The Hasse principle for plane cubics

Wouter Zomervrucht, April 19, 2017

1. Introduction

These notes are based mostly on the preprint [1] by Manjul Bhargava.

Let K be a number field and X/K a smooth proper variety. If X has a K-rational point, then X also has a point over each completion of K. In other words, $X(K) \neq \varnothing$ implies $X(\mathbb{A}_K) \neq \varnothing$. We say that X satisfies the *Hasse principle* if the converse implication $X(\mathbb{A}_K) \neq \varnothing \Rightarrow X(K) \neq \varnothing$ holds. This leads to the following three possibilities.

- $X(K) = \varnothing$, $X(\mathbb{A}_K) = \varnothing$: the Hasse principle holds since X is *not locally soluble*,
- $X(K) \neq \varnothing$, $X(\mathbb{A}_K) \neq \varnothing$: the Hasse principle holds since X is *soluble*,
- $X(K) = \varnothing$, $X(\mathbb{A}_K) \neq \varnothing$: the Hasse principle *fails*.

Example 1.1. The Hasse principle is satisfied for

- Severi–Brauer varieties,
- quadrics in $\mathbb{P}^n$ for all $n \geq 1$,
- cubics in $\mathbb{P}^1$. ⧫

Example 1.2. The Hasse principle fails for

- cubics in $\mathbb{P}^3$ (which are degree 3 del Pezzo surfaces),
- cubics in $\mathbb{P}^2$,
- degree 4 hyperelliptic curves,
- complete intersections of two quadrics in $\mathbb{P}^3$. ⧫

The last three items are models of genus one curves. On the other hand, genus zero curves are Severi–Brauer varieties and hence do satisfy the Hasse principle. One may thus say that genus one curves are among the 'simplest' counterexamples to the Hasse principle. The earliest such counterexamples were found in 1940–2 for degree 4 hyperelliptic curves (Lind and Reichardt) and 1957 for plane cubics (Selmer). Besides isolated cases, more recently entire families of counterexamples have been constructed [8].

2. Arithmetic statistics

In the light of the preceding section, it is natural to ask how 'often' the Hasse principle fails. Let's make that question precise. For simplicity we restrict ourselves to the case $K = \mathbb{Q}$.

Let V be the parameter space of ternary cubic forms. It is isomorphic to the affine space $\mathbb{A}^{10}$. So if R is a ring, $V(R) \cong R^{10}$ is the space of ternary cubic forms with coefficients in R. We also fix a compact subset $B \subseteq V(\mathbb{R})$ that is the closure of an open neighborhood of the origin. Then for any $t > 0$ define

$$N(V, B, t) = \#\{f \in V(\mathbb{Z}) \cap tB\}.$$

Similarly, if Φ is a property of ternary cubic forms, set

$$N^{\Phi}(V, B, t) = \#\{f \in V(\mathbb{Z}) \cap tB \colon \Phi \text{ holds for } f\}.$$

In particular we shall use N^{ls}, N^{sol}, and N^{fail} respectively for ternary cubic forms that are locally soluble, that are soluble, and for which the Hasse principle fails.

Theorem 2.1 (Bhargava–Cremona–Fisher [2]). *One has*

$$\lim_{t\to\infty} \frac{N^{\mathrm{ls}}(V,B,t)}{N(V,B,t)} = \prod_p \Big(1 - \frac{p^9 - p^8 + p^6 - p^4 + p^3 + p^2 - 2p + 1}{3(p^2+1)(p^4+1)(p^6+p^3+1)}\Big) \approx 0.97.$$

Proof (sketch). The given factor for each p is the proportion of ternary cubics over $\mathbb{Z}_p$ that have a solution over $\mathbb{Q}_p$. It is computed by reduction to finite fields. There is no contribution from the infinite place as real plane cubics are always soluble. It remains to prove that solubility at the infinitely many different primes are 'independent' events. This is a consequence of Ekedahl's sieve. ■

Theorem 2.2 (Bhargava [1]). *One has*

$$\liminf_{t\to\infty} \frac{N^{\mathrm{fail}}(V,B,t)}{N(V,B,t)} > 0 \qquad \textit{and} \qquad \liminf_{t\to\infty} \frac{N^{\mathrm{sol}}(V,B,t)}{N(V,B,t)} > 0. \qquad \blacklozenge$$

Although conjecturally the limits exist and are independent of B, the explicit lower bounds found by Bhargava do depend on B. For suitably chosen B, Bhargava proves that the Hasse principle fails for a proportion of at least 28% of all ternary plane cubics.

