The Borel Complexity of the Class of Models of First-order Theories

Hongyu Zhu

University of Wisconsin-Madison

Joint work with Uri Andrews, David Gonzalez, Steffen Lempp, and Dino Rossegger.
With thanks to Ali Enayat, Roman Kossak, and Albert Visser.

Table of Contents

Motivation

Question. How complicated is (true) arithmetic?

What could be an answer?

1. Axiomatization:
   - Incompleteness theorem
   - Undefinability of truth
2. Computing a model:
   - Tennenbaum's theorem
   - Non-standard models of PA have infinite Scott rank (Montalbán-Rossegger)
3. Distinguishing the models:
   - Separate Mod(T) from other structures.

Setting

Assume the language $\mathcal{L}$ is (at most) countable and relational. All structures will be assumed to have domain ω. Fix an enumeration $\{\varphi_i\}_{i\in\omega}$ of atomic $\mathcal{L}\cup\omega$-sentences (using ω as constants).

For any $\mathcal{L}$-structure $\mathcal{M}$, its atomic diagram can be encoded by the path $p_{\mathcal{M}} \in 2^\omega$, where $p_{\mathcal{M}}(i) = 1 \iff \mathcal{M} \vDash \varphi_i$ and 0 otherwise. Identify $\mathcal{M}$ with $p_{\mathcal{M}}$.

Throughout, let T be a (consistent) first-order $\mathcal{L}$-theory. Let $\mathrm{Mod}(T) \subseteq 2^\omega$ be the set of all countable models of T. We can analyze its complexity under the usual Polish topology on 2^ω, and it turns out to be always Borel.

Question. What is the relationship between the complexity of $\mathrm{Mod}(T)$ and that of T?

To answer that, we need to understand what "complexity" means in both contexts.

The Borel Hierarchy

Definition

- $\mathbf{\Sigma}^0_1$ = open sets, $\mathbf{\Pi}^0_1$ = closed sets, $\mathbf{\Delta}^0_1$ = clopen sets
- $\mathbf{\Sigma}^0_\alpha$ = countable union of $\mathbf{\Pi}^0_{<\alpha}$ sets, $\mathbf{\Pi}^0_\alpha = \neg\mathbf{\Sigma}^0_\alpha$ = countable intersection of $\mathbf{\Sigma}^0_{<\alpha}$ sets, $\mathbf{\Delta}^0_\alpha = \mathbf{\Sigma}^0_\alpha \cap \mathbf{\Pi}^0_\alpha$. $(\alpha > 1)$
- $\bigcup_{\alpha<\omega_1} \mathbf{\Pi}^0_\alpha$ = Borel.

An important connection to computability theory:

Theorem

For any recursive ordinal α*,*

$$\mathbf{\Sigma}^0_\alpha = \bigcup_{X\in 2^\omega} \Sigma_\alpha(X), \mathbf{\Pi}^0_\alpha = \bigcup_{X\in 2^\omega} \Pi_\alpha(X), \mathbf{\Delta}^0_\alpha = \bigcup_{X\in 2^\omega} \Delta_\alpha(X).$$

Basically, we can effectively recover a Borel set from its code.

Continuous Reduction

The following notions helps us compare the complexities of Borel sets.

Definition

$f : 2^\omega \to 2^\omega$ is a *continuous reduction* from $A \subseteq 2^\omega$ to $B \subseteq 2^\omega$ if f is continuous and $x \in A \iff f(x) \in B$.

Definition

For a pointclass $\mathbf{\Gamma} \subseteq 2^\omega$, say A is $\mathbf{\Gamma}$-*hard* if every $B \in \mathbf{\Gamma}$ continuously reduces to A (i.e. there is a continuous reduction f from A to B); and A is $\mathbf{\Gamma}$-*complete* if in addition $A \in \mathbf{\Gamma}$.

Again, we have an effective characterization of the continuous functions.

