Unwinding Furstenberg's proof of Szemerédi's Theorem

Jeremy Avigad

Department of Philosophy and Department of Mathematical Sciences
Carnegie Mellon University

(joint work with Henry Towsner)

January, 2008

Hilbert's program

The end of nineteenth century brought the use of infinitary, nonconstructive, set-theoretic methods to mathematics.

With the discovery of set-theoretic paradoxes at the turn of the twentieth century, attention focused on the consistency of the new methods.

> *Our principal result is that the infinite is nowhere to be found in reality. It neither exists in nature nor provides a legitimate basis for rational thought.... Operating with the infinite can be made certain only with the finitary. (Hilbert, "On the infinite," 1925)*

Gödel's second incompleteness theorem shows the impossiblity of providing a finitary justification for the new methods.

Hilbert's program

Nonetheless, proof theorists continued trying to understanding infinitary methods more explicit terms:

- formalizability in restricted fragments of set theory
- interpretability in constructive theories

Research like this yielded "reducibility in principle."

In recent years, "proof mining" has aimed to apply these methods to ordinary mathematical developments, to extract additional information from nonconstructive arguments.

Applications of analysis in number theory and combinatorics are particularly interesting, from a proof-theoretic perspective.

van der Waerden's theorem

Theorem. If one colors the natural numbers with finitely many colors, then there are arbitrarily long monochromatic arithmetic progressions.

The theorem has a Π_2 statement:

Theorem. For every k and r there is an n large enough such that if one colors elements of the set $\{1, \ldots, n\}$ with r colors, there is a monochromatic arithmetic progression of length k.

van der Waerden proved this in 1927. Furstenberg and Weiss presented an elegant proof using topological dynamics in 1978.

Szemerédi's theorem

Szemerédi's theorem is a "density" version of van der Waerden's theorem.

Szemerédi's Theorem. Every set S of natural numbers with positive upper Banach density has arbitrarily long arithmetic progressions.

Equivalently:

Theorem. For every k and $\delta > 0$, there is an n large enough, such that if S is any subset of $\{1, \ldots, n\}$ with density at least δ, then S has an arithmetic progression of length k.

History

- 1952: Roth showed existence of three-term arithmetic progressions.
- 1969: Szemerédi showed existence of four-term arithmetic progressions.
- 1974: Szemerédi proved the full theorem.
- 1977: Furstenberg
  - gave an equivalent measure-theoretic statement,
  - gave an ergodic-theoretic proof, and
  - provided a structural analysis of ergodic measure preserving systems.
- 1979: Furstenberg and Katznelson used the structure theorem to give a streamlined proof of an even stronger statement.

Beleznay and Foreman (1996) show that the structure theorem exhausts the countable ordinals.

A proof-theoretic analysis

The fact that such a detour through the infinite can be used to prove an explicit combinatorial statement deserves logical analysis.

Our analysis runs like this:

- The proof based on the structure theorem can be carried out in a restricted theory known as ID_1.
- Proofs in ID_1 can be translated into explicit computational proofs involving functionals defined by recursion over well-founded trees.

Combining these two steps, Towsner and I are working to obtain an explicit combinatorial version of the proof.

For the second step, we are using a Dialectica-style functional interpretation that we have developed.

This talk will focus on the first step.

Overview

Here is what I aim to do:

- Provide a sketchy overview of the transfinite proof.
- Tell you about ID_1.
- Describe the central issues involved in formalizing the proof in ID_1.

The Furstenberg correspondence principle

A *measure preserving system* $\mathcal{X} = (X, \mathcal{B}, \mu, T)$ consists of a finite measure space $(X, \mathcal{B}, \mu)$, and a *measure preserving transformation*, T.

Given a sequence of subsets S_m of $\{1, \ldots, m\}$ of density $\delta > 0$, the "Furstenberg correspondence" yields a separable mps and a set A such that

- $\mu(A) \geq \delta$
- If $\mu(A \cap T^{-n}A \cap T^{-2n}A \cap \ldots \cap T^{-(k-1)n}A) > 0$ for some n then there are arithmetic progressions of length k in the S_m's.

Szemerédi's theorem becomes equivalent to the following:

Theorem. For any measure preserving system $(X, \mathcal{B}, \mu, T)$, any set A of positive measure, and any k, there is an $n > 0$ such that

$$\mu(A \cap T^{-n}A \cap T^{-2n}A \cap \ldots \cap T^{-(k-1)n}A) > 0.$$

Two distinct behaviors

A measure preserving system is *weak mixing* if we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i<n} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| = 0$$

for every A and B.

A weak mixing system exhibits a high degree of randomness.

