Recursive inequalities in applied proof theory

Thomas Powell
(jww Morenikeji Neri)

University of Bath

Type Theory, Constructive Mathematics and Geometric Logic

CIRM, Marseille
2 May 2023

These slides will be available at
`https://t-powell.github.io/talks`

Background: Applied proof theory

What is applied proof theory?

There is a famous quote due to G. Kreisel (*A Survey of Proof Theory II*):

> *"What more do we know when we know that a theorem can be proved by limited means than if we merely know that it is true?"*

In other words, the **proof** of a theorem gives us much more information than the mere **truth** of that theorem.

Applied proof theory is a branch of logic that uses proof theoretic techniques to exploit this phenomenon.

People do applied proof theory without realising it...

PROBLEM. Give me an upper bound on the nth prime number p_n.

1. What is p_n? I know it exists because of Euclid...
2. Specifically, given $p_1, \ldots, p_{n-1}$, I know that $N := p_1 \cdot \ldots \cdot p_{n-1} + 1$ contains a *new* prime factor q, and so $p_n \leq q \leq N$.
3. In other words, the sequence $\{p_n\}$ satisfies
$$p_n \leq p_1 \cdot \ldots \cdot p_{n-1} + 1 \leq (p_{n-1})^{n-1}$$
4. By induction, it follows that e.g. $p_n < 2^{2^n}$.

This is a simple example of applied proof theory in action! From the **proof** that there are infinitely many primes, we have inferred a **bound** on the nth prime.

… but it's not always that simple

Theorem (Littlewood 1914)

The functions of integers

(a) $\psi(x) - x$, *and*

(b) $\pi(x) - li(x)$

change signs infinitely often, where $\pi(x)$ *is the number of prime* $\leq x$, $\psi(x)$ *is the is logarithm of the l.c.m. of numbers* $\leq x$ *and* $li(x) = \int_0^x (1/\log(u))du$.

The original proof is utterly nonconstructive, using among other things a **case distinction on the Riemann hypothesis**. At the time, no numerical value of x for which $\pi(x) > li(x)$ was known.

In 1952, Kreisel analysed this proof and extracted recursive bounds for sign changes (*On the interpretation of non-finitist proofs, Part II*):

> *"Concerning the bound … note that it is to be expected from our principle, since if the conclusion … holds when the Riemann hypothesis is true, it should also hold when the Riemann hypothesis is nearly true: not all zeros need lie on* $\sigma = \frac{1}{2}$*, but only those whose imaginary part lies below a certain bound … and they need not lie on the line* $\sigma = \frac{1}{2}$*, but near it"*

A routine example from modern applied proof theory

Theorem (Kirk and Sims, *Bulletin of the Polish Academy of Sciences* 1999)

Suppose that C is a closed subset of a uniformly convex Banach space and $T : C \to C$ is asymptotically nonexpansive with $\mathrm{int}(\mathrm{fix}(T)) \neq \emptyset$. Then for each $x \in C$ the sequence $\{T^n x\}$ converges to a fixed point of T.

Theorem (P., *Journal of Mathematical Analysis and Applications* 2019)

Let $T : C \to C$ be a nonexpansive mapping in L_p for $2 \leq p < \infty$, and suppose that $B_r[q] \subset \mathrm{fix}(T)$ for some $q \in L_p$ and $r > 0$. Suppose that $x \in C$ and $\|x - q\| < K$, and define $x_n := T^n x$. Then for any $\varepsilon > 0$ we have

$$\forall n \geq f(\varepsilon)(\|Tx_n - x_n\| \leq \varepsilon)$$

where

$$f(\varepsilon) := \left\lceil \frac{p \cdot 2^{3p+1} \cdot K^{p+2}}{\varepsilon^p \cdot r^2} \right\rceil$$

One of over a hundred papers that were written by proof theorists, but published in specialist journals in the area of application.

A one-slide overview of the field

- Origins in the work of Kreisel and the "unwinding" of proofs [Kreisel, 1951, Kreisel, 1952]. Early case studies in number theory.

- Applications in mathematics were brought to maturity by Kohlenbach and his collaborators from late 90s onwards (see the textbook [Kohlenbach, 2008] and the recent survey papers [Kohlenbach, 2017, Kohlenbach, 2019] for an overview):
  - Case studies in nonlinear analysis, ergodic theory, approximation theory, convex optimization, more recently algebra, probability and number theory,
  - logical metatheorems that explain individual applications as instances of general logical phenomena.

  This branch of research commonly referred to as "proof mining".

- But the area is a lot bigger: It is related, more broadly, to constructive mathematics, formal program extraction [Schwichtenberg and Wainer, 2011], constructive/dynamical algebra [Lombardi and Quitté, 2015], higher-order computability [Longley and Normann, 2015] and much more besides!

- Many of the basic techniques are studied in their own right e.g. Dialectica interpretation, negative translations.

Remarks

- Applied proof theory is not defined by the use of any single technique: It's about thinking about and doing mathematics (or computer science) with a proof-theoretic mentality.
- Finding new ways/areas to apply proof-theoretic thinking in a meaningful way is **very hard** (but very rewarding when it works out).
- Research exists on a broad spectrum. Some papers involve sophisticated logical techniques in an explicit way, other just look like maths papers.
- It is a rapidly growing field. There is huge potential in both
    - Expanding to new areas of application (e.g. probability theory),
    - Incorporating new techniques (e.g. theorem provers, automated reasoning).

Plan for rest of talk:

- Introduce, via a simple example, the main theme of the talk: **Recursive inequalities**.
- Explain why these are interesting for both analysts and applied proof theorists.
- Motivate my **quantitative classification project** with Morenikeji Neri. Present our main results so far.
- Give an extended outline of plans for future work involving
  - Stochastic algorithms
  - Computer formalised mathematics
  - Automated reasoning

Recursive inequalities
(a basic example)

Monotone sequences

Let $c \geq 0$ and $\{\mu_n\}$ be a sequence of nonnegative reals satisfying for all $n \in \mathbb{N}$:

$$\mu_{n+1} \leq c\mu_n$$

Question. Does this sequence converge, and if so how fast?

Theorem

1. *If $c < 1$ then $\mu_n \to 0$, with rate of convergence $\mu_n \leq c^n \mu_0$.*
2. *If $c = 1$ then $\mu_n \to \mu$ for some $\mu \geq 0$, but this may not be computable even if $\{\mu_n\}$ is a computable sequence of rationals.*
3. *If $c > 1$ then $\{\mu_n\}$ may not converge at all.*

Item 2 is proven by adapting Specker's famous construction [Specker, 1949].

An application: Banach's fixed point theorem

Suppose that (X, d) is a metric space and $T : X \to X$ a contractive mapping with constant $c \in [0, 1)$ i.e.

$$d(T(x), T(y)) \leq c \cdot d(x, y)$$

for all $x, y \in X$.

Theorem

If x^* is a fixed point of T, and $x_{n+1} := Tx_n$ for some $x_0 \in X$, then $x_n \to x^*$ with rate

$$d(x_n, x^*) \leq c^n \cdot d(x_0, x^*)$$

Proof. Define $\mu_n := d(x_n, x^*)$. Then

$$\mu_{n+1} = d(Tx_n, x^*) = d(Tx_n, Tx^*) \leq c \cdot d(x_n, x^*) = c\mu_n$$

so we can apply the abstract convergence theorem of the previous slide for $c < 1$.

