On Computing the Rank of Elliptic Curves

Jeff Achter

May 1992

> Where man looks up, and proud to claim
> His rank within the social frame,
> Sees a grand system round him roll,
> Himself its centre, sun and soul!
> Far from the shocks of Europe; far
> From every wild, elliptic star.
>
> — Thomas Moore, *Epistle II*.

The theory of elliptic curves is singular in the way it lies at the intersection of so many branches of mathematics. One can arrive at it by studying the sums of cubes, the theory of doubly-periodic meromorphic functions, or the Jacobians of varieties. One can use the resulting theory to answer the original questions, to test more general theorems of algebraic geometry, or to design public-key cryptosystems.

One of the fundamental arithmetic results is the Mordell-Weil theorem, which says that the group of K-rational points on an abelian variety, and in particular the group of $\mathbb{Q}$-rational points on an elliptic curve, is finitely generated. The goal of this paper is to outline the proof of this theorem for elliptic curves with a rational 2-torsion point; to show how the proof yields a partially effective algorithm for computing the rank of the group; and to discuss the results of this algorithm's implementation.

The theory developed here is embodied in a computer program which computes the rank of elliptic curves. I used this program to estimate the rank of over $100,000$ elliptic curves. Much has been made of Mazur's description of such a process, in which one computes descents by day and images by night [Tat]. However, I didn't have an abundance of days and nights in which to perform the computations. Therefore, I ran the computations simultaneously on about sixty SPARCstations. The results of the calculations match up well with those of Brumer and McGuinness [B-M].

This paper should be accessible to an undergraduate student of mathematics, such as myself. Most of the results described here, particularly those in the first section, are readily available in any number of sources. I have particularly relied on the forthcoming book of Silverman and Tate, *Rational Points on Elliptic Curves* [S-T]. In general, an easily-understood version of any theorem presented here may be found in that book. A readable but more erudite version can be found in either [Sil] or the expositions of Cassels [Cas 1] and Tate [Tat].

The first section is an introduction to some of the basic definitions and properties concerning elliptic curves. The second section presents some of the details of the proof of the Mordell-Weil theorem. Special attention is called to aspects of the proof which lead to a partially effective algorithm. The next section outlines a proof of a more general weak Mordell-Weil theorem. The fourth section describes some of the issues involved in implementing heuristics for computing the rank of the Mordell-Weil group. In the final section I present the results of the computations, and attempt to relate the numerical observations to certain conjectures regarding the rank of elliptic curves.

1 Introduction to Elliptic Curves

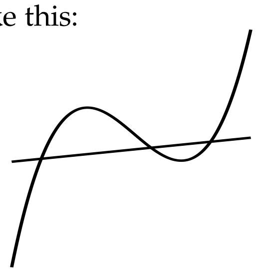
Elliptic curves can be defined in several different ways. One of the most concrete descriptions is that an elliptic curve is a nonsingular cubic in two variables. Such a curve may be described by an equation:

$$ax^3 + bx^2y + cxy^2 + dy^3 + ex^2 + fxy + gy^2 + hx + iy + j = 0$$

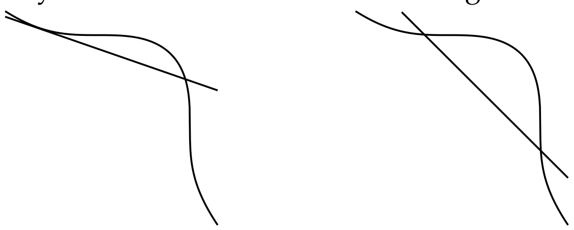
Naturally, the coefficients a through j can be taken from any field. In this discussion we will chiefly be concerned with the rational case, where all of the coefficients lie in $\mathbb{Q}$, the set of rational numbers.[1]

The nonsingularity condition just means that the cubic can't have any cusps or nodes. As it turns out that the theory of singular cubics is in many ways simpler than that of elliptic curves, we really aren't losing anything with this restriction. Once we insist that the curve be nonsingular, a straight line intersecting the curve in two points must intersect it in a third point, as well. It's simply a question of counting them correctly.

The easiest case to think of looks like this:



However, sometimes it may seem that a line isn't intersecting the curve in three places.



In the first case shown above, the line is actually tangent to the cubic at one of the points of intersection. If we simply count the line as intersecting the cubic twice at that point, the general

[1]Elliptic curves can also be defined as projective curves of genus one. If this definition makes sense to you, you can probably safely skip to the Implementation section.

principle is upheld. Slightly trickier is the second case; the line appears to intersect the curve transversely at two points, and no others.

In this case, we need to move into the projective plane. Actually, we only need one point from it. From one viewpoint, we use the projective plane so that geometric theorems work neatly. For example, we might wish to characterize the intersection of lines by saying "any two distinct lines intersect in exactly one point." Unfortunately, this isn't quite true; parallel lines never meet in the affine plane. So to the affine plane we can simply add a "point at infinity" corresponding to each possible line slope. Then even two parallel lines will intersect, since they meet at the point at infinity corresponding to their common slope.

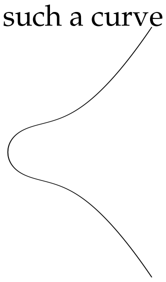
That having been said, we should properly consider the elliptic curve as lying in the projective plane. Even though it may appear that the line in the figure above only intersects the curve twice, it also intersects it at the point at infinity corresponding to all vertical lines. Only one point at infinity must be added to the set of regular points on the elliptic curve in order to have any line intersect the curve in three places. Call that point $\mathcal{O}$.

The important thing to abstract from all of this is that, given a pair of points P and Q lying on an elliptic curve E, we can construct a third point, R, which also lies on the curve. Ultimately, a similar process will yield an actual group law on the points. Let L be the line connecting P and Q. Following the approach of Silverman and Tate, we can define $P * Q$ to be the third point of intersection of L and E. For the reasons outlined above, this does yield a composition law on the set of points on E. This construction will be crucial in defining a group law on the points.

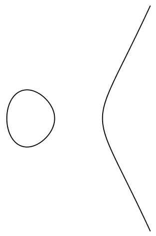
We can write down explicit formulas for computing $P * Q$ in terms of the coefficients of the curve and the coordinates of P and Q. Things are somewhat more manageable if we insist that the curve be in normal form. A nonsingular cubic with at least one rational point is birationally equivalent to a curve of the form $y^2 = x^3 + ax^2 + bx + c$. We can pick a, b and c so that the points on the first curve are in bijection with the points on a curve in normal form, with the possible exception of a few easily identified points. Thus, it suffices to restrict our attention to curves defined by

$$E : y^2 = x^3 + ax^2 + bx + c.$$

Roughly speaking, the real trace of such a curve will either look like this:



or this:



depending on whether $x^3 + ax^2 + bx + c$ has one or three real roots.

In any event, we can now determine explicit formulas for the composition of points. Let $P = (x_0, y_0)$ and $Q = (x_1, y_1)$. Initially let's consider the most straight-forward case, where $x_0 \neq x_1$, and neither P nor Q is $\mathcal{O}$.

We start by constructing the line L between P and Q. It is given by $y = \lambda x + \mu$, where the slope λ is easily calculated to be $(y_1 - y_0)/(x_1 - x_0)$. We can compute μ from λ: $\mu = y_0 - \lambda x_0 = y_1 - \lambda x_1$.

Now, L intersects E in three places. By substituting for y in the equation for E, we find that the intersections are given by the solutions to

$$(\lambda x + \mu)^2 = x^3 + ax^2 + bx + c.$$

This is a cubic in x, and we know that two of the solutions are x_0 and x_1. The third solution will be the desired x_2. So

$$x^3 + (a - \lambda^2)x^2 + (b - 2\lambda\mu)x + (c - \mu^2) = (x - x_0)(x - x_1)(x - x_2).$$

By looking at the coefficients of x^2, we shortly arrive at

$$\begin{aligned} x_2 &= \lambda^2 - a - x_0 - x_1 \\ y_2 &= \lambda x_2 + \mu. \end{aligned}$$

If $P = Q$, then following the discussion at the beginning of this section, we take the line tangent to E at P. Computing the tangent to the curve at a particular point is merely a question of differentiating; the slope is found to be $\lambda = dy/dx = (3x^2 + 2ax + b)/2y$. The calculations for μ, x_2 and y_2 proceed exactly as above with this different value of λ.

What if P and Q are distinct points but have the same x-coordinate? Then the line L between them intersects E in $\mathcal{O}$, the point at infinity. So $P * Q = \mathcal{O}$.

We should describe the situation where one of the points is $\mathcal{O}$. Recall that all vertical lines contain $\mathcal{O}$. So to find $P * \mathcal{O}$ we simply draw the vertical line through P and take the other point of intersection. Thus, $\mathcal{O} * P$ is the other point on E with the same x-coordinate as P.

Finally, since the tangent line to E at $\mathcal{O}$ has a triple intersection at $\mathcal{O}$, we set $\mathcal{O} * \mathcal{O} = \mathcal{O}$.

Thus far we have only presented a rule for combining two points. As mentioned earlier, this doesn't actually define a group law. For example, there is no obvious choice for an identity element. However, we can build on top of this definition to get a proper group law. Let $P + Q = (P * Q) * \mathcal{O}$. In geometric terms, this means take the line connecting P and Q, find its third point of intersection with E, and reflect that point around the x-axis. This does, in fact, define a group law, with $\mathcal{O}$ serving as the identity element. Inverses are given by $-(x, y) = (x, -y)$. The details of this are deferred to [S-T] or the industrious reader. With the explicit formulas in hand the computations are straight forward, if somewhat tedious. Note that this group is automatically abelian, because of the geometric nature of the definition; the line between P and Q is the same as the line between Q and P.

This description has been a little vague about where these points lie. The pictures presented above probably suggested that the points are real, that is, x and y are in $\mathbb{R}$. However, these addition formulas work equally well if x and y are complex solutions to the defining equation of E. If the coefficients a, b and c are rational, then the set of *rational* solutions forms a group, as well. The addition formulas combine the coefficients and coordinates through addition, subtraction, multiplication and division. Fields are closed under these operations; so if the coefficients and coordinates lie in a certain field, then the result is in that field, as well. We write $E(\mathbb{C})$ to denote the group of complex points on E, $E(\mathbb{R})$ for the real points, etc. The focus of this project is the structure of the group $E(\mathbb{Q})$.

2 The Mordell-Weil Theorem

Once the set $E(\mathbb{Q})$ has been endowed with a group law, it makes sense to ask about the structure of that group. It turns out to be easier if we consider the points of finite order and the points of infinite order separately. This section largely provides an overview of results in this area; few theorems are actually proved. Specific proofs are given to the extent that they motivate the algorithm described in section 4.

It makes sense to consider the points of finite and infinite order separately. We say that a point P has m-torsion if $mP = \mathcal{O}$. (Writing mP just means adding P to itself m times.)

As a starting point, we can easily characterize the 2-torsion points, sometimes denoted $E(\mathbb{Q})[2]$. Clearly, $2\mathcal{O} = \mathcal{O}$. Otherwise, let $P = (x_0, y_0)$ be a point such that $2P = \mathcal{O}$, or, equivalently, $P = -P$. Now, the negative of a point is simply that point reflected around the x-axis. So if a point is its own inverse, it must actually lie on the x-axis; $y_0 = 0$. Furthermore, x_0 is a solution to the equation $f(x) = x^3 + ax^2 + bx + c = 0$. Elementary considerations show that a rational solution to this polynomial equation must actually be an integer. Thus, we get a point of order two for every integral root of $f(x)$. The set of 2-torsion points actually forms a subgroup of $E(\mathbb{Q})$. This subgroup, $E(\mathbb{Q})[2]$, is either the trivial group, $\mathbb{Z}/2\mathbb{Z}$, or $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$, depending on whether $f(x)$ has zero, one or three integral roots.

It turns out that in general, for $m \geq 1$, $P \in E(\mathbb{Q})[m]$ will have integer coordinates. Furthermore, let D be the discriminant of $f(x)$, namely, $-4a^3c + a^2b^2 + 18abc - 4b^3 - 27c^2$. Then either $y_0 = 0$ (and $m = 2$), or $y^2|D$. This result is known as the Nagell-Lutz theorem.

