A computational study of a class of recursive inequalities

M. Neri T. Powell

Department of Computer science
University of Bath, Bath, United Kingdom

How do we analyse proofs

- We aim to give computational interpretations to theorems in Mathematics, such as bounds for realisers of existential statement

How do we analyse proofs

- We aim to give computational interpretations to theorems in Mathematics, such as bounds for realisers of existential statement
- We first give computational interpretations to assumptions

How do we analyse proofs

- We aim to give computational interpretations to theorems in Mathematics, such as bounds for realisers of existential statement
- We first give computational interpretations to assumptions
- The proofs are then analysed closely

How do we analyse proofs

- We aim to give computational interpretations to theorems in Mathematics, such as bounds for realisers of existential statement
- We first give computational interpretations to assumptions
- The proofs are then analysed closely
- If a direct interpretation cannot be found, proof interpretations such as the dialectica interpretation often offer equivalent 'metastable' versions of the statement that can be given a computational interpretation

How do we analyse proofs

- We aim to give computational interpretations to theorems in Mathematics, such as bounds for realisers of existential statement
- We first give computational interpretations to assumptions
- The proofs are then analysed closely
- If a direct interpretation cannot be found, proof interpretations such as the dialectica interpretation often offer equivalent 'metastable' versions of the statement that can be given a computational interpretation
- One can sometimes also produces negative results demonstrating that computable realisers cannot be produced

How do we analyse proofs

- We aim to give computational interpretations to theorems in Mathematics, such as bounds for realisers of existential statement
- We first give computational interpretations to assumptions
- The proofs are then analysed closely
- If a direct interpretation cannot be found, proof interpretations such as the dialectica interpretation often offer equivalent 'metastable' versions of the statement that can be given a computational interpretation
- One can sometimes also produces negative results demonstrating that computable realisers cannot be produced
- There are also metatheorems that sometimes tell us what type of computational content we can hope to extract

Convergence

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

Convergence

- $\forall \varepsilon \in \mathbb{Q}_+ \, \exists N \in \mathbb{N} \, \forall n \in \mathbb{N} \, (n \geq N \implies |a_i - a_j| \leq \varepsilon)$
- A natural question to ask is, from a proof of convergence can one find a computable function $f : \mathbb{Q} \to \mathbb{N}$ such that $\forall \varepsilon \in \mathbb{Q}_+ \, \forall n \in \mathbb{N} \, (n \geq N \implies |a_i - a_j| \leq \varepsilon)$

Convergence

- $\forall \varepsilon \in \mathbb{Q}_+ \, \exists N \in \mathbb{N} \, \forall n \in \mathbb{N} \, (n \geq N \implies |a_i - a_j| \leq \varepsilon)$
- A natural question to ask is, from a proof of convergence can one find a computable function $f : \mathbb{Q} \to \mathbb{N}$ such that $\forall \varepsilon \in \mathbb{Q}_+ \, \forall n \in \mathbb{N} \, (n \geq N \implies |a_i - a_j| \leq \varepsilon)$
- Specker showed this was not always possible, through his famous construction of a monotone sequence of rational numbers converging to a non-computable number

Convergence

- $\forall \varepsilon \in \mathbb{Q}_+ \, \exists N \in \mathbb{N} \, \forall n \in \mathbb{N} \, (n \geq N \implies |a_i - a_j| \leq \varepsilon)$
- A natural question to ask is, from a proof of convergence can one find a computable function $f : \mathbb{Q} \to \mathbb{N}$ such that $\forall \varepsilon \in \mathbb{Q}_+ \, \forall n \in \mathbb{N} \, (n \geq N \implies |a_i - a_j| \leq \varepsilon)$
- Specker showed this was not always possible, through his famous construction of a monotone sequence of rational numbers converging to a non-computable number
- $\forall \varepsilon \in \mathbb{Q}_+ \, \forall g : \mathbb{N} \to \mathbb{N} \, \exists n \forall i, j \in [n, n + g(n)](|a_i - a_j| \leq \varepsilon)$

Our analyse of recursive inequalities

- For a contraction mapping T with constant $c \in [0, 1)$ and 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

