1.15 26/07/08 - Finishing the Proof of Nagell-Lutz, Introduction to Descent and the Mordell-Weil Theorem

The first order of business today is to wrap up the proof of the Nagell-Lutz Theorem (1.13.7) by finishing proving that torsion points have integer coordinates (1.14.2). After this, we will start building the theory needed to prove the Mordell-Weil Theorem: the machinery of descent.

1.15.A Completing the Proof of Nagell-Lutz

Let's recall where we are: $E : y^2 = x^3 + ax^2 + bx + c$ is an elliptic curve over $\mathbb{Q}$ with $a, b, c \in \mathbb{Z}$, and for each prime p and integer $r \in \mathbb{Z}_{\geq 1}$ we have defined a subgroup $E_{p,r}(\mathbb{Q}) \subset E(\mathbb{Q})$ consisting of points with $v_p(x) \leq -2r$. The end of the proof last time suggested that torsion points don't like being in the subgroups $E_{p,r}(\mathbb{Q})$. This is indeed correct, and we will see how to formalize this momentarily. However, note first that the above proof gives us something stronger than we were going for; we have actually shown

Proposition 1.15.1. In the above notation, suppose that $P_1, P_2, P_3 \in E_{p,r}(\mathbb{Q})$ (counted with multiplicities!) sum to zero, i.e., are such that $P_1 + P_2 + P_3 = \mathscr{O}$. Then their t-coordinates have the property that

$$v_p(t_1 + t_2 + t_3) \geq 3r.$$

Technically, we haven't checked this in every single case; there are some edge cases left to check, which I leave to the reader.

Exercise 1.15.2. Figure out what edge cases needed to be checked, and verify that this result is true also in these edge cases.

This suggests that something interesting is happening with the t-coordinates. For this, we can consider the map $t : E_{p,r}(\mathbb{Q}) \to p^r\mathbb{Z}_{(p)}$, where $\mathbb{Z}_{(p)} := \{q \in \mathbb{Q} : v_p(q) \geq 0\}$ is the subring of $\mathbb{Q}$ consisting of of rational numbers with "no p in the denominator". This ring has many fascinating properties; for instance, it has unique factorization, and in fact the only prime element it has is p (up to units, of which there are lots!). If you haven't seen these facts before, you should try proving them yourself:

Exercise 1.15.3. Show that the ring $\mathbb{Z}_{(p)}$ above has the claimed properties.

Now could the map $t : E_{p,r}(\mathbb{Q}) \to p^r\mathbb{Z}_{(p)}$ be a group homomorphism? Not quite, because even if $P_1, P_2, P_3 \in E_{p,r}(\mathbb{Q})$ sum to zero, their t-coordinates need not sum to zero in $p^r\mathbb{Z}_{(p)}$, which is certainly a necessary condition for a map to be a group homomorphism; however, 1.15.1 suggests how to fix this. We post-compose with the quotient by the subgroup $p^{3r}\mathbb{Z}_{(p)} \subset p^r\mathbb{Z}_{(p)}$ to get a map

$$[t] : E_{p,r}(\mathbb{Q}) \to p^r\mathbb{Z}_{(p)}/p^{3r}\mathbb{Z}_{(p)}. \tag{1.15.4}$$

This map $[t]$ certainly sends $\mathscr{O}$ to 0, and since negation in (t, s) coordinates is simply $-(t, s) = (-t, -s)$ (check!), it certainly preserves negatives. Therefore, in light of 1.15.1, we see that that 1.15.4 is actually a group homomorphism, thanks to

Lemma 1.15.5. Let A and B be an abelian groups and $\phi : A \to B$ be a map of sets such that

(i) we have $\phi(0) = 0$,
(ii) if $a \in A$, then $\phi(-a) = -\phi(a)$, and
(iii) if $a_1, a_2, a_3 \in A$ are such that $a_1 + a_2 + a_3 = 0$, then also $\phi(a_1) + \phi(a_2) + \phi(a_3) = 0$.

Then ϕ is a group homomorphism.

Exercise 1.15.6. Prove 1.15.5 (c.f. 1.14.14).