Theorem 2.2 is conceptually more complicated than theorem 2.1, since the latter concerns only local behaviour of cubic forms, whereas the former involves global behaviour.

Conjecture 2.3 (Bhargava [1]). *One has*

$$\lim_{t\to\infty} \frac{N^{\mathrm{sol}}(V,B,t)}{N^{\mathrm{ls}}(V,B,t)} = \frac{1}{3}. \qquad \blacklozenge$$

Combined with theorem 2.1 this would imply that of all cubic forms, 3% is not locally soluble, $\frac{1}{3}\cdot 97\%$ is soluble, and $\frac{2}{3}\cdot 97\%$ does not satisfy the Hasse principle.

Remark 2.4. In the above results we allow non-smooth forms. This is not a real issue, however, since statistically 100% of all ternary cubic forms are smooth. ♦

Remark 2.5. As in later sections, we focus on plane cubics. Earlier we mentioned two further models of genus one curves: degree 4 hyperelliptic curves and complete intersections of two space quadrics. If one replaces V by the parameter space $V' \cong \mathbb{A}^5$ of binary quartic forms, or by the parameter space $V'' \cong \mathbb{A}^{20}$ of pairs of quaternary quadratic forms, results analogous to theorems 2.1 and 2.2 hold. The expected ratio in conjecture 2.3 is $1/4$ for both V' and V''. ♦

3. Selmer groups

Theorem 2.2 is proved by passing to the jacobians of the genus one curves in question. We first recall some theory.

Let K be a number field and E/K an elliptic curve. We know from the Mordell–Weil theorem that $E(K)$ is a finitely generated abelian group. Its rank is an important invariant. A common approach towards bounding the rank is as follows. The short exact sequence

$$0 \longrightarrow E[n] \longrightarrow E \xrightarrow{n\cdot} E \longrightarrow 0$$

in étale topology induces a commutative diagram in étale cohomology

$$\begin{array}{ccccccccc} 0 & \longrightarrow & E(K)/nE(K) & \longrightarrow & \mathrm{H}^1(K, E[n]) & \longrightarrow & \mathrm{H}^1(K, E)[n] & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & E(\mathbb{A}_K)/nE(\mathbb{A}_K) & \longrightarrow & \mathrm{H}^1(\mathbb{A}_K, E[n]) & \longrightarrow & \mathrm{H}^1(\mathbb{A}_K, E)[n] & \longrightarrow & 0 \end{array} \tag{3.1}$$

with exact rows.

Definition 3.2. Let $\mathrm{Sel}_n(E) = \ker(\mathrm{H}^1(K, E[n]) \to \mathrm{H}^1(\mathbb{A}_K, E)[n])$ be the *n-Selmer group* of E and $Ш(E) = \ker(\mathrm{H}^1(K, E) \to \mathrm{H}^1(\mathbb{A}_K, E))$ the *Tate–Shafarevich group* of E. ♦

We obtain a new short exact sequence

$$0 \longrightarrow E(K)/nE(K) \longrightarrow \mathrm{Sel}_n(E) \longrightarrow Ш(E)[n] \longrightarrow 0.$$

The Selmer group gives a bound on the rank of E.

