Recent Developments in Proof Mining

Paulo Oliva

Queen Mary, University of London, UK
(pbo@dcs.qmul.ac.uk)

Birmingham, 24 August 2007

Outline

Outline

Proof Mining

Extraction of computational content from (ineffective) mathematical proofs

Proof Mining

Extraction of computational content from (ineffective) mathematical proofs

Proofs often carry more information than what is stated as theorem

Outline

Main Technique

Functional interpretations:

- **Dialectica** (Gödel'1958)
- **Diller-Nahm variant** (Diller/Nahm'1974)
- **Monotone Dialectica** (Kohlenbach'1990)
- **Bounded Dialectica** (Ferreira/O.'2005)

Simple Example

Theorem

$\forall n(f(n+1) > f(n)) \rightarrow \forall k(f(k+2) > f(k))$

Simple Example

Theorem

$\forall n(f(n+1) > f(n)) \rightarrow \forall k(f(k+2) > f(k))$

Proof.

Assume $\forall n(f(n+1) > f(n))$. From that we get both $f(k+1) > f(k)$ and $f(k+2) > f(k+1)$. By transitivity we get $f(k+2) > f(k)$. □

Simple Example

Theorem

$\forall n(f(n+1) > f(n)) \rightarrow \forall k(f(k+2) > f(k))$

Proof.

Assume $\forall n(f(n+1) > f(n))$. From that we get both $f(k+1) > f(k)$ and $f(k+2) > f(k+1)$. By transitivity we get $f(k+2) > f(k)$. □

Can we compute n given k?

Dialectica Interpretation

Interpret

$$\forall n(f(n+1) > f(n)) \to \forall k(f(k+2) > f(k))$$

as

$$\exists \phi \forall k\big(f(\phi k + 1) > f(\phi k) \to f(k+2) > f(k)\big)$$

Dialectica Interpretation

Interpret

$$\forall n(f(n+1) > f(n)) \to \forall k(f(k+2) > f(k))$$

as

$$\exists \phi \forall k \big(f(\phi k + 1) > f(\phi k) \to f(k+2) > f(k)\big)$$

Witness can be produced, e.g.

$$\phi k := \begin{cases} k & \text{if } f(k+1) \leq f(k) \\ k+1 & \text{otherwise} \end{cases}$$

Diller-Nahm Variant

Interpret

$$\forall n(f(n+1) > f(n)) \rightarrow \forall k(f(k+2) > f(k))$$

as

$$\exists \phi \forall k \big(\forall m \in \phi k \; (f(m+1) > f(m)) \rightarrow f(k+2) > f(k)\big)$$

Diller-Nahm Variant

Interpret

$$\forall n(f(n+1) > f(n)) \rightarrow \forall k(f(k+2) > f(k))$$

as

$$\exists \phi \forall k \big(\forall m \in \phi k \ (f(m+1) > f(m)) \rightarrow f(k+2) > f(k)\big)$$

Witness can be produced, e.g.

$$\phi k := \{k, k+1\}$$

Bounded Dialectica Interpretation

Interpret

$$\forall n(f(n+1) > f(n)) \to \forall k(f(k+2) > f(k))$$

as

$$\exists \phi \forall k \big(\forall m \leq \phi k \; (f(m+1) > f(m)) \to f(k+2) > f(k)\big)$$

Bounded Dialectica Interpretation

Interpret

$$\forall n(f(n+1) > f(n)) \rightarrow \forall k(f(k+2) > f(k))$$

as

$$\exists \phi \forall k \big(\forall m \leq \phi k \; (f(m+1) > f(m)) \rightarrow f(k+2) > f(k)\big)$$

Witness can be produced, e.g.

$$\phi k := k + 1$$

Monotone Dialectica Interpretation

Interpret

$$\forall n(f(n+1) > f(n)) \rightarrow \forall n(f(n+2) > f(n))$$

as

$$\exists \phi \exists \psi \leq^* \phi \forall k\big(f(\psi k + 1) > f(\psi k) \rightarrow f(k+2) > f(k)\big)$$

