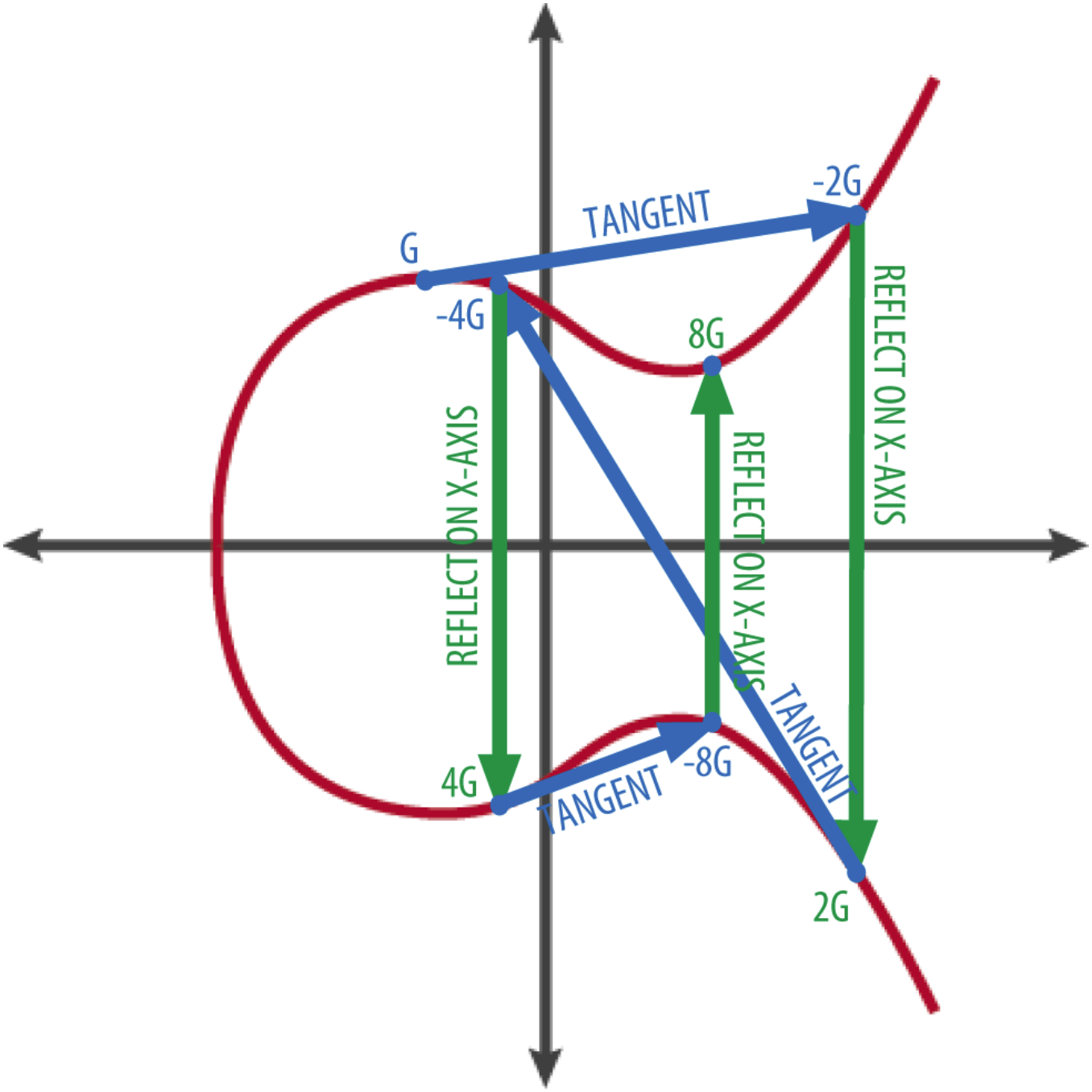
How does ECC go from decimals to integers?

I realise that elliptical curves are tricky, but there's one aspect that no one seems to explain. I've looked, and it's towards the beginning. This is the traverse over a red curve:-



This is only three steps (n=3). For cryptography n must be large. I don't know how large but I cannot believe that all tangential and intersection points occur at coordinates with integer values. They must be decimal. After all the curve is smooth and continuous. The tangents can be all sorts of infinitely precise gradients. We know that decimal encryption can't work due to representational and rounding errors. Assume that point G is at (-2,4). There is no way that point 8G is at (3,3) exactly. And if n is very large?

