The first-order theory of finitely generated fields

Bjorn Poonen

University of California at Berkeley

(on work of Robinson, Ershov, Rumely, Pop, myself, and Scanlon)

Antalya Algebra Days X
June 1, 2008

Distinguishing fields

Example

The first-order sentence

$$(\exists x)(\exists y)\ x^2 + y^2 = -1$$

holds for every finite field, and hence for every field of positive characterstic. But it is false for $\mathbb{Q}$.

To what extent can we distinguish fields by the truth values of first-order sentences?

Some fields cannot be distinguished by first-order sentences.

Example

Any first-order sentence true for $\mathbb{C}$ is true also for any algebraically closed field of characteristic 0.

Corollary (Lefschetz principle)

Many theorems of algebraic geometry proved over $\mathbb{C}$ using analytic methods automatically transfer to any algebraically closed field of characteristic 0.

Example

Elementary model theory shows that any first-order sentence true for $\mathbb{C}$ holds also for any algebraically closed field of sufficiently large positive characteristic (depending on the sentence).

To hope to be able to distinguish fields, we must restrict the class of fields considered.

Finitely generated fields

Every field K has a minimal subfield, isomorphic to either $\mathbb{Q}$ or $\mathbb{F}_p$ for some prime p.

Definition

Call K finitely generated (f.g.) if it is finitely generated as a field extension of its minimal subfield.

- F.g. fields arise naturally as the function fields of varieties over number fields and finite fields.
- The theory of transcendence bases shows that every f.g. field is a finite extension of $\mathbb{Q}(t_1, \ldots, t_n)$ or $\mathbb{F}_p(t_1, \ldots, t_n)$, where n is the absolute transcendence degree.

Three questions about the richness of the arithmetic of f.g. fields

Question (Sabbagh 1980s, Pop 2002)

Given non-isomorphic f.g. fields K and L, is there a sentence that is true for K and false for L?

(Previously Duret had asked for the analogue for f.g. extensions of algebraically closed fields, and he and later Pierce proved some cases of this.)

Question (P. 2007)

Given a f.g. field K, is there a sentence that is true for K and false for all f.g. fields not isomorphic to K?

Question (P. 2007)

Is every reasonable class *of infinite f.g. fields cut out by a single sentence?*

Three questions about the richness of the arithmetic of f.g. fields

Question (Sabbagh 1980s, Pop 2002, proved by Scanlon)

Given non-isomorphic f.g. fields K and L, is there a sentence that is true for K and false for L?

(Previously Duret had asked for the analogue for f.g. extensions of algebraically closed fields, and he and later Pierce proved some cases of this.)

Question (P. 2007, proved by Scanlon)

Given a f.g. field K, is there a sentence that is true for K and false for all f.g. fields not isomorphic to K?

Conjecture (P. 2007, still open)

Every reasonable class *of infinite f.g. fields is cut out by a single sentence.*

Reasonable classes of f.g. fields

Definition (P. 2007, idea due to Hrushovski)

Choose a natural bijection between a recursive $A \subset \mathbb{N}$ and $\{(r, f_1, \ldots, f_m) : r \in \mathbb{N}, f_1, \ldots, f_m \in \mathbb{Z}[x_1, \ldots, x_r]\}$.
Construction of $\mathbb{Z}[x_1, \ldots, x_r]/(f_1, \ldots, f_m)$ yields a map

$$A \to \{\text{f.g. } \mathbb{Z}\text{-algebras}\}.$$

The set of $a \in A$ such that the corresponding $\mathbb{Z}$-algebra is a domain is a recursive subset $B \subset A$. Construction of the fraction field yields a map

$$B \to \{\text{isomorphism classes of f.g. fields}\}.$$

For any class of f.g. fields, define the set of $b \in B$ for which the corresponding field belongs to the class. Call the class reasonable if this subset of B is a first-order definable subset of $(\mathbb{N}, +, \cdot)$.

Rumely's work on the global field case

Definition

A global field is a field of one of the following types:

1. number field: finite extension of $\mathbb{Q}$
2. global function field: f.g. extension of transcendence degree 1 over some $\mathbb{F}_p$.

Rumely gave a positive answer to all three questions restricted to global fields.

- A key step was to build on work of Robinson and Ershov to give uniform definitions of the family of valuation subrings.
- Then, for example, he could distinguish number fields from global function fields by using a first-order sentence saying

"The intersection of all valuation rings in K is not a field."