Can we do anything in the case of noncomputable convergence?

Theorem (Rephrasing of the case $c = 1$)

Suppose that $\{\mu_n\}$ is a sequence of nonnegative reals with $\mu_{n+1} \leq \mu_n$ for all $n \in \mathbb{N}$. Then $\{\mu_n\}$ is Cauchy.

Proof. If this were not true, there would exists some $\varepsilon > 0$ such that

$$\forall N \exists m, n \geq N(|\mu_m - \mu_n| > \varepsilon)$$

Using (weak) choice, there exists $g : \mathbb{N} \to \mathbb{N}$ such that

$$\forall N \exists m, n \in [N, N + g(N)](|\mu_m - \mu_n| > \varepsilon)$$

Define $\tilde{g}(k) := k + g(k)$. Then we have

$$\mu_0 > \mu_{\tilde{g}(0)} + \varepsilon > \mu_{\tilde{g}(\tilde{g}(0))} + 2\varepsilon > \ldots \geq \mu_{\tilde{g}^{(i)}(0)} + i\varepsilon > \ldots$$

which is a contradiction for

$$i := \left\lceil \frac{\mu_0}{\varepsilon} \right\rceil$$

A computational convergence theorem

Theorem (Original theorem)

Suppose that $\{\mu_n\}$ *is a sequence of nonnegative reals with* $\mu_{n+1} \leq \mu_n$ *for all* $n \in \mathbb{N}$. *Then* $\{\mu_n\}$ *is Cauchy.*

Theorem (Computational (or metastable) version)

Suppose that $\{\mu_n\}$ *is a sequence of nonnegative reals with* $\mu_{n+1} \leq \mu_n$ *for all* $n \in \mathbb{N}$. *Then for any* $\varepsilon > 0$ *and* $g : \mathbb{N} \to \mathbb{N}$ *there exists some* $N \leq \Phi(\varepsilon, g)$ *such that*

$$\forall m, n \in [N, N + g(N)](|\mu_m - \mu_n| \leq \varepsilon)$$

where

$$\Phi(\varepsilon, g) := \tilde{g}^{\lceil \mu_0/\varepsilon \rceil}(0)$$

- The extraction of rates of metastability is a standard result in applied proof theory in cases where convergence rates are not possible.
- Metastable convergence was independently discovered by Tao [Tao, 2007], and has a mathematical significance as a “finitary analogue” of Cauchy convergence.

What is going on logically?

$$
\begin{aligned}
&\neg\neg[\{\mu_n\} \text{ is Cauchy}]\\
\iff &\neg\neg[\forall \varepsilon > 0 \exists N \forall m, n \geq N(|\mu_m - \mu_n| \leq \varepsilon)]\\
\iff &\neg[\exists \varepsilon > 0 \forall N \exists m, n \geq N(|\mu_m - \mu_n| > \varepsilon)]\\
\iff &\neg[\exists \varepsilon > 0, g : \mathbb{N} \to \mathbb{N} \forall N \exists m, n \in [N, N + g(N)](|\mu_m - \mu_n| > \varepsilon)]\\
\iff &[\forall \varepsilon > 0, g : \mathbb{N} \to \mathbb{N} \exists N \forall m, n \in [N, N + g(N)](|\mu_m - \mu_n| \leq \varepsilon)]
\end{aligned}
$$

Statement: $\mu_n \leq \mu_{n+1} \implies \{\mu_n\}$ *is Cauchy.*

In general no direct computational interpretation (because of Specker).

Statement: $\mu_n \leq \mu_{n+1} \implies \neg\neg[\{\mu_n\}$ *is Cauchy*$]$.

We can extract a direct realizer i.e. a computable rate of metastability i.e.

$$\forall \varepsilon > 0, g : \mathbb{N} \to \mathbb{N} \exists N \leq \Phi(\varepsilon, g) \forall m, n \in [N, N + g(N)](|\mu_m - \mu_n| \leq \varepsilon)$$

for $\Phi(\varepsilon, g) := \tilde{g}^{\lceil \mu_0/\varepsilon \rceil}(0)$.

$$\boxed{\text{Metastability} \approx \text{negative translation} + \text{Dialectica}}$$

Remember Littlewood and Kreisel...

Theorem (Littlewood 1914)

The functions of integers

(a) $\psi(x) - x$, *and*

(b) $\pi(x) - li(x)$

change signs infinitely often, where $\pi(x)$ *is the number of prime* $\leq x$, $\psi(x)$ *is the is logarithm of the l.c.m. of numbers* $\leq x$ *and* $li(x) = \int_0^x (1/\log(u))du$.

The original proof is utterly nonconstructive, using among other things a **case distinction on the Riemann hypothesis**. At the time, no numerical value of x for which $\pi(x) > li(x)$ was known.

In 1952, Kreisel analysed this proof and extracted recursive bounds for sign changes (*On the interpretation of non-finitist proofs, Part II*):

> *"Concerning the bound ... note that it is to be expected from our principle, since if the conclusion ... holds when the Riemann hypothesis is true, it should also hold when the Riemann hypothesis is nearly true: not all zeros need lie on* $\sigma = \frac{1}{2}$, *but only those whose imaginary part lies below a certain bound ... and they need not lie on the line* $\sigma = \frac{1}{2}$, *but near it"*

This one also needed a route through metastable convergence

Theorem (Kirk and Sims, *Bulletin of the Polish Academy of Sciences* 1999)

Suppose that C is a closed subset of a uniformly convex Banach space and $T : C \to C$ is asymptotically nonexpansive with $\mathrm{int}(\mathrm{fix}(T)) \neq \emptyset$. Then for each $x \in C$ the sequence $\{T^n x\}$ converges to a fixed point of T.

Theorem (P., *Journal of Mathematical Analysis and Applications* 2019)

Let $T : C \to C$ be a nonexpansive mapping in L_p for $2 \leq p < \infty$, and suppose that $B_r[q] \subset \mathrm{fix}(T)$ for some $q \in L_p$ and $r > 0$. Suppose that $x \in C$ and $\|x - q\| < K$, and define $x_n := T^n x$. Then for any $\varepsilon > 0$ we have

$$\forall n \geq f(\varepsilon)(\|Tx_n - x_n\| \leq \varepsilon)$$

where

$$f(\varepsilon) := \left\lceil \frac{p \cdot 2^{3p+1} \cdot K^{p+2}}{\varepsilon^p \cdot r^2} \right\rceil$$

One of several hundred papers that were written by proof theorists, but published in specialist journals in the area of application.

From Tao's blog ("Soft analysis, hard analysis, and the finite convergence principle", 2007)

One can of course keep doing this, achieving various sparsified versions of the pigeonhole principle which each capture part of the infinite convergence principle. To get the full infinite convergence principle, one cannot use any single such sparsified version of the pigeonhole principle, but instead must take **all of them at once**. This is the full strength of the finite convergence principle:

> **Finite convergence principle.** If $\varepsilon > 0$ and $F : \mathbb{Z}_+ \to \mathbb{Z}_+$ is a function and $0 \leq x_1 \leq x_2 \leq \ldots \leq x_M \leq 1$ is such that M is sufficiently large depending on F and ε, then there exists $1 \leq N < N + F(N) \leq M$ such that $|x_n - x_m| \leq \varepsilon$ for all $N \leq n, m \leq N + F(N)$.

