REAL PREIMAGES OF DUPLICATION ON ELLIPTIC CURVES

JOHN CULLINAN

ABSTRACT. Let E be an elliptic curve defined over the real numbers $\mathbf{R}$ and let $P \in E(\mathbf{R})$. In this note we give an elementary proof of necessary and sufficient conditions for the preimages of P under duplication to be real-valued.

1. INTRODUCTION

Let K be a field and E/K an elliptic curve. For the remainder of the paper we fix an algebraic closure $\overline{K}$ of K. It is well known that the points $E(K)$ of E defined over K form an abelian group and the group structure of elliptic curves is a major area of current research with different applications depending on the isomorphism type of K (*e.g.* finite field, number field, p-adic field). Let $m \geq 1$ be a positive integer. Then E admits an endomorphism defined over K, denoted by $[m]$, that maps a point $P \in E$ to $P+\cdots+P$ (m times) $\in E$. If, in addition, m is coprime to the characteristic of K, then a point $P \in E$ has m^2 preimages under $[m]$, with coordinates in $\overline{K}$. Since the group operation on E is defined over K (*i.e.* given by coordinate functions with coefficients in K), it follows that if $P \in E(K)$, then $[m]P \in E(K)$ as well. However, the *preimages* of $P \in E(K)$ might well be defined over non-trivial extensions of K and, in fact, the Galois theory of the fields defined by the preimages of the identity element of E (the *torsion point fields* of E) is one of the most active areas of research in modern number theory. The field of definition of the preimages is precisely what we focus on here.

In this paper we set $m = 2$, take K to be the field of real numbers $\mathbf{R}$, and fix an elliptic curve $E/\mathbf{R}$. Our main theorem is a necessary and sufficient condition for the preimages $[2]^{-1}P$ of $P \in E(\mathbf{R})$ to be real-valued as well. The reason we focus on this special case is twofold. In the proof of the Weak Mordell-Weil Theorem, in order to perform a 2-descent over a number field K, one needs to know the precise conditions under which a point $P \in E(K)$ has all of its preimages under [2] defined over K. For that reason, our result

2010 *Mathematics Subject Classification.* 11G05.

is not a new one, though standard proofs involve Galois cohomology or other complicated machinery. While the Galois cohomology of an elliptic curve is essential to understanding its arithmetic, in this particular instance one can arrive at necessary and sufficient conditions for rationality using purely elementary methods; indeed in [4] a different elementary proof from ours was recently given.

Our other reason for presenting this special case is that our interpretation exploits the visual aspect of the real points on an elliptic curve. Given an elliptic curve $E/\mathbf{R}$, the real points $E(\mathbf{R})$ either form one connected component: $E(\mathbf{R}) = E_0(\mathbf{R})$ or two: $E(\mathbf{R}) = E_0(\mathbf{R}) \cup E_1(\mathbf{R})$. In both cases, the component $E_0(\mathbf{R})$ is known as the the *identity* component (see below for an explanation of the terminology) and is non-compact, while $E_1(\mathbf{R})$ is compact. Our main theorem can be presented visually in terms of these components rather than in terms of algebraic conditions on the coefficients of the equation defining the curve. This how the result is typically stated, such as in [2, Theorem V.1.1] and [4].

We do not assume the reader has a deep familiarity with elliptic curves. In fact, the elementary nature of the proofs is possible precisely because the statements about elliptic curves can be easily translated into statements about the roots of real quartic polynomials. However, we will briefly recall some basic facts and notation in order to state the main result of the paper. In Section 2 we give a quick background on elliptic curves that is needed to translate the statement of the problem into polynomial algebra. We refer the reader to [3] for a thorough treatment of the subject.

Let $\mathbf{P}^2_K$ denote the projective plane over K. In homogeneous coordinates $[x, y, z]$, we identify the affine plane K^2 with the coordinates $[x, y, 1]$ and the points at infinity with $[x, y, 0]$. An elliptic curve E/K is a smooth projective cubic curve with coordinates in $\mathbf{P}^2_K$. One can show [3, Remark III.1.3] that if the characteristic of K does not equal 2 or 3, then we may assume the affine locus of E is given by the equation

$$y^2 = x^3 + ax + b, \tag{1.1}$$

with $a, b \in K$. Together with a single projective point $[0, 1, 0]$ at infinity serving as the identity element $\mathcal{O}$, the solutions to (1.1) with coordinates in $\overline{K}$ form an abelian group. The symbol E refers to the set of all $\overline{K}$-solutions to (1.1) (together with $[0, 1, 0]$), while $E(L)$ refers to the subgroup of E consisting only of solutions with coordinates in the subfield L of $\overline{K}$.