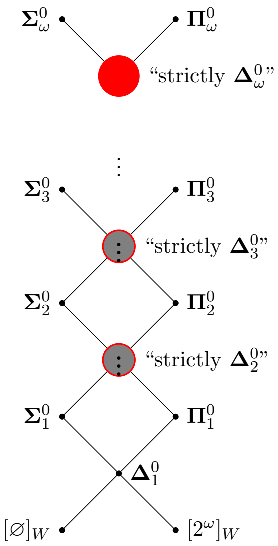
Theorem

$f : 2^\omega \to 2^\omega$ is continuous if and only if it is computable relative to some oracle $X \in 2^\omega$.

The Borel Hierarchy

Here is a picture of Borel sets below $\mathbf{\Pi}^0_\omega$ and $\mathbf{\Sigma}^0_\omega$, under continuous reduction.

Warning: Dangerous notation! For every pointclass $\mathbf{\Gamma}$ on this slide (like $\mathbf{\Sigma}^0_\alpha, \mathbf{\Pi}^0_\alpha$), it is more accurate to say "$\mathbf{\Gamma}$-complete," since $\mathbf{\Sigma}^0_\alpha, \mathbf{\Pi}^0_\alpha, \mathbf{\Delta}^0_\alpha$ are actually downward closed.



Quantifier Alternation Hierarchy

We have a similar hierarchy on the formula side.

Definition

- $\exists_1 = \{\exists \overline{x} R(\overline{x}) : R \text{ quantifier-free}\}$, $\forall_1 = \{\forall \overline{x} R(\overline{x}) : R \text{ quantifier-free}\}$
- $\exists_{n+1} = \{\exists \overline{x} R(\overline{x}) : R \in \forall_n\}$, $\forall_{n+1} = \{\forall \overline{x} R(\overline{x}) : R \in \exists_n\}$

Remark. In the context of arithmetic this is different from Σ^0_n, Π^0_n due to bounded quantifiers. However, over PA, it's safe to use either as long as $n \geq 1$ (by MRDP).

By induction, $\mathrm{Mod}(\varphi)$ is $\mathbf{\Pi}^0_n$ if $\varphi \in \forall_n$. A converse to this is formulated using infinitary logic.

Infinitary Logic

The infinitary logic $\mathcal{L}_{\omega_1\omega}$ is obtained from first-order logic by allowing countable conjunctions and disjunctions (but still referring to only finitely many free variables in a single formula). To form a similar hierarchy, we treat countable conjunctions like universal quantifiers (and similarly for countable disjunctions), resulting in the classes Π^{in}_α and $\Sigma^{\mathrm{in}}_\alpha$.

Theorem (López-Escobar; Vaught)

If $A \subseteq 2^\omega$ is isomorphism-invariant (in particular if $A = \mathrm{Mod}(\varphi)$ for some sentence φ), then it is $\boldsymbol{\Pi}^0_\alpha$ iff it is $\mathrm{Mod}(\psi)$ for some $\psi \in \Pi^{in}_\alpha$.

This gives an upper bound of $\mathrm{Mod}(T)$'s complexity.

Corollary

- $\mathrm{Mod}(T)$ *is always* $\boldsymbol{\Pi}^0_\omega$.
- *If T is axiomatizable by $\forall_n$ (or $\exists_{n-1}$) sentences, then* $\mathrm{Mod}(T)$ *is* $\boldsymbol{\Pi}^0_n$.

The Result for Arithmetic

Theorem (Andrews, Gonzalez, Lempp, Rossegger, Z.)

Mod(TA) *and* Mod(PA) *are both* $\mathbf{\Pi}^0_\omega$*-complete.*

Idea. Given a $\mathbf{\Pi}^0_\omega$ set $P = \bigcap_n P_n$ (where P_n is $\mathbf{\Pi}^0_n$), we will build a continuous reduction f such that $f(p) \models \mathrm{TA}$ when $p \in P$ and $f(p) \not\models \mathrm{PA}$ when $p \notin P$. Thus f witnesses the $\mathbf{\Pi}^0_\omega$-hardness for both Mod(TA) and Mod(PA).

Fact

For each $n \geq 1$, $\mathrm{Th}_{\exists_n}(\mathrm{TA}) + \neg \mathrm{B}\Sigma_n$ *is consistent.*

Take T_n to be any completion of the above theory.