Proposition 3.3. *If E has rank r, then $n^r \leq \#\mathrm{Sel}_n(E)$.* ■

Restricting to $K = \mathbb{Q}$, we may ask statistical questions as before. Given $E/\mathbb{Q}$ there are unique integers $a, b \in \mathbb{Z}$ such that E is isomorphic to the elliptic curve $E_{ab}\colon y^2 = x^3 + ax + b$ and a, b are minimal in the sense that for any prime p either $p^4 \nmid a$ or $p^6 \nmid b$. Then we say that E has *height* $\mathrm{ht}\, E = \max(|a|^3, b^2)$. If φ is a real-valued function on the set of (isomorphism classes of) elliptic curves over $\mathbb{Q}$, then we say that φ has *average value*

$$\mathrm{avg}(\varphi) = \lim_{h\to\infty} \frac{\sum_{\mathrm{ht}\, E \leq h} \varphi(E)}{\sum_{\mathrm{ht}\, E \leq h} 1}.$$

The *density* of a class of elliptic curves is the average value of its indicator function.

Conjecture 3.4 (Goldfeld–Katz–Sarnak). *The average rank of elliptic curves over $\mathbb{Q}$ is $1/2$. More precisely, the classes of rank 0 respectively rank 1 elliptic curves both have density $1/2$.* ♦

The following result is a major step towards this conjecture.

Theorem 3.5 (Bhargava–Shankar [3, 4, 5, 6]). *For elliptic curves over $\mathbb{Q}$,*
- *the average size of Sel_2 is 3,*
- *the average size of Sel_3 is 4,*
- *the average size of Sel_4 is 7,*
- *the average size of Sel_5 is 6.* ♦

Corollary 3.6. *One has $\mathrm{avg}(\mathrm{rk}) \leq 1.05$.*

Proof. If $E/\mathbb{Q}$ has rank r, then $20r - 15 \leq 5^r \leq \#\mathrm{Sel}_5(E)$. Since averaging is linear, we get $20\,\mathrm{avg}(\mathrm{rk}) - 15 \leq \mathrm{avg}(\#\mathrm{Sel}_5) = 6$, hence $\mathrm{avg}(\mathrm{rk}) \leq (6 + 15)/20 = 1.05$. ■

In fact, Bhargava–Shankar [6] prove the sharper bound $\mathrm{avg}(\mathrm{rk}) \leq 0.885$. It follows that the class of rank 0 elliptic curves has positive density. In the other direction, Bhargava–Skinner [7] prove that the class of rank 1 elliptic curves has positive density as well.

4. Selmer elements

As in the previous section, let K be a number field and E/K an elliptic curve.

Definition 4.1. An *n-diagram* for E is a map of K-varieties $C \to X$ that after a finite extension of K becomes isomorphic to the canonical morphism $E \to \mathbb{P}(\mathcal{L}(n \cdot 0))$. An n-diagram $C \to X$ is called *(locally) soluble* if C is. ♦

Proposition 4.2.
- $\mathrm{H}^1(K, E[n]) = \{$*$n$-diagrams for E*$\}$,
- $E(K)/nE(K) = \{$*soluble n-diagrams for E*$\}$,
- $\mathrm{Sel}_n(E) = \{$*locally soluble n-diagrams for E*$\}$.

Proof. An n-diagram is an étale twist of the trivial diagram $E \to \mathbb{P}(\mathcal{L}(n \cdot 0))$. Therefore the first item follows if we prove that $E \to \mathbb{P}(\mathcal{L}(n \cdot 0))$ has automorphism group scheme $E[n]$. But indeed, an automorphism of E (as variety) is a translation over a rational point $P \in E$. Such a translation extends to an automorphism of $\mathbb{P}(\mathcal{L}(n \cdot 0))$ if and only if $n \cdot P \sim n \cdot 0$, which holds if and only if $P \in E[n]$.

