HILBERT’S TENTH PROBLEM AND MAZUR’S CONJECTURE
FOR LARGE SUBRINGS OF \( \mathbb{Q} \)

BJORN POONEN

1. Introduction

Hilbert’s Tenth Problem, in modern terms, was to find an algorithm (Turing machine) to decide, given a polynomial equation \( f(x_1, \ldots, x_n) = 0 \) with coefficients in \( \mathbb{Z} \), whether it has a solution with \( x_1, \ldots, x_n \in \mathbb{Z} \). Y. Matijasevič [Mat70], building on earlier work of M. Davis, H. Putnam, and J. Robinson [DPR61], showed that no such algorithm exists. If one replaces \( \mathbb{Z} \) in both places by a different commutative ring \( R \) (let us assume its elements can be and have been encoded for input into a Turing machine), one obtains a different question, called Hilbert’s Tenth Problem over \( R \), whose answer depends on \( R \). These problems are discussed in detail in [DLPVG00].

In particular, the answer for \( R = \mathbb{Q} \) is unknown. Hilbert’s Tenth Problem over \( \mathbb{Q} \) is equivalent to the general problem of deciding whether a variety over \( \mathbb{Q} \) has a rational point. One approach to proving that Hilbert’s Tenth Problem over \( \mathbb{Q} \) has a negative answer would be to deduce this from Matijasevič’s theorem for \( \mathbb{Z} \), by showing that \( \mathbb{Z} \) is diophantine over \( \mathbb{Q} \) in the following sense:

Definition 1.1. Let \( R \) be a ring, and let \( A \subseteq R^m \). Then \( A \) is diophantine over \( R \) if and only if there exists a polynomial \( f \) in \( m + n \) variables with coefficients in \( R \) such that

\[
A = \{ a \in R^m \mid \exists x \in R^n \text{ such that } f(a, x) = 0 \}.
\]

On the other hand, Mazur conjectures that if \( X \) is a variety over \( \mathbb{Q} \), then the topological closure of \( X(\mathbb{Q}) \) in \( X(\mathbb{R}) \) has only finitely many components [Maz92, Maz95]. This would imply that \( \mathbb{Z} \) is not diophantine over \( \mathbb{Q} \). More generally, Cornelissen and Zahidi [CZ00] have shown that Mazur’s Conjecture implies that there is no diophantine model of \( \mathbb{Z} \) over \( \mathbb{Q} \).

Definition 1.2. A diophantine model of \( \mathbb{Z} \) over \( \mathbb{Q} \) is a set \( A \subseteq \mathbb{Q}^n \) that is diophantine over \( \mathbb{Q} \) with a bijection \( \mathbb{Z} \rightarrow A \) under which the graphs of addition and multiplication on \( \mathbb{Z} \) correspond to subsets of \( A^3 \subseteq \mathbb{Q}^3n \) that are diophantine over \( \mathbb{Q} \).

This is important, because the existence of such a diophantine model, together with Matijasevič’s Theorem, would imply a negative answer for Hilbert’s Tenth Problem over \( \mathbb{Q} \).
This paper studies Hilbert’s Tenth Problem over rings between \( \mathbb{Z} \) and \( \mathbb{Q} \). Such rings are in bijection with subsets of the set \( \mathcal{P} \) of prime numbers. Namely, given \( S \subseteq \mathcal{P} \), one has the ring \( \mathbb{Z}[S^{-1}] \), and conversely, given a subring \( R \) between \( \mathbb{Z} \) and \( \mathbb{Q} \), one has \( R = \mathbb{Z}[S^{-1}] \) where \( S = \mathcal{P} \cap R^\times \).

Using quadratic forms as in J. Robinson’s work, one can show that for any prime \( p \) the ring \( \mathbb{Z}(p) \) of rational numbers with denominators prime to \( p \) is diophantine over \( \mathbb{Q} \) [KR92, Proposition 3.1]. A short argument using this shows that for finite \( S \), Hilbert’s Tenth Problem over \( \mathbb{Z}[S^{-1}] \) has a negative answer.

In this paper, we give the first examples of infinite subsets \( S \) of \( \mathcal{P} \) for which Hilbert’s Tenth Problem over \( \mathbb{Z}[S^{-1}] \) has a negative answer. In fact, we show that there exist such \( S \) of natural density 1, so in one sense, we are approaching a negative answer for \( \mathbb{Q} \). (See Section 6 for the definition of natural density.)

