Fluctuations, effective learnability and metastability in analysis

Ulrich Kohlenbach, Pavol Safarik

Department of Mathematics, Technische Universität Darmstadt, Schlossgartenstraße 7, 64289 Darmstadt, Germany

Dedicated to Professor Sergei Artemov on the occasion of his 60th birthday

Abstract

This paper discusses what kind of quantitative information one can extract under which circumstances from proofs of convergence statements in analysis. We show that from proofs using only a limited amount of the law-of-excluded-middle, one can extract functionals (B, L), where L is a learning procedure for a rate of convergence which succeeds after at most $B(a)$-many mind changes. This (B, L)-learnability provides quantitative information strictly in between a full rate of convergence (obtainable in general only from semi-constructive proofs) and a rate of metastability in the sense of Tao (extractable also from classical proofs). In fact, it corresponds to rates of metastability of a particular simple form. Moreover, if a certain gap condition is satisfied, then B and L yield a bound on the number of possible fluctuations. We explain recent applications of proof mining to ergodic theory in terms of these results.

Keywords: Fluctuations, effective learnability, metastability, proof mining, uniform bounds, functionals of finite type, nonlinear ergodic theory, hard analysis.
2010 MSC: 03F10, 03F60, 47H25, 37A30.

1. Introduction

In this paper we investigate different levels of effective quantitative information on theorems stating the Cauchy property of some sequence (x_n) in a metric space (X, d)

$$(1)\ \forall k \in \mathbb{N}\, \exists n \in \mathbb{N}\, \forall m, \tilde{m} \geq n\ \left(d(x_m, x_{\tilde{m}}) \leq 2^{-k}\right)$$

and also more general Π^0_3-theorems (also with higher type parameters)

$$(2)\ \varphi \equiv \forall k \in \mathbb{N}\, \exists n \in \mathbb{N}\, \forall m \in \mathbb{N}\, \varphi_0(k, n, m),$$

where φ_0 is quantifier-free. Since we refer to real numbers as fast converging Cauchy sequences of rational numbers we have $\leq_{\mathbb{R}} \in \Pi^0_1$ so that (1) has the form (2).

Email address: kohlenbach@mathematik.tu-darmstadt.de, pavol.safarik@gmail.com (Ulrich Kohlenbach, Pavol Safarik)

Cauchy statements (1) are special forms of finiteness statements expressing that there are only finitely many 2^{-k}-fluctuations (i_l, j_l) with

$$(3)\ j_l > i_l \wedge d(x_{i_l}, x_{j_l}) > 2^{-k}.$$

As with general finiteness statements one can ask for a bound on the height of 2^{-k}-fluctuations, i.e. an $\rho(k)$ above which no such fluctuation occurs (so ρ is a rate of convergence) or for a weaker bound $F(k)$ on the number l of such fluctuations $(i_0, j_0), \ldots, (i_l, j_l)$ with $i_{n+1} \geq j_n$ for $n < l$. As to be expected from standard recursion theoretic facts about finiteness statements (see [35]), even primitive recursive Cauchy sequences (x_n) in $\mathbb{R}$ in general will not admit a computable (in k) bound F on the fluctuations and even in cases where they do, in general there will be no computable rate of convergence ρ.

A yet weaker information than a bound F on the number of fluctuations is a bound on the Kreisel no-counterexample interpretation (called 'metastability' by Tao) of (x_n), namely a functional $\Phi(k, g)$ such that

$$(4)\ \forall k \in \mathbb{N}\, \forall g : \mathbb{N} \to \mathbb{N}\, \exists n \leq \Phi(k, g)\, \forall i, j \in [n; n + g(n)]\ \big(d(x_i, x_j) \leq 2^{-k}\big).$$

We call Φ a *rate of metastability* for (x_n).
Note that already the underlying reformulation

$$(5)\ \forall k \in \mathbb{N}\, \forall g : \mathbb{N} \to \mathbb{N}\, \exists n\, \forall i, j \in [n; n + g(n)]\ \big(d(x_i, x_j) \leq 2^{-k}\big)$$

of the Cauchy property still expresses the full Cauchy property of (x_n). However, the proof of the latter from the former is noneffective, corresponding to the fact that there is no way to pass (even pointwise let alone uniformly) from an effective Φ in (4) to an effective bound on fluctuations F or an effective rate of convergence ρ.

Using unbounded search (over the code of the pair of $\exists n$ and the existential quantifier hidden in $<_{\mathbb{R}}$) one can always obtain a rate of metastability that is computable **relative** to $(d(x_i, x_j)_{i,j \in \mathbb{N}}$, but unless (x_n) is a sequence in a metric space X with a computability structure (e.g. $X := \mathbb{R}^n$ with the Euclidean norm as a computable metric), it makes no sense to talk about the computability of (x_n). Moreover, such an unbounded search does not provide any complexity information and the bound will be highly nonuniform (being dependent on all the data used to define (x_n)). In all the applications to which we refer to below, X is an abstract (completely general) Hilbert space or CAT(0) space and so the computability of (x_n) is not even defined. So to be able to talk about a **computable in the data used to define (x_n)** rate of metastability Φ this rate must only depend on general bounding data in $\mathbb{N}$ or $\mathbb{N}^{\mathbb{N}}$, i.e. Φ must be highly uniform. While this uniformity sometimes can be established by going to ultraproducts of X (see [8]) this not even seems to yield the existence of a **computable** Φ let alone of some complexity information.

General logical metatheorems for strong systems of analysis based on full classical logic guarantee that the extractability of (sub-)recursive (and highly uniform) rates of metastability Φ is always possible for large classes of convergence proofs. This has been applied extensively in the context of nonlinear analysis, fixed point theory and ergodic theory during the last 10 years. One of these results is the extraction of a uniform rate of metastability for the strong convergence in the mean ergodic theorem for uniformly convex Banach spaces X from a proof due to Birkhoff [13] carried out in [29]. This rate only

depends on a norm bound $\mathbb{N} \ni b \geq \|x\|$ of the starting point x, a modulus $\eta \in \mathbb{N}^{\mathbb{N}}$ of uniform convexity of X and the error 2^{-k} but, otherwise, is independent of x, the operator and X. That a computable rate of convergence (even for an effective Hilbert space and a computable operator) in general is impossible has been shown in [7]. However, as recently observed by Avigad and Rute [9], the analysis in [29] can be used to obtain a simple effective (and also highly uniform) bound on the number of fluctuations (for the case of Hilbert spaces this was already obtained with an even better bound in [22]). This raises the question whether there are general logical conditions on convergence proofs to guarantee the extractability of effective bounds on fluctuations. Obviously, any condition guaranteeing the extractability of a computable rate of convergence is a sufficient condition for this. Though not satisfied in the particular case just discussed, let us first consider this in order to see in what sense we might try to liberalize such conditions towards rates of fluctuations. To do so in somewhat more precise terms we fix a formal framework such as intuitionistic arithmetic HA^ω in all finite types (actually we use the so-called weakly extensional variant called WE-HA^ω in [26]) or its extension by an abstract (metric or) normed space $(X, \|\cdot\|)$ resulting in $\mathrm{HA}^\omega[X, \|\cdot\|]$ possibly with further axioms stating that X is uniformly convex or even a Hilbert space (see [26] for details). Let AC be the full schema of choice and $\mathrm{LEM}_\neg$ be the law-of-excluded-middle schema restricted to arbitrary negated formulas $\neg\psi$ (which, in particular, includes the case of existential-free formulas and so, as a very special case, Π^0_1-LEM, i.e. LEM restricted to Π^0_1-formulas). Then from a proof of (1) (for some sequence (x_n) definable by a term t of the system having at most number and function parameters a, f) in

$$\mathrm{HA}^\omega + \mathrm{AC} + \mathrm{LEM}_\neg$$

(and in fact even stronger theories augmented with certain noneffective axioms Ω), the extractability of a rate of convergence ρ that is definable (in the same parameters as t) in Gödel's calculus of primitive recursive functionals of finite type is guaranteed.
This follows from the bound extraction theorem for monotone modified realizability from [24, 26] (and for theories with abstract spaces X in [16]). In the case of $\mathrm{HA}^\omega[X, \|\cdot\|]$ even parameters of types such as $X, \mathbb{N} \to X, X \to X$ are allowed in the definition of the sequence (x_n) in X where then ρ depends additionally on majorants for these parameters (which are natural numbers, in the case of the type X, and number-theoretic functions, in the case of the types $\mathbb{N} \to X$, $X \to X$.)

An important weak principle of classical logic not covered by this is the so-called Markov principle which, extended to all finite types, reads as follows

$$\mathrm{M}^\omega \ : \ \neg\neg\exists \underline{x}^{\underline{\sigma}} \, \varphi_0(\underline{x}) \to \exists \underline{x}^{\underline{\sigma}} \varphi_0(\underline{x}),$$

where φ_0 is a quantifier-free formula (with arbitrary further parameters) and $\underline{\sigma}$ an arbitrary tuple of types. However, M^ω becomes permissible once $\mathrm{LEM}_\neg$ is weakened to the so-called lesser-limited-omniscience-principle LLPO (which is the precise amount of classical logic needed to prove the binary ('weak') König's lemma WKL, which with AC intuitionistically implies König's lemma KL; see [26] for details). So instead of $\mathrm{HA}^\omega([X, \|\cdot\|]) + \mathrm{AC} + \mathrm{LEM}_\neg$ we may also have

$$\mathrm{HA}^\omega([X, \|\cdot\|]) + \mathrm{AC} + \mathrm{M}^\omega + \mathrm{LLPO},$$

where then the extraction of a rate of convergence uses the so-called monotone functional interpretation (see [26]).

The in a sense weakest principle covered by neither of these systems (but provable in their union!) is LEM restricted to Σ^0_1-formulas, which we denote by Σ^0_1-LEM:

$$\Sigma^0_1\text{-LEM}: \quad \exists n \in \mathbb{N}\, \varphi_0(n) \vee \forall m \in \mathbb{N}\, \neg\varphi_0(m),$$

where φ_0 is quantifier-free (but may contain parameters of arbitrary type).
While Σ^0_1-LEM in the presence of AC (even when restricted to numbers) creates highly noncomputable functions (in particular when function parameters are allowed to occur in Σ^0_1-LEM which then makes it possible to climb up the entire arithmetical hierarchy) it remains fairly weak over HA^ω. Nevertheless, $\mathrm{HA}+\Sigma^0_1$-LEM already allows one to prove the Cauchyness of the Specker sequence [39], a primitive recursive monotone decreasing sequence of rational numbers in $[0,1]$ which does not have a computable rate of convergence. In fact, as shown in [40], the principle that every bounded monotone sequence of reals is Cauchy can be proven in $\mathrm{HA}^\omega + \Sigma^0_1$-LEM and for sequences defined by terms of HA^ω using only number parameters even with Σ^0_1-LEM$^-$ (where P^- denotes the restriction of an axiom schema P to number parameters only). This is not obvious and requires a novel proof as the usual argument uses the (by [1] e.g. over HA) strictly stronger principle

$$\Sigma^0_2\text{-DNE} \ : \ \neg\neg\exists n \in \mathbb{N}\, \forall m \in \mathbb{N}\, \varphi_0(n,m) \to \exists n \in \mathbb{N}\, \forall m \in \mathbb{N}\, \varphi_0(n,m)$$

('double-negation-elimination principle' for Σ^0_2-formulas). While Σ^0_2-DNE is limit computable in the sense of Hayashi and Nakana [20], any single instance of Σ^0_1-LEM is even learnable with a single mind change. Note also that bounded monotone sequences of real numbers (say in $[0,1]$) always have the simple fluctuation bound $F(k) := 2^k$. That Σ^0_1-LEM$^-$ has a strictly stronger computational interpretation than Σ^0_2-DNE$^-$ can be spelled out in terms of proof interpretations: Σ^0_1-LEM$^-$ (when added e.g. to HA^ω) admits a modified realizability interpretation by terms that are primitive recursive (in the sense of Gödels T) relative to a Skolem function f_{φ_0} for

$$\forall k \in \mathbb{N} \exists n \in \mathbb{N} \forall m \in \mathbb{N}\ (\varphi_0(k,n) \vee \neg\varphi_0(k,m)),$$

i.e.

$$\forall k, m \in \mathbb{N}\ (\varphi_0(k, f_{\varphi_0}(k)) \vee \neg\varphi_0(k,m)),$$

and much work studying this interpretation in terms of learning theory and so-called 1-backtracking games (in the sense of Coquand's game semantics [14, 12]) has been carried out in recent years e.g. by Aschieri and Berardi (see e.g. [6, 4, 5]). By contrast to this, Σ^0_2-DNE$^-$ does not allow such an interpretation and not even a (in this context) weaker monotone modified realizability interpretation (in the sense of the [24], see also [26]) as shown in [1], where this is used to prove that Σ^0_2-DNE$^-$ not even follows from Π^0_2-LEM$^-$.

Remark 1.1. The various realizability and functional interpretations referred to above are different ways of giving a precise meaning to the informal so-called Brouwer-Heyting-Kolmogorov ('BHK') interpretation of intuitionistic logic. A rather different way of formalizing BKH – based on the concept of operations on formal proofs – has been developed since the mid 90's by Sergei Artemov, first for the propositional case (see e.g. [2]) and very recently (together with Tatiana Yavorskaya) for predicate logic (see [3]).

At a first look, all this might suggest to consider $\mathrm{HA}^\omega + \Sigma^0_1\text{-LEM}^{(-)}$ as a promising framework to guarantee computable bounds on the number of fluctuations for provable Cauchy sequences (while in general not computable rates of convergence). However, this turns out to be mistaken as Σ^0_1-LEM is already the general case: let (x_n) be a sequence of real numbers definable by a term t in HA^ω (which may have variables of arbitrary type as parameters). Suppose that for the extension PA^ω of HA^ω by full classical logic

$$\mathrm{PA}^\omega \vdash \forall k\, \exists n\, \forall m, \tilde{m} \geq n\, (|x_m - x_{\tilde{m}}| \leq_{\mathbb{R}} 2^{-k}).$$

Then by negative translation (see [26])

$$\mathrm{HA}^\omega \vdash \forall k\, \neg\neg \exists n\, \forall m, \tilde{m} \geq n\, (|x_m - x_{\tilde{m}}| \leq_{\mathbb{R}} 2^{-k}).$$

Adapting Friedman's proof for the closure of HA^ω under the Markov rule one can show (this is stated for HA without proof in [20] and we include a proof – also for $\mathrm{HA}^\omega[X, \|\cdot\|]$ – below) that $\mathrm{HA}^\omega + \Sigma^0_1$-LEM is closed under the rule version of Σ^0_2-DNE so that we get

$$\mathrm{HA}^\omega + \Sigma^0_1\text{-LEM} \ \vdash \forall k\, \exists n\, \forall m, \tilde{m} \geq n\, (|x_m - x_{\tilde{m}}| \leq_{\mathbb{R}} 2^{-k}).$$

Moreover, if t contains at most number parameters it also suffices to use the restriction $\Sigma^0_1\text{-LEM}^-$ of Σ^0_1-LEM to Σ^0_1-formulas with number parameters only. All this also holds for the systems $\mathrm{HA}^\omega[X, \|\cdot\|]$ and $\mathrm{PA}^\omega[X, \|\cdot\|]$ and sequences (x_n) in X defined by terms of these systems.

So, as far as Π^0_3-theorems are concerned (also with parameters in $\mathbb{N}, \mathbb{N}^\mathbb{N}$ or – in the case of theories with an abstract normed space X – also in $X, X \to X, \mathbb{N} \to X$), there is no difference in proofs based on full classical logic versus proofs using only Σ^0_1-LEM (at least as long as the proofs can be formalized in systems to which negative translation and Friedman's A-translation apply). So in order to have a computational content stronger than metastability guaranteed, we have to look for more restricted uses of $\Sigma^0_1\text{-LEM}^-$. Looking more carefully into the Σ^0_1-LEM-based proof of the Cauchyness of bounded monotone sequences as given in [40] reveals that one can define a sequence of instances $\Sigma^0_1\text{-LEM}(s(n))$ of $\Sigma^0_1\text{-LEM}^-$

$$\exists m \in \mathbb{N}\, (s(n,m) = 0) \vee \forall m \in \mathbb{N}\, (s(n,m) \neq 0)$$

such that to prove the Cauchy property with error 2^{-k} one only needs the first $n = 0, \ldots, s(t(k))$-many instances of this sequence where t is a simple primitive recursive function. So a more promising approach would be to look at proofs of Cauchy statements which can be formalized as follows:

$$\mathrm{HA}^\omega \ \vdash \forall k\, \big(\forall l \leq t(k)\ \Sigma^0_1\text{-LEM}(s(l)) \to \exists n\, \forall m, \tilde{m} \geq n\, (|x_m - x_{\tilde{m}}| \leq_{\mathbb{R}} 2^{-k})\big)$$

or

$$\mathrm{HA}^\omega[X, \|\cdot\|] \ \vdash \forall k\, \big(\forall l \leq t(k)\ \Sigma^0_1\text{-LEM}(s(l)) \to \exists n\, \forall m, \tilde{m} \geq n\, (\|x_m - x_{\tilde{m}}\| \leq_{\mathbb{R}} 2^{-k})\big),$$

where t may contain parameters of type $\mathbb{N}, \mathbb{N} \to \mathbb{N}$ (and $X, \mathbb{N} \to X, X \to X$ in the extended system).

We show that from such proofs one can always extract effective (and in fact primitive

recursive in the sense of Gödel's T) bounds B, L on the effective learnability of a rate of convergence of (x_n). Here B, L are effective functionals (in the parameters $\underline{a}$ of the problem) where L is the learning procedure and B a bound on the number of necessary steps along this procedure. This is, as we will show, a strictly stronger information than a rate of metastability as the latter can be obtained from (majorants B^*, L^* of) the former, even by a uniform primitive recursive procedure (in the ordinary sense of Kleene). However, there are primitive recursive Cauchy sequences of rational numbers with a primitive recursive (again in the ordinary sense of Kleene) rate of metastability which do not admit any computable bound for the learnability of a rate of convergence. So while (as discussed above) a computable Cauchy sequence in $\mathbb{R}$ always has a computable rate of metastability (by unbounded search) it in general will not have computable bounds B, L on the effective learnability of a rate of convergence.
In fact, the additional information provided by B^*, L^* becomes visible by the particular simple structure of the rate of metastability obtained from B^*, L^* which is guaranteed to be of the form (essentially)

$$(L^*(\underline{a}^*) \circ \tilde{g})^{B^*(\underline{a}^*)}(0),$$

where $f^x(0)$ denotes the x-times iteration of the function f starting from 0 and $\tilde{g}(n) := \max\{g(i) : i \leq n\} + n$. The essential point here is that B^*, L^* do not involve the counterfunction g. It is precisely this form of a rate of metastability that has been observed many times in concrete unwindings in ergodic theory and fixed point theory (see e.g.[7, 29, 30, 27, 31, 32]) and which we can explain now for the first time in terms of the logical structure of the given proof.[1] For Cauchy statements, a metastability rate which has the simple form given above conversely implies that the Cauchy statement is (B^*, L^*)-learnable. Notable exceptions to this restricted format of metastability are the rates of metastability for the ergodic theorem for odd (and even more general) operators in [37] (making a nested use of the iteration procedure) and for the (weak convergence in the) Baillon nonlinear ergodic theorem in [28] (making even nested use of a bar recursive functional). However, both underlying proofs violate precisely our criterion of a bounded use of Σ^0_1-LEM.