The map $\mathrm{H}^1(K, E[n]) \to \mathrm{H}^1(K, E)$ sends $C \to X$ to the E-torsor C. It is trivial if and only if C is soluble. The second and third item then follow from (3.1). ■

If $C \to X$ is a locally soluble n-diagram, then X is locally soluble as well. As X is moreover a Severi–Brauer variety, this implies $X \cong \mathbb{P}^{n-1}$. So an n-diagram is a genus one curve with a map to $\mathbb{P}^{n-1}$ whose image has degree n. For small n this situation has an arithmetic interpretation. A 2-diagram is a double cover of $\mathbb{P}^1$ ramified in 4 points, that is, a degree 4 hyperelliptic curve. As such it is given by a binary quartic form. A 3-diagram is a plane cubic, given by a ternary cubic form. A 4-diagram is a complete intersection of two space quadrics, given by two quaternary quadratic forms. A 5-diagram is an intersection of the five 4×4-subpfaffians of a skewsymmetric 5×5-matrix of quinary linear forms; it is given by a quintuple of alternating quinary quadratic forms.

For $n \geq 6$ no such description is available. That is the reason why theorem 3.5 is known for $n = 2$, 3, 4, and 5 only.

5. Ternary cubic forms

The three models of genus one curves discussed in the introduction correspond to elements of 2-, 3-, and 4-Selmer groups, respectively. In the following we consider only plane cubics and hence $n = 3$. Plane cubics are given by a ternary cubic form. The description of a 3-Selmer element by a ternary cubic form is unique up to linear change in variables.

Let $V \cong \mathbb{A}^{10}$ be the parameter space of ternary cubic forms. Attached to $f \in V$ are polynomial invariants a, b such that the jacobian of f is isomorphic to E_{ab}. Let $V_{ab} \subseteq V$ be the algebraic subspace of forms with given invariants $a, b \in \mathbb{Z}$. We obtain a diagram

$$\begin{array}{ccc} E_{ab}(\mathbb{Q})/3E_{ab}(\mathbb{Q}) & \hookrightarrow & \mathrm{Sel}_3(E_{ab}) \\ \| & & \| \\ V_{ab}(\mathbb{Q})^{\mathrm{sol}}/\mathrm{PGL}_3(\mathbb{Q}) & \hookrightarrow & V_{ab}(\mathbb{Q})^{\mathrm{ls}}/\mathrm{PGL}_3(\mathbb{Q}) \end{array}$$

where $^{\mathrm{sol}}$ and $^{\mathrm{ls}}$ indicate subsets of soluble respectively locally soluble forms.

Proof (of theorem 3.5 for $n = 3$, sketch). Every $\mathrm{PGL}_3(\mathbb{Q})$-orbit of $V_{ab}(\mathbb{Q})^{\mathrm{ls}}$ has a representative in $V_{ab}(\mathbb{Z})^{\mathrm{ls}}$. We need to count the average size of $V_{ab}(\mathbb{Z})^{\mathrm{ls}}/\mathrm{PGL}_3(\mathbb{Q})$.

Let L be a fundamental domain for the action of $\mathrm{PGL}_3(\mathbb{R})$ on $V(\mathbb{R})$. Let P be a fundamental domain for the action of $\mathrm{PGL}_3(\mathbb{Z})$ on $\mathrm{PGL}_3(\mathbb{R})$. The product $F = LP = \{lp\colon l \in L, p \in P\}$ is almost a fundamental domain for $\mathrm{PGL}_3(\mathbb{Z})$ acting on $V(\mathbb{R})$, but the orbit of $f \in V(\mathbb{R})$ is represented $[\mathrm{Stab}_{\mathrm{PGL}_3(\mathbb{R})}(f) : \mathrm{Stab}_{\mathrm{PGL}_3(\mathbb{Z})}(f)]$ times. That index takes only a few values, so by counting with appropriate weights we may pretend that F is a true fundamental domain.

Define the *height* of a form f with invariants a, b to be $\mathrm{ht}\, f = \max(|a|^3, b^2)$, i.e. the height of its jacobian. For $t > 0$ set

$$F_t = \{f \in F\colon \mathrm{ht}\, f \leq t^{12}\}.$$

Intuitively, the number of integral points in F_t should be roughly the volume of F_t. Unfortunately, that is false! While F_t does have finite volume, it is not bounded but has a cusp lingering off to infinity. The cusp contains many integral points. To remedy this, call an integral point $f \in V(\mathbb{Z})$ *generic* if it does not define the identity element of $\mathrm{Sel}_3(E)$. Using a technical averaging trick, one can show that integral points away from the cusp are typically generic, and integral points towards the cusp are typically non-generic. It follows that the number of generic integral points in F_t is roughly equal to $\mathrm{vol}(F_t)$.

