Information flow in proofs by contradiction and effective learnability

Ulrich Kohlenbach
Department of Mathematics
Technische Universität Darmstadt

CIFMA-2022,Humboldt University Berlin, September 27, 2022

Applied Proof Theory: "Proof Mining"

'What more do we know if we have proved a theorem by restricted means than if we merely know that it is true? (G. Kreisel, 50's)

Applied Proof Theory: "Proof Mining"

'What more do we know if we have proved a theorem by restricted means than if we merely know that it is true? (G. Kreisel, 50's)

Input: Noneffective proof p of theorem A.

Applied Proof Theory: "Proof Mining"

'What more do we know if we have proved a theorem by restricted means than if we merely know that it is true? (G. Kreisel, 50's)

Input: Noneffective proof p of theorem A.

Goal: New information on A extracted by computing proof p' of A':

Applied Proof Theory: "Proof Mining"

'What more do we know if we have proved a theorem by restricted means than if we merely know that it is true? (G. Kreisel, 50's)

Input: Noneffective proof p of theorem A.

Goal: New information on A extracted by computing proof p' of A':

- Effective bounds,

Applied Proof Theory: "Proof Mining"

'What more do we know if we have proved a theorem by restricted means than if we merely know that it is true? (G. Kreisel, 50's)

Input: Noneffective proof $\boldsymbol{p}$ of theorem $\boldsymbol{A}$.

Goal: New information on $\boldsymbol{A}$ extracted by computing proof $\boldsymbol{p'}$ of $\boldsymbol{A'}$:

- Effective bounds,
- Algorithms,

Applied Proof Theory: "Proof Mining"

'What more do we know if we have proved a theorem by restricted means than if we merely know that it is true? (G. Kreisel, 50's)

Input: Noneffective proof p of theorem A.

Goal: New information on A extracted by computing proof p' of A':

- Effective bounds,
- Algorithms,
- Independence of the bound from certain data (uniformity).

Applied Proof Theory: "Proof Mining"

'What more do we know if we have proved a theorem by restricted means than if we merely know that it is true? (G. Kreisel, 50's)

Input: Noneffective proof p of theorem A.

Goal: New information on A extracted by computing proof p' of A':

- Effective bounds,
- Algorithms,
- Independence of the bound from certain data (uniformity).

The interpretation I exhibits the **flow of data** in the proof: this uses **new higher order concepts**!

“Proof Mining” in core mathematics

- During the last 20 years this proof-theoretic approach has resulted in **numerous new quantitative results** as well as **qualitative uniformity results** in: number theory, combinatorics, nonlinear analysis, fixed point theory, ergodic theory, topological dynamics, approximation theory, nonsmooth optimization etc.

“Proof Mining” in core mathematics

- During the last 20 years this proof-theoretic approach has resulted in **numerous new quantitative results** as well as **qualitative uniformity results** in: number theory, combinatorics, nonlinear analysis, fixed point theory, ergodic theory, topological dynamics, approximation theory, nonsmooth optimization etc.
- General **logical metatheorems** explain this as instances of logical phenomena (K. 2005, Gerhardy/K. 2008, TAMS).

Direct proofs: Herbrand's Theorem

Let $\exists x\, A_{qf}(x)$ be a sentence with quantifier-free A_{qf} provable in a theory $\mathcal{T}$ which is axiomatized by **purely universal axioms** $\forall y\, T_{qf}(y)$.

Direct proofs: Herbrand's Theorem

Let $\exists x\, A_{qf}(x)$ be a sentence with quantifier-free A_{qf} provable in a theory $\mathcal{T}$ which is axiomatized by **purely universal axioms** $\forall y\, T_{qf}(y)$.

Then there are **finitely many** closed terms (built up from the material in A_{qf}, T_{qf}) $s_1, \ldots, s_k, t_1, \ldots, t_n$ such that

$$\bigwedge_{i=1}^{k} T_{qf}(s_i) \rightarrow \bigvee_{j=1}^{n} A_{qf}(t_j)$$

is a **quasi-tautology**.

Direct proofs: Herbrand's Theorem

Let $\exists x\, A_{qf}(x)$ be a sentence with quantifier-free A_{qf} provable in a theory $\mathcal{T}$ which is axiomatized by **purely universal axioms** $\forall y\, T_{qf}(y)$.

Then there are **finitely many** closed terms (built up from the material in A_{qf}, T_{qf}) $s_1, \ldots, s_k, t_1, \ldots, t_n$ such that

$$\bigwedge_{i=1}^{k} T_{qf}(s_i) \to \bigvee_{j=1}^{n} A_{qf}(t_j)$$

is a **quasi-tautology**.

Hence $\exists x\, A_{qf}(x)$ has a **direct proof** by introducing quantifiers and using contractions.

Example (U. Berger)

Consider open theory $\mathcal{T} := \{\forall x(S(x) \neq 0)\}$ in language with equality, constant $\mathbf{0}$ and two unary function symbols S, f.

Example (U. Berger)

Consider open theory $\mathcal{T} := \{\forall x(S(x) \neq 0)\}$ in language with equality, constant 0 and two unary function symbols S, f.

Proposition

$\mathcal{T} \vdash \exists x\big(f(S(f(x))) \neq x\big).$

Example (U. Berger)

Consider open theory $\mathcal{T} := \{\forall x(S(x) \neq 0)\}$ in language with equality, constant $\mathbf{0}$ and two unary function symbols S, f.