A measure-preserving transformation T of $\mathcal{X}$ induces an isometry $\hat{T}$ of $L^2(\mathcal{X})$, $\hat{T}f = f \circ T$. A measure-preserving system is *compact* if it has the property that for every f in $L^2(X, \mathcal{B}, \mu)$, the orbit

$$\{f, \hat{T}f, \hat{T}^2 f, \ldots\}$$

is totally bounded, i.e. has compact closure.

A compact system exhibits a high degree or regularity.

Lack of randomness implies order

Let $\mathcal{X} = (X, \mathcal{B}, \mu, T)$ be an ergodic measure preserving system.

Lemma (Koopman-von Neumann). If $\mathcal{X}$ is not weak mixing, it has a nontrivial compact T-invariant factor.

Three ways of thinking of a T-invariant factor:

- $(X, \mathcal{B}', \mu, T)$, for a T-invariant sub-σ-algebra $\mathcal{B}' \subseteq \mathcal{B}$
- a homomorphic image, or quotient, of $(X, \mathcal{B}, \mu, T)$
- $\hat{T}$-invariant subspace of $L^2(\mathcal{X})$, containing the constant functions and closed under max.

The Furstenberg structure theorem

The notions of compactness and weak mixing relativize to factors.

Lemma (Furstenberg). If an ergodic measure preserving system $(X, \mathcal{B}, \mu, T)$ is not weak mixing relative to a factor $\mathcal{B}'$, there there is an intermediate factor $\mathcal{B}'' \supsetneq \mathcal{B}'$ such that $(X, \mathcal{B}'', \mu, T)$ is compact relative to $(X, \mathcal{B}', \mu, T)$.

We can iterate this, taking unions at limit stages. If the system is separable, the process comes to an end at a countable ordinal.

The Furstenberg structure theorem

The Furstenberg Structure Theorem. Let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure preserving system. Then there is a transfinite increasing sequence of factors $(\mathcal{B}_\alpha)_{\alpha \leq \gamma}$ such that:

- $\mathcal{B}_0$ is the trivial factor.
- For each $\alpha < \gamma$, $(X, \mathcal{B}_{\alpha+1}, \mu, T)$ is compact relative to $(X, \mathcal{B}_\alpha, \mu, T)$.
- For each limit $\lambda \leq \gamma$, $\mathcal{B}_\lambda = \cup_{\alpha<\lambda} \mathcal{B}_\alpha$.
- $(X, \mathcal{B}, \mu, T)$ is weakly mixing relative to $(X, \mathcal{B}_\gamma, \mu, T)$.

If $\mathcal{B} = \mathcal{B}_\gamma$ the system is *distal*, and $\mathcal{B}_\gamma$ is the *maximal distal factor*.

Theorem (Beleznay and Foreman). For any countable ordinal α, there is a separable measure preserving system such that any Furstenberg tower has height at least α.

The proof of Szemerédi's theorem

If $\mu(A) > 0$, say that A is SZ if for every k,

$$\liminf_{N\to\infty} \frac{1}{n} \sum_{n<N} \mu(A \cap T^{-n}A \cap T^{-2n}A \cap \ldots \cap T^{-(k-1)n}A) > 0.$$

Say a factor is SZ if every element is SZ.

If the set A from the Furstenberg correspondence is SZ, we have the desired conclusion.

The property of being SZ:

- holds of the trivial factor;
- is maintained under compact extensions;
- is maintained under limits; and
- is maintained under weak mixing extensions.

The theory ID_1

Let $\psi(P, x)$ be an arithmetic formula with a new predicate symbol P that occurs only positively.

This determines a monotone operator

$$\Gamma_\psi(S) = \{x \mid \psi(S, x)\},$$

which thus has a least fixed point, $I = \bigcap\{S \mid \Gamma_\psi(S) \subseteq S\}$.

The theory ID_1 has axioms asserting:

- Γ_ψ-closure: $\Gamma_\psi(I) \subseteq I$
- Γ_ψ-induction: if $\Gamma_\psi(S) \subseteq S$, then $I \subseteq S$, for any definable set S.

This theory has the same strength as Kripke-Platek set theory with an axiom of infinity.

Formalizing the proof in ID_1

Issue: ID_1 has variables ranging over the natural numbers, and inductively defined predicates.

Solution: The Furstenberg construction yields a separable measure space. Use the inductively defined predicate to represent the maximal distal factor.

The relevant analytic objects can be dealt with at different levels:

- The measure preserving system, $(X, \mathcal{B}, \mu, T)$
- The measure algebra, $\mathcal{B}$ modulo null sets
- The corresponding L^2 space
- The spectral decomposition of T

"Ordinary" mathematics moves seamlessly across representations.

Formalizing the proof in ID_1

Measure and integral are morally the same,

$$\mu(A) = \int \chi_A d\mu$$

We found it most natural to work exclusively with functions.

Take *simple* functions to be finite rational linear combinations of characteristic functions of simple sets:

$$f = \sum_{i<n} a_i \chi_{[\sigma_i]}$$

where $[\sigma_i]$ are the basic open sets given by the correspondence.