Quadratic forms

Definition

A quadratic form over a field k in $x_1, \ldots, x_n$ is a homogeneous polynomial of degree 2 in $k[x_1, \ldots, x_n]$.

Definition

A quadratic form $q(x_1, \ldots, x_n)$ is isotropic if there exist $a_1, \ldots, a_n \in k$ not all 0 such that

$$q(a_1, \ldots, a_n) = 0.$$

It is called anisotropic otherwise.

Example

$x^2 - 3y^2$ over $\mathbb{Q}$ is anisotropic.

Example

$x_1^2 + x_2^2 + x_3^2$ over $\mathbb{Q}$ is anisotropic.

Quadratic forms

Definition

A quadratic form over a field k in $x_1, \ldots, x_n$ is a homogeneous polynomial of degree 2 in $k[x_1, \ldots, x_n]$.

Definition

A quadratic form $q(x_1, \ldots, x_n)$ is isotropic if there exist $a_1, \ldots, a_n \in k$ not all 0 such that

$$q(a_1, \ldots, a_n) = 0.$$

It is called anisotropic otherwise.

Example

$x^2 - 3y^2$ over $\mathbb{Q}$ is anisotropic.

Example

$x_1^2 + x_2^2 + x_3^2 + x_4^2$ over $\mathbb{Q}$ is anisotropic.

Quadratic forms

Definition

A quadratic form over a field k in $x_1, \ldots, x_n$ is a homogeneous polynomial of degree 2 in $k[x_1, \ldots, x_n]$.

Definition

A quadratic form $q(x_1, \ldots, x_n)$ is isotropic if there exist $a_1, \ldots, a_n \in k$ not all 0 such that

$$q(a_1, \ldots, a_n) = 0.$$

It is called anisotropic otherwise.

Example

$x^2 - 3y^2$ over $\mathbb{Q}$ is anisotropic.

Example

$x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 \quad$ over $\mathbb{Q}$ is anisotropic.

Quadratic forms

Definition

A quadratic form over a field k in $x_1, \ldots, x_n$ is a homogeneous polynomial of degree 2 in $k[x_1, \ldots, x_n]$.

Definition

A quadratic form $q(x_1, \ldots, x_n)$ is isotropic if there exist $a_1, \ldots, a_n \in k$ not all 0 such that

$$q(a_1, \ldots, a_n) = 0.$$

It is called anisotropic otherwise.

Example

$x^2 - 3y^2$ over $\mathbb{Q}$ is anisotropic.

Example

$x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2 \quad$ over $\mathbb{Q}$ is anisotropic.

Quadratic forms

Definition

A quadratic form over a field k in $x_1, \ldots, x_n$ is a homogeneous polynomial of degree 2 in $k[x_1, \ldots, x_n]$.

Definition

A quadratic form $q(x_1, \ldots, x_n)$ is isotropic if there exist $a_1, \ldots, a_n \in k$ not all 0 such that

$$q(a_1, \ldots, a_n) = 0.$$

It is called anisotropic otherwise.

Example

$x^2 - 3y^2$ over $\mathbb{Q}$ is anisotropic.

Example

$x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2 + x_7^2$ over $\mathbb{Q}$ is anisotropic.

Quadratic forms

Definition

A quadratic form over a field k in $x_1, \ldots, x_n$ is a homogeneous polynomial of degree 2 in $k[x_1, \ldots, x_n]$.

Definition

A quadratic form $q(x_1, \ldots, x_n)$ is isotropic if there exist $a_1, \ldots, a_n \in k$ not all 0 such that

$$q(a_1, \ldots, a_n) = 0.$$

It is called anisotropic otherwise.

Example

$x^2 - 3y^2$ over $\mathbb{Q}$ is anisotropic.

Example

$x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2 + x_7^2 + x_8^2$ over $\mathbb{Q}$ is anisotropic.

Building anisotropic forms over function fields

Example

For any field k, the form $x^2 + ty^2$ over $k(t)$ is anisotropic.

Lemma

If $q(x_1, \ldots, x_n)$ is anisotropic over k, then the form

$$q(x_1, \ldots, x_n) + tq(y_1, \ldots, y_n)$$

(in twice as many variables) is anisotropic over $k(t)$.

Sketch of proof.

The t-adic valuations of the two halves are even and odd, respectively, so they cannot cancel. $\square$

Example

The form $(x_1^2 + t_1x_2^2) + t_2(y_1^2 + t_1y_2^2)$ over $k(t_1, t_2)$ is anisotropic.