Our analyse of recursive inequalities

- For a contraction mapping T with constant $c \in [0,1)$ and 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
- $f(\varepsilon) = \lceil log_c(\frac{\varepsilon}{\mu_0}) \rceil$

Our analyse of recursive inequalities

- For a contraction mapping T with constant $c \in [0,1)$ and 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
- $f(\varepsilon) = \lceil log_c(\frac{\varepsilon}{\mu_0})\rceil$
- We aim to study the convergence properties of $\{\mu_n\}$ and $\{\beta_n\}$ satisfying, $\mu_{n+1} \leq \mu_n - \alpha_n\beta_n + \gamma_n$

Our analyse of recursive inequalities

- For a contraction mapping T with constant $c \in [0,1)$ and 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
- $f(\varepsilon) = \lceil log_c(\frac{\varepsilon}{\mu_0})\rceil$
- We aim to study the convergence properties of $\{\mu_n\}$ and $\{\beta_n\}$ satisfying, $\mu_{n+1} \leq \mu_n - \alpha_n\beta_n + \gamma_n$

(a) $\sum_{i=0}^{\infty} \alpha_i = \infty$

Our analyse of recursive inequalities

- For a contraction mapping T with constant $c \in [0,1)$ and 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
- $f(\varepsilon) = \lceil log_c(\frac{\varepsilon}{\mu_0}) \rceil$
- We aim to study the convergence properties of $\{\mu_n\}$ and $\{\beta_n\}$ satisfying, $\mu_{n+1} \leq \mu_n - \alpha_n\beta_n + \gamma_n$

(a) $\sum_{i=0}^{\infty} \alpha_i = \infty$

(bI) $\sum_{i=0}^{\infty} \gamma_i < \infty$

Our analyse of recursive inequalities

- For a contraction mapping T with constant $c \in [0, 1)$ and 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
- $f(\varepsilon) = \lceil log_c(\frac{\varepsilon}{\mu_0})\rceil$
- We aim to study the convergence properties of $\{\mu_n\}$ and $\{\beta_n\}$ satisfying, $\mu_{n+1} \leq \mu_n - \alpha_n\beta_n + \gamma_n$

(a) $\sum_{i=0}^{\infty} \alpha_i = \infty$

(bI) $\sum_{i=0}^{\infty} \gamma_i < \infty$

(bII) $\gamma_n/\alpha_n \to 0$ as $n \to \infty$

Our analyse of recursive inequalities

- For a contraction mapping T with constant $c \in [0,1)$ and 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
- $f(\varepsilon) = \lceil log_c(\frac{\varepsilon}{\mu_0}) \rceil$
- We aim to study the convergence properties of $\{\mu_n\}$ and $\{\beta_n\}$ satisfying, $\mu_{n+1} \leq \mu_n - \alpha_n\beta_n + \gamma_n$

(a) $\sum_{i=0}^{\infty} \alpha_i = \infty$

(bI) $\sum_{i=0}^{\infty} \gamma_i < \infty$

(bII) $\gamma_n / \alpha_n \to 0$ as $n \to \infty$

- Both conditions are a strengthening of $\gamma_n \to 0$ as $n \to \infty$

Applications of our abstract study of recursive inequalities

- Alber et al studied a general gradient decent algorithm for convex optimisation on Hilbert spaces

Applications of our abstract study of recursive inequalities

- Alber et al studied a general gradient decent algorithm for convex optimisation on Hilbert spaces
- We observed proving convergence of their algorithm required an application of case (I) of the recursive inequality we analysed

Applications of our abstract study of recursive inequalities

- Alber et al studied a general gradient decent algorithm for convex optimisation on Hilbert spaces
- We observed proving convergence of their algorithm required an application of case (I) of the recursive inequality we analysed
- We were also able to explain why the authors could not find a rate of convergence for their result

Applications of our abstract study of recursive inequalities

- Alber et al studied a general gradient decent algorithm for convex optimisation on Hilbert spaces
- We observed proving convergence of their algorithm required an application of case (I) of the recursive inequality we analysed
- We were also able to explain why the authors could not find a rate of convergence for their result
- We surveyed the proof mining literature and were able to demonstrate how many know results can be seen as special cases of our analysis