We have now produced a homomorphism $[t]$ out of $E_{p,r}(\mathbb{Q})$. What is the kernel of $[t]$? It's the set of all points whose t-coordinates actually lie in $p^{3r}\mathbb{Z}_{(p)}$, but we have a name for such points: these are

precisely the elements of $E_{p,3r}(\mathbb{Q})$. Therefore, $\ker[t] = E_{p,3r}(\mathbb{Q})$, and the First Isomorphism Theorem gives us an *injective homomorphism* of abelian groups

$$E_{p,r}(\mathbb{Q})/E_{p,3r}(\mathbb{Q}) \hookrightarrow p^r\mathbb{Z}_{(p)}/p^{3r}\mathbb{Z}_{(p)}.$$

Since the right side is isomorphic to $p^r\mathbb{Z}/p^{3r}\mathbb{Z}$ (check!), it is cyclic of order p^{2r}. This is sufficient to deduce

Proposition 1.15.7. In the above notation, if $P \in E_{p,r}(\mathbb{Q})$, then $p^{2r}P \in E_{p,3r}(\mathbb{Q})$.

I like to think of 1.15.7 to be a strengthening of 1.14.12; while the latter says that the denominators are "persistant", this one says that the denominators actually get *worse* as you keep adding points–unless you get to $\mathscr{O}$ first. We are now ready to prove 1.14.2.

Proof of 1.14.2. Suppose that the result is not true, and pick a point $P \in \operatorname{Tors} E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$ such that $x(P) \notin \mathbb{Z}$ of least possible order $n \in \mathbb{Z}_{\geq 2}$ among such points. (This is possible by the well-ordering principle!) We will use this point to get a contradiction.

The condition that $x(P) \notin \mathbb{Z}$ just says that there is a prime p such that $P \in E_{p,1}(\mathbb{Q})$; in fact, let $r := -v_p(x(P))/2$ (which belongs to $\mathbb{Z}_{\geq 1}$ by 1.14.8), so that $P \in E_{p,r}(\mathbb{Q}) \smallsetminus E_{p,r+1}(\mathbb{Q})$. Now we split into two cases:

(a) Suppose that $p \mid n$, and let $n = pm$ for some $m \in \mathbb{Z}_{\geq 1}$. Let $Q := pP$. Since P has exact order n, the point Q has exact order $m < n$. On the one hand, since $E_{p,r}(\mathbb{Q})$ is a subgroup (1.14.12) and $P \in E_{p,r}(\mathbb{Q})$, we must also have $Q \in E_{p,r}(\mathbb{Q}) \subset E_{p,1}(\mathbb{Q})$. On the other hand, since $m < n$ and n was chosen to be minimal, if $Q \neq \mathscr{O}$, then Q has integer coordinates, and hence $Q \notin E_{p,1}(\mathbb{Q})$. The only way out of this crisis is if $Q = \mathscr{O}$, which means that $m = 1$ and $n = p$. But if this happens, then using that $[t]$ is a homomorphism, we see that

$$0 = [t](Q) = [t](pP) = p \cdot [t](P) \in p^r\mathbb{Z}_{(p)}/p^{3r}\mathbb{Z}_{(p)},$$

forcing $p \cdot t(P) \in p^{3r}\mathbb{Z}_{(p)}$, which is equivalent to $P \in E_{p,3r-1}(\mathbb{Q})$. This contradicts that $P \notin E_{p,r+1}(\mathbb{Q})$ since $3r - 1 \geq r + 1$.

(b) Finally, suppose that $\gcd(p, n) = 1$. Then Bézout's Lemma implies there are $a, b \in \mathbb{Z}$ such that $an + bp^{2r} = 1$. Applying this to P and recalling that $nP = \mathscr{O}$, we conclude that $P = b(p^{2r}P)$. Now by 1.15.7, we have that $p^{2r}P \in E_{p,3r}(\mathbb{Q})$. Since $E_{p,3r}(\mathbb{Q})$ is a subgroup (1.14.12), this implies that $P = b(p^{2r}P) \in E_{p,3r}(\mathbb{Q})$, which is again a contradiction to $P \notin E_{p,r+1}(\mathbb{Q})$.

■

This completes the proof of 1.14.2 and hence the Nagell-Lutz Theorem (1.13.7).

1.15.B The Mordell-Weil Theorem and The Machinery of Descent

The following discussion has been adapted from [2] §3.1]. The primary goal for the rest of the course is to prove, as promised, the famous "finite basis" theorem of Mordell.

Theorem 1.15.8 (Mordell). Let $E/\mathbb{Q}$ be an elliptic curve. The group $E(\mathbb{Q})$ is finitely generated.