Note that this gives an effective method for determining the torsion group of an elliptic curve. First, there is an easy algorithm to check if a point has finite order; simply take P, $2P$, $3P$, etc., until an m is found such that either $mP = \mathcal{O}$, mP doesn't have integer coordinates, or the y-coordinate of mP doesn't divide the discriminant. Since there are only finitely many integers which divide D, the algorithm wil terminate after finitely many steps.

```
IS_TORSION(a,b,c,P)                /* Is P a torsion point on y^2 = x^3 + ax^2 + bx + c? */
Q=(x,y) ← O
repeat
  Q ← Q + P
until (Q = O) or (Q isn't an integer point) or (y^2 ∤ D)
if (Q = O)
  return( true )
else
  return( false )
```

Additionally, we only need test a finite set of points for finite order.

```
TORSION_GROUP(a,b,c)              /* Find torsion subgroup of y^2 = x^3 + ax^2 + bx + c. */
torsion ← {O}
D ← −4a^3c + a^2b^2 + 18abc − 4b^3 − 27c^2
for each divisor y of D
  for each integer solution x of x^3 + ax^2 + bx + c − y^2
    P ← (x,y)
```

```
    if is_torsion(a,b,c,P)
      torsion ← torsion ∪ {P}
  return( torsion )
```

This is a pretty good characterization of the finite part of the group. However, the group $E(\mathbb{Q})$ could contain elements of infinite order, as well. That this can happen is evidenced by

$$y^2 = x^3 + 17.$$

Clearly, the point $P = (1/4, 33/8)$ lies on the curve, since $(1/4)^3 + 17 = 1089/64 = (33/8)^2$. But its coordinates aren't integers, while all points of finite order have integer coordinates. So P must be of infinite order.

A priori, it's not obvious that we will be able to discern any structure, at all, on the torsion-free portion of the $E(\mathbb{Q})$. It turns out that, even though the group $E(\mathbb{Q})$ may not be finite, it is always *finitely generated*. There exists a finite set of points so that any other point is equal to some linear combination of that set of points; $E(\mathbb{Q}) \cong E(\mathbb{Q})_{tor} \oplus \mathbb{Z}^r$. In other words, there is a set of points P_1, P_2, ..., P_r so that for a $Q \in E(\mathbb{Q})$, we can find integers $c_1, \ldots, c_r$ and a $T \in E(\mathbb{Q})_{tor}$ so that $Q = T + c_1 P_1 + c_2 P_2 + \cdots + c_r P_r$. The rank of an elliptic curve is r, the size of a smallest torsion-free generating set.

This theorem was proved by Mordell in 1922, and subsequently generalized to arbitrary abelian varieties over number fields by Weil. An elementary proof of this theorem proceeds in two sections. The first part is the "weak" Mordell-Weil theorem. Let $2E(\mathbb{Q})$ denote the subgroup obtained by doubling all the points in $E(\mathbb{Q})$: $2E(\mathbb{Q}) = \{2P | P \in E(\mathbb{Q})\}$. The weak Mordell-Weil theorem states that the index $[E(\mathbb{Q}) : 2E(\mathbb{Q})]$ is finite.

For the second part, it is necessary to introduce the notion of a height function. The height function may be thought of as a measure of how big a point is. In order to prove the Mordell-Weil theorem, it's necessary to show that the height function behaves nicely with respect to the group law on the points. Once these two pieces are established, the theorem follows easily.

The proof given here assumes that there is a rational point of order two on the curve. While the theorem is true without this assumption, the proof is somewhat simpler if we can stay in the rational numbers. Now, if E is a curve given in Weierstrass form, then a rational point of order two looks like $(x_0, 0)$ with x_0 an integer. We can make a change of coordinates, $(x, y) \mapsto (x - x_0, y)$ which sends $(x_0, 0)$ to $(0, 0)$. Obviously, such a change won't affect the structure of the group $E(\mathbb{Q})$. Thus, given the restriction that we're considering curves with a rational two-torsion point, we can assume that E is given by $y^2 = x^3 + ax^2 + bx$.

2.1 Weak Mordell-Weil Theorem

The goal is to show that the image of the multiplication-by-two map $[2] : E(\mathbb{Q}) \to E(\mathbb{Q})$ has finite index. The approach taken here is to break the map [2] into two pieces, and show that the image at each step has finite index. In general, one can show that a subgroup H of some group G has finite index by exhibiting a one-to-one map from G/H into some finite group. This, in turn, can be done by showing that H is the kernel of a map from G to a finite group. Such a method will be used here.

Initially, we need to find a group G and a pair of maps f, g so that the following diagram commutes:

$$\begin{tikzcd} E(\mathbb{Q}) \arrow[r, "f"] \arrow[dr, "{[2]}"'] & G \arrow[d, "g"] \\ & E(\mathbb{Q}) \end{tikzcd}$$

In other words, $g(f(P)) = [2](P) = 2P$. For any particular curve E, we'll pick a related elliptic curve $\overline{E}$ and maps $\phi : E(\mathbb{Q}) \to \overline{E}(\mathbb{Q})$ and $\psi : \overline{E}(\mathbb{Q}) \to E(\mathbb{Q})$. These will be given by explicit formulas. Some of the motivation for the specific choices of $\overline{E}$, ϕ and ψ is given in [S-T].

Given a curve $E : y^2 = x^3 + ax^2 + bx$, define the curve $\overline{E}$ by

$$\begin{aligned} \overline{E}: \quad y^2 &= x^3 + \overline{a}x^2 + \overline{b}x \\ \overline{a} &= -2a \\ \overline{b} &= a^2 - 4b. \end{aligned}$$

In a sense, E and $\overline{E}$ are duals of each other. Applying the construction twice yields $\overline{\overline{E}} : y^2 = x^3 + (-2 \cdot -2a)x^2 + (4a^2 - 4(a^2 - 4b))x$, or $y^2 = x^3 + 4ax^2 + 16b$. This is almost the same as the original curve, E; for if (x_0, y_0) is a point on E, then $(4x_0, 8y_0)$ must be a point on $\overline{\overline{E}}$. It's similarly easy to identify a point on $\overline{\overline{E}}$ with a point on E. In fact, the two curves are isomorphic via this identification. This will be useful in constructing the maps ϕ and ψ.

Define the map $\phi : E \to \overline{E}$ by $(x, y) \mapsto (\frac{y^2}{x^2}, y\frac{x^2-b}{x^2})$. This is well-defined everywhere on E except $\mathcal{O}$, the point at infinity, and the two-torsion point $T = (0, 0)$, which is the only point on E with a zero x-coordinate. We can extend the definition to everywhere on E.

$$\begin{aligned} \phi : \quad E &\to \overline{E} \\ P &\mapsto (\frac{y^2}{x^2}, y\frac{x^2 - b}{x^2}) \quad P = (x, y) \neq T, \mathcal{O} \\ & \qquad \overline{\mathcal{O}} \qquad\qquad P = \mathcal{O} \text{ or } P = T. \end{aligned}$$

Here, $\overline{\mathcal{O}}$ is the point at infinity on $\overline{E}$.

One can use an identical construction to get a map from $\overline{E}$ to $\overline{\overline{E}}$. What we actually want, however, is a map from $\overline{E}$ back to the original curve, E. Thus, we combine the above map with $(x, y) \mapsto (x/4, y/8)$ to define ψ.

$$\begin{aligned} \psi : \quad \overline{E} &\to E \\ P &\mapsto (\frac{\overline{y}^2}{4\overline{x}^2}, \overline{y}\frac{\overline{x}^2 - \overline{b}}{8\overline{x}^2}) \quad P = (\overline{x}, \overline{y}) \neq \overline{T}, \overline{\mathcal{O}} \\ & \qquad \mathcal{O} \qquad\qquad P = \overline{\mathcal{O}} \text{ or } P = \overline{T}. \end{aligned}$$

The maps ϕ and ψ are partially characterized by the following proposition.

Proposition 2.1 *Let E, $\overline{E}$, ϕ and ψ be defined as above.*

1. *The maps ϕ and ψ are homomorphisms.*

2. *The kernel of ϕ is $\{\mathcal{O}, T\}$, and the kernel of ψ is $\{\overline{\mathcal{O}}, \overline{T}\}$.*

3. *The composition of the maps is multiplication by two, i.e., $\psi \circ \phi = [2]$, and $\phi \circ \psi = \overline{[2]}$.*

The proof of this proposition is simple, but tedious; the approach is merely indicated here. Proving that ϕ is a homomorphism is a matter of verifying that the group law is preserved: $\phi(P + Q) = \phi(P) + \phi(Q)$. This, in turn, is a matter of turning the crank, and confirming that the rational functions obtained through addition and application of ϕ match up as they should. The price of having such concrete definitions of ϕ and the group law is that many cases must be separately checked. Of course, if ϕ is a homomorphism, then ψ is too, because of the way it was constructed.

Showing that $\ker \phi = \{\mathcal{O}, T\}$ is trivial, given the definition of ϕ; all other points are mapped to "affine points" on $\overline{E}$. The same argument works for ψ.

Finally, $\psi \circ \phi(P) = 2P$. This, too, can be verified through purely algebraic computations.

Thus far, we have broken multiplication by two into two maps, as promised. It remains to show that the index of the image of each map is finite, that is, $[\overline{E}(\mathbb{Q}) : \phi(E(\mathbb{Q}))]$ and $[E(\mathbb{Q}) : \psi(\overline{E}(\mathbb{Q}))]$ are finite. As before, the approach is to (seemingly arbitarily) define a map, and then show that it behaves nicely. The following definitions and lemmas will be stated for the $\phi : E \to \overline{E}$ half of the problem; the analogous statements for ψ can be made and proven in the same way.

We can define a map $\overline{\alpha} : \overline{E}(\mathbb{Q}) \to \mathbb{Q}^*/\mathbb{Q}^{*2}$. Recall that $\mathbb{Q}^*$ is the multiplicative group of rational units, and $\mathbb{Q}^{*2}$ is the subgroup consisting of perfect squares. So $\mathbb{Q}^*/\mathbb{Q}^{*2}$ is like the nonzero rational numbers, with two elements identified if their quotient is the square of a rational number.

$$\begin{aligned} \overline{\alpha}: \quad \overline{E}(\mathbb{Q}) &\to \mathbb{Q}^*/\mathbb{Q}^{*2} \\ \overline{\mathcal{O}} &\mapsto 1 \pmod{\mathbb{Q}^{*2}} \\ \overline{T} &\mapsto \overline{b} \pmod{\mathbb{Q}^{*2}} \\ (\overline{x}, \overline{y}) &\mapsto \overline{x} \pmod{\mathbb{Q}^{*2}}. \end{aligned}$$

The following proposition characterizes the behavior of $\overline{\alpha}$.

Proposition 2.2 *Let E, $\overline{E}$, ϕ and $\overline{\alpha}$ be as above.*

1. *The map $\overline{\alpha} : \overline{E}(\mathbb{Q}) \to \mathbb{Q}^*/\mathbb{Q}^{*2}$ is a homomorphism of groups.*
2. *The kernel of $\overline{\alpha}$ is $\phi(E(\mathbb{Q}))$, and $\overline{\alpha}$ induces a natural injection $\dfrac{\overline{E}(\mathbb{Q})}{\phi(E(\mathbb{Q})} \hookrightarrow \dfrac{\mathbb{Q}^*}{\mathbb{Q}^{*2}}$.*
3. *Let $p_1, \ldots, p_r$ be the distinct primes dividing $\overline{b}$. Then the image of $\overline{\alpha}$ is contained in the subgroup of $\mathbb{Q}^*/\mathbb{Q}^{*2}$ with representatives*

$$\{(-1)^{\epsilon_0} p_1^{\epsilon_1} p_2^{\epsilon_2} \cdots p_r^{\epsilon_r} | \epsilon_i \in \{0, 1\}\} \subset \mathbb{Q}^*/\mathbb{Q}^{*2}.$$

4. *The index $[\overline{E}(\mathbb{Q}) : \phi(E(\mathbb{Q}))]$ is finite.*

Proving the first part of this proposition is a routine calculation. Things start getting interesting with the proof of the second part. What it means is that any point in $\overline{E}(\mathbb{Q})$ whose x-coordinate is a square must come from $E(\mathbb{Q})$, via ϕ.