This principle is easily proven by appealing to the first pigeonhole principle with the sparsified sequence $x_{i_1}, x_{i_2}, x_{i_3}, \ldots$ where the indices are defined recursively by $i_1 := 1$ and $i_{j+1} := i_j + F(i_j)$. This gives an explicit bound on M as $M := i_{\lfloor 1/\varepsilon \rfloor + 1}$. Note that the first pigeonhole principle corresponds to the case $F(N) \equiv 1$, the second pigeonhole principle to the case $F(N) \equiv k$, and the third to the case $F(N) \equiv N$. A particularly useful case for applications is when F grows exponentially in N, in which case M grows tower-exponentially in $1/\varepsilon$.

The main points

- Sequences of nonnegative reals satisfying recursive inequalities may or may not converge, depending on the parameters of those inequalities.
- In cases that they do converge, they may or may not do so with computable rates.
- The convergence properties of sequences of reals can be used to prove convergence results for general metric spaces. Rates of convergence can be applied directly (and this is useful).
- Where no computable rates of convergence do not exist in general, one can often instead provide rates of metastability (and this is also useful!).

Recursive inequalities
(interesting examples)

There are plenty of examples!

Recursive inequalities can be found everywhere numerical analysis and optimization. Their computational analysis is a widespread phenomenon in applied proof theory.

- First clear example in [Kohlenbach and Lambov, 2004] (I think):

$$\boxed{a_{n+1} \leq (1 + b_n)a_n + c_n \text{ with } \sum_{i=0}^{\infty} b_i < \infty \text{ and } \sum_{i=0}^{\infty} c_i < \infty}$$

  Rates of metastability for $\{a_i\}$ calculated and then used in the main result.

- One of several examples from last year is [Sipoş, 2022]:

$$\boxed{a_{n+1} \leq (1 - \alpha_n)a_n + \alpha_n\beta_n \text{ with } \sum_{i=0}^{\infty} \alpha_i = \infty \text{ and } \limsup_{n\to\infty} \beta_n \leq 0}$$

  where $\{\beta_n\}$ could be negative. A rate of metastability for $\{a_n\}$ is required.

I found explicit recursive inequalities in 25 papers in applied proof theory. There are thousands of such papers in ordinary mathematics.

The following examples are from my own research...

An example due to [Alber and Guerre-Delabriere, 1997]

X is a normed space. A mapping $T : X \to X$ is ψ-weakly contractive on some closed, convex $C \subseteq X$ if

$$\|Tx - Ty\| \leq \|x - y\| - \psi(\|x - y\|)$$

for $x, y \in C$, where $\psi : [0, \infty) \to [0, \infty)$ is a nondecreasing function with $\psi(0) = 0$ and $\psi(t) > 0$ for $t > 0$.

- Any contraction mapping is ψ-weakly contractive for $\psi(t) := (1 - c)t$.
- For $X = \mathbb{R}$, the sin function is ψ-weakly contractive on $[0, 1]$ for $\psi(t) := t^3/8$, by considering its Taylor expansion:

$$|\sin(x) - \sin(y)| \leq |x - y| - \frac{|x - y|^3}{8}$$

Reduction to a recursive inequality

Suppose that x^* is a fixed point of T, and the algorithm $\{x_n\}$ is defined according to the Krasnoselskii-Mann method:

$$x_{n+1} = (1 - \alpha_n)x_n + \alpha_n T x_n$$

for $x_0 \in C$ and $\{\alpha_n\} \subset [0, 1]$ with $\sum_{i=0}^{\infty} \alpha_i = \infty$. Then

$$\begin{aligned}
\|x_{n+1} - x^*\| &= \|(1 - \alpha_n)x_n + \alpha_n T x_n - x^*\| \\
&\leq (1 - \alpha_n)\|x_n - x^*\| + \alpha_n \|T x_n - T x^*\| \\
&\leq (1 - \alpha_n)\|x_n - x^*\| + \alpha_n(\|x_n - x^*\| - \psi(\|x_n - x^*\|)) \\
&= \|x_n - x^*\| - \alpha_n \psi(\|x_n - x^*\|)
\end{aligned}$$

i.e. $\mu_n := \|x_n - x^*\|$ satisfies the recursive inequality

$$\boxed{\mu_{n+1} \leq \mu_n - \alpha_n \psi(\mu_n)}$$

which is a generalisation of $\mu_{n+1} \leq c\mu_n$.

Rates of convergence

Theorem (Essentially [Alber and Guerre-Delabriere, 1997])

Suppose that $\{\mu_n\}$ *is a sequence of nonnegative reals satisfying*

$$\mu_{n+1} \leq \mu_n - \alpha_n \psi(\mu_n)$$

where $\{\alpha_n\} \subset [0,1]$ *with* $\sum_{i=0}^{\infty} \alpha_i = \infty$ *and* $\psi : [0,\infty) \to [0,\infty)$ *is a nondecreasing function with* $\psi(0) = 0$ *and* $\psi(t) > 0$ *for* $t > 0$.

Then $\mu_n \to 0$ *with rate*

$$\mu_n \leq \Psi^{-1}\left(\Psi(\mu_0) - \sum_{i=0}^{n-1} \alpha_i\right)$$

for

$$\Psi(x) := \int^x \frac{dt}{\psi(t)}$$

Rates of convergence exist, but are not so simple any more...

It gets trickier

If we instead consider mappings $T : X \to X$ that are *asymptotically* ψ-weakly contractive in some sense e.g.

$$\|T^n x - T^n y\| \leq \|x - y\| - \psi(\|x - y\|) + l_n$$

for $l_n \to 0$ as $n \to \infty$, the corresponding recursive inequality becomes

$$\boxed{\mu_{n+1} \leq \mu_n - \alpha_n \psi(\mu_n) + \alpha_n l_n}$$

This is done in [Alber et al., 2006], but explicit rates of convergence not given.

This can be tackled using ideas from proof theory

Theorem (Adapted from [P. and Wiesnet, 2021])

Suppose that $E \subseteq X$ is convex, $\{A_n\}$ is quasi asymptotically ψ-weakly contractive w.r.t q and with modulus σ. Moreover, from any starting point x_0 define the sequence $\{x_n\}$ by

$$x_{n+1} = (1 - \alpha_n)x_n + \alpha_n A_n x_n$$

where $\{\alpha_n\} \in [0, \alpha]$ is some sequence of nonnegative reals with $\sum_{n=0}^{\infty} \alpha_n = \infty$. Suppose that $\|x_n - q\|$ is bounded by $c > 0$. Then $x_n \to q$, with rate of convergence

$$\|x_n - q\| \leq F^{-1}\left(2\Psi(c) - \sum_{i=0}^{n-2} \alpha_i\right)$$

where $F : (0, \infty) \to \mathbb{R}$ is any strictly increasing and continuous function satisfying

$$F(\varepsilon) \geq 2\Psi\left(\frac{\varepsilon}{2}\right) - \alpha \cdot \sigma\left(\frac{1}{2}\min\left\{\psi\left(\frac{\varepsilon}{2}\right), \frac{\varepsilon}{\alpha}\right\}, c\right)$$

and Ψ is given by

$$\Psi(s) := \int^{s} \frac{dt}{\psi(t)}$$

A final example (simple gradient descent)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex, differentiable function, and x^* be a point where f attains its minimum. Define

$$x_{n+1} := x_n - \alpha_n \nabla f(x_n)$$

for some initial x_0, where $\sum_{i=0}^{\infty} \alpha_i = \infty$ and $\sum_{i=0}^{\infty} \alpha_i^2 < \infty$. Assuming that $L > 0$ is such that $\|\nabla f(x_n)\| \leq L$ for all $n \in \mathbb{N}$, we have

$$\boxed{\mu_{n+1} \leq \mu_n - \alpha_n \beta_n + \tfrac{1}{2} L^2 \alpha_n^2 \quad \text{and} \quad \beta_n - \beta_{n+1} \leq L^2 \alpha_n}$$

for $\mu_n := \frac{1}{2}\|x_n - x^*\|^2$ and $\beta_n := f(x_n) - f(x^*)$.

