Advances in proof mining

Andrei Sipoș

September 14, 2020
University of Bucharest Lectureship Competition
București, România

Proof mining

This, as the title says, is a talk on *proof mining*:

- an applied subfield of mathematical logic
- first suggested by G. Kreisel in the 1950s (under the name "proof unwinding"), then given maturity by U. Kohlenbach and his collaborators starting in the 1990s
- goals: to find explicit and uniform witnesses or bounds and to remove superfluous premises from concrete mathematical statements by analyzing their proofs
- tools used: primarily proof interpretations (modified realizability, negative translation, functional interpretation)
- the adequacy of the tools to the goals is guaranteed by *general logical metatheorems*

I also want to advertise the

Proof Theory Blog

- available at https://prooftheory.blog/
- an initiative of Anupam Das and Thomas Powell
- welcoming of contributions from all sorts of researchers, in or adjacent to proof theory
- I am currently writing an ongoing series on proof mining, entitled:

What proof mining is about

The general situation

In nonlinear analysis and optimization, one is typically given a metric space (X, d)...
(you can imagine e.g. a Hilbert space – since that is often the case)
...and wants to find some special kind of point in it, let's say a fixed point of a self-mapping $T : X \to X$.
We denote the fixed point set of T by $Fix(T)$.

Iterations

One typically does this by building iterative sequences (x_n), e.g. the *Picard iteration*: let $x \in X$ be arbitrary and set for any n, $x_n := T^n x$. We know that if T is a contraction, this converges strongly to a fixed point of T, but in other cases we'll have only weaker forms of convergence...

...like weak convergence itself...

...but most importantly **asymptotic regularity**:

$$\lim_{n \to \infty} d(x_n, Tx_n) = 0.$$

Intuition:

- convergence: "close to a fixed point"
- asymptotic regularity: "close to *being* a fixed point"
  (the iteration is then an **approximate fixed point sequence**)

Rates

In the case of asymptotic regularity:

$$\forall \varepsilon \; \exists N \; \forall n \geq N \; d(x_n, Tx_n) \leq \varepsilon.$$

what proof mining seeks is to find a *rate of asymptotic regularity*: an explicit formula for N in terms of the ε and of (as few as possible of) the other parameters of the problem.

The statement is $\forall\exists\forall$, a case generally excluded by the metatheorems which pertain to classical logic, and rightfully so, since there exist explicit examples ("Specker sequences") of sequences of computable reals with no computable limit and thus with no computable rate of convergence.

Rates continued

In some cases, however, the sequence $(d(x_n, Tx_n))$ is nonincreasing, which gets rid of the last $\forall$. In others, like the one in the sequel, more work is needed.

At the other end of logical complexity, purely universal sentences help us when they show up in proofs that we're analyzing since they lack computational content and thus it doesn't matter whether their subproofs conform to the requirements of the metatheorems (an observation first due to Kreisel).

Consistent feasibility problems

Consider now a subset C of X and a family of mappings $(T_i : C \to C)_{1 \leq i \leq N}$. Assume (at first) that

$$\bigcap_{i=1}^{N} Fix(T_i) \neq \emptyset$$

and that the problem at hand (a *consistent feasibility* or *image recovery problem*) is to find a point in that set.

Usually, what one does is to consider either of the following two constructions:

- a convex combination $T := \sum_{i=1}^{N} \lambda_i T_i$ leading to what is called the *parallel algorithm* or the *method of averaged projections*
- a composition $T := T_N \circ \ldots \circ T_1$ – the *cyclic algorithm* or the *method of alternating projections*

then prove that $Fix(T)$ is equal to the set above and apply a common iteration (Picard, Mann) to T.

More general feasibility problems

One may still consider iterating T in the case where

$$\bigcap_{i=1}^{N} Fix(T_i) = \emptyset.$$

Here, we distinguish two cases:

- *intermediate feasibility* where $Fix(T) \neq \emptyset$ (the interesting case is where T does not inherit any interesting properties of the T_i's)
- *inconsistent feasibility* where $Fix(T) = \emptyset$ (one has to assume, though, an approximate fixed point condition, the showing of which being usually the whole meat of the proof, as we shall see later)

Classes of mappings

We will work with the following two conditions on a map $T : C \to C$.

Definition

The map T is called **nonexpansive** if for all $x, y \in C$, we have that $d(Tx, Ty) \leq d(x, y)$.

The definition below assumes that X is a Hilbert space.

Definition (Browder and Petryshyn, 1967)

Let $k \in [0, 1)$. The map T is called k-**strictly pseudocontractive** if for all $x, y \in C$, we have that:

$$\|Tx - Ty\|^2 \leq \|x - y\|^2 + k\|(x - Tx) - (y - Ty)\|^2.$$

We can see that nonexpansive $\Leftrightarrow$ 0-strictly pseudocontractive.

The result of López-Acedo and Xu

The concrete result that we shall analyse is the following.

Theorem (López-Acedo and Xu, 2007)

Assume X is Hilbert. Let $k \in [0, 1)$ and suppose that each T_i is k-strictly pseudocontractive, with $\bigcap_{i=1}^{N} Fix(T_i) \neq \emptyset$. Let (λ_i^n) and (t_n) be such that:

$$\sum_{n=0}^{\infty} (t_n - k)(1 - t_n) = \infty, \qquad \sum_{j=0}^{\infty} \sqrt{\sum_{i=1}^{N} |\lambda_i^{(j+1)} - \lambda_i^{(j)}|} < \infty.$$

Then any sequence (x_n) that satisfies

$$x_{n+1} := t_n x_n + (1 - t_n) \sum_{i=1}^{N} \lambda_i^{(n)} T_i x_n$$

is, for each i, T_i-asymptotically regular.

What is interesting about this proof is the fact (noticed before by Leuștean in a paper of Tan/Xu) that it consists of two parts:

- a non-constructive part where it is proven that the corresponding limit inferior is 0, which is a $\forall\exists$ statement
  - thus, one might extract a **modulus of liminf**
- a constructive part which uses the above where the actual asymptotic regularity is shown – as seen before, this is a $\forall\exists\forall$ statement
  - thus, one might extract a rate of asymptotic regularity by plugging in the modulus of liminf

This is what we did (see Ann. Pure Applied Logic, 2017).

