MA5219 LOGIC AND FOUNDATION OF MATHEMATICS I

Lecture 23: Ramsey's theorems

Tin Lok Wong

5 November, 2018

The aim of this lecture is to define an indicator for PA over PA.

Roughly speaking, the reason for the unprovability of the sentence in Theorem 22.8(2) is: to achieve a long distance, one needs to go so far out in the structure that is beyond the reach of the theory involved. For instance, in the case of $I\Delta_0(\exp)$ as in Observation 22.4, to achieve arbitrarily long distances, one needs arbitrarily high towers of exponentials, which are beyond the reach of $I\Delta_0(\exp)$. Indicators for PA give much faster-growing functions. In the Hardy Hierarchy of functions, for example, indicators for $I\Delta_0(\exp)$ are at level ω^3, while indicators for PA go up to level ε_0. It is not a coincidence that the ordinal ε_0 appears here again in our study of PA: in a very precise sense, the ordinal of PA is ε_0. However, we will not go into this here.

As suggested in the previous paragraph, to find an indicator for PA, one needs to define very fast-growing functions. However, it is conceivable that even if the growth rate of the function is right, still one may not be able to extract a cut satisfying PA from it, because PA expresses rather more than the totality of Σ_1-definable functions. The case of $I\Delta_0(\exp)$ is simpler since it is equivalent to a set of Π_1 sentences (cf. the proof of Proposition 22.3), and so it is automatically preserved in cuts by Proposition 22.2(2). The same is not true for PA. For instance, the cut $2^a_{\mathbb{N}}$ defined in our proof of Proposition 22.1 does not satisfy PA because it is not closed under iterated exponentiation, while PA proves the totality of iterated exponentiation. It is actually not clear at first sight why indicators for PA should exist.

Nevertheless, it is a fact that having the right growth rate guarantees a cut satisfying PA. The proof is non-trivial, and we will not study it. What we will do is simply: define a function with the right growth rate, then verify in the next lecture that this particular function works.

We need fast-growing functions. Thus, in combinatorics, Ramsey theory is a natural place to start. Let us begin with a standard example.

Example 23.1. Whenever the edges of the complete graph K_6 on 6 vertices are coloured red or blue, there must be a monochromatic triangle (K_3).

Proof. Take any vertex v_0 in the graph. It lies on 5 edges. Since there are only 2 colours, at least three of these edges must have the same colour by the Finite Pigeonhole Principle. Without loss of generality, suppose this colour is blue. Let v_1, v_2, v_3 be distinct vertices which share a blue edge with v_0. If some edge between them is blue, say between v_1 and v_2, then v_0, v_1, v_2 are the vertices of a blue triangle. If no edge between them is blue, then v_1, v_2, v_3 are the vertices of a red triangle. □

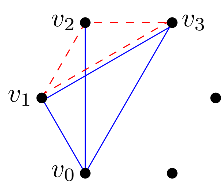

Figure 23.1: A proof that $R^2_2(3) \leqslant 6$

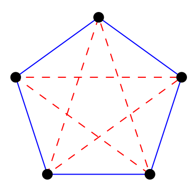

Figure 23.2: A proof that $R^2_2(3) > 5$

As shown in Figure 23.2, there is a 2-colouring of the edges of K_5 that does not give a monochromatic triangle.

Moral. In a large enough structure, there must be some regularity.

In general, the existence of a monochromatic K_m is guaranteed when we colour the edges of K_n, for sufficiently large $n \in \mathbb{N}$, using finitely many colours. This generalizes to *hypergraphs*, i.e., graphs in which (hyper)edges have more than two endpoints. This generalization is the Finite Ramsey Theorem. To state it concisely, we introduce some standard pieces of notation from set theory.

Notation.

- If $n \in \mathbb{N}$, then $\{x \in \mathbb{N} : x < n\}$ is denoted n.
- If $S \subseteq \mathbb{N}$ and $k \in \mathbb{N}$, then $[S]^k = \{X \subseteq S : \mathrm{card}(X) = k\}$.

