Proof mining: a journey from deterministic convergence to probabilistic

M. Neri

Department of Computer Science
University of Bath, Bath, United Kingdom
kejineri.github.io
Logic Colloquium 2023

Applied proof theory in one slide

- Has its origins in the early 1950s in the form of Kreisel's proof unwinding program, with papers describing how proof-theoretic tools can be used to extract bounds in areas such as number theory [Kre51; Kre52].
- Explicit notable case studies came from Kreisel and Luckhardt in the 60s and late 80s respectively (Littlewood's theorem and Roth's theorem [Kre68; Luc89]).
- Due to work from Kohlenbach and his collaborators in the late 90s, this program evolved into what is known today as proof mining with case studies arising all over maths. One can read the textbook [Koh08] for detailed discussions and the recent survey papers [Koh17; Koh19].
- Furthermore, logical metatheorems have been developed that explain a lot of these case studies as being part of general phenomena.

Applied proof theory in one slide

- Has its origins in the early 1950s in the form of Kreisel's proof unwinding program, with papers describing how proof-theoretic tools can be used to extract bounds in areas such as number theory [Kre51; Kre52].
- Explicit notable case studies came from Kreisel and Luckhardt in the 60s and late 80s respectively (Littlewood's theorem and Roth's theorem [Kre68; Luc89]).
- Due to work from Kohlenbach and his collaborators in the late 90s, this program evolved into what is known today as proof mining with case studies arising all over maths. One can read the textbook [Koh08] for detailed discussions and the recent survey papers [Koh17; Koh19].
- Furthermore, logical metatheorems have been developed that explain a lot of these case studies as being part of general phenomena.

Applied proof theory in one slide

- Has its origins in the early 1950s in the form of Kreisel's proof unwinding program, with papers describing how proof-theoretic tools can be used to extract bounds in areas such as number theory [Kre51; Kre52].
- Explicit notable case studies came from Kreisel and Luckhardt in the 60s and late 80s respectively (Littlewood's theorem and Roth's theorem [Kre68; Luc89]).
- Due to work from Kohlenbach and his collaborators in the late 90s, this program evolved into what is known today as proof mining with case studies arising all over maths. One can read the textbook [Koh08] for detailed discussions and the recent survey papers [Koh17; Koh19].
- Furthermore, logical metatheorems have been developed that explain a lot of these case studies as being part of general phenomena.

Applied proof theory in one slide

- Has its origins in the early 1950s in the form of Kreisel's proof unwinding program, with papers describing how proof-theoretic tools can be used to extract bounds in areas such as number theory [Kre51; Kre52].
- Explicit notable case studies came from Kreisel and Luckhardt in the 60s and late 80s respectively (Littlewood's theorem and Roth's theorem [Kre68; Luc89]).
- Due to work from Kohlenbach and his collaborators in the late 90s, this program evolved into what is known today as proof mining with case studies arising all over maths. One can read the textbook [Koh08] for detailed discussions and the recent survey papers [Koh17; Koh19].
- Furthermore, logical metatheorems have been developed that explain a lot of these case studies as being part of general phenomena.

Overview of talk

- Logical background and preliminaries.
- A demonstration of some recent proof mining about deterministic convergence: I shall present a simple case study which is taken from joint project with Thomas Powell on a quantitative classification of recursive inequalities.
- Present current ongoing work in analysing nonconstructive proofs in probability theory.
- Bringing everything together.

Overview of talk

- Logical background and preliminaries.
- A demonstration of some recent proof mining about deterministic convergence: I shall present a simple case study which is taken from joint project with Thomas Powell on a quantitative classification of recursive inequalities.
- Present current ongoing work in analysing nonconstructive proofs in probability theory.
- Bringing everything together.

Overview of talk

- Logical background and preliminaries.
- A demonstration of some recent proof mining about deterministic convergence: I shall present a simple case study which is taken from joint project with Thomas Powell on a quantitative classification of recursive inequalities.
- Present current ongoing work in analysing nonconstructive proofs in probability theory.
- Bringing everything together.

Overview of talk

- Logical background and preliminaries.
- A demonstration of some recent proof mining about deterministic convergence: I shall present a simple case study which is taken from joint project with Thomas Powell on a quantitative classification of recursive inequalities.
- Present current ongoing work in analysing nonconstructive proofs in probability theory.
- Bringing everything together.

Logical background and preliminaries

Logical background and preliminaries

Let us recall the standard definition of Cauchy convergence

$$P := \forall \varepsilon \in \mathbb{Q}_+ \, \exists N \in \mathbb{N} \, \forall n \in \mathbb{N} \, (n \geq N \implies |a_i - a_j| \leq \varepsilon)$$

A natural question to ask is, from a proof of P can one find a computable function $f : \mathbb{Q} \to \mathbb{N}$ such that

$$\forall \varepsilon \in \mathbb{Q}_+ \, \forall n \in \mathbb{N} \, (n \geq f(\varepsilon) \implies |a_i - a_j| \leq \varepsilon)$$

Specker showed this was not always possible, through his famous construction of a bounded monotone sequence of rational numbers converging to a non-computable number [Spe49].

Logical background and preliminaries

Let us recall the standard definition of Cauchy convergence

$$P := \forall \varepsilon \in \mathbb{Q}_+ \, \exists N \in \mathbb{N} \, \forall n \in \mathbb{N} \, (n \geq N \implies |a_i - a_j| \leq \varepsilon)$$

A natural question to ask is, from a proof of P can one find a computable function $f : \mathbb{Q} \to \mathbb{N}$ such that

$$\forall \varepsilon \in \mathbb{Q}_+ \, \forall n \in \mathbb{N} \, (n \geq f(\varepsilon) \implies |a_i - a_j| \leq \varepsilon)$$

Specker showed this was not always possible, through his famous construction of a bounded monotone sequence of rational numbers converging to a non-computable number [Spe49].

Logical background and preliminaries

Let us recall the standard definition of Cauchy convergence

$$P := \forall \varepsilon \in \mathbb{Q}_+ \, \exists N \in \mathbb{N} \, \forall n \in \mathbb{N} \, (n \geq N \implies |a_i - a_j| \leq \varepsilon)$$

A natural question to ask is, from a proof of P can one find a computable function $f : \mathbb{Q} \to \mathbb{N}$ such that

$$\forall \varepsilon \in \mathbb{Q}_+ \, \forall n \in \mathbb{N} \, (n \geq f(\varepsilon) \implies |a_i - a_j| \leq \varepsilon)$$

Specker showed this was not always possible, through his famous construction of a bounded monotone sequence of rational numbers converging to a non-computable number [Spe49].

Logical background and preliminaries

One can show that Cauchy convergence is equivalent, over classical logic, to

$$\forall \varepsilon \in \mathbb{Q}_+ \, \forall g : \mathbb{N} \to \mathbb{N} \, \exists n \, \forall i, j \in [n, n + g(n)](|a_i - a_j| \leq \varepsilon)$$

This can be seen as an instance of Kreisel's no counterexample interpretation [Kre51; Kre52]and results in a new computational challenge, that is, to find a functional (known as a rate of metastability)

$$\Phi : \mathbb{Q}_+ \times (\mathbb{N} \to \mathbb{N}) \to \mathbb{N}$$

such that

$$\forall \varepsilon \in \mathbb{Q}_+ \, \forall g : \mathbb{N} \to \mathbb{N} \, \exists n \leq \Phi(\varepsilon, g) \forall i, j \in [n, n + g(n)](|a_i - a_j| \leq \varepsilon)$$

Logical background and preliminaries

One can show that Cauchy convergence is equivalent, over classical logic, to

$$\forall \varepsilon \in \mathbb{Q}_+ \, \forall g : \mathbb{N} \to \mathbb{N} \, \exists n \, \forall i, j \in [n, n + g(n)](|a_i - a_j| \le \varepsilon)$$

This can be seen as an instance of Kreisel's no counterexample interpretation [Kre51; Kre52]and results in a new computational challenge, that is, to find a functional (known as a rate of metastability)

$$\Phi : \mathbb{Q}_+ \times (\mathbb{N} \to \mathbb{N}) \to \mathbb{N}$$

such that

$$\forall \varepsilon \in \mathbb{Q}_+ \, \forall g : \mathbb{N} \to \mathbb{N} \, \exists n \le \Phi(\varepsilon, g) \forall i, j \in [n, n + g(n)](|a_i - a_j| \le \varepsilon)$$

Rates of metastability

If $\{a_n\}$ is a monotone sequence and $a \in \mathbb{Q}_+$ is such that, for all n, $|a_n| \leq a$ then, one can show that

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

where $\tilde{g}(n) := n + g(n)$, is such is a rate of metastability for the convergence of $\{a_n\}$.