Monotone Dialectica Interpretation

Interpret

$$\forall n(f(n+1) > f(n)) \rightarrow \forall n(f(n+2) > f(n))$$

as

$$\exists \phi \exists \psi \leq^* \phi \forall k \big(f(\psi k + 1) > f(\psi k) \rightarrow f(k+2) > f(k) \big)$$

Witness can be produced, e.g.

$$\phi k := k + 1$$

Advantages of “Bounds”

Advantages of “Bounds”

- **Uniformity**

  Bounds don’t depend on bounded input

  E.g. compact spaces

Advantages of “Bounds”

- **Uniformity**

  Bounds don’t depend on bounded input

  E.g. compact spaces

- **Ineffective principles become interpretable**

  Witnesses may not be computable but can be bounded

  E.g. WKL

Advantages of “Bounds”

- **Uniformity**

  Bounds don’t depend on bounded input

  E.g. compact spaces

- **Ineffective principles become interpretable**

  Witnesses may not be computable but can be bounded

  E.g. WKL

- **Not much is lost**

  Bounds often give precise witness

  E.g. monotonicity, searchable set

Outline

1. Introduction
   - Proof Mining
   - Functional Interpretations
2. Recent Case Studies
   - Approximation Theory
   - Fixed Point Theory
   - Ergodic Theory

Approximation Theory

- Existence and uniqueness of best approximations

  E.g. approximate continuous functions by polynomials

Approximation Theory

- Existence and uniqueness of best approximations
  E.g. approximate continuous functions by polynomials
- **Existence**: quite often ineffective, non-computational

Approximation Theory

- Existence and uniqueness of best approximations
  E.g. approximate continuous functions by polynomials
- **Existence**: quite often ineffective, non-computational
- **Uniqueness**: of form that proof mining applies

$$\forall n^{\mathbb{N}}, f^{C[0,1]}, p_1^{P_n}, p_2^{P_n} (\|f - p_i\| = \text{best} \rightarrow \|p_1 - p_2\| = 0)$$

Approximation Theory

- Existence and uniqueness of best approximations
  E.g. approximate continuous functions by polynomials
- **Existence**: quite often ineffective, non-computational
- **Uniqueness**: of form that proof mining applies

$$\forall n^{\mathbb{N}}, f^{C[0,1]}, p_1^{P_n}, p_2^{P_n}(\|f - p_i\| = \text{best} \to \|p_1 - p_2\| = 0)$$

$$\forall n, f, p_1, p_2, l \exists k(\|f - p_i\| - \text{best} \leq 2^{-k} \to \|p_1 - p_2\| < 2^{-l})$$

L_1 Approximation

Theorem (Jackson'1921)

For any fixed $n \in \mathbb{N}$ and continuous function $f \in C[0,1]$ there exists a unique polynomial $p_n \in P_n$ such that $\|f - p_n\|_1$ is minimal.

L_1 Approximation

Theorem (Jackson'1921)

For any fixed $n \in \mathbb{N}$ and continuous function $f \in C[0,1]$ there exists a unique polynomial $p_n \in P_n$ such that $\|f - p_n\|_1$ is minimal.

Proof (Cheney'1965).

Mathematically elementary proof (just 2 pages), but logically intricate. Use of classical logic and WKL. □

L_1 Approximation

Theorem (Jackson'1921)

For any fixed $n \in \mathbb{N}$ and continuous function $f \in C[0,1]$ there exists a unique polynomial $p_n \in P_n$ such that $\|f - p_n\|_1$ is minimal.

Proof (Cheney'1965).

Mathematically elementary proof (just 2 pages), but logically intricate. Use of classical logic and WKL. □

How to compute p_n given f and n?