The easy part of the inclusion is that $\phi(E(\mathbb{Q})) \subseteq \ker \overline{\alpha}$. To do this, we must show that if $P \in E(\mathbb{Q})$, then $\phi(P)$ is the square of a rational number. But this is obvious from the definition. Certainly, $\mathcal{O}$ and T get sent to $\overline{\mathcal{O}}$, and $\overline{\alpha}$ takes $\overline{\mathcal{O}}$ to 1 in $\mathbb{Q}^*/\mathbb{Q}^{*2}$. Otherwise, if $P = (x, y)$, then the x-coordinate of $\phi(P)$ is y^2/x^2, which is obviously the square of a rational number.

It is slightly trickier to show that $\ker \overline{\alpha} \subseteq \phi(E(\mathbb{Q}))$; any element of the kernel comes from a point in $E(\mathbb{Q})$. So suppose that $(\overline{x}, \overline{y}) \in \overline{E}(\mathbb{Q})$ gets sent to 1 by $\overline{\alpha}$. Then $\overline{x} = w^2$ for some rational w. We'll try to find a point in $E(\mathbb{Q})$ which would be mapped to $(\overline{x}, \overline{y})$.

Actually, since $\ker \phi$ contains two elements, we should really be looking for two elements $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$, where the P_i are rational points on E which are sent to $(\overline{x}, \overline{y})$. Define these points by

$$\begin{array}{ll} x_1 = \frac{1}{2}(w^2 - a + \frac{\overline{y}}{w}) & y_1 = x_1 w \\ x_2 = \frac{1}{2}(w^2 - a - \frac{\overline{y}}{w}) & y_2 = -x_2 w. \end{array}$$

There are a couple of ways to show that $P_i \in E(\mathbb{Q})$. The first method is purely mechanical. Simply substitute x_i into $x^3 + ax^2 + bx$, and verify that the result is the square of a rational number. This will work, even though the actual computation is something of a nuisance.

A second way uses a little more information about the situation. Initially, we see that $x_1 x_2 = b$.

$$\begin{aligned} x_1 x_2 &= \frac{1}{2}(w^2 - a + \frac{\overline{y}}{w}) \cdot \frac{1}{2}(w^2 - a - \frac{\overline{y}}{w}) \\ &= \frac{1}{4}[(w^2 - a)^2 - \frac{\overline{y}^2}{w^2}] \\ &= \frac{1}{4}[(\overline{x} - a)^2 - \frac{\overline{y}^2}{\overline{x}}] \\ &= \frac{1}{4}\frac{\overline{x}^3 - 2a\overline{x}^2 + a^2\overline{x} - \overline{y}^2}{\overline{x}}. \end{aligned}$$

Now we know that $(\overline{x}, \overline{y})$ is on $\overline{E}(\mathbb{Q})$, so $\overline{y}^2 = \overline{x}^3 - 2a\overline{x}^2 + (a^2 - 4b)\overline{x}$. Applying this knowledge to the last line, we find that $x_1 x_2 = b$.

How does this help us prove that $P_i \in E(\mathbb{Q})$? We have to show that $y_i^2 = x_i^3 + ax^2 + bx$; or, equivalently, that $\frac{y_i^2}{x_i^2} = x_i + a + \frac{b}{x_i}$. Now, $\frac{y_i^2}{x_i^2} = w^2$ by the definition of y_i, and we know that $b = x_1 x_2$. So we've reduced the problem to showing that

$$w^2 = x_1 + a + x_2.$$

But this comes from the definition of x_1 and x_2;

$$x_1 + x_2 = \frac{1}{2}(w^2 - a + \frac{\overline{y}}{w}) + \frac{1}{2}(w^2 - a - \frac{\overline{y}}{w}) = w^2 - a.$$

Thus, $\phi(P_1) = \phi(P_2) = (\overline{x}, \overline{y}) = (w^2, \overline{y})$, and $\ker \overline{\alpha} \subseteq \phi(E(\mathbb{Q}))$. The kernel of $\overline{\alpha}$ is $\phi(E(\mathbb{Q})$, and so $\overline{\alpha}$ induces a one-to-one map $\overline{E}(\mathbb{Q})/\phi(E(\mathbb{Q})) \hookrightarrow \mathbb{Q}^*/\mathbb{Q}^{*2}$.

We now turn to the third part of the proposition. Let $S = \{p_1, p_2, \ldots, p_r\}$ be the set of primes which divide $\overline{b}$. Let R^* be the subgroup of $\mathbb{Q}^*$ given by

$$R^* = \{\pm p_1^{e_1} p_2^{e_2} \cdots p_r^{e_r} | e_i \in \mathbb{Z}\}.$$

The claim is that im $\overline{\alpha} \subseteq R^*/R^{*2} \subset \mathbb{Q}^*/\mathbb{Q}^{*2}$. In other words, let $\overline{P} = (\overline{x}, \overline{y})$ be a rational point on $\overline{E}$. It turns out that we can write $\overline{x} = \frac{m}{e^2}$ and $\overline{y} = \frac{n}{e^3}$, where m, e and n are integers, and $\gcd(m, e) = \gcd(n, e) = 1$. What we have to show is that if $p|m$ and $p \notin S$, then $p^2|m$. (Since e is squared, we don't have to worry about its factors.)

Initially, we substitute $\overline{x} = m/e^2$ and $\overline{y} = n/e^3$ into the equation for $\overline{E}$. We get

$$\begin{aligned}
\frac{n^2}{e^6} &= \frac{m^3}{e^6} + \overline{a}\frac{m^2}{e^4} + \overline{b}\frac{m}{e^2} \\
n^2 &= m^3 + \overline{a}m^2e^2 + \overline{b}e^4 m \\
n^2 &= m \cdot (m^2 + \overline{a}me^2 + \overline{b}e^4).
\end{aligned}$$

Supposes q divides m but not $(m^2 + \overline{a}me^2 + \overline{b}e^4)$. Then $q^2|m$, as well, for if $q|n^2$ then $q^2|n^2$.

Suppose q divides $(m^2 + \overline{a}me^2 + \overline{b}e^4)$ but not m. Then q can't divide $\overline{b}e^4$, either. In particular, q divides neither e nor m, and so we don't have to worry about it.

Finally, suppose q divides both m and $(m^2 + \overline{a}me^2 + \overline{b}e^4)$. Then $q|(m^2 + \overline{a}me^2)$, and the last term $\overline{b}e^4$ must also be a multiple of q. Since $q|m$ we can't have $q|e$. Therefore, q must divide $\overline{b}$.

This gives a fairly complete picture of what kinds of primes can divide m. (We don't have to worry about e, since primes which divide e automatically occur with even exponent in $x = m/e^2$.) For most primes p, an even power of p must divide m. The only way that a prime can *not* have an even power of p dividing m/e^2 is if $p|\overline{b}$. This, of course, is precisely the claim.

The final part of the proposition follows easily. The size of R^*/R^{*2} is easily seen to be 2^{r+1}, and $\overline{E}(\mathbb{Q})/\phi(E(\mathbb{Q}))$ maps injectively into it. So the index $[\overline{E}(\mathbb{Q}) : \phi(E(\mathbb{Q}))] \leq 2^{r+1}$, and is certainly finite.

So now we know that the multiplication map $[2] : E(\mathbb{Q}) \to E(\mathbb{Q})$ factors through the group on a related curve $\overline{E}$ via $E(\mathbb{Q}) \stackrel{\phi}{\to} \overline{E}(\mathbb{Q}) \stackrel{\psi}{\to} E(\mathbb{Q})$. The most recent proposition shows that the images at each step, $\phi(E(\mathbb{Q}))$ and $\psi(\overline{E}(\mathbb{Q}))$ are finite. The ultimate goal is the result that the index $[E(\mathbb{Q}) : 2E(\mathbb{Q})] = [E(\mathbb{Q}) : \psi\phi(E(\mathbb{Q}))]$ is finite. The final step doesn't require any knowledge of elliptic curves; it's a simple question of group theory.

Proposition 2.3 *Let A and B be abelian groups with homomorphisms $\phi : A \to B$ and $\psi : B \to A$, such that $\psi \circ \phi = [2]_A$ and $\phi \circ \psi = [2]_B$. Also, suppose that $[B : \phi(A)]$ and $[A : \psi(B)]$ are finite. Then the index $[A : 2A]$ is finite.*

The proof is straight-forward. Take $a_1, \ldots, a_m$ and $b_1, \ldots, b_n$ to be coset representatives for $\psi(B)$ in A and $\phi(A)$ in B, respectively. One shows that the set $\{a_i + \psi(b_j)\}$ is a complete set of coset representatives for $2A$ in A.

Applying this proposition we find that $[E(\mathbb{Q}) : 2E(\mathbb{Q})]$ is finite, as desired.

2.2 Height functions

The other half of the Mordell-Weil theorem requires a notion of the height of a point. This yields a tool which is useful in establishing a set's finitude.

The classic method of descent uses the absolute value function in order to show that, for example, a given polynomial has no integer solutions. One shows that, given a particular integer solution, one can produce another solution whose absolute value is strictly smaller. This yields a contradiction. On one hand, one claims that there is an infinite sequence of integers with strictly decreasing absolute value. On the other hand, the set of integers with absolute value less than some particular value is finite.

This notion of the size of a solution works quite well for studying integers. However, it is not as well suited to studies of *rational* solutions to equations. The key fact used in the strategy outlined above is that, for any bound B, the set $\{x \in \mathbb{Z} | |x| \leq B\}$ is finite. If we let x take on rational values, however, this assertion is no longer true. For example, one could take the set $\{1, \frac{1}{2}, \frac{1}{3}, \ldots\}$. We need a slightly different measure of the size of a rational number.

Let's define the height of a rational number x to be the maximum of the absolute value of its numerator and denominator. More formally, let $x = \frac{m}{n}$, where m and n are relatively prime. Define the height function by $H(x) = \max\{|m|, |n|\}$. This enjoys the same sort of finiteness property described above. Specifically, for any B, the set $\{x \in \mathbb{Q} | H(x) \leq B\}$ is finite.

This definition can be extended to the rational points on an elliptic curve. Let $H(P) = H((x, y)) = H(x)$. A similar finiteness property still holds, namely, $\{P \in E(\mathbb{Q}) | H(P) \leq B\}$ is finite for any fixed bound B. This is because there are only finitely many choices for the x-coordinate, and any value of x yields at most two points on the curve E.

The height function behaves somewhat multiplicatively on the rationals; it makes sense to compare $H(x)H(y)$ to $H(xy)$. Notationally, however, it's desirable to have a function which acts additively. There is an addition law on points on an elliptic curve, while there's no multiplication function. So define the logarithmic height by $h(x) = \log H(x)$. For any rational number x, $H(x)$ is at least 1, and so $h(x)$ is a nonnegative real number. This definition is extended in the obvious way to rational points on elliptic curves.

So much for the motivation of h. In order to prove the Mordell-Weil theorem, one must establish certain properties of this height function.

Proposition 2.4 *Define the height function $h : E(\mathbb{Q}) \to \mathbb{R}$ as above.*

1. *For every $B \in \mathbb{R}$, the set $\{P \in E(\mathbb{Q}) | h(P) \leq M\}$ is finite.*

2. *For every $P_0 \in E(\mathbb{Q})$ there is a constant κ_0, depending only on P_0 and E, such that $h(P + P_0) \leq 2h(P) + \kappa_0$ for all $P \in E(\mathbb{Q})$.*

3. *There is a constant κ, depending only on E, such that $h(2P) \geq 4h(P) - \kappa$.*

This proposition relates the addition law on the curve to a number-theoretic value. Ultimately, it will allow us to reduce the geometric question to a number-theoretic one, and then use number theory in order to obtain an answer.

The first part of this proposition follows from the disccusion above; the proof of the latter two parts is left to [S-T]. Essentially, one works with the concrete descriptions of the addition laws, and attempts to compute the height of the resulting value. This can be a little intricate. The quotients given by the addition formulae aren't necessarily in "lowest form." One must show that there isn't too much cancellation between the numerator and the denominator, in order to put a lower bound on the resulting height.