Proposition

$\mathcal{T} \vdash \exists x\big(f(S(f(x))) \neq x\big).$

Proof: Suppose that

$$\forall x\big(f(S(f(x))) = x\big),$$

then f is **injective**, but also (since $S(x) \neq 0$) surjective on $\{x : x \neq 0\}$ and hence **non-injective**. **Contradiction!** $\square$

An analysis of the above proof extracts Herbrand terms $s_1, \ldots, s_k, t_1, \ldots, t_n$ s.t.

$$(\bigwedge_{i=1}^{k} S(s_i) \neq 0) \rightarrow \bigvee_{j=1}^{n} (f(S(f(t_j))) \neq t_j)$$

is a **quasi-tautology**.

An analysis of the above proof extracts Herbrand terms $s_1, \ldots, s_k, t_1, \ldots, t_n$ s.t.

$$(\bigwedge_{i=1}^{k} S(s_i) \neq 0) \rightarrow \bigvee_{j=1}^{n} (f(S(f(t_j))) \neq t_j)$$

is a **quasi-tautology**.

Indeed:

$$s_1 := f(f(0)),\ \ t_1 := 0,\ \ t_2 := f(0) \text{ or } t_3 := S(f(f(0)))$$

satisfy this.

An analysis of the above proof extracts Herbrand terms $s_1, \dots, s_k, t_1, \dots, t_n$ s.t.

$$\left(\bigwedge_{i=1}^{k} S(s_i) \neq 0\right) \to \bigvee_{j=1}^{n} (f(S(f(t_j))) \neq t_j)$$

is a **quasi-tautology**.

Indeed:

$$s_1 := f(f(0)),\ t_1 := 0,\ t_2 := f(0) \text{ or } t_3 := S(f(f(0)))$$

satisfy this.

One can even extract the finitely many instances of the equality axioms sufficient.

Nonfeasible number of terms

Consider the following fragment of number theory (due to P. Pudlak): $\mathcal{L}(\mathcal{T})$ contains constants $\mathbf{0}, \mathbf{1}$, function symbols $+, 2^{(\cdot)}$, a unary predicate $I(\cdot)$ for being an integer.

Nonfeasible number of terms

Consider the following fragment of number theory (due to P. Pudlak): $\mathcal{L}(\mathcal{T})$ contains constants $\mathbf{0}, \mathbf{1}$, function symbols $+, 2^{(\cdot)}$, a unary predicate $I(\cdot)$ for being an integer.

Non-logical axioms: $x + (y + z) = (x + y) + z$, $y + 0 = y$, $2^0 = 1$, $2^x + 2^x = 2^{1+x}$, $I(0)$, $I(x) \rightarrow I(1 + x)$.

Nonfeasible number of terms

Consider the following fragment of number theory (due to P. Pudlak): $\mathcal{L}(\mathcal{T})$ contains constants $0, 1$, function symbols $+, 2^{(\cdot)}$, a unary predicate $I(\cdot)$ for being an integer.

Non-logical axioms: $x + (y + z) = (x + y) + z,\ y + 0 = y,\ 2^0 = 1,$ $2^x + 2^x = 2^{1+x},\ I(0),\ I(x) \to I(1 + x).$

The conjunction of the universal closure of these non-logical axioms can be written as a sentence $\forall \underline{x}\, A_{qf}(\underline{x})$ with A_{qf} quantifier-free.

We use as an abbreviation: $2_0 := 0,\ 2_{k+1} := 2^{2_k}$.

Nonfeasible number of terms

Consider the following fragment of number theory (due to P. Pudlak): $\mathcal{L}(\mathcal{T})$ contains constants $0, 1$, function symbols $+, 2^{(\cdot)}$, a unary predicate $I(\cdot)$ for being an integer.

Non-logical axioms: $x + (y + z) = (x + y) + z$, $y + 0 = y$, $2^0 = 1$, $2^x + 2^x = 2^{1+x}$, $I(0)$, $I(x) \to I(1 + x)$.

The conjunction of the universal closure of these non-logical axioms can be written as a sentence $\forall \underline{x}\, A_{qf}(\underline{x})$ with A_{qf} quantifier-free.

We use as an abbreviation: $2_0 := 0,\ 2_{k+1} := 2^{2_k}$.

One can show that any **direct** proof of $\vdash \forall \underline{x}\, A_{qf}(\underline{x}) \to I(2_k)$ (without the use of logically involved intermediate concepts used as lemmas) has size $\geq 2_k$.

Pudlak continued: detours via new concepts

With the use of logically complex relations

$$R_0(x) := I(x), \quad R_{n+1}(x) := \forall y \big(R_n(y) \to R_n(2^x + y)\big)$$

in lemmas and modus ponens one can give a short proof of $I(2_k)$ (essentially linear on k):

Pudlak continued: detours via new concepts

With the use of logically complex relations

$$R_0(x) := I(x), \quad R_{n+1}(x) := \forall y \big(R_n(y) \to R_n(2^x + y)\big)$$

in lemmas and modus ponens one can give a short proof of $I(2_k)$ (essentially linear on k): by meta-induction on i one gives short derivations of

$$(*) \ R_i(0) \wedge \forall x \,(R_i(x) \to R_i(1 + x)) :$$

For the induction step observe that by $2^0 = 1$

$$R_{i+1}(0) \leftrightarrow \forall y \,(R_i(y) \to R_i(1 + y))$$

where the right-hand side is the 2nd conjunct from the I.H.

