An exploration of proof mining

Andrei Sipoş[a,b,c]

[a]Research Center for Logic, Optimization and Security (LOS), Department of Computer Science,
Faculty of Mathematics and Computer Science, University of Bucharest,
Academiei 14, 010014 Bucharest, Romania

[b]Simion Stoilow Institute of Mathematics of the Romanian Academy,
Calea Griviţei 21, 010702 Bucharest, Romania

[c]Institute for Logic and Data Science,
Popa Tatu 18, 010805 Bucharest, Romania

Email: andrei.sipos@fmi.unibuc.ro

These notes support a tutorial delivered at Logic Colloquium 2024, but are meant to stand alone. They aim to introduce the informal practice of proof mining through a series of concrete, down-to-earth examples. In a way, they form a natural counterpart to the rather bird's eye view approach taken in the "What proof mining is about" blog post series of mine on the Proof Theory Blog (see, e.g. the first part of it [9]).

The standard references for proof mining are Ulrich Kohlenbach's book [5] and ICM survey [6].

I will assume some familiarity with first-order logic and real analysis.

1 Pure classical first-order logic

The basic idea of proof mining goes like this: one knows about a mathematical result which states that there is an x for which something happens and wants to obtain a concrete formula (a 'witness') for that x. Looking at this idea with a logical eye leads one to the idea of *extraction* theorems (or 'metatheorems'), which assert that, under some conditions, one might simply 'extract' a suitable formula for the x from the (formal) *proof* of the mathematical result. The simplest (and most widely known) such theorem goes back to Herbrand (1930), and concerns proofs which are formalizable in pure classical first-order logic. It is expressed as follows.

Theorem 1.1. *Let σ be a first-order signature which contains at least one constant. Let φ be a quantifier-free σ-formula containing at most one free variable, which we denote by x. Let Γ be a set of purely universal σ-sentences. Assume that $\Gamma \vdash \exists x \varphi$. Then one may extract from that proof a finite list of closed σ-terms $t_1, \ldots, t_n$ such that*

$$\Gamma \vdash \bigvee_{i=1}^{n} \varphi[x := t_i].$$

There exists a standard pedagogical worked example to illustrate this theorem, due to Ulrich Berger. I will now present it purely 'semantically', with no reference to logic, to foreshadow the idea of analysing proofs from mainstream mathematics. So, for now, imagine the following as if it were a result in a mathematics textbook or journal.

Theorem 1.2. *Let M be a set, $z \in M$, $s : M \to M$ and $f : M \to M$. Assume that, for all $x \in M$, $s(x) \neq z$. Then there is an $x \in M$ with $f(s(f(x))) \neq x$.*

Now, a moment's reflection will assure a logically minded reader that this result fits tightly into the conditions of Theorem 1.1. So, the next step would be to look at the proof and to try to apply an algorithm to it in order to get the list of witnesses.

Proof of Theorem 1.2. Assume towards a contradiction that, for all $x \in M$, $fsfx = x$ (we will drop the brackets, since all functions are unary and there is no risk of confusion).

We first claim that f is injective. Let $x_1, x_2 \in M$ be such that $fx_1 = fx_2$. Then $fsfx_1 = fsfx_2$, so, by our assumption, $x_1 = x_2$.

Next, we claim that the image through f of $M \setminus \{z\}$ is the whole M. Let $y \in M$. Then, by our assumption, $y = fsfy$ and we know, by our hypothesis, that $sfy \neq z$, so we are done.

Thus, there is a $t \in M \setminus \{z\}$ with $ft = fz$, which contradicts the injectivity of f. □

The proof above is simple enough that, to 'extract the terms' from it, one does not necessarily need to know the exact algorithm used in a proof of Herbrand's theorem (also, there may be, and indeed they are, several such proofs and algorithms), but one can just informally 'inspect' the proof and the terms will immediately come out of it.

The proof goes by way of contradiction, so we shall look at the instances of the assumption 'for all $x \in M$, $fsfx = x$' that were actually used in the proof.

First of all, to obtain that t, we applied our second claim with $y := fz$, so $t = sfy = sffz$. But, in the proof of that claim, the assumption was used for y, so our first extracted term is fz.

Secondly, we used the injectivity claim with $x_1 := t = sffz$ and $x_2 := z$, and the assumption is used in the proof of that claim for x_1 and x_2 themselves, so our remaining terms are $sffz$ and z.

Therefore, we have obtained the following list of terms: fz, $sffz$, z.

Herbrand's theorem concerns relatively simple formulas, containing just one existential quantifier, so we may naturally ask ourselves whether we can go higher in terms of formula complexity. Let's look at universal-existential sentences, i.e. of the form $\forall x \exists y \varphi$. The goal would be, as before, to obtain a 'formula' (or, rather, a finite list of them, as before) for the y, but, this time, **in terms of** x. Here, a simple trick does the job: we add a new constant c to our signature, and we get that $\Gamma \vdash \exists y(\varphi[x := c])$. We may now apply Herbrand's theorem to obtain a list of witnesses for the y. Notice that these terms will be over the extended signature containing the additional constant c. If we substitute in these terms c by x, we obtain what we wanted: a list of possible 'formulas' for y in terms of x.

Unfortunately, this is the highest we can get. Consider a case where we have one universal quantifier *after* an existential one. Say our signature contains one unary relation (predicate) symbol P and, since we already have this requirement in Herbrand's theorem, one constant c. We may formulate the valid sentence known as the **drinker's paradox**:

$$\exists y \forall z(Py \to Pz).$$

However, there is only one term over this signature, namely c, and one may easily give a counterexample for which c does not validate the sentence. For example, take the universe to be $\{0, 1\}$, P to contain just 1 and c to be 1. Then, clearly, in this structure, it is not the case that $\forall z(Pc \to Pz)$, since it is not true that $P1 \to P0$.

We keep this in mind for later.

