On Pedagogy in Math

There sometimes exists an attitude in math that concepts are challenging to motivate absent a powerful application of them—something to point to and say, "look: this is why these ideas are useful". It's almost as if the great (and at several times unexpected) applications of math have forced the field to always justify itself, to always reassure pupils that everything they learn from it will not only serve them well but will be crucial in real, adult life. But his pedagogical plight is probably self-imposed; just as there is a time to emphasize the practical virtues of math, there are arguably many more times where educators should urge their students to embrace a more curious and artistic mindset—the same one that French mathematician Henri Poincaré espouses when he says [1],

> The scientist does not study nature because it is useful; he studies it because he delights in it, and he delights in it because it is beautiful. If nature were not beautiful, it would not be worth knowing, and if nature were not worth knowing, life would not be worth living. Of course I do not here speak of that beauty that strikes the senses, the beauty of qualities and appearances; not that I undervalue such beauty, far from it, but it has nothing to do with science; I mean that profounder beauty which comes from the harmonious order of the parts, and which a pure intelligence can grasp.

Granted, reciting this quote to a room full of bored, restless, and unmotivated students would very rarely change their disposition toward math. But the same can be said of any explanation of math's practical merits, and, if anything, emphasizing how fascinating math can be in its own right, removed from its practical merits, might be *more* likely to inspire interest than to hammer home its applications. After all (at least in principle) every student can appreciate and can try to grasp what is beautiful about math, but not all students, as it turns out, will actually use it. It's hard to imagine a student scoffing at the idea of attending an art class, complaining that there will never come a time where they will need to know what a primary or complementary color is or how to hold a paintbrush or fire pottery in a kiln or shade with a pencil—they know that that's not the point; they know that they are going to art class to have fun and that they might as well try and see what they get out of it. The same can and should be said for math. This mindset not only frees educators from justifying their subject but reminds them of the better ways to teach it. Would an educator that knows, deep down, that what makes math worth teaching is how it explores and represents and distills beautiful relationships cudgel their students over the head with formulas and their memorization and mindless application? It would seem very unlikely. It is the hope of the writer that this introduction to an introduction on elliptic curves will remind both those already fond and hopefully soon to be fond of math that most of what makes a subject interesting is not the subject itself, but how it is expressed and approached by educators and pupils alike. Inasmuch as this writer follows through on their end of the bargain, it is hoped that the reader will follow through on theirs.

Introduction to Elliptic Curves and Their Group Structure

Charles Kolozsvary Math 370

26 May 2023

Thus marks the exploration of a well-known and beautiful object in math: the elliptic curve. An elliptic curve is the set of solutions to a cubic polynomial equation of two variables over a field[1] ("over", here, means that the cubic polynomial takes members of the field as input). E.g., if K is a field and E is the cubic polynomial equation $y^2 = 3x^3 + 4x^2 + x - 3$, then $E(K)$ would be the set of points $(x, y) \in K$ that satisfy E. Using a change of variables, most cubic polynomials can be written as

$$y^2 = x^3 + ax + b$$

for variables x and y and constants a and b. Written in this manner, the polynomial is said to be in Weierstrass form. Consider the following example elliptic curve

$$E_1(\mathbb{R}) = \{(x, y) \in \mathbb{R}^2 \mid y^2 = x^3 - x + 1\}.$$

What can be said about it? We can immediately tell that $E_1(\mathbb{R})$ is symmetric about the x-axis since y takes both the positive and negative square root of $x^3 - x + 1$. Over $\mathbb{R}$, the square root of a negative number is not defined, so we should inquire where $x^3 - x + 1$ becomes negative, and thus, $E_1(\mathbb{R})$ is undefined. Let $g(x) = x^3 - x + 1$; we notice that $g(-1) = 1$ and $g(-2) = -5$. Therefore, by the intermediate value theorem, there exists some $x_0 \in \mathbb{R}$ such that $-2 < x_0 < -1$ and $g(x_0) = 0$. Thus, the domain of $E_1(\mathbb{R})$ is $[x_0, \infty)$. We can verify these observations by plotting $E_1(\mathbb{R})$ (see figure 1).