The exponent-1 case of the Finite Ramsey theorem is simply the Finite Pigeonhole Principle.

Finite Pigeonhole Principle.

$$\forall m, r \geqslant 1 \ \ \forall f\colon [(m-1)r+1]^1 \to r \ \ \exists H \in [(m-1)r+1]^m \ \ \forall x, y \in [H]^1 \ \ f(x) = f(y).$$

Moreover, this theorem is provable in PA.

Proof sketch. If there are r colours and at most $(m-1)$ elements per colour, then there are at most $(m-1)r$ elements in the domain of the colouring function f. □

One can thus view the Finite Ramsey Theorem as a higher-exponent version of the Finite Pigeonhole Principle. Some standard terminology from Ramsey theory will be handy.

Definition. Let $k, m, n, r \geqslant 1$ and S be a set.

- A set H is *homogeneous* for a function $f\colon [S]^k \to r$ if $H \subseteq S$ and $f(X) = f(Y)$ for all $X, Y \in [H]^k$.
- Following Erdős and Rado, we abbreviate by $n \to (m)^k_r$ the assertion

$$\forall f\colon [n]^k \to r \ \ \exists H \in [n]^m \ \ \forall X, Y \in [H]^k \ \ f(X) = f(Y).$$

Finite Ramsey Theorem (FRT)**.** $\forall k, m, r \geqslant 1 \ \ \exists n = \mathrm{R}^k_r(m) \ \ n \to (m)^k_r$, that is,

$$\forall k, m, r \geqslant 1 \ \ \exists n = \mathrm{R}^k_r(m) \ \ \forall f\colon [n]^k \to r \ \ \exists H \in [n]^m \ \ \forall X, Y \in [H]^k \ \ f(X) = f(Y).$$

Moreover, this theorem is provable in PA.

Using the coding we discussed at the beginning of the previous lecture, one can express FRT in $\mathcal{L}_\mathrm{A}(\exp)$ over PA. For example, we can define $[n]^k$ as the set of all subsets of n from which there exists a bijection to k. We will only prove the FRT in $\mathbb{N}$. The verification that this proof can be carried out in PA is tedious but straightforward. So we omit it.

Proof of FRT in $\mathbb{N}$. We proceed by induction on k. The $k = 1$ case is the Finite Pigeonhole Principle. For the induction step, let $k \geqslant 1$ such that

$$\forall m, r \geqslant 1 \ \ \exists n = \mathrm{R}^k_r(m) \ \ n \to (m)^k_r.$$

Pick $m, r \geqslant 1$. Define $\ell = \mathrm{R}^1_r(m)$ and $d_\ell = k$. Using the induction hypothesis, for each $i < \ell$, let $d_i = \mathrm{R}^k_r(d_{i+1}) + 1$. We show that setting $n = d_0$ works.

Take any $f\colon [n]^{k+1} \to r$. To find the required homogeneous set, we define in the following

$$n = S_0 \supseteq S_1 \supseteq S_2 \supseteq \cdots \supseteq S_\ell$$

with sizes $d_0, d_1, \ldots, d_\ell$ and minima $a_0, a_1, \ldots, a_\ell$ respectively. In addition, each $S_{i+1} \subseteq S_i \setminus \{a_i\}$. Suppose $i < \ell$ such that S_i is defined. Define $f_i\colon [S_i \setminus \{a_i\}]^k \to r$ by setting $f_i(X) = f(X \cup \{a_i\})$ for each $X \in [S_i \setminus \{a_i\}]^k$. Since $\mathrm{card}(S_i \setminus \{a_i\}) = d_i - 1 = \mathrm{R}^k_r(d_{i+1})$, we get $S_{i+1} \in [S_i \setminus \{a_i\}]^{d_{i+1}}$ homogeneous for f_i. Let $g(i)$ be the unique element in $\{f_i(X) : X \in [S_{i+1}]^k\}$.

