Computability in ergodic theory

Jeremy Avigad

Department of Philosophy and Department of Mathematical Sciences
Carnegie Mellon University

(joint work with Philipp Gerhardy and Henry Towsner)

August, 2007

Ergodic theory

A *discrete dynamical system* consists of a structure, $\mathcal{X}$, and an map T from $\mathcal{X}$ to $\mathcal{X}$.

In ergodic theory, $\mathcal{X}$ is assumed to be a finite measure space $(X, \mathcal{B}, \mu)$, and T is assumed to be a *measure preserving transformation*, i.e. $\mu(T^{-1}A) = \mu(A)$ for every $A \in \mathcal{B}$.

Call $(X, \mathcal{B}, \mu, T)$ a *measure preserving system*.

- These can model physical systems (e.g. Hamilton's equations preserve Lebesgue measure).
- They can model probabilistic processes.
- They have applications to number theory and combinatorics.

The metamathematics of ergodic theory

Ergodic theory emerged from seventeenth century dynamics and nineteenth century statistical mechanics.

Since Poincaré, the emphasis has been on characterizing structural properties of dynamical systems, especially with respect to long term behavior (stability, recurrence).

Today, the field uses structural, infinitary, and nonconstructive methods that are characteristic of modern mathematics.

These are often at odds with computational concerns.

The metamathematics of ergodic theory

Central questions:

- To what extent can the methods and objects of ergodic theory be given a direct computational interpretation?
- How can we locate the "constructive content" of the nonconstructive methods?

I will focus on two case studies:

- the von Neumann and Birkhoff ergodic theorems; and
- the Furstenberg structure theorem.

The ergodic theorems

Consider the orbit $x, Tx, T^2x, \ldots$, and let $f : \mathcal{X} \to \mathbb{R}$ be some measurement. Consider the averages

$$\frac{1}{n}(f(x) + f(Tx) + \ldots + f(T^{n-1}x)).$$

For each $n \geq 1$, define $A_n f$ to be the function $\frac{1}{n}\sum_{i<n} f \circ T^i$.

Theorem (von Neumann). For every f in $L^2(\mathcal{X})$, $(A_n f)$ converges in the L^2 norm.

Theorem (Birkhoff). For every f in $L^1(\mathcal{X})$, $(A_n f)$ converges pointwise almost everywhere, and in the L^1 norm.

If $\mathcal{X}$ is *ergodic*, then $(A_n f)$ converges to the constant function $\int f \, d\mu$.

Bounding the rate of convergence

Can we compute a bound the rate of convergence of a $(A_n f)$?

In other words: can we compute a function $r : \mathbb{Q} \to \mathbb{N}$ such that for every rational $\varepsilon > 0$,

$$\|A_m f - A_{r(\varepsilon)} f\| < \varepsilon$$

whenever $m \geq r(\varepsilon)$?

Krengel (et al.): convergence can be arbitrarily slow.

But computability is a different question.

Observations

If $(a_n)_{n \in \mathbb{N}}$ is a sequence of reals that decreases to 0, no matter how slowly, one can compute a bound on the rate of convergence from (a_n).

But there are bounded, computable, decreasing sequences (b_n) of rationals that do not have a computable limit.

There are also computable sequences (c_n) of rationals that converge to 0, with no computable bound on the rate of convergence.

Conclusion: at issue is not the *rate* of convergence, but its *predictability*.

A negative result

Theorem (A-Simic). There are a computable measure-preserving transformation of $[0, 1]$ under Lebesgue measure and a computable characteristic function $f = \chi_A$, such that if $f^* = \lim_n A_n f$, then $\|f^*\|_2$ is not a computable real number.

In particular, f^* is not a computable element of $L^2(\mathcal{X})$, and there is no computable bound on the rate of convergence of $(A_n f)$ in either the L^2 or L^1 norm.

A positive result

An measure-preserving transformation T gives rise to an isometry $\hat{T}$ on $L^2(\mathcal{X})$,

$$\hat{T} f = f \circ T.$$

Riesz showed that the von Neumann ergodic theorem holds, more generally, for any nonexpansive operator $\hat{T}$ on a Hilbert space (i.e. satisfying $\|Tf\| \leq \|f\|$ for every f in $\mathcal{H}$.)

Theorem (A-G-T). Let $\hat{T}$ be a nonexpansive operator on a separable Hilbert space and let f be an element of that space. Let $f^* = \lim_n A_n f$. Then f^*, and a bound on the rate of convergence of $(A_n f)$ in the Hilbert space norm, can be computed from f, $\hat{T}$, and $\|f^*\|$.