I had a look at the wiki article regarding elliptic curve point multiplication, but using an example of 42.57 degrees didn't clear it up any. I think that there's a crucial step I'm missing before the explanation of Integers in ECC. So,

Q: How do traverses over elliptical curves go from decimal coordinates to integer values that can then be used for practical cryptography?

The fact that I've identified this curve as type "red" indicates my lack of ECC knowledge or general mathematics ability. Is it possible to explain

elliptic-curves

asked Jun 26 '17 at 23:24

Paul Uszak
6,785 1 14 34

In crypto, you're doing group operations on a finite field, so calculation of things like slopes $\lambda = (y_2 - y_1)/(x_2 - x_1) \mod p$ is going to be integral, even if in $\mathbb{R}^n$ it wouldn't be. – user47922 Jun 26 '17 at 23:35

1 also note that $\ldots/(x_2 - x_1) \mod p$ is not doing a division but multiplying by the inverse of $(x_2 - x_1)$ in Z_p – Biv Jun 26 '17 at 23:36

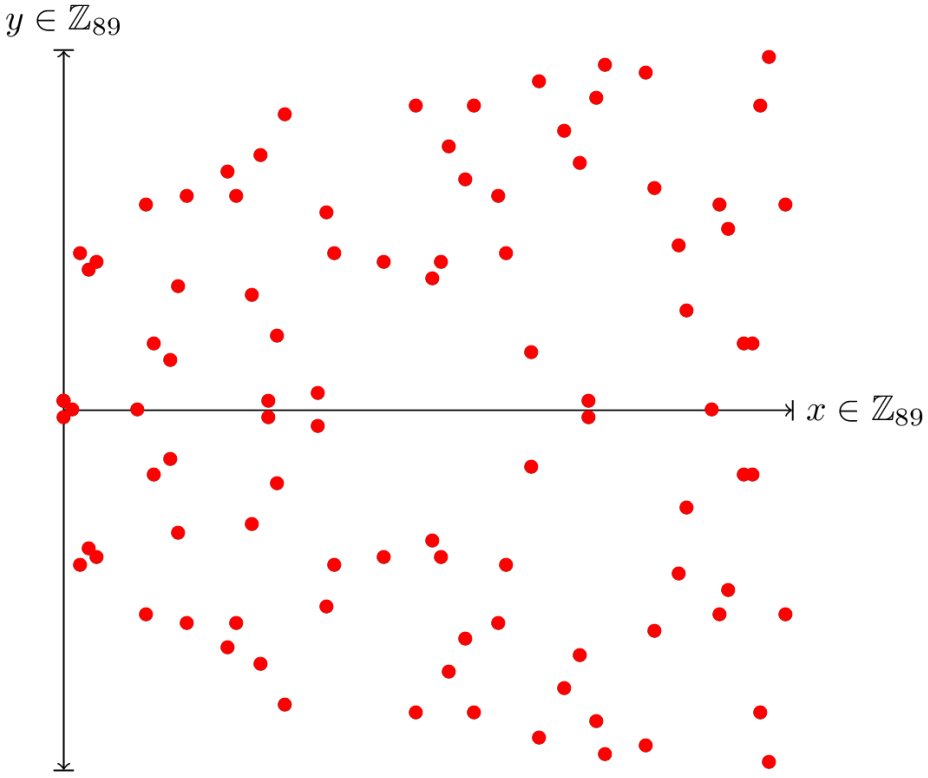
4 There is not a single decimal in an elliptic curve computation over a finite field. You could even say that there is not a single integer. Try to wrap your head around this: all operations are in the finite field, not with integers. They may seem the same (at first sight confusingly so), because the finite field can be represented as the integers modulo a prime. What seems like *division*, is actually *multiplying by an inverse*, which has little to do with decimals. – CurveEnthusiast Jun 27 '17 at 2:11

1 Why are you guys so afraid of the word "division"? It is perfectly proper to use it over any field. – fkraiem Jun 27 '17 at 5:15

4 @fkraiem `Why are you guys so afraid of the word "division"?` Because for a neophyte, this is likely to be taken as a division as in $\mathbb{R}$. – Biv Jun 27 '17 at 13:34