Nonlinear spaces

In addition to linear spaces such as Hilbert or Banach spaces, there has recently been a renewed focus in nonlinear spaces.
We say that:

- a **geodesic** in X is a mapping $\gamma : [0,1] \to X$ such that for any $t, t' \in [0,1]$ we have that

$$d(\gamma(t), \gamma(t')) = |t - t'| d(\gamma(0), \gamma(1))$$

- X is **geodesic** if any two points of it are joined by a geodesic
- X is **CAT(0)** if it is geodesic and for any geodesic $\gamma : [0,1] \to X$ and for any $z \in X$ and $t \in [0,1]$ we have that

$$d^2(z, \gamma(t)) \leq (1-t) d^2(z, \gamma(0)) + t d^2(z, \gamma(1)) - t(1-t) d^2(\gamma(0), \gamma(1))$$

Intuition: curvature at most 0.
Also: CAT(0) spaces are *uniquely* geodesic, so denote $\gamma(t)$ by $(1-t)\gamma(0) + t\gamma(1)$.

Firmly nonexpansive mappings

Assume now that (X, d) is a CAT(0) space. We call a map $T : X \to X$ to be **firmly nonexpansive** if for any $x, y \in X$ and any $t \in [0, 1]$ we have that

$$d(Tx, Ty) \leq d((1-t)x + tTx, (1-t)y + tTy).$$

- important in convex optimization, as primary examples include:
  - projections onto closed, convex, nonempty subsets
  - resolvents (of nonexpansive mappings, of convex lsc functions)
- introduced in a nonlinear context by Ariza-Ruiz/Leuştean/López-Acedo (Trans. AMS 2014)
- they satisfy the slightly weaker property (P_2) (though equivalent to f.n.e. in Hilbert spaces): for all $x, y \in X$,

  $$2d^2(Tx, Ty) \leq d^2(x, Ty) + d^2(y, Tx) - d^2(x, Tx) - d^2(y, Ty)$$

- in particular, even (P_2) implies nonexpansiveness: for any $x, y \in X$, $d(Tx, Ty) \leq d(x, y)$

Results in CAT(0) spaces

The problem of intermediate feasibility was studied in CAT(0) spaces only for $n = 2$, and for mappings satisfying property (P_2) in the case of the cyclic algorithm:

- asymptotic regularity: Ariza-Ruiz/López-Acedo/Nicolae (JOTA 2015)
- an explicit rate: Kohlenbach/López-Acedo/Nicolae (Optimization 2017)

If one defines (where b is a bound on the distance between the initial point x and a given fixed point p):

$$k_b(\varepsilon) := \left\lceil \frac{2b}{\varepsilon} \right\rceil, \qquad \Phi_b(\varepsilon) := k_b(\varepsilon) \cdot \left\lceil \frac{2b(1 + 2^{k_b(\varepsilon)})}{\varepsilon} \right\rceil + 1,$$

then the rate (as N in terms of an ε) is given by $\Phi_b(\varepsilon)$.

The trick

A trick that has been used in Hilbert spaces to pass from compositions to convex combinations was to put on X^n the following scalar product that makes it into a Hilbert space:

$$\langle (x_1, ..., x_n), (y_1, ..., y_n) \rangle := \sum_{i=1}^{n} \lambda_i \langle x_i, y_i \rangle.$$

The diagonal of X^n, denoted by Δ_X, is, then, a subspace isometric to X. If we put Q to be the projection onto Δ_X and U to be the operator given by $U(x_1, ..., x_n) := (T_1 x_1, ..., T_n x_n)$, then one sees that $Q \circ U$ is an operator on Δ_X that is the pushforward by isometry of T.

This idea originated with Pierra in 1984.

Moving to CAT(0) spaces

Our goal: to adapt it to CAT(0) spaces in order to study the intermediate feasibility problem for the same case (with $n = 2$).

Let (X, d) be a metric space and $\lambda \in (0, 1)$. We define $d_\lambda : X^2 \times X^2 \to \mathbb{R}_+$, for any $(x_1, x_2), (y_1, y_2) \in X^2$ by:

$$d_\lambda((x_1, x_2), (y_1, y_2)) := \sqrt{(1-\lambda)d^2(x_1, y_1) + \lambda d^2(x_2, y_2)}.$$

Then:

- (X^2, d_λ) is a metric space
- if (X, d) is complete, geodesic or CAT(0), then (X^2, d_λ) is also complete, geodesic or CAT(0), respectively

Asymptotic regularity

Therefore, let $T_1, T_2 : X \to X$ be (P_2) mappings and set $T := (1 - \lambda)T_1 + \lambda T_2$. Then, by carefully using the geodesic structure of CAT(0) spaces, one may define operators Q and U similar to the ones presented before and prove that they satisfy the required properties.

Thus, by applying the corresponding result for alternating projections, one can prove that if $Fix(T) \neq \emptyset$, then T is asymptotically regular with morally the same rate obtained by Kohlenbach/López-Acedo/Nicolae for compositions.

These results appeared in Journal of Convex Analysis, 2018.

A more elaborate problem

Let us consider now the inconsistent feasibility case that we mentioned before, in a Hilbert space X.
Consider, then, $n \geq 1$ and let $C_1, \ldots, C_n$ be closed, convex, nonempty subsets of X with a not necessarily nonempty intersection.
(Of course, one doesn't care here about convergence, since there may be nothing interesting to converge to...)
Conjecture (Bauschke/Borwein/Lewis '95): asymptotic regularity still holds.

This was proved by Bauschke (Proc. AMS '03).

