zeta(2k)$ where ζ denotes the Riemann zeta function.
     - Proof: We have already shown above that G_{2k} is weakly modular, so we just need to show it is holomorphic at ∞ and compute its value there.
     - First, for τ in the fundamental domain D, so that $|\mathrm{Re}(\tau)| \leq 1/2$ and $|\tau| \geq 1$, we have $|m\tau+n| = (m\tau+n)(m\overline{\tau}+n) = m^2|\tau|^2 + 2mn\,\mathrm{Re}(\tau) + n^2 \geq m^2 - mn + n^2 = |m\rho+n|$ where $\rho = e^{2\pi i/3}$.
     - But as we have already shown, the series for $G_{2k}(\rho)$ converges absolutely, and thus $G_{2k}(\tau)$ converges absolutely and uniformly to a holomorphic function on D, hence by $\Gamma(1)$-invariance the same is true on all of $\mathbb{H}$.
     - For the value at ∞, since G_{2k} is periodic we may assume without loss that $\tau \to \infty$ inside D, and by uniform convergence we may take the limit as $\mathrm{im}(\tau) \to \infty$ term-by-term. The terms with $m \neq 0$ have limit zero, while the terms with $m = 0$ are constant in τ and have sum $\sum_{n\neq 0} \dfrac{1}{n^{2k}} = 2\sum_{n=1}^{\infty} \dfrac{1}{n^{2k}} = 2\zeta(2k)$, as claimed.
  2. More explicitly, the Fourier expansion of G_{2k} is $G_{2k}(\tau) = 2\zeta(2k) + 2\dfrac{(2\pi i)^{2k}}{(2k-1)!}\sum_{n=1}^{\infty}\sigma_{2k-1}(n)q^n$ where $\sigma_d(n)$ is the sum of the dth powers of the positive divisors of n.

- Proof: We start with the standard formula $\pi \cot \pi z = \frac{1}{z} + \sum_{d=1}^{\infty} \frac{2z}{z^2 - d^2}$, which can be extracted from Euler's product formula for the sine function.
- Since $\pi \cot \pi z = \pi i \frac{e^{i\pi} + e^{-i\pi}}{e^{i\pi} - e^{i\pi}} = \pi i \frac{q+1}{q-1} = \pi i - \frac{2\pi i}{1-q} = -i\pi - 2\pi i \sum_{n=0}^{\infty} q^n$, we see that $\frac{1}{z} + \sum_{d=1}^{\infty} \frac{2z}{z^2 - d^2} = -i\pi - 2\pi i \sum_{n=0}^{\infty} q^n$.
- Differentiating $k-1$ times and then separating terms yields $\sum_{d \in \mathbb{Z}} \frac{1}{(z+d)^k} = \frac{1}{(k-1)!}(-2\pi i)^k \sum_{n=1}^{\infty} n^{k-1} q^n$.
- Now, for the Eisenstein series we have $G_{2k}(\tau) = \sum_{(m,n) \neq (0,0)} \frac{1}{(m\tau + n)^{2k}} = 2\zeta(2k) + 2\sum_{m=1}^{\infty} \sum_{n \in \mathbb{Z}} \frac{1}{(m\tau + n)^{2k}}$.
- Substituting $z = m\tau$ in the formula above then yields

$$\begin{aligned} G_{2k}(\tau) &= 2\zeta(2k) + 2\sum_{m=1}^{\infty} \frac{1}{(2k-1)!}(-2\pi i)^{2k} \sum_{n=1}^{\infty} n^{2k-1} q^{mn} \\ &= 2\zeta(2k) + \frac{2(-2\pi i)^{2k}}{(2k-1)!} \sum_{mn=d} n^{2k-1} q^d \\ &= 2\zeta(2k) + \frac{2(-2\pi i)^{2k}}{(2k-1)!} \sum_{n=1}^{\infty} \sigma_{2k-1}(n) q^n. \end{aligned}$$

3. The discriminant $\Delta = g_2^3 - 27 g_3^2$ for the usual $g_2 = 60 G_4$ and $g_3 = 140 G_6$, is a cusp form of weight 12.
   - Proof: Since g_2 is a modular form of weight 4 and g_3 is a modular form of weight 6, both g_2^3 and g_3^2 are modular forms of weight 12, hence so is Δ.
   - To see that Δ is a cusp form we compute its value at ∞.
   - By (1) we have $g_2(\infty) = 120\zeta(4) = \frac{4}{3}\pi^4$ and $g_3(\infty) = 280\zeta(6) = \frac{8}{27}\pi^6$, so $\Delta(\infty) = (\frac{4}{3}\pi^4)^3 - 27(\frac{8}{27}\pi^6)^2 = 0$, using the well-known values $\zeta(4) = \frac{\pi^4}{90}$ and $\zeta(6) = \frac{\pi^6}{945}$ first calculated by Euler using his product formula for sine.
4. Let f be a nonzero modular form of weight $2k$. Then $\sum_{P \in \Gamma(1) \backslash \mathbb{H}^*} \frac{1}{e_P} \mathrm{ord}_P(f) = \frac{k}{6}$, where e_P is the order of the stabilizer of $\Gamma(1)$ at P, which inside D is 2 at $P = i$ and 3 at $P = e^{2\pi i/3}$ and 1 elsewhere.
   - The equation can also be written as $\mathrm{ord}_\infty(f) + \frac{1}{2}\mathrm{ord}_i(f) + \frac{1}{3}\mathrm{ord}_\rho(f) + \sum_{P \in \mathrm{int}(D)} \mathrm{ord}_P(f) = \frac{k}{6}$ for $\rho = e^{2\pi i/3}$. (The order at ∞ is the order of the Fourier expansion $\tilde{f} = \sum_{n=0}^{\infty} c_n q^n$ for f.)
   - We note as usual that up to $\Gamma(1)$-equivalence, f has only finitely many zeroes and poles since it is meromorphic on the compact region $X(1) = \Gamma(1) \backslash \mathbb{H}^*$.
   - The idea is similar to our results about elliptic functions: we integrate $\frac{1}{2\pi i} \frac{f'}{f}$ around the boundary of $X(1)$ and apply Cauchy's integral theorem.
   - The "alterations" to the formula (namely, the $1/e_P$ factors) come from the fact that near i and ρ, the contour only encloses $1/2$ (respectively $1/3$) respectively of a path encircling that point, so only $1/2$ (respectively $1/3$) of the pole residue is counted.
   - Proof: Let R be large and consider the contour C_R that follows the counterclockwise boundary of D, except cuts off along the line $\mathrm{im}(z) = R$, and also follows a sixth-circle arc of radius $1/R$ around $e^{2\pi i/3}$, a half-circle arc of radius $1/R$ around i, and a sixth-circle arc of radius $1/R$ around $e^{i\pi/3}$. Additionally, we include corresponding semicircles of radius $1/R$ around any zero or pole on the boundary of D (there are necessarily two of these). These contributions will cancel out, so we will now ignore them.
   - Consider the integral $\frac{1}{2\pi i} \int_{C_R} \frac{f'(z)}{f(z)}\, dz$, which by Cauchy's residue theorem equals the sum of the residues of $\frac{f'}{f}$ over all points inside C_R, which $R \to \infty$ is the sum of the zero and pole orders at interior points of D.