When K is the field $\mathbf{R}$ of real numbers, then the real points $E(\mathbf{R})$ of E come in one or two connected components, depending on whether the discriminant of the polynomial $x^3 + ax + b$ is negative or positive, respectively. The component of $E(\mathbf{R})$ containing the identity $\mathcal{O}$ is called the identity

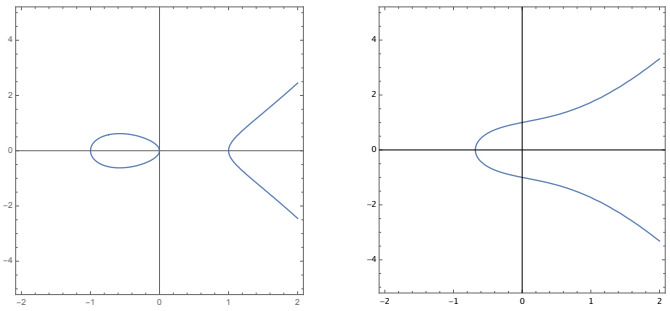

FIGURE 1. Real points of $y^2 = x^3 - x$ and $y^2 = x^3 + x + 1$, respectively

component and denoted $E_0(\mathbf{R})$, while the other (compact) component (for elliptic curves with positive discriminant) is denoted $E_1(\mathbf{R})$.

Let $P \in E(\mathbf{R})$ be non-trivial, *i.e.* $P \neq \mathcal{O}$. When $E(\mathbf{R}) = E_0(\mathbf{R})$ has one component, then *a priori* P has either 0 or 2 preimages under duplication, while if $E(\mathbf{R})$ has two components, then P has 0 or 4 preimages. In terms of graphs, the real points of an elliptic curve can be visualized as one of the two cases in Figure 1. Our main theorem can then be expressed in terms of these components.

Theorem 1.1. *Let $E/\mathbf{R}$ be an elliptic curve with Weierstrass model $y^2 = x^3 + ax + b$ and let $P \in E(\mathbf{R})$ be non-trivial. If $E(\mathbf{R})$ has one connected component, then P has two real preimages under duplication. If $E(\mathbf{R})$ has two components then P has four real preimages if $P \in E_0(\mathbf{R})$ and zero real preimages otherwise.*

2. Background on Elliptic Curves

This section is not meant to be exhaustive, but rather to collect some salient facts about elliptic curves for completeness. In fact, the reader with expertise in this area can easily skip this section, while a non-specialist may find the exposition too terse. The point of this brief section is to show that in order to determine whether the preimages under the duplication map are real or complex, it suffices to determine whether a certain quartic polynomial has real or complex roots.

Setting some additional notation, let $P/m \subset E$ be the set of preimages in $\overline{K}$ of P by $[m]$. We remind the reader of the non-trivial fact [3, Theorem III.6.1] that the set P/m consists of m^2 points of E when m and $\mathrm{char}(K)$ are coprime. The preimages of $\mathcal{O}$ by $[m]$ are called the *m-torsion points* of E and denoted by $E[m]$. In this paper, we are concerned with determining

the field of definition of the set $P/2$, given that P has coordinates in $\mathbf{R}$. Thus, $P/2$ always consists of 4 points defined over $\mathbf{C}$ and we will determine conditions that ensure when they are in fact defined over $\mathbf{R}$.

For the remainder of the paper we focus on the special case $m = 2$ and work in coordinates. Because $\mathbf{R}$ has characteristic zero, we assume E is given by the Weierstrass equation

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

with $a, b \in \mathbf{R}$. We set $f(x) = x^3 + ax + b$ and remind the reader that the smoothness of E implies $f(x)$ has no repeated roots, *i.e.* that the discriminant of f is non-zero:

$$\mathrm{disc}(f) = -4a^3 - 27b^2 \neq 0.$$

Let $P \in E(\mathbf{R})$ be non-trivial and set $P = (t, u)$. Let $Q \in P/2$ have coordinates $Q = (x, y)$. The duplication map $[2] : E \to E$ is then given by explicit rational functions [3, III.2.3]:

$$t = \frac{x^4 - 2ax^2 - 8bx + a^2}{4x^3 + 4ax + 4b} \tag{2.1}$$

$$u = \frac{x^6 + 5ax^4 + 20bx^3 - 5a^2x^2 - 4abx + (-a^3 - 8b^2)}{8y(x^3 + ax + b)}. \tag{2.2}$$

