Quantitative results for stochastic processes

Thomas Powell
(jww Morenikeji Neri)

University of Bath

Mathematisches Kolloquium

TU Darmstadt
17 July 2024

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 an extremely 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 completely routine example – from my own work done here at Darmstadt!

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

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

Modern applied proof theory

- 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).
- There are now hundreds of papers proving new theorems that were obtained using proof theoretic ideas and methods, the majority published in specialised journals in the areas of application, including nonlinear analysis, ergodic theory, convex optimization, . . . (see the proof mining bibliography).
- In parallel, there are logical metatheorems that explain individual applications as instances of general logical phenomena (the first in [Kohlenbach, 2005] and the most recent in [Neri and Pischke, 2024]).
- Now starting to expand and establish new connections with automated reasoning and formal mathematics [Koutsoukou-Argyraki, 2021, Neri and Powell, 2023].

What people working in applied proof theory might do

- Use logical methods to establish quantitative versions of known results in mainstream (non-logic) mathematics.
- Show that a collection of theorems are all instances of a more general, abstract theorem.
- Define new classes of mappings or new types of spaces.
- Develop sophisticated logical systems for reasoning about specific mathematical objects.
- Study a hitherto unexplored area of mathematics to see if proof theoretic methods might be effective and useful. This is **very hard** but **very rewarding** when it works.
- Make important contributions to core logic, including computability theory and theoretical computer science.
- Recently: Formalise their work in a proof assistant (e.g. Lean) or consider automated methods.

Applied proof theory is characterised by thinking about and doing mathematics from a proof-theoretic perspective.

The rest of this talk

One line summary: Applied proof theory has never really tackled probability in a systematic way. We've now started to do this and it is a <u>lot</u> of fun.

- Metastable monotone convergence.
- Metastable martingale convergence.
- Almost-martingales in stochastic optimization.
- The future: Proof mining in probability theory.

Metastable monotone convergence

Monotone convergence theorem

Theorem (First year analysis)

Let $K > 0$ and suppose that $\{x_n\}$ is a monotone sequence of reals with $|x_n| \leq K$ for all $n \in \mathbb{N}$. Then $\{x_n\}$ converges.

Is there a computable, uniform rate of convergence for all sequences in this class? I.e. a function $\phi_K(\varepsilon)$ such that

$$\forall \varepsilon > 0 \, \forall i, j \geq \phi_K(\varepsilon)(|x_i - x_j| < \varepsilon)$$

Absolutely not:

- There exist monotone sequences of *rationals* with $|x_n| \leq 1$ which do not possess a computable rate of convergence (Specker sequences - already mentioned in Nicholas' introduction).
- For any function $\phi(\varepsilon)$ we can construct a montone sequence of rationals in $[0, 1]$ which do not convergence with rate $\phi(\varepsilon)$.

We need to consider a different notion of quantitative convergence.

A logical approach to the MCT – The statement

The following steps are entirely logic-based (i.e. have nothing to do with convergence):

$$
\begin{aligned}
MCT &:= \forall \varepsilon > 0\, \exists n\, \forall i, j \geq n\, (|x_i - x_j| < \varepsilon) \\
&\iff \forall \varepsilon > 0\, \exists n\, \forall k\, \forall i, j \in [n; n+k]\, (|x_i - x_j| < \varepsilon) \\
&\iff \neg\neg \forall \varepsilon > 0\, \exists n\, \forall k\, \forall i, j \in [n; n+k]\, (|x_i - x_j| < \varepsilon) \\
&\iff \neg \exists \varepsilon > 0\, \forall n\, \exists k\, \exists i, j \in [n; n+k]\, (|x_i - x_j| \geq \varepsilon) \\
&\iff \neg \exists \varepsilon > 0\, \exists g : \mathbb{N} \to \mathbb{N}\, \forall n\, \exists i, j \in [n; n+g(n)]\, (|x_i - x_j| \geq \varepsilon) \\
&\iff \forall \varepsilon > 0\, \forall g : \mathbb{N} \to \mathbb{N}\, \exists n\, \forall i, j \in [n; n+g(n)]\, (|x_i - x_j| < \varepsilon) \ := MCT^*
\end{aligned}
$$

Theorem (Metastable monotone convergence theorem – first version)

Let $K > 0$ and suppose that $\{x_n\}$ is a monotone sequence of reals with $|x_n| \leq K$ for all $n \in \mathbb{N}$. Then for any $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$ there exists some $n \in \mathbb{N}$ such that $|x_i - x_j| < \varepsilon$ for all $i, j \in [n; n+g(n)]$.

Question: Can we compute n in ε and g?

A logical approach to the MCT – The proof

Suppose that $\{x_n\} \subset [-K, K]$ is monotone but not Cauchy. Then there is some $\varepsilon > 0$ such that for all $n \in \mathbb{N}$, we can find $k \in \mathbb{N}$ with:

$$\exists i, j \in [n; n+k] \, (|x_i - x_j| \geq \varepsilon)$$

Let $g : \mathbb{N} \to \mathbb{N}$ be a function that finds such a k in terms of n i.e. for all $n \in \mathbb{N}$:

$$\exists i, j \in [n; \tilde{g}(n)] \, (|x_i - x_j| \geq \varepsilon)$$

for $\tilde{g}(x) := x + g(n)$.

Then iterating $\tilde{g}$, for all $e \in \mathbb{N}$:

$$\exists i, j \in [\tilde{g}^{(e)}(0); \tilde{g}^{(e+1)}(0)] \, (|x_i - x_j| \geq \varepsilon) \tag{$*$}$$

In other words, in each of the intervals

$$[0; \tilde{g}(0)], [\tilde{g}(0); \tilde{g}^{(2)}(0)], [\tilde{g}^{(2)}(0); \tilde{g}^{(3)}(0)], \ldots$$

the sequence $\{x_n\}$ experiences a distinct ε-jump (or *fluctuation*).

