Elliptic curves, Logic, and Diophantine stability

B. Mazur

June 20, 2015

Contents

Part I

The Program for these lectures

1 Around Hilbert's Tenth Problem

Some years ago, Karl Rubin and I worked on a problem in the arithmetic of elliptic curves that was needed to answer a general question in logic: given an infinite, but finitely generated, commutative ring A is there an algorithm to determine—in finite time—whether a polynomial in finitely many variables with coefficients in A has a solution or not. We didn't answer that question unconditionally, but rather assumed a standard conjecture in the arithmetic of elliptic curves, and proved (dependent, of course, on an immense amount of prior work—classical work of Julia Robinson/Davis/Putnam/ Matiyasevich as well as recent work of Poonen and Eisentrager) that: no, there is no such algorithm for any finitely generated commutative ring A (of infinite cardinality).

This type of work, of course, has its origin in Hilbert's classical Tenth Problem:

> Given a diophantine equation with any number of unknown quantities and with rational integral numerical coefficients: To devise a process according to which it can be determined in a finite number of operations whether the equation is solvable in rational integers.

We tend to interpret Hilbert's problem broadly in terms of algorithmic processes, and there's some consensus about what that means. The simple answer to Hilbert's question (for $\mathbf{Z}$) is "no," stemming from "classical work" (Julia Robinson/Davis/Putnam/ and Matiyasevich). To say that there is no such algorithm is in no way a completely *negative* statement, given the format of Matiyasevich's proof. For it banks on

- the known fact that there are subsets S of the integers that are undecidable in the simple sense that although there may be an algorithm to list the elements of S (S would then be called **computably enumerable** or **listable**) there is no algorithm to list the elements of the complement of S in $\mathbf{Z}$, so we don't have a way of computing whether or not a given integer is in S; and
- being able to define any such S by a diophantine method.

That is, diophantine formulations capture all listable sets.

So—conditionally on a standard conjecture in the arithmetic of elliptic curves—the theorem that Karl Rubin and I contributed to would then say, for example, that the class of diophantine problems over any ring that is infinite, and finitely generated, is as rich as, for example, the Halting Problem of Alan Turing.

The pressing question is the analogue of Hilbert's Problem for the field $\mathbf{Q}$ and more generally for subfields in $\bar{\mathbf{Q}}$. For example, one can ask the question *from the top*, i.e., for subfields of $\bar{\mathbf{Q}}$ which are the fixed field of a given automorphism of $\bar{\mathbf{Q}}$. The notable success here is for the subfield of real algebraic numbers, i.e., the fixed field of complex conjugation—this field being first-order decidable, by a theorem of Fried, Haran and Völklein.

Julia Robinson had shown that there is a "first-order definition" of $\mathbf{Z}$ in $\mathbf{Q}$ and using Matiyasevich's result one can conclude that there is NO algorithm to decide the truth or falsity of first-order sentences in $\mathbf{Q}$. It is still an open question whether $\mathbf{Z}$ can be defined by diophantine means in $\mathbf{Q}$.

Bjorn proved the following relatively short first order definition of $\mathbf{Z}$ in $\mathbf{Q}$. A rational number t is an integer if for all pairs of rational numbers a, b there are seven rational numbers $x_1, x_2, \ldots, x_7$ such that

$$(a+\sum_{i=1}^{4} x_i^2)\cdot(b+\sum_{i=1}^{4} x_i^2)\cdot(x_1^2-ax_2^2-bx_3^2+abx_4^2-1)^2 = -\prod_{n=0}^{2309}((n-t-2x_1)^2-4ax_5^2-4bx_6^2 = 4abx_7^2-4)^2.$$

2 Diophantine Stability Issues

To prove unconditionally that Hilbert's diophantine problem has a negative answer for any finitely generated commutative ring A (of infinite cardinality) one would like to use the existence of elliptic curves containing rational points of infinite order over a given number field K having a certain 'stability' property as one passes from the field K to larger number fields.

Karl Rubin and I showed that a standard conjecture regarding 2-Selmer groups will imply the desired existence described above. More generally, call the stability property that enters in this discussion "diophantine-stability:" Let L/K be a field extension, and

$$P(X_1, X_2, \ldots, X_n)$$

a polynomial with coefficients in K (or more generally a system of such polynomials). Say that the polynomial P is 'diophantine-stable' for the extension L/K if P acquires no *new* zeroes over L; i.e., if for $a_1, a_2, \ldots, a_n \in L$ we have $P(a_1, a_2, \ldots, a_n) = 0$ then the elements $a_1, a_2, \ldots, a_n$ are all in K. The property having been so important in one context, it is natural to look at it in broader terms.

> *Fix* a variety V over, say, a number field K. Is there a nontrivial field extension L/K for which it is diophantine-stable?

If there is a curve in V that is isomorphic to an open subvariety of the projective line over K the answer is clearly no. Is this the only obstruction to a positive answer to the above question? I.e.,

Question: If V is a variety over a number field K such that for every nontrivial extension L/K, there are new points, i.e., $V(L) - V(K))$ is not empty, does V contain a subvariety over K that is isomorphic to an open subvariety of the projective line over K?

Karl Rubin and I (with help from Michael Larson) have recently proved that the answer to this question is yes, if V is a curve. To formulate this more precisely,

Let $K \subset \bar{\mathbf{Q}}$ be a number field and V an irreducible algebraic variety over K.

Definition 1. *A field extension L/K is* **generated by a point of V over K** *if (it "is"; i.e., if) equivalently:*

- *L is generated over K by the coordinates of the image of some point of an affine open subvariety of V when embedded in some affine space $\mathbf{A}^N$, the embedding being defined over K.*
- *L/K is an extension of V such that there is a point $x \in V(L)$ which is not contained in the subset $V(L') \subset V(L)$ for any proper sub-extension L'/K.*
- *$L = K(x)$ for some point $x \in V(\bar{\mathbf{Q}})$.*

If V is a variety over K we will sometimes say that 'L/K belongs to V' over K if it is generated by a point of V over K. Denote by $\mathcal{L}(V; K)$ the set of field extensions of K belonging to V. That is:

$$\mathcal{L}(V; K) := \{K(x)/K \ ; \text{for } x \in V(\bar{\mathbf{Q}})\}.$$

A vertical companion to the classical Hilbert Tenth problem (and its various variants) might be to simply *fix* one variety V and ask decidability questions about the collection of field extensions $\mathcal{L}(V; K)$.

For example, if V contains a (nonempty) affine open subvariety of the projective line $\mathbf{P}^1$ over K, then $\mathcal{L}(V;K)$ consists of *all* number field extensions of K. It seems natural to us to conjecture the converse. We prove this conjecture for irreducible varieties of dimension one. Specifically:

Theorem 1. *Let V be an irreducible projective curve over the number field K. Then every field extension L/K belongs to V if and only if V is birationally isomorphic (over K) to the projective line.*

Moreover, we show that for any curve of positive genus there are many extension fields do not belong to it:

Theorem 2. *Let X be an irreducible curve over a number field K whose normalization and completion is not of genus 0. Then there is a finite extension K'/K such taht for any positive integer n, there are infinitely many primes ℓ where, for each of them, there are infinitely many cyclic extension fields L/K' of degree ℓ^n such that $X(K') = X(L)$.*

We show this by relating curves to abelian varieties, via their jacobians.

Here are some natural related questions:

Question 1. *Let X and Y be two irreducible smooth projective curves over a number field $K \subset \bar{\mathbf{Q}}$. If $\mathcal{L}(X;K') = \mathcal{L}(Y;K')$ for all number field extensions K'/K, is it true that X and Y are isomorphic over $\bar{\mathbf{Q}}$?*

If one restricts to curves X, Y with X of genus zero, if $\mathcal{L}(X;K) = \mathcal{L}(Y;K)$ then $X \cong Y$ over K. (This is easy to see.)

Also, it is tempting to think that for a fixed cyclic field extensions of large prime degree diophantine stability is not so rare. E.g., we might wonder:

Question 2. *Fixing a curve X over a number field K, is it the case that for any prime degree $\ell >>_{X,K} 0$ there is a significantly large quantity (e.g., a positive density) of cyclic degree ℓ extensions L/K for which X acquires no new rational point?*

3 The rarity of rational points, and some comments about elliptic curves over Q

An affirmative answer to the question just asked conforms to a general sense that—all in all–rational points are rare and when they come in profusion they do so for some eventually graspable reason, and not because they happen. It is a 'minimalist view.'

I'll be discussing this in these lectures, reviewing aspects of the logical vocabulary we used, the basics of elliptic curves, and the various tools for examining aspects of the problem; specifically Selmer groups over arbitrary number fields.

Here are a few comments about Question 2 connected to elliptic curves over $\mathbf{Q}$ (but I'll give a general intro to elliptic curves in a later lecture). Examples of elliptic curves over $\mathbf{Q}$ non-diophantine-stable for cyclic extensions of $\mathbf{Q}$ of order ℓ (even for relatively small primes ℓ) seem to be quite rare over $\mathbf{Q}$. Chantal David, Jack Fearnley and Hershy Kisilevsky [**?**] conjecture that for a fixed elliptic curve over $\mathbf{Q}$ and $\ell \geq 7$, there are only finitely many such extensions. For $\ell = 3$ and 5, following random matrix heuristics, they make these conjectures: if

$$N_{E,\ell}(x) := |\{\chi \text{ of order } \ell \mid \text{cond}(\chi) \leq x \text{ and } L(E,\chi,1) = 0\}|$$

they conjecture that:

$$\begin{aligned} \log N_{E,3}(x) &\sim \frac{1}{2}\log(x), \\ \log N_{E,5}(x) &<<_\epsilon x^\epsilon. \end{aligned}$$

They exhibit one example with $\ell = 11$, namely, the elliptic curve $E := 5906B1$ (using Cremona's classification)[1].

The way David, Fearnley and Kisilevsky proceed is by studying the nonvanishing of values at $s = 1$ of L-functions $L(E,\chi,1)$ for characters χ of order $\ell > 2$ and of varying conduction $N = N_\chi$. This latter question is equivalent to the more combinatorial-seeming question of nonvanishing of:

$$\sum_{a=1}^{N} \chi(a) \cdot [\frac{a}{N}]_E,$$

i.e., the weighted sums

of the *"real" modular symbol*[2] *attached to* E

$$\frac{a}{N} \mapsto [\frac{a}{N}]_E$$

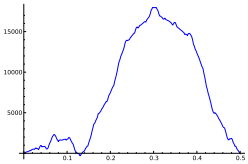
which sends $\mathbf{Q}/\mathbf{Z}$ to rational numbers of bounded denominator.

To study the distribution of these modular symbols, Karl Rubin, William Stein and I have made some computations that I want to describe. Let E be an elliptic curve over $\mathbf{Q}$ with L-function $L(e,s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}$ and with $[\frac{a}{N}]_E$ the "real" modular symbol attached to E. Fixing $N = p$ a large prime, form the function for $0 \leq \tau \leq 1/2$,

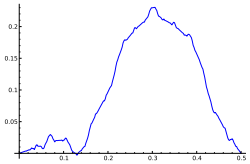
$$G_{E,p}(\tau) = \sum_{0 \leq \frac{a}{p} \leq \tau} [\frac{a}{p}]_E.$$

[1] They show that $L(E,\chi,1) = 0$ where χ is a character of order 11 and of conductor 23 (i.e., χ has the smallest possible conductor for characters of its order). This implies that one has diophantine instability for the cyclic field extension L/K of degree 11 cut out by χ.

[2] The L-functions for an elliptic curve over $\mathbf{Q}$ and all of its twists does many things at once: it records local data, as we'll mention below; but also for the family of twists of E its values at $s = 1$ are given in terms of integrals of a differential form, and therefore are expressible in terms of periods, of a particular sort, of the elliptic curve, and when appropriately normalized, these are the "modular symbols,"' a combinatorial tool computable by a variant of a continued fraction process–hence very quickly.

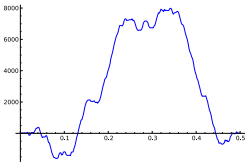


(a) $G_{E,p}(\tau)$

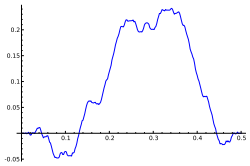


(b) $g_E(\tau)$

Figure 1: $E = 11a$; $p = 100,003$



(a) $G_{E,p}(\tau)$



(b) $g_E(\tau)$

Figure 2: $E = 37a$; $p = 100,019$

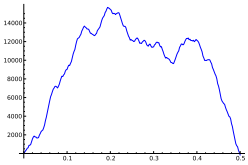
It is natural to try to compare this with the (convergent) function:

$$g_E(\tau) \;=\; \frac{1}{2\pi i}\sum_{n=1}^{\infty}\frac{a_n}{n^2}(\sin(2\pi i n\tau)/2\pi$$

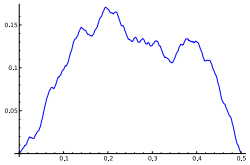
Specifically:

Conjecture 3.

$$\lim_{p\to\infty} G_{E,p}(\tau) \;\stackrel{??}{=}\; g_E(\tau).$$



(a) $G_{E,p}(\tau)$



(b) $g_E(\tau)$

Figure 3: $E = 37b$; $p = 100,043$

Part II

Logic and Number Theory

4 Listable sets of integers

Nowadays one has a large number of different *processes* in our experience (i.e., *successes*). From algorithms to find the maxima of functions on convex polytopes (e.g.: Linear programming) to procedures for factoring numbers into product of primes. The basic questions we tend to ask about these have to do with running time.