Moreover $R_{i+1}(x)$ implies using it twice (**contraction!**)

$$R_i(y) \to R_i(2^x + y) \text{ and } R_i(2^x + y) \to R_i(2^x + (2^x + y))$$

and so by syllogism (cut)

$$R_i(y) \to R_i(2^{1+x} + y), \text{ i.e. } R_{i+1}(1 + x).$$

Since

$$R_k(0) \to (R_{k-1}(0) \to (R_{k-2}(0) \to \ldots \to R_0(2_k) \ldots))$$

$R_0(2_k)$, i.e. $I(2_k)$ follows by k modus ponens applications using $(*)$.

Since

$$R_k(0) \rightarrow (R_{k-1}(0) \rightarrow (R_{k-2}(0) \rightarrow \ldots \rightarrow R_0(2_k)\ldots))$$

$R_0(2_k)$, i.e. $I(2_k)$ follows by k modus ponens applications using $(*)$.

Challenge in automated deduction: **guess useful intermediate lemmas** to **speed up** proofs!

Since

$$R_k(0) \to (R_{k-1}(0) \to (R_{k-2}(0) \to \ldots \to R_0(2_k)\ldots))$$

$R_0(2_k)$, i.e. $I(2_k)$ follows by k modus ponens applications using $(*)$.

Challenge in automated deduction: **guess useful intermediate lemmas** to **speed up** proofs!

Compression of proofs by use of **nested quantifiers**.

No fixed finite number of terms at all

Proposition

Let $f : \mathbb{N} \to \mathbb{N}$ be an arbitrary function. Then

$$\forall k \in \mathbb{N} \exists n \geq k (f(n) \leq f(n^2)).$$

No fixed finite number of terms at all

Proposition

Let $f : \mathbb{N} \to \mathbb{N}$ be an arbitrary function. Then

$$\forall k \in \mathbb{N} \exists n \geq k (f(n) \leq f(n^2)).$$

Proof: Consider

$$S_k := \{m \in \mathbb{N} : \exists n \geq k (f(n) = m)\}.$$

Since $S_k \neq \emptyset$ and $S_k \subseteq \mathbb{N}$, S_k has a smallest element $m_k \in S_k$. Let $n \geq k$ be such that $f(n) = m_k$. Then $f(n) \leq f(\tilde{n})$ for all $\tilde{n} \geq k$ which in particular applied to n^2. $\square$

Discussion 1

The proof uses **induction** along a predicate which is not quantifier-free:

Herbrand's theorem is **not valid** here!

Discussion 2

The above proofs establishes a stronger statement:

$$(+)\ \forall k \in \mathbb{N}\, \exists n \geq k\, \forall m \geq k\, (f(n) \leq f(m)).$$

Discussion 2

The above proofs establishes a stronger statement:

$$(+)\ \forall k \in \mathbb{N}\, \exists n \geq k\, \forall m \geq k\, (f(n) \leq f(m)).$$

This is essentially noneffective: there is a (low complexity) computable function f s.t. there is **no computable** $\alpha(k)$ which produces an $n = \alpha(k)$ with $(+)$.

Discussion 3

However, a solution n for k can be **learned with $(f(k)+1)$-many mind changes**:

Discussion 3

However, a solution n for k can be **learned with $(f(k)+1)$-many mind changes**:

First take $n_0 := k$. If one runs into a **counterexample**

$$m_0 \geq k \wedge f(n_0) > f(m_0)$$

change your mind from n_0 to $n_1 := m_0$ and so on.

Discussion 3

However, a solution n for k can be **learned with $(f(k)+1)$-many mind changes**:

First take $n_0 := k$. If one runs into a **counterexample**

$$m_0 \geq k \wedge f(n_0) > f(m_0)$$

change your mind from n_0 to $n_1 := m_0$ and so on.

Such a mind change can happen at most $(f(k)+1)$-many times since otherwise

$$f(k) = f(n_0) > \ldots > f(n_{f(k)+1}) \geq 0 \quad (\mathit{Contradiction!}).$$

The no-counterexample interpretation

The problem with **(+)** is that it has the form $\forall\exists\forall$ instead of $\forall\exists$.

The no-counterexample interpretation

The problem with **(+)** is that it has the form $\forall\exists\forall$ instead of $\forall\exists$.

However, statements of the form

$$\textbf{(1)}\ \forall k\, \exists n\, \forall m\, A_{qf}(k, n, m)$$

can be equivalently written in $\forall\exists$-form (using a **2nd order quantifier**)

$$\textbf{(2)}\ \forall k\, \forall g\, \exists n\, A_{qf}(k, n, g(n)).$$

The no-counterexample interpretation

The problem with $(+)$ is that it has the form $\forall\exists\forall$ instead of $\forall\exists$.

However, statements of the form

$$(1)\ \forall k\, \exists n\, \forall m\, A_{qf}(k,n,m)$$

can be equivalently written in $\forall\exists$-form (using a **2nd order quantifier**)

$$(2)\ \forall k\, \forall g\, \exists n\, A_{qf}(k,n,g(n)).$$

$(1) \Rightarrow (2)$ is trivial while $(2) \Rightarrow (1)$ follows by contradiction: let (2) be given but

$$\neg(1)\ \exists k\, \forall n\, \exists m\, \neg A_{qf}(k,n,m).$$

Then (choice function)

$$\exists k\, \exists g\, \forall n\, \neg A_{qf}(k,n,g(n))$$

which contradicts (2).