This is an instance of a recursive inequality where it is known that $\beta_n \to 0$, but no general rates have been given.

The main points

- We know all about convergence and computability for $\mu_{n+1} \leq c\mu_n$.
- More interesting mappings, algorithms or spaces give rise to more complex recursive inequalities.
- In these cases, rates of convergence are often either not given or not known.
- Ideas from logic can:
    1. Produce rates when they do exist.
    2. Prove that in some cases, computable rates don't exist.
    3. In the second case, produce computable rates of metastability instead.

The systematic study of a class of recursive inequalities

Main idea

Let's take a general class of recursive inequalities and ask the following questions:

- Under precisely which conditions do we get convergence results? Can we strengthen standard results? (**Analysis**)
- Where we suspect that general rates of convergence might not exist, can we prove this properly? (**Computability**)
- Can we analyse the proofs to produce computable rates of convergence or metastability (**Proof theory, constructive maths**)?
- Can we use this to prove new theorems in "proper" mathematics? (**Applied proof theory**)

Our starting point

A general recursive inequality

$\{\mu_n\}$, $\{\alpha_n\}$, $\{\beta_n\}$ and $\{\gamma_n\}$ are sequences of nonnegative reals satisfying

$$\boxed{\mu_{n+1} \leq \mu_n - \alpha_n \beta_n + \gamma_n}$$

where $\sum_{i=0}^{\infty} \alpha_i = \infty$ and $\gamma_n \to 0$ as $n \to \infty$.

All of our examples were of this kind:

- Banach fixed point theorem: $\mu_{n+1} \leq \mu_n - (1-c)\mu_n$ (i.e. $\alpha_n = 1$ and $\gamma_n = 0$)
- Asymptotically ψ-weakly contractive mappings: $\mu_{n+1} \leq \mu_n - \alpha_n \psi(\mu_n) + \alpha_n l_n$
- Gradient descent: $\mu_{n+1} \leq \mu_n - \alpha_n (f(x_n) - f(x^*)) + \frac{1}{2} L^2 \alpha_n^2$

Question: What can we say about convergence of $\{\mu_n\}$ and $\{\beta_n\}$?

We looked into two main categories based on:

1. $\sum_{i=0}^{\infty} \gamma_i < \infty$
2. $\gamma_n / \alpha_n \to 0$ as $n \to \infty$.

One subclass ("gradient-descent type") was the following

Theorem ([Alber and Iusem, 2001])

$\{\mu_n\}$, $\{\alpha_n\}$, $\{\beta_n\}$ and $\{\gamma_n\}$ are sequences of nonnegative reals satisfying

$$\boxed{\mu_{n+1} \leq \mu_n - \alpha_n \beta_n + \gamma_n}$$

where $\sum_{i=0}^{\infty} \alpha_i = \infty$ and $\sum_{i=0}^{\infty} \gamma_i < \infty$. Then whenever there exists $\theta > 0$ such that

$$\beta_n - \beta_{n+1} \leq \theta \alpha_n \quad \text{all } n \in \mathbb{N}$$

then $\mu_n \to \mu$ for some $\mu \geq 0$ and $\beta_n \to 0$.

We can use this to prove that $f(x_n) \to f(x^*)$ for the simple gradient descent algorithm $x_{n+1} := x_n - \alpha_n \nabla f(x_n)$ discussed earlier.

Actually, it can be used to prove a lot more e.g. convergence of

$$x_{n+1} := P_C\left(x_n - \alpha_n \cdot \frac{u_n}{\|u_n\|}\right) \quad u_n \in \partial_{\varepsilon_n} f(x_n)$$

where $\delta_\varepsilon f(x)$ is the ε-subderivative of f, $x \in H$ and C a closed, convex subset of some Hilbert space H etc .. [Alber et al., 1998]

Noncomputability of "gradient-descent type" convergence

In [Alber et al., 1998], one of the many places where variants of this recursive inequality is used, the authors write, about their proof of $f(x_n) \to f(x^*)$:

> *This result does not give any information on the asymptotic behavior of $\{f(x_n)\}$ outside the subsequence $\{x_{l_n}\}$ [...]*

Theorem ([Neri and P., 2023])

For any sequence of positive reals $\{\alpha_n\}$ bounded above by $\alpha > 0$ with $\sum_{i=0}^{\infty} \alpha_i = \infty$, together with $\theta > 0$, there exist sequences of positive reals $\{\mu_n\}$ and $\{\beta_n\}$, computable in $\{\alpha_n\}$ and θ and satisfying

$$\mu_{n+1} \leq \mu_n - \alpha_n \beta_n \quad \text{and} \quad \beta_n - \beta_{n+1} \leq \theta \alpha_n$$

such that $\mu_n \to \mu$ and $\beta_n \to 0$, but neither with a computable rate of convergence.

Our result offers a formal explanation of why no such information can be extracted.

Strengthening convergence results

That $\beta_n \to 0$ holds for "gradient-descent type" recursive inequalities can be reduced to the following result:

Theorem (cf. Proposition 2 of [Alber et al., 1998])

Suppose that $\{\alpha_n\}$ and $\{\beta_n\}$ are sequences of nonnegative real numbers with $\sum_{i=0}^{\infty} \alpha_i = \infty$ and $\sum_{i=0}^{\infty} \alpha_i \beta_i < \infty$. Then whenever there exists $\theta > 0$ such that the following condition holds:

$$\beta_n - \beta_{n+1} \leq \theta \alpha_n \text{ for all } n \in \mathbb{N}$$

then $\beta_n \to 0$.

We characterised exactly what kind of additional condition was required to establish convergence:

Theorem ([Neri and P., 2023])

Suppose that $\{\alpha_n\}$ and $\{\beta_n\}$ are sequences of nonnegative real numbers with $\sum_{i=0}^{\infty} \alpha_i = \infty$ and $\sum_{i=0}^{\infty} \alpha_i \beta_i < \infty$. Then $\beta_n \to 0$ if and only if there exists $\theta > 0$ such that

$$\limsup_{N \to \infty} \left\{ \beta_n - \beta_m - \theta \sum_{i=n}^{m-1} \alpha_i \,\middle|\, N \leq n < m \right\} \leq 0$$

Constructivising convergence

Theorem ([Neri and P., 2023])

Suppose that $\{\alpha_n\}$ *and* $\{\beta_n\}$ *are sequences of nonnegative real numbers and* r *is a rate of divergence for* $\sum_{i=0}^{\infty} \alpha_i = \infty$*. Let* $\varepsilon > 0$ *and* $g : \mathbb{N} \to \mathbb{N}$ *be arbitrary. Then if* $N_1, N_2 \in \mathbb{N}$ *and* $\theta > 0$ *are such that, setting* $N := \max\{N_1, N_2\}$*, we have*

$$\sum_{i=N_1}^{r(N+g(N),\varepsilon/4\theta)} \alpha_i \beta_i \leq \frac{\varepsilon^2}{8\theta}$$

and

$$\beta_n - \beta_m \leq \theta \sum_{i=n}^{m-1} \alpha_i + \frac{\varepsilon}{4} \text{ for all } N_2 \leq n < m \leq r\left(N + g(N), \frac{\varepsilon}{4\theta}\right)$$

then we can conclude that $\beta_n \leq \varepsilon$ *for all* $n \in [N, N + g(N)]$.