We also have quite a number of guaranteed non-successes:

- There is no finite algorithm to determine, given a finite presentation of a group, whether or not the group is trivial. Or whether two finite presentations present isomorphic groups.
- The *recognition problem for manifolds in dimension four or higher* is unsolvable (it being related directly to the recognition problem for finitely presented groups).

And even when one looks for interesting Diophantine examples, they often come in formats somewhat different from the way Hilbert's Problem is posed. For example,

- we have a (deep) decision procedure to determine whether any given elliptic curve over the rational field $\mathbf{Q}$ has *finitely many* or *infinitely many* solutions. But this distinction

  $$\textit{finitely many} \leftrightarrow \textit{infinitely many}$$

  is not a distinction that Hilbert formulates.
- And, sometimes, we're interested not in answering this question for any single elliptic curve but, for whole families of them. For example, the *congruent number problem* is the problem of determining those positive integers n that can be expressed as the area of a right triangle with three rational number sides. This turns out to be *equivalent* to asking that the elliptic curve

  $$y^2 = x^3 - n^2 x$$

  have infinitely many rational points.

- And, we sometimes try to find single processes that work even allowing for variation of the exponents involved.

  As in:

  1. *Catalan-type Problems*

     For a given integer k find all *perfect powers* that differ by k.

     $$Y^n - X^m \quad = \quad k$$

     Example: *the only two consecutive perfect powers are:*

     $$8 = 2^3 \text{ and } 9 = 3^2,$$

     or as in:

  2. *Fermat's Last Theorem.*

So you might ask why—except for historical reasons—might one be interested in pursuing the question as Hilbert posed it. The answer (which is already enough to spark my interest) is that it is a problem that has led to the most magnificent developments in mathematical logic, and in the intersection of mathematical logic and number theory. But also, thanks to relatively recent work (of Denef, Denef-Lipschitz, Pheidas, Shalapentokh and Poonen) Hilbert's Problem calls for the answers to new *kinds of* questions in number theory, and specifically in the arithmetic of elliptic curves.

So, back to Hilbert's Tenth Problem!

Hilbert is particular in the type of solutions (rational integers) he seeks. Nevertheless, in considering "Hilbert's 10th Problem" we often specifically interpret *Diophantine equation, process* and sometimes generalize the type of solutions being considered. We then end up with a question roughly of the following form:

Let A be a commutative ring. Does there exist a finite algorithm to determine whether any finite

system of polynomial equations in finitely many variables with coefficients in A has a solution in A or not?

INPUT: A finite collection of polynomial equations

$$f_i(X_1, X_2, X_3, \ldots, X_n)$$

with integer coefficients.

OUTPUT: "Yes," or "No," answering the question of whether or not there is an n-tuple of integers $(a_1, a_2, a_3, \ldots, a_n)$ such that

$$f_i(a_1, a_2, a_3, \ldots, a_n) = 0$$

for all i.

One standard way of refining the above question is to "reset" it as a problem related to *listable* sets and *Diophantine sets*.

Listable sets of integers

(synonyms: *recursively enumerable, computably enumerable*)

I'll start with some examples of sets that are easy to "list"

- $$2, 3, 5, 7, 11, 13, 17, 19, 23, \ldots$$
- $$2!, 3!, 4!, 5!, \ldots$$

Discuss what is meant by *easy*

Generally, a subset $\mathcal{L} \subset \mathbf{Z}$ is called **listable** if there exists a finite computer program whose output gives a sequence $\alpha_1, \alpha_2, \alpha_3, \ldots$ of integers such that the set $\mathcal{L}$ is precisely this collection of numbers; i.e.,

$$\mathcal{L} = \{\alpha_1, \alpha_2, \alpha_3 \ldots\}.$$

A computer algorithm that does job this will be called a computer algorithm that "lists $\mathcal{L}$."

Note, though, that–even if the computer spits out a "new" integer every second— the ordering in which the integers in $\mathcal{L}$ come via the computer's list may be helter-skelter in terms of absolute values. Therefore if you suspect that a given number, say 2, is *not* in $\mathcal{L}$ and need to have a definite guarantee of the truth of your suspicion, well (... if you are right!) running the helter-skelter computer algorithm for any finite length of time will be of no help to you.

- A more useful finite computer program might be, for example, a program that for each integer N will, after some guaranteed time (e.g., no greater than $N^{N^{\cdots^N}}$ hours) [3]) actually produces a *complete* list of *all* integers of absolute value $\leq N$ that are in $\mathcal{L}$. (Call such a program a *deluxe program*.)
- Somewhat intermediary to the above two types of computer programs (helter-skelter, and deluxe) would be a *pair* of computer programs, one of which spits out the elements of $\mathcal{L}$ and the other spits of the elements of the complement of $\mathcal{L}$. Supplied with such a pair of programs you might, at the very least, run the first program by day, and the second by night, for then you are guaranteed to know—in some (perhaps unspecified, but) finite time whether or not 2 is in your set $\mathcal{L}$.

The Halting Problem

A set $\mathcal{L}$ that has the property that it and its complement are both listable is called *recursive*.

If you have such a recursive set, then, as mentioned—listing the set $\mathcal{L}$ by day and its complement $\mathbf{N} - \mathcal{L}$ by night—you are guaranteed that for every $N \in \mathbf{N}$ you will know at some finite time whether or not N is in your set.

There exist recursively enumerable sets that are *not* recursive. (The computer algorithms that list such sets are necessarily quite helter-skelter!)

This is a consequence of the famous 1936 theorem of Alan Turing that was phrased in terms of the

[3] N successive exponentials, or choose any recursively formulatable estimate you like

halting problem for algorithms. Turing showed that there exists no universal algorithm to tell you whether or not any finite computer algorithm will terminate finitely, when run.

More specifically, the so-called *halting set*

$$\mathbf{H} := \{\text{The set of couples } (P, x)$$

where P is a program and

x is a possible input to program P and

such that *Program* P will eventually *halt*

if run with input $x\}$

is recursively enumerable—once you code, in a computable way, the (P, x)'s as a subset of natural numbers—BUT the *complement* of this set is not recursively enumerable.

Diophantine sets

Roughly, a *Diophantine subset* of integers (or of natural numbers) is a subset that can be defined using the *seemingly very restricted* vocabulary of polynomials.

Here is one way of formulating this concept over a fairly general ring.

Let A be a commutative noetherian integral domain, the main example being $A = \mathbf{Z}$.

Definition: Let $\mathcal{D} \subset A$ be a subset of the ring A.
Say that $\mathcal{D}$ is **Diophantine in** A if there exists a finite set of polynomials with coefficients in A, in finitely many variables

$$f_i(T; X_1, X_2, \ldots, X_n) \in A[T, X_1, X_2, \ldots, X_n]$$
$$(i = 1, 2, \ldots, m)$$

such that for $\alpha \in A$ the system of polynomial equations

$$f_i(\alpha; X_1, X_2, \ldots, X_n) = 0$$

has a simultaneous solution

$$(X_1, X_2, \ldots, X_n) = (a_1, a_2, \ldots, a_n) \in A^n$$

if and only if

$$\alpha \in \mathcal{D} \subset A.$$

If this happens say that the set of polynomials **cut out** $\mathcal{D}$.

Notice the evident proposition:

Proposition: If $A = \mathbf{Z}$ (or, more, generally, a countable ring) and $\mathcal{D} \subset A$ is Diophantine, then $\mathcal{D}$ is listable.

Moreover, any set of polynomials

$$\{f_i(\alpha; X_1, X_2, \ldots, X_n)\}_i$$

(for $i = 1, 2, \ldots, m$) that "cut out" $\mathcal{L}$ leads to a computer algorithm that lists $\mathcal{L}$.

Remarks:

(1) The collection of Diophantine subsets of an integral domain A is closed under finite union and intersection.

Proof: It suffices to do this for two Diophantine sets $D, E \subset A$:

Let the systems of polynomials

$$\{f_i(t; X_1, \dots)\}_i$$

and

$$\{g_j(t; Y_1, \dots)\}_j$$

cut out D and E respectively.

- The "union" of the two systems, (viewed as polynomials in t and the independent variables X_μ and Y_ν) cuts out $D \cap E$.

- The "product"system given by

  $$\{f_i(t; X_1, \dots) \cdot g_j(t; Y_1, \dots)\}_{i,j}$$

  cuts out $D \cup E$.

Diophantine sets are closed, as well, under polynomial mappings.

Mention: scheme-theoretic definition

(3) For us the most important ring is $A = \mathbf{Z}$. In this context you can replace any finite system of polynomials $\{f_i(t; X_1, \dots)\}_i$ that "cut out" a set D by a single polynomial

$$\Sigma_i f_i(t; X_1, \dots)^2.$$

One is now faced with the task of building a Diophantine vocabulary.

Here is a list of subsets of $\mathbf{Z}$ that are Diophantine (and easily proven to be).

1. Lagrange's Theorem says that any positive whole number is expressible as a sum of four squares.

E.g

$$4001 = (20)^2 + 1^2 + 0^2 + 0^2$$

well ... that might have been too easy an example ...

In our Diophantine vocabulary, this means that the polynomial

$$f(t; X_1, X_2, X_3, X_4) := \quad t - \Sigma_{i=1}^{4} X_j^2$$

cuts out the set of positive integers; so the set of positive numbers is Diophantine.

2. Therefore it follows, by easy steps, that these sets are too:

   - the set of numbers $\geq a$ for any given $a \in \mathbf{Z}$,
   - the set of numbers $\leq b$ for any given $b \in \mathbf{Z}$,
   - any finite subset of $\mathbf{Z}$,
   - the complement of any finite subset of $\mathbf{Z}$.

3. So, if D is Diophantine, then any set obtained from D by removing and adding finite sets is also Diophantine.

4. Arithmetic progressions are Diophantine; as are the set of all squares, all cubes, all n-th powers for any given n.

5. Composite numbers.

6. For any fixed (say, nonsquare, positive) integer d, consider the set of integers t that come in solutions of the Pell equation

   $$t^2 - ds^2 = 1$$

   (this being a set that grows roughly exponentially).

The evolution of Hilbert's problem as developed through the work of

Martin Davis
Julia Robinson
Hillary Putnam
Yuri Matiyasevich:

The culminating theorem is due to Matiyasevich:

Theorem Every listable subset of $\mathbf{Z}$ is Diophantine.

Thus *listable* and *Diophantine* are equivalent conditions for subsets of $\mathbf{Z}$.

Since there exist listable subsets of $\mathbf{Z}$ that are not recursive—i.e., such that their complements are *not listable*, Matiyasevich's Theorem gives a negative answer to Hilbert's question above, but does far more than just that.

For example:

1. It certainly shows that there are systems of polynomials over $\mathbf{Z}$ that admit no "deluxe computer program" as described.

2. The result also implies that relatively benign subsets of $\mathbf{Z}$ can be Diophantinely described, as well. This is not as clear as one might think even for the most familiar subsets. For example:

   - There is a system of polynomials that cut out the set of factorials $1!, 2!, 3! \ldots$ The fact that this set is Diophantine played a big role in the development of the subject.

     By the way, to get such a polynomial one starts by finding a Diophantine way of expressing the binomial coefficients $\binom{n}{m}$ and then dealing with the—to me surprisingly unpromising—formula

     $$m! = \lim_{n\to\infty} \frac{n^m}{\binom{n}{m}}.$$

     (!!!)

     The *factorial* operation has quite a powerful effect if one allows it to be used as a piece of equipment to generate recursively enumerable sets. For example, the set of numbers $\alpha > 1$ such that the expression

     $$\alpha \cdot X_1 + (\alpha - 1)! \cdot X_2 = 1$$

     has a zero for integers X_1, X_2 is precisely the set of prime numbers. But regarding prime numbers, more relevant for our story is the fact that...

   - There is a polynomial over $\mathbf{Z}$ whose set of positive values is the set of *exactly all* prime numbers for integral substitution of its variables. A specific such polynomial is given by **[JSWW76]**:

$$(k+2)\{1-[wz+h+jq]^2-[(gk+2g+k+1)(h+j)+hz]^2-[2n+p+q+ze]^2[16(k+$$

$1)^3(k+2)(n+1)^2+1f^2]^2-[e^3(e+2)(a+1)^2+1o^2]^2-[(a^21)y^2+1-x^2]^2x-[16r^2y^4(a^2-1)+1-u^2]^2-[((a+u^2(u^2-a))^2-1)(n+4dy)^2+1-(x+cu)^2]^2-[n+l+vy]^2-[(a^2-1)l^2+1-m^2]^2-[ai+k+1-l-i]^2-[p+l(a-n-1)+b(2an+2a-n^2-2n-2)-m]^2-[q+y(a-p-1)+s(2ap+2a-p^2-2p-2)-x]^2-[z+pl(a-p)+t(2ap-p^2-1)-pm]^2\}$

Comments on the History

This work ranges frrom 1944 when Emil Post said that Hilbert's tenth problem "begs for an unsolvability proof" to 1970 when Matijasevic clinched the theorem.

