Proof Mining in Analysis: Computability and Complexity

Paulo Oliva

Progress Report

BRICS

BRICS Ph.D. School
Department of Computer Science
University of Aarhus
Denmark

May 2001 **Supervisor:** Ulrich Kohlenbach

Proof Mining in Analysis: Computability and Complexity

Paulo Oliva

May 4, 2001

Contents

Abstract

Proof Mining consists in extracting from prima facie non-constructive proofs constructive information. For that purpose many techniques mostly based on negative translation, realizability, functional interpretation, A-translation and combinations thereof have been developed. After introducing various different approaches to Proof Mining we present an application of 'monotone functional interpretation' to a modification of a classical proof from 1921 of uniqueness of L_1-approximation of continuous functions by polynomials of degree $\leq n$. From that proof we extracted the first effective uniform rate of strong uniqueness (also called 'uniform modulus of uniqueness') for L_1-approximation, which existence had been studied for 80 years but always in a non-constructive manner. For that reason, previous results only stated the dependencies of the modulus but explicit constants had never been obtained. We also show how this modulus of uniqueness can be used to compute the actual best L_1-approximation of the given continuous functions on a compact interval. Finally, we analyze the complexity of the resulting algorithm using the tools of Computable Analysis (cf. [Ko91] and [Wei00]). The first part of this work (the extraction of the uniform modulus of uniqueness) has been concluded, the last two parts are under development. This report also includes indications of future areas of work.

1 Introduction

Together with the development of Set Theory by Cantor at the end of the 19th century Kronecker started to argue for a constructive approach to mathematics. He would be followed by Poincaré, Borel and finally by Brouwer with his thesis (1907) which founded modern intuitionism. Brouwer questioned the fact that some abstract notions used in mathematics have a priori validity. These mainly include the use of classical logic[1] and the use of non-intuitive objects and principles. The mathematics making use of classical logic and these cited objects and principles (which is used in daily life by mathematicians) was then labelled *classical mathematics*, $\mathcal{A}$, and the mathematics suggested by Brouwer was named *intuitionistic mathematics*, $\mathcal{A}_i$.

As an attempt to justify the use of classical mathematics Hilbert presented his famous Consistency Program (CP). Roughly speaking, Hilbert restricted $\mathcal{A}$ to what he called *finitary mathematics* $\mathcal{A}_f$ which was supposed to be accepted as consistent per se.[2] The CP consisted then in proving (in $\mathcal{A}_f$) that $\mathcal{A}$ is consistent (i.e. that not every formula is a theorem of $\mathcal{A}$). Once the goal of the CP had been achieved we would have proven that classical mathematics is conservative over intuitionistic mathematics for so-called *real sentences* (i.e. Π^0_1-sentences), and the mathematical community could safely continue to use classical mathematics.

The natural first step to prove the consistency of $\mathcal{A}$ would be to prove the consistency of the arithmetical part of $\mathcal{A}$, namely first order Peano Arithmetic (**PA**), and that had to be done in a system like **PRA**.

[1] Classical logic is characterised by the law of excluded middle ($A \vee \neg A$) or proof by contradiction ($\neg\neg A \to A$)

[2] The system $\mathcal{A}_f$ was never defined precisely but some of its properties were described by Hilbert. Afterwards, following a work of Skolem (1923), Hilbert and Bernays (1934) defined the system **PRA** (Primitive Recursive Arithmetic) which came to be accepted as a finitary arithmetical system.

That was shown, however, to be an impossible task by Gödel in his second incompleteness theorem. The theorem basically stated that any arithmetic system containing the standard functions of number theory could not prove its own consistency, which immediately implies that even **PA** could not prove its own consistency. One could still ask which kind of abstract notion has to be added to **PA** in order to prove that **PA** is consistent. Just before 1940 Gentzen showed that by using $TI_{<\varepsilon_0}$ (transfinite induction up to ε_0) one could prove the consistency of classical arithmetic. By $TI_{<\varepsilon_0}$ it is possible to show that any proof in **PA** can be transformed into a cut free proof, and since a contradiction (e.g. $0 = 1$) cannot be obtained via a cut free proof, the system **PA** is consistent.

A different approach was taken by Gödel around the same time (but published only in 1958). He defined a system **T** based on functionals of finite type, and as abstract notion he introduced primitive recursion at every finite type. Moreover he defined an interpretation (nowadays called 'functional interpretation', '*Dialectica* interpretation' or 'D-interpretation') of **HA** (Heyting Arithmetic) in **T**, i.e., whenever **HA** proves a formula A then **T** proves the interpretation of A. Together with a translation of **PA** into **HA** [3] one gets an interpretation of classical arithmetic in **T**. Since **T** could be shown to be consistent, and the translation of $0 = 1$ was the identity, we obtain that **PA** is consistent. The *Dialectica* interpretation was present in Gödel's work since its very beginning. In lectures from 1933, 1938 and 1941 Gödel introduces progressively the ideas that would culminate with the formalization of his system **T** and the D-interpretation published in 1958 in the journal *Dialectica* (in German). When an English translation of that article (in 1965) was suggested Gödel took up the job of revising the translation by adding and rewriting footnotes. Since Gödel was never satisfied with the result the English translation was published posthumously only in 1990 in his *Collected Works*.[4]