A bound on the number of fluctuations in general is a still strictly stronger quantitative information than bounds on the effective learnability: we construct a primitive recursive sequence of rational numbers in $[0, 1]$ which has primitive recursive bounds on the learnability of its Cauchy rate but does not admit a computable bound on the number of its fluctuations. Together with the already discussed Specker sequences (which have trivial fluctuation bounds but no effective rates of convergence) we get the (w.r.t. effectivity) strictly decreasing hierarchy of quantitative data for the convergence of Cauchy sequences (x_n):

1. A rate ρ of convergence of (x_n).

2. A bound F on the number of approximate fluctuations of (x_n).

[1] Note that minor modifications of the above format which show up in these bounds, e.g. $L^*_1 \circ \tilde{g} \circ L^*_2$, can easily be reduced to this format.

3. Function(al)s B, L (see the next section for a precise definition) of the learnability of a rate of convergence for (x_n) by B-many mind changes by a learning procedure L.

4. A rate Φ of metastability for (x_n).

While – as discussed above – the extractability of effective data for the levels 1, 3 and 4 of this hierarchy is guaranteed by relatively easy to check logical **a-priori** conditions on the framework in which a Cauchy statement is proven, this seems to be different for level 2: we give a kind of gap condition to be satisfied by L in '3.' which suffices to convert the information provided by B, L into a bound F on the fluctuations. Since to check this condition requires the inspection of the extracted data B, L this is an **a-postiori** condition (which is reminiscent of the growth conditions used in Luckhardt's [35] extraction of bounds on the number of solutions in Roth's theorem).

The above hierarchy apparently fits well to distinguish the computational content that recently has been obtained from proofs in the context of ergodic theory:

As discussed already, Avigad and Rute [9] observed that the extraction of a rate of metastability from Birkhoff's proof of the mean ergodic theorem in uniformly convex Banach spaces carried out in [29] can in fact been used to even obtain a uniform effective fluctuation bound. This corresponds to second level of our hierarchy which by [7] cannot be further improved effectively to the first level. We explain this in terms of our gap condition satisfied by (B, L).

The logical condition needed to assure the extractability of a primitive recursive (in the ordinary sense of Kleene) data B, L for the third level is e.g. satisfied in the proofs of the strong convergence of so-called Halpern iterations in Hilbert spaces (due to Wittmann [42]) and – more generally – in CAT(0) spaces (due to Saejung in [36]). This follows from the analysis of Saejung's proof given in [31] where a primitive recursive (in the ordinary sense) rate of metastability is extracted (the analysis of Wittmann's proof in [27] also shows this in the Hilbert case). A special form of the Halpern iteration (covered by [42, 36]) is given by

$$x_{n+1} := \frac{1}{n+2} x_0 + \left(1 - \frac{1}{n+2}\right) T x_n$$

which can be viewed as a nonlinear generalization of the ergodic average

$$\frac{1}{n+1} \sum_{i=0}^{n} T^i x_0$$

in the mean ergodic theorem with which it coincides for **linear** nonexpansive maps T. As a corollary we obtain the extractability even of primitive recursive learnability data B, L in this case. However, the aforementioned gap condition does not hold for the data we extracted from the proofs. So, as it stands, a fluctuation bound does not seem to follow. Of course, this does not rule out at all that a different proof might yield better data which possibly could satisfy the gap condition and that, consequently, effective fluctuation bounds might result.

While the strong convergence of the ergodic average in general is known to fail for nonlinear nonexpansive maps (whereas weak convergence holds by a deep theorem of Baillon

[10]), it does hold e.g. for odd operators as shown again by Baillon [11] and – for a much more general class of mappings – by Wittmann [41]. Recently, the second author [37] extracted a (primitive recursive in the sense of Kleene) rate of metastability (i.e. level 4) from Wittmann's proof. However, Wittmann's proof does not satisfy the condition sufficient to guarantee level-3 learnability data and, in fact, the extracted bound has a structure similar to the one in our example of a primitive recursive sequence of rational numbers in $[0,1]$ separating the levels 3 and 4 given in section 4 below.

2. Fluctuations versus effective learnability

To be specific, let us use in the following the language of (intuitionistic) arithmetic in all finite types HA^ω (more precisely the system WE-HA^ω, see [26]) as well as its extension $\mathrm{HA}^\omega[X, \|\cdot\|]$ by an abstract normed space X in the sense of [25, 16, 26] in order to be able to cover also the aforementioned recent applications of proof mining to ergodic and fixed point theory which need this enriched language. Everything we say extends mutatis mutandis also to theories where more conditions on X are prescribed (e.g. X being a uniformly convex or a Hilbert space) and convex subsets C of X being added as well as to metric structures X such as metric, W-hyperbolic and CAT(0) spaces (see [26] for all this). The type of natural numbers $\mathbb{N}$ is usually denoted by 0 while 1 denotes the type of functions $\mathbb{N} \to \mathbb{N}$.

$\mathcal{S}^{\omega,X}$ denotes the full set-theoretic model of these theories over the base types $\mathbb{N}$ and X. Occasionally, we will need the relation 'x^* majorizes x' (short: $x^*\ maj\ x$) due to W.A. Howard (for the finite types over $\mathbb{N}$) which is defined in the usual hereditary way by induction on the type of x starting from

$$x^*\ maj_0\ x\ :\equiv x^* \geq x$$

for x^*, x of type 0 and

$$x^*\ maj_X\ x\ :\equiv x^* \geq \|x\|$$

for x of type X and x^* of type 0.

Remark 2.1. Throughout this paper, we will use several encodings. We use j for the Cantor pairing function and j_1 and j_2 for the corresponding projections. Moreover, we use $\langle \underline{a} \rangle$ for both

- a surjective primitive recursive encoding of tuples of a given length l, with the corresponding projections $j_1, \ldots, j_l$ and
- a surjective sequence encoding (which then includes the length of the encoded sequence) with primitive recursive functions for length lh, concatenation $*$, and projection $(\cdot)_{(\cdot)}$ (i.e. $(n)_k$ for the $k+1$-th element of a finite sequence encoded by n and 0 for $k \geq lh(n)$,) when there is not danger of confusion we use also the simpler notation n_k). For details see [26].
- For both the tuple and the sequence encoding we assume the coding to be increasing in each component and that $\langle a_0, \ldots, a_{k-1} \rangle \geq a_i$ for $i < k$.

For a specific encoding satisfying these requirements see [26], where the sequence encoding is denoted by $\langle a_0, \dots, a_{k-1} \rangle$ while the k-tuple encoding is denoted by $\nu^k(a_0, \dots, a_{k-1})$. Whether we mean a tuple or a sequence coding should be mostly clear from the context (roughly, we mean tuple encoding whenever the length is fixed and sequence encoding otherwise), but whenever this is relevant we say also explicitly which encoding is meant.

Definition 2.2 (the number of fluctuations)**.** For a sequence $x_{(\cdot)}$ in some metric space (X, d) and an $\varepsilon > 0$ let $\mathrm{Fluc}(n, i, j)$ denote that there are n fluctuations whose indexes are encoded into i and j.

$$\begin{aligned} \mathrm{Fluc}(n,i,j) :\equiv \mathrm{Fluc}_{x_{(\cdot)},\varepsilon}(n,i,j) :\equiv \quad & \mathrm{lh}(i) = \mathrm{lh}(j) = n \quad \wedge \\ & \forall k < n \ (i_k < j_k) \quad \wedge \\ & \forall k < n-1 \ (j_k \leq i_{k+1}) \quad \wedge \\ & \forall k < n \ (d(x_{i_k}, x_{j_k}) > \varepsilon). \end{aligned}$$

We call b a bound on the number of ε-fluctuations of $x_{(\cdot)}$, iff

$$\forall n > b \forall i, j \neg \mathrm{Fluc}(n, i, j).$$

We call b effective if it is computable in $\varepsilon \in \mathbb{Q}^*_+$ and $x_{(\cdot)}$.

In the Language identification in the limit model for inductive inference, the notion of learnable with an existence of a mind change bound was introduced in the sixties (see e.g. [17]). We define a similar concept in the context of general formal theories like PA^ω. Since we require both the learning procedure and the bound on mind changes to be recursive (effective) we call this property *effective learnability*.

Remark 2.3. The proof-theoretic study of learnability by finitely many (though not necessarily effectively bounded) mind changes in analysis has been initiated by Hayashi (see e.g. [18, 19]) who (with Nakata) established the close relation of this concept to limit computability (see e.g. [20]). The concept of mind change for Cauchy statements is also implicit in section 5.1 of [44] (Proof of Lemma 31.c). Effective learnability concepts for functionals $F : D \to \mathbb{N}^{\mathbb{N}}$ (with $D \subseteq \mathbb{N}^{\mathbb{N}}$) have recently been investigated in [21].

On the one hand we would like the learning procedure to be as simple as possible and on the other hand we would like to formalize that it can access as much finite information as is available (at a given learning step). In the case of monotone formulas (which is a rather rich class of statements including, in particular, all Cauchy statements), there is a straightforward answer (see Definition 2.4 and Proposition 2.5) allowing us to simplify the theory of learnability, if we assume monotonicity. We give a more general definition (Definition 2.9), which coincides with Definition 2.4 in the monotone case, a few pages later.

Definition 2.4 ((B, L)-learnable monotone formulas)**.** Consider a Σ^0_2 formula φ with the only parameters $\underline{a}^\sigma$, i.e.

$$\varphi \ \equiv \ \exists n^0 \forall x^0 \ \varphi_0(x, n, \underline{a}),$$

which is monotone in n, i.e.

$$\forall n^0 \ \forall n' \geq n \ \forall x^0 \ \big(\varphi_0 \ (x, n, \underline{a}) \to \varphi_0(x, n', \underline{a})\big).$$

We call such a formula φ *(B,L)-learnable*, if there are function(al)s B and L such that the following holds (in the full set-theoretic model $\mathcal{S}^{\omega,X}$):

$$\exists i \leq B(\underline{a})\ \forall x\ \varphi_0(x, c_i, \underline{a}),$$

where

$$c_0 := 0,$$
$$c_{i+1} := \begin{cases} L(x,\underline{a}), & \text{for the } x \text{ with } \neg\varphi_0(x,c_i,\underline{a})\ \wedge\ \forall y < x\ \varphi_0(y,c_i,\underline{a}) \text{ if it exists} \\ c_i, & \text{otherwise.} \end{cases}$$

We call such a φ *effectively learnable (with effectively bounded many mind changes)* if it is (B,L)-learnable with computable functionals B and L.[2]

In section 4.2 (Proposition 4.13) we will construct a φ which is true for all parameters $\underline{a} \in \mathbb{N}^{\mathbb{N}}$ but which is not (B,L)-learnable with computable B, L.

This definition is very intuitive in the sense that it formalizes the concept of an (effective) learning process L which learns the witness in an effectively bounded number of attempts in a very straightforward way.

Moreover, this definition allows the learner, i.e. the function L to use the least amount of non-computable information possible, namely only the smallest counterexample to the learners last candidate for the witness. Nevertheless, we will show that this amount of information is, in a sense, already exhaustive. More precisely, we have the following (we use in the rest of this section a surjective sequence coding denoted by $\langle\cdots\rangle$ with primitive recursive functions $lh, *, (n)_k$ as discussed in Remark 2.1):

Proposition 2.5. *Consider a monotone formula φ as above. Suppose there are B and L' s.t.*

$$\exists i \leq B(\underline{a})\ \forall x\ \varphi_0(x, c'_i, \underline{a}),$$

where this time L' can access all reasonable information, i.e.

$$c'_0 := 0,$$
$$c'_{i+1} := \begin{cases} L'(\langle x_0,\ldots,x_i\rangle, \langle c'_0,\ldots,c'_i\rangle, \underline{a}), & \textit{for those } x_j, c'_j,\ j \leq i \textit{ with} \\ & \neg\varphi_0(x_j,c'_j,\underline{a})\ \wedge\ \forall y < x_j\ \varphi_0(y,c'_j,\underline{a}) \\ & \textit{if each exists,} \\ c'_i, & \textit{otherwise.} \end{cases}$$

Then φ is (B,L)-learnable in the original sense (as defined in Definition 2.4), where L is primitive recursively definable in B, L' and the characteristic function of φ_0 (and so, in particular as $\xi(B,L')$ for a closed term ξ of the system at hand).

Proof. W.l.o.g we can assume that there is a $j \leq B(\underline{a})$ s.t. $\forall i < j\ (c'_{i+1} > c'_i)$ and $\forall i \geq j\ (c'_{i+1} = c'_i)$ (we can actually primitive recursively define a learner which satisfies

[2]Note that c_i is the i-th attempt to produce a *c*andidate for a valid n, while B is a *B*ound on the number of such attempts produced before a valid candidate is *L*earned by the procedure L .

this property whenever we have B and L' as above). We set

$$L(x,\underline{a}) := L(\langle\rangle,\langle\rangle,x,\underline{a}) := \begin{cases} 0, & \text{if } \varphi_0(x,0,\underline{a}),\\ L(\langle X(x,0,\underline{a})\rangle,\langle 0\rangle,x,\underline{a}), & \text{otherwise,}\end{cases}$$

$$L(\underbrace{\langle x_0,\dots,x_i\rangle}_{\underline{x}},\underbrace{\langle c_0,\dots,c_i\rangle}_{\underline{c}},x,\underline{a}) := \begin{cases} c_i, & \text{if } \bigvee_{j\le i}\varphi_0(x_j,c_j,\underline{a})\vee i\ge B(\underline{a}),\\ l' := L'(\langle\underline{x}\rangle,\langle\underline{c}\rangle,\underline{a}), & \text{if } x=x_i\vee\varphi_0(x,l',\underline{a}),\\ & \quad \bigwedge_{j\le i}\neg\varphi_0(x_j,c_j,\underline{a})\wedge i<B(\underline{a}),\\ L(\langle\underline{x},X(x,l',\underline{a})\rangle,\langle\underline{c},l'\rangle,x,\underline{a}) & \text{otherwise,}\end{cases}$$

where $X(x,c,\underline{a}) := \min\{x' \le x \ : \ \neg\varphi_0(x',c,\underline{a})\}$ (or x if there is no such x').
We show by induction on i that $\forall i\ (c'_i = c_i)$. This is obvious for $i=0$, moreover if $\forall x\ \varphi_0(x,c'_i,\underline{a})$ then also $\forall x\ \varphi_0(x,c_i,\underline{a})$ and so $c'_{(\cdot)} = c_{(\cdot)}$ (both) by the induction hypothesis $\forall j\le i(c'_j = c_j)$.
Otherwise, we have the smallest counterexample x'_i to c'_i, and since by our hypothesis $c'_i = c_i$ we have also $x_i = x'_i$ for the smallest counterexample to c_i (note that, by the (B,L')-learnability of φ, we have $i < B(\underline{a})$). So, by the monotonicity of φ, we obtain for all $j<i$ that $x'_j \le x_i$, so $X(x_i,c'_j,\underline{a}) = x'_j$ and by definition of L we get in total that

$$\begin{aligned} c_{i+1} = L(x_i,\underline{a}) &= L(\langle X(x_i,0,\underline{a})\rangle,\langle 0\rangle,x_i,\underline{a})\\ &= L(\langle x'_0\rangle,\langle 0\rangle,x_i,\underline{a})\\ &= L(\langle x'_0,X(x_i,L'(\langle x'_0\rangle,\langle 0\rangle,\underline{a}),\underline{a})\rangle,\langle 0,L'(\langle x'_0\rangle,\langle 0\rangle,\underline{a})\rangle,x_i,\underline{a})\\ &= L(\langle x'_0,X(x_i,c'_1,\underline{a})\rangle,\langle 0,c'_1\rangle,x_i,\underline{a})\\ &= L(\langle x'_0,x'_1\rangle,\langle 0,c'_1\rangle,x_i,\underline{a})\\ &= \dots\\ &= L'(\langle\underline{x}'\rangle,\langle\underline{c}'\rangle,\underline{a}) = c'_{i+1}.\end{aligned}$$

□

So from now on we will simply use L in the form which suits us best.

Speaking of a Cauchy sequence $a_{(\cdot)}$, we would say that it has an effectively learnable rate of convergence, if there is a recursive computation for a bound b from the sequence $a_{(\cdot)}$ (resp. the parameters used in defining $a_{(\cdot)}$) and an $\varepsilon>0$, such that there is a procedure to learn an ε-Cauchy point with at most b computable corrections (computable in a counterexample x, which in turn may not be computable itself!).

Remark 2.6. In Definition 2.4, even the condition that x is the smallest counterexample, i.e. $\forall y<x\ \varphi_0(y,c_i,\underline{a})$, is not really necessary. Of course, in such case, for a given learner, the sequence $c_{(\cdot)}$ is not unique and we need to specify what actually is bounded by B. Fortunately, there are only two natural options. Either we say that B is a bound on any sequence $c_{(\cdot)}$ (i.e. B is independent on the choice of the counterexamples) or we say that B is a bound for some sequence $c_{(\cdot)}$ (i.e. for at least one suitable choice of counterexamples). It seems rather obvious that the second option makes little sense, since if there was any bound and learner at all, then $B(\underline{a})=1$ would be a correct bound

for the same learner as well (simply by choosing the right counterexample as the first input). Moreover, a definition in the new sense that B is a bound on any sequence $c_{(\cdot)}$ (any choice of $x_{(\cdot)}$) would be equivalent to Definition 2.4.

- Any given bound B for all such sequences is obviously, in particular, a bound for the one we used in the original definition. Therefore any formula (B, L)-learnable in the new sense is, in particular, (B, L)-learnable in the old sense.
- On the other hand, given B and L satisfying our original definition, we could modify L to L' in such a way, that it actually looks for the smallest counterexample and uses that for its computation, assuring that we in fact generate the same sequence $c_{(\cdot)}$ after all (e.g. set $L'(x, c, \underline{a}) := L(\min\{x' \leq x \ : \ \neg\varphi_0(x', c, \underline{a}) \ \wedge \ \forall y < x' \ \varphi_0(y, c, \underline{a})\}, c, \underline{a})$ if such an x' exists and $L'(x, c, \underline{a}) := L(x, c, \underline{a})$ otherwise). In other words, any formula (B, L)-learnable in the old sense, is (B, L')-learnable in the new sense.

