Lecture 9: Effective Borel sets

Suppose $U \subseteq \mathbb{N}^{\mathbb{N}}$ is open. The there exists a set $W \subseteq \mathbb{N}^{<\mathbb{N}}$ such that

$$U = \bigcup_{\sigma \in W} N_\sigma.$$

Using a standard (effective) coding procedure, we can identify finite sequence of natural numbers with a natural number, and thus can see W as a subset of $\mathbb{N}$.

If we provide a Turing machine with oracle W, we can semi-effectively test for membership in U as follows. Assume we want to determine whether some $\alpha \in \mathbb{N}^{\mathbb{N}}$ is in U. Write α on another oracle tape, and start scanning the W oracle. If we retrieve a σ that coincides with an initial segment of α, we know $\alpha \in U$. On the other hand, if $\alpha \in U$, then we will eventually find some $\alpha\restriction_n$ in W. If $\alpha \notin U$, then the search will run forever. In other words, given W, U is *semi-decidable*, or, extending terminology from subsets of $\mathbb{N}$ to subsets of $\mathbb{N}^{\mathbb{N}}$, U is **recusively enumerable** relative to W.

Similarly, we can identify a closed set F with the code for the tree

$$T_F = \{\alpha\restriction_n : \alpha \in F, n \in \mathbb{N}\}.$$

Then we determining whether $\alpha \in F$ is *co-r.e.* in (the code of) T_F. If $\alpha \notin F$ we will learn so after a finite amount of time.

These simple observations suggest the following general approach to Borel sets.

- Borel sets can be coded by a single infinite sequence in $\mathbb{N}^{\mathbb{N}}$ (or $2^{\mathbb{N}}$).
- Given the code, we can describe the Borel set effectively, by means of oracle computations.
- The connection between degrees of unsolvability and definability results in a close correspondence between arithmetical sets (Σ^0_n) and Borel sets of finite order ($\boldsymbol{\Sigma}^0_n$).

In this lecture we will fully develop this correspondence. Later, we will see that it even extends beyond the finite level.

Borel codes

We fix a computable bijection $\pi : \mathbb{N}^{<\mathbb{N}} \to \mathbb{N}$. Furthermore, let $\langle .,. \rangle$ be the standard coding function for pairs,

$$\langle x, y \rangle = \frac{(x+y)(x+y+1)}{2} + y.$$

Borel codes are defined inductively.

Definition 9.1:

(a) A real $\gamma \in \mathbb{N}^{\mathbb{N}}$ is a $\mathbf{\Sigma}^0_1$- **code** if $\gamma(0) = 2$ and for all $n \geq 1$, $\gamma(n) \in \{0, 1\}$. In this case we say that γ *codes* the open set

$$U = \bigcup_{\gamma(\pi(\sigma)=1)} N_\sigma,$$

or γ is a $\mathbf{\Sigma}^0_1$-*code* for U.

(b) If γ is a $\mathbf{\Sigma}^0_n$-code, then γ' with $\gamma'(0) = 3$, $\gamma'(n+1) = \gamma(n)$ is a $\mathbf{\Pi}^0_n$-code. If γ codes $A \subseteq \mathbb{N}^{\mathbb{N}}$, then we say γ' codes $\neg A$.

(c) If for each $m \geq 0$, γ_m is a $\mathbf{\Pi}^0_n$ code for $A_m \subseteq \mathbb{N}^{\mathbb{N}}$, then γ given by $\gamma(0) = 4$ and

$$\gamma(\langle m, n \rangle + 1) = \gamma_m(n)$$

is a $\mathbf{\Sigma}^0_{n+1}$ code, and it codes the set $\bigcup_m A_m$.

Hence the first position in each code indicates the kind of set it codes – an open set, a complement, or a union.

We also define the *set of Borel codes of finite order*

$$\mathrm{Bc}_\omega = \{\gamma \in \mathbb{N}^{\mathbb{N}} : \gamma \text{ is a } \mathbf{\Sigma}^0_n\text{- or } \mathbf{\Pi}^0_n\text{-code, for some } n \geq 1\}.$$

The following is a straightforward induction.