2.3 Mordell-Weil Theorem

Given the theorem of the height function and the weak Mordell-Weil theorem, we can prove the desired theorem.

Proposition 2.5 *Let E be an elliptic curve with a rational point of order two. Then $E(\mathbb{Q})$ is finitely generated.*

The weak Mordell-Weil theorem says that the group $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite, so we can take a finite set of representatives $\mathcal{Q} = \{Q_1, Q_2, \ldots, Q_n\}$ for the cosets of $2E(\mathbb{Q})$ in $E(\mathbb{Q})$.

Recall that the second part of proposition 2.4 says for each Q_i there is a κ_i, so that $h(Q_i + P) \leq 2h(P) + \kappa_i$ for all rational points P. Since there are only finitely many Q_i, we can take κ' to be the maximum of the κ_i. Thus, $h(Q_i + P) \leq 2h(P) + \kappa'$ for all $Q_i \in \mathcal{Q}$ and $P \in E(\mathbb{Q})$. Let κ be the constant from the same proposition such that $h(2P) \geq 4h(P) - \kappa$. Then the set $\mathcal{R} = \{P \in E(\mathbb{Q}) | h(P) \leq \kappa' + \kappa\}$ is finite, as well. As we will see shortly, $\mathcal{Q} \cup \mathcal{R}$ generates all of $E(\mathbb{Q})$.

To prove this claim, we must show that any $P \in E(\mathbb{Q})$ can be written as a combination of elements of $\mathcal{Q} \cup \mathcal{R}$.

In a way, this proof is inductive. Indeed, if the error-terms κ and κ_0 were zero, one could prove the theorem by induction on the height. Even with the error term, however, the proof is elementary.

The point P must have a coset representative Q_{i_1} in $E(\mathbb{Q})/2E(\mathbb{Q})$; $P - Q_{i_1} \in 2E(\mathbb{Q})$. So there's a P_1 such that $P - Q_{i_1} = 2P_1$. Similarly, we can find an index i_2 and a point P_2 so that $P_1 - Q_{i_2} = 2P_2$. In fact, we can continue this chain as far as we like. Substituting after the m^{th} iteration, we find

$$\begin{aligned} P &= Q_{i_1} + 2(Q_{i_2} + 2(Q_{i_3} + \cdot + P_m)) \\ &= Q_{i_1} + 2Q_{i_2} + 4Q_{i_3} + \cdots + 2^{m-1}Q_{i_m} + 2^m P_m. \end{aligned}$$

Each of the Q_{i_j} is in $\mathcal{Q}$, of course. If we can show that we can pick a (finite) m so that $P_m \in \mathcal{Q} \cup \mathcal{R}$, we're finished. The strategy is to show that the height decreases enough at each step to force P_m into $\mathcal{R}$. We want to show that at, say, the j^{th} step of the chain, the size is decreasing. Proposition 2.4 gives us

$$4h(P_j) \leq h(2P_j) + \kappa = h(P_{j-1} - Q_{i_j}) + \kappa \leq 2h(P_{j-1}) + \kappa' + \kappa.$$

We can isolate $h(P_j)$.

$$\begin{aligned} h(P_j) &\leq \frac{1}{2}h(P_{j-1}) + \frac{\kappa' + \kappa}{4} \\ &= \frac{3}{4}h(P_{j-1}) - \frac{1}{4}(h(P_{j-1}) - (\kappa' + \kappa)). \end{aligned}$$

If $h(P_{j-1}) \leq \kappa' + \kappa$, then we're done, since $P_{j-1} \in \mathcal{R}$. Otherwise, $h(P_j) \leq 3/4h(P_{j-1})$. The height is decreasing by a factor of $3/4$ at every step. After some finite number of steps, the height of $h(P_m)$ will be less than the bound $\kappa' + \kappa$, and so P_m will be in $\mathcal{R}$. This concludes the proof of the Mordell-Weil theorem.

2.4 Effectivity

The proof of the Mordell-Weil theorem, while correct, in a sense is not the best possible. The proof is nonconstructive; it tells us that the rank is finite, without specifying how to find it. Almost every step seems computable. However, at the last (and crucial) step, there is no known algorithm guaranteed to solve the problem which arises. In any given case, however, any one of several informal heuristics may prove sufficient to compute the rank. What this requires is a slightly closer analysis of what is happening in the proof of the weak Mordell-Weil theorem.

Before that, we should review some facts about the structure of $E(\mathbb{Q})$. There is a torsion part of the group. This has a 2-torsion component, $E(\mathbb{Q})[2]$, and another (finite) component, $E(\mathbb{Q})_{tor-2}$. As for the torsion-free part, we know that it is finitely generated as a $\mathbb{Z}$-module. If r is the rank of the group, then the torsion-free portion $E(\mathbb{Q})_\infty$ is isomorphic to $\mathbb{Z}^r$. In summary, we have

$$E(\mathbb{Q}) \cong E(\mathbb{Q})[2] \oplus E(\mathbb{Q})_{tor-2} \oplus E(\mathbb{Q})_\infty \cong (\mathbb{Z}/2\mathbb{Z})^s \oplus E(\mathbb{Q})_{tor-2} \oplus \mathbb{Z}^r,$$

where s is the number of rational points of order two on E, and $E(\mathbb{Q})_{tor-2}$ is a finite group with order relatively prime to two. The weak Mordell-Weil theorem says something about $E(\mathbb{Q})/2E(\mathbb{Q})$. This group looks like

$$\begin{aligned}
E(\mathbb{Q})/2E(\mathbb{Q}) &\cong (\frac{\mathbb{Z}/2\mathbb{Z}}{2\mathbb{Z}})^s \oplus E(\mathbb{Q})_{tor-2}/2E(\mathbb{Q})_{tor-2} \oplus (\mathbb{Z}/2\mathbb{Z})^r \\
&\cong (\mathbb{Z}/2\mathbb{Z})^s \oplus (\mathbb{Z}/2\mathbb{Z})^r \\
&\cong (\mathbb{Z}/2\mathbb{Z})^{s+r}.
\end{aligned}$$

So $[E(\mathbb{Q}) : 2E(\mathbb{Q})] = s + r$. It's easy to compute s. Recall that a point has order two only if its y-coordinate is zero. We're working with the representation $y^2 = x^3 + ax^2 + bx$ for E. So points of order two are given by solutions to $x^3 + ax^2 + bx = x(x^2 + ax + b) = 0$. It's not surprising that we find $(0, 0) \in E(\mathbb{Q})[2]$. The other two solutions will be rational if and only if

$$x = \frac{-a \pm \sqrt{a^2 - 4b}}{2}$$

is rational, that is, if and only if $a^2 - 4b$ is a perfect square. This gives us an easy way to determine s.

Now, if one is careful in the proof of proposition 2.3, and then applies the result to the specific case here, one can show that

$$[E(\mathbb{Q}) : 2E(\mathbb{Q})] = \frac{[E(\mathbb{Q}) : \psi(\overline{E}(\mathbb{Q}))][\overline{E}(\mathbb{Q}) : \phi(E(\mathbb{Q}))]}{[\ker \psi : \ker \psi \cap \phi(E(\mathbb{Q}))]}.$$

Essentially, this consists of guarding against overcounting in the enumeration of coset representatives. We can see that $\overline{T} \in \phi(E(\mathbb{Q}))$ if and only if $\overline{b} = a^2 - 4b$ is a perfect square. This means that $\ker \psi \cap \phi(E(\mathbb{Q}))$ is either $\{\overline{\mathcal{O}}\}$ or $\{\overline{\mathcal{O}}, \overline{T}\}$ depending on whether $a^2 - 4b$ is not or is a perfect square, respectively. The index $[\ker \psi : \ker \psi \cap \phi(E(\mathbb{Q}))]$ is therefore $2 - s$. In summary,

$$[E(\mathbb{Q}) : 2E(\mathbb{Q})] = 2^{r+s} = \frac{[E(\mathbb{Q}) : \psi(\overline{E}(\mathbb{Q}))][\overline{E}(\mathbb{Q}) : \phi(E(\mathbb{Q}))]}{2^{2-s}}.$$

This gives us an explicit formula for the rank.

$$2^r = \frac{[E(\mathbb{Q}) : \psi(\overline{E}(\mathbb{Q}))][\overline{E}(\mathbb{Q}) : \phi(E(\mathbb{Q}))]}{4}.$$

By the isomorphism in proposition 2.2, $\overline{\alpha}(\overline{E}(\mathbb{Q})) \cong \overline{E}(\mathbb{Q})/\phi(E(\mathbb{Q}))$. Similarly, $\alpha(E(\mathbb{Q})) \cong E(\mathbb{Q})/\psi(\overline{E}(\mathbb{Q}))$. So we get the extremely useful formula

$$2^r = \frac{\#\alpha(E(\mathbb{Q})) \cdot \#\overline{\alpha}(\overline{E}(\mathbb{Q}))}{4}.$$

The upshot of all of this is that we can compute the rank of an elliptic curve if we can compute the size of im α and im $\overline{\alpha}$.

As before, this discussion will focus on $\overline{\alpha}$; the argument for α proceeds in the same way. We know that $\overline{\alpha}$ maps $\overline{E}(\mathbb{Q})$ into R^*/R^{*2}. To determine that image, we have to figure out which points in $\overline{E}(\mathbb{Q})$ give rise to nontrivial elements of R^*/R^{*2}. Evidently, $\overline{\mathcal{O}}$ gets sent to the identity element. Also, $\overline{T}$ will be mapped into R^*/R^{*2}. It remains to find out which rational numbers, modulo squares, can occur as the $\overline{x}$-coordinates of points on $\overline{E}$. As before, write $\overline{x} = m/e^2, \overline{y} = n/e^3$. Then $n^2 = m \cdot (m^2 + \overline{a}me^2 + \overline{b}e^4)$.

Let q be the squarefree part of m. We saw above that q must divide $(m^2 + \overline{a}me^2 + \overline{b}e^4)$ and $\overline{b}$, as well. Let $qq' = \overline{b}$, $qm' = q$. Then

$$n^2 = m'q \cdot (q^2m'^2 + \overline{a}qm'e^2 + qq'e^4).$$

Since q divides everything on the left side, it divides n^2, too. Because q is squarefree and $q|n^2$, we have $q|n$. Write $qn' = n$.

$$\begin{aligned}(qn')^2 &= m'q \cdot (q^2m'^2 + \overline{a}qm'e^2 + qq'e^4) \\ n'^2 &= m' \cdot (qm'^2 + \overline{a}m'e^2 + q'e^4).\end{aligned}$$

At this point, m' and $(qm'^2 + \overline{a}m'e^2 + q'e^4)$ are relatively prime, and their product is a square. This means each of them is a square, and thus n' has a factorization $n' = MN$, where

$$\begin{aligned} M^2 &= m' \\ N^2 &= qm'^2 + \overline{a}m'e^2 + q'e^4. \end{aligned}$$

Thus, we can eliminate m'.

$$\begin{aligned} M^2N^2 &= M^2(qM^4 + \overline{a}M^2e^2 + q'e^4) \\ N^2 &= qM^4 + \overline{a}M^2e^2 + q'e^4. \end{aligned}$$

So if there is a point $\overline{E}(\mathbb{Q})$ which gives a nontrivial element of R^*/R^{*2}, it will correspond to a solution to the last equation given above, where $qq' = \overline{b}$. This is how we can attempt to craft an algorithm to determine the rank of an elliptic curve. Simply take all integer factorizations $\overline{b} = qq'$, and see whether the resulting equation has a nontrivial solution. The curve corresponding to such an equation is often called a covering space. Every equation with an integral solution, or covering space with an integral point, corresponds to an element of $\alpha(\overline{E}(\mathbb{Q}))$. So we can compute $\#\alpha$ and $\#\overline{\alpha}$, and thus compute r.

This would work fine, were we able to determine whether the diophantine equations which arise admit solutions. Unfortunately, there is no known finite decision procedure which will determine if such a solution exists. This is not to say that no such decision procedure exists. While it is known that no algorithm can solve arbitrary diophantine equations, this doesn't rule out a method for solving this specific class of diophantine problems.