As far as monotone formulas are concerned, we have yet another nice property.

Proposition 2.7. *A monotone Σ^0_2-formula φ (see also Definition 2.4) that is (B, L)-learnable (uniformly in the parameters $\underline{a}$) is also (B^*, L^*)-learnable (uniformly in majorants $\underline{a}^*$ of $\underline{a}$) for any majorants B^*,L^* of B, L, i.e. (in $\mathcal{S}^{\omega,X}$)*

$$\forall \underline{a}^*, \underline{a}\big(\underline{a}^* \ \text{maj} \ \underline{a} \to \exists i \leq B^*(\underline{a}^*) \ \forall x^0 \ \varphi_0(x, c^*_i, \underline{a})\big),$$

where

$$c^*_0 := 0,$$

$$c^*_{i+1} := \begin{cases} L^*(x, \underline{a}^*), & \text{for the } x \text{ with } \neg\varphi_0(x, c^*_i, \underline{a}) \ \wedge \ \forall y < x \ \varphi_0(y, c^*_i, \underline{a}) \text{ if it exists} \\ c^*_i, & \text{otherwise.} \end{cases}$$

Proof. Note that $B^*, L^*, \underline{a}^* \ maj \ B, L, \underline{a}$ implies that

$$B^*(\underline{a}^*) \geq B(\underline{a}) \wedge \forall x^0, y^0(x \geq y \to L^*(x, \underline{a}^*) \geq L(y, \underline{a}))).$$

Now assume that $c^*_i \geq c_i$. Then by the monotonicity of φ we have for all x

$$\neg\varphi_0(x, c^*_i, \underline{a}) \to \neg\varphi_0(x, c_i, \underline{a})$$

and so the smallest counterexample x_i to c_i is smaller (or equal) than the smallest counterexample x^*_i to c^*_i and so $c^*_{i+1} = L^*(x^*_i, \underline{a}^*) \geq L(x_i, \underline{a}) = c_{i+1}$. Inductively, we get $c^*_i \geq c_i$ for all $i \leq B(\underline{a})$. □

Remark 2.8. If the parameters $\underline{a}$ have all types of degree ≤ 1, then φ in the above proposition is learnable in B^*, L^* uniformly in $\underline{a}$ since $a^M \ maj \ a^1$, where $a^M(n) := \max\{a(i) : i \leq n.\}$.
In this sense, we can extend the term effectively learnable as follows. A monotone Σ^0_2-formula $\varphi(\underline{a})$ is *effectively learnable with finitely many mind changes uniformly in majorants $\underline{a}^*$ of the parameters $\underline{a}$* if it is (B^*, L^*)-learnable (uniformly in majorants $\underline{a}^*$ of the parameters $\underline{a}$ by computable functionals B^*, L^* and all elements of $\underline{a}^*$ are of type level at most one). Note that this means that in the System $\mathsf{HA}^\omega[X, \|\cdot\|]$, $\underline{a}$ could include parameters of types like $X, \mathbb{N} \to X, X \to X$.

There are several ways to generalize our learnability definition. Of course one can drop the monotonicity condition, but we can also allow higher or abstract types for n and x (for x we would need to consider all sequences $c_{(\cdot)}$ since we can provide only a counterexample x, not the smallest counterexample x – see also Remark 2.6). The question here is, what kind of information we do allow the learning function(al) L to access. At the moment, it seems that there is not such a nice and definitive answer as in the monotone case. However, we will stick with a (not necessarily unique) definition (see Definition 2.9), which

1. generalizes Definition 2.4, i.e. if a monotone formula (assuming the bound variables to be of type 0) is learnable according to our new definition, it is also learnable in the sense of Definition 2.4 and vice-versa,
2. while keeping the arguments of the learner L simple, still is equivalent to the case where the learner has access to the full finitary information in the sense of Proposition 2.5 (see Remark 2.10)
3. fits very nicely into the hierarchy of different concepts for computational information (see Proposition 2.16),
4. makes effective learnability guaranteed by very clear logical conditions on the provability of the learned formula (see Theorem 2.11).

The second point seems a very natural requirement and is the cause for the main difference to the monotone case, which is that in Definition 2.9 the learning process may depend on a whole tuple of all counterexamples used so far, rather than only on the last one (last in the sense of number of guesses/candidates, not the index of the counterexample).

Although we do not treat the case of learnability for higher type objects in this paper, the following definition easily applies to this case as well and, therefore, is written in this generality:

Definition 2.9 ((B,L)-learnability for general (not necessarily monotone) formulas)**.** Consider an $\exists\forall$ formula φ with the only parameters $\underline{a}^{\underline{\sigma}}$, i.e.

$$\varphi \;\equiv\; \exists n^{\rho}\forall x^{0}\; \varphi_0(x, n, \underline{a}).$$

We call such a formula φ (B, L)-*learnable*, if there are function(al)s B and L such that the following holds:

$$\exists i \leq B(\underline{a})\; \forall x\; \varphi_0(x, c_i, \underline{a}),$$

where

$$c_0 := 0^{\rho},$$

$$c_{i+1} := \begin{cases} L(\langle x_0, \ldots, x_i\rangle, \underline{a}), & \text{for those } x_j,\ j \leq i \text{ with} \\ & \neg\varphi_0(x_j, c_j, \underline{a}) \;\wedge\; \forall y < x_j\; \varphi_0(y, c_j, \underline{a}) \\ & \text{if each exists,} \\ c_i, & \text{otherwise.} \end{cases}$$

We call such a φ *effectively learnable*, if it is (B, L)-learnable, σ_i and ρ have type level at most one, and B and L are computable.

Remark 2.10. Again, this definition already captures (so to say in a primitive recursive way) the case where the learner could access the previous values of $c_{(\cdot)}$ as well. Simply consider

$$L(\overbrace{\langle x_0,\dots,x_i\rangle}^{\underline{x}:=},\underline{a}) :=$$
$$L'\big(\langle\underline{x}\rangle,\big\langle \underbrace{L'(\langle x_0\rangle,\langle 0\rangle,\underline{a})}_{c_1':=},\underbrace{L'(\langle x_0,x_1\rangle,\langle 0,c_1'\rangle,\underline{a})}_{c_2':=},\dots,L'(\langle\underline{x}\rangle,\langle 0,c_1',\dots,c_i'\rangle,\underline{a})\big\rangle,\underline{a}\big).$$

Of course, one could consider weaker concepts, like a learner which can access only the last counterexample (as in the monotone case). We considered also a learner of the kind $L'(x,c,\underline{a})$ (i.e. a learner who is allowed to use in addition only the lastly learned solution candidate), which doesn't seem to be equivalent to any of the other two concepts.

Let us make the properties of our learnability definition discussed above more transparent by proving the following results which we first briefly motivate: as mentioned already in the introduction (and proved further below in section 3), any classical proof (in a suitable formal system) of a Cauchy statement $\varphi(k) := \exists n\forall i,j\geq n\,(d(x_i,x_j)<2^{-k})$ can be reformulated to use classical logic only up to $\forall l^0(\Sigma^0_1\text{-LEM}(s(l,k)))$ (for some closed term s), i.e.

$$(a)\ \forall k^0\ \big(\forall l^0\,(\Sigma^0_1\text{-LEM}(s(l,k)))\to\varphi(k)\big)$$

follows intuitionistically. This, in particular, applies to the example from section 4.2 of a computable Cauchy sequence in $\mathbb{R}$ which does not possess any computable learnability bounds B,L. So in order to be able to extract such computable data B,L, the Cauchy proof has to be further restricted, namely, to the situation where the proof implicitly contains a bound $t(k)$ on $\forall l^0$, i.e. on the instances of Σ^0_1-LEM$(s(l,k))$ used. This is guaranteed (as we will see) when (a) is strengthened to the (only noneffectively equivalent) form

$$(b)\ \forall k^0\,\exists l^0\,\big(\forall m\leq_0 l\,(\Sigma^0_1\text{-LEM}(s(m,k)))\to\varphi(k)\big).$$

Then from a (semi-intuitionistic) proof of (b) one can extract a term $B(k)$ computing l and two further terms which allow one to build a learning procedure L so that φ is (B,L)-learnable.
In the following $\text{IP}^\omega_\forall$ denotes the independence-of-premise principle for universal premises in all finite types:

$$\text{IP}^\omega_\forall:\quad(\forall\underline{x}\,A_0(\underline{x})\to\exists y\,B(y))\to\exists y\,(\forall\underline{x}\,A_0(\underline{x})\to B(y)),$$

where A_0 is a quantifier-free formula and $\underline{x},y$ are variables of arbitrary types.

Theorem 2.11. *Given that*

$$\mathsf{HA}^\omega[X,\|\cdot\|]+\text{AC}+\text{M}^\omega+\text{IP}^\omega_\forall\vdash$$
$$\forall\underline{a}\,\exists l^0\,\big(\forall m\leq_0 l\exists u^0\forall v^0\,\big(\psi_0(u,m,\underline{a})\vee\neg\psi_0(v,m,\underline{a})\big)\to\exists n^0\forall x^0\varphi_0(x,n,\underline{a})\big),$$

where φ_0,ψ_0 are quantifier-free formulas (containing at most the parameters $\underline{a}$ free), then $\exists n\forall x\varphi_0(x,n,\underline{a})$ is (valid in $\mathcal{S}^{\omega,X}$) (B,L)-learnable (in the sense of Definition 2.9 and,

for monotone formulas, in the sense of Definition 2.4) by functionals given by closed terms of $\mathsf{HA}^\omega[X, \|\cdot\|]$.
To B, L *one can construct majorants* B^*, L^* *given by closed terms of* HA^ω *such that if* $\exists n \forall x \varphi_0(x, n, \underline{a})$ *is monotone (as in Definition 2.4), then it is even learnable in* B^*, L^* *uniformly in majorants* $\underline{a}^*$ *of the parameters* $\underline{a}$.

Proof. Suppose that

$$\mathsf{HA}^\omega[X, \|\cdot\|] + \mathrm{AC} + \mathrm{M}^\omega + \mathrm{IP}^\omega_\forall \vdash$$
$$\forall \underline{a} \exists l^0 \, \big(\forall m \le l \exists u \forall v \, \big(\psi_0(u, m, \underline{a}) \vee \neg\psi_0(v, m, \underline{a})\big) \to \exists n \forall x \varphi_0(x, n, \underline{a})\big).$$

Then by the soundness of the Gödel functional ('Dialectica') interpretation for $\mathsf{HA}^\omega[X, \|\cdot\|] + \mathrm{AC} + \mathrm{M}^\omega + \mathrm{IP}^\omega_\forall$ (see [26]) we obtain that (note that since we do not need bar recursion to interpret $\mathsf{HA}^\omega[X, \|\cdot\|]$ we do not have to go through the model of strongly majorizable functionals and so do not need to assume any smallness condition on the types of $\underline{a}$ to pass to $\mathcal{S}^{\omega, X}$)

$$\mathcal{S}^{\omega, X} \vDash \exists l, V, N \forall U, x \, \Big(\forall m \le l\big(\psi_0(Um, m, \underline{a}) \vee \neg\psi_0(VxU, m, \underline{a})\big) \to \varphi_0(x, N(U), \underline{a})\Big).$$

where '$\exists l, V, N$' is witnessed (uniformly in $\underline{a}$) by closed terms t, s_V, s_N of $\mathsf{HA}^\omega[X, \|\cdot\|]$. The result when the terms s_V, s_N are applied to $\underline{a}$, we conveniently name V and N.
To show the learnability, let $U_{\underline{v}}$ (where $\underline{v}$ is a $t(\underline{a})$-tuple) denote the function

$$U_{\underline{v}}(i) := \begin{cases} v_i & \text{if } \ i < t(\underline{a}), \\ 0 & \text{otherwise,} \end{cases}$$

set $B(\underline{a}) := t(\underline{a})$ and define L in N and V via a sequence of $t(\underline{a})$-tuples $\underline{v}^{(\cdot)}$. More precisely to compute $L(\langle \underbrace{x_0, x_1, \ldots, x_i}_{\underline{x} :=} \rangle, \underline{a})$ for some i we need to define the tuples $\underline{v}^{(0)}, \ldots, \underline{v}^{(i)}$ as follows.

$\underline{v}^0$ Set $\underline{v}^0 := 0, \ldots, 0$ and $c_1 := L(\langle x_0 \rangle, \underline{a}) := N(U_{\underline{v}^0})$.

$\underline{v}^1$ If $\forall x \varphi_0(x, c_1, \underline{a})$ holds, then there is nothing to be done[3]. Otherwise, we have in particular (provided that x_1 is the minimal counterexample)

$$\exists m \le t(\underline{a})\big(\neg\psi_0(U_{\underline{v}^0} m, m, \underline{a}) \wedge \psi_0(Vx_1(U_{\underline{v}^0}), m, \underline{a})\big)$$

and so we can denote the least such an m by m_0 (put $m_0 := 0$ in case such an m does not exist) and define $\underline{v}^1$ as $\underline{v}^0$ except that we set $v^1_{m_0} := Vx_1(U_{\underline{v}^0})$. Furthermore, we set $c_2 = L(\langle x_0, x_1 \rangle, \underline{a}) := N(U_{\underline{v}^1})$. Note that we have

$$\neg\psi_0(U_{\underline{v}^0} m_0, m_0, \underline{a}) \wedge \psi_0(v^1_{m_0}, m_0, \underline{a}). \tag{v0}$$

[3] Of course this is undecidable, however the conclusion discussed next is. In this sense if the conclusion is wrong for the x_1 given as input to L, we can simply set $\underline{v}^1 = \underline{v}^0$ and $L(\langle x_0, x_1 \rangle, \underline{a}) := c_1$ (or even 0 for all that it matters).

$\underline{v}^2$ Now, if $\forall x\varphi_0(x, c_2, \underline{a})$ then we are finished. Otherwise, similarly as before we have

$$\exists m \leq t(\underline{a})\big(\neg\psi_0(U_{\underline{v}^1}m, m, \underline{a}) \wedge \psi_0(Vx_2(U_{\underline{v}^1}), m, \underline{a})\big)$$

and we can denote the least such an m by m_1 and define $\underline{v}^2$ as $\underline{v}^1$ except that we set $v^2_{m_1} := Vx_2(U_{\underline{v}^1})$. As before this means that

$$\neg\psi_0(U_{\underline{v}^1}m_1, m_1, \underline{a}) \wedge \psi_0(v^2_{m_1}, m_1, \underline{a}), \tag{v1}$$

so in particular we obtain that $m_1 \neq m_0$ by (v0) as $U_{\underline{v}^1}m_1 = v^1_{m_1}$. We set $c_3 = L(\langle x_0, x_1, x_2\rangle, \underline{a}) := N(U_{\underline{v}^2})$ and continue.

$\underline{v}^3$ Again, if $\forall x\varphi_0(x, c_3, \underline{a})$ then we are finished. Otherwise, as before, we have that

$$\exists m \leq t(\underline{a})\big(\neg\psi_0(U_{\underline{v}^2}m, m, \underline{a}) \wedge \psi_0(Vx_3(U_{\underline{v}^2}), m, \underline{a})\big)$$

and we can denote the least such an m by m_2 and define $\underline{v}^3$ as $\underline{v}^2$ except that we set $v^3_{m_2} := Vx_3(U_{\underline{v}^2})$. As before this means that

$$\neg\psi_0(U_{\underline{v}^2}m_2, m_2, \underline{a}) \wedge \psi_0(v^3_{m_2}, m_2, \underline{a}), \tag{v2}$$

so in particular we obtain that $m_2 \neq m_1$ by (v1). Moreover, by (v0) and (v2) we have $m_2 \neq m_0$, since from $m_1 \neq m_0$ follows that $v^2_{m_0} = v^1_{m_0}$. We set $c_4 = L(\langle x_0, x_1, x_2, x_3\rangle, \underline{a}) := N(U_{\underline{v}^3})$ and continue.

$\underline{v}^{n+1}$ Finally, in general assume that for some n we have that $\forall i < n \forall j < i\ m_i \neq m_j$ and $\forall i \leq n+1 \neg\forall x\varphi_0(x, c_i, \underline{a})$. Then we have also that

$$\forall i < n\big(\neg\psi_0(U_{\underline{v}^i}m_i, m_i, \underline{a}) \wedge \psi_0(Vx_{i+2}(U_{\underline{v}^{i+1}}), m_i, \underline{a})\big). \tag{vi}$$

As usual we have in particular that

$$\exists m \leq t(\underline{a})\big(\neg\psi_0(U_{\underline{v}^n}m, m, \underline{a}) \wedge \psi_0(Vx_{n+1}(U_{\underline{v}^n}), m, \underline{a})\big)$$

and we can denote the least such an m by m_n and define $\underline{v}^{n+1}$ as $\underline{v}^n$ except that we set $v^{n+1}_{m_n} := Vx_{n+1}(U_{\underline{v}^n})$. As before this means that

$$\neg\psi_0(U_{\underline{v}^n}m_n, m_n, \underline{a}) \wedge \psi_0(v^{n+1}_{m_n}, m_n, \underline{a}). \tag{vn}$$

From $\forall 0 < i < n$ $(m_0 \neq m_i)$ it follows that $\forall 0 < i < n$ $(v^n_{m_0} = v^i_{m_0})$. Assume that $m_n = m_0$, then

$$U_{\underline{v}^n}m_n = v^n_{m_n} = v^n_{m_0} = v^1_{m_0}$$

and we obtain a contradiction as $\neg\psi_0(v^1_{m_0}, m_0, \underline{a})$ follows from (vi) and $\psi_0(v^1_{m_0}, m_0, \underline{a})$ follows from (vn). This shows that $m_n \neq m_0$, similarly one shows that

$$\forall i < n\ (m_n \neq m_i).$$

As usual, we set $c_{n+2} = L(\langle x_0, \ldots, x_{n+1}\rangle, \underline{a}) := N(U_{\underline{v}^{n+1}})$.