But I'll begin in 1960, when Julia Robinson, sharpening work of Martin Davis, and Hillary Putnam, showed that if there exists a *roughly exponential function* defined in a diophantine way; i.e., a Diophantine set $\mathcal{F}$ of couples (a, b) in $\mathbf{N} \times \mathbf{N}$ with two properties:

(a) If $(a, b) \in \mathcal{F}$ then $a < b^b$.

(b) For each positive integer k there is an $(a, b) \in \mathcal{D}$ with $b > a^k$.

then all listable sets would be Diophantine.

In 1970, Matiyasevich provides a Diophantine definition of a set $\mathcal{F}$ as required by J.R.: he defined his $\mathcal{F}$ to be the collection of pairs (a, b) such that

$$b = F_{2a}$$

where F_n is the nth Fibonacci number, thereby completing the proof that all recursively enumerable sets are Diophantine and establishing the fact that Hilbert's tenth problem (over $\mathbf{Z}$) is unsolvable.

(I find this quotation of Matiyasevich illuminating:)

"The idea was as follows. A universal computer science tool for representing information uses words rather than numbers. However, there are many ways to represent words by numbers. One such method is naturally related to Diophantine equations. Namely, it is not difficult to show that every 2×2 matrix

$$\begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}$$

with the m's being non-negative integers and the determinant $m_{11}m_{22} - m_{12}m_{21}$ equal to 1 can be represented, in a unique way, as a product of matrices

$$M_0 := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

and

$$M_1 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

It is evident that any product of such matrices has non-negative integer elements and the determinant equals 1. This implies that we can uniquely represent a word in the two-letter alphabet M_0, M_1 by the four-tuple

$$(m_{11}, m_{12}, m_{21}, m_{22})$$

such that the numbers evidently satisfy the Diophantine equation

$$m_{11}m_{22} - m_{12}m_{21} \ = \ 1.$$

Under this representation of words by matrices, the operation of concatenation-of-words corresponds to matrix multiplication and thus can be easily expressed as a system of Diophantine equations, opening up a way of transforming an arbitrary system of word equations into "equivalent" Diophantine equations. Many decision problems about words had been shown undecidable, so it was quite natural to try to attack Hilbert's tenth problem by proving the undecidability of systems of word equations."

...

SEE BELOW for material. Start with some examples of sets that are easy to "list"

- $$2, 3, 5, 7, 11, 13, 17, 19, 23, \ldots$$

- $$2!, 3!, 4!, 5!, \ldots$$

See [M] and [Sh100b] for good expository accounts of this notion *listable*. Naively,

- a *listable* subset of $\mathbf{Z}$ (synonyms: *recursively enumerable, computably enumerable*) is a subset $\mathcal{L} \subset \mathbf{Z}$ for which there exists a finite computer program whose output gives a sequence $\alpha_1, \alpha_2, \alpha_3, \ldots$ of integers such that the set $\mathcal{L}$ is precisely this collection of numbers; i.e.,

  $$\mathcal{L} = \{\alpha_1, \alpha_2, \alpha_3 \ldots\}.$$

  A computer algorithm that does job this will be called a computer algorithm that "lists $\mathcal{L}$."

  Note, though, that the ordering in which the integers in $\mathcal{L}$ come via the computer's list may be helter-skelter in terms of absolute values. Therefore if you suspect that a given number, say 2, is *not* in $\mathcal{L}$ and need to have a definite guarantee of the truth of your suspicion, well (... if you are right!) running the helter-skelter computer algorithm for any finite length of time will be of no help to you.

- A more useful finite computer program might be, for example, a program that for each integer N will, after some guaranteed time (e.g., no greater than $N^{N^{\cdots^N}}$ hours[4]) actually produces a *complete* list of *all* integers of absolute value $\leq N$ that are in $\mathcal{L}$. (Call such a program a *deluxe program*.)

[4] N successive exponentials, or choose any recursively formulatable estimate you like

- Somewhat intermediary to the above two types of computer programs (helter-skelter, and deluxe) would be a *pair* of computer programs, one of which spits out the elements of $\mathcal{L}$ and the other spits of the elements of the complement of $\mathcal{L}$. Supplied with such a pair of programs you might, at the very least, run the first program by day, and the second by night, for then you are guaranteed to know—in some (perhaps unspecified, but) finite time whether or not 2 is in your set $\mathcal{L}$.

5 The Halting Problem

As mentioned, a set $\mathcal{L}$ that is listable by a finite computer algorithm will be referred to as *listable* or *recursively enumerable*. And a set $\mathcal{L}$ that has the property that it and its complement are both listable is called *recursive*. What we will be using below is the fact (cf. [Sm]) that there exist recursively enumerable sets that are *not* recursive. (The computer algorithms that list such sets are necessarily quite helter-skelter!) The existence of recursively enumerable sets that are not recursive is a consequence of the famous 1936 theorem of Alan Turing that was phrased in terms of the *halting problem for algorithms*. Turing showed that there exists no universal algorithm to tell you whether or not any finite computer algorithm will terminate finitely, when run. More specifically, the so-called *halting set*

$$\mathbf{H} := \{\text{The set of couples } (P, x) \text{ where } P \text{ is a program and } x \text{ is a possible input to program } P \text{ and such that } \textit{Program } P \text{ will eventually } \textit{halt} \text{ if run with input } x\}$$

is recursively enumerable, (i.e., there is fairly evidently a computable function that lists all of the pairs (P, x) it contains) but the complement of this set is not recursively enumerable.

6 Diophantine sets

Roughly, a *Diophantine subset* of integers (or of natural numbers) is a subset that can be defined using the *seemingly very restricted* vocabulary of polynomials. For the classical notion of Diophantine subset of a commutative ring see **[DL78], [Den80]**. Here is one way of formulating this concept. Let A be a commutative noetherian integral domain, the main example being $A = \mathbf{Z}$.

Definition 2. *Let $\mathcal{D} \subset A$ be a subset of the ring A.*
Say that $\mathcal{D}$ is **Diophantine in** *A if there exists a finite set of polynomials with coefficients in A, in finitely many variables*

$$f_i(T; X_1, X_2, \ldots, X_n) \in A[T, X_1, X_2, \ldots, X_n]$$

$(i = 1, 2, \ldots, m)$ such that when specializing to some value $T = \alpha \in A$ we have that the system of polynomial equations

$$f_i(\alpha; X_1, X_2, \ldots, X_n) = 0$$

(for $i = 1, 2, \ldots, m$) has a simultaneous solution

$$(X_1, X_2, \ldots, X_n) = (a_1, a_2, \ldots, a_n) \in A^n$$

if and only if

$$\alpha \in \mathcal{D} \subset A.$$

If this happens say that the set of polynomials **cut out** $\mathcal{D}$.

Notice the evident proposition:

Proposition 1. *If $A = \mathbf{Z}$ (or, more, generally, a countable ring) and $\mathcal{D} \subset A$ is Diophantine, then $\mathcal{D}$ is listable. Moreover, any set of polynomials $\{f_i(\alpha; X_1, X_2, \ldots, X_n)\}_i$ (for $i = 1, 2, \ldots, m$) that "cut out" $\mathcal{L}$ leads to a computer algorithm that lists $\mathcal{L}$.*

Proof: Choose some ordering of A^{n+1} (e.g., lexicographical based on an ordering of A; if $A = \mathbf{Z}$ my preference is for the evident ordering of $\mathbf{Z}$: that is, $-n$ precedes $+n$ and otherwise it is nondecreasing in the absolute value of n) and run through the $n+1$-tuples $(\alpha; a_1, a_2, \ldots, a_n) \in A^{n+1}$ computing $f_i(\alpha; a_1, a_2, \ldots, a_n)$ for $i = 1, 2, \ldots$: every time you get a *hit*—i.e., every time that $f_i(\alpha; a_1, a_2, \ldots, a_n) = 0$ for $i = 1, 2, \ldots$ you record the "α" if it hasn't been previously recorded giving a (possibly empty, of course) sequence $\alpha_1, \alpha_2, \ldots$ listing $\mathcal{D}$.

Remarks: (1) The collection of Diophantine subsets of an integral domain A is closed under finite union and intersection.

Proof: It suffices to do this for two Diophantine sets $D, E \subset A$: if the systems of polynomials $\{f_i(t; X_1, \ldots)\}_i$ and $\{g_j(t; Y_1, \ldots)\}_j$ cut out D and E respectively, then the "union" of the two systems, (viewed as polynomials in t and the independent variables X_μ and Y_ν) cuts out $D \cap E$ while the system given by $\{f_i(t; X_1, \ldots) \cdot g_j(t; Y_1, \ldots)\}_{i,j}$ cuts out $DS \cup E$.

(2) A more general (and, perhaps, algebro-geometrically more natural) way of thinking of Diophantine set is the following:

Let S be an integral noetherian scheme—say an affine scheme $S = \mathrm{Spec}(A)$ where A is a noetherian integral domain— and T an S-scheme of finite type. Let $\mathcal{T} = T(S)$ the set of S-valued points of the S-scheme T. A subset $\mathcal{D} \subset \mathcal{T}$ is **Diophantine** if there is a morphism of S-schemes of finite type $f : X \to T$ such that

$$\mathcal{D} = f(X(S)) \subset T^(S) = \mathcal{T}.$$

To relate the above to the previous definition let $S = \mathrm{Spec}(A)$ and let $T = \mathrm{Spec}(A[t])$ denote the affine line over $\mathrm{Spec}(A)$. So the set $\mathcal{T}$ of A-rational points of T, i.e.,

$$\mathcal{T} = T(A) = \mathrm{Hom}_A(A[t], A).$$

is simply the set A. Diophantine subsets of the ring A are nothing more than the images of the sets of A-rational points,

$$X(A) \longrightarrow T(A) = A,$$

where $X \to T$ range through all morphisms of finite type of A-schemes (of finite type) X.

A vague, but general question, then for any scheme T of finite type over such a base S would be:

To give a useful algorithmic characterization of the subsets $\mathcal{D} \subset \mathcal{T}$ that are Diophantine.

(3) For us the most important ring is $A = \mathbf{Z}$, and scheme T is the affine line. In this context you can replace any finite system of polynomials $\{f_i(t; X_1, \dots)\}_i$ that "cut out" a set D by a single polynomial

$$\Sigma_i f_i(t; X_1, \dots)^2.$$

Here is a list of subsets of $\mathbf{Z}$ that are Diophantine (and easily proven to be).

1. Lagrange proved that any positive whole number is expressible as a sum of four squares.

   E.g
   $$401 = (20)^2 + 1^2 + 0^2 + 0^2$$

   well ... that might have been too easy an example ...

   Lagrange's Theorem says, in our vocabulary, that the polynomial

   $$f(t; X_1, X_2, X_3, X_4) := \quad t - \Sigma_{i=1}^4 X_j^2$$

   cuts out the set of positive integers; so the set of positive numbers is Diophantine.

2. Therefore it follows, by easy steps, that these sets are too:

   - the set of numbers $\geq a$ for any given $a \in \mathbf{Z}$,
   - the set of numbers $\leq b$ for any given $b \in \mathbf{Z}$,
   - any finite subset of $\mathbf{Z}$,
   - the complement of any finite subset of $\mathbf{Z}$.

3. So, if D is Diophantine, then any set obtained from D by removing and adding finite sets is also Diophantine.

4. Arithmetic progressions are Diophantine; as are the set of all squares, all cubes, all n-th powers for any given n.

5. Composite numbers.

6. For any fixed (say, nonsquare, positive) integer d, consider the set of integers t that come in solutions of the Pell equation
$$t^2 - ds^2 = 1$$
(this being a set that grows roughly exponentially).

7 Davis/(Julia) Robinson/Putnam/ Matiyasevich

Here, in a nutshell, is the general status of this question we inherited from Hilbert and from "classical work" of Martin Davis, Julia Robinson, Hillary Putnam and Yuri Matiyasevich. The culminating theorem is Matiyasevich's:

Theorem 4. *Every listable subset of* $\mathbf{Z}$ *is Diophantine.*

Thus *listable* and *Diophantine* are equivalent conditions for subsets of $\mathbf{Z}$. Since there exist listable subsets of $\mathbf{Z}$ that are not recursive—i.., such that their complements are *not listable*, Theorem 4 gives a negative answer to Hilbert's question above, but does far more than just that.

For example:

1. It certainly shows that there are systems of polynomials over $\mathbf{Z}$ that admit no "deluxe computer program" as described.

2. The result also implies that relatively benign subsets of $\mathbf{Z}$ can be Diophantinely described, as well. This is not as clear as one might think even for the most familiar subsets. For example:

   - There is a system of polynomials that cut out the set of factorials $1!, 2!, 3! \ldots$ The fact that this set is Diophantine played a big role in the development of the subject[5].
     The *factorial* operation has quite a powerful effect if one allows it to be used as a piece of equipment to generate recursively enumerable sets. For example, the set of positive numbers α such that the expression
     $$(\alpha + 1) \cdot X_1 + \alpha! \cdot X_2 = 1$$
     has a zero for integers X_1, X_2 is precisely the set of prime numbers. But regarding prime numbers, more relevant for our story is the fact that...