Extracting rates of metastabilities using proof-theoretic methods are standard results in applied proof theory, for example, [Koh05; KK15; NP22].

Rates of metastability

If $\{a_n\}$ is a monotone sequence and $a \in \mathbb{Q}_+$ is such that, for all n, $|a_n| \leq a$ then, one can show that

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

where $\tilde{g}(n) := n + g(n)$, is such is a rate of metastability for the convergence of $\{a_n\}$.

Extracting rates of metastabilities using proof-theoretic methods are standard results in applied proof theory, for example, [Koh05; KK15; NP22].

A simple case of a big analysis
(jww Thomas Powell)

Recursive inequalities: A simple example

Convergence results about sequences of real numbers satisfying recursive inequalities play a big role in analysis.

Suppose (M, d) is a metric space and $T : M \to M$ is a contraction mapping with constant $c \in [0, 1)$, that is,

$$d(Tx, Ty) \leq cd(x, y) \; \forall x, y \in M$$

If x^* a fixed point of T, the distance $\mu_n := d(T^n x_0, x^*)$ satisfies

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

and thus converges to 0.

Furthermore, one can show that real numbers satisfying this recursive inequality converge to 0 with a rate given by

$$f(\varepsilon) = \lceil log_c(\frac{\varepsilon}{\mu_0}) \rceil$$

Recursive inequalities: A simple example

Convergence results about sequences of real numbers satisfying recursive inequalities play a big role in analysis.

Suppose (M, d) is a metric space and $T : M \to M$ is a contraction mapping with constant $c \in [0, 1)$, that is,

$$d(Tx, Ty) \leq cd(x, y)\ \forall x, y \in M$$

If x^* a fixed point of T, the distance $\mu_n := d(T^n x_0, x^*)$ satisfies

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

and thus converges to 0.

Furthermore, one can show that real numbers satisfying this recursive inequality converge to 0 with a rate given by

$$f(\varepsilon) = \lceil log_c(\frac{\varepsilon}{\mu_0})\rceil$$

Recursive inequalities: A simple example

Convergence results about sequences of real numbers satisfying recursive inequalities play a big role in analysis.

Suppose (M, d) is a metric space and $T : M \to M$ is a contraction mapping with constant $c \in [0, 1)$, that is,

$$d(Tx, Ty) \leq cd(x, y)\,\forall x, y \in M$$

If x^* a fixed point of T, the distance $\mu_n := d(T^n x_0, x^*)$ satisfies

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

and thus converges to 0.

Furthermore, one can show that real numbers satisfying this recursive inequality converge to 0 with a rate given by

$$f(\varepsilon) = \lceil log_c(\frac{\varepsilon}{\mu_0}) \rceil$$

Recursive inequalities: A simple example

Convergence results about sequences of real numbers satisfying recursive inequalities play a big role in analysis.

Suppose (M, d) is a metric space and $T : M \to M$ is a contraction mapping with constant $c \in [0, 1)$, that is,

$$d(Tx, Ty) \leq cd(x, y) \, \forall x, y \in M$$

If x^* a fixed point of T, the distance $\mu_n := d(T^n x_0, x^*)$ satisfies

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

and thus converges to 0.

Furthermore, one can show that real numbers satisfying this recursive inequality converge to 0 with a rate given by

$$f(\varepsilon) = \lceil log_c(\frac{\varepsilon}{\mu_0}) \rceil$$

Recursive inequalities: A simple example

Convergence results about sequences of real numbers satisfying recursive inequalities play a big role in analysis.

Suppose (M, d) is a metric space and $T : M \to M$ is a contraction mapping with constant $c \in [0, 1)$, that is,

$$d(Tx, Ty) \leq cd(x, y) \; \forall x, y \in M$$

If x^* a fixed point of T, the distance $\mu_n := d(T^n x_0, x^*)$ satisfies

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

and thus converges to 0.

Furthermore, one can show that real numbers satisfying this recursive inequality converge to 0 with a rate given by

$$f(\varepsilon) = \lceil log_c(\frac{\varepsilon}{\mu_0}) \rceil$$

Recursive inequalities: A more involved example

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ be a convex, smooth function.

Consider the usual steepest descent algorithm defined by

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

for some initial x_0, where the step sizes $\{\alpha_n\}$ satisfy

$$\sum_{i=0}^{\infty} \alpha_i = \infty \quad \text{and} \quad \sum_{i=0}^{\infty} \alpha_n^2 < \infty$$

It is a classic result that $f(x_n) \to f(x^*)$, with f attaining its minimum value at x^*.

Recursive inequalities: A more involved example

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ be a convex, smooth function.

Consider the usual steepest descent algorithm defined by

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

for some initial x_0, where the step sizes $\{\alpha_n\}$ satisfy

$$\sum_{i=0}^{\infty} \alpha_i = \infty \quad \text{and} \quad \sum_{i=0}^{\infty} \alpha_n^2 < \infty$$

It is a classic result that $f(x_n) \to f(x^*)$, with f attaining its minimum value at x^*.

Recursive inequalities: A more involved example

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ be a convex, smooth function.

Consider the usual steepest descent algorithm defined by

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

for some initial x_0, where the step sizes $\{\alpha_n\}$ satisfy

$$\sum_{i=0}^{\infty} \alpha_i = \infty \quad \text{and} \quad \sum_{i=0}^{\infty} \alpha_n^2 < \infty$$

It is a classic result that $f(x_n) \to f(x^*)$, with f attaining its minimum value at x^*.

Recursive inequalities: A more involved example

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ be a convex, smooth function.

Consider the usual steepest descent algorithm defined by

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

for some initial x_0, where the step sizes $\{\alpha_n\}$ satisfy

$$\sum_{i=0}^{\infty} \alpha_i = \infty \quad \text{and} \quad \sum_{i=0}^{\infty} \alpha_n^2 < \infty$$

It is a classic result that $f(x_n) \to f(x^*)$, with f attaining its minimum value at x^*.

Recursive inequalities: A more involved example

Establishing the convergence of the algorithm:

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

$$\sum_{i=0}^{\infty} \alpha_i = \infty \quad \text{and} \quad \sum_{i=0}^{\infty} \alpha_n^2 < \infty$$

is done by setting $\beta_n = f(x_n) - f(x^*)$ and showing (through tedious calculations)

$$\sum_{n=1}^{\infty} \alpha_n \beta_n < \infty$$

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

for some $\theta > 0$.

Recursive inequalities: A more involved example

Establishing the convergence of the algorithm:

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

$$\sum_{i=0}^{\infty} \alpha_i = \infty \quad \text{and} \quad \sum_{i=0}^{\infty} \alpha_n^2 < \infty$$

is done by setting $\beta_n = f(x_n) - f(x^*)$ and showing (through tedious calculations)

$$\sum_{n=1}^{\infty} \alpha_n \beta_n < \infty$$

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

for some $\theta > 0$.

A result about real numbers

Theorem (cf. Proposition 2 of [AIS98])

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 $\beta_n \to 0$.

Ineffectiveness of this result

Theorem (cf. Theorem 3.6 of [NP22])

For any sequence of non-negative rationals $\{\alpha_n\}$ with $\sum_{i=0}^{\infty} \alpha_i = \infty$ and rational $\theta > 0$, there exist sequences of positive rationals $\{\beta_n\}$, computable in $\{\alpha_n\}$ and θ and satisfying

$$\beta_n - \beta_{n+1} \leq \theta \alpha_n$$

and $\sum_{i=0}^{\infty} \alpha_i \beta_i < C$, for some rational C, such that $\beta_n \to 0$, but without a computable rate of convergence.