In particular, if $\hat{T}$ arises from an ergodic transformation T, then f^* is computable from T and f.

A constructive mean ergodic theorem

When there is no computable bound on the rate of convergence, is there anything more we can say?

The assertion that the sequence $(A_n f)$ converges can be represented as follows:

$$\forall \varepsilon > 0 \, \exists n \, \forall m \geq n \, (\|A_m f - A_n f\| < \varepsilon).$$

This is classically equivalent to the assertion that for any function K,

$$\forall \varepsilon > 0 \, \exists n \, \forall m \in [n, K(n)] \, (\|A_m f - A_n f\| < \varepsilon).$$

A constructive mean ergodic theorem

Theorem (A-G-T). Let $\hat{T}$ be any nonexpansive operator on a Hilbert space, let f be any element of that space, and let $\varepsilon > 0$, and let K be any function. Then there is an $n \geq 1$ such that for every m in $[n, K(n)]$, $\|A_m f - A_n f\| < \varepsilon$.

In fact, we provide a bound on n expressed solely in terms of K and $\rho = \|f\|/\varepsilon$ (and independent of $\hat{T}$).

As special cases, we have the following:

- If $K = n^{O(1)}$, then $n(f, \varepsilon) = 2^{2^{O(\rho^2 \log\log \rho)}}$.
- If $K = 2^{O(n)}$, then $n(f, \varepsilon) = 2^{1}_{O(\rho^2)}$.
- If $K = O(n)$ and $\hat{T}$ is an isometry, then $n(f, \varepsilon) = 2^{O(\rho^2 \log \rho)}$.

A constructive pointwise ergodic theorem

The following is classically equivalent to the pointwise ergodic theorem:

Theorem (A-G-T). For every f in $L^2(\mathcal{X})$, $\lambda_1 > 0$, $\lambda_2 > 0$, and K there is an $n \geq 1$ satisfying

$$\mu(\{x \mid \max_{n \leq m \leq K(n)} |A_n f(x) - A_m f(x)| > \lambda_1\}) \leq \lambda_2.$$

We provide explicit bounds on n in terms of f, λ_1, λ_2, and K.

Hard and soft analysis

On his blog, Terence Tao recently emphasized the distinction between "hard" and "soft" analysis.

"Hard" (or "quantitative," or "finitary") analysis deals with the cardinality of finite sets, the measure of bounded sets, the value of convergent integrals, the norm of finite-dimensional vectors, etc.

"Soft" analysis deals with infinitary objects, like sequences, measurable sets and functions, σ-algebras, Banach spaces, etc.

"To put it more symbolically, hard analysis is the mathematics of ε, N, $O()$, and $\leq$; soft analysis is the mathematics of 0, ∞, $\in$, and $\rightarrow$."

Tao independently observed that the methods described here provide "hard" analogues of "soft" results.

Hard and soft analysis

Theorem (Tao). Let $T_1, \ldots, T_l$ be commuting measure preserving transformations of $\mathcal{X}$, and $f_1, \ldots, f_l \in L^\infty(\mathcal{X})$. Then the sequence of "diagonal averages"

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T_1^n x) \cdots f_l(T_l^n x)$$

converges in the L^2 norm.

When $l = 1$, this is essentially the mean ergodic theorem.

Tao's method: run the "Furstenberg correspondence" in reverse, and prove a finitary combinatorial statement by induction on l.

When $l = 1$, this statement is an instance of our constructive MET.

Mixing properties

Ergodicity is equivalent to

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

for every A and B.

A system is *mixing* if we have

$$\lim_{n\to\infty} \mu(T^{-n} A \cap B) = \mu(A)\mu(B).$$

A system is *weak mixing* if we have

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

Compactness

A system is said to be *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.

A weak mixing system exhibits a high degree of randomness.

Can we decompose an arbitrary system into a combination of the two?

The Furstenberg structure theorem

Let $\mathcal{X} = (X, \mathcal{B}, \mu, T)$ be a measure preserving system. (Henceforth, assume T is invertible.)

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 factor:

- $(X, \mathcal{B}', \mu, T)$, for a T-invariant sub-σ-algebra $\mathcal{B}' \subseteq \mathcal{B}$
- A homomorphic image, or quotient, of $\mathcal{X}$.
- A (suitable) $\hat{T}$-invariant subspace of $L^2(\mathcal{X})$.