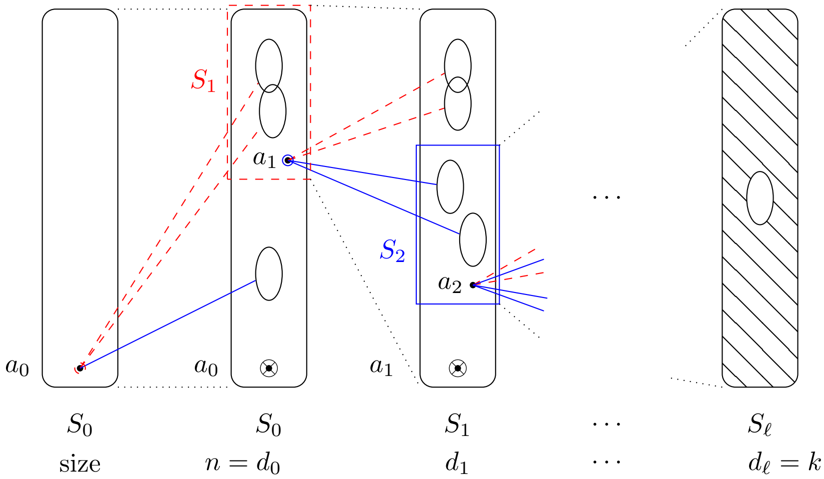

Figure 23.3: The induction step in our proof of the Finite Ramsey Theorem

At the end, we have a function $g\colon \ell \to r$. Since $\ell = \mathrm{R}^1_r(m)$, we have $I \in [\ell]^m$ homogeneous for g. Let c be the unique element in $\{g(i) : i \in I\}$, and $H = \{a_i : i \in I\}$. If $i_0, i_1, \dots, i_k$ is a strictly increasing sequence in I, then

$$\begin{aligned} & f(\{a_{i_0}, a_{i_1}, \dots, a_{i_k}\}) \\ &= f_{i_0}(\{a_{i_1}, a_{i_2}, \dots, a_{i_k}\}) && \text{as } a_{i_1}, a_{i_2}, \dots, a_{i_k} \in S_{i_0} \setminus \{a_{i_0}\}; \\ &= g(i_0) && \text{by the definition of } g \text{, as } i_j > i_0 \text{ implies } a_{i_j} \in S_{i_j} \subseteq S_{i_0+1}; \\ &= c && \text{as } i_0 \in I. \end{aligned}$$

So H is homogeneous for f. □

The proof of FRT can be viewed as a nested application of the Finite Pigeonhole Principle. As more and more things pile up, one is bound to get some repetition. In the case of FRT, this repetition is the homogeneous set.

Definition. Let $k, m, r \geqslant 1$. Then $\mathrm{R}^k_r(m)$ denotes the least $n \in \mathbb{N}$ such that $n \to (m)^k_r$. These $\mathrm{R}^k_r(m)$'s are called *Ramsey numbers*.

For the Ramsey numbers, our proof of FRT gives an upper bound of the form a tower of exponentials of height roughly the same as the exponent. (Erdős and Hajnal has a lower bound of a similar form.) Therefore, Ramsey numbers themselves do not produce functions that grow fast enough for PA.

One way to produce faster-growing functions is to impose (artificially) some extra conditions on the homogeneous set, so that bigger intervals are needed to guarantee its existence. This is where the notion of relative largeness comes into the picture.

Definition. A set H of natural numbers is *relatively large* if $\operatorname{card}(H) \geqslant \min(H)$.

One peculiar feature of the notion of relative largeness is that shifting a relatively large set upwards may not result in a relatively large set. So the notion depends not only on the cardinality of the set but also on the position. In particular, our proof of FRT cannot give relatively large homogeneous sets because the minimum of H is determined at the final stage when we apply the Finite Pigeonhole Principle to g. However, we need it at the beginning to determine how big ℓ and hence n should be. We thus need a different proof.