This leads to the following definition of L:

$$L(x,\underline{a}) := L(\underbrace{\langle x_0, x_1, \dots, x_i\rangle}_{\underline{x}:=}, \underline{a}) := N(U_{\underline{v}^i}).$$

Note that since N and U are total, so is L. Moreover, if the values of i, $\underline{x}$, and c_i satisfy the conditions from Proposition 2.5, then L behaves as described above.
Finally, since there can be only $t(\underline{a})$ many different m_i's, it can happen at most $t(\underline{a})$ many times that $\forall x \varphi_0(x, c_i, \underline{a})$ does not hold, where c_i is defined as in Definition 2.9 with L as above. Hence φ is (B, L)-learnable in the sense of Definition 2.9 and hence - for monotone formulas - also $(B, \tilde{L})$-learnable for some $\tilde{L}$ primitive recursive in B, L (and φ_0) by Proposition 2.5.
The second claim follows from the fact that t, s_V, s_v have majorants t^*, s_V^*, s_N^* given by closed terms of HA^ω (see [26]) which then yield majorants B^*, L^* of B, L. Now apply Proposition 2.7. □

Remark 2.12. Assume that φ_0 in the theorem comes from a Cauchy statement

$$j_1(x), j_2(x) \geq n \to \|a_{j_1(x)} \widehat{-} a_{j_2(x)}\|(k+1) \leq_{\mathbb{Q}} 2^{-k},$$

where – referring to the representation of real numbers by number theoretic functions f representing fast Cauchy sequences of rationals – $\widehat{f}(k+1)$ is a 2^{-k-1}-rational approximation to f (see [26] for details). Then φ is monotone and a counterexample x to n satisfies $x \geq n$ (using that for the Cantor pairing function $x \geq j_i(x)$). Assume also that ψ_0 is monotone in u (which always can be arranged by taking $\psi_0'(u, m, \underline{a}) :\equiv \exists \tilde{u} \leq u\, \psi_0(u, m, \underline{a})$).
Then the complicated iteration used in defining L can be avoided by taking simply

$$L^*(\langle x_0, \dots, x_i\rangle, \underline{a}^*) := N^*_{\underline{a}^*}(\lambda k. V^*_{\underline{a}^*}(x_i, \lambda n. x_i)),$$

where $\lambda \underline{a}^*. N^*_{\underline{a}^*}, \lambda \underline{a}^*. V^*_{\underline{a}^*}$ are majorants of

$$\tilde{N}_{\underline{a}}(f) := \max\left\{\max\left\{N(v^0 * 0) : lh(v) = t\underline{a} \wedge \forall l \leq t\underline{a}(v_l \leq f(l))\right\}, f(l) : l \leq t\underline{a}\right\}$$

and

$$\tilde{V}_{\underline{a}}(x, f) := \max\left\{\max\left\{V(x, v^0 * 0) : lh(v) = t\underline{a} \wedge \forall l \leq t\underline{a}(v_l \leq f(l))\right\}, f(l) : l \leq t\underline{a}\right\}$$

with N, V, t as in Theorem 2.11. Note that with N, V also $\tilde{N}, \tilde{V}$ satisfy the claim in the proof and that L^* (for counterexamples $x_0, \dots, x_i$) is an upper bound for the L defined in terms of $\tilde{N}, \tilde{V}$ as an elementary calculation shows (using that – by the form of φ_0 – a counterexample x to n has to satisfy $x \geq n$). By monotonicity, φ is then also (L^*, B^*)-learnable (uniformly in majorants $\underline{a}^*$ of $\underline{a}$) where B^* is some majorant of B.

Remark 2.13. The theorem remains valid if arbitrary $\mathcal{S}^{\omega, X}$-true purely universal sentences are added as axioms to $\mathrm{HA}^\omega[X, \|\cdot\|]$. The part about the majorizing terms B^*, L^* even remains valid – using monotone functional interpretation – if one adds sentences of the form $\Delta :\equiv \forall a^\delta \exists b \leq_\rho s a \forall c^\tau F_0(a, b, c)$ with quantifier-free F_0 and closed s as axioms which covers the case of the binary ('weak') König's lemma WKL (which together with AC even implies König's lemma KL); see [26] for extensive details on all this.

Using the representation of real numbers from [26], each sequence of type $0 \to 1$ can be viewed as a name of a sequence (a_n) of reals. Now define $\tilde{a}_n := \max_{\mathbb{R}}\big(0, \min_{i\leq n} a_i\big)$. Let $\mathrm{PCM}_{ar}(a_n)$ denote the statement that the monotone decreasing sequence $(\tilde{a}_n)$ in $[0,1]$ is Cauchy (see [26] for details)

$$\mathrm{PCM}_{ar}(a_n):\ \forall k\in\mathbb{N}\ \underbrace{\exists n\in\mathbb{N}\forall m\geq n\ (|\tilde{a}_m-\tilde{a}_n|\leq 2^{-k})}_{\mathrm{PCM}_{ar((a_n),k):\equiv}}$$

(if (a_n) is already a decreasing sequence in $[0,1]$, then $(\tilde{a}_n)=(a_n)$). The usual classical proof of $\mathrm{PCM}(a_n)$ uses Σ^0_2-DNE, but it can be converted into a proof that only needs the weaker Σ^0_1-LEM (see Proposition 3.3 below). In [40], an explicit such proof is constructed exhibiting a concrete sequence of instances of Σ^0_1-LEM sufficient for this. From this proof one can read off the following even more detailed fact:

Proposition 2.14. *There is a primitive recursive functional (in the ordinary sense) Φ such that (using the Cantor pairing function)*

$$\mathrm{HA}^\omega \vdash \forall a^{0\to 1}_{(\cdot)}, k^0\ \big(\forall m\leq j(2^k-1,2^k)\,\Sigma^0_1\text{-LEM}(\Phi(a_{(\cdot)},m)\to\ \mathrm{PCM}_{ar}(a_{(\cdot)},k)\big).$$

Proof. The crucial step in Toftdal's proof in [40] is to show by induction on k that

$$\forall k\exists i\in\{1,\dots,2^k\}\exists n\forall m\left(\frac{i-1}{2^k}\leq\tilde{a}_{n+m}\leq\frac{i}{2^k}\right),$$

where in the induction step Σ^0_1-LEM is used in the form (note that, based on our representation of real numbers, $<_{\mathbb{R}}\in\Sigma^0_1$)

$$\exists n\ \left(\tilde{a}_n<\frac{2i-1}{2^{k+1}}\right)\vee\neg\exists n\ \left(\tilde{a}_n<\frac{2i-1}{2^{k+1}}\right).$$

So to establish $\mathrm{PCM}_{ar}(a_{(\cdot)},k)$ one needs only the instances

$$\exists n\ \left(\tilde{a}_n<\frac{i}{2^l}\right)\vee\neg\exists n\ \left(\tilde{a}_n<\frac{i}{2^l}\right)$$

for $i\leq l-1$ and $l\leq 2^k$, i.e. the codes $j(i,l)$ of the instances used can be bounded by $t(k):=j(2^k-1,2^k)$. The construction of Φ is clear. □

While the usual classical proof of PCM_{ar} only needs Σ^0_1-induction (but Σ^0_2-DNE), the above proof due to Toftdal needs an instance of the Σ^0_2-induction rule (Σ^0_2-IR) which, apparently, is the price to be paid for using only Σ^0_1-LEM (instead of Σ^0_2-DNE). Classically, Σ^0_2-IR is quite strong and proves (relative to PRA) the same Π^0_3-sentences as Π^0_2-IA (see e.g. [38][Theorem 3.11]) and so, in particular, the totality of the Ackermann function. In our intuitionistic context, however, it is weak and the functional interpretation used (without negative translation!) in the proof of Theorem 2.11 (and the corollary below) to extract B, L solves Σ^0_2-IR using only ordinary primitive recursion in the form of R_0. One can also modify Toftdal's proof so that only the Π^0_1-FAC principle ('Π^0_1-finite-axiom-of-choice') is used (which classical is equivalent to Π^0_1-CP which is relative to PRA Π^0_3-conservative of Σ^0_1-IA): w.l.o.g. we may assume that (a_n) is a sequence of rational

numbers (for, otherwise, replace it by $r_n := \min_{i\le n}(\widehat{a}_i + 2^{-i})$ using the representation of reals from [26]): by Σ^0_1-LEM we have

$$\forall i \le 2^k \exists n \forall m \, \big(a_n < \frac{i}{2^k} \vee a_m \ge \frac{i}{2^k}\big).$$

Hence by Π^0_1-FAC

$$\exists n \forall i \le 2^k \, \big(a_{(n)_i} < \frac{i}{2^k} \vee \forall m (a_m \ge \frac{i}{2^k})\big).$$

Now let $i_0 \le 2^k$ be least s.t. $a_{(n)_{i_0}} < \frac{i_0}{2^k}$ (if existent; otherwise $\forall m(a_m = 1)$ and we are done). Then for $l := (n)_{i_0}$

$$\forall m \ge l \, (|a_m - a_l| < 2^{-k}).$$

As a corollary we obtain that Theorem 2.11 also holds with the original assumption being replaced by $\mathrm{PCM}_{ar}(s(\underline{a}, l), t(\underline{a}))$, where $s(\underline{a}, l)^{0\to 1}$ represents for each $l \in \mathbb{N}$ some sequence of reals defined by a closed term s in $\underline{a}$:

Corollary 2.15. *Given that*

$$\mathrm{HA}^\omega[X, \|\cdot\|] + \mathrm{AC} + \mathrm{M}^\omega + \mathrm{IP}^\omega_\forall \vdash$$
$$\forall \underline{a} \, \exists k^0, l^0 \, \big(\, \forall m \le l \; \mathrm{PCM}_{ar}(s(\underline{a}, m), k) \to \exists n^0 \forall x^0 \, \varphi_0(x, n, \underline{a})\big),$$

where s is a closed term and φ_0 as in Theorem 2.11, then $\exists n^0 \forall x \, \varphi_0(x, n, \underline{a})$ is (valid in $\mathcal{S}^{\omega, X}$) (B, L)-learnable (uniformly in $\underline{a}$) by functionals given by closed terms of the system $\mathsf{HA}^\omega[X, \|\cdot\|]$.
To B, L one can construct majorants B^*, L^* given by closed terms of HA^ω such that if $\exists n \forall x \varphi_0(x, n, \underline{a})$ is monotone (see Definition 2.4) then it is even learnable in B^*, L^* uniformly in majorants $\underline{a}^*$ of the parameters $\underline{a}$.

Remark. Of course, instead of sequences in $[0, 1]$ one can also consider sequences in any compact interval $[-C, C]$, where then the functionals B, L will additionally depend on C.

Likewise, instead of decreasing sequences we may also have increasing ones or, if the Cauchy property is changed into the existence of an approximate infimum

$$\exists n \, \forall m (a_n \le a_m + 2^{-k}),$$

also arbitrary sequences in $[-C, C]$.

The next proposition shows how to convert any majorants (B^*, L^*) for a (B, L)-learnable formula into a rate of metastability. This not only guarantees a highly uniform (and for computable (B^*, L^*)) computable rate of metastability but, moreover, such a rate which has a particularly simple form (see the remark and discussion after the proposition): [4]

[4] Note that in our examples, $\underline{a}$ will be data related to an abstract normed of Hilbert space for which (in contrast to $\underline{a}^*$) computability is not even defined.

Proposition 2.16. *Let $\exists m^0 \forall k^0\, \varphi_0(n,m,k,\underline{a})$ be a formula in the language of HA^ω (or $\mathsf{HA}^\omega[X,\|\cdot\|]$) with φ_0 being quantifier-free that is (B,L)-learnable in the sense of Definition 2.9 (uniformly in n and $\underline{a}$) and let B^*, L^* be majorants of B, L.*
Then a rate of metastability Ω (valid in $\mathcal{S}^{\omega,X}$)

$$\forall n^0\, \forall g^1\, \exists m \leq_0 \Omega(g,n)\ \varphi_0(n,m,g(m),\underline{a}) \tag{metastable}$$

for[5]

$$\forall n^0 \exists m^0 \forall k^0\ \varphi_0(n,m,k,\underline{a}) \tag{φ}$$

is given by $\Omega(B^*,L^*,\underline{a}^*)$ (uniformly in majorants $\underline{a}^*$ of the parameters $\underline{a}$), where $\tilde{g}(c) := \max\big(c, \max_{c'\leq c}(g(c'))\big)$ and

$$\Omega := \lambda B^*, L^*, \underline{a}^*, g, n\ .\ C(L^*,g,n,B^*(n,\underline{a}^*),\underline{a}^*),$$

$$C(i) := C(L^*,g,n,i,\underline{a}^*) := \begin{cases} 0, & \text{if } \ i=0, \\ L^+(\langle \overbrace{\tilde{g}(C(i-1)+1),\ldots,\tilde{g}(C(i-1)+1)}^{i\times}\rangle, n, \underline{a}^*), & \text{otherwise,} \end{cases}$$

with $L^+(x) := \max\{L^*(x), x\}$.
Note that Ω is defined using only recursion R_0 of type 0 and hence is primitive recursive in the usual sense of Kleene.

Proof. We reason in $\mathcal{S}^{\omega,X}$. Since φ_0 is quantifier-free, there is a closed term f with $f(n,m,k,\underline{a}) = 0 \leftrightarrow \varphi_0(n,m,k,\underline{a})$. Hence, we have that (for $\underline{a}^*$ majorizing $\underline{a}$) and for the succession c_i from Definition 2.9

$$\forall n\ \exists i \leq B^*(n,\underline{a}^*)\ \forall k\ \ \big(f(n,c_i,k,\underline{a}) = 0\big),$$

by the assumptions of the proposition, and we need to show that

$$\forall g,n\ \exists m \leq \Omega(B^*,L^*,\underline{a}^*,g,n)\ \ \big(f(n,m,g(m),\underline{a}) = 0\big).$$

Now, fix any g, n and assume towards contradiction that $\Omega(B^*,L^*,\underline{a}^*,g,n)$ is not a rate of metastability, i.e. that

$$\forall m \leq \Omega(B^*,L^*,\underline{a}^*,g,n)\ \ \big(f(n,m,g(m),\underline{a}) \neq 0\big). \tag{1}$$

By induction on i we obtain that

$$\forall i \leq B^*(n,\underline{a}^*)\,\big(c_i \leq C(L^*,g,n,i,\underline{a}^*)\big). \tag{2}$$

The case $i=0$ is trivial as $c_0 = C(L^*,g,n,0,\underline{a}^*) = 0$. Next, suppose that for some $1 \leq i \leq B^*(n,\underline{a}^*)$ the following holds

$$\forall j < i\ \big(c_j \leq C(L^*,g,n,j,\underline{a}^*)\big). \tag{3}$$

[5]Note that in order to talk about metastability, we need one of the parameters to have type 0 and we treat it separately.

Denoting the smallest k s.t. $f(n,m,k,\underline{a}) \neq 0$ by x_m (if this does not exist, we have $c_i = c_{i-1}$ and are done), we obtain by (3) that[6]

$$c_i \leq L^*(\langle x_{c_0}, \ldots, x_{c_{i-1}}\rangle, n, \underline{a}^*) \leq L^*(\langle \tilde{g}(c_0), \ldots, \tilde{g}(c_{i-1})\rangle, n, \underline{a}^*) \leq C(L^*, g, n, i, \underline{a}^*),$$

since by (1) we have (using that $C(i)$ is nondecreasing in i) that

$$\forall i \leq B^*(n,\underline{a}^*)\ \big(m \leq C(L^*,g,n,i,\underline{a}^*) \to f(n,m,g(m),\underline{a}) \neq 0\big)$$

and so, in particular, that

$$\forall i \leq B^*(n,\underline{a}^*)\ \big(m \leq C(L^*,g,n,i,\underline{a}^*) \to x_m \leq g(m) \leq \tilde{g}(m)\big).$$

Finally, we can infer from (2) that

$$\forall i \leq B^*(n,\underline{a}^*)\ \big(c_i \leq \Omega(B^*,L^*,\underline{a}^*,g,n)\big),$$

and therefore and by (1) also

$$\forall i \leq B^*(n,\underline{a}^*)\ \neg\forall k^0\big(f(n,c_i,k,\underline{a}) = 0\big)$$

which is a contradiction.

□

Remark 2.17. Note that Ω has essentially the following form[7]

$$(L_{n,\underline{a}^*} \circ \tilde{g})^{B^*(n,\underline{a}^*)}(0),$$
$$L_{n,\underline{a}^*} := \lambda x\ .\ L^*(x,n,\underline{a}^*).$$

Moreover, if we have such a rate of metastability for some Cauchy statement φ as considered in Remark 2.12 so that φ is monotone and a counterexample x is always greater than the witness candidate, and given any $n,\underline{a}^*$ we have an f^1 and a b^0 such that for all $\underline{a}$ that are majorized by $\underline{a}^*$

$$\forall g \exists m \leq (f \circ \tilde{g})^b(0)\varphi_0(n,m,g(m),\underline{a}), \tag{4}$$

then φ is B,L-learnable (uniformly in n and majorants $\underline{a}^*$ for $\underline{a}$) with

$$B(n,\underline{a}^*) := b, \quad L(x,n,\underline{a}^*) := f(x).$$

To prove this fact, we argue as follows. Fix arbitrary $n,\underline{a}^*$ and consider corresponding f and b. Let

$$g(m) = \min\{x\ : \neg\varphi_0(n,m,x,\underline{a})\},$$

[6]Here we simply assume that our encoding is monotone in its components. If for some reason it was not, we could use a L' which returns the maximal value among all codes coordinatewise bounded by the elements of the encoded input of L^*.

[7]Note that the additional dependency on the number i of iterates in Proposition 2.16 via the length of the sequence $\langle\ldots\rangle$ can also be covered by this normal form, since – by $\tilde{g}(C(i-1)+1) \geq \tilde{g}(i) \geq i$ – the length of the sequence $\langle\ldots\rangle$ can be majorized by $\tilde{g}(C(i-1)+1)$ itself.

if such an x exists and 0 otherwise. Note that due to the monotonicity of φ we have $g(m) \neq 0 \to \tilde{g}(m) = g(m)$. This implies that as long as there is a (the smallest) counterexample x_i to c_i, it holds that

$$c_{i+1} = L(x_i, n, \underline{a}^*) = f(x_i) = (f \circ \tilde{g})c_i = (f \circ \tilde{g})^{i+1}(0). \tag{5}$$

Given all this, assume towards contradiction that

$$\forall i \leq B(n, \underline{a}^*) \exists x \neg \varphi_0(n, c_i, x, \underline{a}), \tag{6}$$

and consider any $m \leq (f \circ \tilde{g})^b(0)$. Due to (5) and (6), we get that $m \leq c_b$ and due to the monotonicity of φ this means that there is a counterexample to m (since there is one for c_b by (6), as $B(n, \underline{a}^*) = b$), which means that $\neg\varphi_0(n, m, g(m), \underline{a})$ by definition of g. This is a contradiction to (4).