Generalised gradient descent methods

Theorem ([Neri and P., 2023])

Let X be a real inner product space with $Y \subseteq X$, and suppose that $f : X \to \mathbb{R}$ is a function. Let $\{\alpha_n\}$ be a sequence of nonnegative reals with $\sum_{i=0}^{\infty} \alpha_i = \infty$, and $\{b_n\}$, $\{c_n\}$ and $\{d_n\}$ sequences of nonnegative reals with $\sum_{i=0}^{\infty} \alpha_i b_i < \infty$, $\sum_{i=0}^{\infty} c_i^2 < \infty$ and $\sum_{i=0}^{\infty} d_i < \infty$. Finally, suppose that $x^* \in Y$, $\{x_n\}$ and $\{u_n\}$ are sequences of vectors with $x_n \in Y$ for all $n \in \mathbb{N}$, and $a, p, \theta > 0$ are constants, which satisfy the following properties for all $n \in \mathbb{N}$:

1. $f(x^*) \leq f(x_n)$ *(x^* acts as a minimizer)*
2. $f(x_n) - f(y) \leq \langle u_n, x_n - y \rangle + b_n$ *for all $y \in Y$ (u_n acts as a gradient)*
3. $\|x_{n+1} - x_n\| \leq c_n$ *(x_n acts as a gradient descent method, property I)*
4. $\langle \alpha_n u_n, x_n - x^* \rangle \leq a \langle x_n - x_{n+1}, x_n - x^* \rangle + d_n$ *(x_n acts as a gradient descent method, property II)*
5. $\|u_n\| \leq p$ *(gradients are bounded)*
6. $pc_n + b_n \leq \theta \alpha_n$

Then $f(x_n) \to f(x^*)$.

Generalised gradient descent methods continued...

Theorem (continued...)

Moreover, if r is a rate of divergence for $\sum_{i=0}^{\infty} \alpha_i = \infty$ and $b, c, d, K > 0$ are such that

$$\sum_{i=0}^{\infty} \alpha_i b_i \leq b, \quad \sum_{i=0}^{\infty} c_i^2 \leq c, \quad \sum_{i=0}^{\infty} d_i \leq d, \quad \textit{and} \quad \|x_0 - x^*\|^2 \leq K$$

Then for all $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$ we have

$$\exists n \leq \Phi(\varepsilon, g)\, \forall k \in [n, n + g(n)]\, (f(x_k) \leq f(x^*) + \varepsilon)$$

where

$$\Phi(\varepsilon, g) := \tilde{h}^{(\lceil 4\theta e/\varepsilon^2 \rceil)}(0)$$

$$\tilde{h}(n) := r\left(n + g(n), \frac{\varepsilon}{2\theta}\right) + 1$$

$$e := \frac{a(c+K)}{2} + b + d$$

Discovered while we were writing the paper:

Annual Reviews in Control 53 (2022) 161–186

Contents lists available at ScienceDirect

Annual Reviews in Control

journal homepage: www.elsevier.com/locate/arcontrol

Review article

Convergence of sequences: A survey☆

Barbara Franci [a,*], Sergio Grammatico [b]

[a] Department of Data Science and Knowledge Engineering, Maastricht University, Maastricht, The Netherlands
[b] Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands

ARTICLE INFO

Keywords:
Convergence

ABSTRACT

Convergent sequences of real numbers play a fundamental role in many different problems in system theory, e.g., in Lyapunov stability analysis, as well as in optimization theory and computational game theory. In this survey, we provide an overview of the literature on convergence theorems and their connection with Féjer monotonicity in the deterministic and stochastic settings, and we show how to exploit these results.

1. Introduction

Why Are Convergence Theorems Necessary?
The answer to this "naive" question is not simple.
cit. Boris T. Polyak, 1987 (Polyak, 1987, Section 1.6.2).

While the answer may have become clearer through the years, since many problems in applied mathematics rely on convergence theorems, it is still not simple. Besides the theoretical investigation, in fact, one fundamental aspect is how convergence theorems can be of practical use, i.e., if the assumptions are plausible for a variety of applications, for instance, in systems theory. Moreover, convergence theorems may also give qualitative information, e.g., if convergence is guaranteed for any initial point and in what sense (strongly, weakly, almost surely, in probability), which affects the range of application. The aim of this paper is to collect these results toward a complete

control in traffic networks (Duvocelle, Meier, Staudigl, & Vuong, 2019) and in modeling the prosumer behavior in smart power grids (Franci & Grammatico, 2020a; Franci et al., 2020; Kannan, Shanbhag, & Kim, 2013; Yi & Pavel, 2019).

1.1. Lyapunov decrease and Féjer monotonicity

In the mathematical literature, many convergence results hold for sequences of numbers while in system and control theory, the state and decision variables are usually *vectors* of real numbers. It is therefore important to understand the deep connection between the two theories. The bridging idea is to associate a real number to the state vector, i.e., via a function, and then prove convergence exploiting the properties of such a function. The most common example of this approach is that of Lyapunov theory where a suitable Lyapunov function is shown to be decreasing along the evolution of the state variable, thus

Contains a huge survey of deterministic and stochastic recursive inequalities…

Table 1
Convergence results for Féjer monotone sequences, deterministic sequences of real numbers and with variable metric (separated by the horizontal lines, respectively). For the applications, MI stands for Monotone Inclusion, VI for variational inequalities, NE for Nash Equilibrium problems, LYAP for Lyapunov analysis and NC for nonconvex optimization.