Mathematicians care about rates...

In [AIS98], this recursive inequality is used to prove the convergences of a general gradient descent algorithm for a general class of non-smooth functions on a Hilbert space.

As in the finite-dimensional smooth case, they are able to show, by setting $\beta_n = f(x_n) - f(x^*)$ again,

$$\sum_{n=1}^{\infty} \alpha_n \beta_n < \infty$$

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

for some $\theta > 0$.

Mathematicians care about rates...

In [AIS98], this recursive inequality is used to prove the convergences of a general gradient descent algorithm for a general class of non-smooth functions on a Hilbert space.

As in the finite-dimensional smooth case, they are able to show, by setting $\beta_n = f(x_n) - f(x^*)$ again,

$$\sum_{n=1}^{\infty} \alpha_n \beta_n < \infty$$

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

for some $\theta > 0$.

...but can't always find them

Furthermore, they calculate a rate of convergence for a subsequence of β_n but admit defeat in finding a general rate of convergence.

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

Thus, we are able to give a logical explanation why they were unable to do this from their proof method. In particular, in establishing $\sum_{n=1}^{\infty} \alpha_n \beta_n < \infty$ they only give a bound, and we have shown that this is not enough. However ...

…but can't always find them

Furthermore, they calculate a rate of convergence for a subsequence of β_n but admit defeat in finding a general rate of convergence.

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

Thus, we are able to give a logical explanation why they were unable to do this from their proof method. In particular, in establishing $\sum_{n=1}^{\infty} \alpha_n \beta_n < \infty$ they only give a bound, and we have shown that this is not enough. However …

...but can't always find them

Furthermore, they calculate a rate of convergence for a subsequence of β_n but admit defeat in finding a general rate of convergence.

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

Thus, we are able to give a logical explanation why they were unable to do this from their proof method. In particular, in establishing $\sum_{n=1}^{\infty} \alpha_n \beta_n < \infty$ they only give a bound, and we have shown that this is not enough. However ...

Quantative analysis

One can show that if one has a rate of convergence for $\sum_{n=1}^{\infty} \alpha_n \beta_n$ (which is a very strong condition), one can obtain a rate of convergence for $\beta_n \to 0$. On the other hand, if we only have a bound for this sum, we can obtain a rate of metastability.

Theorem (cf. Corollary 3.11 of [NP22])

Suppose r is a rate of divergence for $\sum_{i=0}^{\infty} \alpha_i = \infty$

(a) *If $\sum_{i=0}^{\infty} \alpha_i \beta_i < \infty$ with rate of convergence ϕ, then $\beta_n \to 0$ with rate of convergence*

$$\psi(\varepsilon) := \phi\left(\frac{\varepsilon^2}{4\theta}\right)$$

(b) *If $\sum_{i=0}^{\infty} \alpha_i \beta_i < \infty$ with rate of metastability Φ, then $\beta_n \to 0$ with rate of metastability*

$$\Psi(\varepsilon, g) := \Phi\left(\frac{\varepsilon^2}{4\theta}, h\right) \text{ for } h(n) := r\left(n + g(n), \frac{\varepsilon}{2\theta}\right) - n$$

Quantative analysis

One can show that if one has a rate of convergence for $\sum_{n=1}^{\infty} \alpha_n \beta_n$ (which is a very strong condition), one can obtain a rate of convergence for $\beta_n \to 0$. On the other hand, if we only have a bound for this sum, we can obtain a rate of metastability.

Theorem (cf. Corollary 3.11 of [NP22])

Suppose r is a rate of divergence for $\sum_{i=0}^{\infty} \alpha_i = \infty$

(a) *If $\sum_{i=0}^{\infty} \alpha_i \beta_i < \infty$ with rate of convergence ϕ, then $\beta_n \to 0$ with rate of convergence*

$$\psi(\varepsilon) := \phi\left(\frac{\varepsilon^2}{4\theta}\right)$$

(b) *If $\sum_{i=0}^{\infty} \alpha_i \beta_i < \infty$ with rate of metastability Φ, then $\beta_n \to 0$ with rate of metastability*

$$\Psi(\varepsilon, g) := \Phi\left(\frac{\varepsilon^2}{4\theta}, h\right) \text{ for } h(n) := r\left(n + g(n), \frac{\varepsilon}{2\theta}\right) - n$$

A brief summary of [NP22]

We underwent a deep study of the 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 either:

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

Note that from the above recursive inequality, all we can deduce is that $\{\mu_n\}$ converges and a subsequence of $\{\beta_n\}$ converges to 0.

A brief summary of [NP22]

We underwent a deep study of the 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 either:

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

Note that from the above recursive inequality, all we can deduce is that $\{\mu_n\}$ converges and a subsequence of $\{\beta_n\}$ converges to 0.

A brief summary of [NP22]

Our study included:

- Providing a sufficient and necessary condition on $\{\beta_n\}$ and $\{\mu_n\}$ that allow us to deduce further convergence properties (such as convergence to 0), including a generalisation of proposition 2 of [AIS98].
- Providing rates of convergences and metastabilities as well as justifying our rates of metstabilities by constructing Specker-like sequences.
- Using our general results to provide a new optimisation algorithm generalising the work done in [AIS98] as well as quantitative results for this algorithm. Furthermore, we are able to demonstrate how our work arises in further applications such as in the study of asymptotically weakly contractive mappings and accretive operator theory.

A brief summary of [NP22]

Our study included:

- Providing a sufficient and necessary condition on $\{\beta_n\}$ and $\{\mu_n\}$ that allow us to deduce further convergence properties (such as convergence to 0), including a generalisation of proposition 2 of [AIS98].
- Providing rates of convergences and metastabilities as well as justifying our rates of metstabilities by constructing Specker-like sequences.
- Using our general results to provide a new optimisation algorithm generalising the work done in [AIS98] as well as quantitative results for this algorithm. Furthermore, we are able to demonstrate how our work arises in further applications such as in the study of asymptotically weakly contractive mappings and accretive operator theory.

A brief summary of [NP22]

Our study included:

- Providing a sufficient and necessary condition on $\{\beta_n\}$ and $\{\mu_n\}$ that allow us to deduce further convergence properties (such as convergence to 0), including a generalisation of proposition 2 of [AIS98].
- Providing rates of convergences and metastabilities as well as justifying our rates of metstabilities by constructing Specker-like sequences.
- Using our general results to provide a new optimisation algorithm generalising the work done in [AIS98] as well as quantitative results for this algorithm. Furthermore, we are able to demonstrate how our work arises in further applications such as in the study of asymptotically weakly contractive mappings and accretive operator theory.

Probabilistic convergence

Computable Probabilistic Convergence

If $\{Y_n\}$ is a sequence of random variables, then Y_n converges almost surely (a.s) simply means,

$$\mathbb{P}(\{\omega \in \Omega : Y_n(\omega) \text{ converges}\}) = 1$$

It is clear that there is no direct computational interpretation that can be given to this definition, however, one can show that it is equivalent to almost sure uniform (a.s.u) convergence

$$\forall \lambda_1, \lambda_2 \in \mathbb{Q}^+ \, \exists N \in \mathbb{N} \, \mathbb{P}(\{\omega \in \Omega : \forall m \geq N \, |Y_m(\omega) - Y_N(\omega)| \leq \lambda_1\}) > 1 - \lambda_2$$

This result is known as Egorov's theorem. The definition of a.s.u convergence can be given a direct computational interpretation. We call a solution to the computational interpretation a rate of almost sure convergence and such rates have been studied in [ADR12].

Computable Probabilistic Convergence

If $\{Y_n\}$ is a sequence of random variables, then Y_n converges almost surely (a.s) simply means,

$$\mathbb{P}(\{\omega \in \Omega : Y_n(\omega) \text{ converges}\}) = 1$$

It is clear that there is no direct computational interpretation that can be given to this definition, however, one can show that it is equivalent to almost sure uniform (a.s.u) convergence

$$\forall \lambda_1, \lambda_2 \in \mathbb{Q}^+ \, \exists N \in \mathbb{N} \mathbb{P}(\{\omega \in \Omega : \forall m \geq N \, |Y_m(\omega) - Y_N(\omega)| \leq \lambda_1\}) > 1 - \lambda_2$$

This result is known as Egorov's theorem. The definition of a.s.u convergence can be given a direct computational interpretation. We call a solution to the computational interpretation a rate of almost sure convergence and such rates have been studied in [ADR12].