The functional interpretation of Gödel had a very nice feature, namely, from a proof of $\forall x \exists y A_0(x, y)$ (where A_0 is a quantifier free formula) in **PA** (let p denote that proof) by the D-interpretation one can obtain *in a very modular way* a term t and a proof of $\forall x A_0(x, t(x))$ in $\mathbf{HA}^\omega$. That was the seeds of Proof Mining.[5]

In [Kre58] Kreisel asked for a shift of emphasis in proof-theoretic research and announced his program called 'unwinding of proofs' which he described as,

'There is a different general program which does not seem to suffer the defects of [Hilbert's] consistency program: To determine the constructive (recursive) content or the constructive equivalent of the non-constructive concepts and theorems used in mathematics, particularly in arithmetic and analysis."

As a tool to accomplish his program Kreisel ([Kre51], [Kre52]) had developed an interpretation called *no counterexample interpretation* (see section 2.5.2). In order

[3] This is the so-called 'negative translation' (or 'N-translation') which was independently discovered by Kolmogorov, Gödel and Gentzen. Due to this embedding of **PA** into **HA** intuitionistic arithmetic cannot be considered finitary since **PA** is not finitary. Bernays even defends that Gentzen $TI_{<\varepsilon_0}$ is less non-finitary than the whole body of intuitionistic methods in number theory ([Kle52], pg. 498).

[4] For more information on the history of Gödel's Functional Interpretation see [Fef93], [Göd90] and [AF98].

[5] The slogan 'Proof Mining' was recently suggested to Ulrich Kohlenbach by Dana Scott to designate the project which was previously called 'Unwinding of Proofs' by Georg Kreisel described below.

to obtain that interpretation, he extended the so-called ϵ-substitution (developed by Hilbert, Ackerman and Bernays) and afterwards used it to unwind Artin's proof of Hilbert's 17th problem (cf. [Del96] and [Luc96]).

In the following section we introduce various interpretations used as tools for Proof Mining and show how they are combined. In section 3 we present a case study [KO01] where the application of 'monotone functional interpretation' was used to extract new numerically relevant information, namely a uniform modulus of uniqueness from a non-constructive proof of the uniqueness of best L_1-approximation by polynomials of degree $\leq n$. The whole process of extracting the modulus is presented in the appendix. In section 5 we outline future work.

Due to the limit on the number of pages we assume knowledge of classical and intuitionistic logic, and of the arithmetic systems here used. The reader who is not familiar with first-order classical/intuitionistic logic (**PL**/**IL**), Peano/Heyting arithmetic (**PA**/**HA**), extensional Peano/Heyting arithmetic in all finite types (**E–PA**$^\omega$/**E–HA**$^\omega$) or weakly extensional Heyting arithmetic in all finite types (**WE–HA**$^\omega$) is referred to [Bus98], [Sho67] and [TS96].

As a matter of convention we use x, y, z for variables ranging over numbers, n, m denote natural numbers, $\overline{n}, \overline{m}$ the respective numerals and s, t range over terms. When talking about **E–HA**$^\omega$, **WE–HA**$^\omega$ and **E–PA**$^\omega$, $\mathcal{T}$ will denote the set of all finite types. We write x^σ and t^ρ to represent that x is a variable of type σ and t is a term of type ρ, where it is implicitly assumed that $\sigma, \rho \in \mathcal{T}$. Moreover, $\underline{x}$, $\underline{y}$ will stand for tuples of variables and $\underline{t}$ for tuples of terms. We use A, B and C for arbitrary formulas or sentences, A_0 for a quantifier free formulas and A_1 for purely existential formulas.

The predicate $T(x, y, z)$ defined by Kleene [Kle67] (see also [Sho67]) is true whenever the Turing machine M, encoded by x, on input y halts with computation z (i.e. z is the Gödel number of that computation). The predicate $T(x, y, z)$ is decidable (even primitive recursive) and the predicate $T'(x) := \exists y T(x, x, y)$ is undecidable (but recursively enumerable).

The material of the following section on Proof Mining is substantially based on [AF98], [BS95], [BSBar], [Bus95], [Koh98a], [Koh93a], [Tro73] and [Tv88].

2 Proof Mining

The general purpose of Proof Mining is to extract from a given proof of a formula A in a system $\mathcal{A}$ some constructive content. By 'constructive content' we normally mean a realizing term for the existential quantifiers of A. For instance, if $A := \forall x \exists y A_0(x, y)$ we want to read off from the proof of A a term t such that $A_0(x, t(x))$ can be proved in $\mathcal{A}$.