- We can also the contributions to the integral on each component directly using the fractional residue theorem.
- The contributions on the vertical sides of D cancel exactly, the contribution on $\text{im}(z) = R$ tends to $-\text{ord}_\infty(f)$, the contribution on each of the sixth-circles tends to $-\frac{1}{6}\text{ord}_\rho(f)$, and the contribution on the half-circle centered at i tends to $-\frac{1}{2}\text{ord}_\rho(i)$.
- Finally we tabulate the contributions on the left and right halves of the boundary of the circle $|z| = 1$: these do not cancel exactly, but rather because $\frac{f'(Sz)}{f(Sz)} = \frac{2k}{z} + \frac{f'(z)}{f(z)}$, there is an extra contribution of $-\int \frac{2k}{z}\,dz$ over the twelfth-circle from $z = e^{i\pi/2}$ to $z = e^{\pi i/3}$, which integrates to $\frac{k}{6}$.
- Setting this sum equal to the sum of residues earlier yields the result.

5. The modular form G_4 has a simple zero at ρ and is nonzero elsewhere, the modular form G_6 has a simple zero at i and is nonzero elsewhere, and the modular form Δ has a simple zero at ∞ and is nonzero elsewhere.
   - Proof: Note that G_4, G_6, and Δ have no poles, so all of their orders are nonnegative everywhere for the purposes of using the formula $\text{ord}_\infty(f) + \frac{1}{2}\text{ord}_i(f) + \frac{1}{3}\text{ord}_\rho(f) + \sum_{P \in \text{int}(D)} \text{ord}_P(f) = \frac{k}{6}$ in (4).
   - By (4) since the weight of G_4 is 4, the sum of orders is $1/3$. But the only way this can happen is for $\text{ord}_\rho(G_4) = 1$ and for the other orders to be zero.
   - Likewise, since the weight of G_6 is 6, the sum of orders is $1/2$, and the only way this can happen is for $\text{ord}_i(G_6) = 1$ and for the other orders to be zero.
   - Finally, for Δ, since its weight is 12 its sum of orders is 1. We have shown in (3) that $\text{ord}_\infty(\Delta) \geq 1$, so we must have $\text{ord}_\infty(\Delta) = 1$ and the other orders equal to zero.
6. We have $\mathcal{M}_{2k} = 0$ for $k < 0$, and also $\mathcal{M}_0 = \mathbb{C}$, $\mathcal{M}_2 = 0$, $\mathcal{M}_4 = \mathbb{C}G_4$, $\mathcal{M}_6 = \mathbb{C}G_6$, $\mathcal{M}_8 = \mathbb{C}G_4^2$, and $\mathcal{M}_{10} = \mathbb{C}G_4G_6$.
   - Proof: If $f \in \mathcal{M}_{2k}$ is nonzero then since f is holomorphic on $\mathbb{H}$ and at ∞ we have $\text{ord}_P f \geq 0$ at all points P, and so the sum in (4) must be nonnegative. In particular, $\mathcal{M}_{2k} = 0$ for $k < 0$.
   - For $\mathcal{M}_0$ we necessarily have $\text{ord}_P f = 0$ so f is holomorphic and nonvanishing hence constant.
   - For $\mathcal{M}_2$ we cannot have a sum of orders involving terms 1, 1/2, 1/3 equal to 1/6, so $\mathcal{M}_2 = 0$.
   - For $\mathcal{M}_4$ have a sum of orders equal to $1/3$ so as in (5) f has a simple zero at ρ and is nonzero elsewhere. But then f/G_4 is holomorphic and nonvanishing hence constant.
   - Exercise: If $f \in \mathcal{M}_6$, $\mathcal{M}_8$, or $\mathcal{M}_{10}$, show that f/G_6, f/G_4^2, or $f/(G_4G_6)$ is constant, respectively.
7. For any $k \geq 2$, we have $\mathcal{M}_{2k} = \mathcal{S}_{2k} \oplus \mathbb{C}G_{2k}$.
   - Proof: The map $\varphi : \mathcal{M}_{2k} \to \mathbb{C}$ given by $\varphi(f) = f(\infty)$ is linear, with kernel $\mathcal{S}_{2k}$.
   - Also since $\varphi(G_{2k}) = 2\zeta(2k) \neq 0$ per (1), we see φ is onto. The result follows immediately from the first isomorphism theorem.
   - Remark: More generally, by the same proof, $\mathcal{M}_{2k} = \mathcal{S}_{2k} \oplus \mathbb{C}f$ for any $f \in \mathcal{M}_{2k}$ that is not a cusp form.
8. For $k \geq 0$, the map $f \mapsto \Delta \cdot f$ is a vector space isomorphism of $\mathcal{M}_{2k}$ with $\mathcal{S}_{2k+12}$.
   - Proof: Clearly this map is linear and well defined since $\Delta(\infty) = 0$ by (3) hence $(\Delta f)(\infty) = 0$, and multiplying by Δ increases the weight by 12.
   - But since Δ has a simple zero at ∞ and is nonvanishing elsewhere by (5), the map from $\mathcal{S}_{2k+12}$ to $\mathcal{M}_{2k}$ with $g \mapsto g/\Delta$ is a well-defined inverse map.
   - Explicitly, if $g \in \mathcal{S}_{2k+12}$ then g/Δ will be a weakly modular function of weight $(2k+12) - 12 = 2k$, and it is holomorphic at ∞ because dividing a holomorphic function that vanishes at P by one with a simple zero at P still yields a holomorphic function.
9. We have $\dim(\mathcal{M}_{2k}) = \lfloor k/6 \rfloor + 1$ except when $k \equiv 1 \pmod 6$ in which case $\dim(\mathcal{M}_{2k}) = \lfloor k/6 \rfloor$.
   - Proof: By (7) and (8) we have $\dim(\mathcal{M}_{2k+12}) = \dim(\mathcal{S}_{2k+12}) + 1 = \dim(\mathcal{M}_{2k}) + 1$, and now the result follows by a trivial induction using (6).

10. The graded algebra $M = \bigoplus_{k=0}^{\infty} \mathcal{M}_{2k}$ is isomorphic as a graded algebra to $\mathbb{C}[G_4, G_6]$ where G_4 has weight 4 and G_6 has weight 6.
    - Exercise: Show that the number of nonnegative integer solutions (a, b) to $2a + 3b = k$ is equal to $\lfloor k/6 \rfloor + 1$ except when $k \equiv 1 \pmod 6$ in which case it is instead $\lfloor k/6 \rfloor$.
    - Proof: Observe that $G_4^a G_6^b \in \mathcal{M}_{2k}$ whenever $2a + 3b = k$.
    - By the exercise and (9), there are exactly $\dim(\mathcal{M}_{2k})$ elements of the form $G_4^a G_6^b$ in $\mathcal{M}_{2k}$, and they are linearly independent because any linear dependence would imply that G_4^3/G_6^2 is an algebraic meromorphic function hence would be constant, but it is not constant because it has a pole at i and a zero at ρ.
    - The graded algebra statement then follows immediately, since $\mathcal{M}_{2k}$ has a basis given by the elements $G_4^a G_6^b$ that lie in it.

- By using (10), we can obtain some nontrivial identities between Eisenstein series.
    - For example, since $G_8 \in \mathcal{M}_8 = \mathbb{C}G_4^2$ we see that $G_8 = cG_4^2$ for some constant c.
    - Evaluating at ∞ shows that $c = \dfrac{2\zeta(8)}{4\zeta(4)^2} = \dfrac{\pi^8/4725}{\pi^8/2025} = \dfrac{3}{7}$, so we obtain the quite nontrivial fact that $G_8 = \dfrac{3}{7}G_4^2$. (All of the obvious approaches to proving this identity directly are quite messy; e.g., directly via the lattice definition, or via the q-expansions.)
    - Exercise: Using $\zeta(10) = \pi^{10}/93555$, prove that $G_{10} = 5G_4G_6/11$.