More developments

The result of Bauschke was then generalized, as mentioned before:

- from projections onto convex sets to firmly nonexpansive mappings
  - a well-behaved class of mappings which is important in convex optimization, as primary examples include:
    - projections onto closed, convex, nonempty subsets
    - resolvents (of nonexpansive mappings, of convex lsc functions)
  - P_C becomes R, C becomes $Fix(R)$
  - one assumes even less: each mapping needs to have only approximate fixed points
  - this was done by Bauschke/Martín-Márquez/Moffat/Wang in 2012
- even more, from firmly nonexpansive mappings to α-averaged mappings – where $\alpha \in (0, 1)$
  - done by Bauschke/Moursi in 2018
  - firmly nonexpansive mappings are exactly $\frac{1}{2}$-averaged mappings

Past work

Kohlenbach analyzed (Found. Comput. Math., 2019) the proofs of Bauschke '03 and of Bauschke/Martín-Márquez/Moffat/Wang '12.

These proofs are organized as follows:

- first one shows that T has arbitrarily small displacements $\forall \varepsilon \, \exists p \, \|p - Tp\| \leq \varepsilon$
- this fact in conjunction with the fact that T is strongly nonexpansive yields asymptotic regularity (Bruck/Reich '77)
  - strongly nonexpansive mappings subsume firmly nonexpansive mappings and are closed under composition

The analysis of the second part relies on previous work of Kohlenbach on strongly nonexpansive mappings (Israel J. Math., 2016).

On strongly nonexpansive mappings I

Definition (Kohlenbach, 2016)

Let $T : X \to X$ and $\omega : (0, \infty) \times (0, \infty) \to (0, \infty)$. Then T is called **strongly nonexpansive** with modulus ω if for any b, $\varepsilon > 0$ and x, $y \in X$ with $\|x - y\| \leq b$ and $\|x - y\| - \|Tx - Ty\| < \omega(b, \varepsilon)$, we have that $\|(x - y) - (Tx - Ty)\| < \varepsilon$.

On strongly nonexpansive mappings II

Theorem (Kohlenbach, FoCM 2019)

Define, for any ε, b, $d > 0$, $\alpha : (0,\infty) \to (0,\infty)$ *and* $\omega : (0,\infty) \times (0,\infty) \to (0,\infty)$,

$$\varphi(\varepsilon, b, d, \alpha, \omega) := \left\lceil \frac{18b + 12\alpha(\varepsilon/6)}{\varepsilon} - 1 \right\rceil \cdot \left\lceil \frac{d}{\omega\left(d, \frac{\varepsilon^2}{27b+18\alpha(\varepsilon/6)}\right)} \right\rceil.$$

Let $T : X \to X$ *and* $\omega : (0,\infty) \times (0,\infty) \to (0,\infty)$ *be such that* T *is strongly nonexpansive with modulus* ω. *Let* $\alpha : (0,\infty) \to (0,\infty)$ *such that for any* $\delta > 0$ *there is a* $p \in X$ *with* $\|p\| \leq \alpha(\delta)$ *and* $\|p - Tp\| \leq \delta$. *Then for any* ε, b, $d > 0$ *and any* $x \in X$ *with* $\|x\| \leq b$ *and* $\|x - Tx\| \leq d$, *we have that for any* $n \geq \varphi(\varepsilon, b, d, \alpha, \omega)$, $\|T^n x - T^{n+1} x\| \leq \varepsilon$.

Thus, one needs a bound on the p obtained in the first part and a SNE-modulus for T.

Extraction details

The first part of the proof provides the most intricate portion of the analysis, since it uses deep results such as Minty's theorem.

Fortunately, these kinds of arguments only enter the proof through $\forall$-lemmas, so the resulting rate is of low complexity (polynomial of degree eight).

What we did was to update these techniques in order to analyze the proof of Bauschke/Moursi '18 for averaged mappings.

The rate of asymptotic regularity

Our final result looked like this.

Theorem (A.S., 2020, arXiv:2001.01513)

Define, for all $m \geq 2$, ε, b, $d > 0$, $K : (0, \infty) \to (0, \infty)$ and $\{\alpha_i\}_{i=1}^m \subseteq (0,1)$, $\Sigma_{m,\{\alpha_i\}_{i=1}^m,K,b,d}(\varepsilon)$ to be

$$\varphi(\varepsilon, b, d, \omega_{\alpha_1 \star \ldots \star \alpha_m}, \delta \mapsto \Psi(m, \{\alpha_i\}_{i=1}^m, K, \delta)).$$

Let $m \geq 2$, $\alpha_1, \ldots, \alpha_m \in (0,1)$ and $R_1, \ldots, R_m : X \to X$ such that for each i, R_i is α_i-averaged. Put $R := R_m \circ \ldots \circ R_1$. Let $K : (0, \infty) \to (0, \infty)$ be such that for all i and all $\varepsilon > 0$ there is a $p \in X$ with $\|p\| \leq K(\varepsilon)$ and $\|p - R_i p\| \leq \varepsilon$.
Then for any b, $d > 0$ and any $x \in X$ with $\|x\| \leq b$ and $\|x - Rx\| \leq d$, we have that for any $\varepsilon > 0$ and $n \geq \Sigma_{m,\{\alpha_i\}_{i=1}^m,K,b,d}(\varepsilon)$, $\|R^n x - R^{n+1} x\| \leq \varepsilon$.

Jointly firmly nonexpansive mappings

As we've seen, firmly nonexpansive mappings unify various important concepts from convex optimization.

We have gone further with this abstraction and introduced *jointly firmly nonexpansive* families of mappings.

This allowed us to form abstract versions of fundamental tools of convex optimization like the *proximal point algorithm* or *approximating curves*.

(Part of this is joint work with Leuştean and Nicolae and has appeared in 2018 in Journal of Global Optimization; part of it is from 2020 and may be found at arXiv:2006.02167.)

We work in a CAT(0) space X.

If T and U are self-mappings of X and λ, $\mu > 0$, we say that T and U are (λ, μ)-*mutually firmly nonexpansive* if for all x, $y \in X$ and all α, $\beta \in [0,1]$ such that $(1-\alpha)\lambda = (1-\beta)\mu$, one has that

$$d(Tx, Uy) \leq d((1-\alpha)x + \alpha Tx, (1-\beta)y + \beta Uy).$$

If $(T_n)_{n\in\mathbb{N}}$ is a family of self-mappings of X and $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$, we say that (T_n) is *jointly firmly nonexpansive* with respect to (γ_n) if for all n, $m \in \mathbb{N}$, T_n and T_m are (γ_n, γ_m)-mutually firmly nonexpansive.