2 Arithmetic and metastability

Pure first-order logic is not all there is, in the sense that the foundations of mathematics are usually formalized by adding *axioms* on top of classical logic. Gödel's (first) incompleteness theorem tells us that no well-behaved axiom system establishing foundations of mathematics can be complete, so we have to deal with not just one, but a plethora of such (incomplete) systems. It is an empirical fact that almost all systems devised by humans for such a purpose can be sorted by their proof-theoretic strength into a linear **hierarchy**, which is usually thought of as containing three conventional 'levels':

- 'arithmetic', in particular the first-order theory of arithmetic usually known as 'Peano arithmetic', denoted by PA;
- 'analysis', in particular the first-order, two-sorted theory usually known as 'second-order arithmetic', denoted by SOA or Z_2;
- 'set theory', in particular the usual Zermelo-Fraenkel set theory with the axiom of choice, ZFC.

Mathematicians usually regard their work as being, in principle, capable of being formalized into ZFC (with the possible addition of large cardinal axioms for e.g. algebraic geometry), but it is generally thought (see *Friedman's grand conjecture*) that all theorems in mainstream mathematics which are purely arithmetical may be proven in a rather weak fragment of PA. So far, the practice of proof mining seems to confirm this (see Ulrich Kohlenbach's paper [7] on this topic where he calls the phenomenon *proof-theoretic tameness*), and this is very fortunate, since there exist term extraction theorems (which are more or less 'higher' analogues of Herbrand's theorem) which do pertain to such systems and may be applied to concrete mathematical proofs.

Since this is an example-based tutorial, I will not present the details of such theorems. Instead, I will start by illustrating one of the simplest results where proof mining may yield a piece of information which may be regarded as essentially new.

Consider the monotone convergence theorem, as usually taught in introductory real analysis courses. To fix ideas, we state it as follows.

Theorem 2.1. *Let (a_n) be a nonincreasing sequence in the interval $[0,1]$. Then (a_n) is convergent.*

The convergence of the sequence (a_n) is expressed as

$$\exists x \in \mathbb{R}\ \forall \varepsilon > 0\ \exists N \in \mathbb{N}\ \forall n \geq N\ |a_n - x| \leq \varepsilon.$$

The (naive) hope here would be to somehow obtain a *rate of convergence*, that is, a function giving an upper bound on the N in terms of the ε that would be 'uniform', in the sense that it would work for all nonincreasing sequences in the interval $[0,1]$. However, this may be easily seen to be impossible.

Assume towards a contradiction that we do have such a function, which we will denote by φ. Let's consider the special case where the limit of (a_n) is 0, so we would have that

$$\forall \varepsilon > 0\ \exists N \leq \varphi(\varepsilon)\ \forall n \geq N\ a_n \leq \varepsilon,$$

or, more simply,

$$\forall \varepsilon > 0\ \forall n \geq \varphi(\varepsilon)\ a_n \leq \varepsilon.$$

In particular, for $\varepsilon := 1/2$,

$$\forall n \geq \varphi(1/2)\ a_n \leq 1/2.$$

But it is easy to cook up a sequence converging to 0 for which the above does not hold: for example, take, for every $n \in \mathbb{N}$,

$$a_n := \begin{cases} 1, & \text{if } n \leq \varphi(1/2), \\ 0, & \text{otherwise.} \end{cases}.$$

We are led to say that monotone sequences in the unit interval converge 'arbitrarily slow'.

I will now turn to what proof mining may indeed be able to yield. First of all, let's get rid of the initial 'non-numerical' quantifier '$\exists x \in \mathbb{R}$' by replacing convergence with *Cauchyness*, i.e.

$$\forall \varepsilon > 0\ \exists N \in \mathbb{N}\ \forall n, m \geq N\ |a_n - a_m| \leq \varepsilon,$$

which we will now put into the following form, which, for some reason, will be easier to work with:

$$\forall \varepsilon > 0\ \exists N \in \mathbb{N}\ \forall M \in \mathbb{N}\ \forall n, m \in [N, N+M]\ |a_n - a_m| \leq \varepsilon.$$

Note that this maneuvering did not provide us with a statement where we may hope to obtain the intuitive corresponding notion of 'rate of Cauchyness', since finding a bound on the N in the above statement may be immediately seen to be more or less equivalent (exercise!) to finding a rate of convergence. And this should have been to some degree expected, since we may see that the statement exhibits a universal quantifier ('$\forall M$') after an existential one ('$\exists N$'), like in the drinker's paradox sentence which I presented in the previous section. Proof mining metatheorems (for systems based on classical logic) are largely restricted to $\forall\exists$-like statements; still, as Ulrich Kohlenbach once pointed out to me, "it is not the 'techniques of proof mining' which preclude [the extraction of a uniform rate of convergence], but reality who does!"

In order to see what we may actually obtain, since what we aim to do is proof mining, we now look at the *proof* of the Cauchyness statement.

Proof. Assume towards a contradiction that the opposite holds, i.e. that

$$\exists \varepsilon > 0 \; \forall N \in \mathbb{N} \; \exists M \in \mathbb{N} \; \exists n, m \in [N, N+M] \; |a_n - a_m| > \varepsilon.$$

Using that the sequence is nonincreasing, this is equivalent to

$$\exists \varepsilon > 0 \; \forall N \in \mathbb{N} \; \exists M \in \mathbb{N} \; a_N - a_{N+M} > \varepsilon.$$

So, we have an $\varepsilon > 0$ such that

$$\forall N \in \mathbb{N} \; \exists M \in \mathbb{N} \; a_N - a_{N+M} > \varepsilon.$$

In order to obtain a contradiction, we apply the statement above to various N's. The simplest possible N is 0. Let's call this N_0. So, there is an M such that $a_0 - a_M > \varepsilon$. This new number M we'll call N_1 (in order to not be ambiguous, we may take the least such M each time we apply the statement). Using this notation, we get that $a_{N_0} - a_{N_1} > \varepsilon$. We apply our statement again to N_1, so there is an M such that $a_{N_1} - a_{N_1+M} > \varepsilon$. Set now $N_2 := N_1 + M$, so $a_{N_1} - a_{N_2} > \varepsilon$. Let's say we apply the statement repeatedly, $l+1$ times in total, so that, for each $i \leq l$, $a_{N_i} - a_{N_{i+1}} > \varepsilon$. Summing up these inequalities, we get that $a_0 - a_{N_{l+1}} > (l+1)\varepsilon > l\varepsilon$. But $a_0 - a_{N_{l+1}} \leq a_0 \leq 1$, so, if we take an l such that $l\varepsilon \geq 1$, say $l := \lceil 1/\varepsilon \rceil$, we get a contradiction. □