Computable Probabilistic Convergence

Furthermore, we can also obtain a metastable notion of almost sure uniform convergence

$$\forall \lambda_1, \lambda_2 \in \mathbb{Q}^+ \, \forall K : \mathbb{N} \to \mathbb{N} \, \exists N$$
$$\mathbb{P}(\{\omega \in \Omega : \forall m \in [N, K(N)] |Y_m(\omega) - Y_N(\omega)| \leq \lambda_1\}) > 1 - \lambda_2$$

This notion of probabilistic convergence was first studied by Avigad et al. [ADR12; AGT10], with explicit rates being extracted for the pointwise ergodic theorem and a computational version of Egorov's theorem.

Large deviations in the strong law of large numbers

Suppose $X_1, X_2, \ldots$ are independent identically distributed (iid) real-valued random variables with $\mathbb{E}(X_n) = 0$ and $\mathbb{E}(|X_n|) < \infty$ for all n. Define $S_n = \sum_{i=1}^{n} X_i$. The strong law of large numbers states that,

$$\frac{S_n}{n} \to 0$$

almost surely.

What has been done?

- The quantitative content of this theorem has been studied extensively in the probability literature, see [Sie75; Fil83; Luz18] for examples.
- The computability theory and the constructive nature of a popular proof of this result given in [Ete81] has been studied in [Gác10].

What has been done?

- The quantitative content of this theorem has been studied extensively in the probability literature, see [Sie75; Fil83; Luz18] for examples.
- The computability theory and the constructive nature of a popular proof of this result given in [Ete81] has been studied in [Gác10].

Kolmogorov's strong law of large numbers

Dropping the identical distribution assumption means we must include additional assumptions to ensure that we can conclude $\frac{S_n}{n} \to 0$ almost surely.

Theorem (Kolmogorov's strong law of large numbers)

If $\{X_n\}$ are independent and

$$\sum_{n=1}^{\infty} \frac{Var(X_n)}{n^2} < \infty$$

then $\frac{S_n}{n} \to 0$ almost surely

Proof of Kolmogorov's strong law of large numbers

The 'standard' proof of this theorem is as follows:

$$\sum_{n=1}^{\infty} \frac{Var(X_n)}{n^2} < \infty \implies \sum_{n=1}^{\infty} \frac{X_n}{n} < \infty \text{ a.s}$$

By Kolmogorov's convergence criterion (easy to obtain a computational version), and

$$\sum_{n=1}^{\infty} \frac{X_n}{n} < \infty \text{ a.s} \implies \frac{1}{n}\sum_{k=1}^{n} X_k \to 0 \text{ a.s}$$

By Kronecker's lemma (hard to obtain a computational version).

Proof of Kolmogorov's strong law of large numbers

The 'standard' proof of this theorem is as follows:

$$\sum_{n=1}^{\infty} \frac{Var(X_n)}{n^2} < \infty \implies \sum_{n=1}^{\infty} \frac{X_n}{n} < \infty \text{ a.s}$$

By Kolmogorov's convergence criterion (easy to obtain a computational version), and

$$\sum_{n=1}^{\infty} \frac{X_n}{n} < \infty \text{ a.s} \implies \frac{1}{n}\sum_{k=1}^{n} X_k \to 0 \text{ a.s}$$

By Kronecker's lemma (hard to obtain a computational version).

Proof of Kolmogorov's strong law of large numbers

The 'standard' proof of this theorem is as follows:

$$\sum_{n=1}^{\infty} \frac{Var(X_n)}{n^2} < \infty \implies \sum_{n=1}^{\infty} \frac{X_n}{n} < \infty \text{ a.s}$$

By Kolmogorov's convergence criterion (easy to obtain a computational version), and

$$\sum_{n=1}^{\infty} \frac{X_n}{n} < \infty \text{ a.s} \implies \frac{1}{n} \sum_{k=1}^{n} X_k \to 0 \text{ a.s}$$

By Kronecker's lemma (hard to obtain a computational version).

Proof of Kolmogorov's strong law of large numbers

The 'standard' proof of this theorem is as follows:

$$\sum_{n=1}^{\infty} \frac{Var(X_n)}{n^2} < \infty \implies \sum_{n=1}^{\infty} \frac{X_n}{n} < \infty \text{ a.s}$$

By Kolmogorov's convergence criterion (easy to obtain a computational version), and

$$\sum_{n=1}^{\infty} \frac{X_n}{n} < \infty \text{ a.s} \implies \frac{1}{n}\sum_{k=1}^{n} X_k \to 0 \text{ a.s}$$

By Kronecker's lemma (hard to obtain a computational version).

Kronecker's lemma

Theorem

Let $\{x_n\}$ be a sequence of real numbers. If $\sum_{i=1}^{\infty} \frac{x_i}{i} < \infty$, then

$$\frac{1}{n}\sum_{i=1}^{n} x_i \to 0$$

Hence, to be more explicit in our proof of Kolmogorov's strong law of large numbers, we must add the line,

$$\mathbb{P}(\{\omega \in \Omega : \sum_{n=1}^{\infty} \frac{X_n(\omega)}{n} < \infty\}) \leq \mathbb{P}(\{\omega \in \Omega : \frac{1}{n}\sum_{k=1}^{n} X_k(\omega) \to 0\})$$

by Kronecker's lemma.

Kronecker's lemma

Theorem

Let $\{x_n\}$ be a sequence of real numbers. If $\sum_{i=1}^{\infty} \frac{x_i}{i} < \infty$, then

$$\frac{1}{n}\sum_{i=1}^{n} x_i \to 0$$

Hence, to be more explicit in our proof of Kolmogorov's strong law of large numbers, we must add the line,

$$\mathbb{P}\left(\left\{\omega \in \Omega : \sum_{n=1}^{\infty} \frac{X_n(\omega)}{n} < \infty\right\}\right) \leq \mathbb{P}\left(\left\{\omega \in \Omega : \frac{1}{n}\sum_{k=1}^{n} X_k(\omega) \to 0\right\}\right)$$

by Kronecker's lemma.

Computability and Kronecker's lemma

Let A be a recursively enumerable set that is not recursive (e.g the halting set) and $\{a_n\}$ be a recursive enumeration of the elements in A. If we take $x_i = i2^{-a_i}$, we have $\sum_{i=1}^{\infty} \frac{x_i}{i} < \infty$, but

$$\frac{1}{n}\sum_{i=1}^{n} x_i \to 0$$

without a computable rate of convergence.

Computability and the strong law of large numbers

We construct also construct a sequence of computable independent random variables $\{X_n\}$, with 0 expected value, such that

$$\sum_{n=1}^{\infty} \frac{Var(X_n)}{n^2} < \infty$$

but

$$\frac{S_n}{n} \to 0$$

almost surely, without a computable rate of almost sure convergence.

Why is computational Kronecker's lemma hard

We want to show

$$\sum_{n=1}^{\infty} \frac{X_n}{n} < \infty \text{ a.s.u} \implies \frac{1}{n}\sum_{k=1}^{n} X_k \to 0 \text{ a.s.u}$$

The obvious way to do this is

$$\sum_{n=1}^{\infty} \frac{X_n}{n} < \infty \text{ a.s.u} \implies \sum_{n=1}^{\infty} \frac{X_n}{n} < \infty \text{ a.s} \implies$$
$$\frac{1}{n}\sum_{k=1}^{n} X_k \to 0 \text{ a.s} \implies \frac{1}{n}\sum_{k=1}^{n} X_k \to 0 \text{ a.s.u}$$

But the quantitative Egorov's theorem is hard.