Previously, Shlapentokh proved that if \( K \) is a totally real number field or a totally complex degree-2 extension of a totally real number field, then there exists a set of places \( S \) of Dirichlet density arbitrarily close to \( 1 - [K : \mathbb{Q}]^{-1} \) such that if \( \mathcal{O}_{K,S} \) is the subring of elements of \( K \) that are integral at all places outside \( S \), then Hilbert’s Tenth Problem over \( \mathcal{O}_{K,S} \) has a negative answer [Shl97], [Shl00], [Shl02]. But for \( K = \mathbb{Q} \), this gives nothing beyond Matijasevič’s Theorem.

More generally, we prove the following:

**Theorem 1.3.** There exist disjoint recursive sets of primes \( T_1 \) and \( T_2 \), both of natural density 0, such that for any set \( S \) of primes containing \( T_1 \) and disjoint from \( T_2 \), the following hold:

1. There exists an affine curve \( E' \) over \( \mathbb{Z}[S^{-1}] \) such that the topological closure of \( E'(\mathbb{Z}[S^{-1}]) \) in \( E'(\mathbb{R}) \) is an infinite discrete set.
2. The set of positive integers with addition and multiplication admits a diophantine model over \( \mathbb{Z}[S^{-1}] \).
3. Hilbert’s Tenth Problem over \( \mathbb{Z}[S^{-1}] \) has a negative answer.

**Remark 1.4.**

(i) Arguably (3) is the most important of the three parts. We have listed the parts in the order they will be proved.

(ii) A subset \( T \subseteq \mathbb{Z} \) is recursive if and only if there exists an algorithm (Turing machine) that takes as input an integer \( t \) and outputs YES or NO according to whether \( t \in T \).

(iii) We use natural density instead of Dirichlet density in order to have a slightly stronger statement. See [Ser73, VI.4.5] for the definition of Dirichlet density and its relation to natural density.

Previously, Shlapentokh [Shl03] used norm equations to prove that there exist sets \( S \subseteq \mathcal{P} \) of Dirichlet density arbitrarily close to 1 for which there exists an affine variety \( X \) over \( \mathbb{Z}[S^{-1}] \) such that the closure of \( X(\mathbb{Z}[S^{-1}]) \) in \( X(\mathbb{R}) \) has infinitely many connected components. (She also proved an analogous result for localizations of the ring of integers of totally real number fields and totally complex degree-2 extensions of totally real number fields. For number fields with exactly one conjugate pair of nonreal embeddings, she obtained an analogous result, but with density only 1/2.)

Question 4.1 of [Shl03] asked whether over \( \mathbb{Q} \) one could do the same for some \( S \subseteq \mathcal{P} \) of Dirichlet density exactly 1. Part (1) of our Theorem 1.3 gives an affirmative answer (take \( S = \mathcal{P} - T_2 \)). In fact, it was the attempt to answer Shlapentokh’s
Lemma 3.1. The notation

Projective model of $y$ have complex multiplication. For example, these conditions hold for the smooth proper curve over $\mathbb{Z}$.

Assume moreover that $E$ is smooth over $\mathbb{Z}$.

Theorem 1.3 to other number fields and to places other than the real place.

2. ELLIPTIC CURVE SETUP

Let $E$ be an elliptic curve over $\mathbb{Q}$ of rank 1. To simplify the arguments, we will assume moreover that $E(\mathbb{Q}) \simeq \mathbb{Z}$, that $E(\mathbb{R})$ is connected, and that $E$ does not have complex multiplication. For example, these conditions hold for the smooth projective model of $y^2 = x^3 + x + 1$. Let $P$ be a generator of $E(\mathbb{Q})$. Fix a Weierstrass equation $y^2 = x^3 + ax + b$ for $E$, where $a, b \in \mathbb{Z}$.

Let $E' = \text{Spec} \mathbb{Z}[S^{-1}]$ be a Weierstrass equation $y^2 = x^3 - (x^3 + ax + b)$, where $S$ is a finite set of primes such that $E'$ is smooth over $\mathbb{Z}[S^{-1}]^{-1}$. In particular, $2 \in S$. Enlarge $S$ if necessary so that $P \in E'(\mathbb{Z}[S^{-1}])$.

3. DENOMINATORS OF $x$-COORDINATES

For nonzero $n \in \mathbb{Z}$, let $d_n \in \mathbb{Z}_{>0}$ be the prime-to-$S$ part of the denominator of $x(nP)$; that is, $d_n$ is the product one obtains if one takes the prime factorization of the denominator of $x(nP)$ and omits the powers of primes in $S$. Define $d_0 = 0$.

