Proof mining and positive-bounded logic

Andrei Sipoș

Institute of Mathematics of the Romanian Academy & University of Bucharest

September 12, 2016
Colloquium Logicum
Hamburg, Germany

Proof mining

Proof mining (introduced and developed by U. Kohlenbach) aims to obtain quantitative information from proofs of theorems (from various areas of mathematics) of a nature which is not (fully) constructive. A comprehensive reference is:

U. Kohlenbach, *Applied proof theory: Proof interpretations and their use in mathematics*, Springer, Berlin/Heidelberg, 2008.

An extensive survey detailing the intervening research can be found in:

U. Kohlenbach, Recent progress in proof mining in nonlinear analysis, preprint, 2016.

The general question

Proof theory is one of the four main branches of logic and has as its scope of study proofs themselves (inside given logical systems), with a special aim upon consistency results, structural and substructural transformations, proof-theoretic ordinals et al.

The driving question of proof mining / interpretative proof theory is the following:

"What more do we know if we have proved a theorem by restricted means than if we merely know that it is true?"
(posed by G. Kreisel in the 1950s)

Kinds of information

By analysing a specific proof of a mathematical theorem, one could obtain, in addition:

- Terms coding effective algorithms for realizers and bounds of existentially quantified variables;
- Independence of certain parameters or at least continuity of the dependency;
- Weakening of premises.

In order for this to work, we must impose well-behavedness conditions upon the *logical system* and upon the *complexity of the conclusion of the theorem*.

The logical systems

We generally use systems of arithmetic in all finite types, intuitionistic or classical, augmented by restricted non-constructive rules (such as choice principles) and by types referring to (metric/normed/Hilbert) spaces and functionals involving them.

Two such systems are denoted by $\mathcal{A}^\omega_i[X, \langle\cdot,\cdot\rangle, C]$ (intuitionistic) and $\mathcal{A}^\omega[X, \langle\cdot,\cdot\rangle, C]$ (classical).
One typically uses proof interpretations to extract the necessary quantitative information. Metatheorems guaranteeing this fact were developed by Gerhardy and Kohlenbach in the 2000s.

A sample metatheorem is the following, for classical logic, which uses Gödel's functional interpretation, in its "monotone" variant introduced by Kohlenbach, combined with the negative translation.

The main classical metatheorem for proof mining

Theorem (Gerhardy and Kohlenbach, 2008)

Let σ, ρ, τ be types (subject to some restrictions). Let $s^{\sigma \to \rho}$ be a closed term of $\mathcal{A}^\omega[X, \langle \cdot, \cdot \rangle]$ and let $B_\forall(x^\sigma, y^\rho, z^\tau, u^0)$ (resp. $C_\exists(x^\sigma, y^\rho, z^\tau, v^0)$) be a $\forall$-formula with only x, y, z, u free (resp. an $\exists$-formula with only x, y, z, v free). If

$$\mathcal{A}^\omega[X, \langle \cdot, \cdot \rangle] \vdash \forall x^\sigma \forall y \leq_\rho s(x) \forall z^\tau (\forall u^0 B_\forall \to \exists v^0 C_\exists),$$

then there exists an extractable computable functional Φ such that for all $b \in \mathbb{N}$

$$\forall x^\sigma \forall y \leq_\rho s(x) \forall z^\tau (\forall u \leq \Phi(x) B_\forall \to \exists v \leq \Phi(x) C_\exists)$$

holds in every model for the corresponding system.

Proof interpretations

We have mainly two proof interpretations at our disposal:

- monotone modified realizability, which:
  - can extract bounds for all kinds of formulas;
  - does not permit the use of excluded middle;
- monotone functional interpretation (combined with negative translation), which:
  - can extract bounds only for Π_2 (that is, $\forall\exists$) formulas;
  - permits the use of excluded middle.