But a monotone sequence in $[-K, K]$ can experience at most $2K/\varepsilon$ distinct ε-fluctuations, so $(*)$ must fail for some $e \leq \lceil 2K/\varepsilon \rceil$. **Contradiction!**

A logical approach to the MCT – The payoff

Theorem (Metastable monotone convergence theorem)

Take $K, \varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$. Then there exists some $N \in \mathbb{N}$ (depending only on K, ε and g) such that for any monotone sequence $\{x_n\}$ in $[-K, K]$, there exists $n \leq N$ such that $|x_i - x_j| < \varepsilon$ for all $i, j \in [n; n + g(n)]$. Moreover, we can define

$$N_K(\varepsilon, g) := \tilde{g}^{(\lceil 2K/\varepsilon \rceil)}(0)$$

for $\tilde{g}(x) := x + g(x)$.

Theorem (Generalised metastable convergence theorem)

Take $\phi : (0, 1) \to \mathbb{R}$, $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$. Then there exists some $N \in \mathbb{N}$ such that ***for any sequence $\{x_n\}$ in some metric space (X, d) that experiences at most $\phi(\varepsilon)$ distinct ε-fluctuations,*** *there exists $n \leq N$ such that $d(x_i, x_j) < \varepsilon$ for all $i, j \in [n; n + g(n)]$. Moreover, we can define*

$$N_\phi(\varepsilon, g) := \tilde{g}^{(\lceil \phi(\varepsilon) \rceil)}(0)$$

for $\tilde{g}(x) := x + g(x)$.

[Tao, 2007] (Tao's blog)

Soft analysis, hard analysis, and the finite convergence principle

23 May, 2007 in expository, math.CA, math.CO, math.LO, opinion | Tags: finite convergence principle, hard analysis, pigeonhole principle, proof theory, Ramsey theory, soft analysis | by Terence Tao

In the field of analysis, it is common to make a distinction between "hard", "quantitative", or "finitary" analysis on one hand, and "soft", "qualitative", or "infinitary" analysis on the other. "Hard analysis" is mostly concerned with finite quantities (e.g. the cardinality of finite sets, the measure of bounded sets, the value of convergent integrals, the norm of finite-dimensional vectors, etc.) and their *quantitative* properties (in particular, upper and lower bounds). "Soft analysis", on the other hand, tends to deal with more infinitary objects (e.g. sequences, measurable sets and functions, σ-algebras, Banach spaces, etc.) and their *qualitative* properties (convergence, boundedness, integrability, completeness, compactness, etc.). To put it more symbolically, hard analysis is the mathematics of ε, N, $O()$, and $\leq$[1]; soft analysis is the mathematics of 0, ∞, $\in$, and $\to$.

At first glance, the two types of analysis look very different; they deal with different types of objects, ask different types of questions, and seem to use different techniques in their proofs. They even use[2] different axioms of mathematics; the axiom of infinity, the axiom of choice, and the Dedekind completeness axiom for the real numbers are often invoked in soft analysis, but rarely in hard analysis. (As a consequence, there are occasionally some finitary results that can be proven easily by soft analysis but are in fact *impossible* to prove via hard analysis methods; the Paris-Harrington theorem gives a famous example.) Because of all these differences, it is common for analysts to specialise in only one of the two types of analysis. For instance, as a general rule (and with notable exceptions), discrete mathematicians, computer scientists, real-variable harmonic analysts, and analytic number

[Tao, 2007] (Tao's blog)

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 $1/\varepsilon$.

Informally, the above principle asserts that any sufficiently long (but finite) bounded monotone sequence will experience arbitrarily high-quality amounts of metastability with a specified error tolerance ε, in which the duration F(N) of the metastability exceeds the time N of onset of the metastability by an arbitrary function F which is specified in advance.

Let us now convince ourselves that this is the true finitary version of the infinite convergence principle, by deducing them from each other:

[Tao, 2008]: A convergence theorem…

NORM CONVERGENCE OF MULTIPLE ERGODIC AVERAGES FOR COMMUTING TRANSFORMATIONS

TERENCE TAO

ABSTRACT. Let $T_1, \dots, T_l : X \to X$ be commuting measure-preserving transformations on a probability space $(X, \mathcal{X}, \mu)$. We show that the multiple ergodic averages $\frac{1}{N}\sum_{n=0}^{N-1} f_1(T_1^n x) \dots f_l(T_l^n x)$ are convergent in $L^2(X, \mathcal{X}, \mu)$ as $N \to \infty$ for all $f_1, \dots, f_l \in L^\infty(X, \mathcal{X}, \mu)$; this was previously established for $l = 2$ by Conze and Lesigne [3] and for general l assuming some additional ergodicity hypotheses on the maps T_i and $T_i T_j^{-1}$ by Frantzikinakis and Kra [4] (with the $l = 3$ case of this result established earlier in [30]). Our approach is combinatorial and finitary in nature, inspired by recent developments regarding the hypergraph regularity and removal lemmas, although we will not need the full strength of those lemmas. In particular, the $l = 2$ case of our arguments are a finitary analogue of those in [3].

1. Introduction

The purpose of this paper is to establish the following norm convergence result for multiple commuting transformations.

Theorem 1.1 (Norm convergence). *Let $l \geq 1$ be an integer. Assume that $T_1, \dots, T_l : X \to X$ are commuting invertible measure-preserving transformations of a measure space $(X, \mathcal{X}, \mu)$. Then for any $f_1, \dots, f_l \in L^\infty(X, \mathcal{X}, \mu)$, the averages*

$$\frac{1}{N}\sum_{n=0}^{N-1} f_1(T_1^n x) \dots f_l(T_l^n x)$$

are convergent in $L^2(X, \mathcal{X}, \mu)$.