Discussion: What the main results in this section (Theorem 2.11 and Proposition 2.16), taken together, show is that if the proof of a (monotone) Π^0_3-statement (e.g. a Cauchy statement) uses only a bounded (in the parameters) number of unnested Σ^0_1-LEM^--instances but may use induction of unrestricted complexity, then we get a rate of metastability which – as a functional in the counterfunction g – has a very simple structure (namely only a single use of iteration of g). This is remarkable as e.g. HA^ω does, of course, prove $\forall g \exists x\, \psi_0(g, x)$-sentences which need much more complicated functionals in g (namely every type-2 functional definable in Gödel's calculus T arrives in this way). What we have shown, however, is that this cannot happen if $\forall g \exists x\, \psi_0(g, x)$ results as the Herbrand normal form of a Π^0_3-statement $\varphi \equiv \forall n^0 \exists m^0 \forall k^0\, \varphi_0(n, m, k)$ (with g being the function variable playing the role of the Herbrand index function) that is provable in HA^ω from the aforementioned restricted uses of Σ^0_1-LEM^-. To see the difference, let us consider the simple but already illuminating case where the proof of φ does not use classical logic at all. Then one can use modified realizability or functional interpretation to extract a (definable in T) witnessing term t for φ (and hence a bound t^* which is uniform in majorants of the parameters) which then a fortiori is also a rate of metastability for φ which does not use the argument g at all. The 'g-involvement' displayed by a rate of metastability reflects the amount of Σ^0_1-LEM^- used in the proof and the former is simple if the latter is low. Moreover, the extraction of the rate of metastability via the extraction of the learning procedure L^* proceeds without any use of negative translation but with direct functional interpretation (while modified realizability would not be sufficient as we need the functional witnessing V in the proof of Theorem 2.11).

A sort of complementary scenario would be to allow full classical logic in a proof but to restrict the use of induction to a bounded number of unnested instances of Σ^0_1-IA, the latter being used e.g. in the form of PCM_{ar}. Let $\mathrm{G_3A}^\omega$ be the finite type extension of Kalmar-elementary arithmetic (based on quantifier-free induction only but with classical logic) from [23] (see also [26]). Now consider a proof

$$\mathrm{G_3A}^\omega[X, \|\cdot\|] \vdash \forall \underline{a} \big(\mathrm{PCM}_{ar}(t_1(\underline{a}), t_2(\underline{a})) \to \exists n^0 \forall x^0\, \varphi_0(x, n, \underline{a}) \big).$$

Then by negative translation and functional interpretation one can extract a rate of metastability for the conclusion making a single use of the rate of metastability of PCM_{ar} given by a single application of the iteration $\tilde{g}^k(0)$ (see [26] prop.2.26) and terms $t[g]$ of

G_3A^ω which can be majorized by terms using only a fixed number of g-nestings (reasoning as in the proof of Proposition 4.13 below). This again leads to a rate of metastability which can be put into the form in Remark 2.17.

So to get a more complicated rate of metastability (e.g. of the form $\tilde{g}^{(\tilde{g}^x(0))}(0)$) requires a nested use of a **combination** of Σ^0_1-LEM$^-$ and (at least) Σ^0_1-IA$^-$ as provided in our example of a sentence φ in Proposition 4.13 that is not effectively learnable (where Σ^0_1-LEM$^-$ together with Π^0_1-CP$^-$ is used).

3. Cauchy statements and unrestricted use of Σ^0_1-LEM

In the following, we refer to Friedman's so-called A-translation from [15] (see e.g. [26]). Since we work in the context of weakly extensional systems and the quantifier-free rule of extensionality QF-ER is not sound under the A-translation we simply add for the reminder of this section all $\mathcal{S}^\omega$-true (resp. $\mathcal{S}^{\omega,X}$-true) purely universal sentences $\mathcal{P}$ in the language of the respective system as axioms (making the use of QF-ER in proofs superfluous as it only proves universal consequences). This, anyhow, is a common device in proof mining as universal axioms do not contribute to the computational content of a proof (this has been stressed by G. Kreisel since the 50's). We denote the extension of the theory $\mathcal{T}$ by the axioms $\mathcal{P}$ by $\mathcal{T}_*$.

Lemma 3.1. *Friedman's A-translation is sound also for* $\mathsf{HA}^\omega_* + \Sigma^0_1$-$\mathsf{LEM}$. *Similarly for* $\mathrm{HA}^\omega_*[X, \|\cdot\|]$ *(and related extensions) instead of* HA^ω_*.

Proof. Consider the following instance of Σ^0_1-LEM

$$\forall y \varphi_0(\underline{a}, y) \vee \exists y \neg \varphi_0(\underline{a}, y).$$

W.l.o.g assume φ_0 is atomic. It suffices to extend Friedman's proof by showing that

$$\mathsf{HA}^\omega_* + \Sigma^0_1\text{-}\mathsf{LEM} \vdash (\Sigma^0_1\text{-}\mathsf{LEM})^A.$$

This means we need to prove

$$\forall y\big(\varphi_0(\underline{a}, y) \vee A\big) \vee \exists y\big((\varphi_0(\underline{a}, y) \vee A) \to A\big), \tag{1}$$

in $\mathsf{HA}^\omega_* + \Sigma^0_1$-$\mathsf{LEM}$.
Suppose that

1. $\forall y \varphi_0(\underline{a}, y)$ holds. Then also $\forall y\big(\varphi_0(\underline{a}, y) \vee A\big)$ holds and therefore also (1).
2. $\exists y \neg \varphi_0(\underline{a}, y)$ holds. Then fix such a y. For this y we get

   $$(\varphi_0(\underline{a}, y) \vee A) \to A$$

   and so $\exists y\big((\varphi_0(\underline{a}, y) \vee A) \to A\big)$ holds and therefore also (1).

For $\mathrm{HA}^\omega_*[X, \|\cdot\|]$ one just has to observe that still every quantifier-free formula can be written as an atomic formula of the form $t\underline{a} =_0 0$ and that the additional axioms are all purely universal and so easily imply their own A-translation. □

For HA instead of HA^ω_* and $\mathrm{HA}^\omega_*[X, \|\cdot\|]$ (also for Σ^0_{n+1}-LEM and Σ^0_{n+2}-DNE), the next proposition is stated (without proof) in [20].

Proposition 3.2. *The theory* $\mathsf{HA}^\omega_* + \Sigma^0_1\text{-}\mathsf{LEM}$ *is closed under the* $\Sigma^0_2\text{-}\mathsf{DNE}$ *rule. Similarly for* $\mathrm{HA}^\omega_*[X, \|\cdot\|]$.

Proof. Suppose

$$\mathsf{HA}^\omega_* + \Sigma^0_1\text{-}\mathsf{LEM} \vdash \neg\neg\exists x\forall y\ \varphi_0(\underline{a}, x, y),$$

where φ_0 is quantifier free and contains only $\underline{a}$ as free variables (in addition to x, y). Moreover, w.l.o.g we assume that φ_0 is atomic.
Rewriting the negations in terms of "$\to$" and "$\bot$" we obtain that

$$\mathsf{HA}^\omega_* + \Sigma^0_1\text{-}\mathsf{LEM} \vdash \big(\ \exists x\forall y\ \varphi_0(\underline{a}, x, y) \to \bot\ \big) \to \bot,$$

and using Friedman's A-translation (with Lemma 3.1) that

$$\mathsf{HA}^\omega_* + \Sigma^0_1\text{-}\mathsf{LEM} \vdash \Big(\ \exists x\forall y\ \big(\varphi_0(\underline{a}, x, y) \vee A\big) \to A\Big) \to A,$$

for any formula A (not containing x, y free). By setting

$$A :\equiv\ \exists x'\forall y'\varphi_0(\underline{a}, x', y'),$$

(we consider only this A throughout the remainder of the proof) we obtain that $\mathsf{HA}^\omega_* + \Sigma^0_1\text{-}\mathsf{LEM}$ proves

$$\big(\ \exists x\forall y\ (\varphi_0(\underline{a}, x, y) \vee \exists x'\forall y'\varphi_0(\underline{a}, x', y')) \to \exists x'\forall y'\varphi_0(\underline{a}, x', y')\big) \to \exists x'\forall y'\varphi_0(\underline{a}, x', y'). \quad (1)$$

Now the claim follows from

$$\Sigma^0_1\text{-}\mathsf{LEM} \vdash \forall y\big(\varphi_0(\underline{a}, x, y) \vee \exists x'\forall y'\varphi_0(\underline{a}, x', y')\big) \to \big(\forall y\ \varphi_0(\underline{a}, x, y) \vee \exists x'\forall y'\varphi_0(\underline{a}, x', y')\big), \quad (2)$$

since using (2) the statement (1) is equivalent to

$$\big((\exists x\forall y\varphi_0(\underline{a}, x, y) \vee \exists x'\forall y'\varphi_0(\underline{a}, x', y')) \to \exists x'\forall y'\varphi_0(\underline{a}, x', y')\big) \to \exists x'\forall y'\varphi_0(\underline{a}, x', y'),$$

which is equivalent to $\exists x\forall y\varphi_0(\underline{a}, x, y)$.
To show (2) consider the following instance of $\Sigma^0_1\text{-}\mathsf{LEM}$

$$\forall y\ \varphi_0(\underline{a}, x, y) \vee \exists y\ \neg\varphi_0(\underline{a}, x, y).$$

Now suppose that

1. $\forall y\ \varphi_0(\underline{a}, x, y)$ holds, then (2) is trivially true.
2. $\exists y\ \neg\varphi_0(\underline{a}, x, y)$ holds, then for such a y we have

   $$\big(\varphi_0(\underline{a}, x, y) \vee \exists x'\forall y'\varphi_0(\underline{a}, x', y')\big) \to \exists x'\forall y'\varphi_0(\underline{a}, x', y')$$

   and so certainly we have also that

   $$\forall y\big(\varphi_0(\underline{a}, x, y) \vee \exists x'\forall y'\varphi_0(\underline{a}, x', y')\big) \to \exists x'\forall y'\varphi_0(\underline{a}, x', y').$$

   Finally $\big(\forall y\ \varphi_0(\underline{a}, x, y) \vee \exists x'\forall y'\varphi_0(\underline{a}, x', y')\big)$ follows from $\exists x'\forall y'\varphi_0(\underline{a}, x', y')$ so (2) holds as well.

□

It is known, that a Cauchy rate is limit computable (which corresponds to Σ^0_2-DNE which – as mentioned in the introduction – is strictly stronger than Σ^0_1-LEM). However, for every provable Cauchy sequence we have that Σ^0_1-LEM is sufficient:

Proposition 3.3. *If a sequence of real numbers (a_n) (or in some PA^ω_*-definable Polish space) defined by a term of PA^ω_*, can be proved to be Cauchy in PA^ω, then the proof can be carried out already in $\mathsf{HA}^\omega_* + \Sigma^0_1\text{-}\mathsf{LEM}$. Similarly for $\mathsf{PA}^\omega_*[X, \|\cdot\|]$ and sequences in X.*

Proof. Consider a sequence $x_{(\cdot)}$ and suppose

$$\mathsf{PA}^\omega_* \vdash \forall k \exists n \forall i, j > n \ \big(|x_i - x_j| \leq 2^{-k}\big).$$

Then by the Kuroda negative translation (see e.g. [26]) we obtain that

$$\mathsf{HA}^\omega_* \vdash \forall k \neg\neg \exists n \forall i, j > n \ \big(|x_i - x_j| \leq 2^{-k}\big).$$

By Proposition 3.2 this implies that

$$\mathsf{HA}^\omega_* + \Sigma^0_1\text{-}\mathsf{LEM} \vdash \forall k \exists n \forall i, j > n \ \big(|x_i - x_j| \leq 2^{-k}\big)$$

(recall that $\leq_{\mathbb{R}} \in \Pi^0_1$). □

4. Which Cauchy statements are effectively learnable and which are not

Proposition 4.1 (Implications between different bounding information for Cauchy statements)**.** *Let (x_n) be a Cauchy sequence in a metric space (X, d).*

1. *A rate of convergence is a bound on the number of fluctuations.*
2. *A bound for the number of fluctuations is a bound B on the number of mind changes to learn a rate of convergence (with a simple projection function as learning procedure L).*
3. *Primitive recursively (in the ordinary sense of Kleene) in majorants B^*, L^* of functionals B, L such that the Cauchy rate is (B, L)-learnable one can obtain a rate of metastability.*

Proof. Consider a Cauchy sequence $x_{(\cdot)}$.

1. Let b be a rate of convergence, i.e.

   $$\forall k \forall n, m \geq b(k) \ \big(d(x_n, x_m) \leq 2^{-k}\big).$$

   Then $b(k)$ is also a bound on the number of 2^{-k} fluctuations, since any fluctuation has to occur before $b(k)$ (i.e. that one of the indexes of the fluctuation has to be smaller than $b(k)$) and there can be at most $b(k)$ many fluctuations indexed within $[0; b(k)]$.

2. Let b be a bound on the number of 2^{-k} fluctuations, i.e.

$$\forall k \forall n > b(k) \forall i, j \neg \mathrm{Fluc}_{2^{-k}}(n, i, j).$$

Then $b(k)$ is also a bound on the number of mind changes to learn a rate of convergence, since for $L(n) := n$ we have that

$$\forall k\ \exists l \leq b(k)\ \forall n, m > c_l\ \ \big(d(x_n, x_m) \leq 2^{-k}\big). \tag{BE}$$

Formally, $L(i, x_{(\cdot)}, k) := i$, and (where – again – to have φ_0 quantifier-free we officially have to use the 2^{-k-1}-rational approximation $\widehat{d(x_n, x_m)}(k+1)$ of $d(x_n, x_m)$)

$$\varphi_0(j(n, m), c_i, x_{(\cdot)}, k) :\equiv\ \big((n > c_i \wedge m > c_i) \to d(x_n, x_m) \leq 2^{-k}\big),$$

where $j(n, m)$ is the Cantor pairing function. The statement (BE) can be inferred from the fact that each mind change (c_i) corresponds to a (different) fluctuation (as it is based on a counterexample for $d(x_n, x_m) \leq 2^{-k}$, whose both indexes are greater than the last c_i).

3. Follows directly from Proposition 2.16.

□

In the rest of this section we show that the hierarchy in Proposition 4.1 between the four different quantitative notations for Cauchy sequences discussed in the introduction is strict. That an effective bound on the number of fluctuations does not imply an effective rate of convergence, follows already from the existence of Specker sequences [39]. We can also use the following very simple example with a 2^{-k}-fluctuation bound k and no effective rate of convergence, since such a rate would decide the halting problem:

Proposition 4.2 $(\alpha_{(\cdot)})$**.** *We take the Cantor pairing function j and set*

$$\alpha_{j(k,n)} := \begin{cases} 2^{-k}, & \text{if } T(k, 0, n), \\ 0, & \text{otherwise}, \end{cases}$$

where T is the primitive recursive Kleene T-predicate. Then (α_n) is a convergent (towards 0) primitive recursive sequence of rationals in $[0, 1]$ with 2^{-k}-fluctuation bound k which has no computable rate of convergence.

We next construct primitive recursive sequence $\beta_{(\cdot)}$ of rational numbers in $[0, 1]$ with an effectively (even primitive recursively) learnable Cauchy rate (so in particular with a primitive recursive rate of metastability), which has no computable bound on fluctuations (this example is not captured by the rough sketch of Avigad and Rute as here the number of the oscillations is determined by the length of the computation, not by the index of the machine as suggested in [9]). Furthermore, we also give an example of a primitive recursive (in the ordinary sense) sequence $\gamma_{(\cdot)}$ of rational numbers in $[0, 1]$ which (provably in the fragment of PA based on Σ^0_1-IA only) converges to 0 (and so has a primitive recursive in the sense of Kleene rate of metastability for the convergence towards 0) which does not have an effectively learnable Cauchy rate.

4.1. A primitive recursive sequence of rationals with a primitive recursively learnable Cauchy rate but with no computable bound on fluctuations

Definition 4.3 ($\beta_{(\cdot)}$)**.** We fix a primitive recursive surjective encoding of triples which is monotone in the third component satisfying $\langle k,n,m\rangle \geq k,n,m$ and set

$$\beta_{\langle k,n,l\rangle} := \begin{cases} 2^{-k}, & \text{if } T(k,0,n) \wedge l \leq n \wedge l \text{ is even,} \\ 0, & \text{otherwise.} \end{cases}$$

In the next propositions we will show that the sequence $\beta_{(\cdot)}$

- is Cauchy (in fact, it converges to zero) – Proposition 4.4,
- its Cauchy rate is effectively learnable – Proposition 4.7,
- there is no computable (in ε and β) bound on the number of ε-fluctuations – Proposition 4.11.

Proposition 4.4. *The sequence $\beta_{(\cdot)}$ is convergent towards 0, provably in $\mathsf{HA}^\omega + \Sigma^0_1$-LEM$^-$. More precisely we show that*

$$\mathsf{HA}^\omega \vdash \forall k\Big(\forall m \leq k\,\big(\exists u\, T(m,0,u) \vee \forall v\, \neg T(m,0,v)\big) \to \exists n \forall x \geq n\big(\beta_x \leq 2^{-k}\big)\Big).$$

Proof. Consider the terminating computations on input 0 of the Turing machines encoded by $0,\dots,k$. Then for every k there is an n corresponding to the code of the longest such computation. W.l.o.g. we can assume that $n \geq k$ (otherwise set $n := k$). This means we have that

$$n \geq k \;\wedge\; \forall n' \forall k' \leq k\,\big(T(k',0,n') \to n' \leq n\big). \tag{7}$$

Now, set

$$c(k) := \max\{\langle k',n',l'\rangle \;:\; n' \leq n,\; k' \leq k,\; l' \leq n'\}.$$

Then c is even a rate of convergence, since

$$\begin{aligned} &\langle k',n',l'\rangle \;>\; c(k) \;\to\; k' > k \vee (k' \leq k \wedge n' > n) \vee (k' \leq k \wedge n' \leq n \wedge l' > n') \\ &\to\; k' > k \vee \beta_{\langle k',n',l'\rangle} = 0 \;\to\; \beta_{\langle k',n',l'\rangle} < 2^{-k}. \end{aligned}$$

These arguments are constructive, except for the existence of the longest computation n. This existence is a consequence of Σ^0_1-LEM^- and Π^0_1-CP^-, where Π^0_1-CP is the bounded collection principle for Π^0_1-formulas (also called $B\Sigma^0_2$ in the literature) which is easily provable by induction in HA$^\omega$. Consider the following $k+1$ instances of Σ^0_1-LEM^-:

$$\forall j \leq k\,\big(\exists n T(j,0,n) \vee \forall m \neg T(j,0,m)\big)$$

which over HA$^\omega$ implies

$$\forall j \leq k \exists n_j\, \forall m\,\big(T(j,0,n_j) \vee \neg T(j,0,m)\big).$$

By an application of Π^0_1-CP$^-$, this in turn implies

$$\exists n \forall j \leq k (\exists n' \leq n T(j,0,n') \vee \forall m \neg T(j,0,m)),$$

(consider $n = \max\{n_j : j \leq k\}$).
This shows that the convergence is provable in $\mathsf{HA}^\omega + \Sigma^0_1$-$\mathsf{LEM}^-$ and the convergence up to an error 2^{-k} in HA^ω uses only $k+1$ instances of Σ^0_1-LEM^-. □

Lemma 4.5. *For a quantifier-free formula φ_0 with parameters only of type 0, we have that*

$$\mathrm{G_3A}^\omega + \Sigma^0_1\text{-}\mathsf{IA}^- \vdash \forall x^0\ \exists u^0 \forall \tilde{x} \le x \big(\forall y^0 \varphi_0(\tilde{x}, y) \vee \exists \tilde{u} \le u \neg\varphi_0(\tilde{x}, \tilde{u})\big).$$

Proof. See [26] Lemma 3.18. □

Remark 4.6. Σ^0_1-IA$^-$ is (over $\mathrm{G_3A}^\omega$) strictly weaker than Π^0_1-CP$^-$ but the proof in Lemma 4.5 needs Σ^0_2-DNE and so more of classical logic than necessary in the proof based on Π^0_1-CP$^-$. In general, it seems that considering the extraction of computational content from proofs, often some amount of classical logic can be reduced on the cost of more recursion.
If one is interested (only) in a classical proof, we obtain due to Lemma 4.5 (simply consider $\varphi_0(x, y) :\equiv \neg T(x, 0, y)$) a proof of the convergence of $\beta_{(\cdot)}$ without the use of Π^0_1-CP, which can be formalized in $\mathrm{G_3A}^\omega + \Sigma^0_1$-$\mathsf{IA}^-$.