[5] To get such a polynomial one starts by finding a Diophantine way of expressing the binomial coefficients $\binom{n}{m}$ and then dealing with the—to me surprisingly unpromising—formula
$$m! = \lim_{n\to\infty} \frac{n^m}{\binom{n}{m}}.$$

- There is a polynomial over $\mathbf{Z}$ whose set of positive values is the set of *exactly all* prime numbers for integral substitution of its variables. A specific such polynomial[6] is given in [**JSWW76**]:

  $(k+2)\{1-[wz+h+jq]^2-[(gk+2g+k+1)(h+j)+hz]^2-[2n+p+q+ze]^2[16(k+1)^3(k+2)(n+1)^2+1f^2]^2-[e^3(e+2)(a+1)^2+1o^2]^2-[(a^21)y^2+1-x^2]^2x-[16r^2y^4(a^2-1)+1-u^2]^2-[((a+u^2(u^2-a))^2-1)(n+4dy)^2+1-(x+cu)^2]^2-[n+l+vy]^2-[(a^2-1)l^2+1-m^2]^2-[ai+k+1-l-i]^2-[p+l(a-n-1)+b(2an+2a-n^2-2n-2)-m]^2-[q+y(a-p-1)+s(2ap+2a-p^2-2p-2)-x]^2-[z+pl(a-p)+t(2ap-p^2-1)-pm]^2\}$

8 Comments on the History

For a historical (and basic mathematical) account of this work it would be difficult, I think, to do better than the very informative wikipedia entry on *Hilbert's tenth Problem* which has a chart listing work ranging from 1944 when Emil Post said that Hilbert's tenth problem "begs for an unsolvability proof" to 1970 when Matijasevic clinched the theorem.(We're told that this Wikipedia entry was composed by Martin Davis, so it is no surprise that it is that excellent!)

On the way to the final formulation of the theorem there is Martin Davis's formulation of what we'll, for reference, call **Davis Sets**, these being sets of natural numbers Δ such that there exists a polynomial with integral coefficients

$$P(T, K, Y, X_1, X_2, \ldots, X_n)$$

in independent variables $T, K, Y, X_1, X_2, \ldots, X_n$ for some n, such that $a \in \mathcal{A}$ if and only if there is a nonnegative integer y such that for all nonnegative integers $k < y$ the polynomial $P(a, k, y, X_1, X_2, \ldots, X_n)$ has a solution in natural numbers $X_1 = a_1, X_2 = a_2, \ldots, X_n = a_n$. Now *Davis sets* are fairly clearly recursively enumerable. In 1949 Davis proved the converse: that every recursively enumerable subset of the set of natural numbers has the above form; i.e., is *Davis*.

A year later, working independently, Julia Robinson formulated her hypothesis that asserts that—roughly speaking—there exists some function

$$\text{“}Exp:\text{”}\mathbf{N} \quad \longrightarrow \quad \mathbf{N}$$

that behaves at least vaguely like an exponential function and whose graph is Diophantine (a *sloppy exponential* would be enough). Her hypothesis that came to be known as "J.R." and explicitly is:

Hypothesis J.R.: There exists a Diophantine set $\mathcal{F}$ of couples (a, b) in $\mathbf{N} \times \mathbf{N}$ with two properties:

(a) If $(a, b) \in \mathcal{F}$ then $a < b^b$.

[6] As you'll see from its equation, this is not the most efficacious way of finding prime numbers, but ...

(b) For each positive integer k there is an $(a, b) \in \mathcal{D}$ with $b > a^k$.

Using hypothesis J.R., Robinson shows that the set EXP of triples (a, b, c) with $a = b^c$ is Diophantine, and from this that the set of primes, and the set of factorials is Diophantine as well.

In 1959 Martin Davis and Hillary Putnam showed—assuming that there were arbitrarily long arithmetic progressions of prime numbers—that Hypothesis J.R. implies the equivalence of Diophantine and recursively enumerable, and thereby conditionally establishing a solution to Hilbert's Tenth Problem (the "conditions" being *the existence of arbitrarily long arithmetic progressions of primes,* and *J.R.*).

A year later, Robinson showed how to avoid the use of the hypothesis that arbitrarily long arithmetic progressions of primes exist, thereby showing that J.R. alone implies a solution to Hilbert's Tenth Problem.

In 1970, Matiyasevich provides a Diophantine definition of a set $\mathcal{F}$ as required by J.R.: he established his $\mathcal{F}$ as the collection of pairs (a, b) such that

$$b = F_{2a}$$

where F_n is the nth Fibonacci number, thereby completing the proof that all listable sets are Diophantine and establishing the fact that Hilbert's Tenth Problem (over $\mathbf{Z}$) is unsolvable.

9 Some comments by Matiyasevich

(We find this quotation of Matiyasevich illuminating:)

"The idea was as follows. A universal computer science tool for representing information uses words rather than numbers. However, there are many ways to represent words by numbers. One such method is naturally related to Diophantine equations. Namely, it is not difficult to show that every 2×2 matrix

$$\begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}$$

with the m's being non-negative integers and the determinant $m_{11}m_{22} - m_{12}m_{21}$ equal to 1 can be represented, in a unique way, as a product of matrices

$$M_0 := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

and

$$M_1 := \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

It is evident that any product of such matrices has non-negative integer elements and the determinant equals 1. This implies that we can uniquely represent a word in the two-letter alphabet M_0, M_1 by the four-tuple

$$(m_{11}, m_{12}, m_{21}, m_{22})$$

such that the numbers evidently satisfy the Diophantine equation

$$m_{11}m_{22} - m_{12}m_{21} \;=\; 1.$$

Under this representation of words by matrices, the operation of concatenation-of-words corresponds to matrix multiplication and thus can be easily expressed as a system of Diophantine equations, opening up a way of transforming an arbitrary system of word equations into "equivalent" Diophantine equations. Many decision problems about words had been shown undecidable, so it was quite natural to try to attack Hilbert's tenth problem by proving the undecidability of systems of word equations.

My next attempt was to consider a broader class of word equations with additional predicates. Since the ultimate goal was always Hilbert's tenth problem, I could consider only such predicates, which (under suitable coding) would be represented by Diophantine equations. In this way I came to what I have called *equations in words and length*. Reduction of such equations was based on the celebrated Fibonacci numbers. It is well known that every natural number can be represented, in an almost unique way, as the sum of different Fibonacci numbers, none of which are consecutive[7] (this is the so called *Zeckendorf representation*). Thus we can look at natural numbers as words in a two-letter alphabet $\{0, 1\}$ with the additional constraint that there cannot be two consecutive 1's. I managed to show that under this representation of words by numbers both the concatenation of words and the equality of the length of two words can be expressed by Diophantine equations."

The culminating theorem is Matiyasevich's:

Theorem 5. *Every computably enumerable subset of* $\mathbf{Z}$ *is diophantine (relative to* $\mathbf{Z}$*).*

This fundamental result, of course, gives a negative answer to the question above, but does far more than just that.

For example:

1. The result implies that relatively benign subsets of $\mathbf{Z}$ can be diophantinely described, as well. This is not as clear as one might think even for the most familiar subsets, and seems interesting to me: for example, there is a polynomial over $\mathbf{Z}$ whose set of positive values is the set of *exactly all* prime numbers for integral substitution of its variables. A specific such polynomial (taking hardly two dozen lines of print) is given in [**JSWW76**].

[7] E.g., $30 = 1 + 8 + 21$.

2. One is not yet finished mining this for concrete versions of "unsolvable problems" but it clearly will give us a wealth of such problems. See, for example, recent postings of Harvey Friedman; these have possible relations to Mnëv's (1988) result that any scheme over **Z** can be expressed as a moduli space classifying configurations[8] of finite points in $\P^2$. Harvey Friedman poses nine different "Families of Problems" regarding configurations of rational lines in the Euclidean plane, These problems ask for existence or nonexistence of integral intersections (with various properties) of linear configurations. Friedman discusses whether the problems in each of these families can be done in ZFC or whether there are examples of problems in that family that cannot: apparently three of Friedman's problem-families can be solved in ZFC, three cannot, and for the remaining three—if Hilbert's Tenth Problem (over **Q**) is undecidable—then these cannot be done in ZFC.

More recent work (Denef/Denef-Lipschitz/Pheidas/Shalpentokh/Poonen) developed ideas that culminated in the following result:

Theorem 6. *If a certain stability result in the arithmetic of elliptic curves holds*[9] *over K, then for any number field K every recursively enumerable subset of $\mathcal{O}_K$, the ring of integers in K is diophantine (relative to $\mathcal{O}_K$).*

As mentioned earlier, Karl Rubin and I have recently shown that this stability result holds *if you assume the* 2*-primary part of the classical Shafarevich-Tate Conjecture* [**MR09**]. As a consequence we have shown that, conditional on the 2-primary part of the Shafarevich-Tate Conjecture, Hilbert's Tenth problem has a negative answer for the ring of integers in *any* number field.

Since Kirsten Eisenträger has, in her thesis, related Hilbert's Tenth Problem over rings of integers in number fields to a much more general class of rings, one gets—thanks to her work:

Theorem 7. *Conditional on the* 2*-primary part of the Shafarevich-Tate Conjecture, Hilbert's Tenth problem has a negative answer for any commutative ring A that is of infinite cardinality, and is finitely generated over* **Z**.

[8] By a *configuration type* let us mean a number N and a collection of subsets $S_1, S_2, \ldots S_n$ of the set $[1, 2, \ldots, N]$. The configuration space associated to such a type is the space of all ordered sets of N points in $\P^2$ subject to the requirement that the points corresponding to S_1 are collinear, and ditto for $S_2, \ldots, S_n$.

[9] Specifically the *stability result* asserts that for every prime degree Galois extension of number fields L/K there exists an elliptic curve E over K with

$$\text{rank}E(K) = \text{rank}E(L) > 0.$$

Part III

Elliptic curves and techniques for studying their arithmetic

10 The virtues of elliptic curves

The study of *Elliptic curves* has quite a unifying effect—which is a source of joy and surprise. It brings together so many other fields of mathematics, and physics and applied areas.

For example, in their essential role in cryptography, elliptic curves have a certain predominance that warrants publications such as this 1999 government memo:

RECOMMENDED ELLIPTIC CURVES FOR FEDERAL GOVERNMENT USE

July 1999

This collection of elliptic curves is recommended for Federal government use and contains choices of private key length and underlying fields.

§1. Parameter Choices

1.1 Choice of Key Lengths

The principal parameters for elliptic curve cryptography are the elliptic curve E and a designated point G on E called the *base point.* The base point has order r, a large prime. The number of points on the curve is $n = fr$ for some integer f (the *cofactor*) not divisible by r. For efficiency reasons, it is desirable to take the cofactor to be as small as possible.

All of the curves given below have cofactors 1, 2, or 4. As a result, the private and public keys are approximately the same length. Each length is chosen to correspond to the cryptovariable length of a common symmetric cryptologic. In each case, the private key length is, at least, approximately twice the symmetric cryptovariable length.

1.2 Choice of Underlying Fields

For each cryptovariable length, there are given two kinds of fields.

- A *prime field* is the field $GF(p)$ which contains a prime number p of elements. The elements of this field are the integers modulo p, and the field arithmetic is implemented in terms of the arithmetic of integers modulo p.

The first page of that memorandum already gets down to the business of discussing the discrete logarithm problem when posed in terms of the near-cyclic group of rational points of those *preferred* elliptic curves, and specifically, the *difficulty* of computing such logs, which—in this game—is a virtue.