In addition, if $(T_\gamma)_{\gamma>0}$ is a family of self-mappings of X, we say that it is plainly *jointly firmly nonexpansive* if for all λ, $\mu > 0$, T_λ and T_μ are (λ, μ)-mutually firmly nonexpansive.

It is clear that a family (T_γ) is jointly firmly nonexpansive if and only if for every $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$, $(T_{\gamma_n})_{n\in\mathbb{N}}$ is jointly firmly nonexpansive with respect to (γ_n).

Examples

It was shown that examples of jointly firmly nonexpansive families of mappings are furnished by resolvent-type mappings used in convex optimization – specifically, by:

- the family $(J_{\gamma f})_{\gamma > 0}$, where f is a proper convex lower semicontinous function on X and one denotes for any such function g its proximal mapping by J_g;
- the family $(R_{T,\gamma})_{\gamma > 0}$, where T is a nonexpansive self-mapping of X and one denotes, for any $\gamma > 0$, its resolvent of order γ by $R_{T,\gamma}$;
- (if X is a Hilbert space) the family $(J_{\gamma A})_{\gamma > 0}$, where A is a maximally monotone operator on X and one denotes for any such operator B its resolvent by J_B.

Convergence

Our main theorem from the 2018 paper was the following.

Theorem (Leuștean, Nicolae, A.S.)

Assume that X *is complete and let* $T_n : X \to X$ *for every* $n \in \mathbb{N}$ *and* (γ_n) *be a sequence of positive real numbers satisfying* $\sum_{n=0}^{\infty} \gamma_n^2 = \infty$. *Assume that the family* (T_n) *is jointly firmly nonexpansive with respect to* (γ_n) *and that* $F := \bigcap_{n \in \mathbb{N}} Fix(T_n) \neq \emptyset$. *Let* (x_n) *be such that for any* n, $x_{n+1} = T_n x_n$.

Then (x_n) Δ*-converges to a point in* F.

The relationship

Recently, we discovered the following link to the resolvent identity.

Theorem (A.S., 2020)

Let $(T_\gamma)_{\gamma>0}$ *is a family of self-mappings of* X. *Then the following are equivalent:*

- 1. *For all* $\gamma > 0$, T_γ *is nonexpansive.*
  2. *For all* $\gamma > 0$, $t \in [0,1]$ *and* $x \in X$,

$$T_{(1-t)\gamma}((1-t)x + tT_\gamma x) = T_\gamma x.$$

- $(T_\gamma)_{\gamma>0}$ *is jointly firmly nonexpansive.*

The uniform case

In the case where the resolvent mappings arise from a uniform object (e.g. uniformly monotone operator, uniformly convex function), they are **uniformly firmly nonexpansive**, a condition which looks like

$$d^2(Tx, Ty) \leq d^2((1-t)x + tTx, (1-t)y + tTy) - 2(1-t)\varphi(\varepsilon)$$

and the corresponding optimizing point is unique.

In this case, using ideas by Kohlenbach '90 and Kohlenbach/Oliva '03, one may obtain a sufficiently constructive proof in order to get a **rate of convergence**.

We did this in 2018, but this year we found out that we may use a quantitative lemma of Kohlenbach/Powell '20 to get one with weaker restrictions. For example, it is enough to assume that

$$\sum_{n=0}^{\infty} \gamma_n = \infty.$$

Approximating curves

Finally, we have obtained the following result about the asymptotic behaviour at infinity, which subsumes a lot of classical results due to Minty, Halpern, Bruck, Jost, as well as a recent one due to Bačák and Reich.

Theorem (A.S., 2020)

Assume that X is complete. Let $(T_\gamma)_{\gamma>0}$ be a jointly firmly nonexpansive family of self-mappings of X. Put $F := \bigcap_{\gamma>0} Fix(T_\gamma)$. Let $x \in X$, $b > 0$, and $(\lambda_n)_{n\in\mathbb{N}} \subseteq (0, \infty)$ with $\lim_{n\to\infty} \lambda_n = \infty$ and assume that for all n, $d(x, T_{\lambda_n} x) \leq b$. Then $F \neq \emptyset$ and the curve $(T_\gamma x)_{\gamma>0}$ converges to the unique point in F which is closest to x.

Convergence and metastability

Let us see now what we can do when we have no hope of obtaining a rate of convergence. A convergence statement usually looks like

$$\forall \varepsilon \; \exists N \; \forall n \geq N \; d(x_n, x) \leq \varepsilon.$$

In a complete space, this is equivalent to Cauchyness, which can be written like

$$\forall \varepsilon \; \exists N \; \forall M \; \forall i, j \in [N, N+M] \; d(x_i, x_j) \leq \varepsilon.$$

In turn, this is equivalent to a Herbrandized variant of it, called "metastability" by Terence Tao (at the suggestion of Jennifer Chayes), expressed as

$$\forall \varepsilon \; \forall g : \mathbb{N} \to \mathbb{N} \; \exists N \; \forall i, j \in [N, N+g(N)] \; d(x_i, x_j) \leq \varepsilon.$$

As this is a $\forall\exists$ statement (in a generalized sense), by the metatheorems of proof mining one can extract from its proof a *rate of metastability*, i.e. a *bound* $\Theta(\varepsilon, g, \ldots)$ on the N.

Gaspar's work

Theorem (Hillam, 1976)

Let $f : [0,1] \to [0,1]$ *be continuous and* $x \in [0,1]$*. If* $\lim_{n\to\infty}(f^n x - f^{n+1}x) = 0$*, then the sequence* $(f^n x)$ *converges.*

Jaime Gaspar, in his 2011 PhD thesis, has obtained for $(f^n x)$ a rate of metastability having as extra parameters:

- a modulus of uniform continuity for f;
- a rate of convergence of $(f^n x - f^{n+1}x)$ towards 0 (later in the thesis refined to a metastable version).