The plan is to go via the infinite. The key observation is that if the set S_0 in the proof of the FRT is infinite, then one can make S_1, and hence $S_2, S_3, \dots$ infinite. For example, at the base level, this follows from the Infinite Pigeonhole Principle, or what we called the regularity scheme in Lemma 21.1. This progresses up by induction for higher exponents. At the end we get an infinite sequence $a_0, a_1, a_2, \dots$ and thus an infinite homogeneous set by another application of the Infinite Pigeonhole Principle. These show the following.

Infinite Ramsey Theorem (IRT).

$$\forall k, r{\geqslant}1 \ \ \forall f\colon [\mathbb{N}]^k \to r \ \ \exists \text{infinite } H{\subseteq}\mathbb{N} \ \ \forall X, Y{\in}[H]^k \ f(X) = f(Y).$$

Moreover, this theorem is provable *definably* in PA when restricted to $k \in \mathbb{N}$, in the sense that if the f is definable, then one can find a definable H with the required properties. □

As mentioned above, the IRT can be proved in a way similar to our proof of the FRT. In particular, it can be proved by an induction on k. In the PA version, this induction cannot be carried out in PA. For instance, one cannot express in $\mathcal{L}_{\mathrm{A}}(\exp)$ the quantifiers $\forall f$ and $\exists H$ in the statement of the IRT because they range over infinite subsets of natural numbers, not natural numbers. So there is no $\mathcal{L}_{\mathrm{A}}(\exp)$ formula with respect to which we can apply induction in PA. In fact, the complexity of the defining formula (in the sense of the arithmetic hierarchy from Lecture 13, say) for the homogeneous set increases with the exponent k. Therefore, for nonstandard exponents k, the naive extrapolation gives defining formulas that have nonstandardly many quantifier alterations. Such formulas exist in all nonstandard models of PA. However, there may not be a reasonable notion of truth on these formulas.

As mentioned above, the relative largeness of a set depends partially on its position. So we state the relevant Ramsey-type theorem in terms of intervals with possibly nonzero starting point.

Definition. Let $k, m, r \geqslant 1$ and $x, y \in \mathbb{N}$.

- Let $[x, y] = \{v \in \mathbb{N} : x \leqslant v \leqslant y\}$ and $[x, \infty) = \{v \in \mathbb{N} : v \geqslant x\}$.
- We abbreviate by $[x, y] \xrightarrow{*} (m)^k_r$ the assertion

  $\forall f\colon [[x, y]]^k \to r \ \exists H{\subseteq}[x, y] \ \big(\forall X, Y{\in}[H]^k \ \ f(X) = f(Y) \text{ and } \operatorname{card}(H) \geqslant \max\{m, \min(H)\}\big).$

Paris–Harrington–Ramsey Theorem (PH). $\forall k, m, r, x{\geqslant}1 \ \ \exists y{\geqslant}x \ \ [x, y] \xrightarrow{*} (m)^k_r$, that is,

$$\begin{aligned}&\forall k, m, r, x{\geqslant}1 \ \ \exists y{\geqslant}x \ \ \forall f\colon [[x, y]]^k \to r \ \ \exists H{\subseteq}[x, y]\\ &\qquad\qquad \forall X, Y{\in}[H]^k \ \ f(X) = f(Y) \quad \text{and} \quad \operatorname{card}(H) \geqslant \max\{m, \min(H)\}.\end{aligned}$$

Moreover, this theorem is provable in PA when restricted to $k \in \mathbb{N}$.