As noted before, this theorem is equivalent to saying that there are finitely many "seed points" in $E(\mathbb{Q})$ such that every other rational point on E can be obtained by iteratively applying the chord-and-tangent processes to the seeds. This theorem was first conjectured by Henri Poincaré in around 1901, and proven by British mathematician Louis Mordell in 1922. Later, André Weil (the advisor of Élisabeth Lutz mentioned above) generalized the result to abelian varieties over number fields in 1928 in his doctoral thesis; this is why the theorem is sometimes also known as the Mordell-Weil theorem, although the case of elliptic curves over $\mathbb{Q}$ is essentially due to Mordell.

To prove 1.15.8, we will need criteria to study when a group is finitely generated. Suppose that A is a finitely generated abelian group. What can we then say about $A/2A$? Well, if we pick $n \in \mathbb{Z}_{\geq 1}$ and

generators $a_1, \dots, a_n \in A$ of A, then the 2^n elements $\sum_{i \in I} a_i$ as I runs over subsets of $\{1, \dots, n\}$ form coset representatives for all of $A/2A$, and hence $A/2A$ is a *finite* group of size at most 2^n. Therefore, for A to be finitely generated, we certainly need $A/2A$ to be finitely generated. So the first thing we need to show is

Theorem 1.15.9 (Weak Finite Basis)**.** Let $E/\mathbb{Q}$ be an elliptic curve. Then $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite.

Is this condition enough? Sadly, no. Even if we ask for A/pA to be finite for *all* primes p, this is still not enough: the group $A = \mathbb{Q}$ (the additive group of rational numbers) has the property that $A/pA = 0$ for all primes p, and yet A is *not* finitely generated.

Exercise 1.15.10. If you haven't seen this before, show that the additive group of rational numbers $A = \mathbb{Q}$ is *not* finitely generated.

What is missing is a notion of "size". To motivate this, let us look back at our proof of Fermat's Last Theorem for $n = 4$ (1.3.A). Fermat's proof proceeded by *infinite descent* in which we started with one solution and produced a "smaller" one. The well-ordering principle then tells us that this procedure cannot go on forever. For this to work, we need the notion of the "size" of a solution; fortunately, there, we could just use the size of integers.

In general, given an abelian group A, suppose we have a notion of "size" (also known as "height") on A as measured by a function $h : A \to [0, \infty)$. For something like what we did in 1.3.A to work for this size function h, we need it to satisfy some properties. Let's think through what some of these might be.

(a) Firstly, we need to ensure that for any bound say $C_1 \in \mathbb{R}_{\geq 0}$, there are only finitely many points of size at most C_1. Indeed, if there were infinitely many points of small size, then no proof along the lines of 1.3.A is going to work.

(b) Secondly, the height function should work well with translation on the group (i.e., must interact with the group structure). In other words, we need to ask that for all $b \in A$, the map $\tau_b : A \to A$ given by $a \mapsto a + b$ doesn't change heights too much, i.e., that we have $h(a + b) \approx h(a)$. Actually, it suffices to give an upper bound, i.e., to ask that for all $b \in A$ there is a constant $C_2(b) \in \mathbb{R}_{\geq 0}$ (possibly depending on b) such that for all $a \in A$ we have $h(a + b) \leq 2h(a) + C_2(b)$.

(c) Finally, we need to ensure that the size of an element does scale with taking multiples; that $h(na) \gtrsim nh(a)$ for all $a \in A$ and $n \in \mathbb{Z}_{\geq 1}$. Actually, for the purposes of our discussion, we will deal with a function h that grows quadratically and not linearly, so we might replace the scaling factor n on the right side by n^2. Again it suffices to prove things up to overall constants. Therefore, we might require that there be some $C_3 \in \mathbb{R}_{\geq 0}$ such that for all $a \in A$ we have $h(2a) \geq 4h(a) - C_3$.

If the above conditions don't feel very motivated yet, I sure hope they do after you see

Theorem 1.15.11 (Abstract Descent)**.** Let A be an abelian group, and let $h : A \to [0, \infty)$ be a function. Suppose that h satisfies the three conditions:

(a) (Finiteness) For each $C_1 \in \mathbb{R}_{\geq 0}$, the set $\{a \in A : h(a) \leq C_1\}$ is finite.
(b) (Translation) For each $b \in A$, there is a $C_2(b) \in \mathbb{R}_{\geq 0}$ (possibly depending on b), such that for all $a \in A$ we have $h(a + b) \leq 2h(a) + C_2(b)$.
(c) (Scaling) There is a $C_3 \in \mathbb{R}_{\geq 0}$ such that for all $a \in A$, we have $h(2a) \geq 4h(a) - C_3$.

