Proof Theory and Proof Mining III: The Dialectica Interpretation and Applications

Jeremy Avigad

Department of Philosophy and
Department of Mathematical Sciences
Carnegie Mellon University

July 2015

Sequence of Topics

1. Computable Analysis
2. Formal Theories of Analysis
3. The Dialectica Interpretation and Applications
4. Ultraproducts and Nonstandard Analysis

Conservation results

Recall that many central proof theoretic results have the following form:

For any $\varphi \in \Gamma$, if $T_1 \vdash \varphi$, then $T_2 \vdash \varphi'$.

They are used for:

- foundational reduction
- "proof mining"

There are model-theoretic and syntactic approaches.

Cut elimination and normalization can lead to dramatic increases in proof length.

Conservation results

Interpretations:

- Double-negation translations
- The Friedman-Dragalin interpretation
- Semantic interpretations
- Forcing translations
- Realizability interpretations
- Functional interpretations

The last two work best on intuitionistic theories.

Realizability

Realizablility: assign to every formula φ in the source language a formula "e realizes φ," where e contains additional information that "witnesses" the truth of φ.

Remember that if classical arithmetic, PA, proves $\forall x \; \exists y \; R(x, y)$, then intuitionistic arithmetic, HA, proves $\forall x \; \neg \forall y \; \neg R(x, y)$.

Typically, a "realization" of this last formula provides no useful information.

In PA, and other in situations as well, the Friedman-Dragalin translation solves this problem.

The Dialectica interpretation

We will consider *functional interpretation*, which also solves the problem, and has other benefits as well.

In 1958 Gödel published an interpretation of *PA* in the Swiss journal, *Dialectica*. In fact, he had the intepretation already in the early 1930's.

Most directly, it interprets *HA* in PRA^{ω}, thereby "reducing" arithmetic to a quantifier-free theory in the higher-types.

Via the double-negation interpretation, this extends to *PA*.

Paulo Oliva has shown that there is a spectrum of interpretations between the Dialectica interpretation and realizability.

First-order arithmetic

Recall:

Primitive Recursive Arithmetic (PRA):

- defining equations for the primitive recursive functions
- quantifier-free induction

Peano Arithmetic (PA) = PRA + induction, over first-order classical logic

Heyting arithmetic (HA) = PRA + induction, over first-order intuitionistic logic

The double-negation interpretation interprets *PA* in *HA*.

Higher-type arithmetic

Recall:

- $PA^\omega = PRA^\omega +$ induction, over classical logic
- $HA^\omega = PRA^\omega +$ induction, over intuitionistic logic

These are conservative extensions of PA and HA respectively.

The Dialectica interpretation

Here is one consequence of the interpretation:

Theorem. Suppose PA proves $\forall x \, \exists y \; R(x, y)$, where R is primitive recursive. Then there is a term f such that PRA^{ω} proves $R(x, f(x))$.

A computable function $f : \mathbb{N} \to \mathbb{N}$ is said to be a *provably total in* T if it is computed by a Turing machine e such that $T \vdash \forall x \, \exists s \; T(e, x, s)$.

This result shows that the provably total computable functions of arithmetic are exactly the type 1 primitive recursive functionals.

The Dialectica interpretation

In fact, the interpretation can be described most directly as an interpretation of HA^ω in PRA^ω.

It handles additional axioms:

- Markov's principle (MP): $\neg\neg\exists x\, A(x) \to \exists x\, A(x)$
- The axiom of choice (AC):

$$\forall x\, \exists y\, \varphi(x, y) \to \exists f\, \forall x\, \varphi(x, f(x))$$

- Independence of premise ($IP_\forall$):

$$(\varphi \to \exists x\, \psi(x)) \to \exists x\, (\varphi \to \psi(x))$$

  where φ is a $\forall$ formula.

The Dialectica interpretation

Assign to every formula φ in the language of PRA^ω a formula

$$\varphi^D \equiv \exists x\, \forall y\, \varphi_D(x, y)$$

where x and y are sequences of variables and φ_D is quantifier-free.

Idea: $\forall y\, \varphi_D(x, y)$ asserts that x is a "strong" realizer for φ.

Inductively one shows:

Theorem (Gödel). If HA^ω proves φ, there is a sequence of terms t such that quantifier-free PRA^ω proves $\varphi_D(t, y)$.