Proof of PH in $\mathbb{N}$. Suppose we have $k, m, r, x \geqslant 1$ such that $\forall n{\in}\mathbb{N} \ \neg([x, x+n] \xrightarrow{*} (m)^k_r)$. For each $n \in \mathbb{N}$, fix $f_n\colon [[x, x+n]]^k \to r$ which has no relatively large homogeneous set of size at least m. The idea is to define a function $f\colon [[x, \infty)]^k \to r$ such that every restriction of f to $[[x, x+n]]^k$, where $n \in \mathbb{N}$, agrees with f_i for some $i \in \mathbb{N}$. This f has an infinite homogeneous set H by the IRT. Any relatively large finite subset $H_0 \subseteq H$ of size at least m is then homogeneous for some f_i, which is not possible by our choice of the f_i's.

In detail, we define infinite index sets

$$\mathbb{N} = I_0 \supseteq I_1 \supseteq I_2 \supseteq \cdots$$

as follows. Suppose I_n is found, where $n \in \mathbb{N}$.

- There are finitely many function $[[x, x+n]]^k \to r$.
- There are infinitely many functions in $\{f_i : i \in I_n\}$ whose domain includes $[[x, x+n]]^k$.

Therefore, by the Infinite Pigeonhole Principle, there are infinitely many f_i's, where $i \in I_n$, which agree on $[[x, x+n]]^k$. Find an infinite $I_{n+1} \subseteq I_n$ such that

$$\forall i, j{\in}I_{n+1} \ \ f_i{\restriction}[[x, x+n]]^k = f_j{\restriction}[[x, x+n]]^k.$$

Define $f\colon [[x, \infty)]^k \to r$ as follows: for each $X \in [[x, \infty)]^k$, find $n \in \mathbb{N}$ such that $X \in [[x, x{+}n]]^k$, then set $f(X) = f_i(X)$ for some $i \in I_{n+1}$. Our choice of I_{n+1} guarantees f is well-defined. By (a shifted version of) the IRT, we get an infinite $H \subseteq [x, \infty)$ homogeneous for f. Take any finite $H_0 \subseteq H$ containing $\min(H)$ whose size is greater than both m and $\min(H)$, which is possible since H is infinite. Find $n \in \mathbb{N}$ such that $[x, x+n] \supseteq H_0$. Then H_0 is a relatively large homogeneous set for f_i of size at least m, for all $i \in I_{n+1}$, because $f_i{\restriction}[H_0]^k = f{\restriction}[H_0]^k$. □

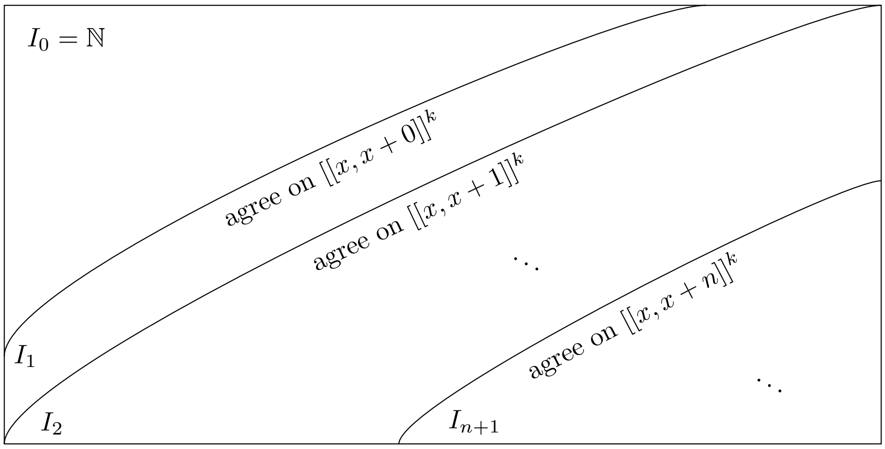

Figure 23.4: Proving the Paris–Harrington–Ramsey Theorem from the Infinite Ramsey Theorem