But going a bit further back:

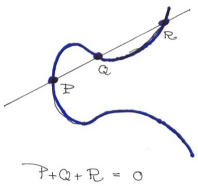
11 In the ancient Greek problems about numbers

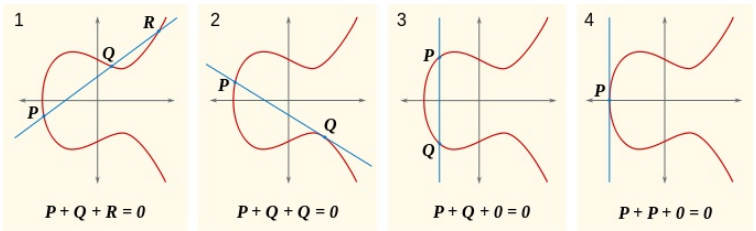
In Diophantus' Problem 24 of Book IV of his treatise (to "divide a number into two numbers so that their product is a cube minus its side") elliptic curves arise implicitly as cubic equations of two variables, and one already sees a hint of (to put it in modern language) the problem of simply *finding points on them.*

In modern vocabulary, Elliptic curves can be represented as smooth plane cubic curves with one point at infinity, and therefore by adroit linear change of variables can be given by an affine equation of the form

$$y^2 = g(x) := x^3 + cx + d,$$

for c, d constants, where the cubic polynomial $g(x)$ has no multiple roots. Such curves then are very algebraic objects, and can be defined over any field k, by taking the constants $c, d \in k$. The "elliptic curve" E itself then is the projective model of this affine curve, and its points rational over the field k is usually denoted $E(k)$ which consists of the single point at infinity –usually called, perversely, 0 or the origin—and all affine points (α, β) each entry in k, satisfying the equation $y^2 = \alpha^3 + c \cdot \alpha + d$. Some readers of Diophantus seem to already find in his treatise hints of what later came to be called the "chord-and-tangent process" for making new points on this curve E (rational over k) from pairs of points in $E(k)$:





This process banks on the fact that our curve is a cubic–i.e., of degree 3— and therefore any straight line (in projective space) will intersect it in exactly three points, counting multiplicity, and depends only on the observation that any line in the projective plane passing through two points with coefficients in a field k will itself be defined over k and hence the third intersection point with

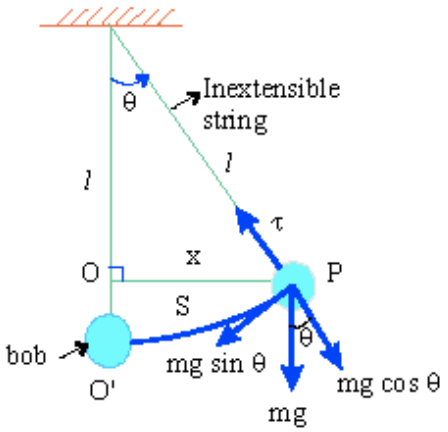
E will have its coordinates in k as well. Defining an addition law of $E(k)$ by stipulating that any 'three' collinear points sum to zero, gives $E(k)$, as it turns out, an abelian group structure:

$$E(k) \times E(k) \longrightarrow E(k)$$

and, taking the algebraic geometric point of view, allows us to think of our elliptic curve as a commutative algebraic group , i.e., an abelian variety.

12 At the time of Abel and Jacobi,

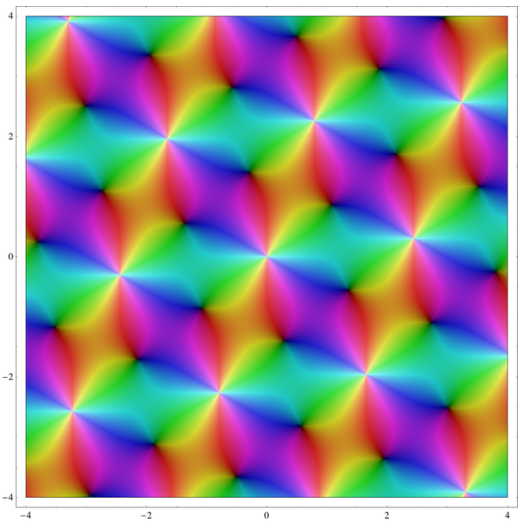
we would already see the trace of elliptic curves *and their periods* in the exact solution of the problem of the 'period' of a simple pendulum as a function of θ, the arc of swing. These show up as integrals over paths on the implicit Riemann surface. (Below k depends only on the physical shape of the pendulum.)



$$Period_k(\theta) = \int_0^{\tan\theta} (\frac{dx}{\sqrt{(1+x^2)(1+k^2x^2)}}.$$

13 From the viewpoint of Weierstrass,

elliptic curves over the complex numbers might arise from double periodic meromorphic functions– $\wp_\Lambda(z)$— on the complex plane. Here is a picture where the color correlates (somehow) to the absolute value of a Weierstrass $\wp$-function.



The structure of doubly-periodic functions being had been viewed (historically) as some companion to the 'singly-periodic' exponential covering ($z \mapsto e^{2\pi i z}$) of the multiplicative group of complex numbers:

$$\mathbf{C} \quad \longrightarrow \quad \mathbf{C}^* \quad \simeq \quad \mathbf{C}/2\pi i\mathbf{Z}.$$

The exponential map is, of course, transcendental. AND it has the following sharper property, thanks to the theorem of Lindemann and Weierstrass that any finite collection of $\mathbf{Q}$-*linearly independent* algebraic numbers have the property that their exponentials and *algebraically independent* over $\mathbf{Q}$.

The corresponding 'doubly-periodic' projections that form quotients of $\mathbf{C}$ by *lattices* $Z \times \mathbf{Z} \approx \Lambda \subset \mathbf{C}$ that produce complex tori,

$$E = E_\Lambda := \mathbf{C}/\Lambda,$$

have the property, first, that the pair consisting of the value, $\wp_\Lambda(z)$, of the $\wp$-function at a point $z \in \mathbf{C}$ and the value of its derivative, ${\wp'}_\Lambda(z)$, at that point determine z modulo Λ, i.e., determine the image of z in E. Moroever, those two (meromorphic) functions $\wp_\Lambda(z)$ and ${\wp'}_\Lambda(z)$ satisfy a cubic polynomial equation. That is, putting $y := {\wp'}_\Lambda$ and $x := \wp_\Lambda$ the parametrization $(x, y) = (\wp_\Lambda, {\wp'}_\lambda)$ satisfies the equation given above,

$$y^2 = x^3 + cx + d,$$

for suitable complex numbers $c = c_\Lambda,\ d = d_\Lambda \ \in \ \mathbf{C}$. We already have here a fascinating, but still simple, structure following the parallel of the exponential function: we have an analytic homomorphism of groups

$$\mathbf{C} \xrightarrow{\pi_\Lambda} E$$

where both domain and range are algebraic groups, and the quotient is formed by passing to the orbits under the action (translation) of Λ, the group of integral points yet another group. As with the exponential function, we have what may be called a *bi-algebraic structure*, where E is defined over a number field (i.e., the c and d are algebraic numbers) the issues of comparison of algebraicity and/or transcendentality of a point and/or its image in the complex plane, under the analytic mapping π_Λ delicate. In fact, we have: the Schneider-Lang Theorem that guarantees that if c_Λ and d_Λ are algebraic, then the nonzero points of $\Lambda \subset \mathbf{C}$ are transcendental.

The variation of possible discrete (rank two) lattices $\Lambda \subset \mathbf{C}$ can be regularized by taking the lattices of the form $\Lambda_\tau := \mathbf{Z} + \tau \cdot \mathbf{Z} \subset \mathbf{C}$where $\tau = x + iy$ has imaginary part positive (i.e., τ lies in the upper half plane $\mathcal{H} \subset \mathbf{C}$). The isomorphism class of $E_\tau := E_{\Lambda_\tau}$ depends only on the orbit of $\tau \in \mathcal{H}$ under the action of $\mathrm{SL}_2(\mathbf{Z})$, the quotient being parametrized by one of the most fascinating functions in the subject, the *elliptic modular function* $j : \mathcal{H} \longrightarrow \mathcal{H}/\mathrm{PSL}_2(\mathbf{Z}) \simeq \mathbf{C}$. The isomorphism class of the elliptic curve E_τ over $\mathbf{C}$ is determined by the value $j(\tau)$ and conversely, given any complex value it is the "j" of a unique isomorphism class of elliptic curves.

Here, again, one has a bi-algebraic structure,

That is, an analytic domain $\mathcal{H} \subset \mathbf{P}^1(\mathbf{C})$ open in an algebraic variety $\mathbf{P}^1(\mathbf{C})$ on which we have an action of an algebraic group (a group scheme) $\mathrm{PGL}_2(\mathbf{C})$ such that the action of its subgroup of real points, $\mathrm{PSL}_2(\mathbf{R})$, preserves $\mathcal{H}$, and the action of its subgroup of integral points, $\mathrm{PSL}_2(\mathbf{Z})$, act discretely giving the bi-algebraic structure,

$$\begin{array}{ccc} \mathcal{H} & \subset & \mathbf{C} \\ j \downarrow & & \\ \mathbf{C} & & \end{array}$$

The range $\mathbf{C}$ when it appears here is sometimes called the **j-line** for, in perhaps a slightly ragged way it classifies all elliptic curves, and can be constituted to parametrize a family of elliptic curves with given j.

AND as in the case of the exponential map, we have a corresponding version of the Lindemann-Weierstass Theorem (The "Ax-Lindemann Conjecture," proved by Klinger, Ullmo and Yafaev. ***)

As is standard nowadays, we systematically consider, and classify, elliptic curves endowed with certain specific properties, or features, such as pairs of elliptic curves together with a chosen point of order N—the completed moduli space for such problems being a curve usually denoted

$$X_1(N) := \mathcal{H}/\Gamma_1(N) \cup \mathit{cusps}$$

or cyclic subgroup of order N,

$$X_0(N) := \mathcal{H}/\Gamma_0(N) \cup \mathit{cusps}$$

or—as we shall see in a moment—other structures as well. These modular curves have natural models over $\mathbf{Q}$. The modular curve $X_0(N)$ may also be interpreted as classifying *cyclic* isogenies $E \to E'$ of elliptic curves of degree N. This is because if you have a cyclic subgroup C_N of order N in E, letting $E' := E/C_N$, the natural projection $\pi : E \to E'$ is a homomorphism with kernel a finite cyclic subgroup of order N (this being the definition of cyclic isogeny) and the converse, too, is clear. Note that there is a natural involution— the Atkin-Lehner involution w of $X_0(N)$ which sends the cyclic isogeny $E \to E'$ to $E' \to E$ (by passing to the dual).

An extremely celebrated theorem, the modularity theorem (Wiles, Taylor, ...) guarantees that any elliptic curve over $\mathbf{Q}$ admits a parametrization by such a modular curve $X_0(N)$ and the parametrization

$$\pi : X_0(N) \to E$$

is defined over $\mathbf{Q}$ unique up to sign if you require that both the integer N and the degree of π be minimal, among all such parametrizations. In the sense that the curves $X_0(N)$ classify all elliptic curves with the requisite cyclic isogeny, and the parametrization π covers E, so in a sense one could say that every elliptic curve over $\mathbf{Q}$ *knows* all elliptic curves—each of its points classifying a finite set of elliptic curves (with particular structure).

14 In the era of Mordell,

the arithmetic of elliptic curves was already in full swing, and any number of a host of questions Mordell himself asked, such as

> What products of two consecutive integers are equal to a product of three consecutive integers?

leads to very interesting questions about elliptic curves. The answer to this question, by the way, known to Mordell half a century ago, is that the only such products are $0, 6$, and 210.

The equation whose integral solutions "solves" Mordell's Question is

$$\mathcal{E} : \quad y^2 + y = x^3 - x$$

and this is an affine model, over $\mathbf{Z}$, of an elliptic curve over $\mathbf{Q}$ which we'll call **Mordell's Elliptic Curve**.

Now if you want to know the answer to Mordell's question, you need only study the *integral* solutions of that equation. For such equations (quadratic expressions of the variable y as equal to cubic expressions of x)—-and in contrast to the general problem of integral solutions as posed by Hilbert's Tenth Problem and as solved negatively by Matyasevich—there is an algorithm allowing one to finitely determine all its integral solutions.

If we return to Mordell's equation and ask for its *rational* rather than only integral solutions, we get quite a different, and beautiful, answer: there are infinitely many rational solutions, and all of them are 'generated' out of the simplest of its solutions: $(x, y) = (0, 0)$.

P_1=[0, 0]

P_2=[1,0]

P_3=[-1,-1]

P_4=[2,-3]

P_5=[1/4, -5/8]

P_6=[6, 14]

P_7=[-5/9, 8/27]

P_8=[21/25, -69/125]

P_9= [-20/49, -435/343]

P_{10}=[161/16, -2065/64]

P_{11}=[116/529, -3612/12167]

P_{12}=[1357/841, 28888/24389]

P_{13}=[-3741/3481, -43355/205379]

Remarks:

1. Notice the growth of the numerators and denominators of these solutions. On the page they trace out the shadow of a parabola (and would do so more strikingly if I put it in smaller type and computed more of them). The equation of the 'limit' parabola is itself an important arithmetic invariant of the elliptic curve (if I normalize for the size of typefont)—determined by the canonical heights of those point, and related to the regulator of the elliptic curve[10] of this elliptic curve. The way Bjorn Poonen and *** use to reconstruct–in a diophantine manner—the ring of integers of a number field K in the ring of integers of an extension field L in the case where one has an elliptic curve over K diophantine-stable for L/K such that $E(K) = E(L)$ is a group of rank one[11] is to work closely with the common structure of their groups of rational points and make close use of the denominators of the x-coordinates of rational points. This follows a long tradition, beginning with the use of Pell's equation, and is a remarkable project.