The system **HA** is designed in such way that, whenever $\exists x A(x)$ is proved (for arbitrary sentences $\exists x A(x)$) then there exists a numeral $\overline{n}$ such that $\mathbf{HA} \vdash A(\overline{n})$. That is the so-called *numerical existence property* (NEP). Therefore, assuming $\mathbf{HA} \vdash \forall x \exists y A(x, y)$, for any given numeral $\overline{n}$ we know by NEP that there exists an $\overline{m}$ s.t. $\mathbf{HA} \vdash A(\overline{n}, \overline{m})$ (where the $\overline{m}$ can be found by unbounded search). Notice that the algorithm for finding $\overline{m}$ given $\overline{n}$ is performed at the meta-level, i.e., there is no term $t \in \mathcal{L}(\mathbf{HA})$ such that $\mathbf{HA} \vdash \forall x A(x, tx)$ since only primitive recursive functions are **HA**-definable.

On the other hand, in the system $\mathbf{PA} := \mathbf{HA} + LEM$ the NEP does not hold in the general setting. A simple example of this fact involves the use of the undecidable predicate $T'(x)$ defined above. Using LEM one can prove in **PA** the sentence $T'(\overline{n}) \vee \neg T'(\overline{n})$, for any fixed numerals $\overline{n}$, i.e. $\mathbf{PA} \vdash \exists x (T(\overline{n}, \overline{n}, x) \vee \forall z \neg T(\overline{n}, \overline{n}, z))$. Suppose we were guaranteed the existence of a numeral $\overline{m}$ such that $\mathbf{PA} \vdash T(\overline{n}, \overline{n}, \overline{m}) \vee \forall z \neg T(\overline{n}, \overline{n}, z)$. Since T is decidable, we would have a general proceeding for solving the halting problem $W := \{x : \exists y T(x, x, y)\}$ in the following way: for a given number n again by unbounded search we find an m such that $\mathbf{PA} \vdash T(\overline{n}, \overline{n}, \overline{m}) \vee \forall z \neg T(\overline{n}, \overline{n}, z)$; then we output $n \in W$ if $\mathbf{PA} \vdash T(\overline{n}, \overline{n}, \overline{m})$ and $n \notin W$ otherwise. Since T is decidable we have that W is a recursive set, a contradiction. That means we are not guaranteed a priory to be able to extract realizing terms for existential sentences A that have been proved in **PA**. In this case it is interesting to investigate which properties a formula A should have so that whenever A is proved in **PA** we are certain to obtain from that proof terms realizing the existential quantifiers of some prenex form of A. Even when no terms can be extracted from the proof of A there are still some possibilities, for instance,[6]

a) look for a variant A' of A (for instance a weakening, i.e. $\mathbf{PA} \vdash A \to A'$, or a translation) such that realizing terms for A' can be extracted,

b) instead of presenting the realizing term we can just extract a set of terms such that one of them is a realizing term for A.

In general, we want to obtain theorems which, given that A has some structure and it is proved in some system **S**, guarantees that realizing terms for A can be extracted. That is normally achieved via a detour, i.e., by first obtaining a proof of a variant A' in some other system **S'**. If **S'** is a system like **HA** realizing terms for A' are easily extractable and hopefully those terms can provide information about the realizing terms of the original A. Therefore, the following sections contain rules of

[6]This same discussion can be extended to stronger systems for analysis (e.g. $\mathbf{WE\text{–}PA}^\omega + AC + WKL$) as it will be seen in section 2.5.1.

the form,

$$\mathbf{S} \vdash A \qquad \Longrightarrow \qquad \mathbf{S'} \vdash A',$$

where in general **S'** is either equal to **S** or is an intuitionistic variant of **S** and A' is a translation of A. Whenever the rule is constructive (in the sense that given a proof p of the sentence A in the system **S** there exists an algorithm for transforming p into a proof of A' in **S'**) we write,

$$\mathbf{S} \vdash A \qquad \stackrel{alg}{\Longrightarrow} \qquad \mathbf{S'} \vdash A'.$$

Note that in all the *soundness theorems* (also called metatheorems) presented below the rule is constructive, and moreover, the algorithm is given by the proof of those metatheorems.

Other applications of the techniques of Proof Mining include underivability and closure results. We give examples of these applications in sections 2.2 and 2.4. For an interesting discussion about constructivism and proof translations see [vD00].

2.1 H-translation (Herbrand disjunction)

If we remove the non-logical axioms from **PA** we get the classical first-order logic **PL** (also called predicate logic). We can show an example where a) and b) from the previous section can be applied.

Definition 2.1 *For each formula* $A := (\forall y_0)\exists x_1 \forall y_1 \ldots \exists x_n \forall y_n A_0(y_0, x_1, y_1, \ldots, x_n, y_n)$ *in the language of* ***PL*** *(where* A_0 *is quantifier free) we define its Herbrand normal form by,*

$$A^H := \exists x_1 \ldots \exists x_n A_0(f_0, x_1, f_1(x_1), \ldots, x_n, f_n(x_1, \ldots, x_n)),$$

where $f_0, \ldots, f_n$ *are new function symbols not present in* A.