At any rate, this is the obstacle to effectivity in the Mordell-Weil theorem.

3 Weak Mordell-Weil: Reprise

What we have just seen for elliptic curves with a rational 2-torsion point actually holds true in a much broader context. If E is *any* elliptic curve, then the group $E(\mathbb{Q})$ is finitely generated. Also, if

K is a number field, that is, a finite extension of $\mathbb{Q}$ contained in $\mathbb{C}$, then $E(K)$ is finitely generated. And we needn't restrict our attention to elliptic curves; if A is any abelian variety defined over a number field K, then $A(K)$ is finitely generated. In this section we will outline a proof of a more general weak Mordell-Weil theorem, namely, $A(K)/mA(K)$ is finite for any integer $m \geq 1$.

Not surprisingly, this theorem will require somewhat more advanced tools. Before, we could describe everything in terms of explicit computations with rational functions. The price of the increased generality of the theorem is an increased abstraction in the method of proof. The method given here will require elements of Galois theory, algebraic number theory, and a little bit of Kummer theory. Most of it follows the approach of Chapter VIII in [Sil]. A more explicit development of the necessary Kummer theory may be found in [Lan].

We should fix some notation before we start. Let A be an abelian variety. By analogy with elliptic curves, let $A(K)$ denote the group of points defined over some field K. We will always take K to be a number field, and $\overline{K}$ its algebraic closure. For any integer m, $[m] : A \to A$ is the multiplication-by-m map; $A(K)[m]$ is the set of points of order m; and $mA(K) = \{mP | P \in A(K)\}$. Finally, let $A[m] = A(\overline{K})[m]$.

The general strategy is to introduce a pairing between $A(K)/mA(K)$ and another group, in this case a Galois group of a certain extension L over K. Then $A(K)/mA(K)$ is finite if and only if Gal(L, K) is finite, that is, if and only if L is a finite extension. To deduce the finitude of $[L : K]$, we will combine the fact that A has good reduction almost everywhere with considerations from algebraic number theory.

As an initial reduction step, we can show that it suffices to assume that all the m-torsion points are defined over K. Let d be the dimension of A. We shall see below that $A(\overline{K})[m] \cong (\mathbb{Z}/m\mathbb{Z})^{2d}$, and thus is finite. Since the coordinates of each point satisfy some algebraic relation (depending on the definition of A), each point $P \in A(\overline{K})[m]$ is defined over a finite extension $K(P)$ over K. The m-torsion points actually lie in a finite extension K' of K, taken to be the compositum of the (finitely many) $K(P)$. The following proposition shows that it suffices to consider $K = K'$ in order to prove the weak Mordell-Weil theorem.

Proposition 3.1 *Let K' be a Galois extension of K such that $A(K')/mA(K')$ is finite. Then $A(K)/mA(K)$ is finite, too.*

One can prove this by showing that there is an exact sequence

$$0 \to \frac{A(K) \cap mA(K')}{mA(K)} \to \frac{A(K)}{mA(K)} \to \frac{A(K')}{mA(K')}.$$

The group $A(K) \cap mA(K')/mA(K)$ injects into the set Map(Gal$(K', K), A[m])$. But Gal(K', K) and $A[m]$ are both finite, and so $A(K) \cap mA(K')/mA(K)$ is finite. Since the outside terms of the exact sequence are finite, so is the middle term, $A(K)/mA(K)$. We actually won't need this until a little later in the proof; but this justifies our assumption that $A[m] \subset A(K)$, and A is defined over K.

3.1 Multiplication by $[m]$

We will need a good understanding of the map $[m]$ in order for the proof to work. By using the techniques of algebraic geometry, we will arrive at the following proposition.

Proposition 3.2 *Let A be an abelian variety of degree d defined over a field F, m a positive integer relatively prime to char$F$. Then the map $[m] : A \to A$ is a surjective, separable morphism of degree m^{2d}. In particular, $\ker[m] = A(F)[m] \cong (\mathbb{Z}/m\mathbb{Z})^{2d}$.*

There isn't space here to properly develop all of the necessary concepts *ab initio*. Elementary algebraic geometry (and quite a bit more) may be found in [Har]. As a starting point, assume Mumford's theorem of the cube [Mum 2]:

Proposition 3.3 *Let D be a divisor on A, $I \subseteq \{1,2,3\}$ and S_I a morphism*

$$\begin{array}{rccc} S_I: & A \times A \times A & \to & A \\ & (x_1, x_2, x_3) & \mapsto & \sum_{i \in I} x_i. \end{array}$$

Then

$$\sum_{i \in I} (-1)^{|I|} S_I^* D = 0.$$

An easy corollary is that if f, g and h are morphisms from A to A, and D is a divisor on A, then

$$(f+g+h)^*D - (f+g)^*D - (f+h)^*D - (g+h)^*D + f^*D + g^*D + h^*D \sim 0.$$

We'll now use this information to characterize $[m]$. Take f to be $[m]$, g the identity function [1], and $h = [-1]$ the inverse map. Clearly, $[m] + [m'] = [m + m']$. Then

$$[m]^*D - [m+1]^*D - [m-1]^*D + [m]^*D + D + [-1]^*D \sim 0.$$

Taking a cue from [Hin] observe that in general, given a group G and a map $f : \mathbb{Z} \to G$ such that $f(n+1) + f(n-1) - 2f(n) = a$, it's true that $f(n) = ((n^2 - n)/2)a + n(f(1) - f(0)) + f(0)$. The proof of this is an easy induction on n. Applying this lemma with $f(n) = [n]^*D$ yields

$$[m]^*D \sim \frac{m^2+m}{2} D + \frac{m^2-m}{2} [-1]^*D.$$

This can be used to show that $[m]$ is surjective. Take D to be an ample divisor on A, $0 \notin D$. Since $[-1]$ is an automorphism of A, $[-1]^*D$ is ample, as well. Applying the formula gives $[m]^*D$ is an ample divisor. Certainly $[m]^*D \cap \ker[m] = \emptyset$. But $[m]^*D$ is ample; therefore, $\ker[m]$ must be of dimension zero. Since $\dim \ker[m] + \dim \text{im}\,[m] = d$, we have $\dim \text{im}\,[m] = d$, and $[m]$ is surjective of finite type. Let $p = \text{char}F$. The map $[m]$ is separable if p does not divide the degree of $[m]$.

It remains to prove the second part of the proposition. On an abelian variety of dimension d, it makes sense to talk about the intersection index of a set of d divisors. In general, if we have a surjective morphism $f : X \to Y$, and divisors $D_1, \ldots, D_d$ on Y, then the intersection numbers are related by

$$(f^*D_1 \cdot f^*D_2 \cdots f^*D_d) = (\text{deg}f)(D_1 \cdots D_d).$$

Now, pick an ample symmetric divisor D on A, that is, $[-1]^*D = D$. Then we know that $[m]^*D \sim m^2 D$. Taking the d-fold self-intersection of D yields

$$\begin{aligned}
\text{deg}[m]\,\overbrace{(D \cdot D \cdot \cdots \cdot D)}^{d \text{ times}} &= ([m]^*D \cdot [m]^*D \cdots [m]^*D) \\
&= (m^2 D \cdot m^2 D \cdots m^2 D) \\
&= m^{2d}(D \cdot D \cdot \cdots \cdot D) \\
\text{deg}[m] &= m^{2d}.
\end{aligned}$$

This gives the size of the m-torsion group, m^{2d}. If m is prime, this is sufficient to characterize the structure of the group. Otherwise, one can prove by induction on the factorization of m that $A[m] \cong (\mathbb{Z}/m\mathbb{Z})^{2d}$. This completes the proof of proposition 3.2.

3.2 The Kummer pairing

If L is any extension of K, then the Galois group Gal(L, K) acts on points $P \in A(L)$ by acting on the coordinates of P. Now, it turns out that the map $[m] : A \to A$ is surjective. This means that, for any point P on the variety, we can find a point Q so that $[m]Q = mQ = P$. With this in mind, we can define the Kummer pairing:

$$\begin{array}{rrcl} \kappa: & \mathrm{Gal}(\overline{K}, K) \times A(K) & \to & A[m] \\ & \sigma \times P & \mapsto & \sigma(Q) - Q \text{ for some } Q : mQ = P. \end{array}$$

It's not obvious that this is actually a well-defined map; for we have some freedom about the choice of Q, and the result might not actually be an m-torsion point. To prove the latter, observe that

$$m(\kappa(\sigma, P)) = m(\sigma(Q) - Q) = m\sigma(Q) - mQ = \sigma(mQ) - P = \sigma(P) - P = \mathcal{O}.$$

The last step follows because $P \in A(K)$, and σ holds K fixed. As for the former contention, suppose that $mQ = mR = P$. Then $m(R - Q) = 0$, and $R' = R - Q \in A[m]$. Writing $R = Q + R'$,

$$\sigma(R) - R = \sigma(Q + R') - (Q + R') = \sigma(Q) - Q + \sigma(R') - R' = \sigma(Q) - Q + R' - R' = \sigma(Q) - Q.$$

The penultimate step works because we assumed that $A[m] \subset A(K)$, so R' is defined over K and σ holds R' fixed. This shows that the pairing is well-defined.

One can further show that $\kappa(\sigma + \tau, P) = \kappa(\sigma, P) + \kappa(\tau, P)$, and $\kappa(\sigma, P + Q) = \kappa(\sigma, P) + \kappa(\sigma, Q)$. In other words, κ is a bilinear pairing.

We can characterize the kernel of this pairing on the left and the right. Let $L = K([m]^{-1}A(K))$ be the compositum of $K(Q)$ for all $Q \in A(\overline{K})$ such that $mQ \in A(K)$.

Proposition 3.4 *The kernel of the Kummer pairing κ is Gal$(\overline{K}, L)$ on the left, and $mA(K)$ on the right.*

As an immediate corollary, we find that the induced pairing

$$\mathrm{Gal}(L, K) \times A(K)/mA(K) \to A[m]$$

is a perfect bilinear pairing. Since $A[m]$ is finite, if we can show that L is a finite extension then we know that $A(K)/mA(K)$ is finite.

First, consider the kernel on the left. Suppose that $\sigma \in \text{Gal}(\overline{K}, L)$. Then $\kappa(\sigma, P) = \sigma(Q) - Q = \mathcal{O}$, since Q is defined over L, and σ holds L fixed. For the other inclusion, suppose that $\sigma \in \text{Gal}(\overline{K}, K)$ is such that $\kappa(\sigma, P) = \mathcal{O}$ for all $P \in A(K)$. Then σ holds $[m]^{-1}P$ fixed for all P. This is precisely how L is defined; σ fixes L, and $\sigma \in \text{Gal}(\overline{K}, L)$.

Second, consider the kernel on the right. Suppose that $\kappa(\sigma, P) = \mathcal{O}$ for all $\sigma \in \text{Gal}(\overline{K}, K)$. Then for any Q such that $mQ = P$, we have $\sigma(Q) - Q = \mathcal{O}$ for all σ. Since Q is held fixed by all the σ, it is defined over K; $P = mQ \in mA(K)$. It's similarly easy to show that any $P \in mA(K)$ is in the kernel on the right.

Now we have to show that L is a finite extension of K. To do this, we will prove that L is ramified at only finitely many places of K, and that the group $\text{Gal}(L, K)$ is abelian of exponent m. This will require information about how the reduced abelian variety $A(K)$ behaves at at residue fields corresponding to nonarchimedean valuations v. Then, we use this knowledge to show that $[L : K]$ is finite.

It's easy to show that L is an abelian extension of exponent m over K. This simply means that $\text{Gal}(L, K)$ is abelian, and every $\sigma \in \text{Gal}(L, K)$ has order dividing m. The kernel on the left of the Kummer pairing is $\text{Gal}(\overline{K}, L)$, and so we have an injection

$$\text{Gal}(L, K) \hookrightarrow \text{Hom}(A(K), A[m]).$$

Recall that $A[m]$ is abelian of exponent m. The set of homomorphisms $\text{Hom}(A(K), A[m])$ forms a group with the law $(\sigma + \tau)(P) = \sigma(P) + \tau(P)$. Since $A[m]$ is abelian, so is $\text{Hom}(A(K), A[m])$. Furthermore, any map from a group into $A[m]$ will necessarily be of order dividing m. Since $\text{Gal}(L, K)$ injects into the group $\text{Hom}(A(K), A[m])$, it too enjoys these properties.