Why is computational Kronecker's lemma hard

We want to show

$$\sum_{n=1}^{\infty} \frac{X_n}{n} < \infty \text{ a.s.u} \implies \frac{1}{n}\sum_{k=1}^{n} X_k \to 0 \text{ a.s.u}$$

The obvious way to do this is

$$\sum_{n=1}^{\infty} \frac{X_n}{n} < \infty \text{ a.s.u} \implies \sum_{n=1}^{\infty} \frac{X_n}{n} < \infty \text{ a.s} \implies$$

$$\frac{1}{n}\sum_{k=1}^{n} X_k \to 0 \text{ a.s} \implies \frac{1}{n}\sum_{k=1}^{n} X_k \to 0 \text{ a.s.u}$$

But the quantitative Egorov's theorem is hard.

Why is computational Kronecker's lemma hard

We want to show

$$\sum_{n=1}^{\infty} \frac{X_n}{n} < \infty \text{ a.s.u} \implies \frac{1}{n}\sum_{k=1}^{n} X_k \to 0 \text{ a.s.u}$$

The obvious way to do this is

$$\sum_{n=1}^{\infty} \frac{X_n}{n} < \infty \text{ a.s.u} \implies \sum_{n=1}^{\infty} \frac{X_n}{n} < \infty \text{ a.s} \implies$$
$$\frac{1}{n}\sum_{k=1}^{n} X_k \to 0 \text{ a.s} \implies \frac{1}{n}\sum_{k=1}^{n} X_k \to 0 \text{ a.s.u}$$

But the quantitative Egorov's theorem is hard.

Why is computational Kronecker's lemma hard

If our rates for Kronecker's lemma were independent of the sequence then one can obtain a rate for

$$\frac{1}{n}\sum_{k=1}^{n} X_k \to 0 \text{ a.s.u}$$

given one for

$$\sum_{n=1}^{\infty} \frac{X_n}{n} < \infty \text{ a.s.u}$$

It turns out the rates one obtains from the 'standard' proof of

Kronecker's lemma are not independent of the sequences, but can be made sufficiently uniform enough for us to be able to avoid using Egorov's theorem.

Why is computational Kronecker's lemma hard

If our rates for Kronecker's lemma were independent of the sequence then one can obtain a rate for

$$\frac{1}{n}\sum_{k=1}^{n} X_k \to 0 \text{ a.s.u}$$

given one for

$$\sum_{n=1}^{\infty} \frac{X_n}{n} < \infty \text{ a.s.u}$$

It turns out the rates one obtains from the 'standard' proof of

Kronecker's lemma are not independent of the sequences, but can be made sufficiently uniform enough for us to be able to avoid using Egorov's theorem.

Metastable Kronecker's lemma

Suppose $\sum_{i=1}^{\infty} \frac{x_i}{i}$ converges with rate of metastability Φ. Then $\frac{1}{n} \sum_{i=1}^{n} x_i$ converges to 0 with rate of metastability

$$\kappa_{\Phi,\{z_n\}}(\varepsilon, g) = \max_{n \leq \Phi(\frac{\varepsilon}{4}, h_{\varepsilon,g})} \max\{n, \lceil \frac{4n|z_n|}{\varepsilon} \rceil\}$$

Where,

$$h_{\varepsilon,g}(n) = g(\max\{n, \lceil \frac{4n|z_n|}{\varepsilon} \rceil\})$$

and $\{z_n\}$ is any sequence such that $|\sum_{i=1}^{n} \frac{x_i}{i}| \leq |z_n|$ for all n.

How does this help?

Intuitively, Markov's inequality tells us that, for all $\varepsilon > 0$

$$\mathbb{P}(\{\omega \in \Omega : |\sum_{k=1}^{n} \frac{X_k(\omega)}{k}| < \frac{\mathbb{E}(|\sum_{k=1}^{n} \frac{X_k}{k}|)}{\varepsilon}\}) > 1 - \varepsilon$$

Thus taking

$$z_n = \frac{\mathbb{E}(|\sum_{k=1}^{n} \frac{X_k}{k}|)}{\varepsilon}$$

and using our metastable Kronecker's lemma, allows us to obtain a rate for

$$\frac{1}{n} \sum_{k=1}^{n} X_k(\omega) \to 0$$

for each ω in as big a proportion of the elements in the sample space as we want. This type of thinking allows us to get rates for the probabilistic analogue of Kronecker's lemma, which in turn gives us rates for Kolmogorov's strong law of large numbers.

How does this help?

Intuitively, Markov's inequality tells us that, for all $\varepsilon > 0$

$$\mathbb{P}(\{\omega \in \Omega : |\sum_{k=1}^{n} \frac{X_k(\omega)}{k}| < \frac{\mathbb{E}(|\sum_{k=1}^{n} \frac{X_k}{k}|)}{\varepsilon}\}) > 1 - \varepsilon$$

Thus taking

$$z_n = \frac{\mathbb{E}(|\sum_{k=1}^{n} \frac{X_k}{k}|)}{\varepsilon}$$

and using our metastable Kronecker's lemma, allows us to obtain a rate for

$$\frac{1}{n} \sum_{k=1}^{n} X_k(\omega) \to 0$$

for each ω in as big a proportion of the elements in the sample space as we want. This type of thinking allows us to get rates for the probabilistic analogue of Kronecker's lemma, which in turn gives us rates for Kolmogorov's strong law of large numbers.

Thinking in approximates is helpful

This type of thinking is very common when doing proof mining or when trying to obtain constructive proofs of results in maths, in the style of Bishop [BB12] for example. Going back to Kreisel [Kre52], where he explained how one could obtain bounds for Littlewood's theorem,

> *"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"*

Almost sure convergence to 0

Let us recall the definition of

$$\frac{S_n}{n} \to 0 \text{ a.s.u}$$

$$\forall \lambda_1, \lambda_2 \in \mathbb{Q}^+ \, \exists N \in \mathbb{N} \mathbb{P}(\{\omega \in \Omega : \forall n \geq N \, |\frac{S_n}{n}| \leq \lambda_1\}) > 1 - \lambda_2$$

This is clearly equivalent to

$$\forall \lambda_1, \lambda_2 \in \mathbb{Q}^+ \, \exists N \in \mathbb{N} \mathbb{P}(\{\omega \in \Omega : \sup_{n \geq N} |\frac{S_n}{n}| > \lambda_1\}) \leq \lambda_2$$

And this is equivalent to

$$\forall \varepsilon \in \mathbb{Q}^+ \, P_{\varepsilon,n} \to 0$$

where

$$P_{n,\varepsilon} = \mathbb{P}(\{\omega \in \Omega : \sup_{m \geq n} |\frac{S_n}{n}| > \varepsilon\})$$

Almost sure convergence to 0

Let us recall the definition of

$$\frac{S_n}{n} \to 0 \text{ a.s.u}$$

$$\forall \lambda_1, \lambda_2 \in \mathbb{Q}^+ \, \exists N \in \mathbb{N} \mathbb{P}(\{\omega \in \Omega : \forall n \geq N \, |\frac{S_n}{n}| \leq \lambda_1\}) > 1 - \lambda_2$$

This is clearly equivalent to

$$\forall \lambda_1, \lambda_2 \in \mathbb{Q}^+ \, \exists N \in \mathbb{N} \mathbb{P}(\{\omega \in \Omega : \sup_{n \geq N} |\frac{S_n}{n}| > \lambda_1\}) \leq \lambda_2$$

And this is equivalent to

$$\forall \varepsilon \in \mathbb{Q}^+ \, P_{\varepsilon,n} \to 0$$

where

$$P_{n,\varepsilon} = \mathbb{P}(\{\omega \in \Omega : \sup_{m \geq n} |\frac{S_n}{n}| > \varepsilon\})$$

Almost sure convergence to 0

Let us recall the definition of

$$\frac{S_n}{n} \to 0 \text{ a.s.u}$$

$$\forall \lambda_1, \lambda_2 \in \mathbb{Q}^+ \, \exists N \in \mathbb{N} \mathbb{P}(\{\omega \in \Omega : \forall n \geq N \, |\frac{S_n}{n}| \leq \lambda_1\}) > 1 - \lambda_2$$