A constructive interpretation of **(2)** is given by a **3rd order** functional $\boldsymbol{\Phi}$ which refutes any attempt $\boldsymbol{g}$ to refute **(1) :**

$$\forall k\, \forall g\, A_{qf}(k, \Phi(k,g), g(\Phi(k,g)))$$

A constructive interpretation of **(2)** is given by a **3rd order** functional $\boldsymbol{\Phi}$ which refutes any attempt $\boldsymbol{g}$ to refute **(1) :**

$$\forall k\, \forall g\; A_{qf}(k, \Phi(k,g), g(\Phi(k,g)))$$

So $\boldsymbol{\Phi}$ solves the **no-counterexample interpretation** (G. Kreisel) of **(1)**.

Back to the example

Instead of

$$(+)\ \forall k \in \mathbb{N} \exists n \geq k \forall m \geq k (f(n) \leq f(m))$$

consider the **equivalent** no-counterexample formulation

$$(++)\ \forall k \in \mathbb{N}\, \forall g : \mathbb{N} \to \mathbb{N}\, \exists n \geq k\, (f(n) \leq f(g_k(n))),$$

where

$$g_k(n) := \max\{g(n), k\}.$$

In contrast to **(+)** its no-counterexample interpretation **(++)** has an effective solution!

In contrast to (+) its no-counterexample interpretation (++) has an effective solution!

Define **recursively**:

$$\Phi(g,k,0) := k, \quad \Phi(g,k,i+1) := g_k(\Phi(g,k,i)).$$

Assume that for all $i \leq f(k)$

$$f(\Phi(i)) > f(g_k(\Phi(i))),$$

i.e.

$$f(\Phi(i)) > f(\Phi(i+1))$$

and hence

$$f(k) = f(\Phi(0)) > \ldots > f(\Phi(f(k)+1)) \geq 0.$$

Contradiction!

Hence there exists an $i \leq f(k)$ with

$$f(\Phi(i)) \leq f(g_k(\Phi(i)))$$

and so (since $\Phi(i) \geq k$) for $\Psi(f,g,k) := \max_{i \leq f(k)} \Phi(g,k,i)$:

$$\exists n \leq \Psi(f,g,k)\ (n \geq k \wedge f(n) \leq f(g_k(n))).$$

For $g(n) := n^2$ we get

$$\exists n \leq k^{2^{f(k)}}\ (n \geq k \wedge f(n) \leq f(n^2)).$$

Comment: The number of potential witnessing data is no longer a **fixed** finite number (e.g. 3), but depends variably on $f(k), k$.

The monotone convergence principle

Let (a_n) be a nonincreasing sequence in $[0,1]$. Then, clearly, (a_n) is convergent and so a Cauchy sequence which we write as:

$$\textbf{(1)}\ \forall \mathbf{k} \in \mathbb{N}\, \exists \mathbf{n} \in \mathbb{N}\, \forall \mathbf{m} \in \mathbb{N}\, \forall \mathbf{i}, \mathbf{j} \in [\mathbf{n}; \mathbf{n}+\mathbf{m}]\ (|\mathbf{a_i} - \mathbf{a_j}| \leq 2^{-\mathbf{k}}),$$

where $[n; n+m] := \{n, n+1, \ldots, n+m\}$.

The monotone convergence principle

Let (a_n) be a nonincreasing sequence in $[0,1]$. Then, clearly, (a_n) is convergent and so a Cauchy sequence which we write as:

$$(1)\ \forall k \in \mathbb{N}\, \exists n \in \mathbb{N}\, \forall m \in \mathbb{N}\, \forall i,j \in [n;n+m]\ (|a_i - a_j| \le 2^{-k}),$$

where $[n;n+m] := \{n, n+1, \ldots, n+m\}$.
Then as above this is equivalent to

$$(2)\ \forall k \in \mathbb{N}\, \forall g \in \mathbb{N}^{\mathbb{N}}\, \exists n \in \mathbb{N}\, \forall i,j \in [n;n+g(n)]\ (|a_i - a_j| \le 2^{-k}).$$

By E. Specker 1949 there exist **computable** such sequences (a_n) even in $\mathbb{Q} \cap [0,1]$ **without computable bound** on '$\exists n$' in **(1).**

By E. Specker 1949 there exist **computable** such sequences (a_n) even in $\mathbb{Q} \cap [0,1]$ **without computable bound** on '$\exists n$' in **(1).**

By contrast, there is a **simple (primitive recursive) bound** $\Phi^*(g,k)$ on **(2)** (also referred to as **'metastability'** by T.Tao):

By E. Specker 1949 there exist **computable** such sequences (a_n) even in $\mathbb{Q} \cap [0,1]$ **without computable bound** on '$\exists n$' in **(1)**.

By contrast, there is a **simple (primitive recursive) bound** $\Phi^*(g,k)$ on **(2)** (also referred to as **'metastability'** by T.Tao):

Proposition

Let (a_n) be any nonincreasing sequence in $[0,1]$ then

$$\forall k \in \mathbb{N} \forall g \in \mathbb{N}^{\mathbb{N}} \exists n \leq \Phi^*(g,k) \forall i,j \in [n; n+g(n)] \, (|a_i - a_j| \leq 2^{-k}),$$

where

$$\Phi^*(g,k) := \tilde{g}^{(2^k-1)}(0) \text{ with } \tilde{g}(n) := n + g(n).$$

Moreover, there exists an $i < 2^k$ such that n can be taken as $\tilde{g}^{(i)}(0)$.