The Dialectica interpretation

Define the translation inductively. Assuming:

- $\varphi^D = \exists x \, \forall y \;\; \varphi_D(x, y)$
- $\psi^D = \exists u \, \forall v \; \psi_D(u, v)$

Define:

- For θ an atomic formula, $\theta^D = \theta_D = \theta$.
- $(\varphi \wedge \psi)^D = \exists x, u \, \forall y, v \; (\varphi_D \wedge \psi_D)$.
- $(\varphi \vee \psi)^D = \exists z, x, u \, \forall y, v \; ((z = 0 \wedge \varphi_D) \vee (z = 1 \wedge \psi_D))$.
- $(\forall z \; \varphi(z))^D = \exists X \, \forall z, y \; \varphi_D(X(z), y, z)$.
- $(\exists z \; \varphi(z))^D = \exists z, x \, \forall y \; \varphi_D(x, y, z)$.

The Dialectica interpretation

The most interesting clause is the one for implication. Assuming:

- $\varphi^D = \exists x\, \forall y\ \varphi_D(x, y)$
- $\psi^D = \exists u\, \forall v\, \psi_D(u, v)$

Define:

- $(\varphi \to \psi)^D = \exists U, Y\, \forall x, v\, (\varphi_D(x, Y(x, v)) \to \psi_D(U(x), v))$.

The last clause is a Skolemization of the formula

$$\forall x\, \exists u\, \forall v\, \exists y\, (\varphi_D(x, y) \to \psi_D(u, v)).$$

Note that $\neg\varphi$ is defined as $\varphi \to \bot$, so

- $(\neg\varphi)^D = \exists Y\, \forall x\, \neg\varphi_D(x, Y(x))$

Interpreting modus ponens

Consider the rule "from φ and $\varphi \to \psi$ conclude ψ."

We are given terms a, b, and c such that PRA^ω proves

$$\varphi_D(a, y)$$

and

$$\varphi_D(x, b(x, v)) \to \psi_D(c(x), v).$$

We need a term d such that PRA^ω proves

$$\psi_D(d, v).$$

Substituting $b(a, v)$ for y in the first hypothesis and a for x in the second, we see that taking $d = c(a)$ works.

Interpreting induction

Induction can be expressed as a rule: "from $\varphi(0)$ and $\varphi(u) \to \varphi(S(u))$ conclude $\varphi(u)$."

Inductively we have a, b, and c and proofs of

- $\varphi_D(0, a, y)$
- $\varphi_D(u, x, b(u, x, y)) \to \varphi_D(S(u), c(u, x), y)$.

We want a term d such that $\varphi_D(u, d(u), y)$.

Define d using primitive recursion, so that

- $d(0) = a$
- $d(S(u)) = c(u, d(u))$.

Interpreting induction

This yields:

- $\varphi_D(0, d(0), y)$
- $\varphi_D(u, d(u), b(u, d(u), y)) \to \varphi_D(S(u), d(S(u)), y)$.

A little more work, we can then prove $\varphi_D(u, d(u), y)$, as desired.

Advantages of the Dialectica interpretation

For example, the formula

$$\forall x\ A(x) \to \forall x\ B(x)$$

translates to

$$\exists f\ \forall x\ (A(f(x)) \to B(x))$$

Markov's principle is verified:

$$\neg\neg\exists x\ A(x) \to \exists x\ A(x)$$

So is the axiom of choice:

$$\forall x\ \exists y\ \varphi(x,y) \to \exists f\ \forall x\ \varphi(x, f(x)).$$

The independence of premise principle ($IP_\forall$) is also interpreted.

Applying the Dialectica interpretation

Recipe:

- Start with a nonconstructive proof.
- Formalize it in $PA^\omega + (QF{-}AC)$.
- Apply a double-negation translation.
- Get a proof in $HA^\omega + (MP) + (IP) + (AC)$.
- Apply the Dialectica interpretation

We will see that certain nonconstructive principles, like weak König's lemma, can also be eliminated.

We will also consider a modification of the D-interpretation, due to Kohlenbach, that makes it easier to extract bounds instead of witnesses.

The no-counterexample translation

Consider, for example, what the ND-translation does to prenex arithmetic formulas, such as

$$\forall x \; \exists y \; \forall z \; A(x, y, z).$$

Negate, Skolemize, and negate again:

$$\forall x, Z \; \exists y \; A(x, y, Z(y))$$

Such a y foils the putative counterexample function, Z.

The Skolemization of this formula is Kreisel's *no-counterexample* interpretation:

$$\exists Y \; \forall x, Z \; A(x, Y(x, Z), z(Y(x, Z))),$$

This works for any number of quantifiers.

The Dialectica interpretation

The D-interpretation works well on restricted theories.

Theorem. If $\widehat{PRA}_i^\omega + (AC) + (MP)$ proves φ, then $\widehat{PRA}_i^\omega \vdash \varphi^D$.

Corollary. If $\widehat{PRA}^\omega + (QF{-}AC)$ proves $\forall x\, \exists y\, \varphi(x, y)$, with φ q.f. in the language of PRA, then so does $\widehat{PRA}^\omega$.

