Computability and uniformity in ergodic theory

Jeremy Avigad

Department of Philosophy and
Department of Mathematical Sciences
Carnegie Mellon University

March, 2015

Convergence in analysis

Many theorems of analysis assert that, under various hypotheses, a certain type of sequence (a_n) converges.

Questions:

- To what extent can one compute a rate of convergence?
- What parameters does the rate of convergence depend on?

When rates are noncomputable and nonuniform, it is especially important to know:

- What other data can be computed?
- What other uniformities can be obtained?

Computability in analysis

A *name* for a real number is a Cauchy sequence (a_n) of rationals such that for every m and $n \geq m$, $|a_n - a_m| \leq 2^{-m}$.

A real number r is *computable* if it has a computable name.

A computable $f : \mathbb{R} \to \mathbb{R}$ takes a name (presented as an oracle) to a name.

Theorem (Specker)

There is a computable, nondecreasing sequence (a_n) of rationals in $[0, 1]$ with no computable limit.

Rates of convergence

Suppose (a_n) is Cauchy:

$$\forall \varepsilon > 0 \, \exists m \, \forall n, n' \geq n \; d(a_{n'}, a_n) < \varepsilon$$

A function $r(\varepsilon)$ satisfying

$$\forall n, n' \geq r(\varepsilon) \; d(a_{n'}, a_n) < \varepsilon$$

is called a *bound on the rate of convergence*.

For computable (a_n), if there is a computable bound on the rate of convergence, then it has a computable limit (but the converse does not necessarily hold).

Finiteness

Let α be an infinite sequence of 0's and 1's.

Three ways to say "there are finitely many 1's":

1. For some n, there are no 1's beyond position n.
2. For some k, there are at most k-many 1's.
3. There are not infinitely many 1's.

These make very different existence claims:

1. $\exists n\, \forall m \geq n\, \alpha(m) \neq 1$
2. $\exists k\, \forall m\, |\{i \leq m \mid \alpha(i) = 1\}| \leq k$
3. $\forall f\, \exists n\, (f(n) > n \rightarrow \alpha(f(n)) \neq 1)$.

(See Bezem, Nakata, Uustalu, "Streams that are finitely red.")

Convergence

Corresponding ways of saying that a sequence (a_n) in a complete space converges:

1. (a_n) is Cauchy.
2. For every $\varepsilon > 0$, (a_n) has finitely many ε-fluctuations.
3. (a_n) is metastably convergent.

These call for three types of information:

1. A bound on the rate of convergence.
2. A bound on the number of fluctuations.
3. A bound on the rate of metastability.

Outline

- Computability, uniformity, and convergence
- Rates of convergence in the mean ergodic theorem
- Bounds on the number of oscilliations
- Metastability
- Pointwise a.e. convergence

Ergodic theory

A *measure-preserving system* $\mathcal{X} = (X, \mathcal{B}, \mu, T)$ consists of:

- a set, X (the "states" of the system)
- a σ-algebra, $\mathcal{B}$ a (the "measurable subsets")
- a finite σ-additive measure, μ; wlog $\mu(X) = 1$
- a *measure-preserving transformation*, T: $\mu(T^{-1}A) = \mu(A)$ for every $A \in \mathcal{B}$

If x is a state, think of Tx as being the state after one unit of time.

The system is said to be *ergodic* if there are no non-trivial T-invariant subsets; in other words, $T^{-1}(A) = A$ implies $\mu(A) = 0$ or $\mu(A) = 1$.

The pointwise ergodic theorem

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 (Birkhoff)

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

The limit, f^*, is T-invariant, that is, $f^* \circ T = f^*$.

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

The mean ergodic theorem

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

$$\langle f, g \rangle = \int fg \; d\mu$$

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 $\|\hat{T} f\| \leq \|f\|$ for every f in $\mathcal{H}$).

Ergodic theory

Applications:

- Stochastic processes ($\mu(A)$ is the probability of being in state A)
- Statistical mechanics
- Physics (e.g. evolution by Hamilton's equations preserves Lebesgue measure)
- Diophantine analysis
- Additive combinatorics

Rates of convergence

Let us focus on the mean ergodic theorem.

Question: can we compute a bound on the rate of convergence of $(A_n f)$ from the inital data (T and f)?

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

$$\|A_n f - A_{n'} f\| < \varepsilon$$

whenever $n, n' \geq r(\varepsilon)$?

Krengel (et al.): convergence can be arbitrarily slow. But computability is a different question.

Noncomputability

Observation (Bishop): the ergodic theorems imply the limited principle of omniscience.