Noneffective proofs, effective learnability by finitely many mind changes

Noneffective proofs, effective learnability by finitely many mind changes

In both examples above the proof of the original statement makes use of the **law-of-excluded middle (LEM)** in the form

$$\Sigma^0_1\text{-LEM}: \ \forall n \in \mathbb{N}\, A_{qf}(k,n) \vee \exists n \in \mathbb{N} \neg A_{qf}(k,n).$$

Noneffective proofs, effective learnability by finitely many mind changes

In both examples above the proof of the original statement makes use of the **law-of-excluded middle (LEM)** in the form

$$\Sigma^0_1\text{-LEM}: \ \forall n \in \mathbb{N}\, A_{qf}(k,n) \vee \exists n \in \mathbb{N} \neg A_{qf}(k,n).$$

In fact, this is the weakest form of LEM sufficient here and - under general assumptions on the proof - for all theorems of the form

$$\forall k \in \mathbb{N}\, \exists n \in \mathbb{N}\, \forall m \in \mathbb{N}\, B_{qf}(k,n,m).$$

NONEFFECTIVE PROOFS, EFFECTIVE LEARNABILITY BY FINITELY MANY MIND CHANGES

In both examples above the proof of the original statement makes use of the **law-of-excluded middle (LEM)** in the form

$$\mathbf{\Sigma^0_1\text{-LEM}:}\ \forall n \in \mathbb{N}\, A_{qf}(k,n) \vee \exists n \in \mathbb{N} \neg A_{qf}(k,n).$$

In fact, this is the weakest form of LEM sufficient here and - under general assumptions on the proof - for all theorems of the form

$$\forall k \in \mathbb{N}\, \exists n \in \mathbb{N}\, \forall m \in \mathbb{N}\, B_{qf}(k,n,m).$$

As a result a **witness** is **not computable** in the parameter k but only **learnable** with a number of mind changes bounded as function $B(k)$ in k, where $B(k)$ corresponds to the number of instances of $\mathbf{\Sigma^0_1}$**-LEM** used.

Effective (B,L)-learnability (K./Safarik 2014)

Effective (B,L)-learnability (K./Safarik 2014)

Under very general assumptions on the proof (but not always) one gets bounds on the no-counterexample interpretation of the form

$$(f_2 \circ \tilde{g} \circ f_1)^{B(k)}(0)$$

with computable $B, f_1, f_2,$ where $f_2 \circ f_1$ essentially is the learning procedure L applied and $B(k)$ is the **number of mind changes**:

Effective (B,L)-learnability (K./Safarik 2014)

Under very general assumptions on the proof (but not always) one gets bounds on the no-counterexample interpretation of the form

$$(f_2 \circ \tilde{g} \circ f_1)^{B(k)}(0)$$

with computable B, f_1, f_2, where $f_2 \circ f_1$ essentially is the learning procedure L applied and $B(k)$ is the **number of mind changes**:

$$\exists n^{\mathbb{N}} \forall m^{\mathbb{N}}\ A_{qf}(k, n, m)$$

is **(B,L)-learnable** if

$$\exists i \leq B(k)\ \forall m\ A_{qf}(k, c_i, m), \text{ where}$$

$$c_0 := 0,$$

$$c_{i+1} := \begin{cases} L(m, k), & \text{for the } m \text{ with } \neg A_{qf}(k, c_i, m) \wedge \forall y < m\ A_{qf}(k, c_i, y) \text{ if } \exists \\ c_i, & \text{otherwise.} \end{cases}$$

Effective fluctuation bounds

Consider the special case for statements expressing the Cauchy property of a sequence.

Effective fluctuation bounds

Consider the special case for statements expressing the Cauchy property of a sequence.

- In the case of monotone sequences $(x_n) \subset [0, C]$ one always has the trivial fluctuation bound $2^k \cdot C$.

Effective fluctuation bounds

Consider the special case for statements expressing the Cauchy property of a sequence.

- In the case of monotone sequences $(x_n) \subset [0, C]$ one always has the trivial fluctuation bound $2^k \cdot C$.
- This might suggest that effective learnability always gives effective fluctuation bounds which, however, is false.

Hierarchy of quantitative Cauchy statements

Hierarchy of quantitative Cauchy statements

1. **rate ρ of convergence $\Rightarrow$**

Hierarchy of quantitative Cauchy statements

1. **rate ρ of convergence $\Rightarrow$**
2. **bound ($b := \rho$) on the number of fluctuations $\Rightarrow$**

Hierarchy of quantitative Cauchy statements

1. **rate ρ of convergence $\Rightarrow$**
2. **bound ($b := \rho$) on the number of fluctuations $\Rightarrow$**
3. **(B, L)-learnability $\Rightarrow$**

Hierarchy of quantitative Cauchy statements

1. **rate ρ of convergence** $\Rightarrow$
2. **bound ($b := \rho$) on the number of fluctuations** $\Rightarrow$
3. **(B, L)-learnability** $\Rightarrow$
4. **rate of metastability Ω**.

Proposition (K./Safarik,2014)

The **hierarchy is strict** in the sense that the existence of computable witnesses for level n not even follows from primitive recursive witnesses for level $n-1$ ($2 \leq n \leq 4$).