Note that this translation is a weakening since $\mathbf{PL} \vdash A \to A^H$. For the sake of legibility we state Herbrand's theorem for Σ^0_3 formulas only.[7]

Theorem 2.2 (Soundness of H-translation) *Let* $A := \exists x \forall y \exists z A_0(x, y, z)$ *be a formula in the language of* ***PL****, where* A_0 *is quantifier free. The following rule holds,*

$$\textbf{\textit{PL}} \vdash A \qquad \stackrel{alg}{\Longrightarrow} \qquad \textbf{\textit{PL}} \vdash \bigvee_{i=1}^{n} \bigvee_{j=1}^{m} A_0(t_i, f_0(t_i), s_j),$$

where suitable terms $t_1, \ldots, t_n$ *and* $s_1, \ldots, s_m$ *(built only from the constants, function symbols and variables of* A^H*) can be extracted from the proof of* A.

We call $A^{H,D} := \bigvee_{i=1}^{n} \bigvee_{j=1}^{m} A_0(t_i, f_0(t_i), s_j)$ a Herbrand disjunction of A.

Example: Take the sentence used above, $A := \forall x(\exists y T(x, x, y) \vee \neg \exists z T(x, x, z))$. Let us consider the following prenex form of A,

$$A_p := \forall x \exists y \forall z (T(x, x, y) \vee \neg T(x, x, z)).$$

[7] For a general statement of the theorem see [Koh98a] and for a detailed exposition about Herbrand's work see [Bus95].

The Herbrand normal form of A_p would be, $A_p^H := \exists y(T(f_0, f_0, y) \vee \neg T(f_0, f_0, f_1(y)))$. Since A_p can be proved in **PL** we are guaranteed by the Soundness of H-translation that terms $t_1, \ldots, t_n$ can be extracted from the proof $\mathbf{PL} \vdash A_p$ such that,

$$A_p^{H,D} := \bigvee_{i=1}^{n} T(f_0, f_0, t_i) \vee \neg T(f_0, f_0, f_1(t_i))$$

can be proved in **PL**. In fact, taking the terms $t_1 = f_0$ and $t_2 = f_1(f_0)$ we have,

$$\mathbf{PL} \vdash T(f_0, f_0, f_0) \vee \neg T(f_0, f_0, f_1(f_0)) \vee T(f_0, f_0, f_1(f_0)) \vee \neg T(f_0, f_0, f_1(f_1(f_0))).$$

Remark 2.3 *Herbrand's theorem can also be extended to open theories, i.e. theories where all non-logical axioms are purely universal. As an application of Herbrand's theorem to Proof Mining we mention Luckhardt's [Luc89] extraction of polynomial bounds from a proof of Roth's theorem.*

2.2 Modified realizability

According to the Brouwer-Heyting-Kolmogorov (BHK) constructive interpretation of logic a proof of A is defined as follows (to mention just some clauses):

- p proves $A \vee B$ iff $p = (p_0, p_1)$ and $p_0 = 0$ implies p_1 proves A and $p_0 = 1$ implies p_1 proves B,
- p proves $A \to B$ iff p is a construction transforming any proof of A into a proof of B,
- p proves $\exists x\, A(x)$ iff $p = (p_0, p_1)$ and p_1 proves $A(p_0)$,

where for prime formulas the notion of proof is supposed to be given. Based on the BHK interpretation Kleene [Kle45] gave a semantics for intuitionistic arithmetic associating to each arithmetical sentence A a notion 'the number n realizes A'. That turned out to be a systematic method of making the constructive content of arithmetical sentences explicit. We present here a version of Kleene's realizability for $\mathbf{E\text{–}HA}^\omega$, called *modified realizability*, which was first formulated by Kreisel [Kre62]. In the following $\underline{x}$ stands for a tuple (possibly empty) which is built from variables in the language of $\mathbf{E\text{–}HA}^\omega$ and tupling operations.

Definition 2.4 (Modified realizability) *With each formula $A \in \mathcal{L}(\mathbf{E\text{–}HA}^\omega)$ associate a formula $\underline{x}$ mr A (read: $\underline{x}$ modified realizes A) of $\mathbf{E\text{–}HA}^\omega$ by induction on the logical structure of A:*

- $\underline{x}\ mr\ A := A$, *if A is a prime formula, i.e. the empty tuple realizes A,*
- $\underline{x}, \underline{y}\ mr\ (A \wedge B) := \underline{x}\ mr\ A \wedge \underline{y}\ mr\ B$,
- $z^0, \underline{x}, \underline{y}\ mr\ (A \vee B) := (z = 0 \to \underline{x}\ mr\ A) \wedge (z \neq 0 \to \underline{y}\ mr\ B)$,
- $\underline{y}\ mr\ (A \to B) := \forall \underline{x}\, (\underline{x}\ mr\ A \to \underline{y}\underline{x}\ mr\ B)$,[8]

[8] The negation of a formula A is understood as an abbreviation for $A \to 0 = 1$, which means that $\underline{y}\ mr\ \neg A := \forall \underline{x}(\underline{x}\ mr\ A \to y(\underline{x})\ mr\ (0 = 1))$, i.e. $\underline{y}\ mr\ \neg A := \forall \underline{x} \neg(\underline{x}\ mr\ A)$.