Theorem (V'yugin)

There is a computable shift-invariant measure μ on 2^ω such that $\lim_{n\to\infty} A_n 1_{[1]}$ is not computable.

Noncomputability

This is essentially a recasting of V'yugin's result:

Theorem (Avigad and 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.

In general, everything is computable from $0'$, and this is sharp.

Computability

Theorem (Avigad, Gerhardy, and Towsner)

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.

Jason Rute and I have shown that this generalizes to a uniformly convex Banach space.

Oscillations

Definition

Say that (a_n) *admits m ε-fluctuations* if there are $i_1 \leq j_1 \leq \ldots \leq i_m \leq j_m$ such that, for each $u = 1, \ldots, m$, $d(a_{i_u}, a_{j_u}) \geq \varepsilon$.

These are also sometimes called ε-jumps, or ε-oscillations.

A moment's reflection shows that (a_n) is Cauchy if and only if for every $\varepsilon > 0$, it admits only finitely many ε-fluctuations.

Call a bound $\varepsilon \mapsto k(\varepsilon)$ on m a *bound on the number of fluctuations*.

Oscillations

Specker's examples show that for (a_n) is a nondecreasing sequence in a closed interval $[a, b]$, there may not be a computable bound on the rate of convergence.

Clearly there is no uniform bound either.

But there are at most $\lceil (b - a)/\varepsilon \rceil$ many ε-fluctuations.

This is easily computable, and very uniform.

(For bounded sequences, bounds on oscillations are closely related to bounds on upcrossings.)

Oscillations

Say the *total variation* of a sequence (a_n) in a metric space is $\sum_n d(a_n, a_{n+1})$.

If the total variation of a sequence is less than B, then (using the triangle inequality) there are at most $\lceil B/\varepsilon \rceil$-many ε-fluctuations.

For the mean ergodic theorem, though, this is too strong. Consider $\mathbb{R}$ as a 1-dimensional Hilbert space, with $Tx = -x$.

The orbit of 1 is

$$1, -1, 1, -1, \ldots$$

and the averages are

$$1, 0, 1/3, 0, 1/5, 0, \ldots$$

and the total variation diverges.

A variational inequality

Theorem (Jones, Ostrovskii, and Rosenblatt)

Let T be any nonexpansive operator on a Hilbert space $\mathcal{H}$, and $x \in \mathcal{H}$. Then for any sequence $n_1 \leq n_2 \leq \ldots$,

$$(\sum_{k=1}^{\infty} \|A_{n_{k+1}} x - A_{n_k} x\|^2)^{1/2} \leq 25\|x\|.$$

This implies that, in particular, the number of ε-fluctuations is at most $(25\|x\|/\varepsilon)^2$.

A variational inequality

Consider the case where T is an isometry.

By the spectral theorem, it suffices to prove the theorem in the case where $Tx = e^{i\theta}x$ is a rotation of the unit circle $\{z \in \mathbb{C} \mid |z| = 1\}$.

To bound $\sum_k \|A_{n_{k+1}}x - A_{n_k}x\|^2$, divide $\mathbb{N}$ into two sets:

- $S = \{k \mid |n_{k+1}\theta - n_k\theta| < 1\}$ (the "short" jumps)
- $L = \{k \mid |n_{k+1}\theta - n_k\theta| \geq 1\}$ (the "long" jumps)

A variational inequality

$\sum_{k \in S} \|A_{n_{k+1}} x - A_{n_k} x\|^2$ is small because the individual differences are small.

$\sum_{k \in L} \|A_{n_{k+1}} x - A_{n_k} x\|^2$ is small because the n_k's increase fast enough.

Jones, Kaufman, Rosenblatt, Wierdl used this analysis, together with ideas from Bourgain, to obtain pointwise variational inequalities.

Uniformly convex spaces

Definition

A Banach space $\mathbb{B}$ is *uniformly convex* if for every $\varepsilon \in (0, 2]$ there exists a $\delta \in (0, 1]$ such that for all $x, y \in \mathbb{B}$, if $\|x\| \leq 1$, $\|y\| \leq 1$, and $\|x - y\| \geq \varepsilon$, then $\|(x + y)/2\| \leq 1 - \delta$.

$L_p(\mathcal{X})$ for $1 < p < \infty$ are uniformly convex, but not $L_1(\mathcal{X})$ or $L_\infty(\mathcal{X})$.

Pisier has shown that any uniformly convex Banach space is isomorphic to one with modulus of uniform convexity $\eta(\varepsilon) = K\varepsilon^p$ for some $p \geq 2$.