Proposition 4.7. *The rate of convergence is effectively learnable in k, i.e. there are total (elementary) recursive functions B and L, s.t. for any k we have that*

$$\forall k\ \exists n \le B(k)\ \forall m > c_n \quad \big(\beta_m \le 2^{-k}\big),$$

where $c_{(\cdot)}$ is defined as in Definition 2.4 with

$$\varphi_0(x, n, k) :\equiv\ x > n \to \beta_x \le 2^{-k}.$$

Proof. Obviously, this follows already from Proposition 4.4. Also, it is easy to see that the rate is (B, L)-learnable with the following B and L:

$$B(k) := k + 1$$
$$L(n, k) := \langle k, n, n\rangle + 1.$$

Let $x \ge L(n, k) > \langle k, n, n\rangle$ be a counterexample. Then (using the definition of $\beta_{(\cdot)}$)

$$j_1(x) \le k \wedge (j_2(x) > n \vee j_3(x) > n) \wedge j_3(x) \le j_2(x)$$

and so

$$j_1(x) \le k \wedge j_2(x) > n \wedge j_3(x) \le j_2(x).$$

The 2nd conjunct implies $j_2(x) > j_2(n)$ and hence $j_1(x) \neq j_1(n)$ if n is a preceding counterexample. However, for numbers $\le k$ this can happen at most k-many times. Hence $B(k) := k + 1$ and L do the job. □

Corollary 4.8. *$\beta_{(\cdot)}$ has a primitive recursive (in the ordinary sense of Kleene) rate of metastability for the convergence towards* 0.

Proof. One can apply Proposition 2.16 to convert the bounds (B, L) from Proposition 4.7 (which are trivially self-majorizing using standard monotonicity properties of the triple coding) into a primitive recursive rate of metastability. Alternatively, one can use that by Lemma 4.5 the convergence of $\beta_{(\cdot)}$ towards 0 is provably in $\mathrm{G_3A}^\omega + \Sigma^0_1$-IA$^-$ and so a fortiori in $\widehat{\mathrm{PA}}^\omega|$+QF-AC (see [26], Prop.3.31). Then proposition 10.54 in [26]

implies the extractability of a primitive recursive rate of metastability for the convergence towards 0 (the latter being essentially the rate of metastability for the statement in Lemma 4.5 which is computed in [26][Prop.3.19]). □

Remark 4.9. One can obtain such a rate of metastability for $\beta_{(\cdot)}$ directly, using previous results of the first author.
By [26] (Prop.13.19) we have that

$$\forall x, f \forall \tilde{x} \leq x \big(\exists y \leq \Phi x f \; T(\tilde{x}, 0, y) \vee \forall z \leq f(\Phi x f) \neg T(\tilde{x}, 0, z)\big), \tag{8}$$

where $\Phi x f \leq \max\{f^i(0) \; : \; i \leq x+1\} =: \Phi^* x f$ (here $f^i(0)$ again denotes the i-times iteration of f). (8) implies

$$\forall z \big(\Phi x f < z \leq f(\Phi x f) \to \forall \tilde{x} \leq x \neg T(\tilde{x}, 0, z)\big). \tag{9}$$

Define $\tilde{f}(n) := \max\{f(n), n\}$ and $f_k(n) := \tilde{f}(\langle k, n, n\rangle + 1)$ and let

$$\Psi(k, f) := \langle k, \Phi^* k f_k, \Phi^* k f_k\rangle + 1.$$

Then Ψ is a rate of metastability for the convergence of $\beta_{(\cdot)}$ towards 0, i.e.:

$$\forall k, f \; \exists n \leq \Psi(k, f) \; \forall z \in [n, \tilde{f}(n)] \;\; (|\beta_z| < 2^{-k}). \tag{10}$$

To prove (10), define $n := \langle k, \Phi k f_k, \Phi k f_k\rangle + 1 \leq \Psi(k, f)$ and let $z \in [n, \tilde{f}(n)]$. Then $z \geq n > \langle k, \Phi k f_k, \Phi k f_k\rangle$ and so one of the following cases holds:

1. $j_1(z) > k$. Then $|\beta_z| \leq 2^{-j_1(z)} < 2^{-k}$.
2. $j_2(z) > \Phi k f_k \wedge j_1(z) \leq k$. Then $\Phi k f_k < j_2(z) \leq z \leq \tilde{f}(n) = f_k(\Phi k f_k)$. Hence, by (9) (applied to $k, j_1(z), f_k, j_2(z)$ for $x, \tilde{x}, f, z$), we get $\neg T(j_1(z), 0, j_2(z))$ and so $\beta_z = 0$.
3. $j_3(z) > \Phi k f_k \wedge j_2(z) \leq \Phi k f_k \wedge j_1(z) \leq k$. Then $j_3(z) > j_2(z)$ and so $\beta_z = 0$.

Lemma 4.10 (Termination causes at least n fluctuations)**.** *Suppose the k^{th}-machine terminates on 0 with computation encoded by n (i.e. $T(k, 0, n)$ holds). Then the sequence $\beta_{(\cdot)}$ contains at least n many 2^{-k}-fluctuations.*

Proof. Consider the tuples of indexes $\underline{i}$, $\underline{j}$, s.t. $i_l := \langle k, n, l\rangle$, $j_l := \langle k, n, l+1\rangle$ and $l + 1 \leq n$. Then we have by definition of $\beta_{(\cdot)}$ (using the monotonicity of the encoding in l) that $\mathrm{Fluc}_{\beta_{(\cdot)}, 2^{-k}}(n, \langle \underline{i}\rangle, \langle \underline{j}\rangle)$. □

Proposition 4.11. *There is no computable bound on the fluctuations of $\beta_{(\cdot)}$.*

Proof. Suppose b_k is a bound on the number of fluctuations by 2^{-k}, then b_k can be used to effectively compute whether the k^{th} Turing machine terminates on input 0 as follows. Let the machine run until the code of the computation reaches b_k (or until it stops). If it terminated, we are done.
Now suppose it terminates with a computation encoded by some $n > b_k$. Then $\beta_{(\cdot)}$ would have at least n many 2^{-k}-fluctuations by Lemma 4.10, which is a contradiction.
Therefore if the machine does not terminate with a code of computation at most b_k it does not terminate at all. □

4.2. Metastability of Cauchy sequences does not imply effective learnability

In the next propositions we define a primitive recursive sequence $\gamma_{(\cdot)}$ of rational numbers in $[0,1]$ (defined in Corollary 4.20 using Definition 4.16) that

- is Cauchy (in fact, it converges to zero) – by Proposition 4.17,
- has a primitive recursive rate of metastability of its convergence towards 0 – by Proposition 4.17,
- has no effectively learnable Cauchy rate – by Corollary 4.20.

We use the upper index as a name extension (like $\underline{k}^K$, meaning a tuple $\underline{k}$ corresponding to a particular K) and as iteration of functions (like $f^n(x)$, meaning we iterate the function f n-many times with the starting point x). When unclear, we use the notation $(f)^n$ to make explicit, that we mean the iteration.

Definition 4.12.

1. For any function $f : \mathbb{N}^2 \to \mathbb{N}$ we define

   $$\widehat{f}(k,n) := 0, \text{ if } f(k,n) = 0 \wedge \forall m < n\ (f(k,m) \neq 0) \text{ and } \widehat{f}(k,n) := 1, \text{ otherwise.}$$

   Note that $\widehat{f}(k,\cdot)$ has a root iff $f(k,\cdot)$ has one but also that it has at most one root.

2. For any f as above and $x, y \in \mathbb{N}$ define

   $$y_{f,x} := \begin{cases} \max\big\{y' \leq y\ :\ \exists x' \leq x\,(y' = \min\{y'' \leq y\ :\ f(x',y'') = 0\ \})\big\}, & \text{if such } y' \text{ exists} \\ x, & \text{otherwise.} \end{cases}$$

   Then for $p = j(y,u)$ define

   $$\varphi_1(f,x,p,z) :\equiv \forall \tilde{x} \leq x \exists \tilde{y} \leq y \forall \tilde{z} \leq z\big(f(\tilde{x},\tilde{y}) = 0 \vee f(\tilde{x},\tilde{z}) \neq 0\big)$$

   and

   $$\varphi_2(f,x,p,z) :\equiv \forall \tilde{y} \leq y_{f,x} \exists \tilde{u} \leq u \forall \tilde{z} \leq z\big(f(\tilde{y},\tilde{u}) = 0 \vee f(\tilde{y},\tilde{z}) \neq 0\big)$$

   and, finally,[8]

   $$\varphi_0 :\equiv \varphi_1 \wedge \varphi_2, \quad \varphi :\equiv \forall f \leq_1 \forall x^0 \exists p \forall z \varphi_0(\widehat{f},x,p,z)$$

Note that the φ_1-part of φ combines a sequence of instances of Σ^0_1-LEM with induction (in the form of Π^0_1-CP) and that the φ_2-part repeats this construction but taking (essentially) the result 'y' from φ_1 as input thereby making it no longer possible to give a computable bound on the number of instances of Σ^0_1-LEM used in total. We will show in the next proposition that φ is not (B,L)-learnable by showing that it implies over a system as weak as $\mathrm{G_3A}^\omega$ (which does not allow for the iteration of a function variable g)

$$\forall g^1 \exists y^0\,(y = g^{g^x(0)}(0))$$

[8] In connection with f we write the type 1 even though it officially is the type $0 \to (0 \to 0)$.

which grows too fast as a functional in g to be derivable from a rate of metastability for φ having the simple form from Remark 2.17 whose existence would follow from the (B, L)-learnability of φ. Here it is crucially used that computable functionals B, L in the function parameter f which can be taken to be bounded by 1 can be effectively majorized by bounds which no longer depend on f.

Proposition 4.13. *φ is provable using Σ^0_1-LEM$^-$ combined with Π^0_1-CP$^-$ (uniformly in f treated as parameter) but is not effectively learnable.*

Proof. We first show that φ is provable: by a suitable instance Σ^0_1-LEM$(t(f))$ of Σ^0_1-LEM$^-$ one obtains

$$\forall x \forall \tilde{x} \leq x \exists y \forall z \big(\exists \tilde{y} \leq y\, \widehat{f}(\tilde{x}, \tilde{y}) = 0 \vee \forall \tilde{z} \leq z\, \widehat{f}(\tilde{x}, \tilde{z}) \neq 0\big)$$

and so by a suitable instance Π^0_1-CP$(s(f))$

$$\forall x \exists y \forall \tilde{x} \leq x \forall z \big(\exists \tilde{y} \leq y\, \widehat{f}(\tilde{x}, \tilde{y}) = 0 \vee \forall \tilde{z} \leq z\, \widehat{f}(\tilde{x}, \tilde{z}) \neq 0\big)$$

which implies

$$\forall x \exists y \forall z \forall \tilde{x} \leq x \exists \tilde{y} \leq y \forall \tilde{z} \leq z\, \big(\widehat{f}(\tilde{x}, \tilde{y}) = 0 \vee \widehat{f}(\tilde{x}, \tilde{z}) \neq 0\big).$$

Now we repeat the same argument with $y_{f,x}$ instead of x.

To show that φ is not learnable we proceed in three steps.

Step 1. We will show that (informally speaking, since formally we cannot express function iteration and the conclusion would have to use ψ_0 and ψ_0^y defined below)

$$\mathrm{G_3A}^\omega \vdash \forall g \forall x\ \big(\exists p \forall z \varphi_0(f_g, x, p, z) \to \exists y'(y' = g^{g^x(0)}(0))\big), \tag{GA}$$

for

$$f_g(b, d) := \begin{cases} 0, & \text{if } \ \mathrm{lh}(d) = b + 1 \wedge d_0 = 0 \wedge \forall i < b\ (d_{i+1} = g(d_i), \\ 1, & \text{otherwise.} \end{cases}$$

Formally, $\exists y'(y' = g^{g^x(0)}(0))$ is to be read as

$$\exists y, u\, (f_g(x, y) = 0 \wedge f_g(y_x, u) = 0), \tag{GA*}$$

where then $y' := u_{y_x}$. Note that we have $f_g =_1 \widehat{f_g}$. To show (GA) fix arbitrary g^1 and x^0. Now assume $\exists p \forall z \varphi_0(f_g, x, p, z)$ and let us fix such a $p = j(y, u)$ to obtain:

$$\forall z \forall \tilde{x} \leq x \exists \tilde{y} \leq y \forall \tilde{z} \leq z \big(f_g(\tilde{x}, \tilde{y}) = 0 \vee f_g(\tilde{x}, \tilde{z}) \neq 0\big) \wedge \tag{11}$$

$$\forall z \forall \tilde{y} \leq y_{f_g, x} \exists \tilde{u} \leq u \forall \tilde{z} \leq z \big(f_g(\tilde{y}, \tilde{u}) = 0 \vee f_g(\tilde{y}, \tilde{z}) \neq 0\big). \tag{12}$$

We now show by quantifier-free induction that

$$\forall x' \leq x \exists y' \leq y\ f_g(x', y') = 0. \tag{13}$$

Note that $\exists y'\, f_g(x', y') = 0$ implies that $y'_{x'} = g^{x'}(0)$. The case $x' = 0$ is trivially satisfied by $y' = \langle 0 \rangle$. Then $y' \leq y$ by (11). So suppose for some $x' < x$ we have $\exists y' \leq y\, f_g(x', y') = 0$. Then we can set

$$z := y' * \langle g(y'_{x'}) \rangle = \langle y'_0, y'_1, \ldots, g(y'_{x'}) \rangle$$

to get $f_g(x'+1,z)=0$ which concludes the proof of (13) since – again by (11) – $z \leq y$. Note, furthermore, that for $y' \leq y$ (and $x' \leq x$),

$$f_g(x',y')=0 \;\rightarrow\; y' \leq y_{f,x}.$$

So we have even that

$$\forall x' \leq x \exists y' \leq y_{f_g,x}\, f_g(x',y')=0. \tag{14}$$

Now, let y^* denote the y' which satisfies (14) for $x'=x$ and note that $y^*_x \leq y^* \leq y_{f_g,x}$. By quantifier-free induction we show that

$$\forall x' \leq y_{f_g,x} \exists u' \leq u\, f_g(x',u')=0. \tag{15}$$

The case $x'=0$ is again trivially satisfied by $u'=\langle 0\rangle \leq u$ (using (12)). So suppose for some $x' < y_{f_g,x}$ that $\exists u' \leq u\, f_g(x',u')=0$. Then we can set

$$z := u' * \langle g(u'_{x'})\rangle = \langle u'_0, u'_1, \ldots, g(u'_{x'})$$

to get $f_g(x'+1,z)=0$, which by (12) implies

$$\exists \tilde{u} \leq u\, f_g(x'+1,\tilde{u})=0$$

and so concludes the proof of (15). Applying (15) we obtain (for $x'=y^*_x$)

$$\exists u' \leq u\, (f_g(x,y^*)=0 \wedge f_g(y^*_x,u')=0),$$

which concludes the proof of (GA^*) and, therefore, also the proof of (GA).