3.3 Reduction

We now have to show that L is unramified outside of some finite set of places, S. This requires a theory of the reduction of an abelian variety. Arguably, such a discussion should take place using the language of schemes. Instead, I shall try to keep the exposition as concrete as possible. Initially we will take a detailed look at the reduction mod p of rational elliptic curves. The concepts introduced will be generalized to the reduction of an abelian variety at a nonarchimedean place.

Let E be an elliptic curve $y^2 = x^3 + ax^2 + bx + c$, where a, b and c are integers, and let p be a prime. We can look at the curve "mod p." Let $\widetilde{a}$ denote the coset corresponding to a in $\mathbb{Z}/p\mathbb{Z}$. Then there is a curve $\widetilde{E} : y^2 = x^3 + \widetilde{a}x^2 + \widetilde{b}x + \widetilde{c}$ defined over $\mathbb{Z}/p\mathbb{Z}$. When will this curve be nonsingular? Recall that a curve is singular if and only if its discriminant D is nonzero. The argument which demonstrates this in the rational case also works in $\mathbb{Z}/p\mathbb{Z}$. The discriminant of $\widetilde{E}$ is simply $\widetilde{D}$, since reduction mod p commutes with the various arithmetic operations. Thus, $\widetilde{E}$ is singular only when $p|D$, for that is precisely when $D \equiv 0 \bmod p$. We say that a curve has good reduction at p if the reduced curve $\widetilde{E}$ is nonsingular. Any curve has good reduction everywhere except at the finitely many primes which divide its discriminant.

We can also reduce a point $P = (x, y) \in E(\mathbb{Q})$, provided that p doesn't divide the denominators of x or y. It's not hard to see that $\widetilde{P} \in \widetilde{E}(\mathbb{Z}/p\mathbb{Z})$. Since all torsion points have integer coordinates, we get a well-defined homomorphism of groups.

$$\begin{array}{rcl} E(\mathbb{Q})_{tor} & \to & \widetilde{E}(\mathbb{Z}/p\mathbb{Z}) \\ (x, y) & \mapsto & (\widetilde{x}, \widetilde{y}) \\ \mathcal{O} & \mapsto & \widetilde{\mathcal{O}}. \end{array}$$

The only element which gets sent to $\widetilde{\mathcal{O}}$ is $\mathcal{O}$, and so the map is actually an injection. As long as $p \nmid D$, $E(\mathbb{Q})_{tor} \hookrightarrow \widetilde{E}(\mathbb{Z}/p\mathbb{Z})$. In particular, $E(\mathbb{Q})[m] \hookrightarrow \widetilde{E}(\mathbb{Z}/p\mathbb{Z})[m]$.

These concepts generalize in a couple of ways. First, we can consider points over any number field K. Recall some basic facts about nonarchimedean valuations, most of which are in the first few pages of [Lan]. The reader is also invited to examine the eminently lucid exposition [Cas 2].

Let $\mathbf{O}$ be the ring of integers of K. Let M_K be a complete set of distinct valuations on K, and $M_K^0 \subset M_K$ the set of nonarchimedean valuations. A $v \in M_K^0$ corresponds to some prime ideal $\mathbf{p}$ of $\mathbf{O}$. Associated to such an ideal is the completion $K_\mathbf{p}$ and the residue field $k = \mathbf{O}/\mathbf{p}$. With this notation, we can still reduce a curve $E(K)$ as above, obtaining $\widetilde{E}$ defined over k_v. Again, the reduced curve $\widetilde{E}$ will be nonsingular so long as $\mathbf{p}$ does not divide D, that is, $\text{ord}_\mathbf{p}(D) = 0$.

Second, we can work with any abelian variety A. Such a variety has an embedding into some projective space $\mathbb{P}^n$. (See [Mum 1] for details.) The variety can thus be described as the intersection of finitely many polynomials $f_i(x_0, x_1, \ldots, x_n)$. The obvious choice for $\widetilde{A}$, the reduction of $A \bmod \mathbf{p}$, is the intersection of the $\widetilde{f}_i$, where $\widetilde{f}$ is obtained from f by simply reducing the coefficients.

The description of singularity here will be in terms of the various properties of polynomials. One probably wouldn't want to actually compute any of the polynomials involved. However, the resulting theory can be used to show that A almost always has good reduction, without recourse to conceptually high-powered machinery.

As before, let d be the dimension of A. The Jacobian matrix J of A is given by $||(\partial f_i/\partial x_j)||$. (The differentiation is purely formal; we are not working with a notion of continuity here.) We say that A is nonsingular at a point P if evaluating J at P yields a matrix of rank $n - d$. Equivalently, J is singular at P if all of the $(n-d) \times (n-d)$ submatrices have determinant zero. Now, the determinant of a matrix is given by a polynomial in the entries of that matrix. Furthermore, even though there are lots of $(n-d) \times (n-d)$ submatrices of J, there are certainly only finitely many of them. So A is singular at a point P if and only if each of the determinant polynomials D_i vanishes at P.

At this point we turn to the classical tool of elimination theory, which can tell us when all of these polynomials have a common zero (see section 9 of [Sch]). Here, it suffices to recall that there is a polynomial expression D in the coefficients of the D_i, which is equal to zero if and only if the D_i have a common nontrivial zero. Since A is nonsingular by definition, D is nonzero. The reduction of $\widetilde{A}$ is accomplished in the expected way; $\widetilde{A}$ is nonsingular if $\widetilde{J}$ has the expected rank, that is, if $\widetilde{D} \not\equiv 0 \bmod v$. Since D has nontrivial valuation at only finitely many of the v, $\widetilde{A}$ is nonsingular except at finitely many places. One can similarly show that $\widetilde{A}$ is actually has a group law on it at almost all places [Ner]. We formalize this information in the following proposition.

Proposition 3.5 *A has good reduction at almost all places; there is a finite set $S \subset M_K$ so that A has good reduction for all $v \in M_K \backslash S$.*

Let $S = \{v : \widetilde{A} \pmod v$ is not an abelian variety $\}$. The following lemma will be useful below.

Proposition 3.6 *Let v be a nonarchimedean valuation, $v \notin S$. Then the reduction map $A[m] \hookrightarrow \widetilde{A}(k_v)$ is injective.*

By proposition 3.5, A will have good reduction at such a place, and thus be an abelian variety of the same dimension. Proposition 3.2 says that the size of $A[m]$ is independent of the field A is defined over, so long as certain degenerate cases are ruled out.

Proposition 3.7 *Let L be defined as above. If A has good reduction at v, then L is unramified at v.*

We can characterize ramification by the effect of the inertia group. Let K' be a Galois extension of K, v a place over K, and v' an extension of v to K'. The group Gal(K', K) acts on the residue fields k and k'. The inertia group $I_{v,v'}$ is the set of elements of Gal(K', K) which act trivially on k'. We say that a subset of k' is unramified at v if it is held fixed by $I_{v,v'}$.

Recall that L is the compositum of fields $K(Q)$, where $mQ \in A(K)$. Certainly if $K(Q)$ is unramified at v for all Q, then L is unramified at v. Using the notation introduced above, we look at the action of the inertia group $I_{v,v'}$ on Q. Take some $\sigma \in I_{v,v'}$. Ultimately, we want to show that σ holds Q fixed, that is, $\sigma(Q) - Q = \mathcal{O}$. By definition of the inertia group, σ acts trivially on $\widetilde{E}_{v'}(k')$. Hence,

$$\widetilde{\sigma(Q) - Q} = \sigma(\widetilde{Q}) - \widetilde{Q} = \widetilde{\mathcal{O}}.$$

Furthermore, $\sigma(Q) - Q$ is a point of order m:

$$m(\sigma(Q) - Q) = \sigma(mQ) - mQ = \sigma(P) - P = \mathcal{O}.$$

Now, $\sigma(Q) - Q$ is a point of order m, which gets sent to $\widetilde{\mathcal{O}}$ under reduction mod v'. But A has good reduction at v, and so the points of order m inject into $\widetilde{E}_{v'}(k')$. This means that $\sigma(Q) - Q = \mathcal{O}$, that is, Q is held fixed by σ. This concludes the proof that K' is unramified outside of S, and thus L is unramified outside S.

3.4 The final step

At this point, we have seen that L is abelian of exponent m, and that it is unramified outside of a finite set of places S. It remains to show that this means L is actually a finite extension of K.

Let R_S be the ring of S-integers, that is,

$$R_S = \{a \in K : v(a) \geq 0 \text{ for all } v \in M_K, v \notin S\}.$$

We can augment S by any finite set of places we choose; for if the resulting field L' is a finite extension of K, and $L \subseteq L'$, then L is a finite extension of K as well. Thus, we can assume that S has been chosen so that R_S is a principal ideal domain. (It will only take finitely many more places to accomplish this, since the class number is finite.) Additionally, assume that all the primitive m^{th} roots of unity ζ^j are contained in K.

Elementary Kummer theory shows that if L is abelian of exponent m, then it is obtained from K by adjoining m^{th} roots of elements of K. The reader is referred to [Lan] for a proof of this fact. Thus, $L \subset K(\sqrt[m]{a} : a \in K)$, and L is the maximal field unramified outside S. The goal is to show that only finitely many m^{th} roots need be adjoined to K in order to get L. We need the following proposition.

Proposition 3.8 *Let v be a discrete valuation on K with $v(m) = 0$. Then v is unramified in $K(\sqrt[m]{a})$ if and only if ord$_v(a) \equiv 0 \bmod m$.*

This makes intuitive sense, when one considers any particular $v = p$ over $K = \mathbb{Q}$. Obviously, the prime ideal $(5) \subset \mathbb{Q}$ will ramify in $\mathbb{Q}(\sqrt[3]{75})$, while it will not ramify in $\mathbb{Q}(\sqrt[3]{375}) = \mathbb{Q}(\sqrt[3]{3})$. The proposition is simply a formalization of this notion.

Since L is unramified at places outside of S, the m^{th} root of some $a \in K$ can be in L only if ord$_v(a) \equiv 0 \bmod m$ for all places $v \notin S$. Actually, since L is the maximal possible extension, the implication works both ways; L is the compositum of all fields $K(\sqrt[m]{a})$ for $a \in K$ such that ord$_v(a) \equiv 0 \bmod m$ for $v \notin S$. It's not hard to see that $K(\sqrt[m]{a}) = K(\sqrt[m]{a^{1+jm}})$. So when adjoining m^{th} roots, it suffices to take one representative for each coset in K^*/K^{*m}. At this point, all we have to do is establish the finitude of

$$T_S = \{a \in K^*/K^{*m} : \text{ord}_v(a) \equiv 0 \pmod{m} \text{ for } v \in M_K, v \notin S\}.$$

Indeed, we will see that $R_S^*/R_S^{*m} \cong T_S$. To see this, consider the natural map $R_S^* \to T_S$ given by $a \mapsto a \pmod{K^{*m}}$. This is a surjection. Let $a \in K^*$ be a representative for an element of T_S. Now, the prime ideals of R_S correspond to the valuations $v \notin S$. Furthermore, a's valuation everywhere outside of S is a multiple of m. So the ideal aR_S is actually the m^{th} power of an ideal in R_S. Since we've chosen S so that R_S is a principal ideal domain, we have $aR_S = b^m R_S$ for some $b \in K^*$, and $a = ub^m$ for some $u \in R_S^*$. Clearly, u and a correspond to the same element in T_S.

Thus, R_S^* maps onto T_S in the natural way.

It's not hard to see that R_S^{*m} gets mapped to the trivial element of T_S. So we know that R_S^*/R_S^{*m} maps onto T_S. It turns out that this map is actually an isomorphism, but we don't really need this fact. Let p_i be a uniformizer for each $v_i \in S$, and $u_1, \ldots, u_t$ be a set of generators for the units in R_S. (Theorem 2.6.4 of [Lan] shows that the group of S-units is finitely generated.) Then

$$\{\zeta^{\epsilon_0} p_1^{\epsilon_1} p_2^{\epsilon_2} \cdots p_r^{\epsilon_r} u_1^{\epsilon_{r+1}} \cdots u_t^{\epsilon_{r+t}} | \epsilon_i \in \{0, \ldots, m-1\}$$

is a set of representatives for for R_S^*/R_S^{*m}, and clearly finite.