A Modular Approach: Proof Interpretations

A Modular Approach: Proof Interpretations

- **Interpret** the formulas A in $P : A \mapsto A^{\mathcal{I}}$,

A Modular Approach: Proof Interpretations

- **Interpret** the formulas A in $P : A \mapsto A^{\mathcal{I}}$,
- Interpretation $C^{\mathcal{I}}$ contains the **additional information**,

A Modular Approach: Proof Interpretations

- **Interpret** the formulas A in $P : A \mapsto A^{\mathcal{I}}$,
- Interpretation $C^{\mathcal{I}}$ contains the **additional information**,
- Construct by **recursion on** P a new proof $P^{\mathcal{I}}$ of $C^{\mathcal{I}}$.
  In particular: solve **modus ponens problem:**

$$\frac{A^I \quad , \quad (A \to B)^I}{B^I}.$$

Constructive reasoning: BHK-interpretation

Constructive reasoning: BHK-interpretation

This interpretation is an **informal attempt** to define a **constructive semantics** for the logical operations and quantifiers:

Constructive reasoning: BHK-interpretation

This interpretation is an **informal attempt** to define a **constructive semantics** for the logical operations and quantifiers:

(I) There is no proof for $\perp$.

Constructive reasoning: BHK-interpretation

This interpretation is an **informal attempt** to define a **constructive semantics** for the logical operations and quantifiers:

(i) There is no proof for $\perp$.

(ii) A proof of $A \wedge B$ is a pair (q, r) of proofs, where q is a proof of A and r is a proof of B.

Constructive reasoning: BHK-interpretation

This interpretation is an **informal attempt** to define a **constructive semantics** for the logical operations and quantifiers:

(I) There is no proof for $\bot$.

(II) A proof of $A \wedge B$ is a pair (q, r) of proofs, where q is a proof of A and r is a proof of B.

(III) A proof of $A \vee B$ is a pair (n, q) consisting of an integer n and a proof q which proves A if $n = 0$ and resp. B if $n = 1$.

CONSTRUCTIVE REASONING: BHK-INTERPRETATION

This interpretation is an **informal attempt** to define a **constructive semantics** for the logical operations and quantifiers:

(I) There is no proof for $\perp$.

(II) A proof of $A \wedge B$ is a pair (q, r) of proofs, where q is a proof of A and r is a proof of B.

(III) A proof of $A \vee B$ is a pair (n, q) consisting of an integer n and a proof q which proves A if $n = 0$ and resp. B if $n = 1$.

(IV) A proof p of $A \to B$ is a construction which transforms any hypothetical proof q of A into a proof $p(q)$ of B.

This interpretation is an **informal attempt** to define a **constructive semantics** for the logical operations and quantifiers:

(I) There is no proof for $\bot$.

(II) A proof of $A \wedge B$ is a pair (q, r) of proofs, where q is a proof of A and r is a proof of B.

(III) A proof of $A \vee B$ is a pair (n, q) consisting of an integer n and a proof q which proves A if $n = 0$ and resp. B if $n = 1$.

(IV) A proof p of $A \to B$ is a construction which transforms any hypothetical proof q of A into a proof $p(q)$ of B.

(V) A proof of $\forall x\, A(x)$ is a construction which produces for every element d of the domain a proof $p(d)$ of $A(d)$.

Constructive reasoning: BHK-interpretation

This interpretation is an **informal attempt** to define a **constructive semantics** for the logical operations and quantifiers:

(I) There is no proof for $\perp$.

(II) A proof of $A \wedge B$ is a pair (q, r) of proofs, where q is a proof of A and r is a proof of B.

(III) A proof of $A \vee B$ is a pair (n, q) consisting of an integer n and a proof q which proves A if $n = 0$ and resp. B if $n = 1$.

(IV) A proof p of $A \to B$ is a construction which transforms any hypothetical proof q of A into a proof $p(q)$ of B.

(V) A proof of $\forall x\, A(x)$ is a construction which produces for every element d of the domain a proof $p(d)$ of $A(d)$.

(VI) A proof p of $\exists x\, A(x)$ is a pair (d, q), where d is an element of the domain and q is a proof of $A(d)$.

Example: a simple case of the modus ponens

Consider proof

$$\frac{A \quad , \quad A \to B}{B},$$

where (for quantifier-free, decidable A_{qf}, B_{qf})

$$A :\equiv \forall k \, \exists n \, \forall m \, A_{qf}(k, n, m) \text{ and } B :\equiv \forall i \, \exists j \, B_{qf}(i, j).$$

Example: a simple case of the modus ponens

Consider proof

$$\frac{A \quad , \quad A \to B}{B},$$

where (for quantifier-free, decidable A_{qf}, B_{qf})

$$A :\equiv \forall k\, \exists n\, \forall m\, A_{qf}(k, n, m) \text{ and } B :\equiv \forall i\, \exists j\, B_{qf}(i, j).$$

Example: Suppose we have constructive proof of:

$$\forall k\, \exists n\, \forall m\, (|r_n - r_{n+m}| \le 2^{-k}) \to \forall i\, \exists j\, (s_j \le 2^{-i}),$$

where $(r_n), (s_n)$ are a nonincreasing sequences in $[0, 1]$.

Example: a simple case of the modus ponens

Consider proof

$$\frac{A \quad , \quad A \to B}{B},$$

where (for quantifier-free, decidable A_{qf}, B_{qf})

$$A :\equiv \forall k \, \exists n \, \forall m \, A_{qf}(k, n, m) \ \text{ and } \ B :\equiv \forall i \, \exists j \, B_{qf}(i, j).$$

Example: Suppose we have constructive proof of:

$$\forall k \, \exists n \, \forall m \, (|r_n - r_{n+m}| \leq 2^{-k}) \to \forall i \, \exists j \, (s_j \leq 2^{-i}),$$

where $(r_n), (s_n)$ are a nonincreasing sequences in $[0, 1]$.

