Computability in ergodic theory

Jeremy Avigad

Department of Philosophy and Department of Mathematical Sciences
Carnegie Mellon University

(joint work with Philipp Gerhardy, Ksenija Simic, and Henry Towsner)

November, 2007
A *discrete dynamical system* consists of a structure, $X$, and an map $T$ from $X$ to $X$:

- Think of the underlying set of $X$ as the set of states of a system.
- If $x$ is a state, $T x$ gives the state after one unit of time.

In ergodic theory, $X$ is assumed to be a finite measure space $(X, \mathcal{B}, \mu)$:

- $\mathcal{B}$ is a $\sigma$-algebra (the “measurable subsets”).
- $\mu$ is a $\sigma$-additive measure, with $\mu(X) = 1$.

$T$ 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.
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’ll start with an overview of some results:

- the von Neumann and Birkhoff ergodic theorems
- negative results
- positive results

Then, as time allows, I’ll present some of the details.
Consider the orbit $x, T x, T^2 x, \ldots$, and let $f : \mathcal{X} \to \mathbb{R}$ be some measurement. Consider the averages

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

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

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

A space is **ergodic** if for every $A$, $T^{-1}(A) = A$ implies $\mu(A) = 0$ or $\mu(A) = 1$.

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

The ergodic theorems

Recall that $L^2(\mathcal{X})$ is the Hilbert space of square-integrable functions on $\mathcal{X}$ modulo a.e. equivalence, with inner product

$$(f, g) = \int fg \, d\mu$$

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

A 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 $\|T f\| \leq \|f\|$ for every $f$ in $\mathcal{H}$.)
Bounding the rate of convergence

Can we compute a bound on the rate of convergence of \((A_n f)\) from the initial data \((T \text{ and } f)\)?

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.

Note that the question depends on suitable notions of computability in analysis (I’ll come back to this).
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.
**Theorem (A-S).** 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

**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 \(K\),

\[ \forall \varepsilon > 0 \ \exists n \ \forall m \in [n, K(n)] \ (\| A_m f - A_n f \| < \varepsilon). \]
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 $n(f, \varepsilon) = 2^{O(\rho^2 \log \log \rho)}$.
- If $K = 2^{O(n)}$, then $n(f, \varepsilon) = 2^{O(\rho^2)}$.
- 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 $f$, $\lambda_1$, $\lambda_2$, and $K$.
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 $\varepsilon$, $N$, $O()$, and $\leq$; soft analysis is the mathematics of $0$, $\infty$, $\varepsilon$, and $\rightarrow$.

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

When $l = 1$, this statement is an instance of our constructive MET.
Thus ends the overview. Now for some of the details:

- Notions of computability in analysis.
- A proof of the mean ergodic theorem.
- Noncomputability of the rate of convergence.
- Computability of the rate of convergence from $\| f^* \|$.
- Our constructive mean ergodic theorem.
**Definition.** A real number $r \in \mathbb{R}$ is *computable* if there is a computable function $\alpha : \mathbb{N} \to \mathbb{Q}$ such that $\lim_{n \to \infty} \alpha(n) = r$ and

$$\forall n \forall m \geq n \ (|\alpha(m) - \alpha(n)| < \frac{1}{2^n}).$$

In other words, $\alpha$ is a computable Cauchy sequence, with an explicit rate of convergence, representing $r$.

**Definition.** A function $f : \mathbb{R} \to \mathbb{R}$ is computable if there is a computable function $F(\alpha, n)$, such that whenever $\alpha$ represents a real number $x$, $\lambda n. F(\alpha, n)$ represents the real number $f(x)$.

**Fact.** With the obvious extension to binary functions, addition and multiplication are computable.

Note: computable implies continuous.
Computability in analysis

What is special about \( \mathbb{R} \)?

- There is a countable dense subset, \( \mathbb{Q} \).
- One can construct \( \mathbb{R} \) as the Cauchy completion of \( \mathbb{Q} \).
- With a natural encoding of \( \mathbb{Q} \), operations we care about are computable.

The idea generalizes to separable metric spaces, and structures built on these.

For example: a computable Hilbert space is given by a countable set \( S \), operations \( + \) and \( x \mapsto q \cdot x \) for \( q \in \mathbb{Q} \), and an inner product \( (x, y) \), such that \( S \) is an inner product space in the usual sense; the corresponding Hilbert space is the Cauchy completion.

**Fact.** A bounded linear operator \( T \) can be defined, equivalently, by its operation on \( S \).
How to handle measure spaces? Think of $[0, 1]$ under Lebesgue measure, or $\{0, 1\}^\omega$ under coin-flipping measure.

Define a countable algebra of “simple” sets $C$, and a $\sigma$-additive measure $\mu$ on those.

Then define the $\sigma$-algebra of measurable functions to be the completion of $C$ under the metric $d(A, B) = \mu(A \triangle B)$, modulo the relation $C \approx D$ given by $\mu(C \triangle D) = 0$.

Alternative approach: define a countable set $C$ of simple functions, with an “integration” operation $f$. Define the $L^1$ space as a completion.

In the usual cases, these turn out to be equivalent. Note that a measurable set of function is only defined up to points of measure 0.