Proposition 9.2: *A set is $\mathbf{\Sigma}^0_n$ ($\mathbf{\Pi}^0_n$) if and only if it has a $\mathbf{\Sigma}^0_n$- ($\mathbf{\Pi}^0_n$) code.*

Note that the definition actually assigns codes to *representations of sets*. A Borel set can have (and has) multiple codes, just as it has multiple representations. We can, for example, represent an open set by different sets W of initial segments.

Moreover, every $\mathbf{\Sigma}^0_1$ set is also $\mathbf{\Sigma}^0_2$, and thus a set has codes which reflect the "more complicated" definition of the $\mathbf{\Sigma}^0_1$ set as a union of closed sets. It is useful to keep this distinction between a Borel set and its Borel representation in mind.

Each Borel code induces a tree structure that reflects how the corresponding Borel set is built up from open sets. A "4" corresponds to a node with infinitely many nodes immediately below it, a "3" to a node with just one immediate extension, and a "2" represents a terminal node, since the open sets are the "building blocks" of the Borel sets and hence are not split further.

The tree of a Borel code is well-founded (i.e. has no infinite path), since a Borel code is defined via a well-founded recursion. The rank of the tree is a countable ordinal.

How hard is it to decide whether a given real is a Borel code? We will see later that this question is quite difficult. In particular, we will extend the set of Borel codes to transfinite orders and see that the set of all Borel codes is *not* Borel. Deciding whether a tree on $\mathbb{N}$ is well-founded will play a fundamental role in this regard.

Borel sets with computable codes

Suppose γ is a computable, $\mathbf{\Sigma}^0_2$-code for an F_σ set F. Then γ is of the form $(4, \gamma')$, where, for $m \geq 0$, the *m-th column* of γ',

$$\gamma'_m(n) = \gamma'(\langle m, n \rangle)$$

is of the form $(3, \alpha_m)$, each α_m being a $\mathbf{\Sigma}^0_1$-code for an open set. Note that γ' and all α_m are computable, too.

We can formulate membership in F as follows.

$$\alpha \in F \quad \Leftrightarrow \quad \exists m \, \forall n \, [\gamma'_m(\pi(\alpha\restriction_n)) = 0].$$

Note that the *inner* predicate $R(x, y)$ given by

$$R(x, y) \quad \Leftrightarrow \quad \gamma'_x(y) = 0$$

is decidable. Hence an F_σ set F with a computable code can be represented in the following form. There exists a recursive predicate $R(x, y)$ such that

$$\alpha \in F \quad \Leftrightarrow \quad \exists m \, \forall n \, R(m, \alpha\restriction_n).$$

In this formulation we drop the coding function π and identify finite sequences directly with natural numbers, and from now on we will continue to do so. It significantly simplifies notation.

On the other hand, if $R(x, y)$ is a recursive predicate, we can define the set

$$W_m = \{\sigma : R(m, \sigma)\}.$$

Then the set $U_m = \bigcup_{W_m} N_\sigma$ is open, and the set F given by

$$\alpha \in F \quad \Leftrightarrow \quad \exists n\, \alpha \in \neg\, U_m \quad \Leftrightarrow \quad \exists m\, \forall n\, \neg R(m, \alpha\restriction_n)$$

is F_σ.

Thus, there seems to be a close connection between F_σ sets with recursive Borel codes and sets definable by Σ^0_2 formulas over recursive predicates. Given that we introduced the notation $\boldsymbol{\Sigma}^0_2$ for F_σ sets earlier, this is perhaps not very surprising.

In this analysis, there seems to be nothing specific about the F_σ used in the example. Indeed, it can be extended to Borel sets of finite order, which we will do next. We introduce the **lightface** Borel hierarchy and show that it corresponds to Borel sets of finite order with recursive codes. Using relativization, we then obtain a complete characterization of Borel sets of finite order: *They are precisely those sets definable by arithmetical formulas, relative to a real parameter.*

The effective Borel hierarchy

Definition 9.3: A set $A \subseteq \mathbb{N}^{\mathbb{N}}$ is

(a) Σ^0_1 if there exists a recursive predicate $R(x)$ such that

$$\alpha \in A \quad \Leftrightarrow \quad \exists n\, R(\alpha\restriction_n),$$

(b) Π^0_n if $\neg A$ is Σ^0_n,

(c) Σ^0_{n+1} if there exists a Π^0_n set P such that

$$\alpha \in A \quad \Leftrightarrow \quad \exists n\, [(n, \alpha) \in P].$$

The following result is at the heart of the effective theory.