These "interpretations" have corresponding metatheorems like the one before which can be used to extract the required quantitative information. In some cases, where no set of restrictions is met, the two may be used in conjunction – see, e.g., Leuştean (2014).

What does “quantitative” mean?

Let us see what kind of information we might hope to extract. An example from nonlinear analysis would be a limit statement of the form:

$$\forall \varepsilon > 0 \exists N_\varepsilon \forall N \geq N_\varepsilon (\|x_n - A_n x_n\| < \varepsilon).$$

What we want to get is a “formula” for N_ε in terms of (obviously) ε and of some other arguments parametrizing our situation. Such a function is called a **rate of convergence** for the sequence.

As the formula above is not in a $\Pi_2/\forall\exists$ form, in some cases we are forced to only quantify its Herbrand normal form and obtain its so-called *rate of metastability* (in the sense of T. Tao).

Metric logics

As we have somehow anticipated, we work with a logic capable of dealing with metric/normed structures. Generally, there have been four major attempts at realising this:

1. continuous logics (Chang and Keisler, 1966)
2. logic of positive bounded formulas (Henson, 1976; Henson and Iovino, 2002)
3. compact abstract theories / CFO (Ben Yaacov et al., 2000s)
4. extensions of higher-order arithmetic systems (Kohlenbach, 2000s)

(4) was until now our topic and will also be our goal; (3) is a special case of (1), which is too general to be useful; (2) was shown to be equivalent to (3) and is preferable in some situations – of it we will talk in the sequel.

Positive-bounded logic – the signatures and models

Let's present the basic ideas of the logic of positive-bounded formulas.

The signatures are many-sorted first-order signatures, with the addition of a designated sort for real numbers. Also, the function symbols must include the canonical normed-space operations for each sort.

The models for such a signature consist of usual many-order first-order models with the property that each underlying set is a normed space w.r.t. the designated functions, the real sort is instantiated with the canonical real number structure and the operations are uniformly continuous on bounded subsets.

Positive-bounded logic – the language

The formulas are built up in this way:

1. the atomic formulas are of the form $t \leq r$ or $r \leq t$, where t is a term of the real sort and $r \in \mathbb{Q}$;
2. we allow conjunctions and disjunctions;
3. finally, we add "bounded" quantifiers, which act like $\exists x(\|x\| \leq r \wedge \varphi)$ and $\forall x(\|x\| \leq r \rightarrow \varphi)$, where $r \in \mathbb{Q}$.

This logic has been studied primarily from a model-theoretic perspective, using the concept of *approximate semantics*. Günzel and Kohlenbach have shown how it may be translated into proof theory.

The class $\mathcal{PBL}$

They defined the class $\mathcal{PBL}$, containing formulas of the following form:

$$\forall l^{\mathbb{N}} \forall_{r_1} x_1^{\tilde{X}} \exists_{s_1} y_1^{\tilde{X}} \ldots \forall_{r_m} x_m^{\tilde{X}} \exists_{s_m} y_m^{\tilde{X}} (T(\underline{x}, \underline{y}, l) =_{\mathbb{R}} 0)$$

where by $\tilde{X}$ we denote both reals and elements of the space, by $\forall_r$ (resp. $\exists_r$) we denote the bounded quantifier from before and the r_i's and s_i's may contain only the variable l.

The positive-bounded formulas may be easily translated into this class by first prenexing them and then performing some simple operations.

The first translation

The quantifier-free part of the prenexed formula is translated as follows:

- $r \leq_{\mathbb{R}} t$ is replaced by $\min\{r, t\} - r =_{\mathbb{R}} 0$;
- $t \leq_{\mathbb{R}} r$ is replaced by $\min\{r, t\} - t =_{\mathbb{R}} 0$;
- $t_1 =_{\mathbb{R}} 0 \vee t_2 =_{\mathbb{R}} 0$ is replaced by $\min\{|t_1|, |t_2|\} =_{\mathbb{R}} 0$;
- $t_1 =_{\mathbb{R}} 0 \wedge t_2 =_{\mathbb{R}} 0$ is replaced by $\max\{|t_1|, |t_2|\} =_{\mathbb{R}} 0$.