Pfister forms

$$\langle\langle a \rangle\rangle := x_1^2 + a x_2^2$$
$$\langle\langle a, b \rangle\rangle := x_1^2 + a x_2^2 + b x_3^2 + ab x_4^2$$
$$\langle\langle a, b, c \rangle\rangle := x_1^2 + a x_2^2 + b x_3^2 + ab x_4^2 + c x_5^2 + ac x_6^2 + bc x_7^2 + abc x_8^2$$

Example

For any field k, the Pfister form $\langle\langle t_1, \ldots, t_n \rangle\rangle$ over $k(t_1, \ldots, t_n)$ is anisotropic.

Example

If $-c$ is a nonsquare in $\mathbb{F}_p$, then $\langle\langle c, t_1, \ldots, t_n \rangle\rangle$ is anisotropic over $\mathbb{F}_p(t_1, \ldots, t_n)$.

Example

$\langle\langle 2, 5, t_1, \ldots, t_n \rangle\rangle$ is anisotropic over $\mathbb{Q}(\sqrt{-1})(t_1, \ldots, t_n)$.

Kronecker dimension is first-order

Definition

Let K be a f.g. field. The Kronecker dimension is defined by

$$\operatorname{Kr.dim} K := \begin{cases} \operatorname{tr.deg}(K/\mathbb{F}_p), & \text{if } \operatorname{char} K = p > 0, \\ \operatorname{tr.deg}(K/\mathbb{Q}) + 1, & \text{if } \operatorname{char} K = 0. \end{cases}$$

Theorem (Voevodsky, Rost, etc.)

Let K be a f.g. field of characteristic not 2 such that $\sqrt{-1} \in K$. Then the largest m for which there exists an anisotropic m-fold Pfister form $\langle\langle a_1, \ldots, a_m \rangle\rangle$ is $\operatorname{Kr.dim} K + 1$.

If K is f.g. and $\operatorname{char} K = 2$, then $[K : K^2] = 2^{\operatorname{tr.deg}(K/\mathbb{F}_2)}$.

Corollary (Pop 2002)

For each $n \in \mathbb{N}$, there exists a first-order sentence ϕ_n that holds for a f.g. field K if and only if $\operatorname{Kr.dim} K = n$.

Defining the constant field

Definition

The constant field k in a f.g. field K is the relative algebraic closure of the minimal subfield in K.

Depending on the characteristic, k is either a finite field or a number field.

Theorem (P. 2007)

There is a first-order formula $\psi(x)$ that in any f.g. field K defines the constant field.

As we know already, we can use Pfister forms to distinguish number fields from finite fields, so we get:

Corollary (P. 2007)

There is a first-order sentence that for f.g. fields K is true if and only if $\operatorname{char} K = 0$.

Theorem (from previous slide)

There is a first-order formula $\psi(x)$ that in any f.g. field K defines the constant field.

Outline of proof in the case char $K = 0$ and $\sqrt{-1} \in K$.

1. There exists an elliptic curve over k such that $E(k)$ is infinite and $E(K) = E(k)$.
2. There exists an infinite subset S of k that is definable in K. (Actually for the next step we need additional constraints on S.)
3. An element $t \in K$ belongs to k *if and only if* $\langle\langle s_1, s_2, t - s_3 \rangle\rangle$ is isotropic for all $s_1, s_2, s_3 \in S$. (One direction is easy: if $t \in k$, then $s_1, s_2, t - s_3$ all belong to the number field k, but $\operatorname{Kr.dim} k + 1$ is only 2, so $\langle\langle s_1, s_2, t - s_3 \rangle\rangle$ must be isotropic.) $\square$

(And all this can be done uniformly, i.e., with first-order formulas independent of K. The char $K = p$ case, which we are omitting, is much more difficult.)

Defining algebraic dependence

The result

Theorem (from two slides ago)

There is a first-order formula $\psi(x)$ that in any f.g. field K defines the constant field.

is the $n = 1$ case of

Theorem (P. 2007)

For each n, there exists a first-order formula $\psi_n(x_1, \ldots, x_n)$ such that for any f.g. field K and any $a_1, \ldots, a_n \in K$, the statement $\psi_n(a_1, \ldots, a_n)$ holds if and only if $a_1, \ldots, a_n$ are algebraically dependent over k.

The proof is a generalization of the $n = 1$ case.

Scanlon's theorem

Recall that we want to prove:

Theorem (Scanlon)

Given a f.g. field K, there is a first-order sentence that is true for K and false for every f.g. field not isomorphic to K.