His main achievement was to fit the original proof of Hillam into a system of lower logical strength by replacing the use of the Bolzano-Weierstrass theorem with that of the infinite pigeonhole principle, thus resulting in a rate of low computational complexity.

Improvements on Gaspar

Theorem (Rhoades, 1974)

Let $f : [0, 1] \to [0, 1]$ *be continuous and* (x_n), $(t_n) \subseteq [0, 1]$ *be such that for all* n,

$$x_{n+1} = (1 - t_n)x_n + t_n f(x_n).$$

If $\lim_{n\to\infty} t_n = 0$, *then the sequence* (x_n) *converges.*

By a slight modification of Gaspar's proof, we have obtained for (x_n) in the above a rate of metastability having as extra parameters a modulus of uniform continuity for f and a rate of convergence of (t_n) towards 0.

- in particular, for $t_n = 1/(n + 1)$ (Franks/Marzec 1971), we obtain an **unconditional** rate of metastability
- also, one can easily extend this to the Ishikawa iteration (Rhoades 1976)

The case of Lipschitz functions

Theorem (Borwein/Borwein, 1991)

Let $L > 0$, $f : [0,1] \to [0,1]$ be L-Lipschitz and (x_n), $(t_n) \subseteq [0,1]$ be such that for all n,

$$x_{n+1} = (1 - t_n)x_n + t_n f(x_n).$$

If there is a $\delta > 0$ such that for all n,

$$t_n \leq \frac{2 - \delta}{L + 1},$$

then the sequence (x_n) converges.

The proof of this theorem relies on a completely new kind of argument, never analyzed before in proof mining.
We managed to extract for (x_n) in the above a rate of metastability having just δ as an extra parameter.

Defining the rate

For all $f : \mathbb{N} \to \mathbb{N}$, we define $\widetilde{f} : \mathbb{N} \to \mathbb{N}$, for all n, by

$$\widetilde{f}(n) := n + f(n).$$

We define $f^M : \mathbb{N} \to \mathbb{N}$, for all n, by

$$f^M(n) := \max_{i \leq n} f(i).$$

In addition, for all $n \in \mathbb{N}$, we denote by $f^{(n)}$ the n-fold composition of f with itself.

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

$$h^g_m(n) := g(m+n), \quad P^{\varepsilon,g}_0 := 0, \quad P^{\varepsilon,g}_{n+1} := P^{\varepsilon,g}_n + \widetilde{h^g_{P^{\varepsilon,g}_n}}^{(\lceil \frac{1}{\varepsilon} \rceil + 1)}(0)$$

$$T_{\varepsilon,\delta} := \left\lceil \log_{\left(1-\frac{\delta}{2}\right)} \varepsilon \right\rceil + 1, \qquad B_{\varepsilon,g,\delta} := T_{\varepsilon,\delta} + \widetilde{g}\left(P^{\varepsilon,g}_{T_{\varepsilon,\delta}}\right) + 1$$

$$\Psi_\delta(\varepsilon, g) := P^{\varepsilon,g}_{B_{\varepsilon,g,\delta}}.$$

An idea of the analysis

The proof is quite intricate and relies on the indices where (x_n) switches direction, labeled (q_r). It divides into two cases:

Case I. There is an r with $q_r = \infty$.

That is, the sequence is monotone from some point on, hence convergent.

Case II. For all r, $q_r < \infty$.

In this case one relies on a previous lemma which essentially says that the monotonicity intervals get exponentially smaller.

In the analyzed proof, the two cases become (using the notations on the previous slide):

Case I. There is an $r \leq B_{\varepsilon,g,\delta}$ with $q_r > P_r^{\varepsilon,g}$.

Case II. For all $r \leq B_{\varepsilon,g,\delta}$, $q_r \leq P_r^{\varepsilon,g}$.

This may be found at arXiv:2008.03934.

The mean ergodic theorem

In the mid-1930s, Riesz proved the following formulation of the classical mean ergodic theorem of von Neumann: if X is a Hilbert space, $T : X \to X$ is a linear operator such that for all $x \in X$, $\|Tx\| \leq \|x\|$, then for any $x \in X$, we have that the corresponding sequence of ergodic averages (x_n), where for each n,

$$x_n := \frac{1}{n+1} \sum_{k=0}^{n} T^k x$$

is convergent.

Rates of metastability were extracted:

- by Avigad/Gerhardy/Towsner 2007 (publ. in Trans. AMS, 2010) – having as an extra parameter an upper bound on $\|x\|$;
- by Kohlenbach/Leuştean 2008 (publ. in Ergodic Theory and Dynamical Systems 2009) – by analyzing a simpler proof of Birkhoff from 1939 which applies to the more general case of uniformly convex Banach spaces.

The work of Kohlenbach and Leuştean

Their main achievement was to replace the use of the greatest lower bound principle by the following quantitative arithmetical version of it:

Lemma

Let $(a_n) \subseteq [0,1]$. *Then for all* $\varepsilon > 0$ *and all* $g : \mathbb{N} \to \mathbb{N}$ *there is an* $N \leq \left(g^M\right)^{\left(\lceil \frac{1}{\varepsilon} \rceil\right)}(0)$ *such that for all* $s \leq g(N)$, $a_N \leq a_s + \varepsilon$.

In the intervening years, proof mining has continued this line of research, yielding for example rates of metastability for **nonlinear** generalizations of ergodic averages.

The multi-parameter mean ergodic theorem

Inspired by Birkhoff's proof, Riesz produced a new one in 1941 that separates more clearly the role played by uniform convexity (i.e. the fact that in such spaces minimizing sequences of convex sets are convergent). One of the advantages of this argument is that it readily generalizes to the following multi-parameter case (a result attributed to Dunford): if $d \geq 1$ and $T_1, \ldots, T_d : X \to X$ are commuting linear operators such that for each l and for each $x \in X$, $\|T_l x\| \leq \|x\|$, then for any $x \in X$, the sequence (x_n), defined, for any n, by

$$x_n := \frac{1}{(n+1)^d} \sum_{k_1=0}^{n} \ldots \sum_{k_d=0}^{n} T_1^{k_1} \ldots T_d^{k_d} x$$

is convergent.