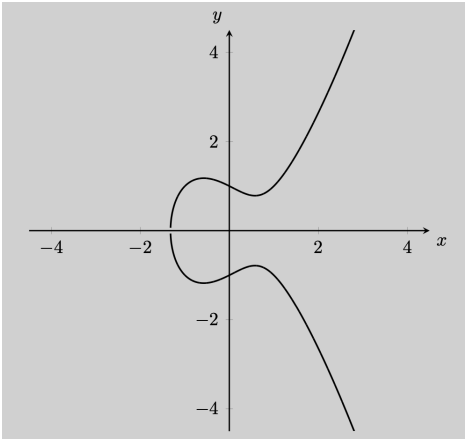

Figure 1: $E_1(\mathbb{R})$ graphed in the xy-plane. Notice that the curve is defined starting somewhere between -2 and -1 and how it is symmetric about the x-axis.

Another Example Elliptic Curve It so happened that $E_1(\mathbb{R})$ was continuous[2], but this will not necessarily be the case. Consider

$$E_2(\mathbb{R}) = \{(x, y) \in \mathbb{R}^2 \mid y^2 = x^3 - 4x + 2\},$$

and let $h(x) = x^3 - 4x + 2$. Plugging in some values for x into $h(x)$, we construct the following table:

[1]For our purposes a field can be understood as a collection of elements where one can add, subtract, multiply, and divide any two elements—excluding division by zero. The real numbers, denoted as $\mathbb{R}$, form what is likely the most familiar field to those who have studied calculus. Other examples include the rational numbers—comprised of every ratio of integers, denoted as $\mathbb{Q}$—and the complex numbers, denoted as $\mathbb{C}$, which are of the form $a + bi$, where $a, b \in \mathbb{R}$ and $i := \sqrt{-1}$.

[2]Continuous isn't technically the correct terminology. Since the elliptic curves discussed do not pass the vertical line test, they are not well defined, and hence they are not functions. Continuity is usually only considered a property of functions, but the word is here invoked to refer to the fact that the "shape" of the curve has no breaks or skips in it.

x	$h(x)$
2	2
1	-1
0	2
-2	2
-3	-13

Inspecting this table and recalling again the intermediate value theorem, we determine there must exist $x_1, x_2, x_3 \in \mathbb{R}$ where

$$-3 < x_1 < -2, \quad 0 < x_2 < 1, \quad 1 < x_3 < 2, \quad \text{and} \quad h(x_1) = h(x_2) = h(x_3) = 0.$$

And so it follows that the domain of $E_2(\mathbb{R})$ is $[x_1, x_2] \cup [x_3, \infty)$, which we can again verify by plotting $E_2(\mathbb{R})$ (see figure 2).

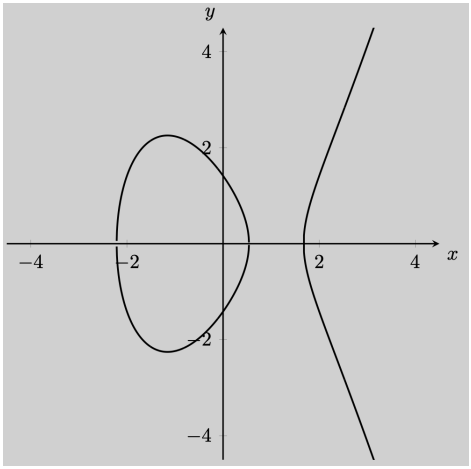

Figure 2: Notice $E_2(\mathbb{R})$ is discontinuous unlike $E_1(\mathbb{R})$ shown in figure 1.

1 Additional Structure Underlying Elliptic Curves

Coming up with example elliptic curves and plotting their solutions can offer only so much insight into these mathematical objects. Another natural avenue of inquiry is to ask what kinds of structures or relationships exist between the members of an elliptic curve E. That is to say that we should ask ourselves, "how might any two members of E be related?" Moreover, how could we, given two members of E, produce a third?