Construction. Starting from $n = 1$, check if $p \in P_n$. If not, proceed to make $f(p)$ a model of T_n (and stop). Otherwise, make the "$\exists_n$-fragment" of $f(p)$ satisfy $\mathrm{Th}_{\exists_n}(\mathrm{TA})$, increment n, and repeat.

General Unbounded Theories

Key. Even if we have committed to a "complete" $\exists_n$-fragment, there is still a "switch" on a higher level that we can turn off to deviate from the target theory.

This allows us to generalize the above argument to a large class of theories.

Definition (Boundedly Axiomatizable Theory)

A theory T is *boundedly axiomatizable* (*bounded* for short) if for some $n < \omega$, T has an axiomatization consisting entirely of $\forall_n$ sentences. (T is *unbounded* otherwise.)

Theorem (AGLRZ)

If T is complete, then T is unbounded $\iff$ $\mathrm{Mod}(T)$ *is $\mathbf{\Pi}^0_\omega$-complete.*

Remark. The assumption of completeness is necessary: there are unbounded theories with a (strictly) $\mathbf{\Delta}^0_\omega$ class of models.

The Unbounded Case

Proof. ($\Leftarrow$) If T is $\forall_n$-axiomatizable then $\mathrm{Mod}(T)$ is $\mathbf{\Pi}^0_n$, thus not $\mathbf{\Pi}^0_\omega$-hard.
($\Rightarrow$) Follow the outline, and check two things.

1. The construction is possible.
   In general, checking whether $p \in P_n$ cannot be done continuously. However, we can approximate it effectively (as we approximate the theory):

Theorem (Solovay; Knight)

Let T be a complete theory. Suppose $R \leq_T X$ is an enumeration of a Scott set $\mathcal{S}$, with functions t_n, $\Delta^0_n(X)$ uniformly in n, such that: for each n, $\lim_s t_n(s)$ is an R-index for $\mathrm{Th}_{\exists_n}(T)$; and for all s, $t_n(s)$ is an R-index for a subset of $\mathrm{Th}_{\exists_n}(T)$. Then X can compute a model $\mathcal{M} \vDash T$ representing $\mathcal{S}$.

   We make extensive use of the uniformity of the above theorem.

2. The desired complete theories T_n's exist.
   We just need $\mathrm{Th}_{\exists_n}(T) = \mathrm{Th}_{\exists_n}(T_n)$ and $T \neq T_n$. If this is not possible, then T is axiomatized by $\mathrm{Th}_{\exists_n}(T)$, a contradiction. □

Examples: Sequential Theories

Corollary

If T is a completion of PA, *then* $\mathrm{Mod}(T)$ *is* $\mathbf{\Pi}^0_\omega$*-complete.*

Proof.

No consistent extension of PA is bounded (Rabin). □

It turns out that there is a large class of unbounded theories:

A theory is *sequential* if, roughly speaking, it is able to encode finite sequences.
(It directly interprets *adjunctive set theory*, i.e. $\varnothing$ exists; and for all x, y, we have $x \cup \{y\}$ exists.)

Theorem (Enayat, Visser)

Every complete sequential theory in a finite language is unbounded.

Examples of sequential theories include $\mathrm{PA}, \mathrm{PA}^-, \mathrm{ZF}$, etc. and their extensions.

The Bounded Case

We can also analyze $\mathrm{Mod}(T)$ when T is bounded, even if it is incomplete:

Theorem (AGLRZ)

For any $n \in \omega$ and any theory T (not necessarily complete):

1. *T is not $\forall_n$-axiomatizable $\Rightarrow \mathrm{Mod}(T)$ is $\mathbf{\Sigma}^0_n$-hard. Thus,*
   T is $\forall_n$-axiomatizable $\iff \mathrm{Mod}(T) \in \mathbf{\Pi}^0_n$.
2. *T is not $\exists_n$-axiomatizable $\Rightarrow \mathrm{Mod}(T)$ is $\mathbf{\Pi}^0_n$-hard. Thus,*
   T is $\exists_n$-axiomatizable $\Leftarrow \mathrm{Mod}(T) \in \mathbf{\Sigma}^0_n$.
3. *T is $\forall_n$- but not $\exists_n$-axiomatizable $\iff \mathrm{Mod}(T)$ is $\mathbf{\Pi}^0_n$-complete.*
4. *T is $\exists_n$-axiomatizable $\Rightarrow \mathrm{Mod}(T) \in \mathbf{\Pi}^0_{n+1}$.*

(3) is immediate from (1) (2), and (4) follows from López-Escobar, so the hard work goes in (1) and (2) (which are proved similarly).