[Tao, 2008] … in metastable form

Theorem 1.6 (Finitary norm convergence). *Let $l \geq 1$ be an integer, let $F : \mathbf{N} \to \mathbf{N}$ be a function, and let $\varepsilon > 0$. Then there exists an integer $M^* > 0$ with the following property: If $P \geq 1$ and $f_1, \ldots, f_l : \mathbf{Z}_P^l \to [-1, 1]$ are arbitrary functions on $\mathbf{Z}_P^l$, then there exists an integer $1 \leq M \leq M^*$ such that we have the "L^2 metastability"*

$$\|A_N(f_1, \ldots, f_l) - A_{N'}(f_1, \ldots, f_l)\|_{L^2(\mathbf{Z}_P^l)} \leq \varepsilon \tag{1}$$

for all $M \leq N, N' \leq F(M)$, where we give $\mathbf{Z}_P^l$ the uniform probability measure.

Remark 1.7. For applications, Theorem 1.6 is only of interest in the regime where $F(M)$ is much larger than M, and P is extremely large compared to l, F, or ε. The key points are that the function F is arbitrary (thus one has arbitrarily high quality regions of L^2 metastability), and that the upper bound M^* on M is independent of P. The $l = 1$ version of this theorem was essentially established (with $\mathbf{Z}_P^l$ replaced by an arbitrary measure-preserving system) in [1].

Later there is a footnote…

[1]In proof theory, this finitisation is known as the *Gödel functional interpretation* of the infinitary statement, which is also closely related to the *Kriesel no-counterexample interpretation* [14], [15] or *Herbrand normal form* of such statements; see [13] for further discussion. We thank Ulrich Kohlenbach for pointing out this connection.

[Avigad and Rute, 2014] One of many examples of explicit bounds on fluctuations and metastable rates (authors are logicians)

Oscillation and the mean ergodic theorem for uniformly convex Banach spaces

JEREMY AVIGAD† and JASON RUTE‡

† *Philosophy and Mathematical Sciences, Carnegie Mellon University, Pittsburgh, USA*
(e-mail: avigad@cmu.edu)
‡ *Department of Mathematics, Pennsylvania State University, State College, PA 16802, USA*

(*Received* 8 *May* 2013 *and accepted in revised form* 20 *August* 2013)

Abstract. Let $\mathbb{B}$ be a p-uniformly convex Banach space, with $p \geq 2$. Let T be a linear operator on $\mathbb{B}$, and let $A_n x$ denote the ergodic average $(1/n)\sum_{i<n} T^n x$. We prove the following variational inequality in the case where T is power bounded from above and below: for any increasing sequence $(t_k)_{k\in\mathbb{N}}$ of natural numbers we have $\sum_k \|A_{t_{k+1}}x - A_{t_k}x\|^p \leq C\,\|x\|^p$, where the constant C depends only on p and the modulus of uniform convexity. For T a non-expansive operator, we obtain a weaker bound on the number of ε-fluctuations in the sequence. We clarify the relationship between bounds on the number of ε-fluctuations in a sequence and bounds on the rate of metastability, and provide lower bounds on the rate of metastability that show that our main result is sharp.

[Kohlenbach and Safarik, 2014] A deep explanation of the underlying logical phenomena

Fluctuations, effective learnability and metastability in analysis

Ulrich Kohlenbach *, Pavol Safarik

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

A R T I C L E   I N F O

Article history:
Available online 20 August 2013

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

MSC:
03F10
03F60
47H25
37A30

Keywords:
Fluctuations
Effective learnability
Metastability
Proof mining
Nonlinear ergodic theory
Hard analysis

A B S T R A C T

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

Metastable convergence: The main points

- There are natural situations where it is impossible to provide computable rates of convergence.
- Where direct rates are not possible, one can often produce either **fluctuation bounds** or **metastable rates** that are both computable and highly uniform.
- Mathematicians outside of logic are very interested in fluctuations and metastability.
- Researchers in applied proof theory have been able to:
  - Extract explicit fluctuation bounds or metastable rates in many different scenarios;
  - Do this in an abstract and general setting;
  - Explain why this is possible from a logical point of view.
- The apparently elementary monotone convergence theorem contained a wealth of riches when analysed from a logical perspective!

In detail: Metastable martingale convergence

Martingales

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}$ be a filtration. Let $\{X_n\}$ be a sequence of real-valued random variables adapted to $\{\mathcal{F}_n\}$ (i.e. X_n is $\mathcal{F}_n$-measurable) such that $\mathbb{E}[|X_n|] < \infty$ for all $n \in \mathbb{N}$.

We call $\{X_n\}$ a *martingale* if

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n \quad \text{almost surely}$$

for all $n \in \mathbb{N}$. It is a *submartingale* if $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \geq X_n$ and a *supermartingale* if $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \leq X_n$.

Example

Suppose that a gambler repeatedly flips a biased coin repeatedly, winning one euro with probability p and losing one euro with probability $1 - p$ each time. Let X_n be their fortune after n coin flips.

- If $p = 1/2$ then $\{X_n\}$ is a martingale.
- If $p > 1/2$ then $\{X_n\}$ is a submartingale.
- If $p < 1/2$ then $\{X_n\}$ is a supermartingale.

Martingale convergence

Martingales are the stochastic analogue of monotone sequences.

Theorem (Our old friend the monotone convergence theorem)

Let $K > 0$ and suppose that $\{x_n\}$ is a monotone sequence of reals with $|x_n| \leq K$ for all $n \in \mathbb{N}$. Then $\{x_n\}$ converges.