- Let us now establish some properties of the modular j-invariant, which will ultimately allow us to prove the uniformization theorem:

- Definition: Let $\tau \in \mathbb{H}$. The modular j-invariant is $j(\tau) = 1728\dfrac{g_2(\tau)^3}{\Delta(\tau)}$.
    - Of course, this is simply the j-invariant of the elliptic curve $y^2 = 4x^3 - g_2(\tau)x - g_3(\tau)$ associated to the lattice $\Lambda = \mathbb{Z} + \mathbb{Z}\tau$.

- Proposition (Properties of $j(\tau)$): Let j be the modular j-invariant defined on the upper-half plane $\mathbb{H}$.
    1. The function $j(\tau)$ is a weakly modular function of weight 0 that has a simple pole at ∞.
        - Proof: As we showed in the proposition above, Δ is nonvanishing on $\mathbb{H}$ and g_2 is holomorphic, so the ratio g_2^3/Δ is also holomorphic.
        - Additionally, since g_2^3 and Δ both have weight 12, the ratio has weight 0.
        - Finally, since g_2 is nonvanishing at ∞ and Δ has a simple zero, we see that j has a simple pole at ∞.
    2. The function j induces a complex analytic isomorphism $j : X(1) \to \mathbb{P}^1(\mathbb{C})$.
        - Proof: This is just a rephrasing of (1), since j is weakly modular and meromorphic on $\mathbb{H}^*$, this means j is a well-defined meromorphic function on $\Gamma(1)\backslash\mathbb{H}^* = X(1)$.
        - But since the total pole order of j is equal to 1, this means $j : X(1) \to \mathbb{P}^1(\mathbb{C})$ is an analytic map of degree 1 between compact Riemann surfaces, meaning it is an isomorphism.
    3. The field of modular functions of weight 0 is equal to $\mathbb{C}(j)$, and the ring of modular functions holomorphic on $\mathbb{H}$ of weight 0 is equal to $\mathbb{C}[j]$.
        - Proof: Suppose f is a modular function of weight 0. Then $f \circ j^{-1} : \mathbb{P}^1(\mathbb{C}) \to \mathbb{P}^1(\mathbb{C})$ is also meromorphic.
        - But by standard results of complex analysis, the only meromorphic functions on the Riemann sphere $\mathbb{P}^1(\mathbb{C})$ are rational functions, so $f \circ j^{-1}(z) = r(z)$ is some rational function in z.
        - Setting $\tau = j^{-1}(z)$ yields $f(\tau) = r(j(\tau))$, meaning f is a rational function in j, as claimed.
        - For the second part, we observe that any rational function with nonconstant denominator must have a zero since $\tau : X(1) \to \mathbb{P}^1(\mathbb{C})$ is onto, so non-polynomial rational functions of j all have at least one pole.

4. (Uniformization) For any $A, B \in \mathbb{C}$ with $A^3 - 27B^2 \neq 0$, there exists a unique complex lattice Λ such that $g_2(\Lambda) = A$ and $g_3(\Lambda) = B$: specifically, it is the period lattice for the elliptic curve $E : y^2 = 4x^3 - Ax - B$.
   - Proof: By (2), since j is onto and $A^3 - 27B^2 \neq 0$, there exists a $\tau \in \mathbb{H}$ such that $j(\tau) = 1728\frac{A^3}{A^3 - 27B^2}$.
   - We will then construct an α such that $\Lambda = \mathbb{Z}\alpha + \mathbb{Z}\alpha\tau$. Note that $g_2(\Lambda) = \alpha^{-4}g_2(\tau)$ and $g_3(\Lambda) = \alpha^{-6}g_3(\tau)$ by the lattice scaling property.
   - If $A = 0$ then $j(\tau) = 0$ so $g_2(\tau) = 0$. We take $\alpha = (g_3(\tau)/B)^{1/6}$: then $g_2(\Lambda) = \alpha^{-4}g_2(\tau) = 0 = A$ and $g_3(\Lambda) = \alpha^{-6}g_3(\tau) = B$.
   - If $B = 0$ then $j(\tau) = 1728$ so $g_3(\tau) = 0$. We take $\alpha = (g_2(\tau)/A)^{1/4}$: then $g_2(\Lambda) = \alpha^{-4}g_2(\tau) = A$ and $g_3(\Lambda) = \alpha^{-6}g_3(\tau) = 0 = B$.
   - If $AB \neq 0$ then $\frac{27g_3(\tau)^2}{4g_2(\tau)^3} = \frac{1728}{j(\tau)} - 1 = \frac{27B^2}{4A^3}$ and so $(g_3(\tau)/B)^{1/6} = (g_2(\tau)/A)^{1/4}$. Taking α to be this quantity yields $g_2(\Lambda) = \alpha^{-4}g_2(\tau) = A$ and $g_3(\Lambda) = \alpha^{-6}g_3(\tau) = B$.
   - It is also straightforward to check that these are the only possible choices of α and τ that will work here.
5. The Fourier expansion of j is $\tilde{j}(q) = q^{-1} + 744 + 196884q + 21493760q^2 + \cdots$ where all of the coefficients are integers.
   - In particular, this expansion gives another reason for the appearance of the scaling constant 1728 in the formula for j: namely, it makes the residue at ∞ equal to 1.
   - The first few coefficients can be worked out explicitly using the q-expansions of the Eisenstein series calculated earlier. (We will not do this calculation explicitly, since it is quite messy.)
   - The fact that the coefficients c_n are integral is more difficult, and in fact they have very many interesting divisibility properties: for example, if $2^a|n$ then $2^{3a+8}|c_n$, and if $3^a|n$ then $3^{2a+3}|c_n$ also.
   - A rather stunningly unexpected observation, first made by McKay in the late 1970s, is that the smallest nontrivial representation of the monster group has dimension 196883, and the second-smallest representation has dimension 21296875 (note in particular that $1 + 196883 + 21296875 = 21493760$).
   - Conway, Norton, and Thompson conjectured that there existed a graded representation of the monster group on an appropriate modular function field arising from the quotient of $\mathbb{H}^*$ (minus some number of points) by an appropriate group. The coefficients of the j-invariant then arise as the traces of low-degree components of this representation.
   - The existence of this "moonshine module" was eventually proven by Borcherds in 1992, thereby establishing this very surprising connection between the j-invariant and the sporadic simple groups. (The name "moonshine" was coined by Conway, who initially thought that the idea that there could be any connection between the j-invariant and representation dimensions was truly outlandish!)
   - We will mention another interesting numerical coincidence related to the j-invariant: as $\tau = (1 + \sqrt{-163})/2$ is a quadratic integer lying in the ring of integers of $\mathbb{Q}(\sqrt{-163})$, which has class number 1, the associated j-invariant is integral.
   - For $\tau = (1 + \sqrt{-163})/2$ we have $q = e^{2\pi i\tau} = -e^{-\pi\sqrt{163}}$, and so $j(-e^{-\pi\sqrt{163}})$ is an integer.
   - But evaluating the q-expansion yields $\tilde{j}(-e^{-\pi\sqrt{163}}) = -e^{\pi\sqrt{163}} + 744 - 196884e^{-\pi\sqrt{163}} + \cdots$, and so since the later terms are all very small, we see that $e^{\pi\sqrt{163}}$ is very close to an integer: indeed, evaluating it numerically yields $e^{\pi\sqrt{163}} \approx 262537412640768744 - 7.5 \cdot 10^{-13}$.
   - The observation that $e^{\pi\sqrt{163}}$ is very close to an integer is often attributed to Ramanujan (though it does not appear in any of his works or notebooks), but was actually first noted by Hermite. This number also appeared in an April Fool's joke (claiming that Ramanujan had predicted it was actually an integer) in Gardner's mathematical games column in Scientific American in 1975.