Tightness

Examples showing the $\exists_n$ result is "tight":

Remark. Using Marker's extension, one can make these work for larger values of n.

Example

Let $\mathcal{L}$ consist of just one unary relation P, and T says P is infinite and coinfinite. Then T is $\exists_1$-axiomatizable and $\aleph_0$-categorical (thus complete). $\mathrm{Mod}(T)$ is $\mathbf{\Pi}^0_2$-complete. [In fact, by our convention, $\mathrm{Mod}(T) \in \mathbf{\Sigma}^0_2 \Rightarrow \mathrm{Mod}(T) = \varnothing$.]

Example

$T = \mathrm{Th}(2 \cdot \mathbb{Q} + 1 + \mathbb{Q}, <, S)$ is axiomatizable by a single $\exists_3$ sentence and $\aleph_0$-categorical. $\mathrm{Mod}(T)$ is $\mathbf{\Sigma}^0_3$-complete.

Example

Use a 2-sorted language to combine a $\exists_2 - \mathbf{\Pi}^0_3$ example and a $\exists_3 - \mathbf{\Sigma}^0_3$ example: this gives a $\exists_3 - \mathbf{\Delta}^0_4$ (strict) example.

Proof of the Bounded Case

Lemma

Suppose $n \in \omega$, and $T^+ \neq T^-$ are complete theories with $\mathrm{Th}_{\exists_n}(T^-) \subseteq \mathrm{Th}_{\exists_n}(T^+)$. Then for any $P \in \mathbf{\Sigma}^0_n$, there is a continuous reduction f such that $f(x) \in \mathrm{Mod}(T^+)$ if $x \in X$, and $f(x) \in \mathrm{Mod}(T^-)$ otherwise.
In particular, $\mathrm{Mod}(T^+)$ is $\mathbf{\Sigma}^0_n$-hard, and $\mathrm{Mod}(T^-)$ is $\mathbf{\Pi}^0_n$-hard.

Remark. This is like stopping the construction for unbounded theories at level n.

Proof.

Given $p \in 2^\omega$, Feed these ingredients to Solovay's theorem:

- R: comes from a fixed oracle.
- t_k for $k < n$: output a fixed R-index of $\mathrm{Th}_{\exists_k}(T^-) = \mathrm{Th}_{\exists_k}(T^+)$.
- t_n: check whether $p \in P$ using $p^{(n-1)}$; keep outputting index of $\mathrm{Th}_{\exists_n}(T^-)$ until the Σ_n outcome (i.e. witness $p \in P$), then switch to $\mathrm{Th}_{\exists_n}(T^+)$.
- t_k for $k > n$: compute membership and then output the correct index. □

Proof of the Bounded Case

Lemma

Let Γ be a family of sentences closed under conjunction and disjunction, A a set of sentences, and φ a sentence such that $A \nvdash \neg\varphi \leftrightarrow \psi$ for any $\psi \in \Gamma$. Then there are complete theories $T^+ \supseteq A \cup \{\varphi\}$, $T^- \supseteq A \cup \{\neg\varphi\}$ with $\mathrm{Th}_\Gamma(T^-) \subseteq \mathrm{Th}_\Gamma(T^+)$.
In fact, we can take T^+ to be any completion of $A \cup \{\varphi\} \cup \mathrm{Th}_\Gamma(A \cup \{\neg\varphi\})$.

Proof.