Theorem (The stochastic analogue: Doob's convergence theorem)

Let $K > 0$ and suppose that $\{X_n\}$ is a sub- or supermartingale with $\mathbb{E}[|X_n|] < K$ for all $n \in \mathbb{N}$. Then $\{X_n\}$ converges almost surely i.e.

$$\mathbb{P}\left(\{\omega \in \Omega \mid \{X_n(\omega)\} \text{ converges}\}\right) = 1$$

Note: Martingales generalise monotone sequences of reals, so we also cannot expect direct rates of (almost sure) convergence...

A logical approach to Doob - The statement

The following steps are logic-based or use continuity properties of $\mathbb{P}$:

$$
\begin{aligned}
& \mathbb{P}\left(\{X_n\} \text{ converges}\right) = 1 \\
\iff & \mathbb{P}\left(\bigcap_{m=0}^{\infty} \bigcup_{n=0}^{\infty} \bigcap_{k=0}^{\infty} \forall i,j \in [n;n+k] \left(|X_i - X_j| < 2^{-m}\right)\right) = 1 \\
\iff & \forall m \left[\mathbb{P}\left(\bigcup_{n=0}^{\infty} \bigcap_{k=0}^{\infty} \forall i,j \in [n;n+k] \left(|X_i - X_j| < 2^{-m}\right)\right) = 1\right] \\
\iff & \forall m, \lambda > 0 \, \exists n \left[\mathbb{P}\left(\bigcap_{k=0}^{\infty} \forall i,j \in [n;n+k] \left(|X_i - X_j| < 2^{-m}\right)\right) > 1 - \lambda\right] \\
\iff & \forall \varepsilon, \lambda > 0 \, \exists n \, \forall k \, \left[\mathbb{P}\left(\forall i,j \in [n;n+k] \left(|X_i - X_j| < \varepsilon\right)\right) > 1 - \lambda\right] \\
\iff & \neg \exists \varepsilon, \lambda > 0 \, \forall n \, \exists k \, \left[\mathbb{P}\left(\forall i,j \in [n;n+k] \left(|X_i - X_j| < \varepsilon\right)\right) \leq 1 - \lambda\right] \\
\iff & \neg \exists \varepsilon, \lambda > 0 \, \exists g : \mathbb{N} \to \mathbb{N} \, \forall n \, \left[\mathbb{P}\left(\forall i,j \in [n;n+g(n)] \left(|X_i - X_j| < \varepsilon\right)\right) \leq 1 - \lambda\right] \\
\iff & \forall \varepsilon, \lambda > 0 \, \forall g : \mathbb{N} \to \mathbb{N} \, \exists n \, \left[\mathbb{P}\left(\forall i,j \in [n;n+g(n)] \left(|X_i - X_j| < \varepsilon\right)\right) > 1 - \lambda\right]
\end{aligned}
$$

The goal

By analysing the proof of Doob's theorem can we prove the following?

Theorem (Metastable martingale convergence theorem)

Take $K, \varepsilon, \lambda > 0$ and $g : \mathbb{N} \to \mathbb{N}$. Then there exists some $N \in \mathbb{N}$ (depending only on K, ε, λ and g) such that for any sub- or supermartingale $\{X_n\}$ with

$$\sup_{n \in \mathbb{N}} \mathbb{E}[|X_n|] < K$$

there exists $n \leq N$ such that

$$|X_i - X_j| < \varepsilon \quad \text{for all} \quad i, j \in [n; n + g(n))$$

with probability $> 1 - \lambda$. Moreover, we can define

$$N_K(\lambda, \varepsilon, g) := \ldots$$

Maybe there is also a connection with fluctuations?

For $\varepsilon > 0$ define the random variable $J_\varepsilon(X_n)$ to be the maximum number of ε-fluctuations experienced by the sequence $\{X_n\}$.

Theorem ([Neri and P., 2024])

For any $\phi : (0,1) \to \mathbb{R}$, $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$ there exists $N \in \mathbb{N}$ such that for any sequence of random variables $\{X_n\}$ with

$$\mathbb{E}\left[J_\varepsilon(X_n)\right] < \phi(\varepsilon)$$

there exists $n \leq N$ such that

$$|X_i - X_j| < \varepsilon \quad \textit{for all} \quad i, j \in [n; n + g(n)]$$

with probability $> 1 - \lambda$. Moreover, we can define

$$N_K(\lambda, \varepsilon, g) := \tilde{g}^{(\lceil \phi(\varepsilon)/\lambda \rceil)}(0)$$

for $\tilde{g}(x) := x + g(x)$.

Proof of the theorem

Suppose for contradiction that for all $n \in \mathbb{N}$:

$$\mathbb{P}\left(\exists i, j \in [n; n + g(n)](|X_i - X_j| \geq \varepsilon)\right) \geq \lambda \qquad (*)$$

so in particular, for all $e \in \mathbb{N}$:

$$\mathbb{P}\left(A_e\right) \geq \lambda \quad \text{for} \quad A_e := \exists i, j \in [\tilde{g}^{(e)}(0); \tilde{g}^{(e+1)}(0)](|X_i - X_j| \geq \varepsilon)$$

For any $k \in \mathbb{N}$ we have

$$(k+1)\lambda \leq \sum_{e=0}^{k} \mathbb{P}(A_e) = \sum_{e=0}^{k} \mathbb{E}\left(I_{A_e}\right) = \mathbb{E}\left[\sum_{e=0}^{k} I_{A_e}\right] \leq \mathbb{E}\left[J_\varepsilon(X_n)\right] < \phi(\varepsilon)$$

which is a contradiction for

$$k := \left\lceil \frac{\phi(\varepsilon)}{\lambda} \right\rceil$$

Therefore $\mathbb{P}(A_e) < \lambda$ for some $e \leq k$ and therefore $(*)$ fails for some

$$n \leq \tilde{g}^{(k)}(0)$$

Now it should be easy?