0.25 (Dec 7) Modularity and Fermat's Last Theorem

- One of the main results of our discussion is that the j-invariant parametrizes the isomorphism classes of elliptic curves over $\mathbb{C}$.

- We have seen this through both from our algebraic perspective (using explicit isomorphisms of curves) and the analytic perspectives (using the modular j-invariant).
- In the modular lens, more precisely, we showed that the modular function j yields an analytic isomorphism $j : X(1) \to \mathbb{P}^1(\mathbb{C})$. By removing the point at ∞ in both spaces we equivalently get an analytic isomorphism from $Y(1) = \Gamma(1)\backslash\mathbb{H}$ to $\mathbb{C}$.
- The whole motivation to begin with was to find an analytic way to describe the moduli space of all elliptic curves up to isomorphism, which we equivalently phrased in terms of characterizing lattices up to homothety.

- What we would like to do now is broaden our perspective to construct moduli spaces for other objects of interest.

  - As we spent much time discussing, an elliptic curve is a genus 1 curve together with a marked point serving as the identity element in its group of points (i.e., a point of order 1).
  - What if we instead wanted to parametrize curves of genus 1 together with a marked point of order N, for some $N \geq 1$? Or, alternatively, what if we wanted to parametrize curves of genus 1 together with a cyclic subgroup of order N? (The difference between these two situations is that in the first case, we have a specific generator for the cyclic subgroup, while in the second we only know the subgroup itself.) Or we could even seek to parametrize curves together with a pair of generators of the N-torsion group.
  - Suppose we want to study pairs (E, P) where P is a point of order N. Then the appropriate notion of equivalence $(E, P) \sim (E', P')$ is for there to exist an isomorphism $\varphi : E \to E'$ with $\varphi(P) = P'$. On the level of lattices, we have pairs (Λ, P) where $P \in \frac{1}{N}\Lambda/\Lambda$ up to corresponding homothety: a pair (Λ, P) is equivalent to (Λ', P') when there exists a scaling α with $\alpha\Lambda = \Lambda'$ and $\alpha P = P'$.
  - Likewise, to study pairs (E, C) where C is a cyclic subgroup of order N, then the appropriate notion of equivalence $(E, C) \sim (E', C')$ is for there to exist an isomorphism $\varphi : E \to E'$ such that $\varphi(C) = C'$. On the level of lattices, we have pairs (Λ, C) where C is a cyclic subgroup of $\frac{1}{N}\Lambda/\Lambda$ of order N, and with (Λ, C) equivalent to (Λ', C') when there exists a scaling α with $\alpha\Lambda = \Lambda'$ and $\alpha C = C'$.
  - Finally, to study pairs $(E, (P, Q))$ where (P, Q) generate $E[N]$, the equivalence is an isomorphism $\varphi : E \to E'$ with $\varphi(P) = P'$ and $\varphi(Q) = Q'$. For lattices, the pairs are $(\Lambda, (P, Q))$ with $\langle P, Q\rangle = \frac{1}{N}\Lambda/\Lambda$ with $(\Lambda, (P, Q))$ is equivalent to $(\Lambda', (P', Q'))$ when there exists a scaling α with $\alpha\Lambda = \Lambda'$, $\alpha P = P'$, and $\alpha Q = Q'$.
  - Since each of the equivalence classes above requires a scaling with $\alpha\Lambda = \Lambda'$, we may rescale our lattices again to be of the form $\mathbb{Z} + \mathbb{Z}\tau$ for some $\tau \in \mathbb{H}$: then we seek to understand which values of τ yield equivalent pairs (Λ, C), (Λ, P), and $(\Lambda, (P, Q))$.
  - In the first case, we can change our lattice's basis so that C is the subgroup generated by $1/N$ inside $\mathbb{Z} + \mathbb{Z}\tau$, in the second case we can change basis so that $P = 1/N \bmod \Lambda$, and in the third case we can change basis so that $P = 1/N \bmod \Lambda$ and $Q = \tau/N \bmod \Lambda$.
  - Since the corresponding action must still preserve the lattice $\mathbb{Z} + \mathbb{Z}\tau$, the resulting actions on τ must all still lie in $\Gamma(1)$, but not all such maps will preserve the additional data of the subgroup C, the point P, or the ordered basis $\{1/N, \tau/N\}$.

- The relevant matrix groups are as follows:

- Definition: Let N be a positive integer. We define the following subgroups of $SL_2(\mathbb{Z})$, where $*$ indicates the value may be arbitrary.

  1. $\Gamma(N) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SL_2(\mathbb{Z}) : \begin{bmatrix} a & b \\ c & d \end{bmatrix} \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \pmod N \right\}$, the principal congruence subgroup of level N.
  2. $\Gamma_1(N) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SL_2(\mathbb{Z}) : \begin{bmatrix} a & b \\ c & d \end{bmatrix} \equiv \begin{bmatrix} 1 & * \\ 0 & 1 \end{bmatrix} \pmod N \right\}$
  3. $\Gamma_0(N) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SL_2(\mathbb{Z}) : \begin{bmatrix} a & b \\ c & d \end{bmatrix} \equiv \begin{bmatrix} * & * \\ 0 & * \end{bmatrix} \pmod N \right\}$

- Notice that $\Gamma(N)$ is the kernel of the reduction-mod-N map $\Gamma(1) \to SL_2(\mathbb{Z}/N\mathbb{Z})$, and so $\Gamma(N)$ is a normal subgroup of finite index in $\Gamma(1)$. Indeed, the projection map is surjective, as follows from a straightforward Euclidean algorithm calculation, so the index $[\Gamma(1) : \Gamma(N)] = \#SL_2(\mathbb{Z}/N\mathbb{Z}) = N^3 \prod_{p|N}(1 - p^{-2})$.
- Since $\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N)$ we see that $\Gamma_0(N)$ and $\Gamma_1(N)$ also have finite index in $\Gamma(1)$.
- Exercise: Calculate $[\Gamma(1) : \Gamma_1(N)]$ and $[\Gamma(1) : \Gamma_0(N)]$.

- Proposition (Moduli Spaces): Let N be a positive integer and let $\Lambda_\tau = \mathbb{Z} + \mathbb{Z}\tau$ for $\tau \in \mathbb{H}$.