That is, given $a, b, t, u \in \mathbf{R}$, the four elements of $P/2$ have coordinates (x, y) given by the solutions to (2.1) and (2.2) above. Clearing denominators in (2.1), it follows that the x-coordinates of the members of $P/2$ are given by the roots of the quartic polynomial

$$x^4 - 4tx^3 - 2ax^2 + (-4ta - 8b)x + (a^2 - 4tb). \tag{2.3}$$

In the proof of Theorem 1.1, we appeal to the theorem of [1, p. 45] on the classification of the roots of a real quartic. In order to apply those results, we need to first put the polynomial (2.3) into reduced form. Hence, we translate x in (2.3) by t and set the following notation:

$$F(x, t) \stackrel{\mathrm{def}}{=} x^4 - (2a + 6t^2)x^2 - 8u^2x + a^2 - 3t^2a - 9bt - 3u^2t, \tag{2.4}$$

where (t, u) are the coordinates of P and $u^2 = t^3 + at + b$.

The non-trivial 2-torsion points of E are given in coordinates by $(t, 0)$ [3, III.2.3(d)], and since a real cubic polynomial has either one real root or three real roots, we will distinguish between these two cases in Section 3 below. This is manifest algebraically through the discriminant $\mathrm{disc}(f)$ of $f(x)$. Visually, an elliptic curve $E/\mathbf{R}$ with negative discriminant has one connected component and with positive discriminant two components. With all of this notation now in place, we are ready to prove the main theorem of the paper.

3. The Main Result

We briefly recall what is already known. In the standard proof of the Weak Mordell-Weil Theorem using a 2-descent over a number field K, one assumes that the elliptic curve has full 2-torsion defined over K. In the course of the proof, one requires a criterion for checking the K-rationality of the preimages of a non-trivial K-rational point P. Writing E in the form

$$y^2 = (x - e_1)(x - e_2)(x - e_3),$$

one shows that $(t, u) \in E(K)$ has four rational preimages if and only if each $t - e_i$ is a square in K [2, Theorem V.1.1]. In [4], the author gives a beautiful elementary proof of this fact for arbitrary fields using only basic properties of the arithmetic of elliptic curves in contrast to the standard proofs which use more technical machinery. Below, we offer an alternative, elementary proof of the rationality of preimages under duplication that yields a purely visual result over the real numbers. Our intent is not to recreate the proof in [4] but rather to highlight the implication over the real numbers. We also do not assume E has full 2-torsion defined over the real numbers, but rather treat the cases where $f(x)$ has one real root or three real roots separately.

Proof of Theorem 1.1. In [1, p. 45, Theorem], the author determines the nature of the roots of a quartic in terms of its discriminant and coefficients; all references below to [1] are to the Theorem on p. 45. To apply those results, we compute the discriminant Δ of the quartic $F(x, t)$:

$$\Delta = \text{disc } F(x,t) = 2^{12}(t^3 + ta + b)^2 \text{ disc}(f). \tag{3.1}$$

Note that the sign of Δ is determined by the sign of $\text{disc}(f)$, and that $\Delta = 0$ if and only if $t^3 + at + b = 0$ (so P is a 2-torsion point). This leads us to consider three cases separately; in the first two P does not have order 2. Therefore, for cases 1 and 2 we may apply the result of [1, p. 45] since the coefficient of x is non-zero. A different argument will be used in Case 3.

Case 1: $\Delta < 0$. In this case $E(\mathbf{R})$ has one real (identity) component $E_0(\mathbf{R})$ and P has either 2 or 0 preimages under duplication. By the classification of quartics in [1], if $\Delta < 0$, then the quartic $F(x, t)$ has two (distinct) real roots and two imaginary roots; the two real roots serve as the x-coordinates of the preimages. And since P does not have order 2, it follows that $u \neq 0$ and so Equation (2.2) gives y as a rational expression in terms of a, b, x, and u. Therefore, if $E(\mathbf{R})$ has one real component and $P \in E(\mathbf{R})$, then P has two real preimages.

Case 2: $\Delta > 0$. Here $E(\mathbf{R})$ has two real components, the identity $E_0(\mathbf{R})$ and the non-identity $E_1(\mathbf{R})$. We again appeal to the result of [1] quoted

above. There are two possibilities for the roots of the real quartic $x^4 + qx^2 + rx + s$ with positive discriminant Δ: either all are real, or all are imaginary. The case of real roots occurs if and only if $q < 0$ and $s < q^2/4$. In the context of the polynomial $F(x,t)$, the quantities $s - q^2/4$ and q are given by the following:

$$s - q^2/4 = -12t(t^3 + at + b) \qquad \text{and} \qquad q = -2a - 6t^2.$$

Note that the sign of $s - q^2/4$ is determined by that of t since $-12(t^3+at+b) = -12u^2 < 0$. Thus, $s - q^2/4 < 0$ if and only if $t > 0$.