We need a function $\phi_K(\varepsilon)$ such that for any sub- or supermartingale $\{X_n\}$ with

$$\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|] < K$$

we have

$$\mathbb{E}[J_\varepsilon(X_n)] < \phi(\varepsilon)$$

Theorem (Chashka, see Theorem 34 of [Kachurovskii, 1996])

For any $K > 0$ there exists a martingale $\{X_n\}$ with

$$\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|] < K$$

such that

$$\mathbb{E}\left[\sqrt{J_\varepsilon(X_n)}\right] = \infty$$

It turns out you only really get nice fluctuation behaviour for L_2-martingales.

For martingales, *crossings* are far easier to characterise

For $a < b$ define the random variable $U_{N,[a,b]}(X_n)$ to be the maximum number of times $\{X_n\}$ upcrosses the interval $[a, b]$ up to time N.

Theorem (Doob's upcrossing inequality for supermartingales)

$$\mathbb{E}\left[U_{\infty,[a,b]}(X_n)\right] \leq \frac{|a| + \sup_{n\in\mathbb{N}} \mathbb{E}(|X_n|)}{b-a}$$

The intuitive idea: Imagine that $\{X_n\}$ represents a stock, and consider an investment strategy that buys the stock whenever its price falls below a, and sells it whenever its price rises above b. Let Y_N denote your winnings after time N.

- Y_N is at least as good as the number of upcrossings times $(b-a)$

$$Y_N \geq (b-a)U_{N,[a,b]}(X_n) - (X_N - a)^- \qquad (*)$$

- Because $\{X_n\}$ is a supermartingale (i.e. the stock value decreases on average), this strategy can't win on average: $\mathbb{E}[Y_N] \leq 0$.

The inequality then follows by taking expectations on both sides of $(*)$.

Metastability for L_1-bounded crossings $C_{[a,b]}(X_n)$ (= down + upcrossings)

Theorem ([Neri and P., 2024])

For any $\lambda, \varepsilon, L, M > 0$ and $g : \mathbb{N} \to \mathbb{N}$ there exists $N \in \mathbb{N}$ such that for any sequence of random variables $\{X_n\}$ such that

$$\mathbb{P}\left(|X_n| \geq M\right) < \frac{\lambda}{2} \quad \text{and} \quad \mathbb{E}\left[C_{[a,b]}(X_n)\right] < L \ \text{ for } [a,b] \in \mathcal{P}(r,l)$$

where $\mathcal{P}(r,l)$ denotes the partition of $[-r,r]$ into l equal subintervals and

$$r := M\left(1+\frac{2}{p}\right) \quad \text{and} \quad l := p+2 \quad \text{and} \quad p := \left\lceil \frac{8M}{\varepsilon} \right\rceil$$

there exists $n \leq N$ such that

$$|X_i - X_j| < \varepsilon \quad \text{for all} \quad i,j \in [n; n+g(n)]$$

with probability $> 1-\lambda$. Moreover, we can define

$$N_{L,M}(\lambda,\varepsilon,g) := \tilde{g}^{(e)}(0)$$

for $\tilde{g}(x) := x + g(x)$ and

$$e := \frac{2(p+2)L}{\lambda}$$

A metastable martingale convergence theorem

Theorem ([Neri and P., 2024])

Take $K, \varepsilon, \lambda > 0$ and $g : \mathbb{N} \to \mathbb{N}$. Then there exists some $N \in \mathbb{N}$ (depending only on K, ε, λ and g) such that for any sub- or supermartingale $\{X_n\}$ with

$$\sup_{n \in \mathbb{N}} \mathbb{E}[|X_n|] < K$$

there exists $n \leq N$ such that

$$|X_i - X_j| < \varepsilon \quad \text{for all} \quad i, j \in [n; n + g(n)]$$

with probability $> 1 - \lambda$. Moreover, we can define

$$N_K(\lambda, \varepsilon, g) := \tilde{g}^{(e)}(0) \;\; \text{for} \;\; e := c\left(\frac{K}{\lambda\varepsilon}\right)^2$$

where $c > 0$ is a suitable constant that can be defined explicitly.

We can use our general framework to do a lot more

Some of our results on martingales:

stochastic process $\{X_n\}$	iterations of $\tilde{g}$
constant, monotone	K/ε
almost sure monotone	$cK/\lambda\varepsilon$
L_2-martingales	$cK^2/\lambda\varepsilon^2$
L_1-martingales	$cK^2/\lambda^2\varepsilon^2$
L_1-almost-martingales	$cK^2/\lambda^{2(1+r)}\varepsilon^2$ some $r \geq 0$

Notes:

- Most of these rates are optimal in a certain sense, but achieving optimal rates and showing that they are optimal was not easy.
- Similar rates can be obtained in other situations where crossing bounds are present e.g. ergodic theory.

For example, [Hochman, 2009] has many beautiful results on upcrossings…

The Annals of Probability
2009, Vol. 37, No. 6, 2135–2149
DOI: 10.1214/09-AOP460

UPCROSSING INEQUALITIES FOR STATIONARY SEQUENCES AND APPLICATIONS

BY MICHAEL HOCHMAN

Hebrew University of Jerusalem

For arrays $(S_{i,j})_{1\le i\le j}$ of random variables that are stationary in an appropriate sense, we show that the fluctuations of the process $(S_{1,n})_{n=1}^{\infty}$ can be bounded in terms of a measure of the "mean subadditivity" of the process $(S_{i,j})_{1\le i\le j}$. We derive universal upcrossing inequalities with exponential decay for Kingman's subadditive ergodic theorem, the Shannon–MacMillan–Breiman theorem and for the convergence of the Kolmogorov complexity of a stationary sample.

