Computability, constructivity, and convergence in measure theory

Jeremy Avigad

Department of Philosophy and
Department of Mathematical Sciences
Carnegie Mellon University

November, 2011

Computability and constructivity

For most of its history, mathematics was fairly constructive:

- Euclidean geometry was based on geometric construction.
- Algebra sought explicit solutions to equations.
- Analysis, probability, etc. were focused on calculations.

Nineteenth century developments challenged this view:

- Analysis: "arbitrary" limits and functions
- Algebra: representation-free characterizations of infinitary objects, infinitary operations.

Computability and constructivity

Reactions:

- Kronecker: focus on symbolic calculation
- Brouwer: focus on construction in intuition
- Bishop: focus on constructivity

Hilbert's program:

- Preserves nonconstructive, infinitary methods.
- Regains explicit computation at the metalevel.
- Gives Π_1 and Π_2 theorems concrete meaning.

Weyl's 1927 assessment

[Hilbert] succeeded in saving classical mathematics by a *radical reinterpretation of its meaning* without reducing its inventory, namely, by formalizing it, thus transforming it in principle from a system of intuitive results into a game with formulas that proceeds according to fixed rules. ...

He asserted, first of all, that the passage through ideal propositions is a legitimate formal device when real propositions are proved; this even the strictest intuitionist must acknowledge. ...

But Hilbert furthermore pointed with emphasis to the related science of *theoretical physics*. Its individual assumptions and laws have no meaning that can immediately be realized in intuition; in principle, it is not the propositions of physics taken in isolation, but only the theoretical system as a whole, that can be confronted with experience.

Weyl's 1927 assessment

If Hilbert's view prevails over intuitionism, as appears to be the case, *then I see in this a decisive defeat of the philosophical attitude of pure phenomenology,* which thus proves to be insufficient for the understanding of creative science even in the area of cognition that is most primal and most readily open to evidence—mathematics.

Computability and constructivity

Weyl's interpretation of Hilbert's views:

- There are rules that govern mathematical reasoning.
- Mathematics has concrete "observable" consequences.
- Mathematical theorems do not have meaning taken in isolation; rather, the meaning is spread across the system as a whole.

Faced with a nonconstructive development, one can:

- Seek constructive versions (constructive mathematics, computational mathematics).
- Calibrate degree of nonconstructivity (recursion theory, descriptive set theory, reverse mathematics).
- Seek hidden computational interpretations and consequences (proof theory, proof mining).

Overview

We will consider

- the ergodic theorems
- measure theoretic convergence theorems

with respect to

- computability and constructivity
- reverse mathematics
- algorithmic randomness
- metastability

Dynamical systems

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

- Think of the underlying set of $\mathcal{X}$ as the set of states of a system.
- If x is a state, Tx gives the state after one unit of time.

These can be used to model:

- physical systems
- stochastic processes

The issues arise here:

- computation is important
- so is limiting behavior, structural properties

Ergodic theory

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

- $\mathcal{B}$ is a σ-algebra (the "measurable subsets").
- μ is a σ-additive measure, wlog $\mu(X) = 1$.

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*.

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

Computability

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 $m : \mathbb{Q} \to \mathbb{N}$ such that for every rational $\varepsilon > 0$,

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

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

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.

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 there is no computable bound on the rate of convergence of* $A_n 1_{[1]}$.

Theorem (A)

There is a computable shift-invariant measure μ *on* 2^ω *such that there is no computable bound on the complexity of* $\lim_{n\to\infty} A_n 1_{[1]}$.

Noncomputability

This is essentially a recasting of V'yugin's 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.

Theorem (A-Simic, Simic)

Over RCA_0*, the mean and pointwise ergodic theorems are equivalent to* (ACA)*.*

Computability

Theorem (A-Gerhardy-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.

Constructivity

Let (a_n) be a bounded sequence of reals. TFAE:

- (a_n) converges.
- For every $\alpha < \beta$, the sequence has only finitely many upcrossings.

Let $\omega_{\alpha,\beta}(x)$ be the number of upcrossings of $(A_n f(x))_{n\in\mathbb{N}}$.

Theorem (Bishop)

For any f in $L^1(\mathcal{X})$ and $\alpha < \beta$, we have

$$\int_X \omega_{\alpha,\beta}\, d\mu \leq \frac{1}{\beta - \alpha} \int_X (f - \alpha)^+\, d\mu.$$

Nuber also provided an "equal conclusion" version of the pointwise ergodic theorem.

The mean ergodic theorem

The Riesz proof of the mean ergodic theorem shows that if $\mathcal{H}$ is a Hilbert space and $\hat{T}$ is nonexpansive, then $\mathcal{H} = \mathcal{M} \oplus \mathcal{N}$ where

- $\mathcal{M} = \{f \mid \hat{T} f = f\}$
- $\mathcal{N} = \overline{\{\hat{T} g - g \mid g \in \mathcal{H}\}}$

If f is in $\mathcal{M}$, $A_n f = f$ for every n.

If f is in $\mathcal{N}$, then $A_n f \to 0$.

Thus $A_n f$ converges to $P_{\mathcal{M}}(f)$, the projection of f onto $\mathcal{M}$.