Now, the thing is, even though we have said and even shown that a uniform 'formula' for a rate of convergence or Cauchyness cannot exist, one may nevertheless get the inescapable feel of a computational flavour in the above argument, so that one may then be tempted to look into it some more. For we see that, at each iteration i, we 'compute' N_{i+1} as $N_i + M$, where the M is somehow 'given' in terms of N, since we assume that for every N there is an M (and, as said above, we take the least such one). We may think of this dependence of M on the N in the form of a function $g : \mathbb{N} \to \mathbb{N}$, so our assumption, in the way that was actually used in the proof, could be expressed as

$$\exists g : \mathbb{N} \to \mathbb{N} \; \forall N \in \mathbb{N} \; a_N - a_{N+g(N)} > \varepsilon,$$

which is usually called the 'Skolemization' of the form in which the assumption had been written before.

Putting $\widetilde{g}(n) := n + g(n)$, we remark that, for each $i \leq l$, $N_{i+1} = \widetilde{g}(N_i)$, so, for each $i \leq l+1$, $N_i = \widetilde{g}^{(i)}(0)$. In particular, for each $i \leq l$, $N_i \leq \widetilde{g}^{(l)}(0)$. Keeping in mind that we have set $l := \lceil 1/\varepsilon \rceil$, what the argument actually showed is that

$$\forall \varepsilon > 0 \; \forall g : \mathbb{N} \to \mathbb{N} \; \exists N \in \mathbb{N} \; \forall n, m \in [N, \widetilde{g}(N)] \; |a_n - a_m| \leq \varepsilon,$$

with the upper bound on the N given by $\widetilde{g}^{(\lceil 1/\varepsilon \rceil)}(0)$. The property given by the statement above is classically equivalent to the Cauchyness of the sequence, being its 'Herbrandization' (the dual of Skolemization). It is notable in having been rediscovered in the 2000s by Terence Tao during his work in ergodic theory [13], and popularized through his blog [12], where he asked for suggestion for a proper name for it. The name that ultimately stuck was **metastability**, suggested by Jennifer Chayes (then a researcher at Microsoft), and one may then speak of a 'rate of metastability' as a bound on the N in the metastability statement, given in terms of the ε and the g (and perhaps some other parameters of the problem).

The above discussion shows that, if we think of the monotone convergence theorem as referring to metastability instead of convergence or Cauchyness, we may reformulate it into the following quantitative form, which shows that monotone sequences (both nonincreasing and nondecreasing, since the proofs are identical) in the unit interval admit a **uniform** rate of metastability. This also shows that the equivalence between Cauchyness and metastability cannot be constructive, since it would then get upgraded to the level of rates, but, as we have seen, a uniform rate of Cauchyness does not (cannot) exist.

Theorem 2.2. *Let (a_n) be a monotone sequence in the interval $[0,1]$. Then, for any $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$ there is an $N \leq \widetilde{g}^{(\lceil 1/\varepsilon \rceil)}(0)$ such that, for all n, $m \in [N, \widetilde{g}(N)]$, $|a_n - a_m| \leq \varepsilon$.*

I will now make some additional remarks. First of all, we have seen in the course of the proof that not only do we know an upper bound on the N, but we know that the N has to belong to the finite list $\widetilde{g}^{(0)}(0), \ldots, \widetilde{g}^{(\lceil 1/\varepsilon \rceil)}(0)$. This resembles a bit the case of pure first-order logic and Herbrand's theorem,

with a significant difference: the length of the list is **variable**, since it depends on ε (but not the g, so we do have some uniformity there). An underlying logical explanation for this fact has recently been discovered, see the paper [11].

In addition, since the proof only involves an initial finite segment of the sequence under discussion (whose length depends, though, on the ε and the g), we might obtain the following completely 'finitary' form of this result, which Tao called the 'finite monotone convergence principle'.

Theorem 2.3. *For any $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$, and any finite monotone sequence $(a_n)_{n=0}^{\widetilde{g}^{(\lceil 1/\varepsilon \rceil + 1)}(0)}$ there is an $N \leq \widetilde{g}^{(\lceil 1/\varepsilon \rceil)}(0)$ with $\widetilde{g}(N) \leq \widetilde{g}^{(\lceil 1/\varepsilon \rceil + 1)}(0)$ such that, for all n, $m \in [N, \widetilde{g}(N)]$, $|a_n - a_m| \leq \varepsilon$.*

Finally, I will mention that, as I said above, this principle was only 'rediscovered' by Tao, since it was already known in this very form by Georg Kreisel in the 1950s, as it can be seen in the exact paper [8, pp. 49–50] which introduced the program now known as proof mining!

3 Iterations on the unit interval

The examples that we have looked at so far cannot be said to truly represent the 'practice' of proof mining, since they were quite pedagogical in nature. I will now aim to present a proof mining result which has indeed been published in a mathematical journal (see [10]). To keep things grounded, the result I have chosen still 'takes place' on the real line, and does not feature more involved structures like polynomial rings or Banach spaces.

The goal of the kind of nonlinear analysis where proof mining has usually produced results is to find algorithms that compute fixed points of operators. Consider the following theorem concerning the unit interval.