This is clearly equivalent to

$$\forall \lambda_1, \lambda_2 \in \mathbb{Q}^+ \, \exists N \in \mathbb{N} \mathbb{P}(\{\omega \in \Omega : \sup_{n \geq N} |\frac{S_n}{n}| > \lambda_1\}) \leq \lambda_2$$

And this is equivalent to

$$\forall \varepsilon \in \mathbb{Q}^+ \, P_{\varepsilon,n} \to 0$$

where

$$P_{n,\varepsilon} = \mathbb{P}(\{\omega \in \Omega : \sup_{m \geq n} |\frac{S_n}{n}| > \varepsilon\})$$

Almost sure convergence to 0

Let us recall the definition of

$$\frac{S_n}{n} \to 0 \text{ a.s.u}$$

$$\forall \lambda_1, \lambda_2 \in \mathbb{Q}^+ \, \exists N \in \mathbb{N} \mathbb{P}(\{\omega \in \Omega : \forall n \geq N \, |\frac{S_n}{n}| \leq \lambda_1\}) > 1 - \lambda_2$$

This is clearly equivalent to

$$\forall \lambda_1, \lambda_2 \in \mathbb{Q}^+ \, \exists N \in \mathbb{N} \mathbb{P}(\{\omega \in \Omega : \sup_{n \geq N} |\frac{S_n}{n}| > \lambda_1\}) \leq \lambda_2$$

And this is equivalent to

$$\forall \varepsilon \in \mathbb{Q}^+ \, P_{\varepsilon,n} \to 0$$

where

$$P_{n,\varepsilon} = \mathbb{P}(\{\omega \in \Omega : \sup_{m \geq n} |\frac{S_n}{n}| > \varepsilon\})$$

Large deviations in the strong law of large numbers

The quantitative content of the strong law of large numbers has been studied in the probability literature in the form of describing the asymptotics of $P_{n,\varepsilon}$, under different assumptions on the iid random variables $\{X_n\}$.

For example:

- In [Sie75; Fil83], $P_{n,\varepsilon}$ is calculated up to asymptotic equivalence to simple functions. Furthermore, the bounds obtained heavily depend on the distribution of the random variables and very strong assumptions on the random variables are given.
- In [Luz18] an upper bound for is given $P_{n,\varepsilon}$ that does not depend on the distribution of the random variables, for by only assuming further that $Var(X_1) < \infty$ and that $\{X_n\}$ are pairwise independent.

Large deviations in the strong law of large numbers

The quantitative content of the strong law of large numbers has been studied in the probability literature in the form of describing the asymptotics of $P_{n,\varepsilon}$, under different assumptions on the iid random variables $\{X_n\}$.

For example:

- In [Sie75; Fil83], $P_{n,\varepsilon}$ is calculated up to asymptotic equivalence to simple functions. Furthermore, the bounds obtained heavily depend on the distribution of the random variables and very strong assumptions on the random variables are given.
- In [Luz18] an upper bound for is given $P_{n,\varepsilon}$ that does not depend on the distribution of the random variables, for by only assuming further that $Var(X_1) < \infty$ and that $\{X_n\}$ are pairwise independent.

Large deviations in the strong law of large numbers

The quantitative content of the strong law of large numbers has been studied in the probability literature in the form of describing the asymptotics of $P_{n,\varepsilon}$, under different assumptions on the iid random variables $\{X_n\}$.

For example:

- In [Sie75; Fil83], $P_{n,\varepsilon}$ is calculated up to asymptotic equivalence to simple functions. Furthermore, the bounds obtained heavily depend on the distribution of the random variables and very strong assumptions on the random variables are given.
- In [Luz18] an upper bound for is given $P_{n,\varepsilon}$ that does not depend on the distribution of the random variables, for by only assuming further that $Var(X_1) < \infty$ and that $\{X_n\}$ are pairwise independent.

An observation

There are various strong laws of large number results, where the random variables are not assumed to be identically distributed and the independence condition is reduced.

A large class of these results([Pet69; CG92; KP10; CS16] for example) follow a similar proof strategy which can be summarized as the following general theorem :

Theorem

let $\{X_n\}$ be a sequence of nonnegative random variables with respective means $\{\mu_n\}$. Assume $\forall n \ \frac{1}{n}\sum_{i=1}^{n}\mu_i \le W$, for some $W > 0$. Let $z_n = \sum_{i=1}^{n}\mu_i$. Suppose for each $\varepsilon, \delta > 0$, $\alpha > 1$ and $0 \le s \le L_\delta$

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\left|\frac{1}{k^{\pm}_{s,\delta,\alpha}(n)}S_{k^{\pm}_{s,\delta,\alpha}(n)} - \frac{1}{k^{\pm}_{s,\delta,\alpha}(n)}z_{k^{\pm}_{s,\delta,\alpha}(n)}\right| > \varepsilon\right) < \infty$$

Then, $\frac{1}{n}S_n - \frac{1}{n}z_n$ converges to 0 almost surely.

An observation

There are various strong laws of large number results, where the random variables are not assumed to be identically distributed and the independence condition is reduced.

A large class of these results([Pet69; CG92; KP10; CS16] for example) follow a similar proof strategy which can be summarized as the following general theorem :

Theorem

let $\{X_n\}$ be a sequence of nonnegative random variables with respective means $\{\mu_n\}$. Assume $\forall n \; \frac{1}{n}\sum_{i=1}^{n}\mu_i \leq W$, for some $W > 0$. Let $z_n = \sum_{i=1}^{n}\mu_i$. Suppose for each $\varepsilon, \delta > 0$, $\alpha > 1$ and $0 \leq s \leq L_\delta$

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\left|\frac{1}{k^{\pm}_{s,\delta,\alpha}(n)}S_{k^{\pm}_{s,\delta,\alpha}(n)} - \frac{1}{k^{\pm}_{s,\delta,\alpha}(n)}z_{k^{\pm}_{s,\delta,\alpha}(n)}\right| > \varepsilon\right) < \infty$$

Then, $\frac{1}{n}S_n - \frac{1}{n}z_n$ converges to 0 almost surely.

A quantitative observation

One can obtain the following quantitative version of this theorem.

Theorem

Suppose

$$\sum_{n=1}^{\infty} \mathbb{P}(|\frac{1}{k^{\pm}_{s,\delta,\alpha}(n)} S_{k^{\pm}_{s,\delta,\alpha}(n)} - \frac{1}{k^{\pm}_{s,\delta,\alpha}(n)} z_{k^{\pm}_{s,\delta,\alpha}(n)}| > \varepsilon) < \infty$$

and converges to its limit with rate of convergence $\Lambda_{\varepsilon,\delta,\alpha} : \mathbb{R} \to \mathbb{R}$*, independent of* s*. Then,* $P^*_{n,\varepsilon} = \mathbb{P}(\sup_{m \geq n} |\frac{1}{m} S_m - \frac{1}{m} z_m| > \varepsilon)$ *converges to 0 with rate of convergence given by* $\Phi_{\varepsilon,\Lambda}(\lambda) = \alpha^{\Lambda_{\frac{\varepsilon}{3\alpha},\frac{\varepsilon}{3},\alpha}(\frac{\lambda}{2})}$*, where* $\alpha = 1 + \frac{\varepsilon}{3W}$

Not all mathematicians think constructively!

The way this proof strategy is used in practice is noneffective.

A bound is calculated for the sum, then the monotone convergence theorem is used to obtain a rate of convergence.

Calculating rates of convergences directly does not trivially follow from a proof that the sum is bounded and new ideas must be introduced to the proof.

Not all mathematicians think constructively!

The way this proof strategy is used in practice is noneffective.

A bound is calculated for the sum, then the monotone convergence theorem is used to obtain a rate of convergence.

Calculating rates of convergences directly does not trivially follow from a proof that the sum is bounded and new ideas must be introduced to the proof.

Not all mathematicians think constructively!

The way this proof strategy is used in practice is noneffective.

A bound is calculated for the sum, then the monotone convergence theorem is used to obtain a rate of convergence.

Calculating rates of convergences directly does not trivially follow from a proof that the sum is bounded and new ideas must be introduced to the proof.