- Partial results during the 1970's [Björneståł'1975 and Kroó'1978]
- Explicit algorithm extracted from Cheney's 1965 proof [Kohlenbach/O. 2001]

Main Obstacle

Attainment of the infimum (WKL) used in proof of following lemma

Lemma (Original)

$\forall x \in A \, (f(x) \neq 0) \to \ldots$

Main Obstacle

Attainment of the infimum (WKL) used in proof of following lemma

Lemma (Original)

$\forall x \in A\,(f(x) \neq 0) \rightarrow \ldots$

WKL used to obtain distance from zero

$$\forall x \in A\,(f(x) \neq 0) \rightarrow \exists \delta \forall x \in A\,(|f(x)| \geq \delta)$$

Main Obstacle

Attainment of the infimum (WKL) used in proof of following lemma

Lemma (Original)

$\forall x \in A\,(f(x) \neq 0) \to \ldots$

WKL used to obtain distance from zero

$$\forall x \in A\,(f(x) \neq 0) \to \exists \delta \forall x \in A\,(|f(x)| \geq \delta)$$

We showed that the weaker version of the lemma is sufficient

Lemma (Weakening)

$\exists \delta \forall x \in A\,(|f(x)| \geq \delta) \to \ldots$

History (L_1 Approximation)

1921	Jackson	proof of existence and uniqueness
1965	Cheney	elementary proof of uniqueness
1975	Björnestål	ineff. existence of modulus on f, n
1978	Kroó	ineff. existence of modulus on ω_f, n
2001	Kohlenbach/O.	explicity modulus of uniqueness
2002	Oliva	complexity of L_1 approximation

Outline

Banach Theorem (1922)

- (X, d) complete metric space
- $f : X \to X$ is **contractive** if $d(f(x), f(y)) \leq \delta \cdot d(x, y) \quad (\delta < 1)$

Banach Theorem (1922)

- (X, d) complete metric space
- $f : X \to X$ is **contractive** if $d(f(x), f(y)) \leq \delta \cdot d(x, y) \quad (\delta < 1)$

Theorem (Banach'1922)

If $f : X \to X$ is contractive then f has a unique fixed-point.

Banach Theorem (1922)

- (X, d) complete metric space
- $f : X \to X$ is **contractive** if $d(f(x), f(y)) \leq \delta \cdot d(x, y) \quad (\delta < 1)$

Theorem (Banach'1922)

If $f : X \to X$ is contractive then f has a unique fixed-point.

For any $x_0 \in X$, $x_{n+1} := f(x_n)$ converges to the fixed-point

Browder/Göhde/Kirk Theorem (1965)

- $(X, \|\cdot\|)$ uniformly convex Banach space
- $C \subseteq X$ convex, closed and bounded
- $f : C \to C$ is **nonexpansive** if $\|f(x) - f(y)\| \leq \|x - y\|$

Browder/Göhde/Kirk Theorem (1965)

- $(X, \|\cdot\|)$ uniformly convex Banach space
- $C \subseteq X$ convex, closed and bounded
- $f : C \to C$ is **nonexpansive** if $\|f(x) - f(y)\| \leq \|x - y\|$

Theorem (Browder, Göhde, Kirk'1965)

If $f : C \to C$ is nonexpansive then f has a fixed-point.

Browder/Göhde/Kirk Theorem (1965)

- $(X, \|\cdot\|)$ uniformly convex Banach space
- $C \subseteq X$ convex, closed and bounded
- $f : C \to C$ is **nonexpansive** if $\|f(x) - f(y)\| \leq \|x - y\|$

Theorem (Browder, Göhde, Kirk'1965)

If $f : C \to C$ is nonexpansive then f has a fixed-point.