Theorem 3.1. *Let $f : [0,1] \to [0,1]$ be continuous, and (t_n) and (x_n) be sequences in $[0,1]$ such that, for any n, $x_{n+1} = (1-t_n)x_n + t_n f(x_n)$. Assume that $\sum t_n = \infty$ and (x_n) is convergent. Then the limit of (x_n) is a fixed point of f.*

Proof. Let z be that limit, and assume that $f(z) \neq z$, w.l.o.g. $f(z) > z$. Set, for every n, $y_n := f(x_n) - x_n$. Since f is continuous, (y_n) tends to $f(z) - z > 0$, so $\sum t_n y_n = \infty$. But, for every $n \in \mathbb{N}$,

$$x_{n+1} - x_0 = \sum_{k=0}^{n} t_k y_k,$$

where the left hand side tends to $z - x_0$ and the right hand side tends to infinity, a contradiction. □

The above theorem, which refers to so-called **Krasnoselski-Mann** iterative sequences, shows that all we need to do for such sequences is to find criteria under which they are convergent. An example of such a convergence theorem is the following, due to Borwein and Borwein [1].

Theorem 3.2. *Let $L > 0$, $f : [0,1] \to [0,1]$ be L-Lipschitz, (t_n) and (x_n) be sequences in $[0,1]$ such that, for all n, $x_{n+1} = (1-t_n)x_n + t_n f(x_n)$. Let $\delta \in (0,1)$ be such that, for all n, $t_n \leq \frac{2-\delta}{L+1}$. Then (x_n) is convergent.*

This is the theorem that we will analyse from the point of proof mining, and our goal will be to extract from its proof a rate of metastability which is as uniform as possible (in fact, as we will see, it will depend, in addition to the ε and the g, just on the δ). We will now look at that proof. You may skip the parts which are not relevant to the proof mining endeavour and which are concentrated in the following lemmas and definitions.

Lemma 3.3. *Let $L > 0$, $f : [0,1] \to [0,1]$ be L-Lipschitz, x, $x^* \in [0,1]$, $\delta \in (0,1)$ and $t \in [0,1]$ such that $t \leq \frac{2-\delta}{L+1}$ and $x^* = (1-t)x + tf(x)$. Let p be a fixed point of f which is located between x and x^*. Then*

$$|x^* - p| \leq (1-\delta)|x - p|.$$

Proof. Assume w.l.o.g. $x \leq x^*$. Then

$$\begin{aligned} |x^* - p| &= x^* - p = (1-t)(x-p) + t(f(x) - f(p)) \leq (t-1)(p-x) + tL(p-x) \\ &= (t(1+L) - 1)(p-x) \leq (1-\delta)|x-p|. \end{aligned}$$

□

Definition 3.4. *Let $(x_n) \subseteq [0,1]$ and $f : [0,1] \to [0,1]$.*

We say that $(\sigma_n) \subseteq \{\pm 1\}$ is the **sign sequence** *for (x_n) relative to f if $\sigma_0 = 1$ and for all n, if $f(x_n) - x_n \neq 0$, $\sigma_{n+1} = \operatorname{sgn}(f(x_n) - x_n)$ and otherwise $\sigma_{n+1} = \sigma_n$ – note that for all n, if $\sigma_{n+1} = 1$ (respectively -1), then $f(x_n) - x_n \geq 0$ (respectively ≤ 0).*

We say that $(q_n) \subseteq \mathbb{N} \cup \{\infty\}$ is the **switching sequence** *for (x_n) relative to f if, denoting by (σ_n) the sign sequence for (x_n) relative to f, $q_0 = 0$ and for all n, if $q_n = \infty$ then $q_{n+1} = \infty$ else if there is a $k > q_n$ with $\sigma_{k+1} = -\sigma_{q_n+1}$, q_{n+1} is the least such k, else $q_{n+1} = \infty$. Note the following:*

- *for all n with $q_{n+1} < \infty$, we have that $\sigma_{q_{n+1}+1} = -\sigma_{q_n+1}$ and that for all $l \in [q_n + 1, q_{n+1}]$, $\sigma_l = \sigma_{q_n+1}$;*
- *(q_n) is nondecreasing;*
- *for every r, $r \leq q_r$;*
- *for every r with $q_r < \infty$, the finite sequence $(x_n)_{n=q_r}^{q_{r+1}}$ is monotone.*

Lemma 3.5. *Let $L > 0$, $f : [0,1] \to [0,1]$ be L-Lipschitz, (t_n) and (x_n) be sequences in $[0,1]$ such that, for all n, $x_{n+1} = (1 - t_n)x_n + t_n f(x_n)$. Let (q_n) be the switching sequence for (x_n) relative to f. Let $r \geq 1$ with $q_{r+1} < \infty$ and put $n_1 := q_r - 1$ and $n_2 := q_{r+1} - 1$. Let $\delta \in (0,1)$ be such that, for all n, $t_n \leq \frac{2-\delta}{L+1}$. Then:*

(i) for all $n \in [n_1 + 1, n_2 + 1]$, x_n is located between x_{n_1} and x_{n_1+1};

(ii) $|x_{n_2} - x_{n_2+1}| \leq \left(1 - \frac{\delta}{2}\right) |x_{n_1} - x_{n_1+1}|$.

Proof. Let (σ_n) be the sign sequence for (x_n) relative to f. Assume w.l.o.g. that $\sigma_{n_1+2} = -1$, so $\sigma_{n_1+1} = \sigma_{n_2+2} = 1$ and for all $n \in [n_1 + 2, n_2 + 1]$, $\sigma_n = -1$. Then $f(x_{n_1}) - x_{n_1} \geq 0 \geq f(x_{n_1+1}) - x_{n_1+1}$ and $x_{n_1} \leq x_{n_1+1}$, so there is a fixed point of f in $[x_{n_1}, x_{n_1+1}]$. Let p be the least one (using here the continuity of f). By Lemma 3.3, $x_{n_1+1} - p \leq (1 - \delta)(p - x_{n_1}) \leq p - x_{n_1}$, which may also be written as $x_{n_1+1} - p \leq \frac{1}{2}(x_{n_1+1} - x_{n_1})$. Note that (x_n) is nonincreasing between $n_1 + 1$ and $n_2 + 1$.

Claim 1. We have that $x_{n_2} \geq p$.