- $\underline{x}\ mr\ (\forall y^{\rho} A(y)) := \forall y^{\rho}(\underline{x}(y)\ mr\ A(y))$,
- $z^{\rho}, \underline{x}\ mr\ (\exists y^{\rho} A(y)) := \underline{x}\ mr\ A(z)$.

Note that if $B := \underline{x}\ mr\ A$ then the free variables of B are contained in $FV(\underline{x}) \cup FV(A)$. Moreover, the length and the type of the tuple $\underline{x}$ is determined by the logical structure of A.

Theorem 2.5 (Soundness of mr**)** *Let $A \in \mathcal{L}(\textbf{E–HA}^{\omega})$. Then,*

$$\textbf{E–HA}^{\omega} \vdash A \quad \overset{alg}{\Longrightarrow} \quad \textbf{E–HA}^{\omega} \vdash \underline{t}\ mr\ A,$$

where $\underline{t} \in \mathcal{L}(\textbf{E-HA}^{\omega})$ is a tuple of terms which can be extract from the given proof of A.

Example: As promised above we use modified realizability to show an underivability result. Let $A := \forall x(\exists y T(x,x,y) \vee \neg\exists z T(x,x,z))$. By the modified realizability translation A is associated with the formula,

$$\forall x((zx)_0 = 0 \rightarrow T(x,x,(zx)_1) \wedge ((zx)_0 \neq 0 \rightarrow \forall y \neg T(x,x,y))),$$

where z^1 is the only free variable and zx is a pair $((zx)_0,(zx)_1)$. Now, suppose

$$\textbf{E–HA}^{\omega} \vdash \forall x(\exists y T(x,x,y) \vee \neg\exists z T(x,x,z)).$$

By the Soundness of mr there exists a term t^1 such that for any given number n the term $t(n)$ is a pair (m, t_1), such that,

$$(m = 0 \rightarrow T(n,n,t_1(n))) \wedge (m \neq 0 \rightarrow \forall y \neg T(n,n,y)),$$

where m is a number. That cannot happen since such t would solve the halting problem, which means that $\forall x(T'(x) \vee \neg T'(x))$ is not provable in $\textbf{E–HA}^{\omega}$. $\dashv$

In order to obtain a complete characterization of mr we need to extend the system $\textbf{E–HA}^{\omega}$. The axiom of choice for types σ and τ (where $A \in \mathcal{L}(\textbf{E–HA}^{\omega})$) is defined as,

$$AC^{\sigma,\tau}\ :\ \forall x^{\sigma}\exists y^{\tau} A(x,y) \rightarrow \exists Y^{\sigma\rightarrow\tau}\forall x A(x, Y(x))$$

and the independence of premise for type ρ (for $A \in \mathcal{L}(\textbf{E–HA}^{\omega})$ and $\exists$-free[9]) as,

$$IP^{\rho}_{ef}\ :\ (A \rightarrow \exists x^{\rho} B(x)) \rightarrow \exists x^{\rho}(A \rightarrow B(x)).$$

We then defined two axiom schemata: the schema of choice $AC := \bigcup_{\sigma,\tau\in\mathcal{T}}\{AC^{\sigma,\tau}\}$ and the independence of premise schema $IP_{ef} := \bigcup_{\rho\in\mathcal{T}}\{IP^{\rho}_{ef}\}$.

Theorem 2.6 (Characterization of mr**)** *Let $A(\underline{a}) \in \mathcal{L}(\textbf{E–HA}^{\omega})$, where $\underline{a}$ is the set of free variables in A. Then,*

- $\textbf{E–HA}^{\omega} + AC + IP_{ef} \vdash A(\underline{a}) \leftrightarrow \exists \underline{x}\ (\underline{x}\ mr\ A(\underline{a}))$,
- $\textbf{E–HA}^{\omega} + AC + IP_{ef} \vdash A(\underline{a}) \implies \textbf{E–HA}^{\omega} \vdash \underline{t}(\underline{a})\ mr\ A(\underline{a})$,

  for some $\underline{t} \in \mathcal{L}(\textbf{E–HA}^{\omega})$ with $FV(\underline{t}) \subseteq FV(A)$.

An immediate consequence of the theorem above is that formulas in AC and IP_{ef} (although not provable in $\textbf{E–HA}^{\omega}$) have realizing terms in $\textbf{E–HA}^{\omega}$.

More information on realizability can be found in [Kle65], [Kle52], [Tv88], [Tro98] and [Koh98a].

[9] A formula $A \in \mathcal{L}(\textbf{E–HA}^{\omega})$ is called $\exists$-free if built from prime formulas by the use of $\wedge, \rightarrow$ and $\forall$ only.

2.3 N-translation (Negative translation)

The N-translation is a way of interpreting formulas of classical logic in intuitionistic logic. The translation was discovered independently by Kolmogorov, Gödel's and Gentzen and can be easily seen to apply to many classical systems as e.g. **PA** and **WE–PA**$^{\omega}$ (with their usual intuitionistic counterparts). We present below a version of the N-translation due to Kuroda,

Definition 2.7 ([Kur51]) *Let $A \in \mathcal{L}(\boldsymbol{PA})$. The negative translation of A is defined as $(A)^N := \neg\neg(A)^*$, where A^* is defined by induction on the logical structure of A:*

- $(A)^* := A$, *if A is a prime formula,*
- $(A \Box B)^* := A^* \Box B^*$, *where $\Box \in \{\wedge, \vee, \rightarrow\}$,*
- $(\exists x A(x))^* := \exists x (A(x))^*$,
- $(\forall x A(x))^* := \forall x \neg\neg (A(x))^*$.