1. The moduli space parametrizing pairs (E, P) up to equivalence, where E is an elliptic curve and P is a point of order N is the set of pairs $(\Lambda_\tau, P) = (\Lambda_\tau, 1/N + \Lambda_\tau)$ under the equivalence $(\Lambda_\tau, P) \sim (\Lambda_{\tau'}, P')$ when $\tau' \in \Gamma_1(N)\tau$. The corresponding moduli space $Y_1(N)$ is therefore isomorphic to the quotient space $\Gamma_1(N)\backslash\mathbb{H}$.
   - Proof: Suppose $(\Lambda_\tau, P) \sim (\Lambda_{\tau'}, P')$: then $\tau' = \gamma\tau = \dfrac{a\tau + b}{c\tau + d}$ for some $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SL_2(\mathbb{Z})$ and $\alpha = c\tau + d$. In such a case we have $P' = (c\tau + d)/N + \Lambda_{\tau'}$, so in order for this to equal $1/N + \Lambda_{\tau'}$ we require $c \equiv 0 \pmod N$ and $d \equiv 1 \pmod N$. Then $\det\gamma = 1$ requires $a \equiv 1 \pmod N$ while b can be arbitrary, so $\gamma \in \Gamma_1(N)$.
   - Conversely, when $\gamma \in \Gamma_1(N)$ we do have $P' = 1/N + \Lambda_{\tau'}$, so $(\Lambda_\tau, P) \sim (\Lambda_{\tau'}, P')$.
   - The second statement is immediate, since the calculation above shows that the equivalence classes are precisely the right cosets of $\Gamma_1(N)$ acting on $\mathbb{H}$.
2. The moduli space parametrizing pairs (E, C) up to equivalence, where E is an elliptic curve and C is a cyclic subgroup of order N is the set of pairs $(\Lambda_\tau, C) = (\Lambda_\tau, \langle 1/N\rangle + \Lambda_\tau)$ under the equivalence $(\Lambda_\tau, C) \sim (\Lambda_{\tau'}, C')$ when $\tau' \in \Gamma_0(N)\tau$. The corresponding moduli space $Y_0(N)$ is therefore isomorphic to the quotient space $\Gamma_0(N)\backslash\mathbb{H}$.
   - Proof: As in (1) we have $\tau' = \gamma\tau = \dfrac{a\tau + b}{c\tau + d}$ for some $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SL_2(\mathbb{Z})$ and $\alpha = c\tau + d$. Then $C' = \langle (c\tau + d)/N\rangle + \Lambda_{\tau'}$, so in order for this to equal $\langle 1/N\rangle + \Lambda_{\tau'}$ we require $c \equiv 0 \pmod N$ and d to be invertible modulo N. But the latter condition follows automatically from $\det\gamma = 1$ since if $c \equiv 0 \pmod N$, then $\det\gamma \equiv ad \pmod N$: thus, we have $\gamma \in \Gamma_0(N)$.
   - Conversely, when $\gamma \in \Gamma_0(N)$ we do have $C' = \langle 1/N\rangle + \Lambda_{\tau'}$, so $(\Lambda_\tau, P) \sim (\Lambda_{\tau'}, P')$. The second statement is immediate as in (1).
3. The moduli space parametrizing pairs $(E, (P, Q))$ up to equivalence, where E is an elliptic curve and (P, Q) is an ordered basis for $E[N]$ is the set of pairs $(\Lambda_\tau, (P, Q)) = (\Lambda_\tau, (1/N + \Lambda_\tau, \tau/N + \Lambda_\tau))$ under the equivalence $(\Lambda_\tau, (P, Q)) \sim (\Lambda_{\tau'}, (P', Q'))$ when $\tau' \in \Gamma(N)\tau$. The corresponding moduli space $Y(N)$ is therefore isomorphic to the quotient space $\Gamma(N)\backslash\mathbb{H}$.
   - Proof: As in (1) we have $\tau' = \gamma\tau = \dfrac{a\tau + b}{c\tau + d}$ for some $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SL_2(\mathbb{Z})$ and $\alpha = c\tau + d$. Then $P' = (c\tau + d)/N + \Lambda_{\tau'}$ and $Q' = (a\tau + b)/N + \Lambda_{\tau'}$, so in order for these to equal $1/N + \Lambda_{\tau'}$ and $\tau/N + \Lambda_{\tau'}$ respectively, we require $c \equiv 0 \pmod N$, $d \equiv 1 \pmod N$, $a \equiv 1 \pmod N$, and $b \equiv 0 \pmod N$, so that $\gamma \in \Gamma(N)$.
   - Conversely, when $\gamma \in \Gamma(N)$ then clearly equivalence follows, and the second statement is immediate as in (1) and (2).

- As in the case of $Y(1) = \Gamma(1)\backslash\mathbb{H}$, the moduli spaces $Y(N)$, $Y_0(N)$, and $Y_1(N)$ all carry a natural Riemann surface structure owing to their construction as a quotient space of $\mathbb{H}$ by the subgroup $\Gamma(N)$, $\Gamma_0(N)$, or $\Gamma_1(N)$ respectively.
   - We can construct these Riemann surfaces geometrically by noting that each subgroup has a fundamental domain obtained as a union of some copies of the fundamental domain D of $\Gamma(1)$: specifically, the union over an arbitrary choice of coset representatives $\cup_{\gamma \in \Gamma(1)/G} \gamma D$, with (as usual) appropriate identifications made along all of the boundaries.
   - Also as with $Y(1)$, these Riemann surfaces are not compact: we can compactify them by working instead with the quotient spaces $G\backslash\mathbb{H}^*$ for $G = \Gamma(N)$, $\Gamma_0(N)$, $\Gamma_1(N)$, yielding the respective moduli spaces $X(N)$, $X_0(N)$, and $X_1(N)$ respectively.

- By general facts about Riemann surfaces (specifically, GAGA: the principle that complex-analytic maps between compact Riemann surfaces are actually defined by algebraic equations), these moduli spaces are actually algebraic curves over $\mathbb{C}$, but just as in the particular case $X(1) \cong \mathbb{P}^1(\mathbb{C})$, much stronger statements are true.
- Specifically, $X_0(N)$ and $X_1(N)$ are both algebraic curves over $\mathbb{Q}$, while $X(N)$ is an algebraic curve over $\mathbb{Q}(\zeta_N)$.
- To show these facts requires working out what the function fields $\mathbb{C}(X(G))$ look like for each of these congruence subgroups G. We have already shown that $\mathbb{C}(X(1)) = \mathbb{C}(j)$ where j is the j-invariant, and so it suffices only to work out what the extensions $\mathbb{C}(X(N))/\mathbb{C}(X(1))$ are.
- After unwinding the ideas appropriately, in fact all of this follows from the moduli space descriptions of $X(N)$ and $X(1)$ above: the extension $\mathbb{C}(X(N))$ is equal to $\mathbb{C}(j, x(E_j[N]))$ where E_j is the (universal) Tate curve $y^2 = 4x^3 - \frac{27j}{j-1728}x - \frac{27j}{j-1728}$ with j-invariant j.
- Then $\mathbb{C}(X(N))/\mathbb{C}(X(1))$ is Galois with Galois group $\mathrm{Aut}(E_j[N]) \cong SL_2(\mathbb{Z}/N\mathbb{Z})/\{\pm I\}$, which (more or less) is just a rephrasing of the original description of $X(N)$ as the moduli space $\Gamma(N)\backslash\mathbb{H}^*$ and the fact that $X(1)/X(N) \cong SL_2(\mathbb{Z}/N\mathbb{Z})/\{\pm I\}$.
- Roughly speaking, the idea is that the j-invariant parametrizes all elliptic curves, and then the additional data carried in $X(N)$ by the generators of the N-torsion subgroup causes the functions in the kernel of the multiplication-by-N map (i.e., the elements of $E_j[N]$) also to be well-defined on $X(N)$.
- Now, we can try to play the same game with base field $\mathbb{Q}$ instead of $\mathbb{C}$: the idea is that $\mathbb{Q}(j, E_j[N])/\mathbb{Q}(j)$ is also Galois with the same Galois group, and so the corresponding function field extension yields a natural candidate for the function field of $X(N)$.
- By using the Weil pairing, one may then show that the resulting field extension describing $X(N)$ is (isomorphic) to a subfield of $\mathbb{Q}(\zeta_N)$, and that $X_0(N)$ and $X_1(N)$ are both actually defined over $\mathbb{Q}$.

