

Computability and convergence

Jeremy Avigad

Department of Philosophy and
Department of Mathematical Sciences
Carnegie Mellon University

July, 2012

Computability and convergence

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 in analysis challenged this view.

A sequence (a_n) in a metric space is said *Cauchy* if for every $\varepsilon > 0$, there is an m such that for every $n, n' \geq m$, $d(a_n, a_{n'}) < \varepsilon$.

If the space is *complete*, such a sequence always has a limit.

The problem: “arbitrary” convergent sequences need not have computable limits.

Computable 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.

Theorem (Specker)

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

In general, one can always compute a name for the limit from the halting problem.

Conversely, there is a sequence (a_n) such that the halting problem is computable from any such name.

Computable analysis

The Bolzano-Weierstrass theorem (proved by Bolzano in 1817) fares even worse.

Theorem (Folklore?)

There is a computable sequence of rationals in $[0, 1]$ with no computable limit point.

In general, one can always find a limit low relative to $0'$.

Conversely, there is a sequence of rationals such that any computable limit point is a PA degree relative to $0'$.

(See Kreuzer, “The cohesive principle and the Bolzano-Weierstrass principle.”)

Computable analysis

A function from f from \mathbb{R} to \mathbb{R} is *computable* if there is a computable procedure taking any name for x to a name for $f(x)$.

Note: the procedure must work on arbitrary names, not just the computable ones. This is “Type 2” or “Polish style” computability.

Computable functions are necessarily continuous.

Computable analysis

These notions transfer to complete separable metric spaces, and mathematical structures that can be coded as such:

- Spaces of functions
- Hilbert spaces
- Banach spaces
- Measure spaces (measure algebras)
- Spaces of operators, measures, etc.

In modern terms, the nineteenth century tension is this: many existence theorems in analysis are not computably valid.

Grappling with the tension

It appears . . . that there are certain mathematical statements that are merely evocative, which make assertions without empirical validity. There are also mathematical statements of immediate empirical validity, which say that certain performable operations will produce certain observable results. . . . Mathematics is a mixture of the real and the ideal, sometimes one, sometimes the other, often so presented that it is hard to tell which is which. The realistic component of mathematics—the desire for pragmatic interpretation—supplies the control which determines the course of development and keeps mathematics from lapsing into meaningless formalism. The idealistic component permits simplifications and opens possibilities which would otherwise be closed. The methods of proof and objects of investigation have been idealized to form a game, but the actual conduct of the game is ultimately motivated by pragmatic considerations. (Errett Bishop, 1967)

Outline

Topics:

- Background and motivation
- Quantitative information in convergence statements
 - Rates of convergence
 - Oscillation inequalities
 - Metastability
- Case study: the mean ergodic theorem
- Other topics

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.

Rates of convergence

Suppose (a_n) is Cauchy:

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

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*.

If there is a computable bound on the rate of convergence of (a_n) , then (a_n) has a computable limit.

Rates of convergence

The converse does not always hold. For example, there are computable sequences (a_n) that converge to 0, but without a computable bound on the rate of convergence.

(The idea: when the n th Turing machine halts, output $1/n$.)

The Specker example shows that a computable, monotone, bounded sequence of rationals need not have a computable rate of convergence.

Oscillations

Definition

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

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

A bound on the rate of convergence is, a fortiori, a bound on the number of fluctuations.

On the other hand, a nondecreasing sequence in $[0, 1]$ clearly has at most $\lceil 1/\varepsilon \rceil$ many ε -fluctuations.

So, for the Specker sequence, there is a computable bound on the number of fluctuations, but no computable bound on the rate of convergence.

It is not hard to cook up a computable sequence that converges to 0, but with no computable bound on the number of fluctuations.

(Idea: when Turing machine n halts, oscillate by $1/n$ lots of times.)

Uniformity

We just observed that a nondecreasing sequence in $[0, 1]$ has at most $\lceil 1/\varepsilon \rceil$ many ε -fluctuations.

This bound is entirely independent of the sequence (a_n) .

So not only do we get a *computable* version of the monotone convergence theorem, but also a highly *uniform* one.

Generally, theorems depend on parameters (a space, a sequence, a transformation, ...)

Sometimes, bounds are independent of some of these: instead of $\forall p \forall \varepsilon > 0 \exists n \dots$ one has $\forall \varepsilon > 0 \exists n \forall p \dots$

Such uniformities are mathematically useful.

Upcrossings

Oscillations are closely related to upcrossings.