Theorem 2.8 (Soundness of N-translation) *Let A be a formula in the $\mathcal{L}(\boldsymbol{PA}) = \mathcal{L}(\boldsymbol{HA})$. The following rule holds,*

$$\boldsymbol{PA} \vdash A \quad \overset{alg}{\Longrightarrow} \quad \boldsymbol{HA} \vdash (A)^N.$$

The theorem above also holds if one replaces **PA** with **WE–PA**$^{\omega}$ and **HA** with **WE–HA**$^{\omega}$, for instance.

A simple consequence of the Soundness of N-translation is that the consistency of **WE–HA**$^{\omega}$ implies the consistency of **WE–PA**$^{\omega}$, for if **WE–PA**$^{\omega}$ is inconsistent, then **WE–PA**$^{\omega}$ $\vdash 0 = 1$ and by the Soundness of N-translation, **WE–HA**$^{\omega}$ $\vdash (0 = 1)^N$, which implies that **WE–HA**$^{\omega}$ is inconsistent.

The following formulas are derivable in **HA/WE–HA**$^{\omega}$ and are very useful when dealing with the negative translation. For arbitrary formulas A, B:

i) $\neg\neg\forall x A \rightarrow \forall x \neg\neg A$,

ii) $\neg A \leftrightarrow \neg\neg\neg A$,

iii) $\neg\neg(A \rightarrow B) \leftrightarrow (A \rightarrow \neg\neg B)$,

iv) $\neg\neg(A \rightarrow B) \leftrightarrow (\neg\neg A \rightarrow \neg\neg B)$.

Example: Once again we use the sentence $A := \forall x \exists y \forall z (T(x,x,y) \vee \neg T(x,x,z))$ to illustrate the N-translation. Since **PA** $\vdash A$, by the Soundness of the N-translation, we have,

$$\mathbf{HA} \vdash \neg\neg\forall x \neg\neg\exists y \forall z \neg\neg(T(x,x,y) \vee \neg T(x,x,z)).$$

By i) we can shift double negations inwards over $\forall$-quantifiers,

$$\mathbf{HA} \vdash \forall x \neg\neg\neg\neg\exists y \forall z \neg\neg(T(x,x,y) \vee \neg T(x,x,z)),$$

and by ii) we replace three negations by a single one,

$$\mathbf{HA} \vdash \forall x \neg\neg\exists y \forall z \neg\neg(T(x,x,y) \vee \neg T(x,x,z)),$$

and since for prime formulas P we have, $\neg\neg P \leftrightarrow P$, we get,[10]

$$\mathbf{HA} \vdash \forall x \neg\neg\exists y \forall z (T(x,x,y) \vee \neg T(x,x,z)).$$

2.4 A-translation

Via negative translation one obtains from a proof of $\exists y A_0(\underline{x}, y)$ in **PA** a proof of $\neg\neg\exists y A_0(\underline{x}, y)$ in **HA**, where $\underline{x}$ are free variables, i.e.

$$\mathbf{PA} \vdash \exists y A_0(\underline{x}, y) \qquad \overset{alg}{\Longrightarrow} \qquad \mathbf{HA} \vdash \neg\neg\exists y A_0(\underline{x}, y).$$

Since **HA** is closed under the Markov rule,

$$\mathbf{HA} \vdash \neg\neg\exists y A_0(\underline{x}, y) \qquad \Longrightarrow \qquad \mathbf{HA} \vdash \exists y A_0(\underline{x}, y),$$

A_0 quantifier free, we know that **PA** is Π^0_2-conservative over **HA**, i.e.,

$$\mathbf{PA} \vdash \forall x \exists y A_0(\underline{x}, y) \qquad \Longrightarrow \qquad \mathbf{HA} \vdash \forall x \exists y A_0(\underline{x}, y).$$

We still need, however, an *algorithm* for transforming a proof $\mathbf{HA} \vdash \neg\neg\exists y A_0(\underline{x}, y)$ into a proof $\mathbf{HA} \vdash \exists y A_0(\underline{x}, y)$. The A-translation[11] is a simple technique for performing such translation which applies to many intuitionistic theories.