By BHK there is a functional $\Phi : \mathbb{N}^{\mathbb{N}} \times \mathbb{N} \to \mathbb{N}$ such that

$$\forall f \in \mathbb{N}^{\mathbb{N}}, i \in \mathbb{N} \, \big(\forall k, m \, A_{qf}(k, f(k), m) \to B_{qf}(i, \Phi(f, i))\big).$$

Example: a simple case of the modus ponens

Consider proof

$$\frac{A \quad , \quad A \to B}{B},$$

where (for quantifier-free, decidable A_{qf}, B_{qf})

$$A :\equiv \forall k\, \exists n\, \forall m\, A_{qf}(k, n, m) \ \text{ and } \ B :\equiv \forall i\, \exists j\, B_{qf}(i, j).$$

Example: Suppose we have constructive proof of:

$$\forall k\, \exists n\, \forall m\, (|r_n - r_{n+m}| \leq 2^{-k}) \to \forall i\, \exists j\, (s_j \leq 2^{-i}),$$

where $(r_n), (s_n)$ are a nonincreasing sequences in $[0, 1]$.

By BHK there is a functional $\Phi : \mathbb{N}^{\mathbb{N}} \times \mathbb{N} \to \mathbb{N}$ such that

$$\forall f \in \mathbb{N}^{\mathbb{N}}, i \in \mathbb{N}\ \big(\forall k, m\, A_{qf}(k, f(k), m) \to B_{qf}(i, \Phi(f, i))\big).$$

Problem: **useless if no computable f satisfies the premise!**

Solution: use still higher types

Gödel's famous **functional ('Dialectica') interpretation** allows to treat even **noneffective proofs** of $\boldsymbol{A \to B}$ with much **stronger result**:

$$\forall Y\, \Big(\forall k, g\, A_{qf}(k, Y(k,g), g(Y(k,g))) \to B_{qf}(i, \Omega(i, Y))\Big).$$

Solution: use still higher types

Gödel's famous **functional ('Dialectica') interpretation** allows to treat even **noneffective proofs** of $A \rightarrow B$ with much **stronger result**:

$$\forall Y \left(\forall k, g\, A_{qf}(k, Y(k,g), g(Y(k,g))) \rightarrow B_{qf}(i, \Omega(i, Y))\right).$$

Important: Computable (usually low complexity) Y satisfying the premise **can be extracted** even from a **nonconstructive** proof of A**!**

Solution: use still higher types

Gödel's famous **functional ('Dialectica') interpretation** allows to treat even **noneffective proofs** of $A \to B$ with much **stronger result**:

$$\forall Y \left(\forall k, g\, A_{qf}(k, Y(k,g), g(Y(k,g))) \to B_{qf}(i, \Omega(i, Y)) \right).$$

Important: Computable (usually low complexity) Y satisfying the premise **can be extracted** even from a **nonconstructive** proof of A**!**

x: **first order**

g: **second order**

Y: **third order**

Ω: **fourth order**.

Solution: use still higher types

Gödel's famous **functional ('Dialectica') interpretation** allows to treat even **noneffective proofs** of $A \to B$ with much **stronger result**:

$$\forall Y \left(\forall k, g\, A_{qf}(k, Y(k,g), g(Y(k,g))) \to B_{qf}(i, \Omega(i, Y))\right).$$

Important: Computable (usually low complexity) Y satisfying the premise **can be extracted** even from a **nonconstructive** proof of A**!**

x: **first order**

g: **second order**

Y: **third order**

Ω: **fourth order**.

As the formulas interpreted are getting increasingly logically complex, arbitrary high finite order functionals are needed to analyze the flow of information in the proof.

The above analysis of

$$\forall k\, \exists n\, \forall m\ (|r_n - r_{n+m}| \leq 2^{-k}) \rightarrow \forall i\, \exists j\ (s_j \leq 2^{-i}),$$

where $(r_n), (s_n)$ are a nonincreasing sequences in $[0, 1]$, has applications even for $s_n := r_n$:

The above analysis of

$$\forall k\, \exists n\, \forall m\, (|r_n - r_{n+m}| \leq 2^{-k}) \to \forall i\, \exists j\, (s_j \leq 2^{-i}),$$

where $(r_n), (s_n)$ are a nonincreasing sequences in $[0,1]$, has applications even for $s_n := r_n$:

From an (even noneffective) **proof** that the **noneffectively convergent** sequence (r_n) converges to 0 one **extracts** a level-4 functional Ω which transforms the given (see above) rate of metastability Φ for (r_n) into an **effective rate of convergence** of r_n towards $0.$

The above analysis of

$$\forall k\, \exists n\, \forall m\, (|r_n - r_{n+m}| \le 2^{-k}) \to \forall i\, \exists j\, (s_j \le 2^{-i}),$$

where $(r_n), (s_n)$ are a nonincreasing sequences in $[0,1]$, has applications even for $s_n := r_n$:

From an (even noneffective) **proof** that the **noneffectively convergent** sequence (r_n) converges to 0 one **extracts** a level-4 functional Ω which transforms the given (see above) rate of metastability Φ for (r_n) into an **effective rate of convergence** of r_n towards $0.$

Roughly: Ω computes in the parameter i a counterfunction g_i and applies Φ to some modified i' and $g_{i'}$:

$$\Omega(i, \Phi) := \Phi(i', g_{i'}).$$

The above analysis of

$$\forall k\, \exists n\, \forall m\, (|r_n - r_{n+m}| \leq 2^{-k}) \to \forall i\, \exists j\, (s_j \leq 2^{-i}),$$

where $(r_n), (s_n)$ are a nonincreasing sequences in $[0,1],$ has applications even for $s_n := r_n$:

From an (even noneffective) **proof** that the **noneffectively convergent** sequence (r_n) converges to 0 one **extracts** a level-4 functional Ω which transforms the given (see above) rate of metastability Φ for (r_n) into an **effective rate of convergence** of r_n towards $0.$

Roughly: Ω computes in the parameter i a counterfunction g_i and applies Φ to some modified i' and $g_{i'}$:

$$\Omega(i, \Phi) := \Phi(i', g_{i'}).$$

Many new rates of convergence in nonlinear analysis have been obtained in this way!