Definition

Given $\alpha < \beta$, say that a sequence (a_n) of real numbers has m *upcrossings* from α to β if there are $i_1 \leq j_1 \leq \dots \leq i_m \leq j_m$ such that, for each $u = 1, \dots, m$, $a_{i_u} < \alpha$ and $a_{j_u} > \beta$.

If (a_n) is a bounded sequence, (a_n) is Cauchy if and only if for every $\alpha < \beta$, there are only finitely many upcrossings.

A bound $b(\alpha, \beta)$ on the number of upcrossings can be computed from a bound $k(\varepsilon)$ on the number of fluctuations, and vice-versa.

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

The translation is an instance of Kreisel's "no-counterexample interpretation," and provides any convergence statement with a computational meaning.

Moreover, there are often very uniform bounds.

Notice that if $k(\varepsilon)$ is a bound on the number of ε -fluctuations, then $M(F, \varepsilon) = F^{k(\varepsilon)}(0)$ is a bound on the rate of metastability, since one of the intervals

$$[0, F(0)], [F(0), F(F(0))], \dots, [F^{k(\varepsilon)}(0), F^{k(\varepsilon)+1}(0)]$$

must fail to contain an ε -fluctuation.

Metastability

The no-counterexample interpretation is, in turn, special case of the Gödel's *Dialectica* interpretation.

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

In particular, he has shown that strong uniformities hold in very general situations.

He and his students have also extracted particular bounds from many theorems in functional analysis.

Metastability has played a role in work by Terence Tao in ergodic theory and additive combinatorics, including his proof with Ben Green that there are arbitrarily long arithmetic progressions in the primes.

Summary

Given that a sequence converges, we can ask for:

- A bound on the rate of convergence.
- A bound on the number of fluctuations.
- A bound on the rate of metastability.

These are successively weaker.

The last is always computable from the sequence itself.

Beyond computability, we may be interested in quantitative data, and/or uniformities.

Convergence questions

Given a convergence theorem, ask:

- Is there a computable bound on the rate of convergence?
 - If so: give quantitative bounds.
 - If not: determine complexity, missing information.
- Is the rate of convergence uniform in any of the parameters?
- Is there a computable bound on the number of fluctuations?
- Are there uniform bounds on the number of fluctuations?
- Give quantitative bound on the rate of metastability.
- Is the rate of metastability uniform in any of the parameters?

The role of logic

Computable analysis is needed to frame the general question as to computability.

- Analysis: particular rates of convergence and particular uniformities
- Logic: general characterizations of what information can be had

Proof theory and proof mining provide general methods for extracting additional information from proofs.

- Analysis: seek rates and uniformities in particular cases
- Logic: provide general methods for finding them

Methods from model theory and nonstandard analysis should also be useful.

Outline

Topics:

- Background and motivation
- Quantitative information in convergence statements
 - Rates of convergence
 - Oscillation inequalities
 - Metastability
- Case study: the mean ergodic theorem
- Other topics

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} (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 $\mu(A) = 1$.

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

The pointwise ergodic theorem

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

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

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.

It is easy to see that 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} .)

Computability

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 initial data $(T$ and $f)$?

In other words: can we compute a function $r : \mathbb{Q} \rightarrow \mathbb{N}$ 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 there is no computable bound on the rate of convergence of $A_n 1_{[1]}$.

Theorem (Avigad)

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

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 .

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

$$1, -1, 1, -1, \dots$$

and the averages are

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

and the total variation diverges.

Square functions

Theorem (Jones, Ostrovskii, and Rosenblatt)

Let T be any nonexpansive operator on a Hilbert space, and f and element. Then for any sequence $n_1 \leq n_2 \leq \dots$,

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

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

The proof uses the spectral theorem to reduce to the simple case where T is just a rotation. (But additional cleverness is needed even in that special case.)

Uniformly convex spaces

Definition

A Banach space \mathcal{B} is *uniformly convex* if for every $\varepsilon \in (0, 2]$ there exists a $\delta \in (0, 1]$ such that for all $x, y \in \mathcal{B}$, 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 $L_\infty(\mathcal{X})$.

Any function $\eta(\varepsilon)$ returning such a δ for every $\varepsilon \in (0, 2]$ is called a *modulus of uniform convexity*.

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

Uniformly convex spaces

Building on work by Kohlenbach and Leuştean, Jason Rute and I showed:

Theorem

Let $f \in \mathcal{B}$, $\varepsilon > 0$, T nonexpansive. Write $\rho = \|f\|/\varepsilon$. Then $(A_n f)$ admits at most $O(\rho^2 \log \rho \cdot \eta(1/(8\rho))^{-1})$ -many ε -fluctuations.