3 Answers

I think what's going on is that a lot of introductions to elliptic curves start off with the geometric description of the group laws so as to be as visual as possible (it certainly is not intuitive to me at all). And then they switch to elliptic curves over finite fields, which look *very* different than the continuous case over $\mathbb{R}$ *even though the group laws are the same*:



$y^2 = x^3 - 2x + 1$ over $\mathbb{Z}_{89}$

[source]

I know; that doesn't look like a curve at all. So the idea of a "curve" is really weird over a finite field. But doing all of your calculations $\mod p$ (*i.e.,* adding points, calculating "slopes") will always be integral.

Let's go through an example using the curve $E : y^2 = x^3 - 2x + 1$ over $\mathbb{F}_{89}$. Picking $x = 4$ gives us $y^2 = 4^3 - 2 \cdot 4 + 1 = 57 \mod 89$. Now, in general, to take the square root modulo a prime, you have to use the Tonelli-Shanks algorithm. Understanding this involves some number theory, such as quadratic residues. Here's a Python implementation. So like "regular" square roots, you can get two answers, and it turns out that $\sqrt{57} \equiv \{71, 18\} \mod 89$. So $(4, 18)$ may

Let's demonstrate addition. Let $P = (P_x, P_y) = (4, 18)$ be the point we just found. I'll pick $Q = (Q_x, Q_y) = (15, 63)$. (I iterated through all x from 1 to 88 and computed y as above with the modular square root, and just picked one.)

The slope is $\lambda = (Q_y - P_y) \cdot (Q_x - P_x)^{-1} = 45 \cdot 11^{-1} = 45 \cdot 81 \bmod 89 \equiv 85 \bmod 89$

(81 is the modular inverse of $11, \mod 89$.)

Now calculate $\nu = P_y - \lambda P_x = 18 - 85 \cdot 4 \bmod 89 \equiv -322 \bmod 89 \equiv 34 \bmod 89$, and finally,

$$x_3 = \lambda^2 - Q_x - P_x = 85^2 - 15 - 4 \bmod 89 \equiv 86 \bmod 89$$

$$y_3 = -(\lambda x_3 + \nu) = -(85 \cdot 86 + 34) \bmod 89 \equiv 43 \bmod 89$$

So we have $P + Q = (86, 43)$ on $E(\mathbb{F}_{89})$. You can check that this point is on the curve since $86^3 - 2 \cdot 86 + 1 \equiv 43^2 \bmod 89$.

You may like exploring these ideas with Sage as well.

edited Jun 27 '17 at 1:58 — answered Jun 26 '17 at 23:44
user47922

2 This picture is more interesting because it is easier to see the $-(x, y) = (x, -y)$ operation. :) – Biv Jun 26 '17 at 23:48

I approve, it's like a crypto Rohrschach test. – user47922 Jun 26 '17 at 23:52

1 @galvatron Perfect! I think that this is the only link between the visual graph of the curve and it's role in actual cryptology I've ever seen. Suggestion: why don't you rewrite the piece of crap Wikipedia article to this? An example makes it simple for everyone. – Paul Uszak Jun 27 '17 at 12:05

1 It IS a Rohrschach pattern. I can see myself as a baby being abused by cuddly toys... – Paul Uszak Jun 27 '17 at 12:07

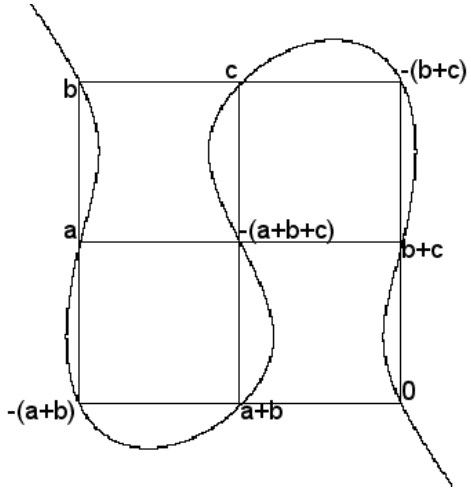
1 In general yes you need to use Tonelli-Shanks to take the square root. It is worth noting that the NIST curves over prime fields have optimized square root algorithms. – puzzlepalace Jun 27 '17 at 18:04

ECC goes from "decimal" (actually, real) to integer by reusing the sames equations (for the curve, point addition or doubling), applied to some finite field instead of the infinite field $(\mathbb{R}, +, *)$. The "curve" becomes a set of finitely many points (as nicely illustrated in that other answer). The intuition that we could keep the geometrical construction does not immediately materialize, especially for point doubling.