Given μ on basic open sets, take elements of L^1 (or L^2, L^∞) to be given by Cauchy sequences in the given norm.

Formalizing the proof in ID_1

In fact, we work exclusively with simple approximations to the relevant objects.

Two major constraints:

- With our representation, pointwise notions have to be avoided.
- Talk of sets (via their characteristic functions) is unnatural, and artificially adds to the axiomatic requirements.

We've invested most of our effort in "purifying" the Furstenberg proof, so it is an argument about (simple approximations to) L^∞ functions.

Formalizing the proof in ID_1

Fix $(X, \mathcal{B}, \mu, T)$. Given a factor, Y, let $Z(Y)$ be the maximal compact extension of Y, i.e. the space spanned by functions that are compact relative to Y.

With a reasonable coding of factors, $Y \mapsto Z(Y)$ is arithmetic. Then the maximal distal factor has a Π^1_1-definition:

$$\mathcal{Y} = \bigcap_Y (Z(Y) \subseteq Y).$$

Details can be found in Beleznay and Foreman.

Issue: to formalize the Furstenberg proof, you need not just the maximal distal factor, but also the structure theorem.

An alternative characterization,

$$f \in \mathcal{Y} \leftrightarrow Y \text{ comes into some well-founded hierarchy } H \ldots$$

is Σ^1_2.

Formalizing the proof in ID_1

Solution: show the map $Y \mapsto Z(Y)$ can be given by a *positive* arithmetic formula, and use the induction principle in ID_1.

Code a factor, Y, as a set of pairs (f, ε) asserting that f is within ε of the factor being represented.

Cast the closure operation as follows: "(f, ε) is in $Z(Y)$ if and only if there is sufficient (positive) information in Y to see that f is within ε of a function that is compact relative to Y."

Corollary (A-T). If X codes a measure-preserving system, the height of the tower is less than or equal to $\omega_1^{CK,X}$. The αth level is computable in $H^X_{2\cdot\alpha}$.

Formalizing the proof in ID_1

Recall that the structure of the Furstenberg proof is as follows:

- By transfinite induction, every factor Y_α is SZ.
- So the maximal distal factor, $\mathcal{Y}$, is SZ.
- $(X, \mathcal{B}, \mu, T)$ is weak mixing relative to $\mathcal{Y}$, so it is SZ too.

Issue: Our "factors" are sets of pairs (f, ε). The property of being SZ is not (a priori) closed under limits.

Solution: Use a strengthening of that property.

Formalizing the proof in ID_1

Recall the finitary statement of Szemerédi's theorem:

Theorem. For every k and $\delta > 0$, there is an n large enough, such that if S is any subset of $\{1, \ldots, n\}$ with density at least δ, then S has an arithmetic progression of length k.

Outline of the Furstenberg proof:

- Suppose for some k and $\delta > 0$ there is no such n.
- Construct the corresponding space $\mathcal{X} = (X, \mathcal{B}, \mu, T)$.
- Every element of $L^2(\mathcal{X})$ is SZ.
- Contradiction.

How do you turn this into a proof that computes an n based on k and δ?

Formalizing the proof in ID_1

Solution: the proof shows

> *For every k and δ, and every μ, there is an n such that if $\mu(A) \geq \delta$, then $\mu(\cap_{i<k} T^{-ik}(A)) > 0$.*

The space of measures μ is compact in the weak-* topology, so, a priori, we know that there is a bound on $n(k, \delta)$ independent of μ.

Proof mining techniques enable one to cast the proof so that we keep track of the bounds.

Conclusions

What we have:

- A computational interpretation of ID_1.
- A detailed sketch of the Furstenberg proof in ID_1.
- A computational (Dialectica) interpretation of the mean ergodic theorem.

What we need to do:

- Apply the computational interpretation to each intermediate lemma.
- Make the result comprehensible.

The mean ergodic theorem is the central nonconstructive component of the Furstenberg proof. It is needed to show that if a system is not weak mixing, it has a nontrivial compact factor.

Conclusions

Two possibilities:

- Worst case: each bit of "unwinding" analytic notions will lead to incomprehensibility.
- Best case: when the analytic notions are specialized to the instances needed here, general baggage is shed, and the resulting concepts are combinatorially natural.

It is too early to make strong claim here, but we are hopeful that the situation is closer to the best-case scenario.

Conclusions

We do not expect to get good bounds from our unwinding. (Note that elementary bounds due to Gowers required significant new ideas.)

What we do hope to get:

- A perspicuous new proof of Szemerédi's theorem.
- Combinatorial ideas that might lead to more general results.
- Possibly logically strong combinatorial statements.
- A better understanding of how ergodic-theoretic methods work, and their combinatorial and computational content.