If $\eta(\varepsilon) = \varepsilon \cdot \tilde{\eta}(\varepsilon)$ with $\tilde{\eta}$ nondecreasing, the conclusion holds with η replaced by $\tilde{\eta}$.

Leuştean extended the result to *power bounded* operators, i.e. assuming $\|T^n f\| \leq C\|f\|$ for every n .

The result is not sharp for Hilbert spaces: we get $O(\rho^3 \log \rho)$ instead of $O(\rho^2)$.

The three of us are working to extend this result.

Metastability

Note that the bound on the number of ε -fluctuations depends only on $\|f\|/\varepsilon$ and η , and not at all on \mathcal{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 an explicit bound on m in terms of F and $\|f\|/\varepsilon$.

Metastability

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 Leuştean extended this to uniformly convex Banach spaces.

In 2007, Tao used metastability to prove a generalization of the mean ergodic theorem to certain “multiple” averages.

Metastability

There are two directions in which one can extend the mean ergodic theorem:

- More general spaces (e.g. reflexive spaces).
- More general averaging schemes.

For example, given a sequence of elements $\alpha_n \in [0, 1]$, Halpern considered the iteration:

$$u_{n+1} = \alpha_{n+1}u_0 + (1 - \alpha_{n+1})Tu_n.$$

For $\alpha_n = 1/(n + 1)$, these are the ergodic averages. With conditions on the α_n , the space, and the operator, these iterates converge too.

Metastability

Kohlenbach analyzed a theorem of Wittman, and obtained a primitive recursive functional bound on the rate of metastability for Halpern iterations on a Hilbert space.

Kohlenbach and Leuştean analyzed a theorem of Saejung, and obtained a much more complex bound on the rate of metastability for Halpern iterations on $CAT(0)$ spaces.

Outline

Topics:

- Background and motivation
- Quantitative information in convergence statements
 - Rates of convergence
 - Oscillation inequalities
 - Metastability
- Case study: the mean ergodic theorem
- Other topics

Measure-theoretic convergence

In measure theory, one can also consider:

- pointwise convergence
- convergence in measure
- convergence in the various L^p norms

For example, the *dominated convergence theorem* says that if a sequence f_n is dominated by an integrable function g and converges pointwise a.e., then it converges in the L^1 norm.

Tao gave a metastable version of the dominated convergence theorem.

Dean, Rute, and I gave a more explicit bound.

Bishop's upcrossing inequalities

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

The pointwise ergodic theorem is equivalent to saying that $\omega_{\alpha,\beta}(x)$ is finite a.e. for every rational $\alpha < \beta$.

Theorem (Bishop)

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

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

Related work

Bishop's upcrossing inequalities were modeled on Doob's analogous result for martingales. (Cf. also Lepingle's theorem.)

There are variations on Bishop's result due to Ivanov, Kachurovskii, Kalikow and Weiss, Hochman, Bourgain.

See especially "Oscillation in ergodic theory" by Jones, Rosenblatt, and Weirld.

There is also a literature on upcrossing inequalities, oscillations, variational inequalities, with respect to martingale convergence theorems, the Lebesgue differentiation theorem, and many other settings.

See papers by Jones and various co-authors, and "A variation norm Carleson theorem" by Oberlin, Seeger, Tao, Thiele, and Wright.

Related work

Kohlenbach and his students have done extensive work on proof mining theorems in fixed-point theory and approximation theory.

Students include Paulo Oliva, Philipp Gerhardy, Branimir Lambov, Eyvind Briseid, Jaime Gaspar, Alexander Kreuzer, and Pavol Safarik.

Related work

There has, of late, been a lot of work in computable aspects of probability and measure theory.

See work by Ackermann, Freer, Gács, Galatolo, Hoyrup, Rojas, Roy, Simpson, V'yugin, Xu.

Algorithmic randomness makes it possible to characterize the counterexamples to a.e. theorems of mathematics.

The ergodic theorems, the Lebesgue differentiation theorem, Lebesgue's theorem, martingale convergence theorems, and more have been studied by this perspective.

See work by Bienvenu, Brattka, Day, Franklin, Freer, Gács, Greenberg, Hoyrup, Kjos-Hanssen, Miller, Miyabe, Nies, Ng, Pathak, Rojas, Rute, Shen, Simpson, Stephan, Towsner, V'yugin.

General questions

When should we expect to have computable bounds on rates of convergence?

When should we expect to have computable bounds on the number of fluctuations?

Can general logical methods help find such bounds, and uniformities?

How and where can the additional information be put to good work?

- Ergodic theory and dynamical systems
- Probability and statistics
- Applications to combinatorics and number theory