Result	Reference	Application	Reference
Proposition 3.1	Bauschke et al. (2011, Proposition 5.4)		
Theorem 3.2	Combettes (2001b, Theorem 3.8)		
Lemma 3.3	Opial et al. (1967) (Opial)	MI - Theorem 6.1	Malitsky and Tam (2020, Theorem 2.5)
		VI - Theorem 6.4	Malitsky (2020, Theorem 1)
Lemma 3.4	Combettes (2001b, Lemma 3.1)	NC - Theorem 6.9	Di Lorenzo and Scutari (2016, Theorem 3)
Corollary 3.5	Scutari and Sun (2019, Lemma 9)		
Lemma 3.6	Bauschke et al. (2011, Lemma 5.31)	VI - Theorem 6.4	Malitsky (2020, Theorem 1)
Corollary 3.7	Malitsky (2015, Lemma 2.8)	VI - Theorem 6.5	Malitsky (2015, Theorem 3.2)
		LYAP - Theorem 6.8	Polyak (1987, Theorem 1.4.1)
Corollary 3.8	Polyak (1987, Lemma 2.2.2)		
Lemma 3.9	Polyak (1987, Lemma 2.2.3)	NE - Theorem 6.7	Kannan and Shanbhag (2012, Theorem 2.4)
Lemma 3.10			
Lemma 3.11	Xu (2003, Lemma 2.1)		
Lemma 3.12	Extension of Xu (2002, Lemma 2.5)	NE - Theorem 6.6	Duvocelle et al. (2019, Theorem 3.1)
Corollary 3.13	Lei, Shanbhag and Chen (2020, Proposition 3)		
Corollary 3.14	Qin, Shang, and Su (2008, Lemma 1.1)		
Corollary 3.15	Xu (1998, Lemma 3)	MI - Theorem 6.3	Dadashi and Postolache (2019, Theorem 3.1)
Proposition 3.16	Alber, Iusem, and Solodov (1998, Proposition 2)		
Lemma 3.17	He and Yang (2013, Lemma 7)		
Lemma 3.18	Maingé (2008, Lemma 2.2)		
Lemma 3.19	Malitsky and Tam (2018, Lemma 2.7)	MI - Theorem 6.2	Malitsky and Tam (2020, Theorem 2.9)
Proposition 5.1	Combettes and Vũ (2013, Proposition 3.2)	MI - Theorem 8.1	Vũ (2013, Theorem 3.1)
Theorem 5.2	Combettes and Vũ (2013, Theorem 3.3)	MI - Theorem 8.1	Vũ (2013, Theorem 3.1)
Corollary 5.3	Combettes and Vũ (2013, Proposition 4.1)		

constructed sequence from such set can be analyzed anyways. On the contrary, in Lyapunov stability analysis, the target set is usually known a priori.

By exploiting the relation between the iterations and a suitable distance-like function, we show in this paper that convergence theorems represent a key ingredient for a wide variety of system-theoretic problems in fixed-point theory, game theory and optimization (Bauschke, Combettes, et al., 2011; Combettes, 2001b; Eremin & Popov, 2009; Facchinei & Pang, 2007; Polyak, 1987). In many cases, the study of iterative algorithms allows for a systematic analysis that follows from the concept of Féjer monotone sequence. The basic idea behind Féjer monotonicity is that at each step, each iterate is closer to the target set than the previous one. In a sense, the distance used for Féjer

1.2. What this survey is about

In this survey, we present a number of convergence theorems for sequences of real (random) numbers. We show how they can be used in combination with (quasi) Féjer monotone sequences or Lyapunov functions to obtain convergence of an iterative algorithm, essentially a discrete-time dynamical system, to a desired solution. Moreover, we present some applications to show how they can be adopted in a variety of settings. Specifically, we present convergence results for both deterministic and stochastic sequences of real numbers and we also include some results on Féjer monotone sequences and with variable metric. We show that these results help proving not only convergence of an iterative algorithm but also the Law of Large Numbers, with

Contains a huge survey of deterministic and stochastic recursive inequalities...

Table 3
Convergence results for deterministic sequences of real numbers divided by their form. In the first line, the most general inequality is presented. NN stands for a sequence of *nonnegative* real numbers, while ✓(✗) indicates if the inequality in the corresponding lemma contains (or not) a term of that column type. C^k is a general "coefficient", whose specific form can be retrieved from the column.

	Seq$(k+1)$ v^{k+1}	$\leq$	Coeff. C^k	Seq(k) v^k	Negative $-\theta^k$	Noise $+\varepsilon^k$
Lemma 3.4	NN		γ		✓	✓
Lemma 3.6	NN		$(1+\delta^k)$		✓	✓
Corollary 3.7	NN		1		✓	✗
Corollary 3.8	NN		$(1+\delta^k)$		✗	✓
Lemma 3.9	Real		γ^k		✗	✓
Lemma 3.10	NN		$(1-\delta^k)$		✓	✓
Lemma 3.11	NN		$(1-\delta^k)$		✗	$\delta^k\beta^k$
Lemma 3.12	NN		$(1-\delta^k)$		✗	$\delta^k\beta^k+\varepsilon^k$
Corollary 3.13	NN		$(1-\delta^k)$		✗	$\delta^k(\beta^k+\eta^k)$
Corollary 3.14	NN		$(1-\delta^k)$		✗	✓
Corollary 3.15	NN		$(1-\delta^k)$		✗	$\eta^k+\varepsilon^k$
Proposition 3.16	NN		1		✗	$a\beta^k$
Lemma 3.17	NN		$(1-\delta^k)$		✗	$\delta^k\gamma^k+\beta^k$
			1		✓	✓
Lemma 3.18	Real		$(1+\delta^k)$		$\delta^k v^{k-1}$	✓
Lemma 3.19	NN		$1/\gamma$		β^{k+1}/γ	β^k/γ

3.2. Convergent sequences of real numbers

We now introduce a number of results on sequences of real numbers. We note that even if the following results are for general sequences of real numbers, their importance for system theory lies on the fact that they can be paired with (quasi) Féjer monotonicity (see Remark 3.5). In Table 3, we summarize the results presented in this section, with emphasis on the auxiliary sequences that may affect convergence.

Let us note that, in the first line of Table 3, C^k is a coefficient which, depending on the form, represents the level of expansion or contraction, ε^k can be seen as an additive noise and θ^k is a "negative term", because of the minus sign, which decreases the value of the sequence v^k. For a graphical interpretation of the effects of those sequences, we also refer to Fig. 4 later on, which is specifically related to Lemma 3.6.

The first lemma that we report is widely used and it has a number of consequences that are widely used as well. We do not include the proof since it is very similar to the proof of the forthcoming Lemma 3.10.

where $\gamma \in (0,1)$, $(\eta^k)_{k\in\mathbb{N}}$ is a decreasing positive sequence such that $\sum_{k=0}^{\infty}(\eta^k)^2 < \infty$, and let $0 \leq v^k \leq \bar{v} < \infty$ for all $k \in \mathbb{N}$. Then, $\sum_{k=1}^{\infty}\eta^k v^k < \infty$.

With the same arguments as for Lemma 3.4, the following corollary can be proven. Interestingly, this result concerns the finite sum of the sequence.

Corollary 3.5 (*Lemma 9, Scutari & Sun, 2019*). *Let $(v^k)_{k\in\mathbb{N}}$ be a real sequence and let $(\theta^k)_{k\in\mathbb{N}}$ and $(\varepsilon^k)_{k\in\mathbb{N}}$ be nonnegative sequences such that $\sum_{k=0}^{\infty}\varepsilon^k < \infty$ and such that*

$$\sum_{n=0}^{N-1} v^{k+N+n} \leq \sum_{n=0}^{N-1} v^{k+n} - \sum_{n=0}^{N-1} \theta^{k+n} + \sum_{n=0}^{N-1} \varepsilon^{k+n}.$$

for $N \in \mathbb{N}$. Then, either $\sum_{n=0}^{N-1} v^{k+n} \to -\infty$, or $\sum_{n=0}^{N-1} v^{k+n}$ converges to a finite value and $\sum_{k=0}^{\infty}\theta^k < \infty$.

Proof. It suffices to set $v_1^k = \sum_{n=0}^{N-1} v^{k+n}$, $\theta_1^k = \sum_{n=0}^{N-1}\theta^{k+n}$ and $\varepsilon_1^k = \sum_{n=0}^{N-1}\varepsilon^{k+n}$ and then apply Lemma 3.4. □

The next lemma is a consequence and a generalization of Lemma 3.4. It has its stochastic counterpart in the well know Robbins–Siegmund Lemma (Lemma 4.1) (Robbins & Siegmund, 1971). It is taken from Bauschke et al. (2011) yet here we provide a different proof. For a graphical interpretation, we refer to Fig. 4.