Proof of claim 1: Assume that $p > x_{n_2}$. Then there is an $n' \in [n_1 + 1, n_2)$ with $x_{n'} \geq p > x_{n'+1}$. By Lemma 3.3, we have that

$$p - x_{n'+1} \leq (1 - \delta)(x_{n'} - p) \leq (1 - \delta)(x_{n_1+1} - p) \leq (1 - \delta)(p - x_{n_1}),$$

so

$$x_{n'+1} \geq (1 - \delta)x_{n_1} + \delta p \geq x_{n_1}.$$

Since $n' + 2 \leq n_2 + 1$, $\sigma_{n'+2} = -1$, so $f(x_{n'+1}) - x_{n'+1} \leq 0$. Then, since $f(x_{n_1}) - x_{n_1} \geq 0$, there is a fixed point q between x_{n_1} and $x_{n'+1}$. But then $x_{n_1} \leq q \leq x_{n'+1} < p \leq x_{n_1+1}$, which contradicts the minimality of p. ∎

Thus, using that either $x_{n_2+1} \geq p$ or $p \geq x_{n_2+1}$, we have that either $x_{n_2+1} \geq p \geq x_{n_1}$ or $x_{n_2} \geq p \geq x_{n_2+1}$.

Claim 2. We have that $x_{n_2+1} \geq (1 - \delta)x_{n_1} + \delta p$.

Proof of claim 2: In the first case above, the statement is obvious. Suppose now that $x_{n_2} \geq p \geq x_{n_2+1}$. Then, by Lemma 3.3,

$$p - x_{n_2+1} \leq (1 - \delta)(x_{n_2} - p),$$

so

$$x_{n_2+1} \geq (2 - \delta)p - (1 - \delta)x_{n_2}.$$

It remains to be shown that $(2 - \delta)p - (1 - \delta)x_{n_2} \geq (1 - \delta)x_{n_1} + \delta p$. Since $p - x_{n_1} \geq x_{n_1+1} - p \geq x_{n_2} - p$, $2p - x_{n_2} \geq x_{n_1}$, which we multiply by $(1 - \delta)$ to obtain the desired inequality. ∎

Let $n \in [n_1+1, n_2+1]$. Then $x_{n_1+1} \geq x_n \geq x_{n_2+1} \geq (1-\delta)x_{n_1} + \delta p \geq x_{n_1}$. Thus, we get (i).
We now prove (ii). We have that

$$x_{n_2} - x_{n_2+1} \leq x_{n_1+1} - x_{n_2+1} \leq (1-\delta)(x_{n_1+1} - x_{n_1}) + \delta(x_{n_1+1} - p)$$
$$\leq (1-\delta)(x_{n_1+1} - x_{n_1}) + \frac{\delta}{2}(x_{n_1+1} - x_{n_1}) = \left(1 - \frac{\delta}{2}\right)(x_{n_1+1} - x_{n_1}).$$

□

We may now turn to the proof of the theorem itself.

Proof of the theorem. Let (q_n) be the switching sequence for (x_n) relative to f. We distinguish two cases.
Case I. There is an r with $q_r = \infty$.
Take r to be minimal with this property. Clearly, $r \geq 1$ and $q_{r-1} < \infty$, so $(x_n)_{n=q_{r-1}}^{\infty}$ is monotone and hence convergent.
Case II. For all r, $q_r < \infty$.
We first show that, for all $r \geq 1$ and all $n \geq q_r$, x_n is between x_{q_r-1} and x_{q_r}. Let $r \geq 1$. We prove that, for all $s \geq r$ and all $n \in [q_s, q_{s+1}]$, x_n is between x_{q_r-1} and x_{q_r}. If $s = r$, this follows immediately from Lemma 3.5. Now let $s \geq r+1$. By the induction hypothesis, for all $m \in [q_{s-1}, q_s]$, x_m is between x_{q_r-1} and x_{q_r} – in particular, x_{q_s-1} and x_{q_s} are. By Lemma 3.5, x_n is between x_{q_s-1} and x_{q_s}, thus also between x_{q_r-1} and x_{q_r}.
Again by Lemma 3.5, we get that, for all $r \geq 1$, $|x_{q_{r+1}-1} - x_{q_{r+1}}| \leq \left(1 - \frac{\delta}{2}\right)|x_{q_r-1} - x_{q_r}|$ and thus, by an easy induction, for all $r \geq 1$, $|x_{q_r-1} - x_{q_r}| \leq \left(1 - \frac{\delta}{2}\right)^{r-1}$. Combining this with the result in the previous paragraph, we get that, for all $r \geq 1$ and all n, $m \geq q_r$, $|x_n - x_m| \leq \left(1 - \frac{\delta}{2}\right)^{r-1}$.
Let $T := \left\lceil \log_{\left(1-\frac{\delta}{2}\right)} \varepsilon \right\rceil + 1$ (note that T depends on δ and ε). Then we have that, for all n, $m \geq q_T$, $|x_n - x_m| \leq \left(1 - \frac{\delta}{2}\right)^{T-1} \leq \varepsilon$. Thus, (x_n) is Cauchy, hence convergent. □