- If $f(C)$ compact $x_{n+1} := \frac{x_n + f(x_n)}{2}$ converges to a fixed-point
- Rate of convergence in general not computable (Kohlenbach)
- Can compute rate of **asymptotic regularity** of x_n

  i.e. how fast $\|x_n - f(x_n)\| \to 0$

Ishikawa Theorem (1976)

- $(X, \|\cdot\|)$ normed linear space
- $C \subseteq X$ convex and bounded

Theorem (Ishikawa'1976)

If $f : C \to C$ *is nonexpansive then* $\lim_{n\to\infty} \|x_n - f(x_n)\| = 0$, *where*

- $(\lambda_n)_{n\in\mathbb{N}} \in [0,1]$ *is divergent in sum and* $\limsup \lambda_n < 1$
- $x_{n+1} = (1-\lambda_n)x_n + \lambda_n f(x_n)$

Ishikawa Theorem (1976)

- $(X, \|\cdot\|)$ normed linear space
- $C \subseteq X$ convex and bounded

Theorem (Ishikawa'1976)

If $f : C \to C$ *is nonexpansive then* $\lim_{n\to\infty} \|x_n - f(x_n)\| = 0$, *where*

- $(\lambda_n)_{n\in\mathbb{N}} \in [0,1]$ *is divergent in sum and* $\limsup \lambda_n < 1$
- $x_{n+1} = (1-\lambda_n)x_n + \lambda_n f(x_n)$

Theorem (Borwein/Reich/Shafrir'1992)

Dropping the boundedness assumption on C *we still have* $\lim_{n\to\infty} \|x_n - f(x_n)\| = r_C(f)$.

History (asymptotic regularity)

1976	Ishikawa	no uniformity
1978	Edels./O'Brien	ineff. uniformity in x_0 (fixed λ)
1983	Goebel/Kirk	ineff. uniformity in x_0 and f $(*)$
1990	Goebel/Kirk	conjecture no uniformity in C
1992	Bor./Rei./Sha.	generalisation of Ishikawa $(*)$
1996	Baillon/Bruck	full uniformity for fixed λ
2001	Kohlenbach	full uniformity
2000	Kirk	uniformity on x_0, f (f direc. nonexp., fixed λ)
2003	Kohlen./Leust.	full uniformity, hyper. spc. and f direc. nonexp.

Outline

Ergodic

Systems or processes with the property that, given sufficient time, they include or affect all points in a given space

Ergodic

Systems or processes with the property that, given sufficient time, they include or affect all points in a given space

Such systems or processes can be represented statistically by a reasonably large selection of points

Ergodic Theory

- Let

  (X, Σ, μ) probability space

  $T : X \to X$ measure preserving transformation

Ergodic Theory

- Let

  (X, Σ, μ) probability space

  $T : X \to X$ measure preserving transformation

- T is **ergodic** if

  $\mu(A) \neq 0 \ \wedge \ \mu(X \backslash A) \neq 0 \quad \Longrightarrow \quad \mu(A \, \Delta \, T^{-1}(A)) \neq 0$

Ergodic Theory

- Let

  (X, Σ, μ) probability space

  $T : X \to X$ measure preserving transformation

- T is **ergodic** if

  $\mu(A) \neq 0 \;\wedge\; \mu(X \backslash A) \neq 0 \quad \Longrightarrow \quad \mu(A \,\Delta\, T^{-1}(A)) \neq 0$

- Study of **ergodic transformations**

Mean Ergodic Theorem (functional analysis)

- $\mathcal{H}$ Hilbert space
- $T : \mathcal{H} \to \mathcal{H}$ nonexpansive linear operator (i.e. $\|Tf\| \leq \|f\|$)
- $S_n f := f + Tf + \ldots + T^{n-1} f$
- $A_n f := \frac{S_n f}{n}$

Mean Ergodic Theorem (functional analysis)

- $\mathcal{H}$ Hilbert space
- $T : \mathcal{H} \to \mathcal{H}$ nonexpansive linear operator (i.e. $\|Tf\| \leq \|f\|$)
- $S_n f := f + Tf + \ldots + T^{n-1} f$
- $A_n f := \frac{S_n f}{n}$

Theorem (von Newmann)

The sequence $A_n f$ converges.