Definition 2.9 ([Fri78]) *Let $A \in \mathcal{L}(\boldsymbol{HA})$. With each formula $F \in \mathcal{L}(\boldsymbol{HA})$ (such that the free variables of A are not bounded in F) associate a formula $(F)^A \in \mathcal{L}(\boldsymbol{HA})$, called the A-translation of F, in the following way: $(F)^A$ results when all prime formulas P in F are replaced by $P \vee A$.*

Theorem 2.10 (Soundness of A-translation) *Let A and F be formulas in $\mathcal{L}(\boldsymbol{HA})$ such that the free variables of A are not bounded in F. The following rule holds,*

$$\boldsymbol{HA} \vdash F \qquad \overset{alg}{\Longrightarrow} \qquad \boldsymbol{HA} \vdash (F)^A.$$

As a corollary of the Soundness of A-translation we obtain an effective closure of **HA** under Markov rule,

Corollary 2.11 *The following rule holds,*

$$\boldsymbol{HA} \vdash \forall x \neg\neg\exists y A_0(x, y) \qquad \overset{alg}{\Longrightarrow} \qquad \boldsymbol{HA} \vdash \forall x \exists y A_0(x, y),$$

for A_0 quantifier free.

[10] In arithmetical systems like **HA** any quantifier free formulas $A_0(x)$ is equivalent to a prime formula $t(x) = 0$, for a suitable term t.

[11] The A-translation was developed by H. Friedman [Fri78] and independently investigated by Dragalin [Dra79]. Variants of the translation were considered in [Lei85] and [Tv88]

Proof: In **HA** any quantifier free formula $A_0(x,y)$ can be written as a prime formula $P(x,y)$ (see footnote 10). In this way $\mathbf{HA} \vdash \forall x \neg\neg\exists y A_0(x,y)$ implies,

$$\mathbf{HA} \vdash ((\exists y P(x,y) \to \perp) \to \perp).$$

By the Soundness of A-translation (for $A := \exists y P(x,y)$) we have,

$$\mathbf{HA} \vdash ((\exists y(P(x,y) \vee \exists y P(x,y)) \to \exists y P(x,y)) \to \exists y P(x,y)),$$

and hence, $\mathbf{HA} \vdash (\exists y P(x,y) \vee \exists y P(x,y) \to \exists y P(x,y)) \to \exists y P(x,y)$. Since $\mathbf{HA} \vdash A \vee A \to A$ for arbitrary A we get that $\mathbf{HA} \vdash \exists y P(x,y)$, which implies, $\mathbf{HA} \vdash \forall x \exists y A_0(x,y)$.

2.4.1 N-translation + A-translation + Modified Realizability

Assume we have proved in **PA** the statement $A := \forall x \exists y A_0(x,y)$, where A_0 is quantifier free. From that proof, by the soundness of N-translation, we obtain a proof of $(A)^N := \forall x \neg\neg\exists y A_0(x,y)$ in **HA**. Using the A-translation, as discussed in the previous section, we obtain a proof of $\forall x \exists y A_0(x,y)$ in **HA**. Finally, by the soundness of modified realizability, we can extract from that proof a term t such that

$$\mathbf{E\text{–}HA}^\omega \vdash t\ mr\ \forall x \exists y A_0(x,y),$$

which means $\mathbf{E\text{–}HA}^\omega \vdash \forall x A_0(x, t(x))$. Notice that we could not apply modified realizability directly to $\forall x \neg\neg\exists y A_0(x,y)$ (without using A-translation first) since the empty tuple realizes $\forall x \neg\neg\exists y A_0(x,y)$. The combination of these three techniques however provides a powerful tool for extracting realizing terms from classical proofs of Π^0_2-formulas. Some drawbacks include,

- A-translation does not seem to be sound for the quantifier free rule of extensionality which would imply that it cannot be applied, e.g., to the system $\mathbf{WE\text{–}HA}^\omega$.
- A-translation does not establish conservation results for the Markov principle for arbitrary formulas (see [CH99]).

The important achievement of this approach however has been the automatization of the whole process by Berger and Schwichtenberg with the *Minlog system*. The system receives as input a completely formalized proof in **PA** of an existential statement A and generates as output realizing terms for A through an iterative process. Due to the high complexity of completely formalizing proofs, however, just simple examples have been treated so far.

Further information on this approach to Proof Mining can be found in [Mur90], and most recently in [BS95] and [BSBar] where a refined combination of negative translation and A-translation is described in order to reduce the complexity of the extracted term. In [BSS01] some proofs in arithmetic have been analyzed. Another application of realizability for automated extraction of programs from constructive proofs is presented in [HN88].

2.5 D-interpretation (Functional interpretation)

As mentioned in the introduction, the D-interpretation was developed by Gödel with the purpose of providing a less 'non-finitary' proof for the consistency of **PA**. After accomplishing its original goal it has been used for a variety of other applications such as establishing conservation results and Proof Mining. We focus here on the latter aspect.