- Now that $X_0(N)$, $X_1(N)$, and $X(N)$ are compact Riemann surfaces, in analogy with what we did for $X(1)$, we can consider meromorphic functions on these surfaces.
  - As with $X(1)$, the best thing to do is to consider the broader family of (weakly) modular forms with respect to the subgroups $\Gamma_0(N)$, $\Gamma_1(N)$, and $\Gamma(N)$.

- Definition: Let Γ be a congruence subgroup of $SL_2(\mathbb{Z})$. A weakly modular function of weight k with respect to Γ is a meromorphic function f on $\mathbb{H}$ such that $f(\gamma\tau) = (c\tau + d)^k f(\tau)$ for all $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \Gamma$.
  - When Γ contains $\Gamma(N)$, we say that f is weakly modular of weight k and level N.
  - Because $\begin{bmatrix} 1 & N \\ 0 & 1 \end{bmatrix} \in \Gamma(N)$, all weakly modular functions of level N are periodic with period N, hence has a Fourier expansion with respect to $q_N = e^{2\pi i\tau/N}$ of the form $\tilde{f}(\tau) = \sum_{n=-\infty}^{\infty} a_n q_N^n$.
  - We say that f is meromorphic at ∞ if the expansion is a Laurent expansion (starting at $n = k$ for some k), and holomorphic at ∞ if the expansion is a power series (starting at $n = 0$).
  - With $\Gamma(1)$, all points of $\mathbb{P}^1(\mathbb{Q})$ were $\Gamma(1)$-equivalent to ∞, but with a proper subgroup Γ, there may be others that are not Γ-equivalent to ∞. In order to have the proper analogy, a modular form with respect to Γ must also be holomorphic at all cusps (Γ-equivalence classes of points in $\mathbb{P}^1(\mathbb{Q})$).

- Definition: Let Γ be a congruence subgroup of $SL_2(\mathbb{Z})$. We say a weakly modular function f of weight k with respect to Γ is a modular form if f is holomorphic on $\mathbb{H}$ and at all cusps, the latter condition meaning that $f(\alpha\tau)$ is holomorphic at ∞ for all $\alpha \in SL_2(\mathbb{Z})/\Gamma$. If in addition f vanishes at all cusps, we say f is a cusp form.
  - We note that the cusp condition only needs to be checked for a finite set of α, namely, any set of coset representatives for Γ in $SL_2(\mathbb{Z})$.
  - As with $\Gamma(1)$, the modular forms $\mathcal{M}_k(\Gamma)$ and cusp forms $\mathcal{S}_k(\Gamma)$ of weight k are vector spaces that fit into graded algebras.

- For proper subgroups of $SL_2(\mathbb{Z})$, we will generally end up with more modular forms than with $\Gamma(1) = SL_2(\mathbb{Z})$ itself.
- For instance, for any $N \geq 2$, the conditionally convergent Eisenstein series $G_2(\tau) = \sum_{m\in\mathbb{Z}} \sum_{n\in\mathbb{Z}'} \frac{1}{(m\tau+n)^2}$ is not weakly modular, but the modified series $G_{2,N}(\tau) = G_2(\tau) - NG_2(N\tau)$ is modular of weight 2 for $\Gamma_0(N)$.
- There is very, very much to say about general modular forms and cusp forms of various levels, which we will not be able to discuss now, aside from mentioning that many of the results we showed for $\Gamma(1)$ extend fairly nicely (e.g., the weight-counting formula leading to a dimension formula for $\mathcal{M}_k$, which can be obtained by using Riemann-Hurwitz on the natural covering map from $\Gamma\backslash\mathbb{H}^*$ to $\Gamma(1)\backslash\mathbb{H}^*$).
- As in the case of $\Gamma(1)$, a natural problem is to construct a convenient basis for the spaces of modular forms and cusp forms of weight k for Γ.
- One way to do this is to use the Hecke operators $\langle l\rangle$ and T_l, which are commuting endomorphisms of $\mathcal{M}_k(\Gamma_1(N))$ which also preserve cusp forms.
- To construct these, note first that $\Gamma_1(N)$ is a normal subgroup of $\Gamma_0(N)$ and the map sending $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ to d is an isomorphism of the quotient with $(\mathbb{Z}/N\mathbb{Z})^\times$. The weight-$k$ action of $\Gamma_0(N)$ preserves $\mathcal{M}_k(\Gamma_1(N))$ and $\mathcal{S}_k(\Gamma_1(N))$, hence we get an action of the quotient group on each.
- For $d \in (\mathbb{Z}/N\mathbb{Z})^\times$ we write the corresponding automorphism of $\mathcal{M}_k(\Gamma_1(N))$ as $\langle d\rangle$, and then when l is a prime not dividing N we define the automorphism $S_l = l^{k-2}\langle l\rangle$.
- We also define the operator T_l on $\mathcal{M}_k(\Gamma_1(N))$ via $\widetilde{T_l f}(q) = \sum_{n=0}^\infty c_{ln}(f)q^n + \sum_{n=0}^\infty l c_n(S_l f)q^{ln}$. It's not at all obvious that this preserves $\mathcal{M}_k(\Gamma_1(N))$ and $\mathcal{S}_k(\Gamma_1(N))$, but it does: geometrically speaking, T_l is obtained as a sum over all of the double cosets of $\Gamma_1(N)$.
- One can also define a natural inner product on the space of cuspforms, called the Petersson inner product: with hyperbolic measure $d\mu(\tau) = y^{-2}dx\,dy$ (for $\tau = x+iy$) and $f, g \in \mathcal{S}_k(\Gamma)$, we take $\langle f.g\rangle_\Gamma = \frac{1}{V_\Gamma}\int_{\Gamma\backslash\mathbb{H}^*} f(\tau)\overline{g(\tau)}(\mathrm{im}\tau)^k d\mu(\tau)$ where $V_\Gamma = \int_{\Gamma\backslash\mathbb{H}^*} d\mu(\tau)$ is the volume of the fundamental domain. (The normalization ensures that the inner product is consistent when changing groups, and the integral converges because the product fg vanishes at each cusp.)
- Then the Hecke operators are normal operators with respect to the Petersson inner product, so since they all commute with one another, by an invocation of the spectral theorem, one then immediately sees that $S_k(\Gamma)$ has an orthonormal basis of Hecke eigenforms.
- We make one final observation using the Hecke operators: if $f \in \mathcal{M}_k(\Gamma_1(N), \chi)$ is a Hecke eigenform and has a Fourier expansion $\tilde{f}(q_N) = \sum_{n=0}^\infty a_n(f)q_N^n$ with $a_1(f) = 1$, then $a_{mn}(f) = a_m(f)a_n(f)$ when $\gcd(m,n) = 1$, and $a_{p^r}(f) = a_p(f)a_{p^r}(f) - \chi(p)p^{k-1}a_{p^{r-2}}(f)$ for all primes p and all $r \geq 2$.