[10] (which happens to be $0.0511114082399688\ldots$ and equal—in this case—to the ratio of the value of the derivative of the L-function divided by the real period, $0.305999773834052\ldots/5.98691729246392$)

[11] (later work of *** allows one to use stability when the rank is more generally just ≥ 1)

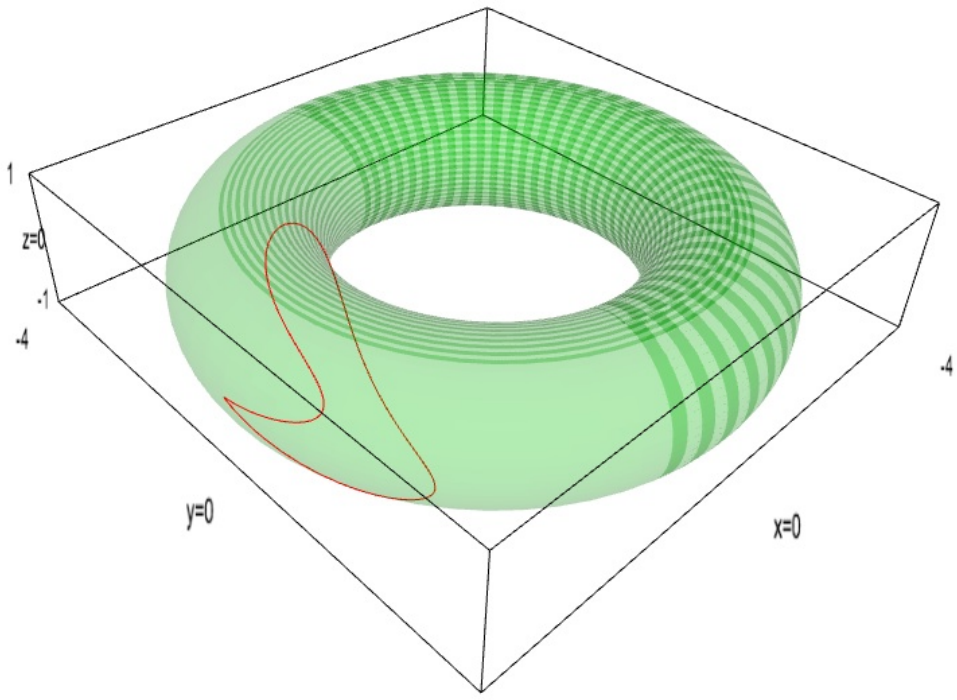
2. **Mordell's elliptic curve** knows all other elliptic curves in the sense described above insofar as it's points classify (symmetrically) pairs $E \to E'$ and $E' \to E$ each of them being a 37-isogeny, and each dual to the other.

3. The curve $\mathcal{E} - 0$ is a quotient of $Y_0(37)$ by the Atkin-Lehner involution. So it comes with a natural covering $\mathcal{H} \to \mathcal{E}$. In this way **Mordell's elliptic curve** inherits the hyperbolic structure of $\mathcal{H}$ In particular, is laced with the image of all the closed geodesics. For any rational number a/b define the vertical line in the upper half place with abscissa a/b:

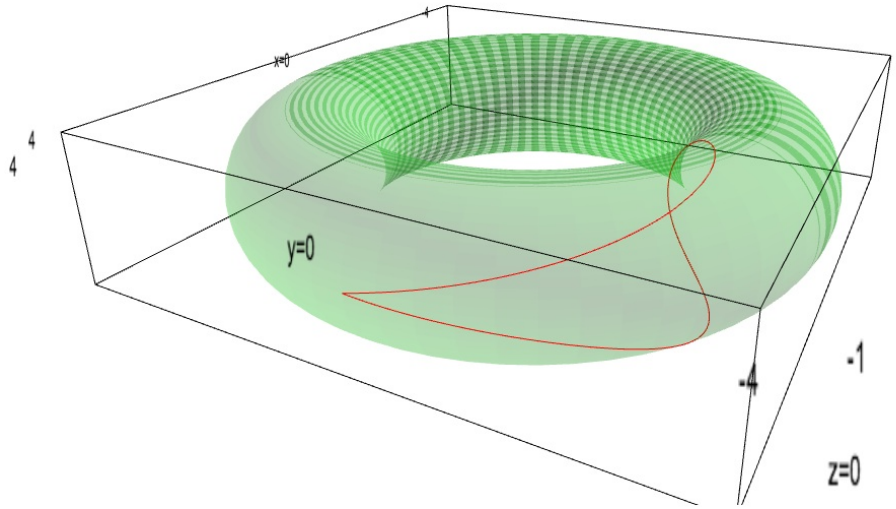
$$I(a/b) := \{a/b + iy \mid 0 \leq y \leq \infty\}$$

and consider its image in $\mathcal{E}$. This is a countably infinity family of loops,starting and ending at 0, on the Riemann surface $\mathcal{E}$ and are actual geodesics in the hyperbolic structure of $\mathcal{E} - 0$. The "modular symbols" discussed previously are *rational numbers with bounded denominators* obtain by integrating a natural differential form over these curves, normalized by division by a period. In the drawings below we draw $\mathcal{E}$ as a recognizable torus, so unfortunately you lose the vision of these loops as geodesics. But here are a few (thanks loads to Hao Chen for these!)

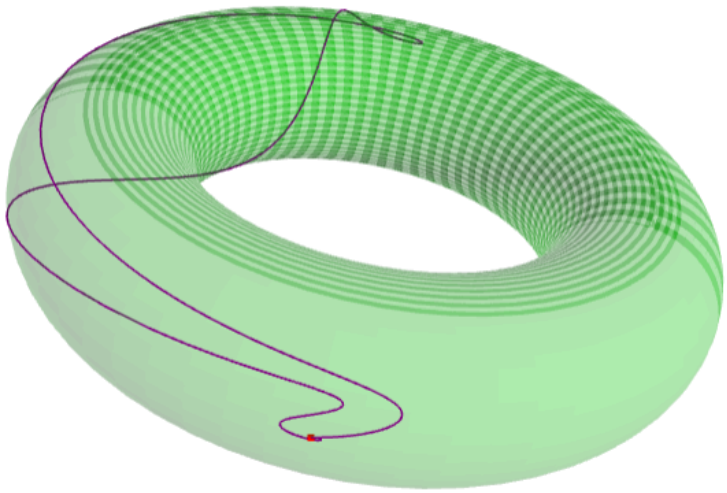
$I(1/7)$



$I(4/5)$



$I(389/4001)$



A fundamental theorem (1922) for any elliptic curve E of Mordell (over $\mathbf{Q}$) extended by Andrei Weil over any number field K says that the group $E(K)$ is a *finitely generated* abelian group (called naturally, the **Mordell-Weil group** of E over K) and so is characterized up to isomorphism by its two invariants:

- its *torsion subgroup*, $T(E, K)$,
- and its *rank* $r(E, K)$.

I.e.,

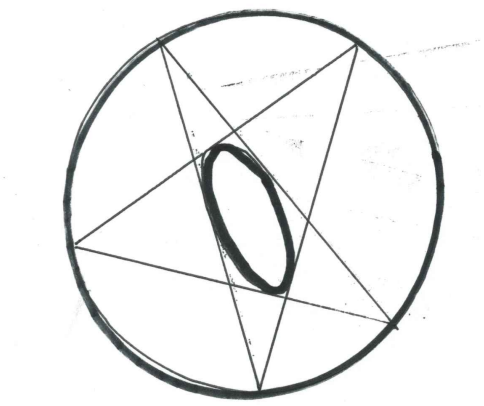
$$E(K) \simeq T(E, K) \bigoplus \mathbf{Z}^{r(E,K)}.$$

This immediately leads to two mathematical projects that are—as it turns out—surprisingly different.

- Study the behavior of torsion $(E, K) \mapsto T(E, K)$,
- Study the behavior of rank $(E, K) \mapsto r(E, K)$,

15 Torsion

Torsion in elliptic curves have, as one of their many neat realizations, periodic arrays in the classical it Poncelet Billiard game where you have a configuration of two conics in the plane (I think of one of them as the "outer conic" comprising the outer profile of the billiard table, encircling the other conic, which we'll call the "inner conic," and which we can think of as an obstruction on the table. The game is to make a shot that bounces multiple times off the rim of the outer conic, but each time it comes back, its path just grazes the inner one, and it makes a finite periodic trajectory this way.



We have a complete classification of torsion, rational over Q for elliptic curves defined over Q. It could be stated this way...

Theorem 8. *$T(E, \mathbf{Q})$ is either cyclic of order ≤ 10, or order 12, or else is a direct product of a cyclic group of order 2 with a cyclic group of order $2, 4$ or 6. Moreover, for each of these structures there is a single rationally-parametrized one parameter family of elliptic curves with that type of torsion subgroup.*

... or in the following more suggestive "minimalist" way:

Theorem 9. *The isomorphy type of a finite group T occurs as the rational torsion group $T(E, \mathbf{Q})$ of some elliptic curves over $\mathbf{Q}$ only when it is forced to occur, by algebraic geometry. Namely, only when the modular curve classifying elliptic curves endowed with such a finite subgroup is isomorphic to $\mathbf{P}^1$. In such a case, there is an infinite rationally parametrized family of elliptic curves over $\mathbf{Q}$ possess T as rational torsion group. .*

Say, then, that an elliptic curve E over a number field of degree d with rational torsion group $T(E, K)$ is **sporadic** if it doe not occur in a rationally parametrized family (over $\mathbf{Q}$) of elliptic curves E over number fields of degree d with rational torsion group $T(E, K)$.

So, there are no sporadic points over number fields of degree 1. Thanks to Merel, Oesterlé, Parent, Kamienny, and very recent progress due to Maarten Derickx, Sheldon Kamienny, William Stein, Michael Stoll, and van der Hoej there's a very promising area to be explored for torsion over fields of degree d over $\mathbf{Q}$.

Fix a positive integer d and let $P(d)$ be the *largest* prime number p such that there exists an elliptic curve (without CM; i.e., without 'extra' endomorphisms) defined over some number field of degree $\leq d$ over $\mathbf{Q}$ and for which there is a point of order p on that elliptic curve, rational over that field.

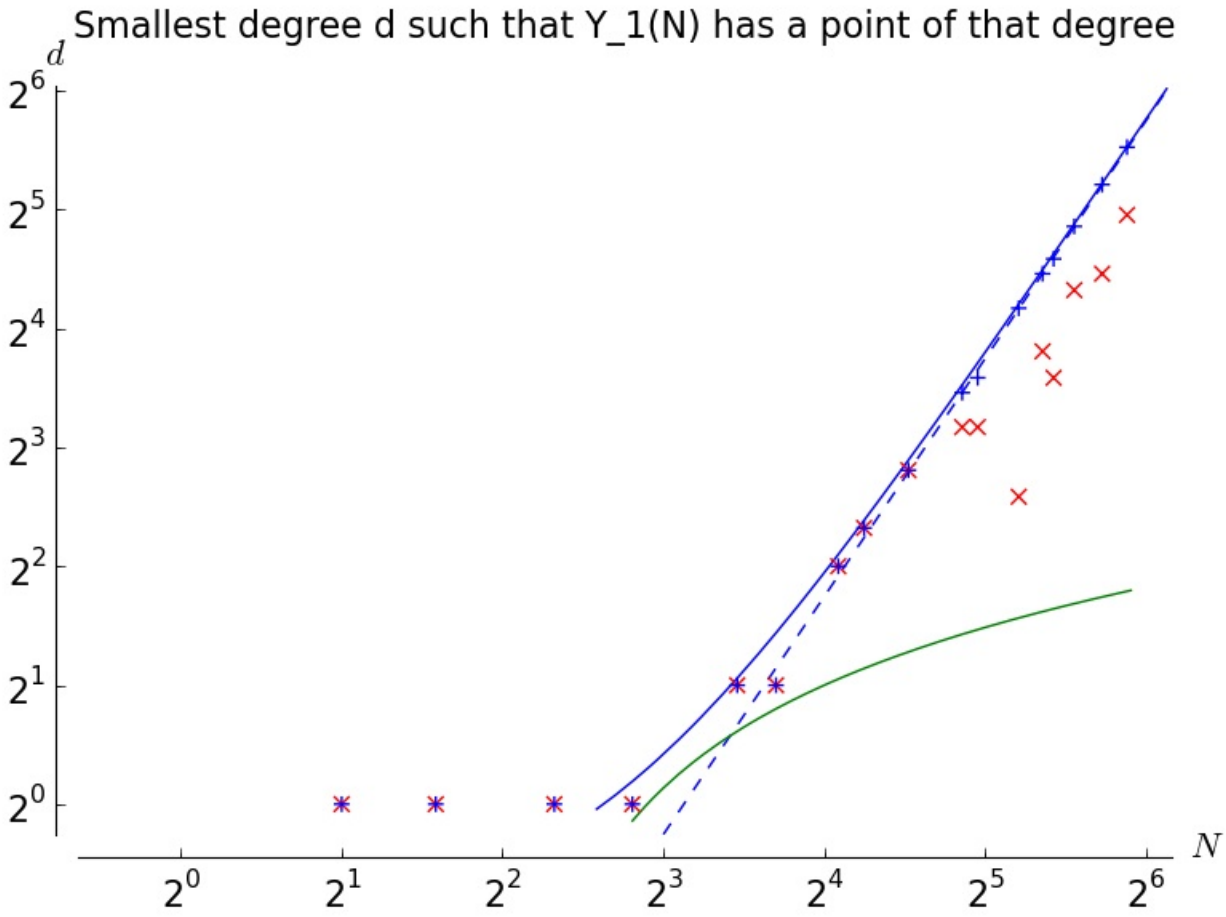
We have the proved bounds, the upper bound being the result of Merel, Oesterlé, Parent :

$$d^{1/2} << P(d) << 3^d.$$

Conjecture 10.

$$P(d) <<_\epsilon d^{1/2+\epsilon}.$$

Here below is a graph computed by Maarten Derickx and Mark van Hoej. It is a log-log plot where the axes are $(x, y) = (\log p, \log d)$, the data points recording examples of 'lowest' degree d for the corresponding p occurs as prime torsion in a non-CM elliptic curve (over a field of degree d). The quotation-marks around the word 'lowest' is meant to signal that the blue data points and the blue extrapolated line corresponds to the lowest d for which there is a rational family of such examples of prime torsion p over fields of degree d. The red data points correspond to the sporadic points. The green curve is the proved (exponential) lower bound relating d to p. Visibly, much more computation needs to be done if we are to be able to surmise any general behavior with some feeling that there is evidence behind our guess.