The Furstenberg structure theorem

The notions of compactness and weak mixing relativize to factors.

Lemma (Furstenberg). If a system $(X, \mathcal{B}, \mu, T)$ is not weak mixing relative to a factor $\mathcal{B}'$, there there is an intermediate factor $\mathcal{B}''$ 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. Let $(X, \mathcal{B}, \mu, T)$ be any measure preserving system. Then there is a transfinite increasing sequence of factors $(\mathcal{B}_\alpha)_{\alpha \leq \gamma}$ such that:

- B_0 is the trivial factor.
- For each $\alpha < \gamma$, $(X, B_{\alpha+1}, \mu, T)$ is compact relative to (X, B_α, μ, T).
- For each limit $\lambda \leq \gamma$, $\mathcal{B}_\lambda = \cup_{\alpha<\lambda} \mathcal{B}_\alpha$.
- Either $\mathcal{B}_\gamma = \mathcal{B}$, or $(X, \mathcal{B}, \mu, T)$ is weakly mixing relative to $(X, \mathcal{B}_\gamma, \mu, T)$.

Szemerédi's theorem

A sturcture theorem has a direct application to combinatorics:

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.

Szemerédi proved the theorem in 1975. Two years later, Furstenberg provided a new proof, by translating the theorem into measure-theoretic terms.

Furstenberg correspondence

Suppose we are given a sequence of subsets S_n of $\{1, \ldots, n\}$ of density $\delta > 0$.

Let T be the shift map on $2^{\mathbb{Z}}$. We can thin out the sequence and define a T-invariant measure μ on $2^{\mathbb{Z}}$ such that for any finite pattern σ, $\mu([\sigma])$ is the limit of the density of that pattern in the S_n'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 such that

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

Furstenberg's proof

If the space is weak mixing, the theorem holds, because the events are close to uncorrelated.

If the space is compact, the theorem holds, because events come close to recurring.

More generally, define a stronger inductive hypotheses, that

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

Analysis of the structure theorem

Tao writes:

> *This ergodic theory argument is the shortest and most flexible of all the known proofs, and has been the most successful at leading to further generalizations of Szemerédi's theorem… On the other hand, the infinitary nature of the argument means that it does not obviously provide any effective bounds for the quantity* $N_{SZ}(k, \delta)$.

But what makes the argument nonconstructive?

Analysis of the structure theorem

Kra writes:

> *Furstenberg's proof relies on a compactness argument, making it difficult to extract any explicit bounds in the finite version of Szemerédi's theorem.*

She seems to be referring to the combinatorial compactness argument implicit in the correspondence principle. Furstenberg writes:

> *However, the ergodic-theoretic approach depends essentially on passing to a limit whereby a set* $\{1, 2, 3, \ldots, N\}$ *is replaced by a measure space, and the translations* $n \to n + a$ *are replaced by measure preserving transformations of this space. In passing to this limit one loses sight of the size* N *of the interval* $\{1, 2, 3, \ldots, N\}$. *As a result this approach is incapable of giving any information regarding [*$N_{SZ}(k, \delta)$*] beyond the fact that it is finite.*

Analysis of the structure theorem

The correspondence principle can be reduced to the task of picking a path through an infinite binary tree computable from S'.

In general, such a path is not computable from S'. But, by the Jockusch-Soare low-basis theorem, there is a path that is low in S'.

There are proof-theoretic techniques for eliminating “weak König’s lemma,” and other techniques for handling mild uses of arithmetic comprehension.

If this were the only use of nonconstructivity, the argument would be pretty tame.

Analysis of the structure theorem

The transfinite iteration should seem suspect. But there is nothing inherently wrong with transfinite recursion.

Define the set of (full) well-founded trees on $\mathbb{N}$ inductively:

- e (one node only) is a well-founded tree;
- If $f(n)$ is a well-founded tree for every n, then one can make these the subtrees of a new root.

Then definition by recursion is constructively valid:

$$F(e) = a$$
$$F(T) = G(\lambda n F(T_{(n)})) \quad \text{if } T \text{ is not } e$$

Analysis of the structure theorem

So where's the problem?

Answer: taking limits, or projections, at each stage. The transfinite iteration then amplifies the problem.

Theorem (Beleznay and Foreman). The Furstenberg structure theorem exhausts the countable ordinals.

Observation (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 $H_{2\cdot\alpha}^X$.