To determine when $q < 0$, note that the hypothesis $\Delta > 0$ implies $\text{disc}(f) = -4a^3 - 27b^2 > 0$, which is only possible if $a < 0$. Hence the square-root $\sqrt{-a/3}$ is real. We then see that $q < 0$ if and only if $|t| > \sqrt{-a/3}$. Together, the two conditions $q < 0$ and $s < q^2/4$ are simultaneously satisfied if and only if $t > \sqrt{-a/3}$.

Finally, consider the location of the roots of $f(x)$ on the real number line. By Rolle's theorem, the three real distinct roots of $f(x)$ are separated by the real roots of $f'(x) = 3x^2 + a$, which are $x = \pm\sqrt{-a/3}$. The component of $E(\mathbf{R})$ to the left of $\sqrt{-a/3}$ is $E_1(\mathbf{R})$ and to the right of $\sqrt{-a/3}$ is $E_0(\mathbf{R})$. Hence, $t > \sqrt{-a/3}$ and $(t,u) \in E(\mathbf{R})$ if and only if $(t,u) \in E_0(\mathbf{R})$.

Case 3: $\Delta = 0$. In this case P is a 2-torsion point of E with coordinates $(t,0)$ and $F(x,t)$ simplifies to

$$F(x,t) = x^4 - (2a + 6t^2)x^2 + a^2 - 3t^2a - 9bt = (x^2 - (a + 3t^2))^2. \quad (3.2)$$

The preimages of P are 4-torsion points of E and we will show that in each of the cases where E has one or two components, there is a unique 2-torsion point in $E_0(\mathbf{R})$ with real preimages. We treat the two cases separately and remark that the result of [1] does not apply to either case since the coefficient of x is 0.

Case 3a: $\Delta = 0$ and $\text{disc}(f) < 0$. Here $E(\mathbf{R}) = E_0(\mathbf{R})$ consists of one component and hence one non-trivial 2-torsion point defined over $\mathbf{R}$. Let $P = (t,0)$ denote the unique non-trivial 2-torsion point of $E(\mathbf{R})$. Observe further that if the cubic polynomial $f(x) = x^3 + ax + b$ has one real root t, then $a + 3t^2 > 0$ (denote the three roots of f by t, $-t/2 \pm is$, note that $a = -3t^2/4 + s^2$, and add $3t^2$). Denote by θ the positive square root of $a + 3t^2$. Since we translated x by t in (2.4) to put $F(x,t)$ into reduced form, we can recover the x-coordinates of the preimages of P from (3.2) as $x = t \pm \theta$. Therefore, the four preimages of P are given by

$$(t + \theta, \pm\sqrt{f(t + \theta)}) \qquad \text{and} \qquad (t - \theta, \pm\sqrt{f(t - \theta)}).$$

Because f has a single real root t, and because $f(x) > 0$ if and only if $x > t$, it follows that $f(t+\theta) > 0$ and $f(t-\theta) < 0$. Hence $f(t+\theta)$ has real square roots and $f(t-\theta)$ does not. Thus, P has two real preimages, given in coordinates by $(t+\theta, \pm\sqrt{f(t+\theta)})$.

Case 3b: $\Delta = 0$ and $\operatorname{disc}(f) > 0$. Since $\operatorname{disc}(f) > 0$, we have that $E(\mathbf{R})$ consists of two connected components and hence $E[2] \subset E(\mathbf{R})$. We will show there is precisely one non-trivial 2-torsion point whose preimages are all real-valued; namely the point belonging to $E_0(\mathbf{R})$. We will also show that none of the preimages of the other two nontrivial 2-torsion points are real-valued.

If the x-coordinates of the 2-torsion points of E are denoted t_1, t_2, and t_3, then by the same reasoning as in **Case 2** above, exactly one of the t_i is greater than $\sqrt{-a/3}$. Without loss of generality we take $t_1 < t_2 < t_3$ and note that $t_1, t_2 \in E_1(\mathbf{R})$, $t_3 \in E_0(\mathbf{R})$, and $t_1 < -\sqrt{-a/3}$ and $t_3 > \sqrt{-a/3}$. For $i \in \{1, 3\}$, let θ_i denote the positive square root of $3t_i^2 + a$. We will now show that $(t_3, 0)$ is the 2-torsion point with real preimages.