First we verify that $A \cup \{\varphi\} \cup \mathrm{Th}_\Gamma(A \cup \{\neg\varphi\})$ is consistent: if not, then for some $\psi \in \Gamma$, we have $A \cup \{\neg\varphi\} \vdash \psi, A \cup \{\varphi\} \vdash \neg\psi$. So by deduction theorem, $A \vdash \neg\varphi \leftrightarrow \psi$, a contradiction.
Take T^+ to be any such completion. Let $\breve{\Gamma} = \{\neg\varphi | \varphi \in \Gamma\}$. By completeness, $\mathrm{Th}_\Gamma(T^-) \subseteq Th_\Gamma(T^+) \iff \mathrm{Th}_{\breve{\Gamma}}(T^+) \subseteq \mathrm{Th}_{\breve{\Gamma}}(T^-)$, so it suffices to verify $A \cup \{\neg\varphi\} \cup \mathrm{Th}_{\breve{\Gamma}}(T^+)$ is consistent. If not, then for some $\psi \in \Gamma$ we have $T^+ \vdash \neg\psi$ and $A \cup \{\neg\varphi\} \vdash \psi$. This contradicts the consistency of T^+. Thus, any completion $A \cup \{\neg\varphi\} \cup \mathrm{Th}_{\breve{\Gamma}}(T^+)$ works as T^-. □

Proof of the Bounded Case

Corollary

For any theory T and any family of sentences Γ closed under conjunction and disjunction, if T is not Γ-axiomatizable (i.e. $\mathrm{Th}_\Gamma(T)$ is not equivalent to T), then there are complete theories T_0, T_1 such that $T \subseteq T_0$, T is inconsistent with T_1, and $\mathrm{Th}_\Gamma(T_0) \subseteq \mathrm{Th}_\Gamma(T_1)$.

Remark. This gives us the "switch."

Proof.

Let $A = \mathrm{Th}_\Gamma(T)$. Choose some sentence φ provable from T but not A. Check that: (1) $A \nvdash \varphi \leftrightarrow \psi$ for any $\psi \in \Gamma$; (2) $T \cup \mathrm{Th}_{\breve{\Gamma}}(A \cup \{\neg\varphi\})$ is consistent.
For (1), if it fails then $T \vdash \varphi \leftrightarrow \psi$, so $\psi \in \mathrm{Th}_\Gamma(T) = A$. Now $A \vdash \varphi$, contradiction.
For (2), if it fails then for some $\psi \in A$, $A \cup \neg\varphi \vdash \neg\psi$, so $A \vdash \varphi$, contradiction.
Now apply the previous lemma (with $\breve{\Gamma}$ in place of Γ above) to a completion T_0 of $T \cup \mathrm{Th}_{\breve{\Gamma}}(A \cup \{\neg\varphi\})$ (as T^+) to obtain T_1 (as T^-). □

Proof of the Bounded Case

Theorem

For any $n \in \omega$ and any theory T (not necessarily complete):

1. *T is not $\forall_n$-axiomatizable $\Rightarrow$ $\mathrm{Mod}(T)$ is $\mathbf{\Sigma}^0_n$-hard.*
2. *T is not $\exists_n$-axiomatizable $\Rightarrow$ $\mathrm{Mod}(T)$ is $\mathbf{\Pi}^0_n$-hard.*

Proof.

1. Apply the corollary to T and $\Gamma = \forall_n$ to obtain T_0, T_1. Then use the first lemma with $T^+ = T_0, T^- = T_1$.
2. Similar to the previous point: $\Gamma = \exists_n, T^+ = T_1, T^- = T_0$.

□

Connection to Infinitary Logic

While infinitary logic is more expressive than first-order logic, it does not do so more efficiently (in terms of quantifier complexity). We give a quick proof:

Theorem (Keisler 1965; Harrison-Trainor, Kretschmer 2023)

If a finitary formula φ is equivalent to a Π_n^{in} formula ψ, then φ is actually equivalent to a finitary $\forall_n$ formula.

Proof.

By compactness, it suffices to show $T = \{\varphi\}$ is $\forall_n$-axiomatizable. This is immediate as $\mathrm{Mod}(T) = \mathrm{Mod}(\psi)$ is $\mathbf{\Pi}_n^0$. □

It turns out that this is, in a sense, equivalent to our theorem. It's also interesting that Keisler used games to approximate formulas, Harrison-Trainor and Kretschmer used arithmetic forcing, while we used (complicated) priority argument, which is effective.