We managed to extract a rate of metastability having as extra parameters d, an upper bound b for $\|x\|$ and a modulus of uniform convexity η for X.

The rate

Define, for any suitable s, β, f, ε, g, u, d, δ, Q, γ, n, η, α, b:

$$\begin{aligned}
p(s) &:= 2s^2 + 2s \\
G(\beta, f) &:= \left((p \circ f)^M\right)^{\left(\left\lceil \frac{1}{\beta} \right\rceil\right)}(0). \\
\Phi(d, \delta, Q) &:= \max\left(Q, \left\lceil \frac{2^d Q}{\delta} \right\rceil\right) \\
h_{\gamma,g,d}(n) &:= \widetilde{g}\left(\Phi\left(d, \frac{\gamma}{2}, n\right)\right) \\
\Psi(d, \gamma, g) &:= \Phi\left(d, \frac{\gamma}{2}, G\left(\frac{\gamma}{2}, h_{\gamma,g,d}\right)\right) \\
u_\eta(\alpha) &:= \frac{\alpha}{2} \cdot \eta(\alpha) \\
\Xi_{\eta,d}(\alpha, g) &:= \Psi\left(d, \frac{u_\eta(\alpha)}{2}, g\right) \\
\Theta_{\eta,d,b}(\varepsilon, g) &:= \Xi_{\eta,d}(\varepsilon/b, g).
\end{aligned}$$

Difficulties encountered

- we first finitized the proof by noticing that the infimum of all the convex combinations of the iterates of x may be effectively replaced by that of just the arithmetic means of pairs of two given ergodic averages;
- thus, we only needed to extend the abovementioned principle of Kohlenbach and Leuștean to double sequences, which we did by means of the Cantor pairing function;
- we also had to use a combinatorial argument to deal with multiple dimensions.

This work may be found at arXiv:2008.03932, but also some further research is needed.

For example, it would be interesting to find out if one can obtain bounds on the number of fluctuations, in the spirit of Avigad/Rute 2015 and Kohlenbach/Safarik 2014.

We now present the most intricate result obtained so far through proof mining (jww U. Kohlenbach, to appear in Commun. Contemp. Math.). Assume that X is complete and that the following makes sense. Let $C \subset X$ be convex, closed, bounded, nonempty and let $T : C \to C$ be nonexpansive. Fix $x \in C$ and put, for all $t \in [0,1)$, $T_t : C \to C$, defined, for all $y \in C$, by

$$T_t y := tTy + (1-t)x.$$

Clearly, each T_t is a t-contraction and so there is a unique x_t with $x_t = T_t x_t$.

- in Hilbert spaces, $\lim_{t \to 1} x_t = P_{Fix(T)} x$ (Browder 1967; Halpern 1967) – a particular case of the previous approximating curve result
  - rate of metastability obtained by Kohlenbach for this and Wittmann's 1992 theorem (Adv. Math., 2011)
- generalization of Wittmann to CAT(0) spaces (Saejung 2010)
  - rate of metastability obtained by Kohlenbach/Leuştean (Adv. Math., 2012)

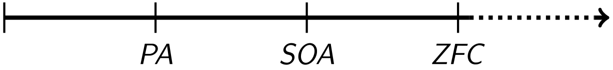
These results were made possible by eliminating strong proof principles used in the original proofs. Why?

The inner workings of proof mining

Remember how proof mining works:

$$S \vdash \forall x \exists y \varphi(x, y) \Rightarrow S' \vdash \forall x \varphi(x, tx).$$

What is S? Recall the Gödel hierarchy:



- $PA \rightsquigarrow$ System T (Gödel, early 1940s, published 1958)
- $SOA \rightsquigarrow$ System $T + BR$ (Spector, 1962)
- ZFC: beyond the range of current interpretative proof theory

The point of the simplifications before was to show that the System T functionals are sufficient for expressing the desired rates. Also, see the recent approach of Ferreira/Leuştean/Pinto (Adv. Math., 2019) via the bounded functional interpretation of Ferreira/Oliva.

Enter Reich

What about extending the Browder-Halpern theorem to more general Banach spaces? (Browder covered the ℓ^p case but left open the L^p one, expect for the L^2 spaces which are Hilbert.)

We have e.g. the following result, central to proving the convergence of numerous nonlinear analysis algorithms.

Theorem (Reich, 1980)

In the framework above, if X is a **uniformly smooth Banach space***, then for all $x \in C$ we have that $\lim_{t \to 1} x_t$ exists and it is a fixed point of T.*

The extraction of a rate of metastability for the above statement has stood as an open problem for 10 years.

The first question to ask is: what property does this $p \in Fix(T)$ satisfy? (We expect it to be relevant, since the corresponding one turned out to be in the Browder analysis.)

Sunny nonexpansive retractions

Let E be a nonempty subset of C and $Q : C \to E$. We call Q a **retraction** if for all $x \in E$, $Qx = x$. If Q is a retraction, we call it **sunny** if for all $x \in C$ and $t \geq 0$, $Q(Qx + t(x - Qx)) = Qx$.

Proposition (Variational Inequality)

A retraction $Q : C \to E$ is sunny and nonexpansive iff for all $x \in C$ and $y \in E$,

$$\langle x - Qx, j(y - Qx) \rangle \leq 0.$$

As a consequence, there is at most one sunny nonexpansive retraction $Q : C \to E$ (Bruck 1973). In addition, $Q = P$ iff X is Hilbert (Bruck 1974).

We may now say that the point p in Reich's theorem satisfies $p = Q_{Fix(T)}x$, where $Q_{Fix(T)} : C \to Fix(T)$ is the unique sunny nonexpansive retraction.

A use of strong principles

The crucial segment defines a function $f : C \to \mathbb{R}_+$, for all $z \in C$, by $f(z) := \limsup_{n\to\infty} \|x_n - z\|$. Let K be the set of minimizers of f. The claim is that there is a $p \in K \cap Fix(T)$.