Lemma 3.6 (*Lemma 5.31, Bauschke et al., 2011*). *Let $(v^k)_{k\in\mathbb{N}}$, $(\theta^k)_{k\in\mathbb{N}}$, $(\varepsilon^k)_{k\in\mathbb{N}}$ and $(\delta^k)_{k\in\mathbb{N}}$ be nonnegative sequences such that $\sum_{k=0}^{\infty}\varepsilon^k < \infty$ and $\sum_{k=0}^{\infty}\delta^k < \infty$ and*

$$v^{k+1} \leq (1+\delta^k)v^k - \theta^k + \varepsilon^k, \text{ for all } k \in \mathbb{N}. \quad (3.2)$$

Then, $\sum_{k=0}^{\infty}\theta^k < \infty$ and $(v^k)_{k\in\mathbb{N}}$ is bounded and converges to a nonnegative variable.

Proof. Define $\beta^k = \prod_{i=1}^{k}(1+\delta^i)$ and note that β^k converges to some $\bar{\beta}$ since $(\delta^k)_{k\in\mathbb{N}}$ is summable. Moreover, it holds that

$$1+\delta^k = \frac{\beta^k}{\beta^{k-1}}$$

… along with general heuristics for using them:

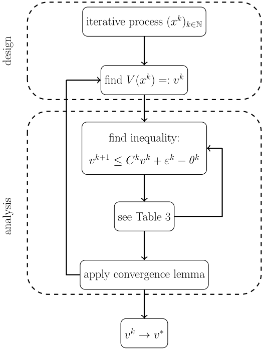

Fig. 7. Schematic representation of how the convergence lemmas for sequences can be used. Given the iterative process, a suitable nonnegative function (Lyapunov or distance-like) should be designed. Then, exploiting the properties of the application at hand, an inequality involving the iterates at times $k+1$ and k can be retrieved. Hence, one should check if the inequality corresponds to a known result (Table 3 for sequences of real numbers) and use the corresponding result to prove convergence. The whole process may take repeated steps to find a suitable function and/or inequality. The same reasoning applies to the stochastic case, in which one should have an expected valued inequality (with $\mathbb{E}[v^{k+1}]$) and refer to Table 4 for a convergence result on stochastic sequences. See also Fig. 8 for an example.

Proof. Let $x^* \in (A+B)^{-1}(0)$. It is possible to show, by using monotonicity and some norm properties, that the following inequality holds:

$$\|x^{k+1} - x^*\|^2 + 2\alpha_k \langle B(x^{k+1}) - B(x^k), x^* - x^{k+1}\rangle +$$
$$+ \left(\frac{1}{2} + \epsilon\right) \|x^{k+1} - x^k\|^2$$

of strong monotonicity of the operator A, they also prove convergence with linear rate, using Lemma 3.19.

Theorem 6.2 (*Theorem 2.9, Malitsky & Tam, 2020*). *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone and μ-strongly monotone and $B : \mathcal{H} \to \mathcal{H}$ be monotone and ℓ-Lipschitz continuous. Suppose $\alpha \in \left(0, \frac{1}{2\ell}\right)$. Then, the sequence $(x^k)_{k\in\mathbb{N}}$ generated by (6.8) converges R-linearly to the unique point $\bar{x} \in \mathcal{X}$ such that $0 \in (A+B)(\bar{x})$.*

Proof. Similarly to the proof of Theorem 6.1 but using strong monotonicity, one obtains the inequality

$$\begin{aligned}(1+2\mu\alpha)\|x^{k+1} - x^*\|^2 &+ 2\alpha\langle B(x^{k+1}) - B(x^k), x^* - x^{k+1}\rangle \\ &+ (1-\alpha\ell)\|x^{k+1} - x^k\|^2 \\ \le \|x^k - x^*\|^2 &+ 2\alpha\langle B(x^k) - B(x^{k-1}), x^* - x^k\rangle \\ &+ \frac{1}{2}\|x^k - x^{k-1}\|^2.\end{aligned} \tag{6.10}$$

Setting $\gamma = (1+2\mu\alpha) > 1$, $v_k := \frac{1}{2}\|x^k - x^*\|^2$ and $\beta_k := \frac{1}{2}\|x^k - x^*\|^2 + 2\alpha\langle B(x^k) - B\left(x^{k-1}\right), x^* - x^k\rangle + \frac{1}{2}\|x^k - x^{k-1}\|^2$, one can apply Lemma 3.19 to conclude that the sequence $(x^k)_{k\in\mathbb{N}}$ converges to the unique solution $\bar{x}$ and with a linear rate. □

Application of Corollary 3.15. As an application of Corollary 3.15, let us consider the *inertial forward–backward* algorithm proposed in Dadashi and Postolache (2019) for approximating a zero of an inclusion problem $x \in (A+B)^{-1}(0)$:

$$\begin{cases} y^k = J_{\alpha_k A}\left(x^k - \alpha_k B x^k\right) \\ x^{k+1} = \nu_k x^k + \beta_k y^k + \gamma_k e^k \end{cases} \tag{6.11}$$

where $J_{\alpha_k A}$ is the resolvent of A (Definition A.1) and e^k is an error vector. By using Corollary 3.15 the authors prove the following result.

Theorem 6.3 (*Theorem 3.1, Dadashi & Postolache, 2019*). *Let B be α-cocoercive and let A be maximally monotone. Let $\nu_k, \beta_k, \gamma_k \in (0,1)$ be such that $\nu_k + \beta_k + \gamma_k = 1$ and*

1. $\lim_{k\to\infty} \gamma_k = 0$, *and* $\sum_{k=1}^{\infty} \gamma_k = \infty$,
2. $\lim_{k\to\infty} e^k = 0$,
3. $0 < a \le \nu_k \le b < 1$ *and* $0 < c \le \beta_k \le d < 1$,
4. $0 < c \le \alpha_k < 2\alpha$ *and* $\lim_{k\to\infty}\left(\alpha_k - \alpha_{k+1}\right) = 0$.

Then, the sequence $(x^k)_{k\in\mathbb{N}}$ generated by (6.11) converges to the point $x^* \in (A+B)^{-1}(0)$, where $x^* = \text{proj}$ … (0)

Thoughts for the future

The analysis of further recursive inequalities

A comprehensive survey paper on recursive inequalities for applied proof theory would certainly be valuable! But there are also plenty of new directions to look at.

Particularly interesting would be **stochastic algorithms**. These rely heavily on things like the *Robbins-Siegmund lemma* (which in turn relies on Martingale theory):

Lemma (Robbins-Siegmund 1971)

Let $\{\mu_n\}$, $\{\delta_n\}$, $\{\varepsilon_n\}$ *and* $\{\theta_n\}$ *be sequences of nonnegative reals such that* $\sum_{i=0}^{\infty} \varepsilon_i < \infty$, $\sum_{i=0}^{\infty} \delta_i < \infty$

$$E[\mu_{n+1} \mid \mathcal{F}_n] \leq (1 + \delta_n)\mu_n + \varepsilon_n - \theta_n \quad a.s.$$

for some filtration $\{\mathcal{F}_n\}$. *Then* $\sum_{i=0}^{\infty} \theta_i < \infty$ *and* $\{\mu_n\}$ *converges a.s.*

- Can we give results of this kind a computational interpretation?
- Are there applications in stochastic optimization?