We formulated our proof of PH as a proof by contradiction, but it does not need to be. To see this, consider all colourings $f\colon [[x, x+n]]^k \to r$, where $n \in \mathbb{N}$, with no relatively large homogeneous set of size at least m. Order them by inclusion. This gives a tree of colourings which is finitely branching. This tree cannot have an infinite path because the union along a path is again a colouring with no relatively large homogeneous set of size at least m. (Note that an infinite set of natural numbers is always relatively large.) Therefore, by some fixed finite level, all the paths in this tree must have terminated. This is the level at which every colouring gets a relatively large homogeneous set of size at least m, as required. Such an argument is often called a *compactness argument* because its key ingredient is the compactness of the tree of colourings with respect to some natural topology. This compactness is used to deduce the existence of an empty level from the absence of infinite paths. We do not adopt this argument since it is slightly more abstract.

With PH at hand, we are ready to define our indicator for PA.

Definition (Bovykin)**.** Over PA, define

$$Y_{\mathrm{PH}}(x,y) = \begin{cases} (\max e \geqslant 1)\big([x,y] \overset{*}{\to} (\mathrm{R}_e^{2e}(3e) + 4e + 1)^{4e+1}_{3e+1}\big), & \text{if this maximum exists;} \\ 0, & \text{otherwise.} \end{cases}$$

Note that $\mathrm{PA} \vdash \forall x, y\, \forall e{\geqslant}y\, \neg\big([x,y] \overset{*}{\to} (\mathrm{R}_e^{2e}(3e) + 4e + 1)^{4e+1}_{3e+1}\big)$. Therefore, by the least number principle from Proposition 22.5, the maximum in the definition of $Y_{\mathrm{PH}}(x,y)$ exists precisely when $[x,y] \overset{*}{\to} (\mathrm{R}_1^{2\times 1}(3\times 1) + 4\times 1 + 1)^{4\times 1+1}_{3\times 1+1}$. This condition can be checked algorithmically because there are only finitely many functions $[[x,y]]^{4\times 1+1} \to 3\times 1+1$ for any given x, y. Thus one can express $Y_{\mathrm{PH}}(x,y) = z$ using a Σ_1 formula by the Church–Turing Thesis. The following shows the implication (i) ⇒ (ii) in property (b) of indicators.

Lemma 23.2. Let $a, b \in M \models \mathrm{PA}$. If $K \subseteq_{\mathrm{e}} M$ satisfying PA with $a \in K < b$, then $Y^M_{\mathrm{PH}}(a,b) > \mathbb{N}$.

Proof. For every $n \in \mathbb{N}$,

$$\begin{array}{lll} & K \models [a,y] \overset{*}{\to} (\mathrm{R}_n^{2n}(3n) + 4n + 1)^{4n+1}_{3n+1} \quad \text{for some } y \in K, & \text{by the version of PH provable in PA;} \\ \therefore & K \models Y_{\mathrm{PH}}(a,y) \geqslant n \quad \text{for some } y \in K, & \text{by the definition of } Y_{\mathrm{PH}}; \\ \therefore & K \models \underbrace{\exists z{\geqslant}n\ Y_{\mathrm{PH}}(a,y) = z}_{\Sigma_1} \quad \text{for some } y \in K, & \\ \therefore & M \models \exists z{\geqslant}n\ Y_{\mathrm{PH}}(a,y) = z \quad \text{for some } y \in K < b, & \text{by Proposition 22.2(2), as } K \subseteq_{\mathrm{e}} M; \\ \therefore & M \models Y_{\mathrm{PH}}(a,b) \geqslant n & \text{as } y < b \text{ implies } Y^M_{\mathrm{PH}}(a,y) \leqslant Y^M_{\mathrm{PH}}(a,b). \end{array}$$

□

The next lecture will be devoted to a proof of the converse of Lemma 23.2. This will complete the verification that $Y_{\mathrm{PH}}(x,y) = z$ is an indicator for PA over PA. The unprovability of $\forall z\, \forall x\, \exists y\, Y_{\mathrm{PH}}(x,y) > z$, and hence PH, in PA will then follow from Proposition 22.8(2).