Step 2. We investigate the terms witnessing the implication (GA). By prenexation we obtain

$$\mathrm{G_3A}^\omega \vdash \forall g,x,p \exists y',z\, \big(\varphi_0(f_g,x,p,z) \rightarrow y' = g^{g^x(0)}(0)\big),$$

and, therefore, by program extraction theorems (see Corollary 3.1.3 in [23]), we get closed terms s and t in $\mathrm{G_3A}^\omega$, s.t.

$$\forall g,x,p\, \big(\varphi_0(f_g,x,p,sgxp) \rightarrow tgxp = g^{g^x(0)}(0)\big). \tag{16}$$

Step 3. Finally, we show that with sufficiently large (in the sense of growth) g, this contradicts the effective learnability of φ. Suppose namely that φ were learnable by computable functionals $B(f,x), L(y,f,x)$. Then also

$$\begin{aligned} &B^*(x) := \sup\{B(f,\tilde{x}) : f \leq_1 1, \tilde{x} \leq x\} \text{ and} \\ &L^*_x(y) := \max\{y, \sup\{L(\tilde{y},f,\tilde{x}) : f \leq_1 1, \tilde{y} \leq y, \tilde{x} \leq x\}\} \end{aligned}$$

are computable (in x resp. in x,y) and majorize B,L. Now by Proposition 2.16 and the fact that f_g is trivially majorized by 1 we get, in particular, that

$$\exists p \leq \Omega(B^*, L^*_x, h_x, x)\; \varphi_0(f_g,x,p,sgxp),$$

for any g and x, by setting

$$h^1_x := \lambda p \,.\, sgxp.$$

So, we obtain together with (16) that

$$\forall g, x \exists p_x \leq \Omega(B^*, L^*_x, h_x, x) \, \big(tgxp_x = g^{g^x(0)}(0)\big).$$

Since s and t are closed terms of $\mathrm{G_3A}^\omega$ and the variables g, x, p_x have types ≤ 1 by normalization arguments (see Corollary 2.2.24 and Remark 2.2.25 in [23]) we know that there is a constant D, s.t. (for any g that majorizes $\lambda n.2^n$)

$$\forall x, v \, \big(\tilde{g}^D(x+v) \geq sgxv, tgxv\big).$$

Since we may assume that $\tilde{g}(n) > n$ and, therefore, $\tilde{g}^x(v) \geq x + v$, this yields

$$\forall x, v \, \big(\tilde{g}^{D+x}(v) \geq sgxv, tgxv\big).$$

As a consequence, we get (using the Ω-definition and that that B^*, L^*_x are selfmajorizing and that $L^*_x(y) \geq y$) that for all x

$$\tilde{g}^{D+x}\left(\Omega(B^*, L^*_x, \tilde{g}^{D+x}, x)\right) \geq \tilde{g}^{D+x}\left(\Omega(B^*, L^*_x, h_x, x)\right) \geq tgxp_x = g^{g^x(0)}(0).$$

By the definition of Ω (see also Remark 2.17)

$$\begin{array}{l} \tilde{g}^{D+x}\big(\Omega(B^*, L^*_x, \tilde{g}^{D+x}, x\big) \leq \tilde{g}^{D+x}\left((L^*_x \circ \tilde{g}^{D+x})^{B^*(x)}(0)\right) \leq \\ \tilde{g}^{D+x}\left((L^*_x \circ \tilde{g})^{(D+x)B^*(x)}(0)\right) \leq (L^*_x \circ \tilde{g})^{\widehat{B^*}(x)}(0) \end{array}$$

and so (for all x)

$$g^{g^x(0)}(0) \leq (L^*_x \circ \tilde{g})^{\widehat{B^*}(x)}(0)$$

where $\lambda x, y.L^*_x(y)$ and $\widehat{B^*}(x) := (D + x)(B^*(x) + 1)$ are fixed total recursive functions that do not depend on g which is not possible for sufficiently fast growing g. □

Corollary 4.14. *Let φ_0 be as in the previous Proposition. If for a quantifier free formula ψ_0 (with no hidden parameters)*

$$\mathrm{G_3A}^\omega + \mathsf{QF\text{-}AC} \vdash \forall f \leq 1 \forall x^0 \big(\exists y^0 \forall z^0 \psi_0(\xi(f), \chi(x), y, z) \to \exists p^0 \forall z^0 \varphi_0(\widehat{f}, x, p, z)\big),$$

where ξ and χ are closed terms of $\mathrm{G_3A}^\omega$, then $\forall f \leq 1 \forall x^0 \exists y^0 \forall z^0 \psi_0(f, x, y, z)$ is also not effectively learnable. Here

$$\text{QF-AC}: \;\; \forall x \, \exists y \, F_0(x, y) \to \exists f \, \forall x \, F_0(x, f(x)),$$

with quantifier-free F_0 and x, y of arbitrary types.

Proof. This follows analogously from [23] (Corollary 3.1.3.) and our Proposition 2.16 as in the proof of Proposition 4.13. □

Remark 4.15. We can prove in $\mathrm{G_3A}^\omega + \Sigma^0_1\text{-}\mathsf{CP}$ (which is included in $\mathrm{G_3A}^\omega + \mathsf{QF\text{-}AC}^{0,0}$) that φ in Proposition 4.13 is actually equivalent to its monotone version,

$$\tilde{\varphi} \equiv \forall f^1 \leq 1 \forall x^0 \exists q \forall z \exists p \leq q \forall \tilde{z} \leq z \varphi_0(\widehat{f}, x, p, z).$$

Since this equivalence holds also pointwise (in f, x), we can use Corollary 4.14 to infer that there is actually a monotone formula, which is not effectively learnable.

Definition 4.16. Define (using a surjective quadruple coding) a primitive recursive sequence of rational numbers in $[0, 1]$ by

$$\gamma(f)_{\langle k,n,i,m\rangle} := \begin{cases} 2^{-k}, & \text{if } \widehat{f}(k,n) = 0 \wedge i \leq n \wedge \widehat{f}(i,m) = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Proposition 4.17. *The sequence $(\gamma(f)_z)_{z\in\mathbb{N}}$ converges to 0 but the formula stating the existence of a Cauchy point for any f*

$$\psi := \forall f^1 \leq_1 1, x^0 \exists z \forall k \geq z \ \big(\gamma(f)_k < 2^{-x}\big).$$

is not effectively learnable. However, there is a primitive recursive (in the ordinary sense of Kleene) rate of metastability for the convergence of $\gamma(f)_{(\cdot)}$ towards 0, which does not depend on f.

Proof. The existence of a metastability rate follows from the fact that $\gamma(f)_{(\cdot)}$ converges to 0 for any $f \leq_1 1$. Moreover, since the proof can be formalized in $\mathrm{G_3A}^\omega + \Sigma^0_1\text{-}\mathsf{IA}$ there is a primitive recursive rate and since f is trivially majorizable it is also clear that there is even a primitive recursive rate which does not depend on f. (In Remark 4.18 below, we actually give such a rate explicitly.)
To show the unlearnability, due to Corollary 4.14 it suffices to show that

$$\mathrm{G_3A}^\omega + \mathsf{QF\text{-}AC}^{0,0} \vdash \forall f \leq 1 \forall x^0 \big(\exists z \forall k \geq z \ \big(\gamma(f)_k < 2^{-x}\big) \to \exists p \forall z' \varphi_0(\widehat{f}, x, p, z')\big).$$

To prove φ, fix arbitrary f^1 and x^0 and suppose that

$$\exists z \forall k \geq z \big(\gamma(f)_k < 2^{-x}\big).$$

Moreover, assume towards contradiction

$$\exists \tilde{x} \leq x \exists a \geq \max(z, x) \widehat{f}(\tilde{x}, a) = 0. \tag{17}$$

Since $a \geq \tilde{x}, z$, this implies that $k := \langle \tilde{x}, a, \tilde{x}, a\rangle \geq z$ and $\gamma(f)_k = 2^{-\tilde{x}}$, which is a contradiction.
Hence we can conclude that

$$\forall \tilde{x} \leq x \big(\exists \tilde{y} < \max(x, z) \widehat{f}(\tilde{x}, \tilde{y}) = 0 \ \vee \ \forall a \widehat{f}(\tilde{x}, a) \neq 0\big),$$

which is equivalent to

$$\forall \tilde{x} \leq x \forall a \exists \tilde{y} < \max(x, z) \ \big(\widehat{f}(\tilde{x}, \tilde{y}) = 0 \ \vee \ \forall \tilde{a} \leq a \, \widehat{f}(\tilde{x}, \tilde{a}) \neq 0\big). \tag{18}$$

Next, set $y := \max(x, z)$ and assume towards contradiction that

$$\exists \tilde{y} \leq y_{\widehat{f},x} \exists a \geq z \widehat{f}(\tilde{y}, a) = 0. \tag{19}$$

Recall that

$$y_{\widehat{f},x} := \begin{cases} \max\big\{y' \leq y \ : \ \exists x' \leq x \, (y' = \min\{y'' \leq y \ : \ \widehat{f}(x', y'') = 0\})\big\}, & \text{if such } y' \text{ exists} \\ x, & \text{otherwise.} \end{cases} \tag{20}$$

Note that if $y_{\widehat{f},x} = x$, then φ follows already from (18). Otherwise, denote the smallest $x' \leq x$ for which $\widehat{f}(x', y_{\widehat{f},x}) = 0$ by $\tilde{x}$. Then $k := \langle \tilde{x}, y_{\widehat{f},x}, \tilde{y}, a\rangle \geq z$ and $\gamma(f)_k = 2^{-\tilde{x}}$, which is a contradiction.
Finally, $\exists p \forall z' \varphi_0(\widehat{f}, x, p, z')$ follows from not-(19) and (18) (with $y := \max(x, z)$, $p := \langle y, z\rangle$).

□

Remark 4.18. Similarly as before, we can give an explicit such rate of metastability for the convergence of $(\gamma(f)_z)$ toward 0. As in Remark 4.9, there is a $\Phi_f xg \leq \Phi^* xg := \max\{g^i(0) : i \leq x + 1\}$ such that

$$\forall x, f, g \forall \tilde{x} \leq x \big(\exists y \leq \Phi_f xg \, (\widehat{f}(\tilde{x}, y) = 0) \vee \forall z \leq g(\Phi_f xg) \, (\widehat{f}(\tilde{x}, z) \neq 0)\big).$$

Define

$$\begin{aligned} \Phi_1 xg &:= \Phi_f\Big(x, \lambda y. g_y\big(\Phi_f(y, g_x)\big)\Big), \\ \Phi_2 xg &:= \Phi_f\big(\Phi_1 xg, g_{\Phi_1 xg}\big), \end{aligned}$$

where $g_y(n) := g(y, n)$. Then

$$\forall x, f, g \ \forall \tilde{x} \leq x \forall \tilde{y} \leq \Phi_1 xg \left\{ \begin{array}{l} \Big((\exists y \leq \Phi_1 xg\big(\widehat{f}(\tilde{x}, y) = 0\big) \vee \forall z \leq g(\Phi_1 xg, \Phi_2 xg) \ \big(\widehat{f}(\tilde{x}, z) \neq 0\big)) \wedge \\ \ (\exists u \leq \Phi_2 xg \, (\widehat{f}(\tilde{y}, u) = 0) \vee \forall z \leq g(\Phi_1 xg, \Phi_2 xg) \, (\widehat{f}(\tilde{y}, z) \neq 0))\Big). \end{array} \right.$$

This implies

$$\forall z\big(\Phi_1 xg < z \leq g(\Phi_1 xg, \Phi_2 xg) \to \forall \tilde{x} \leq x \ (\widehat{f}(\tilde{x}, z) \neq 0)\big) \tag{21}$$

and

$$\forall z\big(\Phi_2 xg < z \leq g(\Phi_1 xg, \Phi_2 xg) \to \forall \tilde{x} \leq \Phi_1 xg \ (\widehat{f}(\tilde{x}, z) \neq 0)\big). \tag{22}$$

Define $\tilde{g}(n) := \max\{g(n), n\}$ and $g_k(n, m) := \tilde{g}(\langle k, n, n, m\rangle + 1)$ and

$$\Psi(k, g) := \langle k, \Phi_1^* k g_k, \Phi_1^* k g_k, \Phi_2^* k g_k\rangle + 1,$$

where Φ_i^* is defined as Φ_i but with Φ^* and $(g_k)^M$ instead of Φ_f and g_k. To show

$$\forall k, g, f \ \exists n \leq \Psi(k, g) \ \forall z \in [n, \tilde{g}(n)] \ \ (|\gamma(f)_z| < 2^{-k}),$$

define $n := \langle k, \Phi_1 k g_k, \Phi_1 k g_k, \Phi_2 k g_k\rangle + 1 \leq \Psi(k, g)$ and let $z \in [n, \tilde{g}(n)]$. Then $z \geq n > \langle k, \Phi_1 k g_k, \Phi_1 k g_k, \Phi_2 k g_k\rangle$. Hence one of the following holds:

1. $j_1(z) > k$. Then $|\gamma(f)_z| \leq 2^{-j_1(z)} < 2^{-k}$.
2. $j_2(z) > \Phi_1 k g_k \wedge j_1(z) \leq k$. Then $\Phi_1 k g_k < j_2(z) \leq z \leq \tilde{g}(n) = g_k(\Phi_1 k g_k, \Phi_2 k g_k)$. Hence, by (21) (applied to $j_1(z), k, g_k, j_2(z)$ for $\tilde{x}, x, g, z$), $\gamma(f)_z = 0$.
3. $j_3(z) > \Phi_1 k g_k \wedge j_2(z) \leq \Phi_1 k g_k \wedge j_1(z) \leq k$. Then $j_3(z) > j_2(z)$ and so $\gamma(f)_z = 0$.
4. $j_4(z) > \Phi_2 k g_k \wedge j_3(z) \leq \Phi_1 k g_k \wedge j_2(z) \leq \Phi_1 k g_k \wedge j_1(z) \leq k$. Then $\Phi_2 k g_k < j_4(z) \leq z \leq \tilde{g}(n) = g_k(\Phi_1 k g_k, \Phi_2 k g_k)$. Hence, by (22) (applied to $j_3(z), k, g_k, j_4(z)$ for $\tilde{y}, x, g, z$), $\gamma(f)_z = 0$.

Note that the 2-nested use of primitive recursive iteration hidden in the 2-nested application of Φ_f in the definition of Φ_i very much resembles the basic structure of the rate of metastability extracted from a concrete proof in ergodic theory in [37] (see the discussion in the introduction).

Corollary 4.19. *Let $\gamma(e)_{(\cdot)}$ be defined as $\gamma(f)_{(\cdot)}$ but with $\widehat{f}(x,y) = 0$ being replaced by*

$$P(e,x,y) :\equiv \mathrm{lh}(y) = x+1 \wedge y_0 = 0 \wedge \forall i < x(T(e, y_i, y_{i+1})).$$

Then $(\gamma(e)_n)$ is a sequence of rational numbers in $[0,1]$ that converges to 0 but the formula stating the existence of a Cauchy point of $\gamma(e)_{(\cdot)}$ for any number e

$$\psi := \forall e^0, x^0 \exists z \forall k \geq z \ \big(\gamma(e)_k < 2^{-x}\big).$$

is not effectively learnable.

Proof. First note that the convergence (towards 0) of $\gamma(e)_n$ follows from this property of $\gamma(f)_n$ since taking $f(x,y) := 0$, if $P(e,x,y)$, and $f(x,y) := 1$, otherwise, both sequences coincide (note that $f = \widehat{f}$).
Let e by a code of a total recursive function and define $g(x) := \mu y \, . \, T(e,x,y)$. The arguments of both Proposition 4.13 and Proposition 4.17 then remain valid, except the fact that the set of Gödel numbers e – in contrast to f_g – is not majorizable. We also need the additional assumption $\forall x \, (T(e,x,g(x)))$ in (GA), which expresses that $g(x) := \mu y \, . \, T(e,x,y)]$ defines a total function. This assumption, however, does not contribute in the course of the functional interpretation argument applied to (GA) ('Step 2' in the proof of Proposition 4.13) as it is purely universal. So as before, the learnability would lead to a constant D and total recursive functions $\lambda e, x, y . L^*_{e,x}(y)$, B^* such that

$$\forall e \, \forall g \, \Big(\forall x^0 \, T(e,x,g(x)) \to \forall x^0 \ (L^*_{e,x} \circ \tilde{g})^{\widehat{B^*}(e,x)}(0) \geq g^{g^x(0)}(0)) \Big)$$

where $\widehat{B^*}(e,x) := (D + x + e)(B^*(x) + 1)$.
We can argue similarly as in the proof of Proposition 4.13, since for any fixed g given by some e as above, eventually we have $x > e$, so we can simply choose a total recursive g which grows much faster than $L^*_{x,x}(x)$ and $B^*(x,x)$. □

Corollary 4.20. *Define the primitive recursive sequence of rational numbers in $[0,1]$ (using the Cantor pairing function)*

$$\gamma_{j(e,n)} := \gamma(e)_n \cdot 2^{-e}.$$

This sequence converges to 0 *(provably in* $\mathrm{G_3A}^\omega + \Sigma^0_1\text{-IA}^-$ *and hence with a primitive recursive – in the sense of Kleene – rate of metastability) but the formula stating the existence of a Cauchy point of* $\gamma_{(\cdot)}$

$$\psi := \forall x^0 \exists z \forall k \geq z \, \big(\gamma_k < 2^{-x}\big).$$

is not effectively learnable.

Proof. Let $\rho_e(x)$ be a rate of convergence for $(\gamma(e)_n)_n$ and define $\rho(x) := \max\{\rho_{\tilde{x}}(x) : \tilde{x} \leq x\}$. Then

$$\widehat{\rho}(x) := j(x, \rho(x))$$

is a rate of convergence for (γ_n) towards 0.
Conversely, for any rate ρ of convergence for (γ_n) we have that $\rho_e(x) := \rho(e+x)$ is a rate of convergence of $(\gamma(e)_n)_n$. In particular, the 2^{-e-x}-Cauchy property for (γ_n) implies the 2^{-x}-Cauchy property of $(\gamma(e)_n)_n$. Hence by Corollary 4.19, (γ_n) does not have an effectively learnable Cauchy rate. □

5. When learnability implies fluctuation bounds

In some cases, the effective (B, L)-learnability of a convergence rate (meaning that the convergence rate can be learned by L with $B(\underline{a})$-many mind changes) gives a bound on the number of fluctuations. This, for instance, is the case for bounded monotone sequences. In general, we can say that effective learnability implies the existence of an effective bound of fluctuations, if the learner and the sequence satisfy certain gap conditions. Informally, if

- any two exceptions $i, \tilde{\imath}$ to the Cauchy property for 2^{-k} have distance at least $\Delta_*(\max(i, \tilde{\imath}), k)$, and
- the learning map jumps at most by $J^*(i, k)$ from i, and
- (for any k) $J^*(\cdot, k)$ is asymptotically at most equivalent to $\Delta_*(\cdot, k)$,

then the number of fluctuations is asymptotically bounded by B^*. Below, we discuss an example from ergodic theory, where these conditions are met.

Proposition 5.1 (Gap conditions on the learner)**.** *Let* $a_{(\cdot)}$ *be some sequence in a metric space* (X, d) *and let*

$$\varphi :\equiv \forall k \exists n \forall i \; \overbrace{\forall \tilde{\imath} < i \big(n \leq \tilde{\imath} \to d(a_{\tilde{\imath}}, a_i) \leq 2^{-k}\big)}^{\varphi_0(i,n,k) :\equiv}.$$

be a (B, L)*-learnable formula (which states simply the Cauchy property of the sequence, we use* $\tilde{\imath}$ *simply because the natural choice* j *already denotes the pairing function) and let* B^*, L^* *be majorants of* B, L*.*
Moreover, suppose that there are functions $\Delta_* > 0$*,* J^**, s.t.*

$$\forall n, i, i' \, \Big(\big(\neg\varphi_0(i, n, k) \wedge \neg\varphi_0(i', n, k)\big) \to |i' - i| \geq \Delta_*(i, k)\Big),$$

$$\forall n, i\big(\neg\varphi_0(i,n,k) \to L^*(i,k) - i \leq J^*(i,k)\big),$$

and

$$\big(J^* \in \mathcal{O}(\Delta_*)\big) \equiv \forall k \exists N_k, K_k\ \forall x \geq N_k\ \big(K_k \Delta_*(x,k) \geq J^*(x,k)\big). \tag{23}$$

Then there is a bound on the number of 2^{-k}-fluctuations, which is primitive recursive in B^*, Δ_*, J^* and N_k, K_k (which witness (23)) given by

$$b(k) := B^*(k)\bigg(2 + K_k + \max_{n<N_k}\Big(\frac{J^*(n,k)}{\Delta_*(n,k)}\Big)\bigg).$$

Note that in the case where $N_k = 0$, we get

$$b(k) = (2 + K_k)B^*(k).$$

Proof. For simplicity, assume $N_k = 0$ (otherwise we could just replace every occurrence of K_k by $(K_k + \max_{n<N_k}(\frac{J^*(n,k)}{\Delta_*(n,k)}))$).
Firstly, note that any 2^{-k}-fluctuation between two indexes $\tilde{\imath}$ and i corresponds to a counterexample i. So, by definition there is at most one fluctuation in the interval $[c_l, i_l]$, where i_l is the smallest counterexample to the solution candidate c_l. Moreover, if there is such a fluctuation, its greater index is i_l.
Secondly, from our assumption on Δ_* we get that there are at most

$$\left\lceil \frac{c_{l+1} - i_l}{\Delta_*(i_l,k)} \right\rceil \leq \left\lceil \frac{J^*(i_l,k)}{\Delta_*(i_l,k)} \right\rceil \leq K_k$$

many fluctuations within an interval $[i_l, c_{l+1}]$.
There are at most $B^*(k)$ such pairs of intervals, before a 2^{-k}-Cauchy point is reached, but there might be fluctuations, which arise only when we unite two such intervals.
By incrementing K_k by 1, we already account for any additional fluctuation due to combining the intervals $[c_l, i_l]$ and $[i_l, c_{l+1}]$. This is because if there was a fluctuation within $[c_l, i_l]$, then there cannot be an additional one which results from combining such a pair of intervals, as its greater index would be i_l. There can, however, be an additional fluctuation, when we combine the intervals $[i_l, c_{l+1}]$ and $[c_{l+1}, i_{l+1}]$.