This completes the proof of the weak Mordell-Weil theorem. To trace up the chain of implications, recall that we now know that R_S^*/R_S^{*m} is finite; T_S is finite; $[L : K]$ is finite; Gal(L, K) is finite; $A(K)/mA(K)$ is finite, as desired.

4 Implementation

I have written a program which attempts to compute the rank of $E(\mathbb{Q})$ for an elliptic curve of the form $y^2 = x^3 + ax^2 + bx$. (Recall that setting $c = 0$ ensures that there will be at least one point of order two, $(0, 0)$, on the curve. This, in turn, is necessary for the detailed proof of the Mordell-Weil theorem given above.) In this section I discuss the construction of the program and some of the trade-offs involved in its implementation.

My ultimate goal was to craft a routine which would take a description of an elliptic curve and return the rank of that curve. Early on, I had to decide which computer language I would use for the implementation. Obvious choices include Mathematica, C++, LISP and C. The first three languages are attractive in that they have more-or-less transparent support of arbitrary-precision integer arithmetic. This is useful — in fact, vital — for attempting to find points on the covering spaces.

Joseph Silverman has already written an elliptic curve computation package in Mathematica, and at first I hoped to interface my code with his. However, I find the Mathematica environment quite poor for program development. Also, Mathematica maintains most of its information in such generality that the actual execution time is extremely unsatisfactory. Finally, even though Mathematica is widespread in the academic community, it *is* a private piece of software, and my personal bias is towards making software as freely available as possible.

I held somewhat similar reservations about C++; it simply isn't as available as LISP and C. So for me, the choice came down to LISP or C. At this point, it was a matter of personal preference; I am more familiar with C. In particular, I know I can write faster code in C then I can in LISP. The functional-programming paradigm didn't offer many advantages for this particular project. Thus, the program is written in C.

Since C doesn't have native support for larger (> 32 bit) integer arithmetic, I had to add in my own. There are a few packages available for arbitrary precision arithmetic. Typically these are slowed down by frequent dynamic memory operations. Additionally, different packages are optimized for different sized integers. Pari, for example, performs best on numbers of forty or fifty digits. However, the program for computing the rank really only requires integers on the order of twenty digits, and often the numbers are much smaller. So, using [Knu] as my guide I wrote `hp.c`, a high-precision integer arithmetic library.

The file `numth.c` contains a number of useful number-theoretic algorithms, such as greatest common divisor, primality, factorization, etc. The factorization routine in particular is quite primitive. Profiling experiments reveal that it isn't a bottleneck, so I have focused my attentions elsewhere. However, it is an obvious site for improvements. The package `lists.h` is a very basic set of primitives for maintaining list-like arrays. The lists have a compile-time maximum size; operations supported are creation, insertion and deletion of elements, list-length inquiry, random-access address, and disposal. It is far from general, but it performs the necessary job for the larger program.

The actual algorithm for computing the rank of the Mordell-Weil group and auxiliary routines are found in `rank.c`. At its highest level, the algorithm is quite simple.

```
EC_RANK(a,b)                          /* Returns the rank of y^2 = x^3 + ax^2 + bx. */
r ← rank( α(a,b) )                    /* Find rank of image of φ. */
r̄ ← rank( α(ā,b̄) )                    /* Find rank of image of ψ. */
rank ← r + r̄ − 2
return( rank )
```

The tricky part is determining the size of the images $\phi(E(\mathbb{Q}))$ and $\psi(\overline{E}(\mathbb{Q}))$. Recall that these images inject into $\mathbb{Q}^*/\mathbb{Q}^{*2}$. There's an obvious approach to computing the rank of $\alpha(a,b)$, which falls out of our proof of the weak Mordell-Weil theorem.

```
RANK(α(a,b))
candidates ← sq_free_fac(b)           /* Find square-free factors of b */
image ← ∅                             /* Image of φ(E(Q)) */
for each b1 ∈ candidates
  b2 ← b/b1
  if ( b1 m^4 + am^2 e^2 + b2 e^4 = n^2 ) has an integer solution
    image ← image ∪ {b1}
add in O and T
return( log2( —image— ) )
```

This would work well enough, were the computations described at every line effective. (The logarithm is taken in the last step in order to return the rank of the group, since every element has order 2.) However, there is no known algorithm which will determine if the equation $b_1m^4 + am^2e^2 + b_2e^4 = n^2$ has a nontrivial solution in a finite amount of time. All one can do is hope. Of course, some instances of this problem are easily solved. For example, if b_1, a and b_2 are all negative, there will clearly be no solution; there aren't even any real solutions. Additionally, if the equation has no solution mod p for some prime p, then there is no solution in the integers.

Checking for solutions mod p is a finite process; one must only let x and y take on all of the (finitely many) possible distinct values, and check explicitly for a solution.

If such considerations fail to rule out the existence of a solution, one is forced to try a brute-force check for integral solutions. This is typically done until either a solution is found, or the computation is arbitrarily aborted after some range.

This is why any "algorithm" to compute the rank of an elliptic curve can really only be a heuristic, given the current state of the art. We can only hope that the computer successfully finds points or proves that they don't exist.

The algorithm presented above has a two disadvantages, even beyond the crucial lack of effectivity. First, in practice the algorithm often finds an image set with an impossible number of elements. Since the image of α is a subgroup of R^*/R^{*2}, whose order is a power of 2, the size of the image must be a power of 2, as well. A second, related criticism is that the algorithm ignores the group structure of im α; it merely treats it as a set. It is not implausible that a better algorithm could be crafted by taking advantage of the group structure.

Indeed, this is the case. The following algorithm maintains separate sets for b_1's which definitely give rise to a point in the image, b_1's which definitely don't, and b_1's it couldn't definitely say either about. It takes advantage of the group structure in the following way. Suppose $H \subset G$ is a subgroup. If $h \in H$, $g \in G$ and their product hg is in H, then g must be in H, as well. Similarly, if $k \notin H$, and $kg \in H$, then $g \notin H$.

```
RANK(α(a,b))
S ← nonarch_places(b)          /* Find the nonarchimedean places dividing (b) */
def_in ← ∅
def_out ← ∅
for each g in ⟨S⟩                      /* ⟨S⟩ is the group generated by S */
  if g ∈⟨def_in⟩
    end this iteration of main loop
  for each k ∈def_out                              /* def_out isn't a group */
    if g·k ∈⟨def_in⟩
      def_out ← def_out ∪ {g}
```

```
      end this iteration of main loop
                                                  /* We have to do some actual work */
    b1 ← g
    b2 ← b/b1
    switch( solvable(b1 m^4 + a m^2 e^2 + b2 e^4 = n^2) )   /* One of three things happens */
      case is_sol:
        def_in ← def_in ∪ {g}
      case no_sol:
        def_out ← def_in ∪ {g}
      case cant_tell:
        dont_know ← def_in ∪ {g}
  check again to clean up dont_know
  return( rank(def_in) )
```

This is the algorithm I actually used. Most of the details are self-explanatory. I would only add a note about the implementation of the group law on R^*/R^{*2}. Recall that this group has coset representatives

$$\{(-1)^{\epsilon_0} p_1^{\epsilon_1} p_2^{\epsilon_2} \cdots p_r^{\epsilon_r} | \epsilon_i \in \{0, 1\}\} \subset \mathbb{Q}^*/\mathbb{Q}^{*2}.$$

So on a computer, it makes sense to represent any element of R^*/R^{*2} as an $n+1$-bit word. As for implementing the group law, we have $p \cdot p = 1 \cdot 1 = 1$, $p \cdot 1 = 1 \cdot p = p$ for each prime p. So if we have two elements $g = g_0 g_1 \cdots g_n$ and $h = h_0 h_1 \cdots h_n$, then their product is given by $(gh)_i = g_i + h_i \mod 2$. Bitwise addition mod 2 can be accomplished by exclusive-or'ing the two values together; thus, $gh = XOR(g, h)$.

In computing whether an element is in the span of a set of elements, one essentially performs an orthogonalization operation. I find it interesting that a similar computation arises in the work of Brumer, Kramer and McGuinness [BKM]. As part of their algorithm to compute the 2-Selmer group of an elliptic curve, they wind up inverting a matrix over $\mathbb{Z}/2\mathbb{Z}$.

Even though I wrote the program to be as efficient as I could, it still takes several seconds to compute the rank of an elliptic curve. While this may not be prohibitive to the casual investigator, it becomes a problem when one is looking at a large number of curves. I wanted to compute the rank of about one hundred thousand curves. (The choice of curves is discussed below.) Rather than wait sixty days for one SPARCstation to complete the calculations, I harnessed sixty SPARCstations for one day. These computations lend themselves to a coarse-grained distributed network. Every computer in the network can be assigned a range of curves to look at. Such an assignment is simple, as the family of curves $y^2 = x^3 + ax^2 + bx$ has an obvious parameterization. Each computer can industriously work on its segment of the problem space, without need for communication with other computers. I chose to implement parallelism using Curtis Yarvin's QuaHog system [Yar]. QuaHog daemons are running on most SPARCstations in the Brown University computer science department. Each daemon monitors the load level on its machine. The driver program tells its local QuaHog daemon that it has sixty jobs, say, and roughly describes their resource requirements. The jobs are farmed out appropriately to idle nodes throughout the network. The system is designed to cause as little trauma to other users as possible. If a node where a QuaHog job is running is suddenly getting heavy use, the job will exit gracefully. The master program then resubmits the job, which will then run on another node. Since the basic program was written to be easily restarted, there is fairly graceful process migration.

5 Results and Discussion

The rank of elliptic curves has been an object of computer-aided investigation for a relatively long time. By the 1950's, Birch and Swinnerton-Dyer were letting results from computer experiments guide their conjectures [B-S]. They performed a 2-descent similar to the one described in section 4 on curves of the form $y^2 = x^3 - Ax - B$, $y^2 = x^3 - A$, and $y^2 = x^3 - Ax$. Relying particularly on numerical evidence gained from computing the rank of curves $y^2 = x^3 - Ax$, they formulated a two-part conjecture. A gentle explanation of the Birch-Swinnerton-Dyer conjecture is outlined in [Zag]. Essentially, it relates the local behavior of an elliptic curve E to its global behavior. The

L-series of E is

$$L(E,s) = \prod_p \frac{1}{1 - a(p)p^{-s} + \epsilon(p)p^{1-2s}} \qquad (\mathrm{Re}(s) > \frac{3}{2}).$$

Here, the product is taken over all primes p, $a(p) = \#\widetilde{E}(\mathbb{Z}/p\mathbb{Z})$, and $\epsilon(p)$ is zero if p divides the discriminant of E, and one otherwise. The first part of the BSD conjecture may be stated as follows.

Conjecture 5.1 *The L-series of E continues meromorphically to $s = 1$, and it has a zero there with order equal to the rank of $E(\mathbb{Q})$.*

Birch and Swinnerton-Dyer's calculations bore out this prediction on the curves they examined. Since then, a considerable body of evidence for the conjecture has accumulated. Some of this is numerical; we will consider such results shortly. Some of the evidence is more theoretical; implications of the conjecture have been verified for certain classes of elliptic curves. Note that, if true, the BSD conjecture would provide an effective method for determining the rank of an elliptic curve. We would merely have to compute $L(E,1)$, $L'(E,1)$, $L''(E,1)$, etc., to some desired degree of accuracy.