If we focus on elliptic curves like the ones already discussed (namely, those over $\mathbb{R}$), we can answer the above question geometrically: we draw a line between two given points on $E(\mathbb{R})$, then find where that line intersects the curve a third time (see figure 3).

This process of taking two points on an elliptic curve $E(\mathbb{R})$, drawing a line connecting them, then finding the other point of intersection with $E(\mathbb{R})$, can be thought of as a binary operation, which we'll denote as $\star$. For example, in figure 3, $P \star Q = M$[3].

With this conception of $\star$ in hand, a fairly straightforward question emerges: given the points $P = (\alpha_1, \beta_1)$ and $Q = (\alpha_2, \beta_2)$ on an elliptic curve satisfying $y^2 = x^3 + ax + b$, can we produce an explicit formula for $P \star Q = M = (\alpha_3, \beta_3)$ in terms of P and Q?

1.1 Using P and Q to find M To tackle this question, we should start by determining the equation of the line connecting P and Q in terms of P and Q themselves. We know that its slope is $m = \frac{\beta_2 - \beta_1}{\alpha_2 - \alpha_1}$, and the

[3]This is an intuitive definition of $\star$ that unfortunately (for reasons that will be discussed later) doesn't quite work but is very close to the actual definition.

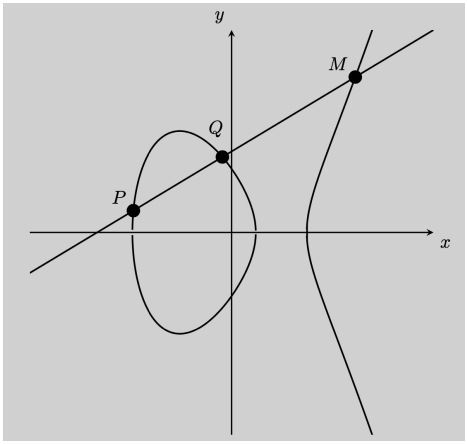

Figure 3: The line connecting points P and Q on the elliptic curve $E_2(\mathbb{R})$ intersects $E_2(\mathbb{R})$ at another point: M.

general equation of a line with slope m intersecting the y axis at h is written as

$$y = mx + h.$$

Therefore, $\beta_1 - m\alpha_1 = h$ since (α_1, β_1) is one of the points on the line connecting $P = (\alpha_1, \beta_1)$ and $Q = (\alpha_2, \beta_2)$ (unsurprisingly, it is also the case that $h = \beta_2 - m\alpha_2$). Determining where else this line intersects the elliptic curve merely involves substituting $y = mx + h$ into $y^2 = x^3 + ax + b$.

$$\begin{aligned} y^2 &= x^3 + ax + b \\ (mx + h)^2 &= x^3 + ax + b \\ m^2x^2 + 2mhx + h^2 &= x^3 + ax + b \\ 0 &= x^3 - m^2x^2 + (a - 2mh)x + b - h^2 = f(x) \end{aligned}$$

Since $P = (\alpha_1, \beta_1)$ and $Q = (\alpha_2, \beta_2)$ are already known to satisfy the equation of the line and the elliptic curve, $f(x)$ must factor as

$$f(x) = (x - \alpha_1)(x - \alpha_2)(x - \alpha_3).$$

If we expand this product of linear factors, the coefficients of each term must equal those of $x^3 - m^2x^2 + (a - 2mh)x + b - h^2$, which allows us to solve for α_3.

$$\begin{aligned} f(x) &= (x - \alpha_1)(x - \alpha_2)(x - \alpha_3) \\ &= \Big(x^2 - (\alpha_1 + \alpha_2)x + \alpha_1\alpha_2\Big)(x - \alpha_3) \\ &= x^3 - \alpha_3x^2 - (\alpha_1 + \alpha_2)x^2 + \alpha_3(\alpha_1 + \alpha_2)x + \alpha_1\alpha_2x - \alpha_1\alpha_2\alpha_3 \\ &= x^3 - (\alpha_1 + \alpha_2 + \alpha_3)x^2 + (\alpha_1\alpha_2 + \alpha_1\alpha_3 + \alpha_2\alpha_3)x - \alpha_1\alpha_2\alpha_3 \\ \implies -m^2 &= -(\alpha_1 + \alpha_2 + \alpha_3) \\ \alpha_3 &= m^2 - \alpha_1 - \alpha_2 \end{aligned}$$

To find β_3, we simply substitute α_3 for x in our equation of the line, which gives us that

$$\beta_3 = m\alpha_3 + h.$$

And thus we have successfully found $M = (\alpha_3, \beta_3)$ from P and Q—though, not entirely; there are some additional cases that we need to consider depending on P and Q.