Update: but not all hope is lost. The last picture in this tutorial shows that, when the finite group is $(\mathbb{Z}_p, +, *)$, we can make a geometric construction of point addition $A + B$ when $A \neq B$. And for doubling point A we can take another point B and say that $2 \times A = 2 \times A + (B - B) = (A + B) + (A - B)$ and compute the later geometrically with three point additions.

Note: Just restricting from $\mathbb{R}$ (reals) to the subset $\mathbb{Z}$ (integers) leads nowhere, because so few points on the curve fall onto integer coordinates.

The geometric construction in the question applies on the Cartesian plane. It lets one establish the formulas for the point addition operation, and its special case doubling, as equations between Cartesian coordinates in the field $(\mathbb{R}, +, *)$. That point addition is an operation on the points of curve; it is internal (result is on the curve), and commutative. We can include an additional point as identity element, and define that elements have an additive inverse by symmetry relative to the horizontal axis; all that is geometrically evident and consistent. Then algebra based on the formulas shows the surprising fact that the point addition operation is associative (reportedly, there's also a non-trivial geometrical proof; what follows is an illustration only). We have a group!



Credit: Thomas Cooper, who created a GNU Octave script to generate this file; Public Domain.

Elliptic curve cryptography replaces $(\mathbb{R}, +, *)$ with a finite field, perhaps $(\mathbb{Z}_p, +, *)$. This is not performed by considering points of the curve that fall on integer points, because there are very few; for example $y^2 = x^3 - 2x + 1$ has only 3 integer solutions: $(x, y) \in \{(0, 1), (0, -1), (1, 0)\}$.

Rather, replacing $(\mathbb{R}, +, *)$ with a finite field is made by using the curve's equation in that finite field, as well as the formulas established for point addition and doubling; that yields an operation that is commutative, associative, with identity element, and an opposite for any element; again, the axioms for a group.

The resulting "curve" looks like an haphazard set of points (with symmetry along an axis inherited from that of the original curve), as nicely illustrated in that other answer. The geometric (or at least, visual) aspect of the construction vanishes, except for opposites. In particular, we can no longer visually define a tangent, ~~or the intersection of the line joining two points of the "curve" with a third point of the "curve"~~. And what used to be a "slope" now involves a multiplicative inverse in the finite field, which computation is performed radically differently (e.g. by the Extended Euclidean algorithm or by exponentiation) from computing the multiplicative inverse in $\mathbb{R}$.

edited Jun 30 '17 at 11:51

answered Jun 27 '17 at 5:47

fgrieu
76.8k 7 158 322

1 "the geometric aspect of the construction vanishes" only if you define geometry in a very narrow way. – fkraiem Jun 27 '17 at 5:57

in uneasy -> is uneasy. – Maarten Bodewes Jun 27 '17 at 7:47

I'm sorry, but your answer hasn't added anything at all to how a continuous curve stops being so. This is exactly the type of explanation you see everywhere, and what lead me to the question. It's not you, it's me, but I'm hoping for an explanation of the stage prior to this step whilst the curve is still smooth and curvy... – Paul Uszak Jun 27 '17 at 12:14

@Paul Usak: I now give two arguments: restricting to integer points give too few points. And computing the slope involves a multiplicative inverse in the finite field. – fgrieu Jun 28 '17 at 9:47

1 If this answer is still live, might I be so bold as to suggest a focus on getting from an algebraic curve Y^2= etc to an integer coordinate such as (86,43)? I think that might help me the most as it's this transition I'm flummoxed with. – Paul Uszak Jun 28 '17 at 19:00

One way to get a better grasp on the transition is to look at the explicit formulas that are derived from the image that you posted:

For $R = P \oplus Q$ on $E : y^2 = x^3 + ax + b$ we get the following formulas for point addition and doubling:

Point addition:

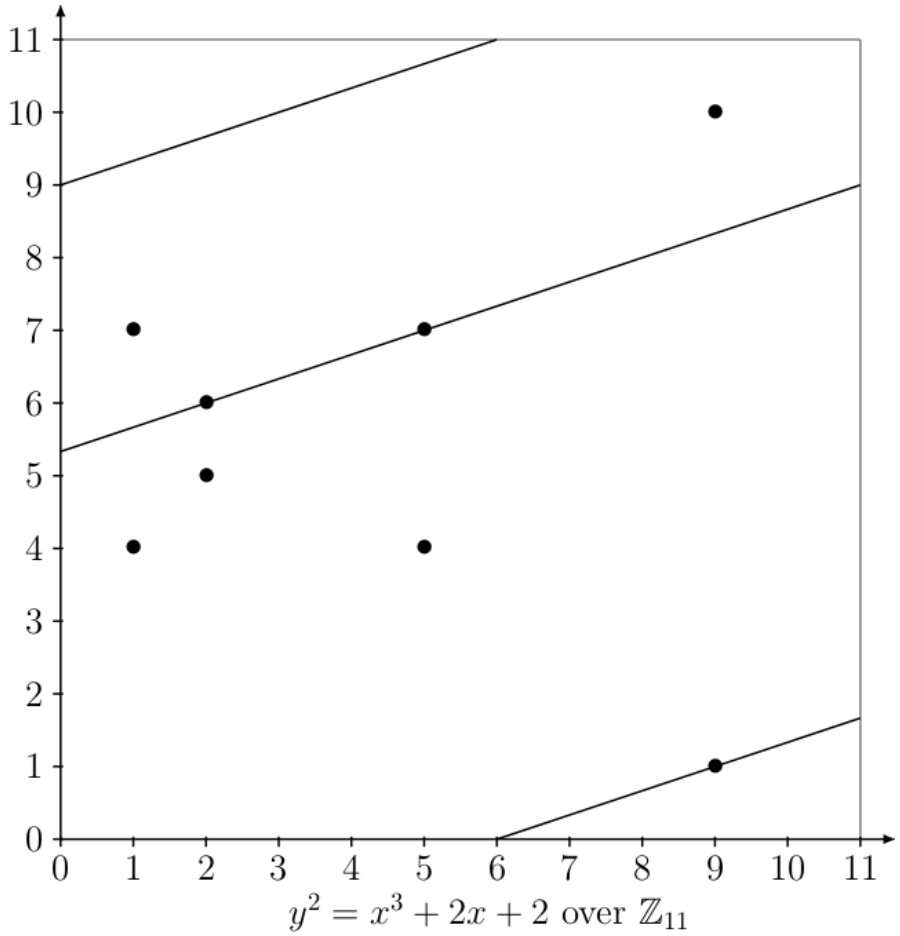
- $\lambda = \frac{y_Q - y_P}{x_Q - x_P}$
- $\nu = y_P - \lambda x_P$.
- $x_R = \lambda^2 - x_P - x_Q$

Point doubling:

- $\lambda = \frac{3x_P^2+a}{2y_P}$
- $\nu = y_P - \lambda x_P.$
- $x_R = \lambda^2 - 2x_P$
- $y_R = -\lambda x_R - \nu$

And it is immediately obvious that all the operations that are performed are available in a field. Hence, if the curve parameter a is in a finite field and taking two points with coordinates in that field the result must also be in it.

That explains why using our formula over a finite field still works. And, when taking the x-axis symmetry into account, it also explains why a line through two points will still end up at a third point over a finite field.



Why elliptic curve points form a group beyond the geometrical intuition is usually proven using advanced techniques from algebraic geometry. Then it is easy to show that this also holds over a finite field.

If you want to go down that rabbit hole I think about the first 50 pages of Hartshorne's book on algebraic geometry might suffice but they are far from an easy read and require knowledge of algebra.

Another interesting take on the problem is this elementary proof of the group law using only the explicit formulas. Unfortunately it is hard to read and imho not very enlightening. But it is a proof that the group law holds.

answered Jul 2 '17 at 0:13

Elias
3,778 1 6 24

Another good book in the "algebraic geometry makes it clear" camp is Silverman's *Arithmetic of Elliptic Curves*, which dedicates the first two chapters to the subject as a tutorial. The same prerequisites apply. – user47922 Jul 2 '17 at 0:29