$(QF{-}AC)$ can be used to prove Σ_1 induction:

$$\exists y\, \varphi(0, y) \wedge \forall x\, (\exists y\, \varphi(x, y) \rightarrow \exists y\, \varphi(x + 1, y)) \rightarrow \forall x\, \exists y\, \varphi(x, y).$$

So this shows that $I\Sigma_1$ is Π_2 conservative over PRA.

Majorizability

Sometimes we only care about bounds.

Definition (Howard). Define $a \leq^*_\tau b$, read *a is hereditarily majorized by b*, by induction on the type of τ:

- $a \leq^*_{\mathsf{N}} b \equiv a \leq b$
- $a \leq^*_{\rho \to \sigma} b \equiv \forall x, y\ (x \leq^*_\rho y \to a(x) \leq^*_\sigma b(y))$

For example, g majorizes f at type $\mathsf{N} \to \mathsf{N}$ if for every x, $g(x)$ is greater than or equal to $f(0), f(1), \ldots, f(x)$.

Proposition. If $x \geq^* y$ and $y \geq z$ (pointwise) then $x \geq^* z$.

Proposition. Every term of PRA^ω has a majorant.

Weak König's lemma

Saying $f \in \{0,1\}^\omega$ is equivalent to $f \leq^* \lambda x.\, 1$.

So, if $\lambda f\ G(f, \vec{x})$ is majorized by $\lambda f\ H(f, \vec{x})$, then for each $f \in \{0,1\}^\omega$, $G(f, \vec{x}) \leq H(\lambda x.\, 1, \vec{x})$.

Theorem. $\widehat{PRA}^\omega + (QF{-}AC) + (WKL)$ is conservative over $\widehat{PRA}^\omega$ for Π_2 sentences.

Proof. Use the Dialectica interpretation. If the source theory proves $\forall x\, \exists y\, A(x, y)$, then $\widehat{PRA}^\omega$ proves the ND-translation of

$$(WKL) \to \forall x\, \exists y\, A(x, y).$$

Majorizability can be used to eliminate the dependence on the hypothesis.

The monotone interpretation

Observations:

- Often, one only cares about bounds, not witnesses.
- Some principles, like (WKL), don't affect bounds.

A "monotone" variant of the Dialectica interpretation, due to Kohlenbach, interprets every formula by one of the form

$$\exists x^* \, \exists x \leq^* x^* \, \forall y \, A(x, y)$$

A metatheorem

The theory $E\!-\!PA^\omega$ is PA^ω with an axiom of extensionality.

The theory $WE\!-\!HA^\omega$ is HA^ω is a weaker extensionality rule.

Let X be a complete separable metric space, K a compact space.

Let $\varphi(n^0, x^1, y^1, m^0)$ be an existential formula that is provably extensional in $x \in X$ and $y \in K$.

Chapter 15 of Kohlenbach, *Applied Proof Theory: Proof Interpretations and their Use in Mathematics* proves a general corollary of the Dialectica interpretation.

A metatheorem

Suppose $E{-}PA^{\omega} + (QF{-}AC^{1,0}) + (QF{-}AC^{0,1}) + (WKL)$ proves

$$\forall n \in \mathbb{N}\, \forall x \in X\, \forall y \in K\, \exists m \in \mathbb{N}\, \varphi(n, x, y, m).$$

Then there is a uniform bound $\Phi(n, x)$ in PRA^{ω}, such that $WE{-}HA^{\omega}$ proves

$$\forall n \in \mathbb{N}\, \forall x \in X\, \forall y \in K\, \exists m \leq \Phi(n, x)\, \varphi(n, x, y, m).$$

Note that the bound doesn't depend on the compact space.

A metatheorem

We can add axioms Γ of the form

$$\forall x \in X' \, \exists y \in K' \, \forall w \in W \, (F(x, y, w) =_{\mathbb{R}} 0)$$

where X' and W are CSM's, K' is compact.

These are then replaced by certain "ε-weakenings" in the target theory:

$$\forall x \in X' \, \forall k, n \in \mathbb{N} \, \exists y \in K' \, \forall i < n \, (|F(x, y, w_i)| < 2^{-k})$$

where $(w_i)_{i \in \mathbb{N}}$ is the countable dense subset of W.

A metatheorem

Allowable axioms include a wide range of principles from analysis:

- Basic properties of continuous functions, integrals, sups, trig functions, etc.
- The fundamental theorem of calculus.
- The Heine-Borel theorem.
- Uniform continuity of continous functions on a compact interval.
- The extreme value theorem.

These are simply eliminated.