1.2 Potential Problems With Finding M from P and Q What happens if $P = Q$, or even more worrisome, if $P = (\alpha, \beta)$ and $Q = (\alpha, -\beta)$ (see figure 5)? If the former is true, we notice that as Q comes closer to P, the line connecting them approaches the tangent line to the curve at P (depicted in figure 4)—this is what the line "connecting" P to itself becomes, and it is therefore the line we will use to find the other point of intersection with the curve.

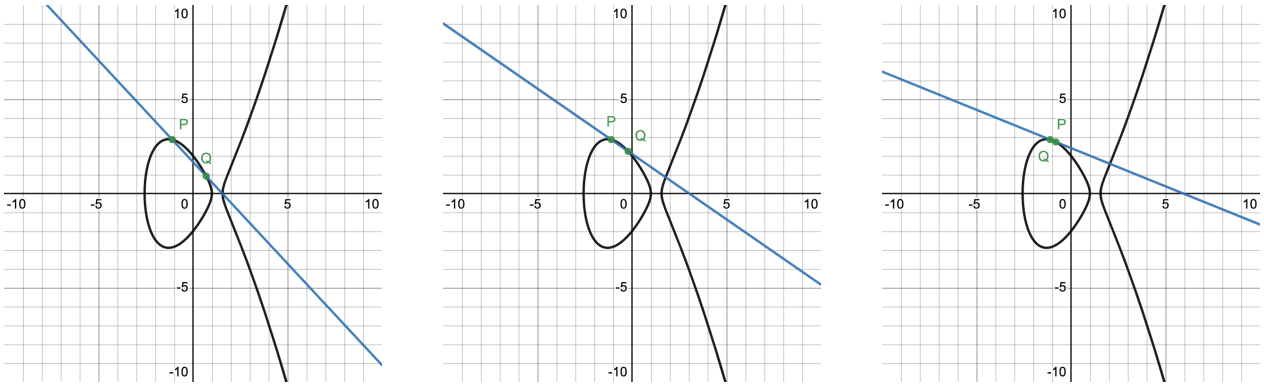

Figure 4

To find the slope of the tangent line to an elliptic curve E at a point P, we only need to recall how to perform implicit differentiation with respect to x.

$$y^2 = x^3 + ax + b$$

$$\frac{d}{dx}y^2 = \frac{d}{dx}(x^3 + ax + b)$$

$$2y\frac{dy}{dx} = 3x^2 + a$$

$$\frac{dy}{dx} = \frac{3x^2 + a}{2y} = m$$

Therefore, if $P = (\alpha, \beta)$ on an elliptic curve which satisfies $y^2 = x^3 + ax + b$, then $P \star P = (\sigma, \tau)$, where

$$\sigma = m^2 - 2\alpha \qquad \tau = m\sigma + h$$

$$m = \frac{3\alpha^2 + a}{2\beta} \qquad h = \beta - m\alpha.$$

To address what $P \star Q$ becomes when P and Q are reflections of each other over the x-axis (as shown figure 5), we must rethink the binary operation $\star$ as we presently know it, and introduce the notion of a group.