1. Introduction. Let us say that a sequence $(X_n)_{n=1}^{\infty}$ of real numbers has k crossings (or upcrossings) of an interval $[s,t]$ if there are indices

$$1 \le i_1 < j_1 < i_2 < j_2 < \cdots < i_k < j_k$$

such that $X_{i_m} < s$ and $X_{j_m} > t$. Allowing X_n to be random, it easily follows that $\lim X_n$ exists a.s. if and only if, for every interval of positive length, the probability of infinitely many crossings of the interval is 0.

There are a number of classical limit theorems in probability that can be formulated and proven in this way, the best known of which is Doob's upcrossing inequality for L^1 martingales [6]: if $(S_n)_{n=1}^{\infty}$ is an L^1 martingale, then, for $s<t$,

$$\mathbb{P}\big((S_n)_{n=1}^{\infty} \text{ has } k \text{ upcrossings of } [s,t]\big) \le \frac{\sup_n \|S_n\|_1}{k(t-s)}$$

(see also Dubins [7]). A similar inequality was proven by Bishop for the time averages $S_n = \frac{1}{n}\sum_{i=1}^{n} X_n$ of an L^1 stationary process $(X_n)_{n=1}^{\infty}$ [1, 2]. Assuming nonnegativity of the process instead of integrability, Ivanov [8] proved the following beautiful result: for every $s<t$,

Almost martingales in stochastic optimization

Almost monotone sequences in (nonstochastic) optimization: An example

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 $t > 0$.

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

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 $x_n \to x^*$.

Proof.

1. Show that $\|x_{n+1} - x^*\| \leq \|x_n - x^*\| - \alpha_n \psi(\|x_n - x^*\|)$
2. Prove that whenever $\boxed{u_{n+1} \leq u_n - \alpha_n \psi(u_n)}$ with $\sum_{i=0}^{\infty} \alpha_i = \infty$ then $u_n \to 0$.

(Rates of convergence for a generalised version of this result obtained in [Powell and Wiesnet, 2021].)

More examples from applied proof theory

Almost-monotone sequences can be found everywhere numerical analysis and optimization (where they are connected to the concept of Fejér monotonicity). Analysing these is a crucial step in obtaining explicit rates of convergence or metastability.

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

$$\boxed{a_{n+1} \leq (1 + b_n)a_n + c_n \text{ with } \sum b_i < \infty \text{ and } \sum c_i < \infty}$$

  Rates of metastability for $\{a_n\}$ calculated and used to produce quantitative results on asymptotically nonexpansive mappings.

- Numerous applied proof theory papers from the last years use variants of the following:

$$\boxed{s_{n+1} \leq (1 - a_n)s_n + a_n r_n + v_n \text{ with } \sum \alpha_i = \infty, \limsup_{n\to\infty} r_n \leq 0, \sum v_n < \infty}$$

  See [Pinto, 2023] for a detailed overview of the many variants.

There are explicit recursive inequalities in 30–40 papers in applied proof theory. There are thousands of such papers in ordinary optimization.

There is a useful survey paper covering mainstream mathematics [Franci and Grammatico, 2022]

Annual Reviews in Control 53 (2022) 161–186

Contents lists available at ScienceDirect

Annual Reviews in Control

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

ELSEVIER

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

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

Contains a huge survey of lemmas involving almost-monotone sequences

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 sequences can be seen as a specific class of Lyapunov function and Féjer monotonicity shows that it is decreasing along the iterates. The

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 applications in model predictive control (Lee & Nedić, 2015) and opinion dynamics (Shi et al., 2013) among others.

We report in Tables 1 and 2 the results for deterministic and

Contains a huge survey of lemmas involving almost-monotone sequences

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.

Lemma 3.4 (*Lemma 3.1, Combettes, 2001b*). *Let $\gamma \in (0,1]$ and let $(v^k)_{k\in\mathbb{N}}$,*

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*

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

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

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

and, for all $k \in \mathbb{N}$

… along with general heuristics for using them:

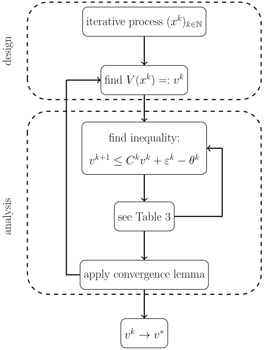

Fig. 7. Schematic representation of how the convergence lemmas for sequences can be used. Given the iterative process, a suitable nonnegative function (Lyapunov or distance-like) should be designed. Then, exploiting the properties of the application at hand, an inequality involving the iterates at times $k+1$ 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

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

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. *$0 < a \leq \nu_k \leq b < 1$ and $0 < c \leq \beta_k \leq d < 1$,*
4. *$0 < c \leq \alpha_k < 2\alpha$ 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 $x^* = \mathrm{proj}$

Almost-supermartingales

Half of [Franci and Grammatico, 2022] deals with results on stochastic sequences, none of which have been considered by applied proof theorists. The following result is particularly important:

Lemma 4.1 (*Robbins–Siegmund Lemma*). *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 $\sum_{k=0}^{\infty}\varepsilon^k < \infty$, $\sum_{k=0}^{\infty}\delta^k < \infty$ and*

$$\mathbb{E}[v^{k+1}|\mathcal{F}_k] \leq (1+\delta^k)v^k + \varepsilon^k - \theta^k \text{ a.s., for all } k \in \mathbb{N} \tag{4.1}$$

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

Proof. The proof follows by rewriting the sequence as in Lemma 3.6. Then, it is possible to show that the sequence

$$y^n = \tilde{v}^k - \sum_{k=0}^{n-1}(\tilde{\varepsilon}^k - \tilde{\theta}^k)$$

is a supermartingale. The claim then follows by the Martingale Convergence Theorem (Theorem 2.4). See Robbins and Siegmund (1971) for technical details. □

It can be found in any text on stochastic optimization, and is used to establish the convergence of algorithms in game theory, convex optimization, machine learning, . . .