A metatheorem

Principles based on *sequential compactness* are not included. But even restricted uses of these can be eliminated, in certain contexts.

There are variants of the metatheorem for

- stronger theories
- weaker theories
- theories in which spaces that are not assumed to be separable

In the last case, the spaces are modeled abstractly.

Applications

There have been a number of applications of proof mining techniques to fields of analysis.

For example:

- uniqueness proofs in approximation theory
- rates of asymptotic regularity and uniformities from proofs of convergence in fixed point theorems
- bounds on rates of metastability (and uniformities) in ergodic theorems.

I will discuss the third here.

Applications

References for the first two topics:

- Kohlenbach, *Applied Proof Theory: Proof Interpretations and their Use in Mathematics* (the locus classicus)
- Articles and surveys on Kohlenbach's web page

Additional exposition:

- Towsner, "A worked example of the functional interpretation"
- slides, "Proof mining," on my web page (from 2004)

Convergence theorems

Recall that many convergence theorems are computationally false:

- We have seen that given $f : \mathbb{N} \to [0, 1]$ nondecreasing, it is generally not possible to compute the limit.
- Similarly for the mean ergodic theorem.

The Dialectica interpretation provides something weaker, but computationally meaningful: *a rate of metastability*.

Let's pause to talk about quantitative data associated with convergence theorems.

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:

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

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*.

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 nth 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 $i_1 \leq j_1 \leq \ldots \leq i_m \leq j_m$ 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

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 $\forall p \, \forall \varepsilon > 0 \, \exists n \ \ldots$ one has $\forall \varepsilon > 0 \, \exists n \, \forall p \ \ldots$.

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 $i_1 \leq j_1 \leq \ldots \leq i_m \leq j_m$ such that, for each $u = 1, \ldots, m$, $a_{i_u} < \alpha$ 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

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

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

This statement 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

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

must fail to contain an ε-fluctuation.

Metastability

In general, the bound on the rate of metastability is computable from the data.

More importantly: it is often very *uniform*.

The Dialectica interpretation predicts / explains the uniformity, and allows us to extract explicit bounds.

Later, we will see that ultraproduct methods also allow us to predict the uniformity.

Ergodic theory recap

Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system, and let $f : X \to \mathbb{R}$ be a measurable function. For every $n \geq 1$, let

$$(A_n f)(x) = \frac{1}{n} \sum_{i<n} f(T^i x).$$

The pointwise ergodic theorem says that for f in L^1, $(A_n f)$ converges pointwise a.e.

The mean ergodic theorem says that for f in L^2, $(A_n f)$ converges in the L^2 norm.

In general, the rate of convergence cannot be computed from T and f. But it *can* be computed from T, f, and $\|f^*\|$.

A metastable ergodic theorem

The following is an equivalent statement of the ergodic theorem:

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

The Dialectica interpretation enables us to extract a bound on n that is expressed solely in terms of F and $\rho = \|f\|/\varepsilon$ (and independent of $\hat{T}$).

We will see later that this uniformity can also be obtained using a compactness argument.

History

- Variations on the no-counterexample interpretation played an implicit role in the Green-Tao theorem on arithmetic progressions in the primes, and in a quantitative proof of Szemerédi's theorem by Tao.
- Avigad, Gerhardy, Towsner (2010) gave a metastable ergodic theorem
- Tao (2008) used metastability in "Norm convergence of multiple ergodic averages," and coined the term.
- Kohlenbach and Leuştean (2009): extended AGT result to uniformly convex Banach spaces
- Walsh (2012): used metastability to generalize Tao's result to nilpotent group actions

History

Kohlenbach, Leuștean, and others have obtained vast generalizations of these results, involving:

- more general forms of averaging and iteration
- more general spaces, such as $CAT(0)$ spaces

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

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

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.

Oscillations and metastability

By the very nature of the statement, if a convergence theorem is true, the metastable version holds computationally.

What makes metastability useful is that the bounds are usually very *uniform* in the data.

This is a very general phenomena:

- It is explained by Kohlenbach's general metatheorems.
- We will see that it is also explained by an ultraproduct argument.

Oscillations and metastability

Recall also that a bound on the number of ε-fluctuations is a stronger piece of data.

Sometimes such bounds are available, but this seems to be a less general phenomenon, and such results are harder to obtain.

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 $(25\|x\|/\varepsilon)^2$.

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 $\|(x+y)/2\| \leq 1 - \delta$.

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).

Summmary

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.

Summary

Bounds on the rate of metastability are generally very uniform, and can be obtained using proof mining methods.

Bounds on the number of fluctuations are harder to obtain.

Kohlenbach and Safarik, "Fluctuations, effective learnability and metastability in analysis," provides an analysis of the relationship between the two.