Proposition 9.4: *Let $A \subseteq \mathbb{N}^{\mathbb{N}}$. Then*

$$A \text{ is } \Sigma^0_n\ (\Pi^0_n) \quad \Leftrightarrow \quad A \text{ is } \boldsymbol{\Sigma}^0_n\ (\boldsymbol{\Pi}^0_n) \text{ and has a computable } \boldsymbol{\Sigma}^0_n\text{- } (\boldsymbol{\Pi}^0_n\text{-}) \text{ code.}$$

Proof. We proceed by induction on the Borel complexity.

Suppose A is Σ^0_1. Let R be recursive such that $A = \{\alpha \colon \exists n\, R(\alpha\restriction_n)\}$. If we let $W = \{\sigma : R(\sigma)\}$, then

$$A = \bigcup_{\sigma \in W} N_\sigma,$$

and hence is an open set. Furthermore,

$$\gamma(n) = \begin{cases} 2 & n = 0, \\ 1 & n \geq 1 \,\&\, R(n-1), \\ 0 & n \geq 1 \,\&\, \neg R(n-1), \end{cases}$$

is a computable Borel code for A.

If A is $\boldsymbol{\Sigma}^0_1$ with a computable, $\boldsymbol{\Sigma}^0_1$-code γ, then γ is of the form $(2, \gamma')$, γ' coding a representation of A as a union of basic open cylinders. Then

$$\alpha \in A \quad \Leftrightarrow \quad \exists n\, [\gamma'(\alpha\restriction_n) = 1].$$

Hence we can set $R(\sigma) = \gamma'(\sigma)$.

If A is Π^0_n, then $\neg A$ is Σ^0_n. By induction hypothesis, $\neg A$ has a computable $\boldsymbol{\Sigma}^0_n$-code γ. Then $(3, \gamma)$ is a computable $\boldsymbol{\Pi}^0_n$-code for A.

Conversely, if $\gamma = (3, \gamma')$ is a computable $\boldsymbol{\Pi}^0_n$-code for a $\boldsymbol{\Pi}^0_n$ set A, then γ' is a computable $\boldsymbol{\Sigma}^0_n$-code for the $\boldsymbol{\Sigma}^0_n$ set $\neg A$. By induction hypothesis, $\neg A$ is Σ^0_n and hence A is Π^0_n.

Finally, assume that A is Σ^0_{n+1}. Let P be Π^0_n such that

$$\alpha \in A \quad \Leftrightarrow \quad \exists n\, [(n, \alpha) \in P].$$

By induction hypothesis, there exists P is $\boldsymbol{\Pi}^0_n$ with a computable $\boldsymbol{\Pi}^0_n$-code $\gamma = (3, 4, \dots)$. Let

$$P_m = \{\beta : (m, \beta) \in P\} = P \cap N_{\langle m \rangle}.$$

Then each P_m is $\boldsymbol{\Pi}^0_n$, since the Borel levels are closed under finite intersections, and we have

$$A = \bigcup_m P_m.$$

Therefore, A is $\boldsymbol{\Sigma}^0_{n+1}$. Furthermore, each P_m has a computable $\boldsymbol{\Pi}^0_n$-code γ_m, which can be computed uniformly in m, and thus $\gamma^* = (4, (\gamma_m(n))_{m,n})$ is a computable, $\boldsymbol{\Sigma}^0_{n+1}$-optimal code for A.

For the converse, let A be $\mathbf{\Sigma}^0_{n+1}$ with a computable $\mathbf{\Sigma}^0_{n+1}$-code $\gamma = (4, \gamma')$. Then each of the columns of γ' is a computable $\mathbf{\Pi}^0_n$-code for a $\mathbf{\Pi}^0_n$ set P_m. Let $P'_m = \{(m, \alpha)\colon \alpha \in P_m\}$. P'_m is $\mathbf{\Pi}^0_n$, too. This can be seen as follows. $\mathbb{N} \times \mathbb{N}^{\mathbb{N}}$ is homeomorphic to $\mathbb{N}^{\mathbb{N}}$. $\{m\} \times P_m$ is Π^0_n in $\mathbb{N} \times \mathbb{N}^{\mathbb{N}}$, by replacing each set S in the definition of P_m by $\{m\} \times S$ (note that $\{m\}$ is clopen in $\mathbb{N}$). Borel complexities are preserved under homeomorphic images. (We will discuss the closure properties of Borel sets in detail later.)