Improving known rates

In [Luz18], it is shown that if $\{X_n\}$ is a sequence of pairwise independent, identically distributed random variables with $\mathbb{E}(X_1) = 0$, $Var(X_1) = \sigma^2 < \infty$ and $\mathbb{E}(|X_1|) = \tau < \infty$ then for all $\beta > 1$

$$P_{n,\varepsilon} = \mathbb{P}(\{\omega \in \Omega : \sup_{m \geq n} |\frac{S_m}{m}| > \varepsilon\}) = O(\frac{log(n)^{\beta-1}}{n})$$

Using our quantitative general theorem we can get, for fixed $\varepsilon > 0$ $\forall \lambda > 0$ and $\forall n \geq \Phi(\lambda)$

$$P_{n,\varepsilon} \leq \lambda$$

where, $\Phi(\lambda) = \frac{36\alpha^2\sigma^2}{\lambda\varepsilon^2(\alpha-1)}$, with $\alpha = 1 + \frac{2\varepsilon}{3\tau}$.
Which implies

$$P_{n,\varepsilon} = O(\frac{1}{n})$$

Improving known rates

In [Luz18], it is shown that if $\{X_n\}$ is a sequence of pairwise independent, identically distributed random variables with $\mathbb{E}(X_1) = 0$, $Var(X_1) = \sigma^2 < \infty$ and $\mathbb{E}(|X_1|) = \tau < \infty$ then for all $\beta > 1$

$$P_{n,\varepsilon} = \mathbb{P}(\{\omega \in \Omega : \sup_{m \geq n} |\frac{S_m}{m}| > \varepsilon\}) = O(\frac{log(n)^{\beta-1}}{n})$$

Using our quantitative general theorem we can get, for fixed $\varepsilon > 0$ $\forall \lambda > 0$ and $\forall n \geq \Phi(\lambda)$

$$P_{n,\varepsilon} \leq \lambda$$

where, $\Phi(\lambda) = \frac{36\alpha^2\sigma^2}{\lambda\varepsilon^2(\alpha-1)}$, with $\alpha = 1 + \frac{2\varepsilon}{3\tau}$.

Which implies

$$P_{n,\varepsilon} = O(\frac{1}{n})$$

Improving known rates

In [Luz18], it is shown that if $\{X_n\}$ is a sequence of pairwise independent, identically distributed random variables with $\mathbb{E}(X_1) = 0$, $Var(X_1) = \sigma^2 < \infty$ and $\mathbb{E}(|X_1|) = \tau < \infty$ then for all $\beta > 1$

$$P_{n,\varepsilon} = \mathbb{P}(\{\omega \in \Omega : \sup_{m \geq n} |\frac{S_m}{m}| > \varepsilon\}) = O(\frac{log(n)^{\beta-1}}{n})$$

Using our quantitative general theorem we can get, for fixed $\varepsilon > 0$ $\forall \lambda > 0$ and $\forall n \geq \Phi(\lambda)$

$$P_{n,\varepsilon} \leq \lambda$$

where, $\Phi(\lambda) = \frac{36\alpha^2\sigma^2}{\lambda\varepsilon^2(\alpha-1)}$, with $\alpha = 1 + \frac{2\varepsilon}{3\tau}$.
Which implies

$$P_{n,\varepsilon} = O(\frac{1}{n})$$

Linear bounds are optimal

For every $\delta > 0$, we can obtain a sequence of pairwise iid random variables with mean 0, finite first moment and finite variance, such that,

$$\mathbb{P}(\sup_{m \geq n} |\frac{1}{m} S_m| > \varepsilon) \geq \frac{\omega}{n^{1+\delta}}$$

for some $\omega > 0$.

Our construction, however, does not rule out the possibility that $\mathbb{P}(\sup_{m \geq n} |\frac{1}{m} S_m| > \varepsilon) = O(\frac{1}{n log(n)})$, for example.

Linear bounds are optimal

For every $\delta > 0$, we can obtain a sequence of pairwise iid random variables with mean 0, finite first moment and finite variance, such that,

$$\mathbb{P}(\sup_{m\geq n} |\frac{1}{m}S_m| > \varepsilon) \geq \frac{\omega}{n^{1+\delta}}$$

for some $\omega > 0$.

Our construction, however, does not rule out the possibility that $\mathbb{P}(\sup_{m\geq n} |\frac{1}{m}S_m| > \varepsilon) = O(\frac{1}{nlog(n)})$, for example.

Thoughts for the future

Thoughts for the future

- We spoke about the role recursive inequalities play in establishing the convergence of algorithms all over mathematics.
- We then spoke about how looking at probabilistic convergence theorems is a fruitful area of potential proof mining.
- We can bring these two ideas together. It turns out that recursive inequalities are used to establish the convergence of stochastic algorithms, see [FG22]. The most famous of these is the Robbins-Siegmund theorem.

Thoughts for the future

- We spoke about the role recursive inequalities play in establishing the convergence of algorithms all over mathematics.
- We then spoke about how looking at probabilistic convergence theorems is a fruitful area of potential proof mining.
- We can bring these two ideas together. It turns out that recursive inequalities are used to establish the convergence of stochastic algorithms, see [FG22]. The most famous of these is the Robbins-Siegmund theorem.

Thoughts for the future

- We spoke about the role recursive inequalities play in establishing the convergence of algorithms all over mathematics.
- We then spoke about how looking at probabilistic convergence theorems is a fruitful area of potential proof mining.
- We can bring these two ideas together. It turns out that recursive inequalities are used to establish the convergence of stochastic algorithms, see [FG22]. The most famous of these is the Robbins-Siegmund theorem.

Robbins-Siegmund 1971

Theorem

Let $\{\mathcal{F}_n\}$ be a filtration on a probability space and $\{X_n\}, \{a_n\}, \{b_n\}, \{c_n\}$ be sequences of nonnegative $\mathcal{F}_n$-measurable random variables satisfying:

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) \leq X_n(1+b_n) + a_n - c_n \textit{ almost surely}$$

$$\sum_{n=1}^{\infty} a_n < \infty \textit{ almost surely}$$

$$\sum_{n=1}^{\infty} b_n < \infty \textit{ almost surely}$$

Then $\{X_n\}$ converges almost surely and $\sum_{n=1}^{\infty} c_n < \infty$ almost surely

Giving this theorem a computational interpretation could result in many applications in stochastic optimisation and stochastic approximation theory.

Thoughts for the future

In parallel with H. Cheval [https://github.com/hcheval] and A. Koutsoukou-Argyraki, Thomas Powell and I [https://github.com/Kejineri/Proof-mining-] have been formalising aspects of applied proof theory in lean.

Why is formalisation useful?

Why is having a comprehensive library on quantitative convergence results for sequences of real numbers satisfying recursive inequalities useful?

- Many Results in many areas of maths such as non-linear analysis, numerical analysis and convex optimization reduce lemmas on recursive inequalities. Thus, Such a library will allow further formalization work to be done.
- In say, an optimization algorithm, reducing $\|f(x_n) - f(x^*)\|$ to some recursive inequality is unusually some routine manipulation of properties of the mapping, algorithm and space it is acting on. One could therefore try to develop algorithms for automating this procedure.

Why is formalisation useful?

Why is having a comprehensive library on quantitative convergence results for sequences of real numbers satisfying recursive inequalities useful?

- Many Results in many areas of maths such as non-linear analysis, numerical analysis and convex optimization reduce lemmas on recursive inequalities. Thus, Such a library will allow further formalization work to be done.
- In say, an optimization algorithm, reducing $\|f(x_n) - f(x^*)\|$ to some recursive inequality is unusually some routine manipulation of properties of the mapping, algorithm and space it is acting on. One could therefore try to develop algorithms for automating this procedure.

Why is formalisation useful?

Why is having a comprehensive library on quantitative convergence results for sequences of real numbers satisfying recursive inequalities useful?

- Many Results in many areas of maths such as non-linear analysis, numerical analysis and convex optimization reduce lemmas on recursive inequalities. Thus, Such a library will allow further formalization work to be done.
- In say, an optimization algorithm, reducing $\|f(x_n) - f(x^*)\|$ to some recursive inequality is unusually some routine manipulation of properties of the mapping, algorithm and space it is acting on. One could therefore try to develop algorithms for automating this procedure.