We will also need the assertion of uniform continuity, coded as:

$$U_m(T, \omega_T) := \forall n^0, b^0, l^0 \forall_b x_1, z_1, y_1, w_1, \ldots y_m, w_m$$

$$\left(\bigwedge_{i=1}^{m} \|x_i - z_i\|, \|y_i - w_i\| <_{\mathbb{R}} 2^{-\omega_T(b,n,l)} \rightarrow\right.$$

$$\left. |T(\underline{x}, \underline{y}, l) - T(\underline{z}, \underline{w}, l)| \leq_{\mathbb{R}} 2^{-n}\right)$$

The Δ-formulas

Another class of formulas they defined, into which the $\mathcal{PBL}$ formulas will be translated, is the one of Δ-formulas, of the following form:

$$\forall \underline{a} \exists \underline{b} \leq \underline{r}\underline{a} \forall \underline{c} B(\underline{a}, \underline{b}, \underline{c})$$

where B is quantifier-free and the possible types of $\underline{a}$, $\underline{b}$ and $\underline{c}$ (which are all variable sequences) are restricted to "admissible" types.

This type of formulas is usually well-behaved with respect to monotone variants of proof interpretations (see the book of Kohlenbach, 2008).

The second translation

A $\mathcal{PBL}$ formula like shown a while ago will be translated into the following Δ-formula:

$$\exists \underline{Y} \forall l^{\mathbb{N}} \forall \underline{x} \left(\bigwedge_{i=1}^{m} Y_i \leq \lambda y, l.s_i \wedge T(\underline{\tilde{x}}, \underline{Y \tilde{x}} l, l) =_{\mathbb{R}} 0 \right)$$

i.e. a Skolemization that also uses a special construction $x \mapsto \tilde{x}$ and where we can remove the boundings of universal quantifiers by reasons of extensionality and uniform continuity.

Showing the validity of this translation is the non-trivial portion of proving Günzel and Kohlenbach's metatheorem *(see the following slide)*.

The metatheorem for $\mathcal{PBL}$

Theorem (Günzel and Kohlenbach, 2016)

Let Θ be a set of $\mathcal{PBL}$-sentences for which the $U_m(T, \omega_T)$ are provable in $\mathcal{A}^\omega[X, \|\cdot\|]$. Let $B_\forall(x,u)$ (resp. $C_\exists(x,v)$) be a $\forall$-formula with only x, u free (resp. an $\exists$-formula with only x, v free). If

$$\mathcal{A}^\omega[X, \|\cdot\|] + \Theta \vdash \forall x^\rho (\forall u^0 B_\forall \to \exists v^0 C_\exists),$$

then there exists an extractable computable functional Φ such that for all x and x^* (where x^* is of a corresponding "number" type and majorizes x) we have that

$$\forall u \leq \Phi(x^*) B_\forall(x,u) \to \exists v \leq \Phi(x^*) C_\exists(x,v)$$

holds in every model for the corresponding system.

Examples of spaces

In their paper, the following classes of spaces axiomatizable by Δ-formulas were considered:

- Banach lattices, with the following subclasses:
  - $L^p(\mu)$ lattices
    - of which, also for atomless μ
  - $C(K)$ lattices
  - BL^pL^q lattices

Our goal will be to adapt an axiomatization of $L^p(\mu)$ as Banach spaces without necessarily considering a lattice structure.

We shall use the following characterization for $L^p(\mu)$ spaces:

Theorem (Lindenstrauss, Pelczynski, Tzafriri – late 1960s)

A Banach space is isomorphic to a $L^p(\mu)$ space iff for all $\varepsilon > 0$ and all finite-dimensional subspaces B of it, there is a finite-dimensional subspace C which contains B and is "$(1+\varepsilon)$-isometric" (gauging by the Banach-Mazur distance) to a finite-dimensional Banach space with the standard p-norm.