The notation $m | n$ means $n = mZ$.

Lemma 3.1.

(a) For any $r \in \mathbb{Z}$, the set \{ $n \in \mathbb{Z} : r | d_n$ \} is a subgroup of $\mathbb{Z}$.
(b) There exists $c \in \mathbb{R}_{>0}$ such that $\log d_n = (c - o(1))n^2$ as $n \to \infty$ (cf. [Ser87], Lemma 8).

Proof. (a) We may reduce to the case where $r = p^e$ for some prime $p \notin S$ and $e \in \mathbb{Z}_{>0}$. Thus it suffices to show that the set $E_\epsilon := \{ Q \in E(\mathbb{Q}_p) : v_p(x(Q)) \leq -\epsilon \} \cup \{ O \}$ is a subgroup of $E(\mathbb{Q}_p)$, where $v_p : \mathbb{Q}_p^* \to \mathbb{Z}$ is the $p$-adic valuation. The set $E_1$ is the kernel of the reduction map $E(\mathbb{Q}_p) \to E(\mathbb{F}_p)$ (extend $E$ to a smooth proper curve over $\mathbb{Z}_p$ to make sense of this). Since $p > 2$, the formal logarithm $\lambda : E_1 \to p\mathbb{Z}_p$ is an isomorphism [Sil82, IV.6.4]. By [Sil82, IV.5.5, IV.6.3], $v_p(\lambda(Q)) = v_p(z(Q))$ for all $Q \in E_1$, where $z = -x/y$ is the standard parameter for the formal group. By [Sil82, pp. 113-114], $x = -2z^{-2} + \cdots \in \mathbb{Z}_p((z))$, so $v_p(x(Q)) = -2v_p(z(Q)) = -2v_p(\lambda(Q))$. Thus $G_\epsilon$ corresponds under $\lambda$ to $p^{e/2}\mathbb{Z}_p$ and is hence a subgroup.

(b) The number $\log d_n$ is the logarithmic height $h(nP)$, except that in the sum defining the height, the terms corresponding to places in $S$ have been omitted. A standard diophantine approximation result (see Section 7.4 of [Ser97])
implies that each such term contributes at most a fraction $o(1)$ of the height, as $n \to \infty$. If $\hat{h}$ is the canonical height, then $h(nP) = \hat{h}(nP) + O(1) = \hat{h}(P)n^2 + O(1)$. Take $c = \hat{h}(P)$, which is positive, since $P$ is not torsion.

Remark 3.2. The bottom of p. 306 in \cite{Aya92} relates the denominators of $x(nP)$ in lowest terms to the sequence of values of division polynomials evaluated at $P$. The study of divisibility properties of the latter sequence is very old: results were claimed in the 19th century by Lucas (but apparently not published), and proofs were given in \cite{War48}.

For $n \in \mathbb{Z}$, let $S_n$ be the set of prime factors of $d_n$. If $m, n \in \mathbb{Z}$, then $(m, n)$ denotes their greatest common divisor.

Corollary 3.3. If $m, n \in \mathbb{Z}$, then $S_{(m, n)} = S_m \cap S_n$. In particular, if $(m, n) = 1$, then $S_m$ and $S_n$ are disjoint.

Proof. Lemma 3.1(a) implies the first statement. Since $P \in E'(\mathbb{Z}[S_{\text{bad}}^{-1}])$, we have $S_1 = \emptyset$, and the second statement follows.

Lemma 3.4. If $\ell$ and $m$ are primes, and $\max\{\ell, m\}$ is sufficiently large, then $\ell \in \mathbb{Z} - (S_\ell \cup S_m)$ is nonempty.

Proof. If $p \mid d_m$, or equivalently $v_p(x(mP)) < 0$, then using the formal logarithm $\lambda$ as in the proof of Lemma 3.1(a), we obtain

$$v_p(d_m) = -v_p(x(\ell mP)) = 2v_p(\lambda(\ell mP)) = 2v_p(\ell \lambda(mP)) = v_p(\ell^2 d_m).$$

If $S_{\ell m} - (S_\ell \cup S_m)$ were empty, then for each $p \mid d_m$, we could apply either this result or the analogue with $\ell$ and $m$ interchanged and hence deduce $d_m \mid \ell^2 m^2 d_\ell d_m$. This contradicts Lemma 3.1(b) if $\max\{\ell, m\}$ is sufficiently large.