- The Modularity Theorem can then be phrased as follows:

- Theorem (Modularity Theorem, Morphisms): Let $E/\mathbb{C}$ be an elliptic curve with rational j-invariant. Then for some positive integer N there exists an onto morphism of algebraic curves $X_0(N) \to E$ defined over $\mathbb{Q}$.
  - The function in this theorem gives what is called a modular parametrization of E. The smallest such N is called the conductor of E.
  - This formulation of the modularity theorem, while fairly direct, does not give any indication why such a morphism should exist, nor why it would be so useful, so we will now shift direction to give another formulation in terms of the traces of Frobenius $a_p(E)$ for primes p, packaged together using the L-function, that highlights the number-theoretic content of this statement.

- So, suppose that E is an elliptic curve with rational j-invariant.
  - Since E is isomorphic to a curve with rational Weierstrass coefficients (namely, the Tate curve with the same j-invariant), we may simply replace E with that curve, and by rescaling we may in fact take the Weierstrass coefficients to be integers: say $y^2+a_1xy+a_3y = x^3+a_2x^2+a_4x+a_6$ for $a_1, a_2, a_3, a_4, a_6 \in \mathbb{Z}$.
  - Consider two such equations to be equivalent if they are related by a rational change of variables.

- For each prime p let $v_p(E)$ be the minimal p-adic valuation of the discriminant of any Weierstrass equation equivalent to the one for E. If we define the global minimal discriminant of E as $\Delta_{\min}(E) = \prod_p p^{v_p(E)}$, then (by a somewhat involved calculation that essentially reduces to the Chinese remainder theorem) there exists a change of variables that minimizes the p-adic valuation of Δ simultaneously for all p.
- Example: The rational elliptic curve $y^2 = x^3 - 1323x + 97254$ with $\Delta = -2^{12}3^{15}67$ is isomorphic to the curve $y^2 + xy = x^3 - x + 2$ with $\Delta = -3^3 67$. This latter model is minimal since any rational change of variable rescales Δ by rational twelfth powers, so the minimal possible discriminant would be $-3^3 67$.

- Now assume that we have written E using a global minimal Weierstrass equation with discriminant Δ. We can then consider the reduction of E modulo p: when $p \nmid \Delta$ the reduction is a nonsingular elliptic curve (we say that E has good reduction modulo p), and when $p|\Delta$ the reduction is a singular curve (we say that E has bad reduction modulo p).
  - We have three different types of bad reduction, depending on the type of singularity:
    * The group of points on $E/\mathbb{F}_p$ is isomorphic to the multiplicative group $\mathbb{F}_p^*$, in which case we say E has split multiplicative reduction modulo p. This occurs when the singularity of E is a node whose double tangent line is defined over $\mathbb{F}_p$.
    * The group of points on $E/\mathbb{F}_p$ is isomorphic to the multiplicative group of $(p+1)$st roots of unity in $\mathbb{F}_{p^2}^*$, in which case we say E has nonsplit multiplicative reduction modulo p. This occurs when the singularity of E is a node whose double tangent line is not defined over $\mathbb{F}_p$ (its slope is then in $\mathbb{F}_{p^2}$).
    * The group of points on $E/\mathbb{F}_p$ is isomorphic to the additive group $\mathbb{F}_p$, in which case we say E has additive reduction modulo p. This occurs when the singularity of E is a cusp (i.e., with a triple tangent line).
  - Using the behavior at the primes of bad reduction, we can give a formula for the conductor of E: it is $N = \prod_p p^{f_p}$ where f_p is 0 at primes of good reduction, 1 at primes of multiplicative reduction, and 2 at primes of additive reduction (plus an additional bounded term for $p = 2$ and $p = 3$, whose behavior is somewhat more complicated due to possible wild ramification there).

- Now we can define the L-function associated to an elliptic curve.
  - Recall from our proof of the Weil conjectures that for $E/\mathbb{F}_p$ the zeta function $Z_E(T) = \exp(\sum_{n=1}^{\infty} \#E(\mathbb{F}_{p^n})T^n/n)$ is a rational function of the form $\dfrac{L_p(T)}{(1-T)(1-pT)}$ where $L_p(T) = (1-\alpha T)(1-\beta T) = 1 - a_p T + pT^2$ for $a_p = \alpha + \beta = \mathrm{tr}(\varphi_p) = p + 1 - \#F(\mathbb{F}_p)$ equal to the trace of Frobenius, and where α, β are complex conjugates of absolute value $\sqrt{p}$.
  - We now extend the definition of this local factor $L_p(T)$ to the situation where E has bad reduction at p by taking $L_p(T) = \begin{cases} 1-T & \text{when } E \text{ has split multiplicative reduction} \\ 1+T & \text{when } E \text{ has nonsplit multiplicative reduction modulo } p. \\ 1 & \text{when } E \text{ has additive reduction} \end{cases}$
  - In fact, from the calculations noted above for the point counts in the three cases, with $a_p = p + 1 - \#E_{ns}(\mathbb{F}_p)$, we still actually have $L_p(T) = 1 - a_p T + \chi_{\mathrm{triv},\Delta}(p)T^2$, where $\chi_{\mathrm{triv},\Delta}$ is the trivial character modulo Δ (which is 1 on integers relatively prime to Δ and 0 on integers not relatively prime to Δ).

- Definition: Let E be an elliptic curve over $\mathbb{Q}$ in global minimal Weierstrass form. The Hasse-Weil L-function associated to E is the Euler product $L(s,E) = \prod_p L_p(p^{-s})^{-1} = \prod_p (1 - a_p p^{-s} + \chi_{\mathrm{triv},\Delta}(p)p^{1-2s})^{-1}$, where L_p is the local factor described above, $a_p = p + 1 - \#E_{ns}(\mathbb{F}_p)$, and $T = p^{-s}$ is the corresponding local variable.
  - By expanding the Euler product, we may write it as a Dirichlet series $L(s,E) = \sum_{n=1}^{\infty} a_n(E) n^{-s}$ for some appropriate integer coefficients $a_n(E)$ obtained from the series expansions.
  - For primes of good reduction, the Euler factor is simply $(1-\alpha p^{-s})^{-1}(1-\beta p^{-s})^{-1} = \sum_{k=0}^{\infty} [\sum_{j=0}^{k} \alpha^j \beta^{k-j}] p^{-ks}$, and for primes of bad reduction the Euler factor is even simpler: it is either $\sum_{k=0}^{\infty} p^{-ks}$, $\sum_{k=0}^{\infty} (-1)^k p^{-ks}$, or 1.
  - From the Weil conjectures via the Hasse bound, we know that $|\alpha| = |\beta| = p^{1/2}$, and so the Euler product converges for $\mathrm{Re}(s) > 3/2$.