We will now, for the first time, go deep into the actual workings of proof mining, and show how one can obtain a quantitative version of the proof above. The argumentation below closely follows my thought processes from January-February 2020, when I worked on this result.
We recall that our goal is to find a rate of metastability, so we will assume that we have an ε and a g and try to work with them.
The main point will be to 'finitize' the case distinction. Instead of having an r with $q_r = \infty$, and thus an infinite monotone sequence from one point on, we will only need a monotone fragment which is 'sufficiently large', in the spirit (and, hopefully, the letter) of the finite monotone convergence principle.
To give some intuition, consider the case where $q_1 > \widetilde{g}^{(\lceil 1/\varepsilon \rceil + 1)}(0)$. Then $(x_i)_{i=0}^{\widetilde{g}^{(\lceil 1/\varepsilon \rceil + 1)}(0)}$ is a finite monotone sequence and, thus, by the finite monotone convergence principle, there is an $N \leq \widetilde{g}^{(\lceil 1/\varepsilon \rceil)}(0)$ such that, for all n, $m \in [N, \widetilde{g}(N)]$, $|x_n - x_m| \leq \varepsilon$.
Now, look at the case where $q_1 \leq \widetilde{g}^{(\lceil 1/\varepsilon \rceil + 1)}(0)$. Take $P := \widetilde{g}^{(\lceil 1/\varepsilon \rceil + 1)}(0)$. We will cut off the terms of the sequence (x_n) before rank P, so we set, for every n, $y_n := x_{n-P}$. This will perturb the g in a way that we do not yet know, so assume that we work, for the sequence (y_n), with a different function h, whose form is to be determined. We will want, similarly to the above, to have that $(y_i)_{i=0}^{\widetilde{h}^{(\lceil 1/\varepsilon \rceil + 1)}(0)}$ is a finite monotone sequence, so we will want to have that $P + \widetilde{h}^{(\lceil 1/\varepsilon \rceil + 1)}(0) < q_2$. We set $P_2 := P + \widetilde{h}^{(\lceil 1/\varepsilon \rceil + 1)}(0)$. By the finite monotone convergence principle, there is an $N' \leq \widetilde{h}^{(\lceil 1/\varepsilon \rceil)}(0)$ such that, for all n, $m \in [N', \widetilde{h}(N')]$, $|y_n - y_m| \leq \varepsilon$. If we set $N := P + N'$, we have that

$$N \leq P + \widetilde{h}^{(\lceil 1/\varepsilon \rceil)}(0) \leq P + \widetilde{h}^{(\lceil 1/\varepsilon \rceil + 1)}(0) = P_2,$$

so there is an $N \leq P_2$ such that, for all n, $m \in [N, P + \widetilde{h}(N')]$, $|x_n - x_m| \leq \varepsilon$. We would want that $P + \widetilde{h}(N') = \widetilde{g}(N)$, i.e. $P + N' + h(N') = P + N' + g(P + N')$. This can be arranged if we take, for any n, $h(n) := g(P + n)$. We will generally consider a family of such functions (h_m), where, for every m and n, $h_m(n) := g(m + n)$, so the h we consider here is in fact h_P. In particular, $P_2 = P + \widetilde{h_P}^{(\lceil 1/\varepsilon \rceil + 1)}(0)$.
We will now consider the most general case of the argument above. Put, for every r, $P_{r+1} = P_r + \widetilde{h_{P_r}}^{(\lceil 1/\varepsilon \rceil + 1)}(0)$. We notice that $P = 0 + \widetilde{h_0}^{(\lceil 1/\varepsilon \rceil + 1)}(0)$, so we can put $P_0 := 0$ and then $P = P_1$.

The cases which we considered before where the cases where $q_1 > P_1$, and the one where $q_2 > P_2$. We will now consider the general case where there is an r such that $q_r > P_r$, and we take the minimal such one. Since $q_0 = 0$, we must have $r \geq 1$, so $P_r = P_{r-1} + \widetilde{h_{P_{r-1}}}^{(\lceil 1/\varepsilon \rceil + 1)}(0)$. In addition, we must have $q_{r-1} \leq P_{r-1}$. We have, therefore, that $(x_{P_{r-1}+i})_{i=0}^{\widetilde{h_{P_{r-1}}}^{(\lceil 1/\varepsilon \rceil + 1)}(0)}$ is a finite monotone sequence and, thus, by the finite monotone convergence principle, there is an $N' \leq \widetilde{h_{P_{r-1}}}^{(\lceil 1/\varepsilon \rceil)}(0)$ such that, for all n, $m \in [P_{r-1} + N', P_{r-1} + \widetilde{h_{P_{r-1}}}(N')]$, $|x_n - x_m| \leq \varepsilon$. Now, if we take $N := P_{r-1} + N'$, we have that

$$N \leq P_{r-1} + \widetilde{h_{P_{r-1}}}^{(\lceil 1/\varepsilon \rceil)}(0) = P_r$$

and

$$P_{r-1} + \widetilde{h_{P_{r-1}}}(N') = P_{r-1} + N' + h_{P_{r-1}}(N') = N + g(P_{r-1} + N') = N + g(N) = \widetilde{g}(N),$$

so, for all n, $m \in [N, \widetilde{g}(N)]$, $|x_n - x_m| \leq \varepsilon$.

We know, thus, that, if there is an r such that $q_r > P_r$, then the N we need is bounded above by P_r. If we just assume the existence of an r, this does not help us at all in establishing a bound on the N, so we need to do a 'second finitization', meaning that we would like this r to be bounded by a quantity B, for then we would have $N \leq P_r \leq P_B$. This B is still unknown, and in order to determine it, we will have to look at the 'finitary' case which corresponds to the second case in the original proof of the theorem, which will take the form: for every $r \leq B$, $q_r \leq P_r$. Making the proof of this case go through will show us how we may choose the B.

Using the same kind of argument as before, we get that, for every r with $1 \leq r < B$ and every $n \in [q_r, q_B]$, x_n is between $x_{q_r - 1}$ and x_{q_r}, and then that, for every r with $1 \leq r < B$ and all n, $m \in [q_r, q_B]$, $|x_n - x_m| \leq \left(1 - \frac{\delta}{2}\right)^{r-1}$. We would also like to get, using the final argument from the qualitative proof, that, for all n, $m \in [q_T, q_B]$, $|x_n - x_m| \leq \left(1 - \frac{\delta}{2}\right)^{r-1}$ (remember that T was an exact quantity given in terms of δ and ε). Thus, we need that $T < B$, so $B \geq T + 1$.

In the end, we want to get an N such that, for all n, $m \in [N, \widetilde{g}(N)]$, $|x_n - x_m| \leq \varepsilon$ (of course, together with a tractable bound on the N). For that, using the above, we only need that $[q_T, q_B] \supseteq [N, \widetilde{g}(N)]$, i.e. $q_T \leq N \leq \widetilde{g}(N) \leq q_B$.

We will take $N := P_T$. Since $T \leq B$, $P_T \leq P_B$.