Before Scanlon's work, it was understood by the experts that it would suffice to prove the uniform definability of the family of valuation rings in K, as Rumely had done for global fields.

Scanlon had two insights that enabled him to build upon the previous work to complete the proof:

- The first, which ultimately is more a matter of elegance and convenience than a key element of the proof, is that it suffices to show that $\mathbb{N}$ (with $+$ and $\cdot$) is bi-interpretable in K (with parameters).
- The second, which is the key, is that a "local-global theorem for Brauer groups of function fields", due to Faddeev in characteristic 0 and generalized to arbitrary characteristic by Auslander and Brumer, can be adapted to prove uniform definability of a family of enough valuation rings of K to prove the bi-interpretability.

Bi-interpretability

To say that $\mathbb{N}$ is bi-interpretable in K (with parameters) means all of the following:

1a. There is an interpretation of $\mathbb{N}$ in K: i.e., there is a definable subset $\mathbb{N}_K$ of K^m for some m with a bijection $\mathbb{N}_K \to \mathbb{N}$ such that the subsets of $(\mathbb{N}_K)^3$ corresponding to the graphs of $+$ and $\cdot$ (subsets of $\mathbb{N}^3$) are definable subsets of K^{3m}. (The actual definition is slightly more general: it allows $\mathbb{N}_K \to \mathbb{N}$ to be a *surjection* inducing a definable equivalence relation.)

1b. There is an interpretation of K in $\mathbb{N}$.

2a. Let $K_{\mathbb{N}_K}$ be the interpreted copy of K in K obtained by composing the interpretations in 1a and 1b; then the identification $K_{\mathbb{N}_K} \to K$ is K-definable.

2b. The analogous identification $\mathbb{N}_{K_\mathbb{N}} \to \mathbb{N}$ is $\mathbb{N}$-definable.

Proof of Scanlon's theorem, *assuming bi-interpretability*.

- The uniform definition of algebraic dependence (P. 2007) lets one uniformly interpret some global field in each infinite f.g. field (with parameters).
- Combining this with the uniform interpretation of $\mathbb{N}$ in global fields (Rumely) gives a uniform interpretation $\mathbb{N}_L$ of $\mathbb{N}$ in all infinite f.g. fields L.
- There are formulas defining a copy $K_{\mathbb{N}}$ (with $+$ and $\cdot$) of K in $\mathbb{N}$ (recursion).
- Transferring this from $\mathbb{N}$ to $\mathbb{N}_L$ yields a uniformly L-definable field $K_{\mathbb{N}_L}$, still isomorphic to K.
- The interpretations $K_{\mathbb{N}}$ and $\mathbb{N}_K$ are those in Scanlon's bi-interpretation; in particular, there is a formula η that over K defines an isomorphism $K_{\mathbb{N}_K} \to K$.
- Use the sentence that over L says that for some values of the parameters η defines an isomorphism $K_{\mathbb{N}_L} \to L$: this sentence is true for $L = K$, but not for any other f.g. field L! $\square$

Outline of Scanlon's proof of bi-interpretability.

- The proof is by induction on $n := \operatorname{tr.deg}(K/k)$, with Rumely's work as the base case.
- Choose algebraically independent $t_1, \ldots, t_{n-1}$ and let K_1 be the set of $\alpha \in K$ such that $t_1, \ldots, t_{n-1}, \alpha$ are algebraically dependent.
  (K_1 is definable in K, by P. 2007.)
  Then K_1 is a field with $\operatorname{tr.deg}(K_1/k) = n - 1$, and K is the function field of a curve over K_1.
- Adapt Fadeev-Auslander-Brumer to construct a definable family consisting of most of the valuation rings between K_1 and K.
- more work...

$\square$

Some non-uniformity

- There is no definition of $\mathbb{F}_p$ in $\mathbb{F}_{p^2}$ that is uniform in p. (This follows from work of Chatzidakis, Macintyre, and van den Dries 1992.)
- There is no uniform existential definition of the constant field in f.g. fields, even if one fixes the characteristic.

Uniform bi-interpretability?

- The formulas constructed by Rumely and P. are uniform; i.e., the same formulas work for all f.g. fields K.
- But as Scanlon points out, some parts of his proof are not uniform.
- If the bi-interpretability statement could be made uniform, one could prove the remaining conjecture, that every reasonable class of infinite f.g. fields can be cut out by a single sentence.