- By directly multiplying out the Euler factors we have $a_{mn}(E) = a_m(E)a_n(E)$ whenever $\gcd(m,n) = 1$, so in particular $a_1(E) = 1$.
- Additionally, using the explicit descriptions of the Euler factors above, it is also easy to check that $a_p(E) = a_p$ and also that $a_{p^r}(E) = a_p(E)a_{p^{r-1}}(E) - \chi_{\text{triv},\Delta}(p)pa_{p^{r-2}}(E)$ for each $r \geq 2$.
- And now the key: notice that these are *exactly* the same recurrence conditions as those on the coefficients of the q-expansion of a Hecke eigenform in $\mathcal{S}_2(\Gamma_0(N))$! (The fact that the eigenform is a cusp form is just saying that $a_0 = 0$ here.)
- This is, in fact, a rephrasing of the modularity theorem from before:

- Theorem (Modularity Theorem, L-Functions): Let $E/\mathbb{Q}$ be an elliptic curve with conductor N. Then there exists some Hecke eigenform $f \in \mathcal{S}_2(\Gamma_0(N))$ such that $a_n(f) = a_n(E)$ for all integers n, where $\tilde{f}(q_N) = \sum_{n=1}^{\infty} a_n q_N^n$ is the Fourier series for f and $L(s,E) = \sum_{n=1}^{\infty} a_n(E)n^{-s}$ is the L-series for E.
  - Let us attempt to give an extremely sketchy explanation of why the existence of an onto morphism $X_0(N) \to E$ defined over $\mathbb{Q}$ implies the existence of this Hecke eigenform with the same coefficients.
  - First, the existence of this morphism implies the existence of a Hecke eigenform $f \in \mathcal{S}_2(\Gamma_0(N))$, by very general facts about cusp forms of weight 2 for $\Gamma_0(N)$. Then because the coefficient recurrences are the same, it suffices to show that $a_p(f) = a_p(E)$ for primes p.
  - Now, if f is a Hecke eigenform, $a_p(f)$ is obtained by calculating $T_p f$, and this may in turn be computed inside the divisor group $\text{Pic}^0(X_0(N))$. One can then show that T_p acts on $\text{Pic}^0(X_0(N))$ via $\varphi_p + \widehat{\varphi_p}$ (this is essentially a result known as the Eichler-Shimura relation), and then passing from $\text{Pic}^0(X_0(N))$ to $\text{Pic}^0(E) \cong E$ yields $\text{tr}(\varphi_p) = a_p(E)$.

- Since it will lead us into our historical application of modularity, we will mention one other formulation involving Galois representations:

- Definition: A d-dimensional l-adic Galois representation is a continuous homomorphism $\rho : \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to GL_d(L)$ where L is a finite-degree extension of the l-adic rational field $\mathbb{Q}_l$.
  - We view two such representations ρ, ρ' as equivalent when there exists some matrix $A \in GL_d(L)$ for which $\rho'(\sigma) = A^{-1}\rho(\sigma)A$.
  - We mention also that any such field L is obtained as a completion of some number field $K/\mathbb{Q}$ at some prime ideal of its ring of integers lying above l.
  - We have previously obtained such representations arising from the Galois action on the Tate module of an elliptic curve.
  - We can also construct l-adic Galois representations in a similar manner by using the l-power torsion in the Picard group $\varprojlim_n \text{Pic}^0(X_1(N))[l^n]$, which is a rank-$2g$ $\mathbb{Z}_l$-module where g is the genus of $X_1(N)$. The action of the Hecke operators then decompose this module into g independent submodules of rank 2, corresponding to Hecke eigenforms, which allow us to associate a Hecke eigenform with an l-adic Galois representation.

- Theorem (Modularity Theorem, Representations): Let E be an elliptic curve over $\mathbb{Q}$ with conductor N. Then for all primes l and some Hecke eigenform $f \in \mathcal{S}_2(\Gamma_0(N))$ with rational Fourier coefficients, the associated l-adic Galois representation for f is equivalent to the l-adic Galois representation for E.
  - The connection between this version of modularity and the one for L-functions is that the local Euler factor is equal to the characteristic polynomial of the Frobenius map.
  - In fact, because the characteristic polynomial of Frobenius acting on $T_l(E)$ is independent of the choice of $l \neq p$ (as we proved during our discussion of the Weil conjectures), and an analogous statement holds on the modular form side, the theorem above for all l follows from the much weaker version asserting that the representations are equivalent for only a single value of l.
  - In fact, it is this "weaker" version that was proven for semistable curves by Taylor and Wiles in 1995 (fixing the gap in Wiles' earlier 1993 paper), and then for all elliptic curves by Breuil, Conrad, Diamond, and Taylor in 2001.

- Let us close our discussion by explaining the some of the history of modularity and how it is used in Wiles' proof of Fermat's conjecture.
  - In 1955, Taniyama stated a preliminary version of the modularity conjecture, and worked with Shimura to give a more precise improved statement, which became known as the Taniyama-Shimura conjecture. In the 1960s Weil gave reformulations and identified the level of the associated modular form as the conductor of the elliptic curve.
  - The connection of modularity to Fermat's conjecture is via the Frey-Hellegouarch curve, which was first studied by Hellegouarch in 1976 (although that work did not connect the curve to modularity). Frey constructed the same curve in 1982 and observed that it would have various unusual properties. These observations were extended by Serre in 1985 and Frey in 1986 indicating that this curve could give a counterexample to the Taniyama-Shimura conjecture.
  - Explicitly, suppose that we have a nonzero integer solution to $a^p + b^p + c^p = 0$ for p prime, and define the Frey-Hellegouarch curve $E : y^2 = x(x - a^p)(x + b^p)$, whose discriminant is $\Delta = a^p b^p c^p = (abc)^p$.
  - Then E necessarily has good reduction at all primes not dividing abc and multiplicative reduction at all primes dividing abc, meaning that E is semistable (the idea is that when E has additive reduction, then working over an extension field can change the reduction behavior[18], but when E has multiplicative reduction, that does not occur).
  - The number field generated by the p-torsion of E is ramified only at 2 and at p, and so the associated level of its mod-p Galois representation is 2. But $\mathcal{S}_2(\Gamma_0(2)) = 0$ because the corresponding modular curve $X_0(2)$ has genus 0, so there cannot be a modular form having the same p-adic Galois representation as E.
  - Serre's proof that Taniyama-Shimura would imply Fermat's conjecture was incomplete, and the missing portion became known as the epsilon conjecture. The epsilon conjecture was proven by Ribet in 1990, and so the full argument that Taniyama-Shimura implies Fermat's conjecture had been completed.
  - In 1993, after working in secret for six years on Taniyama-Shimura, Wiles announced that he had proven the semistable case of the conjecture, which (by the previous results) would establish Fermat's conjecture. A gap was discovered during the subsequent peer review, but after an additional year's work in collaboration with Taylor, a corrected proof was announced and published in 1995.
  - Wiles and Taylor-Wiles consider the mod-3 Galois representation $\overline{\rho}_{E,3} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to GL_2(\mathbb{F}_3)$. If this representation is irreducible, then by results of Langlands and Tunnell, the 3-adic representation arises from a cusp form, and then (here is the extremely technical and difficult part) if the mod-3 representation is modular, then the 3-adic representation is modular.
  - The technical hypotheses are met when E is semistable, and so a contradiction arises if the mod-3 Galois representation is irreducible. If it is not, then (with some substantial additional effort) Taylor and Wiles show that the mod-5 Galois representation can be used instead.
  - The contribution of Breuil, Conrad, Diamond, and Taylor is to remove the semistability requirement, thereby showing that the argument works for all elliptic curves.

[18] For example, $y^2 = x^3 - 9x$ has additive reduction at $p = 3$ over $\mathbb{Q}$, over $\mathbb{Q}(\sqrt{-3})$ it is isomorphic to $y^2 = x^3 - x$ which has good reduction at the ramified prime $P = (1 + e^{i\pi/3})$ lying above 3.