Since $t_3 > \sqrt{-a/3}$, it follows from (3.2) that the x-coordinates of the preimages of $(t_3, 0)$ are real-valued and given by $t_3 \pm \theta_3$. For the y-coordinates, we must determine when

$$y^2 = x^3 + ax + b = \frac{x^4 - 2ax^2 - 8bx + a^2}{4t_3} > 0. \tag{3.3}$$

By (3.1), we have $\operatorname{disc}\, x^4 - 2ax^2 - 8bx + a^2 = 2^{12}b^2 \operatorname{disc}(f) \geq 0$; note that since $\operatorname{disc}(f) > 0$ it follows that $-2a > 0$.

If $b \neq 0$, then by [1] none of the roots of $x^4 - 2ax^2 - 8bx + a^2$ are real-valued and the sign of its leading term is positive. Hence

$$x^4 - 2ax^2 - 8bx + a^2 > 0$$

for all x and so by (3.3) $y^2 > 0$ if and only if $t_3 > 0$. If $b = 0$, then $y^2 = (x^2 - a)^2/4t_3$ and so again $y^2 > 0$ if and only if $t_3 > 0$. In both cases taking square roots shows explicitly that the y-coordinates of the preimages are real-valued as well. This shows that the unique 2-torsion point of E belonging to $E_0(\mathbf{R})$ has four real preimages under duplication.

To finish the proof of **Case 3b** we must show that neither of the points $(t_1, 0), (t_2, 0) \in E_1(\mathbf{R})$ have any real-valued preimages. Note that it is not possible for $t_2 = \sqrt{-a/3}$ for then t_2 would be simultaneously a root of f and f', contradicting the separability assumption of f. Thus, we know $|t_2| < \sqrt{-a/3}$ and additionally $t_3 < -\sqrt{-a/3}$. Because of the range of values of t_2, it follows that $3t_2^2 + a < 0$, so its square root is imaginary, and hence the x-coordinates of the preimages of $(t_2, 0)$ are not real-valued.

The x-coordinates of the preimages of $(t_1, 0)$ are given by $t_1 \pm \theta_1$. It therefore remains to show that both $f(t_1 \pm \theta_1)$ are negative, and hence that the y-coordinates, $\sqrt{f(t_1 \pm \theta_1)}$, of the preimages are not real-valued. But for any $w < t_1$, we have $f(w) < 0$, hence $f(t_1 - \theta_1) < 0$.

It is routine to check that $f(t_1 + \theta_1) = \theta_1^2(3t_1 + 2\theta_1)$. Hence, to finish the proof of Theorem 1.1 it suffices to show $3t_1 + 2\theta_1 < 0$. Since $t_1 + t_2 + t_3 = 0$, we may write

$$\begin{aligned} 3t_1 + 2\theta_1 &= 3t_1 + 2\sqrt{3t_1^2 + a} \\ &= 3t_1 + 2\sqrt{3t_1^2 - t_1^2 - t_1t_2 - t_2^2} \\ &= 3t_1 + 2\sqrt{(2t_1 + t_2)(t_1 - t_2)}. \end{aligned}$$

Because $t_1 < t_2$ and $(2t_1 + t_2)(t_1 - t_2) > 0$, the quantity $(2t_1 + t_2)(t_1 - t_2)$ is maximized when $2t_2 + t_1 = 0$. Setting $t_2 = -t_1/2$ yields

$$3t_1 + 2\theta_1 = 3t_1 + 3|t_1| = 0.$$

However, if $t_2 = -t_1/2$ then $t_3 = -t_1/2$ as well, contradicting the separability assumption of f. It follows that

$$2\sqrt{(2t_1 + t_2)(t_1 - t_2)} < 3|t_1|$$

and hence that $3t_1 + 2\theta_1 < 0$. This completes the proof of Theorem 1.1. □

Acknowledgements. We would like to thank Yuri Zarhin for helpful comments and the referee for their careful reading of the manuscript which greatly improved the exposition of this article.

References

[1] Dickson, L.E. Elementary Theory of Equations. New York, Wiley, 1914.
[2] Lang, S. Elliptic Curves Diophantine Analysis. Grundlehren der Mathematischen Wissenschaften **231**, *Springer-Verlag, Berlin-New York,* 1978.
[3] Silverman, J. H. The arithmetic of elliptic curves. Graduate Texts in Mathematics, **106**. Springer, 2009.
[4] Zarhin, Y. Division by 2 on elliptic curves. Preprint, posted 30 May, 2016. `http://arxiv.org/pdf/1605.09279v1.pdf`.

Department of Mathematics, Bard College, Annandale-On-Hudson, NY 12504, USA
E-mail address: `cullinan@bard.edu`