At this point there are three more issues to be dealt with, all of which are essentially local in nature: we should replace $\mathrm{PGL}_3(\mathbb{Z})$-orbits by $\mathrm{PGL}_3(\mathbb{Q})$-orbits, we should restrict to locally soluble forms, and we should restrict to forms with minimal invariants. These items can be dealt with by counting with weights $w(f) = \prod_p w_p(f)$, where each w_p is given by congruence conditions at p-powers.

A careful computation now shows that the 3-Selmer groups has 3 non-identity elements on average, hence $\mathrm{avg}(\#\mathrm{Sel}_3) = 4$. ∎

6. Solubility of Selmer elements

In this section we deduce theorem 2.2 from theorem 3.5.

Theorem 6.1. *Let $n = 2$, 3, or 4. Of all n-Selmer elements, ordered by the height of their jacobian, a positive proportion is not soluble.*

Proof. The point of the proof is that (by the method in corollary 3.6) the 5-Selmer group gives a strictly better bound on the average rank of elliptic curves than the n-Selmer group. Indeed, as 100% of all elliptic curves have trivial rational n-torsion,

$$\begin{aligned}\mathrm{avg}(\#E(\mathbb{Q})/nE(\mathbb{Q})) = \mathrm{avg}(n^{\mathrm{rk}\, E}) &\leq \mathrm{avg}\Big(\frac{(5^{\mathrm{rk}\, E} - 5)(n^2 - n)}{20} + n\Big) \\ &\leq \mathrm{avg}\Big(\frac{(\#\mathrm{Sel}_5(E) - 5)(n^2 - n)}{20} + n\Big) = \frac{n^2 - n}{20} + n.\end{aligned}$$

The proportion of soluble n-Selmer elements, i.e. of n-Selmer elements that lie in the subgroup $\#E(\mathbb{Q})/nE(\mathbb{Q})$, is bounded above by

$$\frac{\mathrm{avg}(\#E(\mathbb{Q})/nE(\mathbb{Q}))}{\mathrm{avg}(\#\mathrm{Sel}_n(E))} \leq \frac{(n^2 - n)/20 + n}{\sigma(n)} < 1.$$

∎

Clearly also a positive proportion of non-identity n-Selmer elements is not soluble.

Theorem 6.2. *Let $n = 2$, 3, 4, or 5. Of all non-identity n-Selmer elements, ordered by the height of their jacobian, a positive proportion is soluble.*

Proof. Bhargava–Skinner [7] have shown that a positive proportion of all elliptic curves have rank at least 1. This implies that $\operatorname{avg}(n^{\operatorname{rk} E} - 1)$ is strictly positive. Therefore, the desired proportion is at least

$$\frac{\operatorname{avg}(\#E(\mathbb{Q})/nE(\mathbb{Q}) - 1)}{\operatorname{avg}(\#\operatorname{Sel}_n(E) - 1)} \geq \frac{\operatorname{avg}(n^{\operatorname{rk} E} - 1)}{\sigma(n) - 1} > 0. \qquad \blacksquare$$

Remark 6.3. Theorems 6.1 and 6.2 do not immediately imply the result of theorem 2.2. There are two issues to circumvene. First, in the above theorems we have ordered plane cubics by the height of their jacobian instead of the size of their coefficients. Secondly, many plane cubics may represent the same 3-Selmer element. ♦

We recall some notations from section 5. Let $V \cong \mathbb{A}^{10}$ be the parameter space of ternary cubic forms. For any $a, b \in \mathbb{Z}$ there is a bijection $\operatorname{Sel}_3(E_{ab}) = V_{ab}(\mathbb{Z})^{\mathrm{ls}}/\mathrm{PGL}_3(\mathbb{Q})$. Moreover, we have constructed a fundamental domain F for the action of $\mathrm{PGL}_3(\mathbb{Z})$ on $V(\mathbb{R})$. Denoting

$$F_t = \{f \in F \colon \operatorname{ht} f \leq t^{12}\}$$

we proved that the number of generic integral points in F_t is roughly equal to $\operatorname{vol}(F_t)$, which is approximately $c_1 t^{10}$ for some constant $c_1 > 0$.