2 Groups and Abstraction in Math

It turns out that the points on the elliptic curves we have been investigating have a rather significant underlying structure. That structure is what is known as a group in mathematics. Vaguely, you can think of a group as an abstraction of symmetry and regularity itself. Groups emerged from the work of Évariste Galois in the mid-19th century when mathematicians were attempting to solve polynomials of degree five and greater using explicit formulas, and it wasn't until several decades after that time until the definition of a group was settled upon—one that was neither too strict or too general so that it may be useful. Despite the fact that the modern notion of a group is a human invention that has only existed for a few centuries, groups have proven to be a powerful tool in chemistry, physics, cryptography, and many other fields. Yet this should not be too surprising; a lot of mathematics concisely captures very general patterns that both predictably and unpredictably give rise to powerful models that interpret natural phenomena.

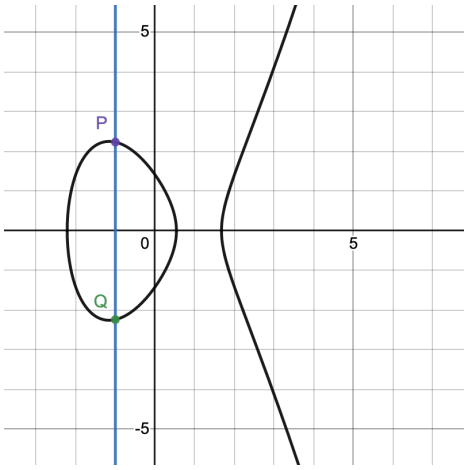

Figure 5: What is $P \star Q$ when P is directly above Q? The line connecting them does not intersect the elliptic curve E at another point—at least not until we view E differently.

For example, consider the Arabic numeral 5. What does it represent? Not just five oranges, five geese, or five pairs of scissors, but any five distinct things. With the abstraction of numbers at our disposal, the fact that "if one drops three apples into a bucket already containing seven apples, then the bucket now contains ten apples'"can be simply expressed as "3 + 7 = 10".

The relevant abstraction that groups capture is the fact that certain collections of objects behave nicely with one another. More concretely, a group is a set of elements and a binary operation such that (1) there exists a "do nothing" element[4], (2) all elements have an inverse element (so that the output of the binary operation between an element and its inverse is the identity), and (3) the binary operation is associative. Associativity means that no matter how we arrange the parenthesis in an expression, we get the same result.

One of the simplest groups is that of the integers, denoted $\mathbb{Z}$, under the binary operation of addition; there exists a "do nothing" integer which doesn't change any other integer when added to it (zero), for every integer there is an opposite integer (its negative), and for any $a, b, c \in \mathbb{Z}$,

$$a + (b + c) = (a + b) + c.$$

The group of integers under addition, sometimes written as $(\mathbb{Z}, +)$, is an example where the binary operation is commutative, that is, for any two integers $a, b \in \mathbb{Z}$, $a + b = b + a$. When this is the case, we say that the group is abelian (the name coming from the 19th century mathematician Niels Henrik Abel).

3 Group Structure of Elliptic Curves

So then, how are elliptic curves groups? For reasons we'll continue to elaborate upon, the members of the elliptic curve are the elements of the group, and the binary operation is nearly the same as the $\star$ we have already discussed. The alteration we'll make to $\star$ is that we reflect its output across the x-axis (depicted in figure 6). This alteration is required to ensure that $\star$ is an associative binary operation and that the identity of the group behaves as desired, though, this in and of itself does not make it clear why $\star$ is associative. It is actually anything but obvious, geometrically speaking, that $\star$ is associative. It would not be conceptually difficult (just computationally tedious) to verify that $\star$ is associative using the explicit formulas we derived earlier. Granted, the formulas would be changed to reflect the alteration just made to $\star$; namely, the y component of $P \star Q$—what was denoted τ in section 1.2—would be multiplied by -1.

[4]This is often referred to the identity or neutral element. The number 0 acts as an identity element under addition for virtually every set of numbers. Actually, it might be more accurate to say that in any setting where the relevant operation is addition, 0 will denote the identity or neutral element *by definition*—0 is commonly referred to as the *additive identity*, as opposed to, say, the *multaplicative identity* that is usually denoted by the number 1.