A quantitative Robbins-Siegmund theorem

Theorem (Neri and P., coming soon...)

Let $\{X_n\}$, $\{A_n\}$, $\{B_n\}$ and $\{C_n\}$ be nonnegative stochastic processes adapted to some filtration $\mathcal{F}_n$ such that

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \leq (1 + A_n)X_n - B_n + C_n$$

almost surely for all $n \in \mathbb{N}$. Suppose that $K > \mathbb{E}[X_0]$ and that $\rho, \tau : (0, 1) \to [1, \infty)$ are monotone and satisfy

$$\mathbb{P}\left(\prod_{i=0}^{\infty}(1 + A_n) \geq \rho(\lambda)\right) < \lambda \text{ and } \mathbb{P}\left(\sum_{i=0}^{\infty} C_n \geq \sigma(\lambda)\right) < \lambda$$

for all $\lambda \in (0, 1)$. Then for any $\varepsilon, \lambda > 0$ and $g : \mathbb{N} \to \mathbb{N}$ there exists some $n \leq N_{K,\rho,\sigma}(\lambda, \varepsilon)$ such that

$$|X_i - X_j| < \varepsilon \quad \text{for all} \quad i, j \in [n; n + g(n)]$$

with probability $> 1 - \lambda$, where

$$N_{K,\rho,\sigma}(\lambda, \varepsilon) := \tilde{g}^{(e)}(0) \text{ for } e := c\left(\frac{\rho\left(\frac{\lambda}{8}\right) \cdot \left(K + \sigma\left(\frac{\lambda}{16}\right)\right)}{\lambda\varepsilon}\right)^2$$

where $c > 0$ is a suitable constant that can be defined explicitly.

The future: Proof mining in probability theory

Progress so far

Covered in this talk:

- A broad understanding of martingales (and related things) from a computational perspective.
- A quantitative Robbins-Siegmund theorem, plus a toolkit for obtaining metastable rates for general almost-supermartingales.

Recent work by collaborators:

- A beautiful "proof-theoretically tame" logical system for probability, and a metatheorem that guarantees the extractability of numerical information that is *independent* of the underlying probability space [Neri and Pischke, 2024].
- New convergence rates for strong laws of large numbers [Neri, 2024].

[Neri and Pischke, 2024]

PROOF MINING AND PROBABILITY THEORY

MORENIKEJI NERI[a] AND NICHOLAS PISCHKE[b]

[a] Department of Computer Science, University of Bath,
Claverton Down, Bath, BA2 7AY, United Kingdom,
[b] Department of Mathematics, Technische Universität Darmstadt,
Schlossgartenstraße 7, 64289 Darmstadt, Germany,

E-mails: mn728@bath.ac.uk, pischke@mathematik.tu-darmstadt.de

ABSTRACT. We extend the theoretical framework of proof mining by establishing general logical metatheorems that allow for the extraction of the computational content of theorems with prima facie "non-computational" proofs from probability theory, thereby unlocking a major branch of mathematics as a new area of application for these methods. Concretely, we devise proof-theoretically tame logical systems that, for one, allow for the formalization of proofs involving algebras of sets together with probability contents as well as associated Lebesgue integrals on them and which, for another, are amenable to proof-theoretic metatheorems in the style of proof mining that guarantee the extractability of effective and tame bounds from larges classes of ineffective existence proofs in probability theory. Moreover, these extractable bounds are guaranteed to be highly uniform in the sense that they will be independent of all parameters relating to the underlying probability space, particularly regarding events or measures of them. As such, these results, in particular, provide the first logical explanation for the success and the observed uniformities of the previous ad hoc case studies of proof mining in these areas and further illustrate their extent. Beyond these systems, we provide extensions for the proof-theoretically tame treatment of σ-algebras and associated probability measures using an intensional approach to infinite unions. Lastly, we establish a general proof-theoretic transfer principle that allows for the lift of quantitative information on a relationship between different modes of convergence for sequences of real numbers to sequences of random variables.

Keywords: Proof mining; Metatheorems; Probability theory; Egorov's theorem; Dominated convergence theorem.
MSC2020 Classification: 03F10, 03F35, 28A12, 28A25, 60A10

1. INTRODUCTION

One of the fundamental driving questions of proof theory is the following: What is the computational content of a mathematical theorem and how can it be exhibited? Proof mining, which emerged as a subfield of mathematical logic in the 1990s through the work of Ulrich Kohlenbach and his collaborators[1] aims at answering that question by extracting this computational content from theorems with proofs as they are found in the mainstream mathematical literature. This

[Neri, 2024]

QUANTITATIVE STRONG LAWS OF LARGE NUMBERS

MORENIKEJI NERI

Department of Computer Science, University of Bath
E-mail: mn728@bath.ac.uk

Abstract. Using proof-theoretic methods in the style of proof mining, we give novel computationally effective limit theorems for the convergence of the Cesaro-means of certain sequences of random variables. These results are intimately related to various Strong Laws of Large Numbers and, in that way, allow for the extraction of quantitative versions of many of these results. In particular, we produce optimal polynomial bounds in the case of pairwise independent random variables with uniformly bounded variance, improving on known results; furthermore, we obtain a new Baum-Katz type result for this class of random variables. Lastly, we are able to provide a fully quantitative version of a recent result of Chen and Sung that encompasses many limit theorems in the Strong Laws of Large Numbers literature.