If further $A/2A$ is finite, then A is finitely generated.

A function $h : A \to [0, \infty)$ on an abelian group A satisfying (a), (b), and (c) as above is called a *height function*. To motivate the proof of 1.15.11, suppose we take a finite set $S \subset A$ of coset representatives for $A/2A$. This means that given any $a \in A$, there is a $b_1 \in S$ and $a_1 \in A$ such that $a = b_1 + 2a_1$. Now we hope that a_1 has smaller "size" than a (as measured by h), and indeed

$$h(a_1) \leq \frac{1}{4}h(2a_1) + \frac{1}{4}C_3 \leq \frac{1}{4}\left(2h(a) + C_2(-b_1)\right) + \frac{1}{4}C_3 = \frac{1}{2}h(a) + \frac{1}{4}\left(C_2(-b_1) + C_3\right),$$

where the first step uses (c) and the second uses (b). Since there are only finitely many possibilities for what $b_1 \in S$ can be, we might as well take $C_2 := \max_{b \in S} C_2(-b)$ and let $C := (C_2 + C_3)/4$ to get the more uniform bound

$$h(a_1) \leq \frac{1}{2}h(a) + C. \tag{1.15.12}$$

This tells us that (up to the constant term C), the height of a_1 is less than half that of a. In particular, we can iterate this procedure to get points of smaller and smaller height, until we land in a set of height small enough that is independent of the $a \in A$ we began with. Since the set of such points will be finite by (a), this will be enough to prove that A is finitely generated. Let's make this into a formal proof.

Proof. Let $S \subset A$ be a finite set of coset representatives for $A/2A$, and let $C_2 := \max_{b \in S} C_2(-b)$ and $C := (C_2 + C_3)/4$ as above. Let $T := \{a \in A : h(a) \leq 2C + 1\}$, which is finite by (a). (We'll see where this bound in the definition of T comes from in just a moment.) We claim that $S \cup T$ generate A.

To see this, suppose we are given an $a \in A$; we need to show that a lies in the subgroup of A generated by $S \cup T$. As above, since S forms a complete set of representatives, there is a $b_1 \in S$ and $a_1 \in A$ such that $a = b_1 + 2a_1$. Iterating this process, we obtain recursively the elements $a_n \in A$ and $b_n \in S$ such that for each $n \in \mathbb{Z}_{\geq 1}$ we have $a_{n-1} = b_n + 2a_n$ (where $a_0 := a$). By an iteration of 1.15.12, we see immediately that for each $n \geq 1$,

$$h(a_n) \leq \frac{1}{2^n} h(a) + \left(1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^{n-1}}\right) C \leq \frac{1}{2^n} h(a) + 2C.$$

This tells us that if $N \gg 1$ (specifically, $N \geq \log_2 h(a)$ if $h(a) \neq 0$; if $h(a) = 0$ then $a \in T$ and there is nothing to show), then $a_N \in T$. It then follows from the unfolding

$$a = b_1 + 2b_2 + \cdots + 2^{N-1} b_N + 2^N a_N$$

that a lies in the subgroup of A generated by $S \cup T$, as needed. ∎

Therefore, to prove 1.15.8 for an elliptic curve $E/\mathbb{Q}$ by invoking 1.15.11 for $A = E(\mathbb{Q})$, we need to do two things: we need to prove the weak finite basis theorem (1.15.9), and then we need to come up with a function $h : E(\mathbb{Q}) \to [0, \infty)$ that satisfies properties (a), (b), (c) of 1.15.11. We'll tackle the second of these issues first next time.

Remark 1.15.13. Why is 1.15.11 called the "descent" theorem? Well, the terminology really does come from Fermat's method of infinite descent which we saw in 1.3.A. The idea is that we start with a solution (here, a), and produce from it a "smaller" solution (here, a_1). We keep continuing this way to get smaller and smaller solutions (here, a_n). If we get to a point where the smaller solution is smaller than is possible for that particular Diophantine equation (as happened in 1.3.A), then we are done–we have shown there are no solutions. What happens more generally is that we keep iterating this procedure to get to a point where the last iterate in this process is so small that it lies in some finite set of solutions that we understand (here, T). Then, up to the finitely many terms b_i which go into getting to this small solution, this finite set of "small" solutions generates every other.

Unimportant Remark 1.15.14. There is nothing special about the number 2 above, and it could have been replaced by any integer $m \geq 2$ (c.f. [8, Thm. VIII.3.1]). It just that the choice of $m = 2$ is very convenient, as we shall see next time.