Formalizing applied proof theory (and nonlinear analysis!)

There are lots of big efforts on formalising program extraction (in Minlog, Coq, Isabelle, Nuprl, ...).

However, I'm aware of only three people developing libraries of formal proofs specifically for the "proof mining" branch of applied proof theory:

- H. Cheval: `https://github.com/hcheval`
- A. Koutsoukou-Argyraki: [Koutsoukou-Argyraki, 2021]
- M. Neri: `https://github.com/mneri123/Proof-mining-`

Building a comprehensive library on convergence results for sequences of reals (along with rates of convergence/metastability) would be extremely useful:

- Many results in both areas reduce to lemmas on recursive inequalities. Formalizing these provide a solid base for more extensive formalization work.
- This would not need to rely on advanced libraries: It's enough to have the basic theory of real numbers, infinite series etc.
- Some convergence proofs could be given to good undergraduate students for projects.

Some initial progress:

```
lemma abstract_lemma1 (θ : nnseq )(α : nnseq)(K: {x: ℝ // x > 0}) (r: ℕ → {x: ℝ // x > 0} → ℕ )
(N : {x:ℝ // x > 0 }  → ℕ ) (ϕ : {x:ℝ // x > 0 }  → {x:ℝ // x > 0 })
(h1 : ∀(n:ℕ), (θ.1 n) < K) (h2: RoD r α)
(h3: ∀ ε : {x:ℝ // x > 0 } , ∀ n ≥ N(ε), (ε:ℝ) < θ.1 (n + 1)  → θ.1 (n + 1) ≤ θ.1 n - (α.1 n)*ϕ(ε)):
RoC (λ ε : {x: ℝ // x > 0},  (r (N ε) ⟨K/(ϕ ε), div_pos K.2 (ϕ ε).2 ⟩+1)) θ  :=
begin
have
H1 : ∀ ε : {x:ℝ // x > 0 },∀ n ≥ N(ε), θ.1 n ≤ ε  → θ.1 (n + 1) ≤ ε,
by_contradiction p1,
push_neg at p1,
cases p1 with ε p2,
cases p2 with n p3,
have p5 : ε < ε,
calc ε.1 < θ.1 (n+1): (p3.2).2
... ≤ θ.1 n - (α.1 n)*ϕ(ε): h3 ε n p3.1 (p3.2).2
... ≤ θ.1 n : sub_le_self (θ.1 n) (mul_nonneg (α.2 n) (le_of_lt (ϕ ε).2 ))
... ≤ ε :(p3.2).1,
exact (lt_self_iff_false ε).mp p5,
have H2 : ∀ ε : {x :ℝ // x > 0}, ∃ n ∈ finset.Ico (N ε) ((r (N ε) ⟨↑K / ↑(ϕ ε), _⟩)+1), θ.1 (n + 1) ≤
by_contradiction,
push_neg at h,
```

Automating the reduction to (quantitative) lemmas

The reduction of e.g. $\{\|x_n - q\|\}$ to some recursive inequality often uses little more than routine calculations and properties of the mapping, algorithm and underlying space.

- Can we develop algorithms for automating this procedure?
- Are there new logics for reasoning about abstract spaces that would be helpful?
- This could also then automate bound extraction?

Conclusion

Three possible directions for future research that each reinforce the other:

1. The proof theoretic analysis of new recursive inequalities, particularly in the stochastic setting.
2. A formal library of lemmas on convergent sequences of real numbers.
3. Automating the reduction of concrete algorithms to recursive inequalities.

THANK YOU!

References I

Alber, Y., Chidume, C., and Zegeye, H. (2006).
Approximating fixed points of total asymptotically nonexpansive mappings.
Fixed Point Theory and Applications, 2006:1687–1812.

Alber, Y. and Guerre-Delabriere, S. (1997).
Principle of weakly contractive maps in Hilbert spaces.
In Gohberg, I. and Lyubich, Y., editors, *New Results in Operator Theory and its Applications*, volume 98, pages 7–22.

Alber, Y. and Iusem, A. N. (2001).
Extension of subgradient techniques for nonsmooth optimization in Banach spaces.
Set-Valued Analysis, 9:315–335.

Alber, Y., Iusem, A. N., and Solodov, M. V. (1998).
On the projected subgradient method for nonsmooth convex optimization in a Hilbert space.
Mathematical Programming, 81(1):23–35.

Franci, B. and Grammatico, S. (2022).
Convergence of sequences: A survey.
Annual Reviews in Control, 53:161–186.

References II

Kohlenbach, U. (2008).
Applied Proof Theory: Proof Interpretations and their Use in Mathematics.
Springer Monographs in Mathematics. Springer.

Kohlenbach, U. (2017).
Recent progress in proof mining in nonlinear analysis.
IFCoLoG Journal of Logics and their Applications, 10(4):3361–3410.

Kohlenbach, U. (2019).
Proof-theoretic methods in nonlinear analysis.
In *Proceedings of the International Congress of Mathematicians 2018*, volume 2, pages 61–82. World Scientific.

Kohlenbach, U. and Lambov, B. (2004).
Bounds on iterations of asymptotically quasi-nonexpansive mappings.
In *Proceedings of the International Conference on Fixed Point Theory and Applications*, pages 143–172. Yokohama Publishers.

References III

Koutsoukou-Argyraki, A. (2021).
On preserving the computational content of mathematical proofs: Toy examples for a formalising strategy.
In *Connecting with Computability. CiE 2021*, volume 12813 of *LNCS*, pages 285–296. Springer.

Kreisel, G. (1951).
On the interpretation of non-finitist proofs, Part I.
Journal of Symbolic Logic, 16:241–267.

Kreisel, G. (1952).
On the interpretation of non-finitist proofs, Part II: Interpretation of number theory.
Journal of Symbolic Logic, 17:43–58.

Lombardi, H. and Quitté, C. (2015).
Commutative Algebra: Constructive Methods, volume 20 of *Algebra and Applications*.
Springer Netherlands.

Longley, J. and Normann, D. (2015).
Higher-Order Computability.
Theory and Applications of Computability. Springer-Verlag Berlin Heidelberg.

References IV

Neri, M. and Powell, T. (2023).
A computational study of a class of recursive inequalities.
Preprint, available at https://arxiv.org/abs/2207.14559.

Powell, T. and Wiesnet, F. (2021).
Rates of convergence for asymptotically weakly contractive mappings in normed spaces.
Numerical Functional Analysis and Optimization, 42(15):1802–1838.

Schwichtenberg, H. and Wainer, S. (2011).
Proofs and Computations.
Perspectives in Logic. Cambridge University Press.

Sipoş, A. (2022).
Abstract strongly convergent variants of the proximal point algorithm.
Computational Optimization and Applications, 83(1):349–380.

Specker, E. (1949).
Nicht konstruktiv beweisbare Sätze der Analysis.
Journal of Symbolic Logic, 14:145–158.

References V

Tao, T. (2007).
Soft analysis, hard analysis, and the finite convergence principle.
Essay posted 23 May 2007.
Appeared in: 'T. Tao, Structure and Randomness: Pages from Year One of a Mathematical Blog. AMS, 298pp., 2008'.