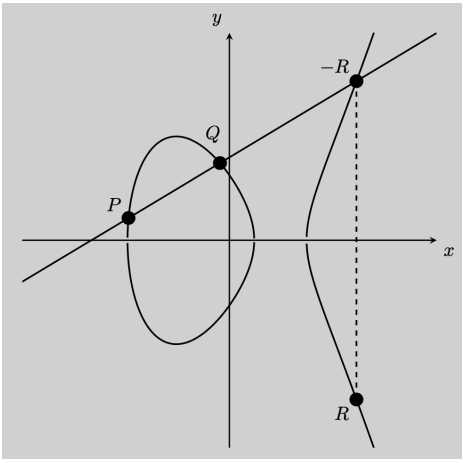

Figure 6: The actual definition of $\star$: $P \star Q = R$, not what is now denoted $-R$.

But what, then, would the identity element of the elliptic curve be? Is there a point O on an elliptic curve E such that for any other point P on E, $O \star P = P$? Try as you might to discover such a point, it does not exist while considering E with traditional euclidean geometry. Instead, we need to consider E with what is known as projective geometry. In doing so, such a point O becomes part of E: it is the point infinitely far off in the y direction. Under the axioms of projective geometry, all vertical lines intersect O—in the same way that if you were to stare at two railroad tracks which you know to be parallel, they would nonetheless seem to intersect at the horizon (applying this projective lens to the xy-plane produces some interesting results; see figure 7 to view what the parabola $y = x^2$ becomes). O acts as the identity element because for

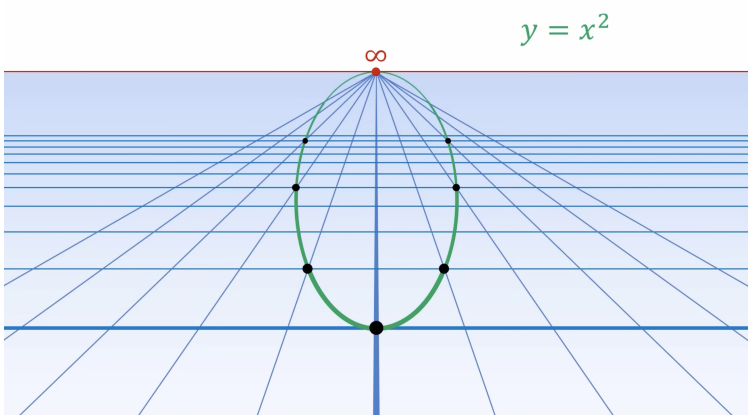

Figure 7: Looking at the xy-plane with a new perspective. All vertical lines intersect at an infinitely far point. This means a parabola like $y = x^2$ is actually an ellipse viewed through projective geometry. Image curtesy of Bill Shillito [2].

any point P above the x-axis, the line between P and O would be vertical, and so, the other point which the curve intersects is the one directly below P, which becomes P after we reflect over the x-axis—this is why we altered $\star$ in the manner we did; it ensures that O "does nothing"; i.e, that $O \star P = P$, as depicted in figure 8.

Identifying O as the identity also lets us see that the inverse of each element on the elliptic curve is itself reflected across the x-axis. The inverse of a point P is denoted $-P$, as also depicted in figure 8.

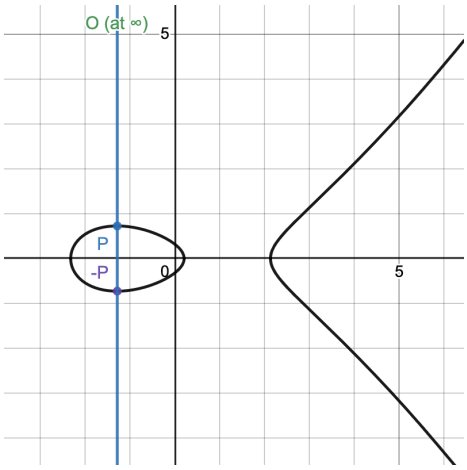

Figure 8: Graph demonstrating $P \star O = P$ and that $P \star -P = O$.