A similar argument shows that $P^*_m = \{(k, \alpha)\colon k \neq m \text{ or } (k = m \,\&\, \alpha \in P_m)\}$ is $\mathbf{\Pi}^0_n$ for each n. Now let $P^* = \bigcup_m P^*_m$. Then P^* is $\mathbf{\Pi}^0_n$, and we can effectively and uniformly in m compute an $\mathbf{\Pi}^0_n$-code for it. By induction hypothesis, P^* is Π^0_n, and we have

$$\alpha \in A \quad \Leftrightarrow \quad \exists m\, (m, \alpha) \in P^*,$$

as desired. □

Relativization

Using relativized computations via oracles, we can define a relativized version of the effective Borel hierarchy. This way we can capture *all* Borel sets of finite order, not just the ones with computable codes.

Definition 9.5: Let $\gamma \in \mathbb{N}^{\mathbb{N}}$. A set $A \subseteq \mathbb{N}^{\mathbb{N}}$ is

(a) $\Sigma^0_1(\gamma)$ if there exists a predicate $R(x)$ *recursive* in γ such that

$$\alpha \in A \quad \Leftrightarrow \quad \exists n\, R(\alpha\restriction_n),$$

(b) $\Pi^0_n(\gamma)$ if $\neg A$ is $\Sigma^0_n(\gamma)$,

(c) $\Sigma^0_{n+1}(\gamma)$ if there exists a $\Pi^0_n(\gamma)$ set P such that

$$\alpha \in A \quad \Leftrightarrow \quad \exists n\, [(n, \alpha) \in P].$$

A straightforward relativization gives the following analogue of Proposition 9.4.

Proposition 9.6: *Let $A \subseteq \mathbb{N}^{\mathbb{N}}$ and $\gamma \in \mathbb{N}^{\mathbb{N}}$. Then*

$$A \text{ is } \Sigma^0_n(\gamma)\ (\Pi^0_n(\gamma)) \quad \Leftrightarrow \quad A \text{ is } \mathbf{\Sigma}^0_n\ (\mathbf{\Pi}^0_n) \text{ and has a } \mathbf{\Sigma}^0_n\text{- } (\mathbf{\Pi}^0_n\text{-}) \text{ code recursive in } \gamma.$$

We can now present the *fundamental theorem of effective descriptive set theory*.

Theorem 9.7: *A set $A \subseteq \mathbb{N}^{\mathbb{N}}$ is $\boldsymbol{\Sigma}^0_n$ ($\boldsymbol{\Pi}^0_n$) if and only if it is $\Sigma^0_n(\gamma)$ ($\Pi^0_n(\gamma)$) for some $\gamma \in \mathbb{N}^{\mathbb{N}}$.*

Proof. If A is $\boldsymbol{\Sigma}^0_n$, then by Proposition 9.2 it has a $\boldsymbol{\Sigma}^0_n$-code γ, and by Proposition 9.6 A is $\Sigma^0_n(\gamma)$. The other direction follows immediately from Proposition 9.6.

The argument for $\boldsymbol{\Pi}^0_n$ is completely analogous. □

Definability in Arithmetic

One of the fundamental insights of recursion theory is the close relation between computability and definability in arithmetic. The recursively enumerable subsets of $\mathbb{N}$ are precisely the sets Σ_1-definable over the standard model of arithmetic, $(\mathbb{N}, +, \cdot, 0, 1)$, and **Post's Theorem** uses this result to establish a rigid connection between levels of arithmetical complexity and computational complexity.

As indicated above, we can use this relation to give a characterization of the Borel sets of finite order in terms of definability. Since we are dealing with subsets of $\mathbb{N}^{\mathbb{N}}$, that is, with *sets* of functions on $\mathbb{N}$ rather than just functions on $\mathbb{N}$, we will work in the framework of *second order arithmetic*.