Remark 3.5. Our Lemma 3.1(a) is a special case of Lemma 9 of \cite{Sil88} (except for the minor differences that \cite{Sil88} requires $E$ to be in minimal Weierstrass form and considers the full denominator instead of its prime-to-$S_{\text{bad}}$ part). The method of proof is the same. These results may be viewed as elliptic analogues of Zsigmondy's Theorem; see \cite{Eve02}.

4. Definition of $T_1$ and $T_2$

For each prime number $\ell$, let $a_\ell$ be the smallest $a \in \mathbb{Z}_{>0}$ such that $d_\ell > 1$. By Lemma 3.1(b), $a_\ell = 1$ exists, and $a_\ell = 1$ for all $\ell$ outside a finite set $L$ of primes. Baker's method, Chapter 8] lets us compute the finite set $E'(\mathbb{Z}[S_{\text{bad}}^{-1}])$, so the set $L$ and the values $a_\ell$ for $\ell \in L$ are computable.

Let $p_1 = \max S_{a_\ell}$ where $a = a_\ell$. For primes $\ell$ and $m$ (possibly equal), Lemma 3.1(a) lets us define $p_{\ell m} = \max (S_{\ell m} - (S_\ell \cup S_m))$ when $\max\{\ell, m\}$ is sufficiently large. Let $\ell_1 < \ell_2 < \ldots$ be a sequence of primes outside $L$. (The $\ell_i$ will be constructed in Section 4 with certain properties, but for now these properties are not relevant.) The plan will be to force $\ell_i P \in E'(\mathbb{Z}[S_{\ell_i}^{-1}])$ for all $i$, by requiring each $S_{\ell_i}$ to be contained in $S$. On the other hand, we must require other primes to lie outside $S$ to make sure that not too many other multiples of $P$ end up in $E'(\mathbb{Z}[S_{\ell_i}^{-1}])$.

Let $T_1 = S_{\text{bad}} \cup \bigcup_{i \geq 1} S_{\ell_i}$. Let $T_2^p$ be the set of $p_1$ for all primes $\ell \in \{\ell_1, \ell_2, \ldots\}$. If $\ell_1$ is sufficiently large, we may define $T_2 = \{p_{\ell_i, \ell_j} : 1 \leq j \leq i\}$ and $T_2^p = \{p_{\ell_i, \ell_j} : \ell \in L, i \geq 1\}$. Finally, let $T_2 = T_2^p \cup T_2^* \cup T_2^e$. 

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
5. Properties of \( T_1 \) and \( T_2 \)

**Proposition 5.1.** The sets \( T_1 \) and \( T_2 \) are disjoint.

**Proof.** By definition of \( p_i \) and \( p_{im} \), \( S_{had} \cap T_2 = \emptyset \). If \( \ell \neq \ell_i \), then \( (\ell \ell_i, \ell_i) = 1 \), so Corollary 3.3 implies \( p_i \not\in S_{\ell_i} \). Thus \( T_1 \cap T_2 = \emptyset \). If \( i \not\in \{ j, k \} \), then Corollary 3.3 implies \( p_{\ell_i} \ell \not\in S_{\ell_i} \), while if \( i \in \{ j, k \} \), then \( p_{\ell_i} \ell \not\in S_{\ell_i} \) by definition of \( p_{\ell_i} \). Thus \( T_1 \cap T_2 = \emptyset \).

**Proposition 5.2.** If \( S \) contains \( T_1 \) and is disjoint from \( T_2 \), then \( E'(\mathbb{Z}[S^{-1}]) \) is the union of \( \{ \pm \ell_i P : i \geq 1 \} \) and some subset of the finite set \( \{ rP : r \mid \prod_{\ell \in L} \ell^{\alpha_{\ell}-1} \} \).

**Proof.** Because the equation of \( E \) relates the \( x \)- and \( y \)-coordinates, a point \( nP \) belongs to \( E'(\mathbb{Z}[S^{-1}]) \) if and only if \( S_n \subseteq S \). In particular, \( S_{\ell_i} \subseteq T_1 \subseteq S \), so \( \pm \ell_i P \in E'(\mathbb{Z}[S^{-1}]) \).

Any point outside
\[
\{ \pm \ell_i P : i \geq 1 \} \cup \{ rP : r \mid \prod_{\ell \in L} \ell^{\alpha_{\ell}-1} \}
\]
is \( nP \) for some \( n \) divisible by one of the following:
- \( \ell^{\alpha_{\ell}} \) for some \( \ell \) not in the sequence \( \ell_1, \ell_2, \ldots \),
- \( \ell_i \ell_j \) for some \( 1 \leq j \leq i \), or
- \( \ell_i \ell_i \) for some \( \ell \in L \) and \( i \geq 1 \).