An application: polynomial rate in Bauschke's solution of 'zero displacement conjecture'

Consider a Hilbert space H and nonempty closed and convex subsets $C_1, \ldots, C_N \subseteq H$ with metric projections P_{C_i}, define $T := P_{C_N} \circ \ldots \circ P_{C_1}$. In 2003 Bauschke proved the 'zero displacement conjecture':

$$\|T^{n+1}x - T^n x\| \to 0 \quad (x \in H).$$

Previously only known for $N = 2$ or $Fix(T) \neq \emptyset$ (or even $\bigcap_{i=1}^N C_i \neq \emptyset$) or C_i half spaces etc.

An application: polynomial rate in Bauschke's solution of 'zero displacement conjecture'

Consider a Hilbert space H and nonempty closed and convex subsets $C_1, \ldots, C_N \subseteq H$ with metric projections P_{C_i}, define $T := P_{C_N} \circ \ldots \circ P_{C_1}$. In 2003 Bauschke proved the 'zero displacement conjecture':

$$\|T^{n+1}x - T^n x\| \to 0 \quad (x \in H).$$

Previously only known for $N = 2$ or $Fix(T) \neq \emptyset$ (or even $\bigcap_{i=1}^{N} C_i \neq \emptyset$) or C_i half spaces etc.

Proof uses abstract theory of maximal monotone operators: Minty's theorem, Brezis-Haraux theorem, Rockafellar's maximal monotonicity and sum theorems, Bruck-Reich theory of strongly nonexpansive mappings, conjugate functions, normal cone operator...).

An application: polynomial rate in Bauschke's solution of 'zero displacement conjecture'

Consider a Hilbert space $\boldsymbol{H}$ and nonempty closed and convex subsets $\boldsymbol{C_1, \ldots, C_N \subseteq H}$ with metric projections $\boldsymbol{P_{C_i}}$, define $\boldsymbol{T := P_{C_N} \circ \ldots \circ P_{C_1}}$. In 2003 Bauschke proved the 'zero displacement conjecture':

$$\boldsymbol{\|T^{n+1}x - T^n x\| \to 0 \quad (x \in H).}$$

Previously only known for $\boldsymbol{N = 2}$ or $\boldsymbol{Fix(T) \neq \emptyset}$ (or even $\bigcap_{i=1}^{N} \boldsymbol{C_i} \neq \emptyset$) or $\boldsymbol{C_i}$ half spaces etc.

Proof uses abstract theory of maximal monotone operators: Minty's theorem, Brezis-Haraux theorem, Rockafellar's maximal monotonicity and sum theorems, Bruck-Reich theory of strongly nonexpansive mappings, conjugate functions, normal cone operator...).

Logical metatheorems guarantee an effective rate of convergence which only depends on $\boldsymbol{\varepsilon, N, b \geq \|x\|, K \geq \|c_i\|}$ for some $\boldsymbol{c_i \in C_i}$

Theorem (K. FoCM 2019)

$$\Phi(\varepsilon, N, b, K) := \left\lceil \frac{18b + 12\alpha(\varepsilon/6))}{\varepsilon} - 1 \right\rceil \left\lceil \left(\frac{D}{\omega(D, \tilde{\varepsilon})} \right) \right\rceil$$

is a **rate of asymptotic regularity** in Bauschke's result, where

$$\tilde{\varepsilon} := \frac{\varepsilon^2}{27b + 18\alpha(\varepsilon/6)}, \quad D := 2b + NK, \quad \omega(D, \tilde{\varepsilon}) := \frac{1}{16D}(\tilde{\varepsilon}/N)^2.$$

$$\alpha(\varepsilon) := \frac{(K^2 + N^3(N-1)^2K^2)N^2}{\varepsilon}.$$

Here $b \geq \|x\|$ and $K \geq \left(\sum_{i=1}^{N} \|c_i\|^2\right)^{\frac{1}{2}}$ for some $(c_1, \ldots, c_N) \in C_1 \times \ldots \times C_N$.

References

1) Kohlenbach, U., Applied Proof Theory. Springer Monograph in Mathematics. Springer 2008, xix+536pp.

2) Kohlenbach, U., Recent progress in proof mining in nonlinear analysis. IFCoLog Journal of Logics and their Applications, vol.10, Issue 4, pp. 3361-3410 (2017).

3) Kohlenbach, U., Proof-theoretic Methods in Nonlinear Analysis. Proc. ICM 2018, B. Sirakov, P. Ney de Souza, M. Viana (eds.), Vol. 2, pp. 61-82. World Scientific 2019.

4) Kohlenbach, U., A polynomial rate of asymptotic regularity for compositions of projections in Hilbert space. Foundations of Computational Mathematics vol. 19, pp. 83-99 (2019).

5) Kohlenbach, U., Safarik, P., Fluctuations, effective learnability and metastability in analysis. Ann. Pure and Applied Logic vol. 165, pp. 266-304 (2014).