Examples of formalization

This is the example of a sequence of computable numbers that converge to 0 without a computable rate of convergence.

```
def Specker_seq (α: nnseqR):= ∀ ϕ: N → N, (RoC ϕ α) → ¬ computable(ϕ)
@[reducible]
def s (α: nnseqQ): N → Q :=
λ n, if h: ∃ m : N, s_prop n m then α.1 (nat.find h) else 0
```

```
def  s_prop(n :N ) : N → Prop:= (λ (m : N),
      m ≤ n ∧
        (∃ (x : N), evaln n (of_nat_code m) 0 = some x) ∧
          ∀ (l : N), l < n → ∀ (w : N), ¬evaln l (of_nat_code m) 0 = some w)
```

Theorem ([Kohlenbach and Powell., 2020])

Lemma 3.4. *Let (θ_n) and (α_n) be sequences of nonnegative reals such that $\sum_{i=0}^{\infty} \alpha_i$ diverges, and suppose that for any $\varepsilon > 0$ there exist some $\delta > 0$ and $N \in \mathbb{N}$ such that*

$$(*) \quad \forall n \geq N(\varepsilon < \theta_{n+1} \to \theta_{n+1} \leq \theta_n - \alpha_n \cdot \delta).$$

Then $\theta_n \to 0$ as $n \to \infty$. Moreover, if:

(i) $K \in (0, \infty)$ satisfies $\theta_n < K$ for all $n \in \mathbb{N}$,
(ii) $r : \mathbb{N} \times (0, \infty) \to \mathbb{N}$ is a rate of divergence for $\sum_{i=0}^{\infty} \alpha_i$,
(iii) $N : (0, \infty) \to \mathbb{N}$ and $\varphi : (0, \infty) \to (0, \infty)$ witness property $(*)$ in the sense that for all $\varepsilon > 0$ we have

$$\forall n \geq N(\varepsilon)(\varepsilon < \theta_{n+1} \to \theta_{n+1} \leq \theta_n - \alpha_n \cdot \varphi(\varepsilon)),$$

then $\Psi_{K,r,N,\varphi}(\varepsilon) := r(N(\varepsilon), K/\varphi(\varepsilon)) + 1$ is a rate of convergence for (θ_n).

Examples of formalization

```
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  :=
```

Thank you!

References I

References

[Spe49] Ernst Specker. “Nicht Konstruktiv Beweisbare Sätze der Analysis”. In: *Journal of Symbolic Logic* 14 (1949), pp. 145–158.

[Kre51] Georg Kreisel. “On the Interpretation of Non-Finitist Proofs, Part I”. In: *Journal of Symbolic Logic* 16 (1951), pp. 241–267. DOI: 10.1017/S002248120010058l.

[Kre52] Georg Kreisel. “On the Interpretation of Non-Finitist Proofs, Part II: Interpretation of Number Theory”. In: *Journal of Symbolic Logic* 17 (1952), pp. 43–58. DOI: 10.2307/2267457.

[Kre68] Georg Kreisel. “A Survey of Proof Theory”. In: *The Journal of Symbolic Logic* 33.3 (1968), pp. 321–388. ISSN: 00224812. URL: http://www.jstor.org/stable/2270324 (visited on 05/30/2023).

References II

[Pet69] Valentin Vladimirovich Petrov. "On the strong law of large numbers". In: *Theory of Probability & Its Applications* 14.2 (1969), pp. 183–192.

[Sie75] D Siegmund. "Large deviation probabilities in the strong law of large numbers". In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 31.2 (1975), pp. 107–113.

[Ete81] Nasrollah Etemadi. "An elementary proof of the strong law of large numbers". In: *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 55.1 (1981), pp. 119–122.

[Fil83] James Allen Fill. "Convergence Rates Related to the Strong Law of Large Numbers". In: *The Annals of Probability* 11.1 (1983), pp. 123–142. DOI: 10.1214/aop/1176993663. URL: https://doi.org/10.1214/aop/1176993663.

References III

[Luc89] Horst Luckhardt. “Herbrand-Analysen Zweier Beweise Des Satzes von Roth: Polynomiale Anzahlschranken”. In: *The Journal of Symbolic Logic* 54.1 (1989), pp. 234–263. ISSN: 00224812. URL: http://www.jstor.org/stable/2275028.

[CG92] Tapas K Chandra and A Goswami. “Cesaro uniform integrability and the strong law of large numbers”. In: *Sankhyā: The Indian Journal of Statistics, Series A* (1992), pp. 215–231.

[AIS98] Yakov Alber, Alfredo Noel Iusem, and M. V. Solodov. “On the projected subgradient method for nonsmooth convex optimization in a Hilbert space”. In: *Mathematical Programming* 81.1 (1998), pp. 23–35. DOI: 10.1007/BF01584842.

References IV

[Koh05] Ulrich Kohlenbach. "Some computational aspects of metric fixed point theory". In: *Nonlinear Analysis* 61.5 (2005), pp. 823–837. DOI: 10.1016/j.na.2005.01.075.

[Koh08] Ulrich Kohlenbach. *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*. Springer Monographs in Mathematics. Springer, 2008. DOI: 10.1007/978-3-540-77533-1.

[AGT10] JEREMY AVIGAD, PHILIPP GERHARDY, and HENRY TOWSNER. "LOCAL STABILITY OF ERGODIC AVERAGES". In: *Transactions of the American Mathematical Society* 362.1 (2010), pp. 261–288. DOI: 10.1090/S0002-9947-09-04814-4.

[Gác10] Péter Gács. "A constructive law of large numbers with application to countable Markov chains". In: *Boston university* (2010).

References V

[KP10] VM Korchevsky and VV Petrov. "On the strong law of large numbers for sequences of dependent random variables". In: *Vestnik St. Petersburg University: Mathematics* 43 (2010), pp. 143–147.

[ADR12] Jeremy Avigad, Edward T. Dean, and Jason Rute. "A metastable dominated convergence theorem". In: *Journal of Logic & Analysis* 4.3 (2012), pp. 1–19. DOI: 10.4115/jla.2012.4.3.

[BB12] Errett Bishop and Douglas Bridges. *Constructive analysis*. Vol. 279. Springer Science & Business Media, 2012.

[KK15] Ulrich Kohlenbach and Angeliki Koutsoukou-Argyraki. "Rates of convergence and metastability for abstract Cauchy problems generated by accretive operators". In: *Journal of Mathematical Analysis and Applications* 423.2 (2015), pp. 1089–1112.

References VI

[CS16] Pingyan Chen and Soo Hak Sung. "A strong law of large numbers for nonnegative random variables and applications". In: *Statistics & Probability Letters* 118 (2016), pp. 80–86. ISSN: 0167-7152. DOI: https://doi.org/10.1016/j.spl.2016.06.017. URL: https://www.sciencedirect.com/science/article/pii/S0167715216300980.

[Koh17] Ulrich Kohlenbach. "Recent progress in proof mining in nonlinear analysis". In: *J. Appl. Logics–IfCoLoG J. Logics Their Appl* 4.10 (2017), pp. 3361–3410.

[Luz18] NUNO Luzia. "A SIMPLE PROOF OF THE STRONG LAW OF LARGE NUMBERS WITH RATES". In: *Bulletin of the Australian Mathematical Society* 97.3 (2018), pp. 513–517. DOI: 10.1017/S0004972718000059.

References VII

[Koh19] Ulrich Kohlenbach. “Proof-theoretic methods in nonlinear analysis”. In: *Proceedings of the International Congress of Mathematicians 2018*. Vol. 2. World Scientific, 2019, pp. 61–82. DOI: 10.1142/9789813272880_0045.

[FG22] Barbara Franci and Sergio Grammatico. “Convergence of sequences: A survey”. In: *Annual Reviews in Control* (2022). Article in press. DOI: 0.1016/j.arcontrol.2022.01.003.

[NP22] Morenikeji Neri and Thomas Powell. “A computational study of a class of recursive inequalities”. In: *arXiv preprint arXiv:2207.14559* (2022).