Lemma 6.1(i) implies then that \( S_n \) contains a prime of \( T_2^1, T_2^2 \), or \( T_2^2 \), respectively, so \( S_n \not\subseteq S \).

6. Natural density

The **natural density** of a subset \( T \subseteq \mathcal{P} \) is defined as
\[
\lim_{X \to \infty} \frac{\# \{ p \in T : p \leq X \}}{\# \{ p \in \mathcal{P} : p \leq X \}},
\]
if the limit exists. One defines **upper natural density** similarly, using \( \text{lim sup} \) instead of \( \text{lim} \).

**Lemma 6.1.** If \( \alpha \in \mathbb{R} - \mathbb{Q} \), then \( \{ \ell \alpha \mod 1 : \ell \text{ is prime} \} \) is equidistributed in \([0, 1]\). That is, for any interval \( I \subseteq [0, 1] \), the set of primes \( \ell \) for which \( (\ell \alpha \mod 1) \) belongs to \( I \) has natural density equal to the length of \( I \).

**Proof.** See p. 180 of [Vin54].

Let \( y(\ell P) \in \mathbb{Q} \) denote the \( y \)-coordinate of \( \ell P \in E(\mathbb{Q}) \).

**Corollary 6.2.** If \( I \subseteq \mathbb{R} \) is an interval with nonempty interior, then the set of primes \( \ell \) for which \( y(\ell P) \in I \) has positive natural density.

**Proof.** Since \( E(\mathbb{R}) \) is a connected compact 1-dimensional Lie group over \( \mathbb{R} \), we can choose an isomorphism \( E(\mathbb{R}) \to \mathbb{R}/\mathbb{Z} \) as topological groups. Since \( P \) is of infinite order, its image in \( \mathbb{R}/\mathbb{Z} \) is represented by an irrational number. The subset of \( E(\mathbb{R}) \) having \( y \)-coordinate in \( I \) corresponds to a nontrivial interval in \( \mathbb{R}/\mathbb{Z} \). Now apply Lemma 6.1.
7. Construction of the $\ell_i$

For prime $\ell$, define

$$\mu_\ell = \sup_{X \in \mathbb{Z}_{\geq 2}} \frac{\#\{p \in S_\ell : p \leq X\}}{\#\{p \in \mathcal{P} : p \leq X\}}.$$ 

The supremum is attained for some $X \leq \max S_\ell$, so $\mu_\ell$ is computable for each $\ell$.

**Lemma 7.1.** For any $\epsilon > 0$, the natural density of $\{\ell : \mu_\ell > \epsilon\}$ is 0.

**Proof.** For $X \in \mathbb{R}$, let $\pi(X) := \#\{p \in \mathcal{P} : p \leq X\}$. If $\ell$ is a prime and $\mu_\ell > \epsilon$, then we can choose $X_\ell \in \mathbb{Z}_{\geq 2}$ such that

$$\frac{\#\{p \in S_\ell : p \leq X_\ell\}}{\pi(X_\ell)} > \epsilon.$$

For $M \in \mathbb{Z}_{\geq 2}$, let $U_M$ be the set of primes $\ell$ such that $\mu_\ell > \epsilon$ and $X_\ell \in [M, 2M)$. If $\ell \in U_M$, then

$$\#\{p \in S_\ell : p \leq 2M\} \geq \#\{p \in S_\ell : p \leq X_\ell\} > \epsilon \pi(X_\ell) \geq \epsilon \pi(M).$$

But the $S_\ell$ are disjoint by Corollary 3.3, so

$$\pi(2M) \geq \sum_{\ell \in U_M} \#\{p \in S_\ell : p \leq 2M\} \geq \epsilon \pi(M) \#U_M.$$