Since, for all $r \leq B$, $q_r \leq P_r$, and $T \leq B$, $q_T \leq P_T = N$. We only need to guarantee that $\widetilde{g}(N) \leq q_B$, i.e. that $\widetilde{g}(P_T) \leq q_B$. Since $B \leq q_B$, we only need that $\widetilde{g}(P_T) \leq B$.

Combining this requirement with the one from before, we see that, if we take $B := T + \widetilde{g}(P_T) + 1$, we are done.

Note that, in both 'finitary' cases, P_B is a bound on the N, so this will be our desired rate of metastability, which depends on the ε, the g and the δ.

This final quantitative result and its proof are expressed in the paper [10] as follows.

Theorem 3.6. *Let $L > 0$, $f : [0,1] \to [0,1]$ be L-Lipschitz, (t_n) and (x_n) be sequences in $[0,1]$ such that for all n, $x_{n+1} = (1 - t_n)x_n + t_n f(x_n)$.*

Define, for any suitable ε, g, δ, m, n:

$$\begin{aligned}
h_m^g(n) &:= g(m+n) \\
P_0^{\varepsilon,g} &:= 0 \\
P_{n+1}^{\varepsilon,g} &:= P_n^{\varepsilon,g} + \widetilde{h_{P_n^{\varepsilon,g}}^g}^{(\lceil \frac{1}{\varepsilon} \rceil + 1)}(0) \\
T_{\varepsilon,\delta} &:= \left\lceil \log_{\left(1 - \frac{\delta}{2}\right)} \varepsilon \right\rceil + 1 \\
B_{\varepsilon,g,\delta} &:= T_{\varepsilon,\delta} + \widetilde{g}\left(P_{T_{\varepsilon,\delta}}^{\varepsilon,g}\right) + 1 \\
\Psi_\delta^{\mathrm{KM}}(\varepsilon, g) &:= P_{B_{\varepsilon,g,\delta}}^{\varepsilon,g}.
\end{aligned}$$

Let $\delta \in (0,1)$ be such that for all n, $t_n \leq \frac{2-\delta}{L+1}$.

Let $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$. Then there is an $N \leq \Psi_\delta^{\mathrm{KM}}(\varepsilon, g)$ such that, for all n, $m \in [N, N + g(N)]$, $|x_n - x_m| \leq \varepsilon$.

Proof. We may now drop ε, g, δ where they show up as indices or arguments. It is immediate that:

- for all n, $P_n \leq P_{n+1}$;
- $\left(1 - \frac{\delta}{2}\right)^{T-1} \leq \varepsilon$.

Let (q_n) be the switching sequence for (x_n) relative to f. Note that (q_n) is strictly increasing and for all r, $r \leq q_r$ and if $q_r < \infty$, then $(x_n)_{n \in [q_r, q_{r+1})}$ is monotone. We distinguish two cases.

Case I. There is an $r \leq B$ with $q_r > P_r = P_{r-1} + \widetilde{h_{P_{r-1}}}^{(\lceil \frac{1}{\varepsilon} \rceil + 1)}(0)$.

Take r to be minimal with this property. Clearly, $r \geq 1$ and $q_{r-1} \leq P_{r-1}$, so

$$(x_{P_{r-1}+i})_{i=0}^{\widetilde{h_{P_{r-1}}}^{(\lceil \frac{1}{\varepsilon} \rceil + 1)}(0)}$$

is a subsequence of $(x_n)_{n \in [q_{r-1}, q_r)}$ and is thus monotone. By Theorem 2.3, there is an $N' \leq \widetilde{h_{P_{r-1}}}^{(\lceil \frac{1}{\varepsilon} \rceil)}(0)$ such that, for all n, $m \in [P_{r-1} + N', P_{r-1} + N' + h_{P_{r-1}}(N')]$, $|x_n - x_m| \leq \varepsilon$.

Put $N := P_{r-1} + N'$. Then

$$N \leq P_{r-1} + \widetilde{h_{P_{r-1}}}^{(\lceil \frac{1}{\varepsilon} \rceil)}(0) \leq P_{r-1} + \widetilde{h_{P_{r-1}}}^{(\lceil \frac{1}{\varepsilon} \rceil + 1)}(0) = P_r \leq P_B = \Psi^{\mathrm{KM}}.$$

In addition,

$$h_{P_{r-1}}(N') = g(P_{r-1} + N') = g(N),$$

so, for all n, $m \in [N, N + g(N)]$, $|x_n - x_m| \leq \varepsilon$.

Case II. For all $r \leq B$, $q_r \leq P_r$.

We first show that for all $r \in [1, B-1]$ and all $n \in [q_r, q_B]$, x_n is between $x_{q_r - 1}$ and x_{q_r}. Let $r \in [1, B-1]$. We prove that for all $s \in [r, B-1]$ and all $n \in [q_s, q_{s+1}]$, x_n is between $x_{q_r - 1}$ and x_{q_r}. If $s = r$, this follows immediately from Lemma 3.5.(i). Now let $s \geq r+1$. By the induction hypothesis, for all $m \in [q_{s-1}, q_s]$, x_m is between $x_{q_r - 1}$ and x_{q_r} – in particular, $x_{q_s - 1}$ and x_{q_s} are. By Lemma 3.5.(i), x_n is between $x_{q_s - 1}$ and x_{q_s}, thus also between $x_{q_r - 1}$ and x_{q_r}.

By Lemma 3.5.(ii), we get that for all $r \in [1, B-1]$, $|x_{q_{r+1}-1} - x_{q_{r+1}}| \leq \left(1 - \frac{\delta}{2}\right) |x_{q_r - 1} - x_{q_r}|$ and thus, by an easy induction, for all $r \in [1, B-1]$, $|x_{q_r - 1} - x_{q_r}| \leq \left(1 - \frac{\delta}{2}\right)^{r-1}$. Combining this with the result in the previous paragraph, we get that, for all $r \in [1, B-1]$ and all n, $m \in [q_r, q_B]$, $|x_n - x_m| \leq \left(1 - \frac{\delta}{2}\right)^{r-1}$.