Subgradient decent

- Suppose that H is a real valued Hilbert space, $C \subseteq H$ a closed, convex subset of H, and $f : H \to \mathbb{R}$ a convex and continuous function

Subgradient decent

- Suppose that H is a real valued Hilbert space, $C \subseteq H$ a closed, convex subset of H, and $f : H \to \mathbb{R}$ a convex and continuous function
- The ε-subdifferential of f at $x \in H$ is defined by: $\partial_\varepsilon f(x) := \{u \in H \,|\, f(y) - f(x) \geq \langle u, y - x \rangle - \varepsilon \quad \text{for all } y \in H\}$

Subgradient decent

- Suppose that H is a real valued Hilbert space, $C \subseteq H$ a closed, convex subset of H, and $f : H \to \mathbb{R}$ a convex and continuous function
- The ε-subdifferential of f at $x \in H$ is defined by: $\partial_\varepsilon f(x) := \{u \in H \,|\, f(y) - f(x) \geq \langle u, y - x\rangle - \varepsilon \quad \text{for all } y \in H\}$
- $x_{n+1} = P_C\left(x_n - \frac{\alpha_n}{\nu_n} u_n\right) \quad \text{for} \quad u_n \in \partial_{\varepsilon_n} f(x_n) \quad \text{with} \quad u_n \neq 0$

Subgradient decent

- Suppose that H is a real valued Hilbert space, $C \subseteq H$ a closed, convex subset of H, and $f : H \to \mathbb{R}$ a convex and continuous function
- The ε-subdifferential of f at $x \in H$ is defined by: $\partial_\varepsilon f(x) := \{u \in H \mid f(y) - f(x) \geq \langle u, y - x \rangle - \varepsilon \quad \text{for all } y \in H\}$
- $x_{n+1} = P_C\left(x_n - \frac{\alpha_n}{\nu_n} u_n\right) \quad \text{for} \quad u_n \in \partial_{\varepsilon_n} f(x_n) \quad \text{with} \quad u_n \neq 0$
- where $\{\alpha_n\}$ satisfies $\sum_{i=0}^{\infty} \alpha_i = \infty$ and $\sum_{i=0}^{\infty} \alpha_i^2 < \infty$, $\{\varepsilon_n\}$ is a sequence of nonnegative error terms with $\varepsilon_n \leq \mu \alpha_n$ for some $\mu > 0$ and $\nu_n := \max\{1, \|u_n\|\}$. The algorithm halts if $0 \in \partial_{\varepsilon_n} f(x_n)$ at any point

Metastable subgrafdient decent

Let $x^* \in C$ be a minimizer of f on C, and suppose that $\{x_n\}$ is an infinite sequence generated by the algorithm, whose components satisfy all of the properties outlined above. Suppose that $\rho > 1$ is such that $\|u_n\| \leq \rho$ for all $n \in \mathbb{N}$. Then $f(x_n) \to f(x^*)$. Moreover, if r is a rate of divergence for $\sum_{i=0}^{\infty} \alpha_i = \infty$ and $L, K > 0$ are such that $\sum_{i=0}^{\infty} \alpha_i^2 \leq L$ and $\|x_0 - x^*\|^2 \leq K$, then for all $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$ we have

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

where

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

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

$$e := \frac{\rho(L + K)}{2} + (\mu + 2\rho)L$$

$$\theta := \rho + \mu$$

Future work and concluding remarks

- We are currently working on giving computational interpretations to probabilistic convergence results about sequences of random variables (The Martingale convergence theorem and the Robbins Siegmund theorem)

[1] https://github.com/mneri123/Proof-mining-

Future work and concluding remarks

- We are currently working on giving computational interpretations to probabilistic convergence results about sequences of random variables (The Martingale convergence theorem and the Robbins Siegmund theorem)
- We have started formalising aspects of applied proof theory of the Lean theorem prover [1]

[1] https://github.com/mneri123/Proof-mining-