In 1939, Garrett Birkhoff gave a short and elegant proof that the mean ergodic theorem holds for uniformly convex spaces.

Uniformly convex spaces

Theorem (Avigad and Rute)

Let $p \geq 2$ and let $\mathbb{B}$ be any p-uniformly convex Banach space. Let T be a linear operator on $\mathbb{B}$ satisfying $B_1\|y\| \leq \|T^n y\| \leq B_2\|y\|$ for every n and $y \in \mathbb{B}$, for some $B_1, B_2 > 0$. Then for any x in $\mathbb{B}$ and any increasing sequence $(t_k)_{k\in\mathbb{N}}$,

$$\sum_k \|A_{t_{k+1}}x - A_{t_k}x\|^p \leq C\|x\|^p$$

for some constant C (depending only on B_1, B_2, K, and p).

The difficulty is that the spectral theorem does not apply here.

Uniformly convex spaces

Theorem (Pisier)

Suppose $2 \le p < \infty$. The following are equivalent:

1. *$\mathbb{B}$ is isomorphic to a p-uniformly convex Banach space.*
2. *There is a constant, C, such that if $(M_n)_{n \ge 0}$ is any martingale in $L^p(X;\mathbb{B})$, then*

$$\|M_0\|^p_{L^p(X;\mathbb{B})} + \sum_{n \ge 0} \|M_{n+1} - M_n\|^p_{L^p(X;\mathbb{B})} \le C \sup_{n \ge 0} \|M_n\|^p_{L^p(X;\mathbb{B})}.$$

The idea behind our proof:

- For $L^p(\mathbb{R}^n)$, Jones, Kaufman, Rosenblatt, and Wierdl later use martingales instead of spectral analysis.
- They obtain a key result for $l^p(\mathbb{R})$, and then use transfer.
- We use Pisier's theorem to obtain the result for $l^p(\mathbb{B})$.
- Use a novel transfer argument to pass from $l^p(\mathbb{B})$ to $\mathbb{B}$.

Metastability

Recall that (a_n) is Cauchy if

$$\forall \varepsilon > 0 \; \exists m \, \forall n, n' \geq m \; d(a_n, a_{n'}) < \varepsilon$$

But in general m is not computable from (a_n) and ε.

The statement above is equivalent to

$$\forall \varepsilon > 0, F \; \exists m \, \forall n, n' \in [m, F(m)] \; d(a_n, a_{n'}) < \varepsilon.$$

Given $\varepsilon > 0$ and F, one can find such an m by blind search.

Call $M(F, \varepsilon)$ a *bound on the rate of metastability* if it is a bound on such an m.

Metastability

Note that the bound on the number of ε-fluctuations in the last theorem depends only on $\|x\|/\varepsilon$, B_1, B_2, K, p (and otherwise not on $\mathbb{B}$ or T).

The metastable formulation of the mean ergodic theorem says that for any function F,

$$\forall \varepsilon > 0\, \exists m\, \forall n, n' \in [m, F(m)]\, (\|A_n f - A_{n'} f\| < \varepsilon).$$

The results above give a uniform and explicit bound on m.

We will see that this uniformity holds much more generally.

Metastability

The metastable translation of a convergence statement is an instance of Kreisel's "no-counterexample interpretation," which is, in turn, special case of the Gödel's *Dialectica* interpretation.

Ulrich Kohlenbach has developed extensive "proof mining" methods based on these ideas.

Without knowing about Jones, Rosenblatt, and Ostrovskii's result, Gerhardy, Towsner, and I gave such bounds for the Hilbert space case in 2007.

Kohlenbach and Leuștean extended this to uniformly convex Banach spaces.

Kohlenbach and his students have extended the analysis to many other settings.

Metastability

Metastability played a role in the Tao-Green proof that there are arbitrarily long arithmetic progressions in the primes.

In 2007, Tao used metastability to prove a generalization of the mean ergodic theorem to certain "multiple" averages. There are now other proofs that do not use metastability.

In 2012, Miguel Walsh used metastability to generalize Tao's result to nilpotent group actions.

In a blog post, Tao has given an alternate proof of Walsh's theorem, using nonstandard analysis.

Metastability

In these applications, it is the *uniformity* given by a metastable bound that proves useful.

José Iovino and I noticed that such uniformities are easily obtained using a compactness argument.

Ultraproducts in analysis

Let I be any infinite set, D be a nonprincipal ultrafilter on I.