16 Density questions having to do with rank

Let K be a fixed number field and consider the collection of all elliptic curves defined over K. The most natural 'first question' that is somewhat of a statistical nature that you might ask about Mordell-Weil rank is:

> Does $r(E;K)$ admit a finite upper bound (for fixed K and all elliptic curves over K)?

Here, far from actually having a resolution of this yes or no question, we don't even seem to enjoy a uniform consensus about guesses for what the truth is here, even for the field $\mathbf{Q}$. (There are number theorists who believe yes, and others who believe no.) The following chart, which I got off the web, tabulates world's record large ranks for elliptic curves over $\mathbf{Q}$—so far— with the year of their discovery and the winners.

rank $\geq$	year	Author(s)
3	1938	$Billing$
4	1945	$Wiman$
6	1974	$Penney - Pomerance$
7	1975	$Penney - Pomerance$
8	1977	$Grunewald - Zimmert$
9	1977	$Brumer - Kramer$
12	1982	$Mestre$
14	1986	$Mestre$
15	1992	$Mestre$
17	1992	$Nagao$
19	1992	$Fermigier$
20	1993	$Nagao$
21	1994	$Nagao - Kouya$
22	1997	$Fermigier$
23	1998	$Martin - McMillen$
24	2000	$Martin - McMillen$
28	2006	$Elkies$

To see what's involved in the last entry (Elkies elliptic curve) of this table:

Elkies elliptic curve:

$$\mathcal{E}: \quad Y^2 + XY + Y \;=\; X^3 - X^2 -$$

$$2006776241557552658503320820$$

$$933854275093023031217895650 2X$$

$$+$$

$$3448161179503055646703298569039072037 4855$$

$$9443593191803612660082962919394487322434 29$$

Independent points of infinite order:

P1 = [-212415009125438107329213 7463, 259854492051899599030515511070780628911531]
P2 = [233450986603470175688475 4537, 188720041954944691808683165528036279315 31]
P3 = [-167173605406236906387903 8663, 251709377261144287808506947241319126049131]
P4 = [213913026013915666649298 2137, 366395091714397292024214596929412975275 31]
P5 = [153470676446712072388547 7337, 854295853460176942890210328627810727995 31]
P6 = [-273107948787567703334157 5063, 262521815484332191641284072623902143387531]

P7 = [277572626684457164970545853 7, 1284575547401406024886948769908264036993 1]
P8 = [149438572932718895754183381 7, 8848660552773340598611649451404923341145 1]
P9 = [186843822862088735850906525 7, 5923740321443770871272514039305935858913 1]
P10 = [200894510882574377486654253 7, 4769067788012555288215175078154142471153 1]
P11 = [234836054091802516965163293 7, 1749293000620055785734033247644880436353 1]
P12 = [-14720840070904811744700086 63, 2466434506535037141999474415497597984691 31]
P13 = [292412860770806121336328893 7, 2835026443148887850148835647476737589953 1]
P14 = [537499389106606189329393453 7, 2861889084272633864511750319164798937315 31]
P15 = [170969076823335452333400855 7, 7189883497468608946615970052921598092163 1]
P16 = [245095401135359314407259518 7, 4445228173532634357049262550610714736531]
P17 = [296925470927355916746467493 7, 3276689307536627080133368254316046968753 1]
P18 = [271191493494169260133288293 7, 2068436612778381698650413981506590613531]
P19 = [200785860779968545287783289 37, 27796085411378066046560517256246240300915 31]
P20 = [215808245024073477431781069 7, 3499437340196402680996966224180090125473 1]
P21 = [200464545824705902240322493 7, 4804932978070464552243986699988847546753 1]
P22 = [297574945094799626494709133 7, 3339898982607532232020893441010485786913 1]
P23 = [-21024904676862851501473478 63, 2595763914598757895716773931716872032275 31]
P24 = [311583179915063034902194537, 1681043852299806035401094729156601534739 31]
P25 = [277393100834186523144377181 7, 1263216283464992100241411627376927581345 1]
P26 = [215658118814376840936346138 7, 3512509296402290889700415051637517808733 1]
P27 = [386633049987241250881565913 7, 1211977556559442262930369267150258473225 31]
P28 = [223086828977357602377867873 7, 2855876003059748566338702060076864002853 1]

We seem to have no new entries for the above table in the *past eight years*, but our knowledge, and the precision of our expectations, about densities of ranks is extending daily thanks to people in this room!

17 The computable upper bound, and the constraint of *parity*

- **A theorem:** For every prime number p there is a *computable* number $r_p(E,K)$—called the **reduced mod p-Selmer rank**—that constitutes an upper bound for the Mordell-Weil rank:
$$r(E,K) \leq r_p(E,K).$$
- **A Conjecture:**
$$r(E,K) \equiv r_p(E,K) \mod 2,$$
i.e., the Mordell-Weil rank is of the same *parity* as the reduced mod p-Selmer rank (for every p).
- **A Fact:** We have (at least) the beginning of an understanding of statistical questions regarding the parity of reduced mod p-Selmer rank (and this conjecturally translates to a similar understanding of the statistics of Mordell-Weil rank).

Let's make some guesses now about rank, following the minimalist instinct. However, at this point it pays

- to repeat that *parity* is indeed a constraint and something that one must take careful account of, before making guesses, and

- to note that to do statistics about infinitely many instances one must say how one orders them. The ordering arrangement doesn't have to be a full linear ordering, but at the very least it should be given by an increasing system of finite subsets of the objects that are being studied, where the union of all these finite subsets is the whole. Then, one can talk about densities, or probabilities of features.

We will discuss statistics for the following two types of families.

(a) *All* elliptic curves defined over a fixed number field K. This infinite collection is "ordered" by the size of the absolute value of the norm of the conductor.

(b) *All* quadratic twists of a given elliptic curve E over a given field K. This boil down to considering the class of elliptic curves expressible by the equations

$$E^{(d)}: \quad dy^2 = x^3 + ax + b$$

for $a, b, d \in K$, with a, b fixed and d an integer of K, varying (mod squares). This infinite collection is "ordered" by the maximum size of the absolute value of the norm of any prime ideal dividing d.

The minimalist instinct then suggests:

Question 3. *Is it true that, in either of these cases, if we consider the statistics of the sub-collection with even Mordell-Weil rank parity, it is* 100% *likely that the Mordell-Weil rank of a member of that family is* 0*? And as for the statistics of the sub-collection with odd Mordell-Weil rank parity, is it* 100% *likely that the Mordell-Weil rank of a member of that family is* 1*?*

(For the second type of family, at least for those over $K = \mathbf{Q}$, this was already conjectured by Dorian Goldfeld in 1979.)

Of course, to connect these expectations with a general sense of the average rank, we should either know or guess something about the density of parity.

18 All elliptic curves over a fixed number field

For the first type of family described above, i.e. for all elliptic curves defined over a fixed number field K, we expect that the distribution of even/odd parities is 50/50; i.e., half are even and half are odd, when the count is made according to the ordering that we described.

This would suggest the following target:

Conjecture 11. *The average Mordell-Weil rank for all elliptic curves over any fixed number field K is* 1/2.

In 1992 Armand Brumer showed (by analytic means, and conditional on standard conjectures) that the average rank of elliptic curves over $K = \mathbf{Q}$ is bounded above by 2.3.

More recently we have the magnificent achievement of Arul Shankar and Manjul Bhargava who established that it is bounded above by 0.99.

NOTE: I should include the most up-to-date announcements of Manjul and his co-authors!

This is by a formidable new tack on the geometry-of-numbers approach to counting mathematical objects related to this problem. Things are moving and we might hope for continued progress here in the coming years.

19 Quadratic twist families

Here we have some classical work by Heath-Brown for a specific family, and by Swinnerton-Dyer (with a recent improvement by Dan Kane) for the special case of elliptic curves over $\mathbf{Q}$ that have particular features related to their 4-torsion. Importantly, they establish finite average values of Mordell-Weil ranks for these families.

But, conceiving the problem for more general number fields one encounters a (surprising) new feature in the nature of parity itself. This is described in recent work of Zev Klagsbrun, Karl Rubin and myself. We deal with the mod 2-Selmer rank parity for a quadratic twist family over a number field K. This, then, is conjecturally the Mordell-Weil rank parity. We show that in the case where the number field K has at least one real embedding, the distribution of even/odd parities is 50/50. But even if you fix a specific elliptic curve E but allow your self to consider different choices of field K over which you gather parity statistics, the proportions of even/odd can change dramatically. For example, take the elliptic curve (labelled 50B1 by Cremona)

$$E: \quad y^2 + xy + y \;=\; x^3 + x^2 - 3x - 1.$$

By judicious choices of fields K one can obtain quadratic twist families whose mod 2-Selmer rank parity ratios take on a dense set of numbers in the range $(0, 1)$.

Part IV

L-functions and Selmer groups

Let E be an elliptic curve over a number field K. Here are the basic tools we have to understand its arithmetic. I'll formulate this for $K = \mathbf{Q}$, and then comment on how this description does or does not extend to other number fields and elliptic curves.

- **The local study.**

  We want to reduce E modulo p systematically for primes p. For all but finitely many primes we can get its reduction mod p— an elliptic curve over $\mathbf{F}_p$—by simply reducing the coefficents of its equation mod p. We define the basic local invariant $a_p(E) := 1+p-|E(\mathbf{F})|_p$. Knowledge of these a_p's for primes p gives us knowledge of the isogeny class of each of the elliptic curves $E_{/\mathbf{F}_p}$. Even for the remaining finitely many primes, there is a natural definition of an "a_p." To summarize:

  The local information for the arithmetic of E is given by the function on primes:

  $$p \mapsto a_p \sim \text{isogeny class of the reduction of } E \text{ mod } p.$$

- **Taking all the local studies together.** We put this together to get the **L-function** as Dirichlet series.

  $$L(E, \mathbf{Q}, s) = \prod_p \frac{1}{1 - a_p p^- s + p^{1-2s}} = \sum_{n \in \mathbf{Z}} a_n n^{-s}$$

where the coefficients a_n are easily expressible in terms of the a_p for p prime.

So far, we can do something analogous for all E and all K, giving a Dirichlet series $L(E, K, s)$ convergent in right half-plane.

- **The implications for global arithmetic.**

  Conjecture 12. *The Dirichlet series $L(E, K, s)$ extends to an analytic function in a region including the point $s = 1$ and has a zero at $s = 1$ of order equal to the rank of $E(K)$.*

  The extension of $L(E, \mathbf{Q}, s)$ to an entire function on the complex plane is, of course, one of the great achievements of modern number theory, and follows from the modularity theorem of ***. The full conjecture holds in the case where the order of zero at $s = 1$ is 0, or 1, thanks to ****

The restriction of use of this basic tool of arithmetic is first, that the analytic extension is known for only a limited class of number fields; and the remainder of the conjecture, when known to be true, is at the moment known only in cases where the order of vanishing is 0 or 1.

20 Some words about the methods for proving diophantine stability

20.1 Descent

The standard method—perhaps the only fully proved method—of finding upper bounds for $r(E, K)$ for specific elliptic curves E over specific fields K (or when extended to abelian varieties over number fields K) is the *method of descent* that seems to have been already present in some arguments due to Fermat and has been elaborated and refined ever since. These days "descent" is done via computation of what are called **Selmer groups**. Here is the "shape" of the descent method as it has evolved in present times. One should note that there are two virtues to this classical method. It is 'elementary' in the sense that its ingredients are hardly anything more than Galois cohomology and basic algebraic number theory. Also it works for all number fields. I'll explain it first for elliptic curves when the base field K is the rational field $\mathbf{Q}$, and then discuss the differences that one encounters over general fields and for general abelian varieties.

Remember, though, that I want to explain the diophantine stability features that it helps with, so we will also be considering the relative theory when one passes from our base field to a cyclic extension of prime degree ℓ over the base field.

20.2 Selmer groups

For simplicity, fix the elliptic curve E over $\mathbf{Q}$ and a prime ℓ to illustrate the method. The basic exact sequence of $\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q})$-modules given by multiplication by ℓ,

$$0 \to E[\ell] \to E(\bar{\mathbf{Q}}) \xrightarrow{\times \ell} E(\bar{\mathbf{Q}}) \to 0$$

gives us, after passing to cohomology, an injection

$$E(\mathbf{Q})/\ell E(\mathbf{Q}) \ \hookrightarrow \ H^1(\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}), E[\ell]).$$

The $\mathbf{F}_\ell$-vector space $H^1(\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}), E[\ell])$ is infinite dimensional, and we want to capture the subspace $E(\mathbf{Q})/\ell E(\mathbf{Q})$, thereby computing—after a tiny bit of work—the Mordell-Weil rank of E over $\mathbf{Q}$.