Since f is convex and continuous, C is closed convex bounded nonempty, and X is uniformly smooth, hence reflexive, we have that (!) $K \neq \emptyset$. Let $y \in K$ and $z \in C$. Then:

$$\begin{aligned} f(Ty) = \limsup_{n\to\infty} \|x_n - Ty\| &\leq \limsup_{n\to\infty} (\|x_n - Tx_n\| + \|Tx_n - Ty\|) \\ &\leq \limsup_{n\to\infty} (\|x_n - Tx_n\| + \|x_n - y\|) \\ &\leq \limsup_{n\to\infty} \|x_n - Tx_n\| + \limsup_{n\to\infty} \|x_n - y\| \\ &= f(y) \leq f(z), \end{aligned}$$

so $Ty \in K$. Now, since K is a closed convex bounded nonempty T-invariant subset of a uniformly smooth space, we have that (!) there is a $p \in K \cap Fix(T)$.

On uniqueness

We try to find an alternative path to the claim. Of course, *a posteriori* the point in $K \cap Fix(T)$ is unique, as it is simply the limit p of the sequence (x_n), characterized by $f(p) = 0$.
Is there a way of obtaining this uniqueness *a priori*?

Answer: Yes, if we use the following proposition which holds if the space is in addition **uniformly convex** (still covering the L^p case).

Proposition (Zălinescu, JMAA 1983)

Let X be uniformly convex with modulus η and $b \geq \frac{1}{2}$. Then there is a $\psi_{b,\eta} : (0,2] \to (0,\infty)$ such that for all $\varepsilon \in (0,2]$ and all x, y in the closed ball of radius b with $\|x - y\| \geq \varepsilon$, one has that

$$\left\| \frac{x+y}{2} \right\|^2 + \psi_{b,\eta}(\varepsilon) \leq \frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2.$$

In 2018, Bačák and Kohlenbach have obtained an *explicit* formula for $\psi_{b,\eta}$.

Removal of comprehension axioms

The proof may then be further simplified as follows:

The first (tedious) step is to replace the ideal elements (limits, fixed points) by approximate ones. For example, it turns out that in the previous argument, only arbitrarily good minimizers are needed.

The second step is to replace the $\limsup$'s by *approximate* $\limsup$'s (whose existence may be shown, using ideas from Kohlenbach 2000, to be equivalent to Π^0_2-IA), in a process known as *arithmetization* (this is possible mainly because the $\limsup$'s are used pointwise and not as an operator in itself).

On complexity and tameness

After all the simplifications have been done, the extraction process proceeds smoothly and yields a purely numerical term. A close analysis of the term shows that the functional can actually be defined in T_1, and it is an open question whether it is actually in T_0 or whether some different proof may produce a T_0-definable rate of metastability, similarly to all the rates obtained in proof mining so far (*proof-theoretic tameness*).

From the introduction to the paper:
"The enormous complexity of the final bound reflects the profound combinatorial and computational content of Reich's deep theorem."

Applications

- The rate of metastability thus obtained can be used as an input to a previous partial analysis by Kohlenbach/Leuştean of a proof of Shioji/Takahashi (Proc. AMS, 1997) for the convergence in our setting of the Halpern iteration.
- In addition, a slightly modified argument (using a resolvent construction) works also if one replaces the nonexpansive mapping T with a more general pseudocontraction (required to be uniformly continuous), i.e. one that satisfies, for all x, $y \in C$,
$$\langle Tx - Ty, j(x-y)\rangle \leq \|x-y\|^2.$$
- This more general bound completes an analysis of Körnlein/Kohlenbach of a proof of Chidume/Zegeye (Proc. AMS, 2004) for the convergence of the Bruck iteration.

Axiomatizing L^p spaces

My next result also pertains to L^p spaces, as it provides an axiomatization of them that yields a metatheorem of the sort used in proof mining.

This follows earlier work of Günzel/Kohlenbach (Adv. Math., 2016) which shows that one can find such a metatheorem for any space that is axiomatizable in positive-bounded logic (the premier logic for metric spaces).

(This result was published in J. Symb. Logic in 2019.)

Examples of spaces

In the Günzel/Kohlenbach paper, the following subclasses of Banach lattices were considered:

- the class of all Banach lattices
- $L^p(\mu)$ lattices
  - out of which, the case of atomless μ
- $C(K)$ lattices
- BL^pL^q lattices

The goal was to adapt an axiomatization of $L^p(\mu)$ spaces simply as Banach spaces (without considering an additional lattice structure).

Abstract characterization of L^p spaces

We started from the following characterization for $L^p(\mu)$ spaces:

Theorem (Lindenstrauss, Pelczynski, Tzafriri, late 1960s)

A Banach space is isomorphic to a $L^p(\mu)$ space iff for all $\varepsilon > 0$ and all finite-dimensional subspaces B of it, there is a finite-dimensional subspace C which contains B and is "$(1+\varepsilon)$-isometric" (gauging by the Banach-Mazur distance) to a finite-dimensional Banach space with the standard p-norm.

We then adapted it (using ideas of Henson/Raynaud 2007) to one where the corresponding objects have dimension/norm bounds.

Theorem (A.S., 2019)

A Banach space X is isomorphic to a $L^p(\mu)$ space iff for all $x_1,\ldots,$ x_n in X of norm at most 1 and for all $N \in \mathbb{N}_{\geq 1}$, there is a subspace $C \subseteq X$ and $y_1,\ldots, y_n$ in C of norm at most 1 such that C is of dimension at most $(4nN+1)^n$, it is isometric to $\mathbb{R}_p^{\dim_{\mathbb{R}} C}$ and for all i, $\|x_i - y_i\| \leq \frac{1}{N}$.