Although implicit, the above is not appropriate for a logical treatment because we have no a priori bound on the dimension of C – which would leed to infinitely-long formulas / unbounded quantifiers.

The useful characterization

The trick is to use portions of the proof of the "only if" direction (found in an 1973 book by Lindenstrauss and Tzafriri) + some ideas of Henson and Raynaud (2007) which give us the following alternate characterization as an intermediate step of the proof.

Theorem (A.S.)

A Banach space X is isomorphic to a $L^p(\mu)$ space iff for all $x_1, \ldots, x_n$ in X of norm less than 1 and for all $N \in \mathbb{N}_{\geq 1}$, there is a subspace $C \subseteq X$ and $y_1, \ldots, y_n$ in C of norm less than 1 such that C is of dimension at most $(4nN+1)^n$, it is isometric to $\mathbb{R}_p^{\dim_{\mathbb{R}} C}$ and for all i, $\|x_i - y_i\| \leq \frac{1}{N}$.

As you can see the C is now bounded and the approximation was "permuted". The proof also gives us the norm bounds which are useful for the final axiomatization.

The first set of axioms

We can translate the condition from before into a crude set of "first-order" axioms (i.e. the $A_{n,N}$'s):

$$\psi_m(\underline{z}) := \forall \underline{\lambda} \left(\left\| \sum_{i=1}^m \lambda_i z_i \right\| = \left(\sum_{i=1}^m |\lambda_i|^p \right)^{\frac{1}{p}} \right)$$
$$\psi'_{m,n}(\underline{y}, \underline{z}) := \bigwedge_{k=1}^n \left(\exists \underline{\lambda} \left(y_k = \sum_{i=1}^m \lambda_i z_i \right) \right)$$
$$\psi''_{n,N}(\underline{x}, \underline{y}) := \bigwedge_{k=1}^n \left(\|x_k - y_k\| \leq \frac{1}{N+1} \wedge \|y_k\| \leq 1 \right)$$
$$\varphi_{n,m,N}(\underline{x}) := \exists \underline{y} \exists \underline{z} \left(\psi_m(\underline{z}) \wedge \psi'_{m,n}(\underline{y}, \underline{z}) \wedge \psi''_{n,N}(\underline{x}, \underline{y}) \right)$$
$$\phi_{n,N}(\underline{x}) := \bigvee_{0 \leq m \leq (4nN+1)^n} \varphi_{n,m,N}(\underline{x})$$
$$A_{n,N} := \forall \underline{x} \left(\left(\bigwedge_{k=1}^n \|x_k\| \leq 1 \right) \to \phi_{n,N}(\underline{x}) \right)$$

Why are they Δ?

Still, the question remains as to why we can treat the axioms from before as Δ axioms with which we can work. What we must do is bound the existential quantifiers.

Firstly, we can consider the generators being of norm lesser than 1 through a operation (so we do not add complexity to the formulas); then, by the first formula, we can see that each z_i must be of norm equal to one; the y_k's are already bounded by 1; finally, the second formula combined with the first and the x-bound leads to a similar bound on the λ's.

The final axiom

Therefore, the Δ-formulation of the $A_{n,N}$'s, denoted by B, is defined as follows:

$$\psi(m,z) := \forall \lambda^{1(0)(0)} \left(\left\| \sum_{i=1}^{m} |\lambda(i)|_{\mathbb{R}} \cdot_X z(i) \right\| =_{\mathbb{R}} \left(\sum_{i=1}^{m} |\lambda(i)|_{\mathbb{R}}^{p} \right)^{1/p} \right)$$
$$\psi'(m,n,y,z,\lambda) := \forall k \preceq_0 (n-1)$$
$$\left(y(k+1) =_X \sum_{i=1}^{m} \lambda(i) \cdot_C z(i) \right)$$
$$\psi''(n,N,x,y) := \forall k \preceq_0 (n-1)$$
$$\left(\left\| \widetilde{x(k+1)} - y(k+1) \right\| \leq_{\mathbb{R}} \frac{1}{N} \wedge \|y(k+1)\| \leq_{\mathbb{R}} 1 \right)$$
$$\varphi(n,m,N,x,y,z,\lambda) := \psi(m,z) \wedge \psi'(m,n,y,z,\lambda) \wedge \psi''(n,N,x,y)$$
$$B := \forall n^0, N^0 \geq 1 \forall x^{X(0)} \exists y, z \preceq_{X(0)(0)} 1_{X(0)(0)} \exists \lambda^{1(0)(0)(0)} \in [-2,2]$$
$$\exists m \preceq_0 (4nN+1)^n \varphi(n,m,N,x,y,z,\lambda)$$

The final axiom

Note that we have used an $\exists$-quantifier instead of the disjunction and also universal quantifiers for n and N instead of an infinite list of axioms – this results in a slightly stronger system (since we dont have here an ω-rule), though not for standard models which appear in the conclusion of our metatheorem.

The soundness theorem

The above theorems and discussion certify that this new theory with the axiom B added (plus a constant for p with the constraint $1 \leq p$) accurately captures the $L^p(\mu)$ spaces.

Theorem (A.S.)

Let X be a Banach space and $p \geq 1$. Denote by $\mathcal{S}^{\omega,X}$ its associated set-theoretic model and let the constant c_p in our extended signature take as a value the canonical representation of the real number p. Then $\mathcal{S}^{\omega,X}$ is a model of $\mathcal{A}^\omega[X, \|\cdot\|, \mathcal{C}, L^p]$ iff X is isomorphic to some $L^p(\Omega, \mathcal{F}, \mu)$ space.

The metatheorem

Theorem (A.S.)

Let $B_\forall(x,u)$ (resp. $C_\exists(x,v)$) be a $\forall$-formula with only x,u free (resp. an $\exists$-formula with only x,v free). If

$$\mathcal{A}^\omega[X,\|\cdot\|,L^p] \vdash \forall x^\rho(\forall u^0 B_\forall \to \exists v^0 C_\exists),$$

then there exists an extractable computable functional Φ such that for all x and x^* (where x^* is of a corresponding "number" type and majorizes x) we have that

$$\forall u \leq \Phi(x^*) B_\forall(x,u) \to \exists v \leq \Phi(x^*) C_\exists(x,v)$$

holds in every $L^p(\mu)$ Banach space.

Applications

This new axiomatization, being essentially a comparison theorem with the p-normed Euclidean spaces, is particularly application-friendly. Given an existing proof on $L^p(\mu)$ spaces, one can translate it into this system because:

- when narrowing it on the $\mathbb{R}^n_p$ case, the statements about integrals become statements about sums of real numbers, which are most probably provable in $\mathcal{A}^\omega$ and/or universal formulas which can be freely added to the system with no regard to the bound extraction;
- the approximation argument involves a series of ε-style bounds, which are also probably formalizable.

Worked example

As an example, we have formalized the proof on the validity of the canonical modulus of uniform convexity for this type of spaces (in the case $p \geq 2$).

Theorem (A.S.)

Provably in our system (plus the axiom $2 \leq_{\mathbb{R}} c_p$ and other two universal ones), the function $\eta : (0, 2] \to (0, \infty)$, defined, for any ε, by $\eta(\varepsilon) := 1 - (1 - (\frac{\varepsilon}{2})^p)^{1/p}$, is a modulus of uniform convexity.

In this case, the intermediate result about real numbers is primarily Clarkson's inequality.

All these results can be found in a detailed form at arXiv:1609.02080.

Thank you for your attention.