4 Rational Points on Elliptic Curves

The last topic we will cover in some capacity is how one would find rational points on an elliptic curve. Based on the explicit formulas for $P \star Q$ we derived in sections 1.1 and 1.2, so long as we find a single point on $E(\mathbb{Q})$, we can find $2P$, $3P$, and so on, each of which will also be members of $E(\mathbb{Q})$ since computing $P \star Q$ only involves adding, subtracting, multiplying, and dividing—which recall, are precisely the operations defined on $\mathbb{Q}$. So at least for curves where finding a first rational point is simple, we can seemingly find many other such rational points.

Let's consider again the polynomial $y^2 = x^3 - x + 1$, but now over $\mathbb{Q}$; let

$$E_1(\mathbb{Q}) = \{(x, y) \in \mathbb{Q}^2 \mid y^2 = x^3 - x + 1\}.$$

We can find our first rational point without too much difficulty; $1^3 - 1 + 1 = 1$, therefore $P = (\alpha, \beta) = (1, 1) \in E_1(\mathbb{Q})$. We'll calculate $2P$ using the formulas we derived.

$$\begin{aligned} m &= \frac{3 \cdot \alpha^2 + a}{2\beta} = \frac{3-1}{2} = 1 \\ h &= \beta - m\alpha = 1 - 1 = 0 \\ \sigma &= m^2 - 2\alpha = -1 \\ \tau &= -(m\sigma + h) = -(-1 + 0) = 1 \\ \implies 2P &= (-1, 1). \end{aligned}$$

We can also compute $3P = 2P \star P = (-1, 1) \star (1, 1)$.

$$\begin{aligned} m &= \frac{1-1}{1--1} = 0 \\ h &= 1 - 0 = 1 \\ \sigma &= m^2 - \alpha_1 - \alpha_2 = 0 \\ \tau &= -(m\sigma + h) = -1 \\ \implies 3P &= (0, -1) \end{aligned}$$

Performing additional calculations, we find that $4P = (3, -5)$, and $5P = (5, 11)$. For $n \in \{1, 2, 3, 4, 5\}$,

$nP \in \mathbb{Z}^2$, but this trend ends when $6 = n$: $6P = (\frac{1}{4}, \frac{7}{8}) \notin \mathbb{Z}^2$. Figure 9 plots P through $4P$, $E_1(\mathbb{R})$, and various lines to help trace what finding multiples of P looks like geometrically.

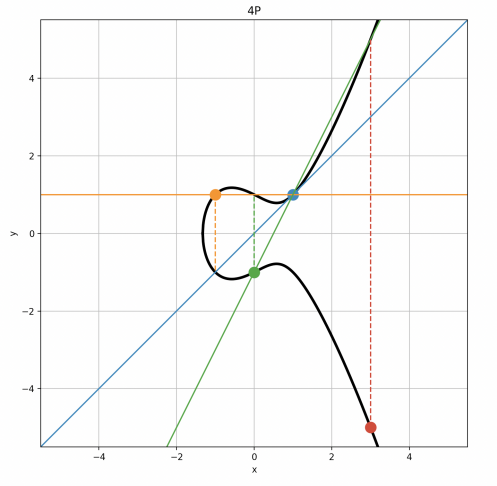

Figure 9: The black curve is $E_1(\mathbb{R})$, $P = (1, 1)$ is the blue dot, $2P = (-1, 1)$ is the yellow dot, $3P = (0, -1)$ is the green dot, and $4P = (3, -5)$ is the red dot. The color of the line used to find $P \star kP$ for $k \in \{1, 2, 3\}$ is that of the point kP. Where each line intersects $E_1(\mathbb{R})$ at the point other than P and kP, a dotted line connects that intersection with its reflection over the x-axis—giving the point $(k+1)P$, which shares its color with the dotted line.

We can continue to find multiples of P for larger $n \in \mathbb{Z}_{\geqslant 1}$ (See figure 10). For n even moderately large, $E_1(\mathbb{R})$ becomes covered by the multiples of P within the displayed region. Yet do not be under any illusion that even an infinite number of multiples of P would give all of $E_1(\mathbb{R})$. The real numbers are infinitely more dense than the rationals, and while it might look like the multiples of P saturate $E_1(\mathbb{R})$, there still remain an uncountably infinite number of points on $E_1(\mathbb{R})$ which the multiples of P do not give. We have not shown that these multiples of P even make up all of $E_1(\mathbb{Q})$, but it turns out that the group of finite points on an elliptic curve is finitely generated—that is, there are a finite number of points in $E_1(\mathbb{Q})$ which generate all of $E_1(\mathbb{Q})$.