Locally,over $\mathbf{Q}_p$ for any prime p we have the same story,

$$E(\mathbf{Q}_p)/\ell E(\mathbf{Q}_p) \ \hookrightarrow \ H^1(\mathrm{Gal}(\bar{\mathbf{Q}}_p/\mathbf{Q}_p), E[\ell]).$$

and the global and local pictures fit neatly together in that the projection

$$H^1(\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}), E[\ell]) \to H^1(\mathrm{Gal}(\bar{\mathbf{Q}}_p/\mathbf{Q}_p), E[\ell])$$

sends $E(\mathbf{Q})/\ell E(\mathbf{Q})$ to

$$E(\mathbf{Q}_p)/\ell E(\mathbf{Q}_p) \ \hookrightarrow \ H^1(\mathrm{Gal}(\bar{\mathbf{Q}}_p/\mathbf{Q}_p), E[\ell]).$$

$$\begin{array}{ccc} E(\mathbf{Q})/\ell E(\mathbf{Q}) & \longrightarrow & H^1(\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}), E[\ell]) \\ \downarrow & & \downarrow \\ E(\mathbf{Q}_p)/\ell E(\mathbf{Q}_p) & \longrightarrow & H^1(\mathrm{Gal}(\bar{\mathbf{Q}}_p/\mathbf{Q}_p), E[\ell]) \end{array}$$

It is natural, then to try to at least approximately 'cut out' the subspace $E(\mathbf{Q})/\ell E(\mathbf{Q})$ by using all this local information together. That is the purpose of the *Selmer group*, $S_\ell(E)$.

Definition 3. *$S_\ell(E) \subset H^1(\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}), E[\ell])$ is the intersection over all prime p of the inverse images of the images*

$$E(\mathbf{Q}_p)/\ell E(\mathbf{Q}_p) \subset \ H^1(\mathrm{Gal}(\bar{\mathbf{Q}}_p/\mathbf{Q}_p), E[\ell]).$$

What we have done, then, is to impose a **local condition** for each prime p: that the cohomology classes giving elements of the Selmer group reduce to specific subgroups in local cohomology. The

subgroups will be called "local conditions." The Selmer group is the subgroup consisting of all cohomology classes in this infinite dimensional vector space $H^1(\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}), E[\ell])$ that satisfies **all these local conditions**.

Let us call these local conditions our **base local conditions** noting that if we imposed other local conditions, we will probably get different groups. AND nothing stops us from *defining* 'Selmer groups' with any local conditions we want—*artificial Selmer groups* so to speak. We will eventually do this, modifying our base Selmet conditions at finitely many primes.

The base Selmer group has four key properties, the last two being conjectures:

- $S_\ell(E)$ is *computable* in theory and is a finite dimensional $\mathbf{F}_\ell$ vector space; i.e., there is indeed a finite algorithm that computes it.
- The subspace $E(\mathbf{Q})/\ell E(\mathbf{Q})$ is contained in $S_\ell(E)$.
- **Conjecture:** the dimensions of $E(\mathbf{Q})/\ell E(\mathbf{Q})$ and in $S_\ell(E)$ have the same parity.
- **Conjecture:** If $\ell >> 0$ $E(\mathbf{Q})/\ell E(\mathbf{Q}) = S_\ell(E)$.

20.3 The *relative theory (for elliptic curves)*

Here we consider a cyclic extension $L/\mathbf{Q}$. The issue for us is whether or not $\mathrm{rank}(L) > \mathrm{rank}(\mathbf{Q})$. The Galois group $\mathrm{Gal}(L/\mathbf{Q})$ acts on the finite dimensional $\mathbf{Q}$-vector space $E(L) \otimes \mathbf{Q}$. Diophantine stability here requires that the action be trivial, i.e, for any Dirichlet character χ of order ℓ that cuts out this cyclic field extension, the χ-component of the $\mathrm{Gal}(L/\mathbf{Q})$-representation $E(L) \otimes \mathbf{Q}$ vanish. We view

$$\chi : G_{\mathbf{Q}} \to \mu_\ell$$

as a homomorphism of the Galois group $G_{\mathbf{Q}} := \mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}$ onto the group of ℓ-roots of unity (we won't say where these ℓ-th roots of unity lie), and L is the field fixed by the kernel of χ.

The χ-twisted Selmer group. Given such a Dirichlet character χ we consider the packet of local characters

$$\{\chi_p : G_{\mathbf{Q}_p} \to \mu_\ell\}_p$$

obtained from it. There is a procedure for twisting the local Selmer condition at p by a local character $G_{\mathbf{Q}_p} \to \mu_\ell\}_p$ at p and we do this, guided by the packet of local characters coming from a global character χ. By imposing those "χ-twisted" local conditions attached to the local characters χ_p related to the global character χ on classes in the *same* infinite dimensional vector space $H^1(\mathrm{Gal}(\bar{\mathbf{Q}}/\mathbf{Q}), E[\ell])$ one defines a χ-twisted Selmer group $S_\ell(E;\chi)$ with one important property for us:

> If, for any χ cutting out $L/\mathbf{Q}$, we have $S_\ell(E;\chi) = 0$, then $\mathrm{rank}(L) = \mathrm{rank}(\mathbf{Q})$.

The aim, then, is starting with any $S_\ell(E;\chi)$ which has positive dimension, to modify the local χ_p's at one prime p at a time—by multiplying χ by a local character ψ_p at a very judicious choice of prime p and prove that $S_\ell(E;\chi\cdot\psi_p)$ has lower dimension. (in fact, one lower dimension). Keep going, to end up with an artificial, perhaps, collection of local conditions giving trivial Selmer group.

In a general situation, two obstacles stand in the way of this plan:

1. **Enough critical primes** To identify the judicious primes p above that perform this 'lowering of dimension' for us. We call them **critical primes** p and their basic features are that p is of good reduction for E and ℓ divides $p-1$ (no problem finding primes of this sort) and that the action of ϕ_p, the Frobenius element at p on the $\mathbf{F}_\ell$-vector space $E[\ell]$ have a *one-dimensional subspace of fixed vectors*; colloquially a 'unique' fixed eigenvector. Here—given some other hypotheses that will obtain when $\ell >> 0$)—we make use of Global Duality to guarantee that between the strictest local condition at p and the most relaxed local condition at p, the corresponding Selmer groups differ in size by one dimension.

   Moreover, we engineered our choice of prime p so that the reduction homomorphism mapping $S_\ell(E,\chi)$ is onto the one-dimensional Selmer local condition. In this set-up, any change of local condition subgroup at p will define an "artificial global Selmer group" of dimension one less than dim $S_\ell(E,\chi)$.

   Iterating this process a finite number of times, leads us to a modification of the base local conditions at finitely many critical primes, such that teh artifically constructed Selmer group is zero.

2. **Enough silent primes** In the account we gave, we modified local conditions for the construction of our Selmer group, a single place at a time, to keep lowering dimension. Why, at the end of our process, is there a **global** Dirichlet character whose corresponding local characters give us the local Selmer conditions we end up with? The answer is: there needn't be such: we'll call such non-globalizable systems of local characters "semi-local." Here is where "silent primes" enter. For $\ell >> 0$, there are primes $p \equiv 1$ moduloℓ ($p \neq 2$ and of good reduction for E) such that ϕ_p has no nonzero fixed vectors in its action on $E[\ell]$. For these, the local cohomology group vanishes.These we'll call silent primes, for the twisting the local condition at such primes by ψ_p doesn't change the local condition, hence the Selmer group. But judicious twisting by silent primes will turn semi-local characters to global ones.

3. Result: we end up with a global character Ψ of order ℓ such that

$$S_\ell(E,\Psi) = 0.$$

   In fact quite a number of characters Ψ—but, unfortunately—not a positive density of them.

20.4 The *relative theory (for absolutely simple abelian varieties)*

The issue of **critical primes** and **silent primes** becomes more delicate in the context of abelian varieties, and we thank Michael Larsen for writing an appendix to our paper that provides what is needed. To make things simple, we'll discuss what is needed when $End(A) = \mathbf{Z}$.

Theorem 1. *(Larsen) If A is an abelian variety over a number field K with $\mathrm{End}_{\bar{K}} A = \mathbf{Z}$, then:*

here exists a positive density set of primes ℓ for which:

1. **"Silent elements"** *there exists $g_0 \in \mathrm{Gal}(\bar{K}/K^{\mathrm{ab}})$ such that $A[\ell]^{g_0}$, and*
2. **"Critical elements"** *there exists $g_1 \in \mathrm{Gal}(\bar{K}/K^{\mathrm{ab}})$ such that $A[\ell]^{g_1} \cong \mathbf{F}_\ell$.*

We apply this theorem, using the Cebotarev density theorem, to find our silent primes and critical primes; i.e., primes such that their corresponding Frobenius elements are silent, or critical elements. There are two steps, both interesting, in the proof of this theorem. The first is a proposition about general irreducible representations of simply connected, split semisimple algebrac groups over $\mathbf{F}_\ell$ (for $\ell >> 0$).

Larsen

Proposition 2. *(Larsen) For every positive integer n, there exists a positive integer N such that if ℓ is a prime congruent to 1 (mod N), G is a simply connected, split semisimple algebraic group over $\mathbf{F}_\ell$, and $\rho\colon G(\mathbf{F}_\ell) \to \mathrm{GL}_n(\mathbf{F}_\ell)$ is an almost faithful absolutely irreducible representation such that $(\mathbf{F}_\ell^n)^{\rho(g_0)} = (0)$ for some $g_0 \in G(\mathbf{F}_\ell)$, then there exists $g_1 \in G(\mathbf{F}_\ell)$ such that*

$$\dim(\mathbf{F}_\ell^n)^{\rho(g_1)} = 1.$$

The key to the proof of this is to find the appropriate element in the image of a principal homomorphism of SL_2 into G.

Michael Larsen applies his proposition to the Galois representations associated to A, an abelian variety over a number field K with $\mathrm{End}_{\bar{K}}(A) = \mathbf{Z}$

Theorem 2. *If A is an abelian variety (of dimension d) over a number field K with $\mathrm{End}_{\bar{K}}(A) = \mathbf{Z}$, then there exists a positive density set of primes ℓ for which there exist elements $g_0, g_1 \in \mathrm{Gal}(\bar{K}/K^{\mathrm{ab}})$ such that $A[\ell]^g \cong 0$ or $\mathbf{F}_\ell$ respectively.*

Of course, here, we are looking for elements in $\mathrm{Aut}_{\mathbf{F}_\ell}(A[\ell]) \sim \mathrm{GL}_{2d(\mathbf{F}_\ell)}$ in the image,

$$\bar{\Gamma}_\ell \subset \mathrm{GL}_{2d(\mathbf{F}_\ell)},$$

of the Galois group G_K. This for $\ell >> 0$.

Step 1: The first task is to relate this image, $\bar{\Gamma}_\ell$—up to an index bounded independent of ℓ to a the $\mathbf{F}_\ell$ points of a simply connected split semisimple algebraic group. Let $G_\ell/\mathbf{Q}_\ell$ denote the Zariski-closure of Γ_ℓ. By a theorem of Serre Course84-85,Ribet-1-29-81, by replacing K with a larger number field if necessary, we may assume that G_ℓ is connected for all ℓ. Further enlarging K, by another theorem of Serre Vigneras, we may assume that for all $\ell \gg 0$ there exists an absolutely irreducible connected reductive subgroup H_ℓ of GL_n over $\mathbf{F}_\ell$ such that $\bar{\Gamma}_\ell$ is a subgroup of index $\leq B(n)$ of $H_\ell(\mathbf{F}_\ell)$, where $B(n)$ does not depend on ℓ. By Proposition **??**, for all ℓ sufficiently

large, the rank of G_ℓ equals the rank of H_ℓ, and there exists a number field L such that for every sufficiently large ℓ, H_ℓ is split whenever ℓ splits completely in L.

Let $\tilde{G}_\ell$ and $\tilde{H}_\ell$ denote the simply connected cover of the derived group of G_ℓ and H_ℓ respectively. There exists an integer r depending only on n such that every element $g \in G_\ell(\mathbf{Q}_\ell)$ (resp. $h \in H_\ell(\mathbf{F}_\ell)$) can be written $g^{rn} = \det(g)^r g_0$ (resp. $h^{rn} = \det(h)h_0$) and g_0 (resp. h_0) lies in the derived group of $G_\ell(\mathbf{Q}_\ell)$ (resp. $H_\ell(\mathbf{F}_\ell)$). Thus g_0 lies in the image of $\tilde{G}_\ell(\mathbf{Q}_\ell) \to G_\ell(\mathbf{Q}_\ell)$, and likewise for h_0.

Step 2: Use the rather detailed classification theorems of representations of reductiv groups; results of Serre, Ribet, Larsen and Pink to conclude.