Mean Ergodic Theorem (functional analysis)

- $\mathcal{H}$ Hilbert space
- $T : \mathcal{H} \to \mathcal{H}$ nonexpansive linear operator (i.e. $\|Tf\| \leq \|f\|$)
- $S_n f := f + Tf + \ldots + T^{n-1} f$
- $A_n f := \frac{S_n f}{n}$

Theorem (von Newmann)

The sequence $A_n f$ converges.

What about rate of convergence?

Rate of Convergence

Convergence:

$$\forall \varepsilon^{\mathbb{Q}^*_+} \exists n^{\mathbb{N}} \forall m \geq n (\|A_m f - A_n f\| < \varepsilon)$$

Rate of Convergence

Convergence:

$$\forall \varepsilon^{\mathbb{Q}^*_+} \exists n^{\mathbb{N}} \forall m \geq n (\|A_m f - A_n f\| < \varepsilon)$$

Rate of convergence $r(\varepsilon)$ is such that

$$\forall \varepsilon^{\mathbb{Q}^*_+} \forall m \geq r(\varepsilon) (\|A_m f - A_{r(\varepsilon)} f\| < \varepsilon)$$

Rate of Convergence

Convergence:

$$\forall \varepsilon^{\mathbb{Q}_+^*} \exists n^{\mathbb{N}} \forall m \geq n (\|A_m f - A_n f\| < \varepsilon)$$

Rate of convergence $r(\varepsilon)$ is such that

$$\forall \varepsilon^{\mathbb{Q}_+^*} \forall m \geq r(\varepsilon) (\|A_m f - A_{r(\varepsilon)} f\| < \varepsilon)$$

Not computable in general!

Rate of Convergence (n.c.i)

Look at the no-counterexample interpretation of

$$\forall \varepsilon^{\mathbb{Q}^*_+} \exists n^{\mathbb{N}} \forall m \geq n (\|A_m f - A_n f\| < \varepsilon)$$

i.e.

$$\forall \varepsilon^{\mathbb{Q}^*_+}, M \exists n^{\mathbb{N}} (M(n) \geq n \rightarrow \|A_m f - A_n f\| < \varepsilon))$$

or, equivalently

$$\forall \varepsilon^{\mathbb{Q}^*_+}, K \exists n^{\mathbb{N}} \forall m \in [n, K(n)] \, (\|A_m f - A_n f\| < \varepsilon))$$

Rate of Convergence (n.c.i)

Look at the no-counterexample interpretation of

$$\forall \varepsilon^{\mathbb{Q}^*_+} \exists n^{\mathbb{N}} \forall m \geq n (\|A_m f - A_n f\| < \varepsilon)$$

i.e.

$$\forall \varepsilon^{\mathbb{Q}^*_+}, M \exists n^{\mathbb{N}} (M(n) \geq n \rightarrow \|A_m f - A_n f\| < \varepsilon))$$

or, equivalently

$$\forall \varepsilon^{\mathbb{Q}^*_+}, K \exists n^{\mathbb{N}} \forall m \in [n, K(n)] \, (\|A_m f - A_n f\| < \varepsilon))$$

Classically equivalent! Computationally better!

Avigad/Gerhardy

$$\forall \varepsilon^{\mathbb{Q}^*_+}, K \exists n^{\mathbb{N}} \forall m \in [n, K(n)] \, (\|A_m f - A_n f\| < \varepsilon))$$

Extraction of bound n given ε and K

Uses elimination of monotone Skolem functions (due to Kohlenbach)

Summary

- Approximation theory
  - *modulus of uniqueness*
  - classical logic, weak König's Lemma
- Fixed point theory
  - *modulus of asymptotic regularity*
  - convergence of bounded monotone sequences
- Ergodic theory
  - *modulus of convergence* (n.c.i.)
  - convergence of bounded monotone sequences