References

[1] Henri Poincaré. *Science et Méthode*. Paris: E. Flammarion, 1908, p. 22.

[2] Bill Shillito. *Putting algebraic curves in perspective*. Dec. 2019. URL: https://www.youtube.com/watch?v=XXzhqStLG-4.

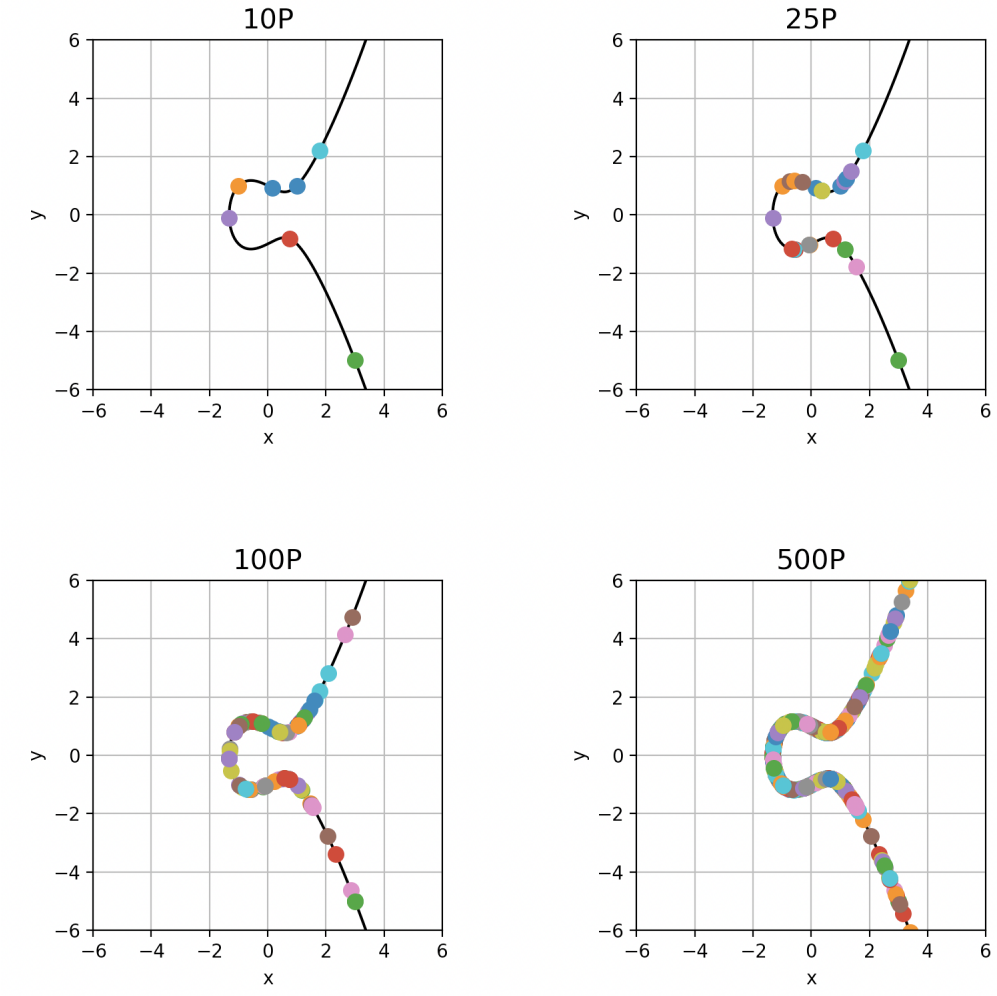

Figure 10: Plotting an increasing number of the multiples of P. The title of each subplot indicates the largest multiple of P calculated. N.b. that not all multiples may be within the region displayed.