It is the use of a projection that makes the proof nonconstructive.

Constructivity

Theorem (Spitters)

Constructively, for f in $L^1 \cap L^\infty$, the following are equivalent:

1. $P_{\mathcal{N}}(f)$ *exists.*
2. $(A_n f)$ *converges in the L^2 norm.*

However:

Theorem (A-Simic)

Each of the following is equivalent to (ACA) over RCA_0:

- *The mean ergodic theorem ("for every f, $P_{\mathcal{N}}(f)$ exists")*
- *"for every f, if $P_{\mathcal{M}}(f)$ exists then $P_{\mathcal{N}}(f)$ exists."*
- *"for every f, if $P_{\mathcal{M}}(f) = 0$ then $P_{\mathcal{N}}(f)$ exists."*

Algorithmic randomness

A *Martin-Löf test*, or *effectively null G_δ set*, is a set $\bigcap_n G_n$ where

- the G_n's are uniformly recursively open, and
- for each n, $\mu(G_n) < 2^{-n}$.

ω is Martin-Löf random if it avoids (passes) every Martin-Löf test.

Theorem (V'yugin)

Let μ be a computable shift-invariant measure on 2^ω, and let f be computable. Then $(A_n f)$ converges at every Martin-Löf random.

Hoyrup and Rojas have generalized this to effectively presented measure spaces.

Algorithmic randomness

When the space is ergodic, the conclusion also holds when f is the characteristic function of a recursively open set.

This is due to Franklin, Greenberg, Miller, and Ng, and, independently, Bienvenu, Day, Hoyrup, Mezhirov and Shen.

Algorithmic randomness

(G_n) is a *Schnorr test* if, moreover, $\mu(G_n)$ is uniformly comptable.

A point ω of a measure preserving system $(X, \mathcal{B}, \mu, T)$ is *typical* (or *generic*) if $(A_n f)x$ converges to $\int f$ for every continous f.

Theorem (Gács, Hoyrup, Rojas)

The following are equivalent:

- *x is typical whenever T and μ are computable and μ is ergodic.*
- *x is Schnorr random.*

Theorem (A)

For every shift-invariant measure μ on 2^ω, ergodic or not, there is a computable typical point.

Metastability

If (a_n) is a nondecreasing sequence of real numbers in $[0, 1]$, then

$$\forall \varepsilon > 0 \, \exists m \, \forall n, n' \geq m \; |a_n - a_{n'}| < \varepsilon$$

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

The statement above is equivalent to

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

But there is always an $m \leq F^{\lfloor 1/\varepsilon \rfloor}(0)$ satisfying the conclusion.

Replacing convergence by "metastable convergence" provides a computable bound that is, moreover, *uniform* in (a_n).

Metastability

Notes:

- This is an instance of Kreisel's "no-counterexample interpretation," which is, in turn, a special case of the Gödel's *Dialectica* interpretation.
- Kohlenbach has developed extensive "proof mining" methods based on these ideas.
- Instances of metastability play a role in Tao's work in ergodic theory and ergodic Ramsey theory, where the uniformities are important.

Metastability

Returning to the mean ergodic theorem, the assertion that the sequence $(A_n f)$ converges can be represented as follows:

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

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

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

A constructive mean ergodic theorem

Theorem (A-Gerhardy-Towsner)

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 F be any function. Then there is an $m \geq 1$ such that for every n, n' in $[m, F(m)]$, $\|A_n f - A_{n'} f\| < \varepsilon$.

In fact, we provide a bound on m expressed solely in terms of F and $\rho = \|f\|/\varepsilon$. Notably, the bound is independent of X and $\hat{T}$.

As special cases, we have the following:

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

A constructive pointwise ergodic theorem

The following is classically equivalent to the pointwise ergodic theorem:

Theorem (A-Gerhardy-Towsner)

For every f in $L^2(\mathcal{X})$, $\lambda_1 > 0$, $\lambda_2 > 0$, and F there is an $m \geq 1$ satisfying

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

We provide explicit bounds on m in terms of f, λ_1, λ_2, and F.

Qualitatively different bounds can be obtained using Bishop's upcrossing inequality.

Overview

We will consider

- the ergodic theorems
- measure theoretic convergence theorems

with respect to

- computability and constructivity
- reverse mathematics
- algorithmic randomness
- metastability

Convergence theorems

Recall that a sequence (f_n) of measurable functions

- converges *pointwise a.e.* if for almost every x, $(f_n(x))$ converges.
- converges in *mean* if it converges in the L^1 norm,

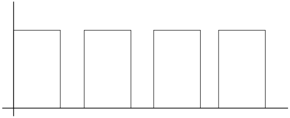
$$\|f\| = \int |f| \, d\mu.$$

Convergence in mean does not imply pointwise convergence.

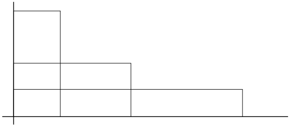
Convergence theorems

Nor does pointwise convergence imply convergence in mean.