Let p be a prime, $p \equiv 5 \bmod 8$. The rank of the curve $y^2 = x^3 + px$ cannot be bigger than 1. Conjecturally this bound is always reached, and the rank of such a curve is 1. Bremner and Cassels produced points of infinite order for all such p less than 1000 [B-C]. Taking advantage of the structure of this particular class of curves, they actually carried out a second descent. (The method of section 4 is sometimes called a first descent; finding points on E is reduced to finding points on each of a finite set of curves. In a second descent, one examines each of those covering spaces by a further descent process.) This was necessary, as a straight brute force search might not encounter a generator in a reasonable amount of time. Indeed, Bremner and Cassels have characterized the generators as sometimes being "rather large." As an example they cite the generator for $y^2 =$

$x^3 + 877x$, whose x-coordinate is

$$\frac{375,494,528,127,162,193,105,504,069,942,092,792,346,201}{6,215,987,776,871,505,425,463,220,780,697,238,044,100}.$$

Using a computer search, Bremner extended this result for such curves with $1000 < p < 5000$ [Bre]. He also computed $L'(E,1)$ for all $1000 < p < 20,000$, and found it to be nonzero. Thus, the computer produced numerical evidence that the BSD conjecture is compatible with the conjecture for the rank of $y^2 = x^3 + px$.

Computers are now sufficiently powerful that we can investigate the behavior of large classes of curves. Zagier and Kramarz examined the L-series for elliptic curves $x^3 + y^3 = m$ (or, equivalently, $y^2 = x^3 - 432m^2$) [Z-K]. They assumed that the BSD conjecture is true, and thus were able to compute a rank for each of the curves. Before their work, there was apparently a sort of folk conjecture that the rank of an elliptic curves is as low as possible. The BSD conjecture says that the parity of the rank is even or odd depending on whether the sign of the functional equation is $+$ or $-$. It was believed that, for any large sample of curves, half of them would have rank zero, and half have rank one, with higher ranks occurring in a vanishingly small percentage of the curves. What they found really didn't support this conjecture. Rather, they found that about a quarter (23.3%) of the curves with an even functional equation actually have rank 2 or higher. Additionally, ranks of 3 or higher occurred in about 2% of the curves with odd rank. While this is a small percentage, it could well be a significant deviation from zero; a set of 70,000 curves was examined.

Perhaps a few words of caution about such computer experiments are appropriate here. The obvious source of error is an error in programming. This seems to have been a particular concern of Birch and Swinnerton-Dyer.

> On the other hand, in translating an elaborate scheme of calculation into a machine program one is bound to make mistakes. Most of these are found before the program

> is used for production runs; they show up because the program grinds to a halt or produces ridiculous results. But a program which is believed to work may still contain logical errors which only have an effect in rare circumstances: and indeed most computers have anomalies which cause them occasionally not to behave in the way that their specifications suggest. In fact, our program for stage (ii) was imperfect in that a very few equivalences were missed by the machine.
>
> For these reasons we believe that results obtained from a computer should not be automatically trusted. In some cases they can be checked because they have properties which were not essentially used in the course of the calculation and which would be unlikely to survive if an error had been made.... But if checks of this sort are not available, results should not be fully trusted until they have been independently reproduced by a different programmer using a different machine. We do not think this sets an unreasonable standard, now that computers are becoming so widely available; and we are satisfied that lower standards have already led to a number of untrue results being published and believed. [B-S]

Statistical experiments such as [Z-K] raise additional methodological questions. Fundamentally, one has to decide if one's sample set of curves is representative of the behavior of all elliptic curves. There really isn't any way to definitively answer this.

Brumer and McGuinness argue that the set of curves with prime conductor is typical of the whole set [B-M]. (The conductor of a curve, like its discriminant, measures its singularity at various reductions.) They had a Macintosh II estimate the rank of the 310,716 elliptic curves with prime conductor less than 10^8. The rank was primarily computed using the parity conjecture and the order of vanishing of the L-series. Additionally, they searched for independent points as a lower bound for the rank. Thus, assuming the BSD conjecture, they found the following rank distribution. Results for curves positive and negative discriminants are shown separately, as Brumer and McGuinness argue that the difference in rank distribution is significant.

rank	0	1	2	3	4	5
$D > 0$	27.87	45.51	21.68	4.62	0.33	0.0
$D < 0$	31.30	46.41	18.71	3.35	0.21	0.0025
all	30.04	46.08	19.80	3.82	0.26	0.0016

This roughly corroborates the findings of Zagier and Kramarz; they found that 23.9% of the curves have rank at least 2, which is far from the earlier near-zero estimate.

I used the algorithm developed in this paper to predict the rank of all nonsingular curves $y^2 = x^3 + ax^2 + bx$ with $-30 \leq a \leq 30$ and $-900 \leq b \leq 900$. In a sense, the difficulty of finding the rank grows linearly in the size of b, and quadratically in the size of a; this explains the shape of the chosen search space. Recall that the key step of computing the rank is finding an integral point on a 2-covering space of the form $N^2 = qM^4 + \overline{a}M^2e^2 + q'e^4$. The program attempts to prove that no such solution exists using certain elementary considerations. After that, it searches for solutions with $|M|, |e| \leq 40$.

Of the 107,100 curves examined, it was able to precisely determine the rank of 16,719. The observed rank distributions match up nicely with those reported by Brumer and McGuinness. For purposes of comparison, separate statistics are given here for curves with positive and negative discriminant. However, it does not appear that the sign of the discriminant systematically influenced the rank of the curve.

rank	0	1	2	3	4
$D > 0$	29.12	45.36	20.99	4.25	0.27
$D < 0$	28.96	45.76	21.09	3.90	0.29
all	29.06	45.54	21.03	4.09	0.28

In my opinion, the similarity between these numbeers and those of Brumer and McGuinness is interesting for a number of reasons. First, it obviously lends support to their hypotheses about the distribution of rank. Second, this may say something about the set of all elliptic curves. Recall that Brumer and McGuinness had examined the set of elliptic curves with (small) prime conductor. I looked at a collection of curves with small discriminant. Furthermore, the numbers above represent only the curves for which I was able to precisely determine the rank. It is intriguing that these two seemingly disparate sets correspond so well. One could tentatively say that the numbers reported above are characteristic of any large set of elliptic curves.

This summarizes the situation for the curves whose rank was successfully computed. What can we say about the remaining ones? In each such case, the algorithm could neither prove nor disprove the existence of an integer point on certain curves. Thus, it can compute lower and upper bounds for them. (Obviously, these lower and upper bounds agree when it has actually computed the rank.) The lower and upper bounds were distributed as follows:

rank	0	1	2	3	4	5	6	7	8
min	46.17	39.02	12.85	1.86	0.11	0.002			
max	4.54	17.87	28.60	26.94	15.44	5.54	0.99	0.074	0.002

Even though the curves where the lower and upper rank bounds differ may seem like failures, we can still infer some information about the Selmer Tate-Shafarevich group associated to the curve, Sel $=$ Sel$(E, \mathbb{Q})$ and Ш $=$ Ш$(E, \mathbb{Q})$. A full discussion of these groups is left to chapter X of [Sil]. Here we merely recall that associated to every isogeny of elliptic curves defined over K $\phi : E(K) \to E'(K)$ there is an exact sequence

$$0 \to E(K) \to E(K) \xrightarrow{\phi} E'(K) \to \text{Sel}(E, K)[\phi] \to \text{Ш}(E, K)[\phi] \to 0.$$

In our case, the isogeny is $[2]$, $K = \mathbb{Q}$ and $E = E'$. Thus, there is an exact sequence

$$0 \to E(\mathbb{Q})/2E(\mathbb{Q}) \to \text{Sel}[2] \to \text{Ш}[2] \to 0.$$

All of the groups in the sequence are finite. The Selmer group is effectively computable. (See [BKM] for an actual algorithm.) So in order to determine what $E(\mathbb{Q})/2E(\mathbb{Q})$, we merely have to figure out what part of Sel[2] comes from $E(\mathbb{Q})/2E(\mathbb{Q})$, and what part of it comes from Ш$[2]$. Unfortunately, there is no algorithm guaranteed to do this.

As Silverman remarks, the group $\text{Ш}[2]$ can be interpreted as the group of 2-covering spaces which have an element mod p for every p. The identity element of this group is the class of covering spaces which actually have a $\mathbb{Q}$-rational point. This intuition explains why the sequence given above is exact.

Covering spaces for which the program is unable to produce an integral point or to rule one out could well correspond to elements of $\text{Ш}[2]$. In view of this, it appears that the family examined tends to have a nontrivial Ш.

Zagier and Kramarz that the percentage of curves with trivial Ш tends to zero asymptotically as one considers a set of elliptic curves. For such curves there is a mod p point on every curve which lifts to a rational point. In view of this, it is a bit surprising that my program only completely determined $E(\mathbb{Q})/2E(\mathbb{Q})$ for about one sixth of the curves. One possibility is that their family of curves is nonrepresentative, of the family $y^2 = x^3 + ax^2 + bx$ if not of all elliptic curves. They present some evidence that the family $x^3 + y^3 = m$ may be anomalous. I feel that the discrepancy is probably to the brevity of the search for integer points. There is simply no way that a first descent will produce the generating point of $y^2 = x^3 + 877x$. As a remedy, one could derive and implement second and third descents. This increases the number of curves one must examine, but (probably) reduces the height of points on each of the curves.

References

[Bre] A. Bremner, "On the equation $y^2 = x(x^2 + p)$," in *Number Theory and Applications*, ed. R. Mollin, Kluwer Academic Publishers, 1989, 3-22.

[B-C] A. Bremner and J.W.S. Cassels, "On the equation $y^2 = x(x^2 + p)$," *Mathematics of Computation*, 42:185 (1984), 257-264.

[B-S] B. Birch and H.P.F. Swinnerton-Dyer, "Notes on elliptic curves (I) and (II)," *J. Reine Angew. Math.*, 212 (1963), 7-25; 218 (1965), 79-108.

[BKM] A. Brumer, K. Kramer, and O. McGuinness, "Calculating Selmer groups of elliptic curves with all 2-torsion points rational," 1991 [to appear].

[B-M] A. Brumer and O. McGuinness, "The behavior of the Mordell-Weil group of elliptic curves," *Bull. Amer. Math. Soc.*, 23:2 (1990), 375-382.

[Cas 1] J.W.S. Cassels, "Diophantine equations with special reference to elliptic curves," *J. London Math. Soc.*, 41 (1966), 193-291.

[Cas 2] J.W.S. Cassels, *Local Fields*, Cambridge University Press, 1986 (New York).

[Har] R. Hartshorne, *Algebraic Geometry*, Springer-Verlag, 1977 (New York).

[Hin] M. Hindry, *Geometrie Arithmetique*, lecture notes, Université Paris, 1991.

[Knu] D. Knuth, *The Art of Computer Programming, Volume 2; Seminumerical Algorithms*, Addison-Wesley, 1981 (Reading).

[Lan] S. Lang, *Fundamentals of Diophantine Geometry*, Springer-Verlag, 1983 (New York).

[Mum 1] D. Mumford, "On the equations defining abelian varieties," *Inventiones Mathematicae*, 1 (1966), 287-354.

[Mum 2] D. Mumford, *Abelian Varieties*, Oxford University Press, 1970 (Bombay).

[Ner] A. Neron, *Geometrie Diophantienne*, lecture notes, Publications Mathematiques D'Orsay, 1967.

[Sch] A. Schinzel, *Selected Topics on Polynomials*, University of Michigan Press, 1982 (Ann Arbor).

[Shi] G. Shimura, "Reduction of algebraic varieties with respect to a discrete valuation of the basic field," *Amer. J. Math.*, 77 (1955), 134-176.

[Sil] J. Silverman, *The Arithmetic of Elliptic Curves*, Springer-Verlag, 1986 (New York).

[S-T] J. Silverman and J. Tate, *Rational Points on Elliptic Curves*, Springer-Verlag, 1992 (New York) [to appear].

[Tat] J. Tate, "The arithmetic of elliptic curves," *Inventiones Mathematicae*, 23 (1974), 179-206.

[Yar] C. Yarvin, *QuaHog*, computer software.

[Zag] D. Zagier, "The Birch-Swinnerton-Dyer conjecture from a naive point of view," *Arithmetic Algebraic Geometry*, ed. G. van derGeer, F. Oort, J. Steebrink, Birkhauser, 1991 (Boston), 377-389.

[Z-K] D. Zagier and G. Kramarz, "Numerical investigations related to the L-series of certain elliptic curves," *J. Indian Math. Soc.*, 52 (1987), 51-69.