Effectiveness

Since our argument is effective, we would like to analyze what oracles are needed for the construction. In particular, the oracle appearing in Solovay's theorem appears cumbersome, but is actually necessary.

Definition

The oracle D *witnesses* the $\mathbf{\Gamma}$-hardness of $Y \subseteq 2^\omega$ if for every Borel code C for a set $X \in \mathbf{\Gamma}$, there exists a Turing operator Φ such that for every $x \in 2^\omega$,

$$x \in X \iff \Phi^{D \oplus C \oplus x} \in Y.$$

If the choice of Φ is independent of C, we say D *uniformly witnesses* the $\mathbf{\Gamma}$-hardness of Y.

Effectiveness for the Unbounded Case

Theorem (AGLRZ)

Suppose T is complete and unbounded, and X is an oracle appropriate for T in the sense of Solovay's theorem. Then X uniformly witnesses the $\mathbf{\Pi}^0_\omega$-hardness of $\mathrm{Mod}(T)$.

Idea. Using the Scott set, we can choose T_n wisely to get rid of the non-uniformity.

For arithmetic, this gives a complete characterization of the witnessing oracles:

Corollary

If T is a completion of PA *and X computes a non-standard model of T, then X uniformly witnesses the $\mathbf{\Pi}^0_\omega$-hardness of* $\mathrm{Mod}(T)$.
On the other hand, if T is a completion of PA^- *and X computes no non-standard model of T, then it does not witness the $\mathbf{\Sigma}^0_2$-hardness of* $\mathrm{Mod}(T)$.

Effectiveness for the Unbounded Case

Corollary

If T is a completion of PA *and X computes a non-standard model of T, then X uniformly witnesses the $\mathbf{\Pi}^0_\omega$-hardness of* $\mathrm{Mod}(T)$*.*
On the other hand, if T is a completion of PA^- *and X computes no non-standard model of T, then it does not witness the $\mathbf{\Sigma}^0_2$-hardness of* $\mathrm{Mod}(T)$*.*

Proof.

The first part follows from the theorem and Solovay's characterization. For the second part, notice there is a $\Sigma_2(X)$-complete set S whose code is computable from X and whose elements are all computable (for example, $\{\chi_{\{n\}} | n \in X''\}$). Hence, if X were to be a witness, any model of T built by the Turing functional will be computable from X, so it has to be the standard model. But the standard model has a Π^{in}_2 Scott sentence, making S $\Pi_2(X)$, contradiction. □

Effectiveness for the Bounded Case

In the bounded case T may not be complete, so we may need a PA degree over T to compute a completion (thus build a model). This is sufficient.

Theorem (AGLRZ)

If T is not $\forall_n$-axiomatizable, and X is PA over T, then X uniformly witnesses the $\mathbf{\Sigma}^0_n$-hardness of $\mathrm{Mod}(T)$.

Proof.

A PA degree is enough to find completion of the theories used in the lemmas, and the proof from there is uniform. □

Further Questions

- Characterize the Wadge degrees occupied by Mod(T)?
  In particular, how do they differ from the degrees that are Scott complexities?
- Can more be said about the $\mathbf{\Pi}^0_\omega$ case when T is incomplete (and not sequential)?
- More analysis on oracles?

Thank you for listening!

References

[1] Uri Andrews, David Gonzalez, Steffen Lempp, Dino Rossegger, and Hongyu Zhu. *The Borel complexity of the class of models of first-order theories*. 2024. arXiv: 2402.10029 [math.LO].

[2] Ali Enayat and Albert Visser. *Incompleteness of boundedly axiomatizable theories*. 2024. arXiv: 2311.14025 [math.LO].

[3] Matthew Harrison-Trainor and Miles Kretschmer. "Infinitary Logic Has No Expressive Efficiency Over Finitary Logic". In: *The Journal of Symbolic Logic* (2023), 1–18. DOI: 10.1017/jsl.2023.19.