Thus by the Prime Number Theorem, $\#U_M = O(1)$ as $M \to \infty$. If $2^{k-1} \leq N < 2^k$, then

$$\#\{\ell : \mu_\ell > \epsilon \text{ and } X_\ell \leq N\} \leq \#U_2 + \#U_4 + \#U_8 + \cdots + \#U_{2^k} = O(k) = O(\log N)$$

as $N \to \infty$. If $\mu_\ell > \epsilon$, then by definition of $X_\ell$,

$$\pi(X_\ell) < \frac{\#S_\ell}{\epsilon} \leq \frac{\log_2 d_\ell}{\epsilon} = O(\ell^2)$$

as $\ell \to \infty$ by Lemma 3.1(b), so $X_\ell = O(\ell^2 \log \ell)$ by the Prime Number Theorem. Combining the previous two sentences shows that

$$\#\{\ell \leq Y : \mu_\ell > \epsilon\} = \#\{\ell \leq Y : \mu_\ell > \epsilon \text{ and } X_\ell \leq O(Y^2 \log Y)\} = O(\log O(Y^2 \log Y)),$$

which is $o(\pi(Y))$ as $Y \to \infty$. \qed

Define the $\ell_i$ inductively as follows. Given $\ell_1, \ldots, \ell_{i-1}$, let $\ell_i$ be the smallest prime outside $L$ such that all of the following hold:

1. $\ell_i > \ell_j$ for all $j < i$,
2. $\mu_{\ell_i} \leq 2^{-i}$,
3. $p_{\ell_i, i} > 2^i$ for all $j \leq i$,
4. $p_{\ell_i, i} > 2^i$ for all $\ell \in L$, and
5. $|y(\ell_i, P) - i| \leq 1/(10i)$.

**Proposition 7.2.** The sequence $\ell_1, \ell_2, \ldots$ is well defined and computable.

**Proof.** By induction, we need only show that for each $i$, there exists $\ell_i$ as above. By Corollary 6.2, the set of primes satisfying (5) has positive natural density. By Lemma 7.1, (2) fails for a set of natural density 0. Therefore it will suffice to show that (1), (3), and (4) are satisfied by all sufficiently large $\ell_i$. 

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
For fixed \( j \leq i \), the primes \( p_{\ell_i} \) for varying values of \( \ell_i \) are distinct by Corollary 8.3, so eventually they are greater than \( 2^j \). The same holds for \( p_{\ell r} \), for fixed \( \ell \in L \). Thus by taking \( \ell \), sufficiently large, we can make all the \( p_{\ell_i \ell} \) and \( p_{\ell r} \), greater than \( 2^j \).

Each \( \ell_i \) can be computed by searching primes in increasing order until one is found satisfying the conditions.

\[ \square \]

8. Recursiveness of \( T_1 \) and \( T_2 \)

The set \( \{ \ell_1, \ell_2, \ldots \} \) is recursive, since it is a strictly increasing sequence whose terms can be computed in order. This is needed for the proofs in this section.

**Proposition 8.1.** The set \( T_1 \) is recursive.

**Proof.** Since \( S_{\text{bad}} \) is finite, it suffices to give an algorithm for deciding whether a prime \( p \notin S_{\text{bad}} \) belongs to \( \bigcup_{i \geq 1} S_{\ell_i} \). We have \( p \in \bigcup_{i \geq 1} S_{\ell_i} \) if and only if \( p \mid d_{\ell_i} \) for some \( i \), which holds if and only if the order \( n_p \) of the image of \( P \) in \( E(F_p) \) divides \( \ell_i \) for some \( i \). The order \( n_p \) can be computed, and \( n_p \neq 1 \), since \( P \in E'(\mathbb{Z}[S_{\text{bad}}^{-1}]) \). So we simply check whether \( n_p \in \{ \ell_1, \ell_2, \ldots \} \).

**Lemma 8.2.** If \( \ell \) is prime, then \( \ell \mid \#E(F_{p_{\ell}}) \).

**Proof.** By definition of \( p_\ell \), the point \( \ell^{\alpha_{\ell}} P \) reduces to 0 in \( E(F_{p_{\ell}}) \) but \( \ell^{\alpha_{\ell} - 1} P \) does not.

\[ \square \]

**Proposition 8.3.** The set \( T_2^a \) is recursive.

**Proof.** If \( p \in T_2^a \), then \( p = p_{\ell} \) for some \( \ell \notin \{ \ell_1, \ell_2, \ldots \} \), and then \( \ell \mid \#E(F_p) \) by Lemma 8.2. Therefore to test whether a prime \( p \notin S_{\text{bad}} \) belongs to \( T_2^a \), compute \( \#E(F_p) \) and its prime factors: one has \( p \in T_2^a \) if and only if there is a prime factor \( \ell \) such that \( \ell \notin \{ \ell_1, \ell_2, \ldots \} \) and \( p_{\ell} = p \).

\[ \square \]

**Proposition 8.4.** The sets \( T_2^b \) and \( T_2^c \) are recursive.

**Proof.** By condition (3) in the definition of \( \ell_i \), if a prime \( p \) belongs to \( T_2^b \), it must equal \( p_{\ell_i \ell} \) for some \( 1 \leq j \leq i \) with \( 2^j < p \). Thus to test whether a prime \( p \) belongs to \( T_2^c \), simply compute \( p_{\ell_i \ell} \) for \( 1 \leq j \leq i < \log_2 p \).