Keywords: Laws of Large Numbers; Large deviations; Limit theorems; Proof mining.
MSC2020 Classification: 60F10, 60F15, 03F99

1. Introduction

Throughout this article, fix a probability space $(\Omega, \mathbf{F}, \mathbb{P})$ which all the random variables we work with will act on.

The classical Strong Law of Large Numbers is the following fundamental result due to Kolmogorov:

Theorem 1.1 (The classical Strong Law of Large Numbers, c.f. Theorem 6.6.1 of [1]). *Suppose $X_1, X_2, \ldots$ are independent, identically distributed (iid) real-valued random variables with $\mathbb{E}(|X_1|) < \infty$. Then,*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}(X_1) \tag{1}$$

almost surely, that is, with probability 1.

For ease write $S_n := \sum_{i=1}^{n} X_i$ and assume $\mathbb{E}(X_1) = 0$. By multiple applications of the continuity of the probability measure, one can show (following the nation in [2]) that the conclusion of the Strong Law of Large Numbers is equivalent to the sequence of real numbers

[Neri and P., 2024]

ON QUANTITATIVE CONVERGENCE FOR STOCHASTIC PROCESSES: CROSSINGS, FLUCTUATIONS AND MARTINGALES

MORENIKEJI NERI AND THOMAS POWELL

ABSTRACT. We develop a general framework for extracting highly uniform bounds on local stability for stochastic processes in terms of information on fluctuations or crossings. This includes a large class of martingales: As a corollary of our main abstract result, we obtain a quantitative version of Doob's convergence theorem for L_1-sub- and supermartingales, but more importantly, demonstrate that our framework readily extends to more complex stochastic processes such as almost-supermartingales, thus paving the way for future applications in stochastic optimization. Fundamental to our approach is the use of ideas from logic, particularly a careful analysis of the quantifier structure of probabilistic statements and the introduction of a number of abstract notions that represent stochastic convergence in a quantitative manner. In this sense, our work falls under the 'proof mining' program, and indeed, our quantitative results provide new examples of the phenomenon, recently made precise by the first author and Pischke, that many proofs in probability theory are proof-theoretically tame, and amenable to the extraction of quantitative data that is both of low complexity and independent of the underlying probability space.

1. Introduction

Applied proof theory (or *proof mining*) [30] is a subfield of logic in which proofs of mathematical theorems are carefully analysed with the aim of strengthening those theorems, typically through obtaining quantitative information or showing that they hold in a generalised, more abstract setting. Proof mining has achieved considerable success over the last decades, primarily in areas related to nonlinear analysis, such as fixed point theory and convex optimization. This paper is part of a new approach to expand proof mining into probability theory. We aim to set the groundwork for new applications involving stochastic processes, particularly stochastic optimization, by providing the first quantitative study of martingales from this perspective. More specifically, we provide quantitative versions of some classic martingale convergence theorems based on locating regions of *local stability* as in [4], which represent instances of a flexible, general method for obtaining bounds on local stability from quantitative information on fluctuations or upcrossings.

1.1. **Background: Proof mining and probability.** Originating in Kreisel's "unwinding" program of the 1950s [36, 37], and brought to maturity from the

Future work

1. Applications of the quantitative martingale and Robbins-Siegmund theorems in stochastic optimization and machine learning.
2. Abstract convergence proofs for generalised classes of algorithms in these areas.
3. Expanding the system of [Neri and Pischke, 2024] to include an abstract, logical treatment of random variables and notions of integrability.
4. Using abstracts convergence results on almost-supermartingales as the basis for a major effort to build a library of computer formalised proofs for stochastic optimization[1]
5. The development of algorithms for automating the reduction to a supermartingale i.e. automatically generating convergence proofs.

and much more ...

THANK YOU!

[1] see [Koutsoukou-Argyraki, 2021] for some speculative ideas on the formalisation of applied proof theory.

References I

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.

Avigad, J. and Rute, J. (2014).
Oscillation and the mean ergodic theorem for uniformly convex Banach spaces.
Ergodic theory and dynamical systems, 35(4):1009–1027.

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

Hochman, M. (2009).
Upcrossing inequalities for stationary sequences and applications.
Annals of Probability, 37(6):2135–2149.

Kachurovskii, A. G. (1996).
The rate of convergence in ergodic theorems.
Russian Mathematical Surveys, 51(4):653–703.

References II

Kohlenbach, U. (2005).
Some logical metatheorems with applications in functional analysis.
Transactions of the American Mathematical Society, 357(1):89–128.

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

Kohlenbach, U. and Safarik, P. (2014).
Fluctuations, effective learnability and metastability in analysis.
Annals and Pure and Applied Logic, 165:266–304.

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.

Neri, M. (2024).
Quantitative strong laws of large numbers.
Preprint, available at `https://arxiv.org/abs/2406.19166v1`.

References IV

Neri, M. and P., T. (2024).
On quantitative convergence for stochastic processes: Crossings, fluctuations and martingales.
Preprint, available at https://arxiv.org/abs/2406.19979.

Neri, M. and Pischke, N. (2024).
Proof mining and probability theory.
Preprint, available at https://arxiv.org/abs/2403.00659.

Neri, M. and Powell, T. (2023).
A computational study of a class of recursive inequalities.
Journal of Logic and Analysis, 15(3):1–48.

Pinto, P. (2023).
On quantitative versions of xu's lemma.
Note, available from author's webpage.

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.

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'.

Tao, T. (2008).
Norm convergence of multiple ergodic averages for commuting transformations.
Ergodic Theory and Dynamical Systems, 28:657–688.