Functions can go off to infinity along the horizontal axis:



or along the vertical axis:



The dominated convergence theorem

In subsystems of second-order arithmetic, an L^1 functional is represented by a Cauchy sequence of simple functions.

One formulation of the dominated convergence theorem (DCT):

> *For every (f_n) and g in L^1, if (f_n) is pointwise Cauchy a.e. and dominated by g, then there is an $f \in L^1$ such that f_n converges to f pointwise a.e. and $(\int f_n)$ converges to $\int f$.*

Yu showed that this is equivalent to (ACA).

The dominated convergence theorem

Another formulation, (DCT'):

> *For every (f_n), g, and f in L^1, if (f_n) converges pointwise a.e. to f and is dominated by g, then $(\int f_n)$ converges to $\int f$.*

Simpson conjectured this is equivalent to $(WWKL)$.

Yet another formulation (DCT''):

> *For every (f_n) and g in L^1, if (f_n) is pointwise Cauchy a.e. and (f_n) and is dominated by g, then $(\int f_n)$ is Cauchy.*

Reverse mathematics

Simpson and Yu defined *Weak Weak König's Lemma*:

$$\forall T \text{ (if } T \text{ is an infinite binary tree and}$$
$$\lim_{n\to\infty} \frac{|\sigma \in T, len(\sigma) = n|}{2^n} > 0, \text{there is a path through } T).$$

It is equivalent to each of the following:

- Every closed set with positive measure has a point.
- For every X, there is a ML random relative to X (essentially Kucera).

Reverse mathematics

Simpson and Yu showed that ($WWKL$) is strictly weaker than (WKL) and not provable in (RCA_0).

$$
\begin{array}{c}
(ACA) \\
\downarrow \\
(WKL) \\
\downarrow \\
(WWKL) \\
\downarrow \\
RCA_0
\end{array}
$$

Simpson conjectured that (DCT') is equivalent to ($WWKL$).

Reverse mathematics

Ed Dean, Jason Rute, and I defined (*2-WWKL*) to be the relativization of (*WWKL*) to trees definable from the Turing jump of any set.

Theorem

Over RCA_0, the following are equivalent:

- (*2-WWKL*)
- *the statement that every G_δ set with positive measure has a point*
- ($B\Sigma_2$) *plus the assertion that 2-randoms exist relative to any* X.

Reverse mathematics

$$
\begin{tikzcd}
 & (\mathit{ACA}) \arrow[dl] \arrow[dddr] & \\
\vdots & & \\
(\mathit{3\text{-}WWKL}) \arrow[d] & & \\
(\mathit{2\text{-}WWKL}) \arrow[dr] & & (\mathit{WKL}) \arrow[dl] \\
 & (\mathit{WWKL}) \arrow[d] & \\
 & \mathit{RCA_0} &
\end{tikzcd}
$$

Reverse mathematics

Theorem (A-Dean-Rute)

Over RCA_0, the following are equivalent:

- (*2-WWKL*)
- (DCT')
- (DCT'')
- *Versions of Egorov's theorem for test functions.*

Idea: given any G_δ set with positive measure, construct a sequence of functions that converge to 0 off that set.

Conversely: given a sequence (f_n) such that $(\int f_n)$ doesn't converge to 0, construct a G_δ set where it fails to converge.

Metastability

In "Norm convergence of multiple ergodic averages for commuting transformations," Tao provided a quantitative version of the dominated convergence theorem.

Theorem (Tao)

For every $M(F, \varepsilon)$, there is an $M'(F', \varepsilon')$ satisfying the following. Given a probability space $\mathcal{X} = (X, \mathcal{B}, \mu)$ and sequence $(f_n) : X \to [0,1]$, if

$$\forall \varepsilon > 0, F, x \; \exists m \leq M(F, \varepsilon) \; \forall n \in [m, F(m)] \; f_n(x) < \varepsilon,$$

then

$$\forall \varepsilon' > 0, F' \; \exists m \leq M'(F', \varepsilon') \; \forall n \in [m, F(m)] \; \int f_n < \varepsilon'.$$

Metastability

Tao's proof is nonconstructive. M' can be computed in principle, say, by blind search.

> *In practice, though, it seems remarkably hard to do; the proof of the Lebesgue dominated convergence theorem, if inspected carefully, relies implicitly on the infinite pigeonhole principle, which is notoriously hard to finitize. (Tao '06)*

Tao took results from reverse mathematics to suggest that the dependence of M' on the parameters is likely to be "fantastically poor."

Metastability

Ed Dean, Jason Rute, and I:

- Gave a constructive proof of a key combinatorial lemma.
- Used that to derive a quantitative version of Egorov's theorem.
- Used that to derive a (mild) strengthening and "Cauchy version" of Tao's theorem.

Theorem (A, Dean, Rute)

If M is in the calculus $\mathsf{G}_\infty\mathsf{A}^\omega$, then M' is a primitive recursive functional (in the sense of Kleene). If M at at level n of Gödel's T, then M' is at level $n + 1$.

Conclusions

Dynamical systems, ergodic theory, measure theory, and so on are often nonconstructive.

But that does not mean that they do not have computational content.

It only means that sometimes you have to look a little harder to find it.