The **language of second order arithmetic**has two kinds of variables: *number variables* $x, y, z, \ldots$ (and sometimes k, l, m, n if they are not used as metavariables), to be interpreted as elements of $\mathbb{N}$, and *function variables* $\alpha, \beta, \gamma, \ldots$, intended to range over functions from $\mathbb{N}$ into $\mathbb{N}$, i.e. elements of Baire space, i.e. reals. The non-logical symbols are the binary function symbols $+, \cdot$, the binary relation symbol $<$, the *application function* symbol ap, and the constants $\underline{0}, \underline{1}$. *Numerical terms* are defined in usual way using $+, \cdot, \underline{0}, \underline{1}$, and involve only number variables. *Atomic formulas* are $t_1 = t_2$, $t_1 < t_2$, and $\mathrm{ap}(\alpha, t_1) = t_2$, where t_1, t_2 are numerical terms.

The **standard model of second order arithmetic** is the structure

$$\mathcal{A}^2 = (\mathbb{N}, \mathbb{N}^{\mathbb{N}}, \mathrm{ap}, +, \cdot, <, 0, 1),$$

where $+$ and $\cdot$ are the usual operations on natural numbers, $<$ is the standard ordering of $\mathbb{N}$. The two domains are connected by the binary operation $\mathrm{ap} : \mathbb{N}^{\mathbb{N}} \times \mathbb{N} \to \mathbb{N}$, defined as

$$\mathrm{ap}(\alpha, x) = \alpha(x).$$

A relation $R \subseteq \mathbb{N}^m \times (\mathbb{N}^{\mathbb{N}})^n$ is *definable over $\mathcal{A}^2$* if there exists a formula φ of second order arithmetic such that for any $x_1, \ldots, x_m \in \mathbb{N}$ and $\alpha_1, \ldots \alpha_n \in \mathbb{N}^{\mathbb{N}}$,

$$R(x_1, \ldots, x_m, \alpha_1, \ldots \alpha_n) \quad \text{iff} \quad \mathcal{A}^2 \models \varphi[x_1, \ldots, x_m, \alpha_1, \ldots \alpha_n].$$

Theorem 9.8: *A set $A \subseteq \mathbb{N}^{\mathbb{N}}$ is $\boldsymbol{\Sigma}^0_n$ ($\boldsymbol{\Pi}^0_n$) if and only if it is definable over $\mathcal{A}^2$ by a Σ^0_n (Π^0_n) formula.*

Here, Σ^0_n (Π^0_n) formula means that we can *only quantify over number variables*, as opposed to Σ^1_n (Π^1_n) formulas, where we can also quantify over function variables.

The proof is a straightforward extension of the standard argument for subsets of $\mathbb{N}$.

To formulate the fundamental Theorem 9.7 in terms of definability, we need the concept of **relative definability**. We add a new constant function symbol $\underline{\gamma}$ to the language. Given a function γ, a relation is **definable in** γ if it is definable over the structure

$$\mathcal{A}^2(\gamma) = (\mathbb{N}, \mathbb{N}^{\mathbb{N}}, \mathrm{ap}, +, \cdot, <, 0, 1, \gamma),$$

where the symbol $\underline{\gamma}$ is interpreted as γ.

Then the following holds.

Theorem 9.9: *A set $A \subseteq \mathbb{N}^{\mathbb{N}}$ is $\boldsymbol{\Sigma}^0_n$ ($\boldsymbol{\Pi}^0_n$) if and only if it is definable in γ by a Σ^0_n (Π^0_n) formula, for some $\gamma \in \mathbb{N}^{\mathbb{N}}$.*

This theorem facilitates the description of Borel sets considerably. As an example, consider the set

$$A = \{\alpha \colon \alpha \text{ eventually constant}\}.$$

We have

$$\alpha \in A \quad \Leftrightarrow \quad \exists n \forall m [m \geq n \Rightarrow \alpha(n) = \alpha(m)]$$

The right hand side is a Σ^0_2-formula. Hence the set A is $\boldsymbol{\Sigma}^0_2$ (even Σ^0_2).