Since $T \leq B - 1$, for all n, $m \in [q_T, q_B]$, $|x_n - x_m| \leq \varepsilon$. Take $N := P_T \leq P_B = \Psi^{\mathrm{KM}}$. Then on one hand $N = P_T \geq q_T$ (since $T \leq B$) and on the other $N + g(N) = \widetilde{g}(P_T) \leq T + \widetilde{g}(P_T) + 1 = B \leq q_B$, so $[N, N + g(N)] \subseteq [q_T, q_B]$, hence, for all n, $m \in [N, N + g(N)]$, $|x_n - x_m| \leq \varepsilon$. □

I will now close with some remarks (originating in discussions with Ulrich Kohlenbach following the delivery of this tutorial) concerning the formalization of the theorem analysed above. Since the theorem does not concern metric structures, one does not need the machinery of abstract types involved in the 'general logical metatheorems' of [4, 2] and [5, Chapter 17], but rather the older metatheorem for Polish spaces [5, Theorem 15.1] (as originally proven in [3]). One only needs to check that the objects mentioned in the theorem – the function f, the sequences (x_n) and (t_n) and the numbers L and δ are suitably encodable. The sequences are elements of the compact space $[0,1]^{\mathbb{N}}$ (and hence the bound will not depend on them), while L and δ are elements of a Polish space. In addition, In order to represent f as an element of a Polish space, one considers (see [5, p. 83]) its restriction to the rational numbers, since the whole f may be recovered from it together with a modulus of uniform continuity, which is here immediately given by the L. Note also that one may replace the quantity δ by a natural number l, taking $\delta := 2^{-l}$, since one only has to witness that $\delta > 0$.

One then has to give a logical explanation for the independence of the final rate on the L. First of all, within the proof itself, the assumption that f is L-Lipschitz (together with the fact that, for all n, $t_n \leq \frac{2-\delta}{L+1}$) is only used to show that, for all n, and for all x, $y \in [0,1]$,

$$(t_n - 1)|x - y| + t_n|f(x) - f(y)| \leq (1 - 2^{-l})|x - y|,$$

which may be considered as a universal assumption. Then the only issue left to be considered is the use of L in the representation of f by way of a modulus of uniform continuity, as mentioned above. We fix an ε and a g, considering them as parameters of the problem. In the case that, for all $n \leq \widetilde{g}(0)$, $t_n < 2^{-n} \cdot \varepsilon$, we have that, for all $n \leq \widetilde{g}(0)$, $|x_{n+1} - x_n| < 2^{-n} \cdot \varepsilon$, so, for all n, $m \in [0, \widetilde{g}(0)]$, $|x_n - x_m| \leq \varepsilon$ – thus, in this case, any number may serve as a rate of metastability. We may assume, therefore, that there is an $n \leq \widetilde{g}(0)$ with $t_n \geq 2^{-n} \cdot \varepsilon \geq 2^{-\widetilde{g}(0)} \cdot \varepsilon$. Using the equation from before, we get that, for all x, $y \in [0,1]$,

$$2^{-\widetilde{g}(0)} \cdot \varepsilon |f(x) - f(y)| \leq t_n|f(x) - f(y)| \leq (2 - 2^{-l} - t_n)|x - y| \leq (2 - 2^{-l})|x - y|,$$

from which we get that, for all x, $y \in [0,1]$,

$$|f(x) - f(y)| \leq \frac{2^{\widetilde{g}(0)}(2 - 2^{-l})}{\varepsilon} \cdot |x - y|,$$

thus giving a Lipschitz constant for f.

References

[1] D. Borwein, J. Borwein, Fixed point iterations for real functions. *Journal of Mathematical Analysis and Applications* 157, no. 1, 112–126, 1991.

[2] P. Gerhardy, U. Kohlenbach, General logical metatheorems for functional analysis. *Transactions of the American Mathematical Society* 360, 2615–2660, 2008.

[3] U. Kohlenbach, Effective moduli from ineffective uniqueness proofs. An unwinding of de la Vallée Poussin's proof for Chebycheff approximation. *Annals of Pure and Applied Logic* 64, no. 1, 27–94, 1993.

[4] U. Kohlenbach, Some logical metatheorems with applications in functional analysis. *Transactions of the American Mathematical Society* vol. 357, no. 1, 89–128, 2005.

[5] U. Kohlenbach, *Applied proof theory: Proof interpretations and their use in mathematics*. Springer Monographs in Mathematics, Springer, 2008.

[6] U. Kohlenbach, Proof-theoretic methods in nonlinear analysis. In: B. Sirakov, P. Ney de Souza, M. Viana (eds.), *Proceedings of the International Congress of Mathematicians 2018 (ICM 2018)*, Vol. 2 (pp. 61–82). World Scientific, 2019.

[7] U. Kohlenbach, Local formalizations in nonlinear analysis and related areas and proof-theoretic tameness. In: P. Weingartner, H.-P. Leeb (eds.), *Kreisel's Interests. On the Foundations of Logic and Mathematics*. College Publications. Tributes vol. 41 (pp. 45–61), 2020.

[8] G. Kreisel, On the interpretation of non-finitist proofs, part II: Interpretation of number theory, applications. *Journal of Symbolic Logic* 17, 43–58, 1952.

[9] A. Sipoş, What proof mining is about, Part I. Blog post, 2020. Available online at: https://prooftheory.blog/2020/06/06/what-proof-mining-is-about-part-i/.

[10] A. Sipoş, Rates of metastability for iterations on the unit interval. *Journal of Mathematical Analysis and Applications*, Volume 502, Issue 1, 125235 [11 pages], 2021.

[11] A. Sipoş, On extracting variable Herbrand disjunctions. *Studia Logica*, Volume 110, Issue 4, 1115–1134, 2022.

[12] T. Tao, Soft analysis, hard analysis, and the finite convergence principle. Essay posted May 23, 2007. Appeared in: T. Tao, *Structure and Randomness: Pages from Year One of a Mathematical Blog*. AMS, 298 pp., 2008.

[13] T. Tao, Norm convergence of multiple ergodic averages for commuting transformations. *Ergodic Theory & Dynamical Systems* 28, 657–688, 2008.