The proof that \( T_2^c \) is recursive is similar, using condition (4).

Thus \( T_1 \) and \( T_2 \) are recursive.

9. The Densities of \( T_1 \) and \( T_2 \)

**Proposition 9.1.** The set \( T_1 \) has natural density 0.

**Proof.** For fixed \( r \in \mathbb{Z}_{>0} \), the set \( \bigcup_{i > r} S_{\ell_i} \) differs from \( T_1 \) in only finitely many primes, so it suffices to show that the former has upper natural density tending to 0 as \( r \to \infty \). By definition of \( \mu_{\ell_i} \), the upper natural density is bounded by \( \sum_{i > r} \mu_{\ell_i} \leq \sum_{i > r} 2^{-i} = 2^{-r} \), which tends to 0 as \( r \to \infty \).

\[ \square \]

**Proposition 9.2.** The sets \( T_2^b \) and \( T_2^c \) have natural density 0.

**Proof.** Suppose \( 2^m \leq X < 2^{m+1} \). By condition (3) defining \( \ell_i \), the only primes of the form \( p_{\ell_i \ell} \) that might be \( \leq X \) are those with \( 1 \leq j \leq i \leq m \). There are at most \( O(m^2) = O((\log X)^2) \) of these, which is negligible compared to \( \pi(X) \). Thus \( T_2^b \) has natural density 0. The proof for \( T_2^c \) is similar.

\[ \square \]
The rest of this section is devoted to the proof that $T_2^0$ has natural density 0. Recall that $T_2^0$ consists of primes of the form $p_2$. If the sequence of $p_2$ grew faster than the sequence of primes $\ell$, then $T_2^0$ would have density 0. But Lemma 8.2 implies only that $p_2$ is at least about the size of $\ell$. The strategy for strengthening this bound will be to show that numbers of the form $\#E(F_p)$ are typically divisible by many primes. For $n \in \mathbb{Z}_{>0}$, let $\omega(n)$ be the number of distinct prime factors of $n$.

**Lemma 9.3.** For any $t \geq 1$, the natural density of $\{ p : \omega(\#E(F_p)) < t \}$ is 0.

**Proof.** For a prime $\ell$, let $E[\ell]$ denote the group of points of order dividing $\ell$ on $E$. Then $\ell \mid \#E(F_p)$ if and only if the image of the Frobenius element at $p$ under $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \text{Aut} E[\ell]$ has a nonzero fixed vector. Since $E$ does not have complex multiplication, the image of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ in $\prod_{E} \text{Aut} E[\ell]$ is open. (This follows from [Ser72].) Thus $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \prod_{E} \text{Aut} E[\ell]$ is surjective for some finite $\mathcal{L} \subseteq \mathcal{P}$. A calculation shows that the fraction of elements of $\text{Aut} E[\ell] \cong \text{GL}_2(\mathbb{F}_\ell)$ having a nonzero fixed vector is

$$\frac{\ell^3 - 2\ell}{(\ell^2 - 1)(\ell^2 - \ell)} = \frac{1}{\ell} + O\left(\frac{1}{\ell^2}\right).$$

The sum of this over $\ell$ diverges, so as $C \to \infty$, the fraction of elements of $\prod_{E \in \mathcal{L}} \text{Aut} E[\ell]$ having fewer than $t$ components with a nonzero fixed vector tends to 0. Applying the Chebotarev Density Theorem (see Théorème 1 of [Ser81] for a version using natural density) and letting $C \to \infty$, we obtain the result. \hfill $\Box$

**Proposition 9.4.** The set $T_2^0$ has natural density 0.

**Proof.** Because of Lemma 9.3 it suffices to show that the upper natural density of

$$T_{2,t}^0 := \{ p \in T_2^0 : \omega(\#E(F_p)) \geq t \}$$

tends to 0 as $t \to \infty$.