Suppose that for each i, (X_i, d_i) is a metric space with a distinguished point a_i.

Let

$$X_\infty = \left\{ (x_i) \in \prod_{i \in I} X_i \mid \sup_i d(x_i, a_i) < \infty \right\} / \sim,$$

where $(x_i) \sim (y_i)$ if and only if $\lim_{i,D} d(x_i, y_i) = 0$.

Call this the "ultraproduct of the spaces X_i." This works for more general metric structures.

Metastability and ultraproducts

Theorem (Avigad and Iovino)

Let C be a collection of pairs $((X,d),(a_n)_{n\in\mathbb{N}})$. Fix a nonprincipal ultrafilter. The following statements are equivalent:

1. *There is a uniform bound on the rate of metastability for the sequences (a_n).*
2. *For any sequence $((X_k,d_k),(a_n^k))_{k\in\mathbb{N}}$ of elements of C, the sequence $(\bar{a}_n)$ in the ultraproduct is Cauchy.*

The first clause means: for every $F:\mathbb{N}\to\mathbb{N}$ and $\varepsilon>0$, there is a b with the following property: for every pair $((X,d),(a_n)_{n\in\mathbb{N}})$ in C, there is an $n\leq b$ such that $d(a_i,a_j)<\varepsilon$ for every $i,j\in[n,F(n)]$.

Metastability and ultraproducts

What this means: if you have a convergence theorem, and

- the class of structures described by the theorem is closed under ultraproducts, and
- and the hypotheses are preserved by ultraproducts,

then there is a uniform bound on the rate of metastability.

There are sufficient syntactic conditions for these conditions to hold.

Metastability and ultraproducts

A strong version of the mean ergodic theorem:

Theorem (Lorch, Riesz, Yosida, Kakutani)

If T is any power-bounded linear operator on a reflexive Banach space $\mathbb{B}$, and x is any element of $\mathbb{B}$, then the sequence $(A_n x)_{x \in \mathbb{N}}$ converges.

Alas, the class of reflexive Banach spaces is not closed under ultraproducts.

But for fixed p, the p-uniformly convex spaces are, confirming the uniformity in that case.

Metastability and ultraproducts

There are other collections of reflexive Banach spaces preserved under ultraproducts: uniformly nonsquare Banach spaces, J-(n, ε) convex Banach spaces, etc.

The result also shows that mere convergence in the Tao / Walsh results implies uniformity.

It also provides short confirmations of other uniformities uncovered by Kohlenbach and students.

Metastability and ultraproducts

If you prove a convergence theorem, you know it is true.

- Closure under ultraproducts then tells you that there are uniform bounds on the rate of metastability.
- Under general computability hypotheses, there is even a computable bound (Rute).

Alternatively, using Kohlenbach's methods:

- If the proof can be carried out in a certain (strong) theory, and the theorem has a certain logical form, you get uniformity and computability at once.
- Precise details of the theory give you more information about the computation.
- Analysis of the proof gives you an explicit bound.
- The methods can also handle non-continuous functions.

Pointwise convergence a.e.

A sequence of functions (f_n) in on a finite measure space converges a.e. if and only if

$$\forall \varepsilon > 0, \lambda > 0 \, \exists m \, \forall n, n' \geq m \; \mu\{x \mid |f_n(x) - f_{n'}(x)| \geq \varepsilon\} < \lambda.$$

One can similarly consider rates of convergence, variational inequalities, and metastable versions.

For example, Jones, Kaufman, Rosenblatt, and Wierdl show:

Theorem

Let $p \geq 2$, $f \in L^p$. Then for any increasing $(t_k)_{k \in \mathbb{N}}$,

$$\left\| \left(\sum_k \sup_{u,v \in [t_k, t_{k+1}]} |A_u f - A_v f|^p \right)^{1/p} \right\|_p \leq C_p \cdot \|f\|_p.$$

Pointwise convergence a.e.

Another one:

Theorem

Let $q > p \geq 2$, $f \in L^q$.

$$\left\| \left(\sup_{(t_k)} \sum_k |A_{t_{k+1}} f - A_{t_k} f|^q \right)^{1/q} \right\|_p \leq C_{p,q} \cdot \|f\|_p .$$

Summary

Given a convergence theorem, you can ask:

- Are there computable / uniform bounds on the rate of convergence?
- Are there computable / uniform bounds on the number of oscillations?
- Are there computable / uniform bounds on the rate of metastability?

General questions:

- Can this information be mined systematically from the original proofs?
- What mathematical / computational / combinatorial information does this provide?