The final axiom

This how this looks like in the higher-typed language of proof mining metatheorems:

$$\psi(m,z) := \forall a^{1(0)} \left(\left\| \sum_{i=1}^{m} |a(i)|_{\mathbb{R}} \cdot_X z(i) \right\| =_{\mathbb{R}} \left(\sum_{i=1}^{m} |a(i)|_{\mathbb{R}}^{p} \right)^{1/p} \right)$$
$$\psi'(m,n,y,z,\lambda) := \forall k \preceq_0 (n-1)$$
$$\left(y(k+1) =_X \sum_{i=1}^{m} \lambda(k+1,i) \cdot_X z(i) \right)$$
$$\psi''(n,N,x,y) := \forall k \preceq_0 (n-1)$$
$$\left(\left\| \widetilde{x(k+1)} - y(k+1) \right\| \leq_{\mathbb{R}} \frac{1}{N} \wedge \|y(k+1)\| \leq_{\mathbb{R}} 1 \right)$$
$$\varphi(n,m,N,x,y,z,\lambda) := \psi(m,z) \wedge \psi'(m,n,y,z,\lambda) \wedge \psi''(n,N,x,y)$$
$$B := \forall n^0, N^0 \geq 1 \forall x^{X(0)} \exists y,z \preceq_{X(0)} 1_{X(0)} \exists \lambda^{1(0)(0)} \in [-1,1]$$
$$\exists m \preceq_0 (4nN+1)^n \varphi(n,m,N,x,y,z,\lambda)$$

The metatheorem

Theorem (A.S., 2019)

Let $B_\forall(x,u)$ (resp. $C_\exists(x,v)$) be a $\forall$-formula with only x,u free (resp. an $\exists$-formula with only x,v free). If

$$\mathcal{A}^\omega[X,\|\cdot\|,L^p] \vdash \forall x^\rho(\forall u^0 B_\forall \to \exists v^0 C_\exists),$$

then there exists an extractable computable functional Φ such that for all x and x^* (where x^* is of a corresponding "number" type $\widehat{\rho}$ and majorizes x) we have that

$$\forall u \leq \Phi(x^*) B_\forall(x,u) \to \exists v \leq \Phi(x^*) C_\exists(x,v)$$

holds in every $L^p(\mu)$ Banach space.

An immediate application of this is a simpler proof for the derivation of the modulus of uniform convexity.

And now for something completely different...

Chebyshev approximation

We have the following classical Chebyshev approximation result.

Theorem (de la Vallée Poussin, Young – 1900s)

For every $n \in \mathbb{N}$ and every continuous $f : [0,1] \to \mathbb{R}$ there is an unique $p \in P_n$ (the set of real polynomials of degree at most n) such that

$$\|f - p\| = \min_{q \in P_n} \|f - q\|$$

(where $\|\cdot\|$ denotes the supremum norm).

Kohlenbach extracted in 1990 a *modulus of uniqueness* – a function Ψ with the property that if p_1 and p_2 are such that $\|f - p_1\|, \|f - p_2\| \leq \min + \Psi(\delta)$, then $\|p_1 - p_2\| \leq \delta$.

He did this by analyzing the uniqueness proof and obtaining an approximate version of it.

Directions to follow

Kohlenbach also suggested in his 1990 thesis to extend the techniques to the following results:

- L_1-best approximation: analyzed by Kohlenbach/Oliva in the early 2000s
- Chebyshev approximation with bounded coefficients
  - a 1971 result of Roulier and Taylor
  - its analysis stood thus for 30 years as an open problem in proof mining

The last one is what we are focusing here on.

The result

Theorem (Roulier and Taylor, 1971)

Let $n,\ m \in \mathbb{N}$ be such that $m \leq n$ and $(k_i)_{i=1}^m \subseteq \mathbb{N}$ be such that $0 < k_1 < \ldots < k_m \leq n$. In addition, let $(a_i)_{i=1}^m$ and $(b_i)_{i=1}^m$ be finite sequences in $\mathbb{R} \cup \{\pm\infty\}$ be such that for all $i \in \{1, \ldots, m\}$, $a_i \leq b_i$, $a_i \neq \infty$ and $b_i \neq -\infty$. If one sets

$$K := \left\{ \sum_{i=0}^{n} c_i X^i \in P_n \;\middle|\; \text{for all } i \in \{1, \ldots, m\},\ a_i \leq c_{k_i} \leq b_i \right\},$$

then for any continuous $f : [0,1] \to \mathbb{R}$ there is a unique $p \in K$ such that

$$\|f - p\| = \min_{q \in K} \|f - q\|.$$

The differences

The essential step was to modify the classical Lagrange interpolation formula for polynomials of degree n:

$$p = \sum_{j=1}^{n+1} \left(\prod_{i \neq j} \frac{X - x_i}{x_j - x_i} \right) \cdot p(x_j)$$

to the case where we have an $r \in \mathbb{N}$ with $r \leq n$ and $(d_i)_{i=1}^{r+1} \subseteq \mathbb{N}$ with $n \geq d_1 > d_2 > \ldots > d_{r+1} = 0$, and the polynomials are of the form

$$p = \sum_{i=1}^{r+1} \eta_i X^{d_i}.$$

The ideas

Using notions from algebraic combinatorics – generalized Vandermonde determinants, partitions, Young tableaux, Schur functions (denoted by $s_\bullet$) – we obtain the following "Lagrange-Schur" formula:

$$p = \sum_{j=1}^{n+1} \left(\prod_{i \neq j} \frac{X - x_i}{x_j - x_i} \right) \cdot p(x_j) \cdot \frac{s_{\lambda^d}(X, x_1, \ldots, \widehat{x_j}, \ldots, x_{r+1})}{s_{\lambda^d}(x_1, \ldots, x_{r+1})},$$

where the additional Schur factors may be easily bounded in the ways we desire.

The final modulus

In the end, we obtain the following modulus of uniqueness:

$$\Psi(\delta) := \frac{\left(\frac{\chi_{\omega,n,M}\left(\frac{L}{2}\right)}{2}\right)^{\frac{n^2}{2}+2n}}{10 \cdot N_n^2(n+1)(nF_n+1)} \cdot \delta,$$

which depends (in addition to δ) on

- the norm of a polynomial p_0 in K;
- the degree n;
- a lower bound L on E;
- a modulus of uniform continuity ω for f;
- the norm of f.

The paper containing this result was recently accepted to Mathematische Nachrichten.

Thank you for your attention.