Suppose $p = p_\ell \in T_{2,t}^0$. By Lemma 8.2, $\ell \mid \#E(F_p)$. By definition of $T_{2,t}^0$, the integer $\#E(F_p)$ is divisible by at least $t - 1$ other primes, so $2^{t-1}\ell \leq \#E(F_p)$. There exists a degree-2 map $E \to \mathbb{P}^1$ over $\mathbb{F}_p$, so $\#E(F_p) \leq 2(p + 1) \leq 4p$. Combining the previous two sentences yields $\ell \leq 2^{3-t}p$. Since every element of $T_{2,t}^0$ is $p_\ell$ for some $\ell$, we have

$$\#\{ p \in T_{2,t}^0 : p \leq X \} \leq \pi(2^{3-t}X) = (2^{3-t} + o(1)) \pi(X)$$
as $X \to \infty$. Thus by definition, the upper natural density of $T_{2,t}^0$ is at most $2^{3-t}$. This goes to 0 as $t \to \infty$. \hfill $\Box$

Thus $T_1$ and $T_2$ have natural density 0.

10. PROOF OF THEOREM 1.3

By Proposition 5.2 $E'(\mathbb{Z}[S^{-1}])$ differs from $\{ \pm \ell_i P : i \geq 1 \}$ by at most a finite set. Since $y(\pm \ell_i P)$ is within 1/10 of $\pm i$, any bounded subset of $\mathbb{R}^2$ contains at most finitely many points of $E'(\mathbb{Z}[S^{-1}])$. Part (1) of Theorem 1.3 follows.

We next construct a diophantine model $A$ of the positive integers over $\mathbb{Z}[S^{-1}]$. The set of nonzero elements of $\mathbb{Z}[S^{-1}]$ is diophantine (see Theorem 4.2 of [Shi94]), and we can represent elements of $\mathbb{Q}$ as fractions of elements of $\mathbb{Z}[S^{-1}]$ with nonzero
LARGE SUBRINGS OF $\mathbb{Q}$

Therefore equations over $\mathbb{Q}$ can be rewritten as systems of equations over $\mathbb{Z}[S^{-1}]$, and there is no harm in using them in our diophantine definitions. In particular, we may use the predicate $x \geq y$, since it can be encoded as $(\exists z_1, z_2, z_3, z_4 \in \mathbb{Q})(x = y + z_1^2 + z_2^2 + z_3^2 + z_4^2)$.

For $i \in \mathbb{Z}_{>0}$, define $y_i := y(\ell_i P)$. Let $A = \{y_1, y_2, \ldots\}$. Then $A$ is diophantine over $\mathbb{Z}[S^{-1}]$, because it consists of the nonnegative elements of the set of $y$-coordinates of $E'(\mathbb{Z}[S^{-1}])$ minus a finite set. We have a bijection $\mathbb{Z}_{>0} \rightarrow A$ taking $i$ to $y_i$.

It remains to show that the graphs of addition and multiplication on $\mathbb{Z}_{>0}$ correspond to diophantine subsets of $A^3$. We know $|y_i - i| \leq 1/(10i) \leq 1/10$, so the idea is that the addition on $\mathbb{Z}_{>0}$ should correspond to the operation of adding elements of $A$ and then rounding to the nearest element of $A$. A similar idea will work for squaring, and we will get multiplication from addition and squaring.

**Lemma 10.1.** Let $m, n, q \in \mathbb{Z}_{>0}$. Then

1. $m + n = q$ if and only if $|y_m + y_n - y_q| \leq 3/10$.
2. $m^2 = n$ if and only if $|y_m^2 - y_n| \leq 4/10$.

**Proof.**

1. The quantity $y_m + y_n - y_q$ differs from the integer $m + n - q$ by at most $1/10 + 1/10 + 1/10$.

2. The quantity $y_m^2 - y_n$ differs from the integer $m^2 - n$ by at most

$$|y_m^2 - m^2| + |y_n - n| \leq \left| \left( m + \frac{1}{10m} \right)^2 - m^2 \right| + \frac{1}{10} \leq \frac{4}{10}.$$  

Lemma 10.1 shows that the two predicates $m + n = q$ and $m^2 = n$ on $\mathbb{Z}_{>0}$ correspond to diophantine predicates on $A$. Building with these, we can show the same for $mn = q$, since

$$mn = q \iff (m + n)^2 = m^2 + n^2 + q + q.$$  

This completes the proof of part (2) of Theorem 1.3. Part (3) follows from (2) and Matijasević’s Theorem.

**Acknowledgements**

I thank Thanases Pheidas and the referee for helpful comments, in particular for simplifying the diophantine definition of multiplication at the end. I thank also Gunther Cornelissen and Graham Everest for suggesting references for Section 3 and Alexandra Shlapentokh for suggesting some improvements in the exposition.

**References**