Lemma 6.4. *There are about $c_2 t^{10}$ elliptic curves of height at most t^{12}, for some constant $c_2 > 0$.*

Proof. For the elliptic curve $y^2 = x^3 + ax + b$ to have height at most t^{12}, one must have $|a| \leq t^4$ and $|b| \leq t^6$, which yields $o(t^4) \cdot o(t^6) = o(t^{10})$ possibilities. $\blacksquare$

Proof (theorem 2.2). As $\operatorname{Sel}_3(E)$ has 3 non-identity elements on average, there are approximately $3c_2 t^{10}$ non-identity 3-Selmer elements of height at most t^{12}. This is precisely the number of locally soluble generic integral points in F_t. By theorem 6.1, a positive proportion $c_3 t^{10}$ of these is not soluble.

Now let $B \subseteq V(\mathbb{R})$ be a compact subset that is the closure of an open neighborhood of the origin. For $r \to \infty$ one has $\operatorname{vol}(F_t \cap rB) \to \operatorname{vol}(F_t)$. Hence there exists some $\delta > 0$ and $r \gg 0$, independent of t, such that

$$\frac{\operatorname{vol}(F_t \cap rtB)}{\operatorname{vol}(F_t)} \geq 1 - c_3/c_1 + \delta$$

for all t. That means that $F_t \cap rtB$ contains at least

$$(1 - c_3/c_1 + \delta)c_1 t^{10} = (c_1 - c_3 + \delta c_1)t^{10}$$

generic integral points. The total number of generic integral points in F_t is $c_1 t^{10}$, of which at least $c_3 t^{10}$ are locally soluble but not soluble. We conclude that $F_t \cap rtB$, and in particular rtB, contains at least $\delta c_1 t^{10}$ locally soluble non-soluble generic integral points. Hence

$$\lim_{t\to\infty} \frac{N^{\mathrm{fail}}(V, B, t)}{N(V, B, t)} \geq \frac{\delta c_1 (t/r)^{10}}{\operatorname{vol}(B) t^{10}} = \frac{\delta c_1}{r^{10} \operatorname{vol}(B)} > 0.$$

The same proof works for the proportion of soluble plane cubics, applying theorem 6.2 instead of 6.1. $\blacksquare$

References

[1] M. Bhargava, *A positive proportion of plane cubics fail the Hasse principle*. Preprint, 2014.

[2] M. Bhargava, J. Cremona and T. Fisher, *The proportion of plane cubic curves over* $\mathbb{Q}$ *that everywhere locally have a point*. Preprint, 2013.

[3] M. Bhargava and A. Shankar, *Binary quartic forms having bounded invariants, and the boundedness of the average rank of elliptic curves*. Preprint, 2010.

[4] M. Bhargava and A. Shankar, *Ternary cubic forms having bounded invariants, and the existence of a positive proportion of elliptic curves having rank* 0. Preprint, 2010.

[5] M. Bhargava and A. Shankar, *The average number of elements in the* 4*-Selmer groups of elliptic curves is* 7. Preprint, 2013.

[6] M. Bhargava and A. Shankar, *The average size of the* 5*-Selmer group of elliptic curves is* 6*, and the average rank is less than* 1. Preprint, 2013.

[7] M. Bhargava and C. Skinner, *A positive proportion of elliptic curves over* $\mathbb{Q}$ *have rank one*. Preprint, 2014.

[8] J.-L. Colliot-Thélène and B. Poonen, *Algebraic families of nonzero elements of Shafarevich–Tate groups*. Journal of the American Mathematical Society 13, pp. 83–99, 2000.