We suspect that this is sharp, at least for limit α. This means that the Furstenberg tower is a wildly noncomputable object.

Analysis of the structure theorem

And yet, the structure theorem can be used to prove an explicit combinatorial result. How does this work?

Our explanation:

- The argument can be carried out in Kreisel's theory ID_1 of arithmetic inductive definitions.
- Proofs in ID_1 have constructive interpretations.

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.

The theory ID_1 adds these axioms:

- $\forall x\ (\psi(P, x) \rightarrow\ P(x))$
- $\forall x\ (\psi(\theta/P, x) \rightarrow \theta(x)) \rightarrow \forall x\ (P(x) \rightarrow \theta(x))$, for each formula θ.

These express that P is the least fixed point of Γ_ψ.

The constructive theory $ID_1^{i,acc}$

For a constructive version, restrict to intuitionistic logic, and insist that the inductive definitions be *accessibility definitions*:

$$\forall y \; (y \prec x \to P(y))$$

These pick out the well-founded part of a primitive recursive relation, $\prec$.

Locating the constructive content of a theory

A Π_2 sentence is one of the form $\forall \bar{x}\ \exists \bar{y}\ R(\bar{x}, \bar{y})$.

Szemerédi's theorem has this form.

Two ways of characterizing the Π_2 consequences of a theory:

1. Every Π_2 sentence provable in T is also provable in a constructive theory, T'.
2. Every Π_2 sentence provable in T is witnessed by an element of a particular class of computable functions, C.

The constructive content of *PA*

Define the set of *finite types*:

- N is a finite type; and
- assuming σ and τ are finite types, so are $\sigma \times \tau$ and $\sigma \to \tau$.

The set of *primitive recursive functionals of finite type* is a set of computable functionals obtained from the use of explicit definition (λ abstraction), application, pairing, and projections, and a scheme of primitive recursion:

$$F(0) = a$$
$$F(n+1) = G(n, F(n))$$

where the range of F may be any finite type.

The constructive content of PA

Theorem. Every Π_2 theorem of PA is provable in HA.

Theorem. Every Π_2 theorem of PA is witnessed by a primitive recursive functional of type $N \to N$.

The constructive content of *PA*

To prove this, first use apply double-negation interpretation to PA. The sentence $\forall \bar{x} \; \exists \bar{y} \; R(\bar{x}, \bar{y})$ becomes $\forall \bar{x} \; \neg\neg\exists \bar{y} \; R(\bar{x}, \bar{y})$

From here, there are two ways of proceeding. One option:

- Use the Friedman-Dragalin A-translation to “repair” the interpretation of Π_2 sentences.
- Apply modified realizability.

Another option:

- Use the Dialectica interpretation to extract a primitive recursive functional.
- Interpret the result in *HA*.

Experience has shown that the second is better for proof mining.

The constructive content of ID_1

Extend the finite types by adding a new base type, Ω, which is intended to denote the set of well-founded (full) trees on $\mathbb{N}$.

Add two new operations:

- *Sup*, of type $(N \to \Omega) \to \Omega$, forms a new tree from a sequence of subtrees;
- Sup^{-1}, of type $\Omega \to (N \to \Omega)$, which returns the immediate subtrees of a nontrivial tree.

Add the principle of recursive definition:

$$F(e) = a$$
$$F(Sup(h)) = G(\lambda n\ F(h(n)))$$

Call these the *primitive recursive tree functionals*.

The constructive content of ID_1

Theorem. Every Π_2 theorem of ID_1 is provable in $ID_1^{i,acc}$.

Theorem. Every Π_2 theorem of ID_1 is witnessed by a primitive recursive tree functional of type $N \to N$.

The constructive content of ID_1

Once again, there are two ways of obtaining this result. In both cases, start with a double-negation translation.

From there, one can apply a method by Buchholz:

- Use a complex forcing translation.
- Apply modified realizability.

Towsner and I have recently developed an alternative route:

- Use a Dialectica interpretation.
- Interpret the result in $ID_1^{i,acc}$.

Conclusions

This provides a strategy for interpreting the Furstenberg proof: formalize in ID_1, and apply the Dialetica translation.

Towsner and I are working on turning this into a *readable* proof.

Goals:

- A perspicuous new proof of Szemerédi's theorem.
- A better understanding of the combinatorial content of the structure theorem.
- New combinatorial methods, which may generalize.
- Possibly combinatorial independences, à la Friedman.
- A better understanding of the use of infinitary methods in combinatorics.