□

We now consider the general form of the structure of Birkhoff's proof of the mean ergodic theorem as analyzed in [29] and the argument used in [9] to convert the rate of metastability obtained in [29] into a bound on the number of fluctuations:

Let $x_{(\cdot)}$ be a sequence in some normed space X (in the case at hand X is a uniformly convex Banach space) and $y_{(\cdot)}$ be a sequence in $\mathbb{R}_+$ definable by terms in $\mathsf{HA}^\omega[X, \|\cdot\|, \ldots]$. Suppose the Cauchyness of $x_{(\cdot)}$ is proved using that $y_{(\cdot)}$ has arbitrarily good approximate infima, i.e.

$$\forall \delta > 0 \exists n \forall k \forall \tilde{k} \leq k\ (y_{\tilde{k}} \geq y_n - \delta) \tag{24}$$

$$\to \forall \varepsilon > 0 \exists m \forall u \forall i, j \in [m, u]\ (\|x_i - x_j\| \leq \varepsilon). \tag{25}$$

This implication is classically equivalent to

$$\forall \varepsilon > 0 \exists \delta > 0\ \Big(\exists n \forall k \forall \tilde{k} \leq k\ (y_{\tilde{k}} \geq y_n - \delta) \to \exists m \forall u \forall i, j \in [m, u]\ (\|x_i - x_j\| \leq \varepsilon)\Big). \tag{+}$$

Suppose now that we are in the situation of Corollary 2.15, i.e.

$$\mathsf{HA}^\omega[X,\|\cdot\|,\ldots]+\mathrm{AC}+\mathrm{M}^\omega+\mathrm{IP}^\omega_\forall \vdash (+)$$

then

$$\mathsf{HA}^\omega[X,\|\cdot\|,\ldots]+\mathrm{AC}+\mathrm{M}^\omega+\mathrm{IP}^\omega_\forall \vdash$$
$$\forall\varepsilon>0\exists\delta>0\forall n\exists m\geq n\forall u\exists k(\forall\tilde{k}\leq k(y_{\tilde{k}}\geq_{\mathbb{R}} y_n-\delta)\to\forall i,j\in[m,u](\|x_i-x_j\|<_{\mathbb{R}}\varepsilon).$$

Hence by monotone functional interpretation one extracts terms $\delta_\varepsilon>0$, m_ε and k_ε (depending additionally only on majorants of the parameters $\underline{a}$ used in the definition of our sequences) s.t. (valid in $\mathcal{S}^{\omega,X}$) for all majorants $\underline{a}^*$ of $\underline{a}$

$$\forall\varepsilon>0\ \forall n,u$$
$$\Big(m_\varepsilon(n)\geq n\wedge\Big(\forall\tilde{k}\leq k_\varepsilon(n,u)\ (y_{\tilde{k}}\geq y_n-\delta_\varepsilon)\to\forall i,j\in[m_\varepsilon(n),u]\ (\|x_i-x_j\|\leq\varepsilon)\Big)\Big).\quad(*)$$

Now define $k^*_\varepsilon(u):=\max\{k_\varepsilon(i,u)\ :\ i\leq u\}$ and consider

$$\forall\varepsilon>0\ \forall n,u\ \Big(\forall\tilde{k}\leq k^*_\varepsilon(u)\ (y_{\tilde{k}}\geq y_n-\delta_\varepsilon)\to\forall i,j\in[m_\varepsilon(n),u]\ (\|x_i-x_j\|\leq\varepsilon)\Big).\quad(**)$$

We can infer $(**)$ from $(*)$ by the following case distinction:[9] Fix $\varepsilon>0$ and n.

Case 1: $u<m_\varepsilon(n)$. Then the conclusion and hence the whole implication is trivially true.

Case 2: $u\geq m_\varepsilon(n)\geq n$.Then $k^*_\varepsilon(u)\geq k_\varepsilon(n,u)$ and so $\forall\tilde{k}\leq k^*_\varepsilon(u)\ (y_{\tilde{k}}\geq y_n-\delta_\varepsilon)$ implies $\forall\tilde{k}\leq k_\varepsilon(n,u)\ (y_{\tilde{k}}\geq y_n-\delta_\varepsilon)$ and so the claim follows as well.

Now suppose w.l.o.g. that $k^*_\varepsilon:\mathbb{N}\to\mathbb{N}$ is injective and for any given u define

$$l_u\ :=\ (k^*_\varepsilon)^{-1}(u).$$

Then $(**)$ applied to $u:=l_u$ yields

$$\forall\varepsilon>0\ \forall n,u\ \Big(\forall\tilde{k}\leq u\ (y_{\tilde{k}}\geq y_n-\delta_\varepsilon)\to\forall i,j\in[m_\varepsilon(n),(k^*_\varepsilon)^{-1}(u)]\ (\|x_i-x_j\|\leq\varepsilon)\Big).\ (\text{-})$$

Now let N_0, N_1, ..., N_{S_ε} be integers s.t. $N_0=0$ and N_{i+1} is the least $m>N_i$ s.t. $y_m<y_{N_i}-\delta_\varepsilon$ as long as such an m exists. Assume that $b\geq y_0$ (for some b) and so $S_\varepsilon\leq\frac{b}{\delta_\varepsilon}$.
By (-) there are no ε-fluctuations of $x_{(\cdot)}$ on the S_ε many intervals $[m_\varepsilon(N_i),(k^*_\varepsilon)^{-1}(N_{i+1})]$ for $i=0,\ldots,S_\varepsilon-1$.
In the intervals $[(k^*_\varepsilon)^{-1}(N_i),m_\varepsilon(N_i)]$ for $i=1,\ldots,S_\varepsilon$ and $[0,m_\varepsilon(N_0)]$ we have to show that if we have for any $N\in\mathbb{N}$ s many fluctuations indexed within $[(k^*_\varepsilon)^{-1}(N),m_\varepsilon(N)]$ (or in $[0,m_\varepsilon(N_0)]$) each indexed by a pair of indexes (i,j) then the highest index of such fluctuation (j_s) has to be greater than (or equal to) some $\varphi_\varepsilon(s,N)$, where φ_ε is such that

$$\exists\tilde{s}\forall n\ \big(\varphi_\varepsilon(\tilde{s},n)>m_\varepsilon(n)\big).$$

[9] We are grateful to P. Oliva for pointing this out to us.

Then, given such an $\tilde{s}$, we have at most

$$\frac{b}{\delta_\varepsilon} + \tilde{s}\left(\frac{b}{\delta_\varepsilon} + 1\right)$$

many fluctuations.

In the case of Birkhoff's proof, the analysis in [29] and the discussion in [9] gives the following data used in [9]:

$$\delta_\varepsilon := \frac{\varepsilon^2}{512b}, \qquad m_\varepsilon(n) := \left\lceil \frac{16b}{\varepsilon} \right\rceil n,$$
$$(k^*_\varepsilon)^{-1}(n) := \left\lfloor \frac{n}{2} \right\rfloor, \qquad \varphi_\varepsilon(s,n) := \big(1 + \frac{\varepsilon}{2b}\big)^s n,$$

and so (for $\varepsilon < 2b$)

$$\tilde{s} \leq \frac{4b \log \left\lceil \frac{16b}{\varepsilon} \right\rceil}{\varepsilon}.$$

The function φ_ε results (see [9] for the calculation) from the fact that

$$\|x_{n+k} - x_k\| \leq 2n\|x\|/(n+k)$$

which is established already in Birkhoff's proof and which – for $n = 1$ – shows that (x_k) has a linear rate of asymptotic regularity.

Remark 5.2. Naturally, we could use the data, which led to the bound of $\tilde{s}$ above, also simply with Proposition 5.1 to obtain a similar fluctuation bound (which has the same structure in ε).

For the case of Halpern iterations (with scalar $1/(n+1)$) mentioned in the introduction, the analysis given in [31] yields (roughly) the following data for Hilbert spaces X (see [31, 33] for the detailed definition of Θ_n):

$$\delta_\varepsilon := \frac{\varepsilon^4}{576(b+1)^4}, \qquad m_\varepsilon(n) \approx \Theta_n(\frac{\varepsilon^2}{4}) \approx n^2,$$
$$(k^*_\varepsilon)^{-1}(n) := \left\lfloor \frac{n \cdot \varepsilon}{3b^2} \right\rfloor.$$

Similar data are also obtained in the recent [34] which is based on the analysis of a different proof for the strong convergence of the Halpern iteration from [43].
However, now the rate of asymptotic regularity roughly is of order (see corollary 6.3 in [31])

$$\|x_{k+1} - x_k\| \leq \frac{b}{\sqrt{k}},$$

which does not lead to a linear (in n) $\varphi_\varepsilon(s,n)$ and even if it would, this would not suffice to dominate $m_\varepsilon(n)$. So as it stands, the analysis does not seem to yield any fluctuation bound for the Halpern iteration (x_k).

Proposition 5.3. *Given a bound B_ε on the number of fluctuations, there is an in B_ε (and the given data $(k^*_\varepsilon)^{-1}$ and m_ε and the majorants $\underline{a}^*$ of their parameters including ε)) primitive recursive φ_ε satisfying the conditions in the proof:*

$$\forall \varepsilon > 0 \forall s, N, i, j \big(i, j \in [(k^*_\varepsilon)^{-1}(N), m_\varepsilon(N)] \ \wedge \ \mathrm{Fluc}_\varepsilon(s,i,j) \ \to j_s \geq \varphi_\varepsilon(s,N)\big), \quad (26)$$

$$\exists \tilde{s} \forall n \ \big(\varphi_\varepsilon(\tilde{s}, n) > m_\varepsilon(n)\big). \quad (27)$$

Proof. Set

$$\varphi_\varepsilon(s,n) := \begin{cases} (k^*_\varepsilon)^{-1}(n) & \text{if } \ s \leq B_\varepsilon, \\ m_\varepsilon(n) + 1 & \text{otherwise.} \end{cases}$$

□

Remark 5.4. In particular, this means that if we know there is for computable (in the majorants $\underline{a}^*$ of the parameters including ε) $(k^*_\varepsilon)^{-1}$ and m_ε no computable φ_ε (in $\underline{a}^*$) satisfying these conditions, then there cannot be a bound on the number of fluctuations, which is computable (in $\underline{a}^*$).

Acknowledgements: This research was supported by the German Science Foundation (DFG Project KO 1737/5-1).

References

[1] Akama, Y., Berardi, S., Hayashi, S., Kohlenbach, U., An arithmetical hierarchy of the law of excluded middle and related principles. Proc. of the 19th Annual IEEE Symposium on Logic in Computer Science (LICS'04), pp. 192-201, IEEE Press (2004).

[2] Artemov, S., Explicit provability and constructive semantics. Bull. Symbolic Logic **7**, pp. 1-36 (2001).

[3] Artemov, S., Yavorskaya, First-order logic of proofs. Technical Report TR-2011005, Cuny PhD Program in Computer Science 2011.

[4] Aschieri, F., A contructive analysis of learning in Peano Arithmetic. To appear in: Ann. Pure Appl. Logic.

[5] Aschieri, F., Learning based on realizability for HA+EM1 and 1-Backtracking games: Soundness and completeness. To appear in: Ann. Pure Appl. Logic.

[6] Aschieri, F., Berardi, S., A new use of Friedman's translation: interactive realizability. in: Logic, Construction, Computation, Ontos-Verlag Series in Mathematical Logic, Berger et al. editors, 2012

[7] Avigad, J., Gerhardy, P., Towsner, H., Local stability of ergodic averages. Trans. Amer. Math. Soc. **362**, pp. 261-288 (2010).

[8] Avigad, J., Iovino, J., Ultraproducts and metastability. arXiv:1301.3063, Preprint 10pp., 2013.

[9] Avigad, J., Rute, J., Oscillation and the mean ergodic theorem. arXiv:1203.1743, Preprint 7pp., 2012.

[10] Baillon, J.B., Un théorème de type ergodique pour les contractions non linéaires dans un espace de Hilbert. C.R. Acad. Sci. Paris Sèr. A-B **280**, pp. 1511-1514 (1975).

[11] Baillon, J.B., Quelques propriétés de convergence asymptotique pour les contractions impaires. C.R. Acad. Sci. Paris Sèr. A-B **283**, pp. 587-590 (1976).

[12] Berardi, S., Coquand, T, Hayashi, S., Games with 1-Backtracking. Ann. Pure Appl. Logic **161**, pp. 1254-1264 (2010).

[13] Birkhoff, G., The mean ergodic theorem. Duke Math. J. **5**, pp. 19-20 (1939).

[14] Coquand, T., A semantics of evidence for classical arithmetic. J. Symbolic Logic **60**, pp. 325-337 (1995).

[15] Friedman, H., Classical and intuitionistically provably recursive functions. In: Müller,G.H., Scott, D.S. (eds.), Higher Set Theory, pp. 21-27. Springer LNM **669** (1978).

[16] Gerhardy, P., Kohlenbach, U., Strongly uniform bounds from semi-constructive proofs, Ann. Pure Appl. Logic **141**, pp. 89-107 (2006).

[17] Gold, E. M., Language identification in the limit. Information and Control **10**, pp. 447-474 (1967).

[18] Hayashi, S., Mathematics based on learning. Proc. 13th Internat. Conf. ALT 2002, Springer Lecture Notes in Artificial Intelligence Vol. 272, pp. 7-21 (2002).
[19] Hayashi, S., Mathematics based on incremental learning – Excluded middle and inductive inference Original Research Article Theoretical Computer Science **350**, pp. 125-139 (2006) (Journal version of [18]).
[20] Hayashi, S., Nakata, M., Towards limit computable mathematics. In: P. Callaghan et al. (eds.), TYPES 2000, Springer LNCS **2277**, pp. 125-144 (2002).
[21] Higuchi, K., Kihara, T., Inside the Muchnik degrees: discontinuity, learnability, and constructivism. Preprint 2012.
[22] Jones, R.L., Ostrovskii, I.V., Rosenblatt, J.M., Square functions in ergodic theory. Ergodic Theory and Dynamical Systems **16**, pp. 267-305 (1996).
[23] Kohlenbach, U., Mathematically strong subsystems of analysis with low rate of growth of provably recursive functionals. Arch. Math. Logic **36**, pp. 31-71 (1996).
[24] Kohlenbach, U., Relative constructivity. J. Symbolic Logic **63**, pp. 1218-1238 (1998).
[25] Kohlenbach, U., Some logical metatheorems with applications in functional analysis. Trans. Amer. Math. Soc. **357**, no. 1, pp. 89-128 (2005).
[26] Kohlenbach, U., Applied Proof Theory: Proof Interpretations and their Use in Mathematics. Springer Monographs in Mathematics. xx+536pp., Springer Heidelberg-Berlin, 2008.
[27] Kohlenbach, U.,On quantitative versions of theorems due to F.E. Browder and R. Wittmann. Advances in Mathematics **226**, pp. 2764-2795 (2011).
[28] Kohlenbach, U., A uniform quantitative form of sequential weak compactness and Baillon's nonlinear ergodic theorem. Communications in Contemporary Mathematics **14**, 20pp. (2012).
[29] Kohlenbach, U., Leuştean, L., A quantitative mean ergodic theorem for uniformly convex Banach spaces. Ergodic Theory and Dynamical Systems **29**, pp. 1907-1915 (2009).
[30] Kohlenbach, U., Leuştean, L., Asymptotically nonexpansive mappings in uniformly convex hyperbolic spaces. Journal of the European Mathematical Society **12**, pp. 71-92 (2010).
[31] Kohlenbach, U., Leuştean, L., Effective metastability of Halpern iterates in CAT(0) spaces. Adv. Math. **231**, pp. 2526-2556 (2012).
[32] Kohlenbach, U., Leuştean, L., On the computational content of convergence proofs via Banach limits. Philosophical Transactions of the Royal Society A **370**, pp. 3449-3463 (2012).
[33] Kohlenbach, U., Leuştean, L., Addendum to [31]. 2pp. (2013).
[34] Körnlein, D., Analysis of a proof due to Xu. Preprint 2013.
[35] Luckhardt, H., Herbrand-Analysen zweier Beweise des Satzes von Roth: Polynomiale Anzahlschranken. J. Symbolic Logic **54**, pp. 234-263 (1989).
[36] Saejung, S., Halpern's iteration in CAT(0) spaces. Fixed Point Theory and Applications **2010** (2010).
[37] Safarik, P., A quantitative nonlinear strong ergodic theorem for Hilbert spaces. J. Math. Anal. Appl. **391**, pp. 26-37 (2012).
[38] Sieg, W., Fragments of arithmetic. Ann. Pure Appl. Logic. **28**, pp. 33-71 (1985).
[39] Specker, E., Nicht konstruktiv beweisbare Sätze der Analysis. J. Symb. Logic **14**, pp. 145-158 (1949).
[40] Toftdal, M., Calibration of Ineffective Theorems of Analysis in a Constructive Context. Master Thesis, Aarhus Universitet, 2004.
[41] Wittmann, R., Mean ergodic theorems for nonlinear operators. Proc. Amer. Math. Soc. **108**, pp. 781-788 (1990).
[42] Wittmann, R., Approximation of fixed points of nonexpansive mappings. Arch. Math. **58**, pp. 486-491 (1992).
[43] Xu, H.-K., Iterative algorithms for nonlinear operators. J. London Math. Soc. **66**, pp. 240-256 (2002).
[44] Ziegler, M., Real hypercomputation and continuity. Theory of Computing Systems vol.41, pp. 177-206 (2007).