Definition 2.12 ([Göd58]) *With each formula $A \in \mathcal{L}(\textbf{WE–HA}^\omega)$ associate a formula $(A)^D := \exists \underline{x} \forall \underline{y} A_D(\underline{x}, \underline{y})$ in $\mathcal{L}(\textbf{WE–HA}^\omega)$, where A_D is quantifier free and $FV(A_D) = FV(A)$, defined inductively in the following way:*

1. $(A)^D := (A)_D := A$, *for prime formulas A,*

and if $(A)^D := \exists \underline{x} \forall \underline{y} A_D(\underline{x}, \underline{y})$ and $(B)^D := \exists \underline{u} \forall \underline{v} B_D(\underline{u}, \underline{v})$ we have,

2. $(A \wedge B)^D := \exists \underline{x}, \underline{u} \forall \underline{y}, \underline{v}\ (A_D(\underline{x}, \underline{y}) \wedge B_D(\underline{u}, \underline{v}))$,
3. $(A \vee B)^D := \exists z^0, \underline{x}, \underline{u} \forall \underline{y}, \underline{v}\ ((z = 0 \to A_D(\underline{x}, \underline{y})) \wedge (z \neq 0 \to B_D(\underline{u}, \underline{v})))$,
4. $(A \to B)^D := \exists \underline{U}, \underline{Y} \forall \underline{x}, \underline{v}\ (A_D(\underline{x}, \underline{Y}\, \underline{x}\, \underline{v}) \to B_D(\underline{U}\, \underline{x}, \underline{v}))$,
5. $(\exists z^\rho\ A(z))^D := \exists z, \underline{x} \forall \underline{y}\ A_D(\underline{x}, \underline{y}, z)$,
6. $(\forall z^\rho\ A(z))^D := \exists \underline{X} \forall z, \underline{y}\ A_D(\underline{X} z, \underline{y}, z)$.

Example: Here we point out one of the differences between D-interpretation and modified realizability. Let us consider the formula $C := \forall x^0 A_0(x) \to \forall y^0 B_0(y)$, where for simplicity A_0 and B_0 are quantifier free. The functional interpretation of C is $(C)^D := \exists X^1 \forall y (A_0(Xy) \to B_0(y))$, i.e., D-interpretation asks for a function X which for any given y makes $A_0(Xy) \to B_0(y)$ true. On the other hand, the modified realizability translation of C is realized by the empty tuple.

Theorem 2.13 (Soundness of D-interpretation) *Let $A(z) \in \mathcal{L}(\textbf{WE–HA}^\omega)$, where z is free. The following rule holds,*

$$\textbf{WE–HA}^\omega \vdash A(z) \quad \stackrel{alg}{\Longrightarrow} \quad \textbf{WE–HA}^\omega \vdash \forall \underline{y} A_D(\underline{t} z, \underline{y}, z),$$

where $\underline{t}$ is a tuple of closed terms in the language of $\textbf{WE–HA}^\omega$ which can be extract from the given proof of $A(z)$.

Markov principle for type σ is defined as,

$$MP^\sigma\ :\ \neg\neg\exists x^\sigma A_0(x) \to \exists x A_0(x),$$

where A_0 is quantifier free. The schema of Markov principle is $MP := \bigcup_{\sigma \in \mathcal{T}} \{MP^\sigma\}$. Note that the functional interpretation of MP^σ,

$$\begin{aligned} (MP^\sigma)^D &= (\neg\neg\exists x^\sigma A_0(x) \to \exists y^\sigma A_0(y))^D, \\ &= \exists Y \forall x (\neg\neg A_0(x) \to A_0(Yx)), \end{aligned}$$

is realized by $Y := \lambda x.x$. Moreover, the AC (mentioned in section 2.2) and the independence of premise for purely universal formulas,

$$IP_\forall^\rho \; : \; (\forall y A_0(y) \to \exists x^\rho B(x)) \to \exists x^\rho (\forall y A_0(y) \to B(x)),$$

where A_0 is quantifier free, also have realizing functionals. Let $IP_\forall := \bigcup_{\rho \in \mathcal{T}} \{IP_\forall^\rho\}$. In this way we can extend the Soundness theorem of D-interpretation to the system **WE–HA**$^\omega + AC + IP_\forall + MP$. The following theorem says that this extension of **WE–HA**$^\omega$ is enough to characterize the D-interpretation:

Theorem 2.14 (Characterization of D-interpretation) *Let $A \in \mathcal{L}($**WE–HA**$^\omega)$. Then,*

$$\textbf{\textit{WE–HA}}^\omega + AC + IP_\forall + MP \vdash A \leftrightarrow (A)^D.$$

Functional interpretation has been extended to a number of other theories, e.g. systems of feasible arithmetic ([CU93]), and used as a tool for extracting programs from classical proofs (see [Koh90]). For a comprehensive treatment of functional interpretation see [AF98].

2.5.1 N-translation + Monotone D-interpretation

A modification of functional interpretation where instead of looking for realizing terms one searches for 'upper bounds' on these terms was defined in [Koh96]. When combined with the negative translation, monotone functional interpretation can be used to translate proofs in **WE–PA**$^\omega$ into proofs in **WE–HA**$^\omega$, in the same manner as the standard functional interpretation. The reason to consider the monotone version is that dependencies in the extracted term can be eliminated and in many cases (due to the monotonicity) the upper bound can be used as a realizing term. To illustrate this we consider the following simple example. Suppose that functional interpretation extracts from a proof of $\forall x, y \exists z A_0(x, y, z)$ a functional $\Phi(x, y)$ bounding the variable z, i.e.,

$$\forall x, y \exists z \leq \Phi(x, y) A_0(x, y, z)$$

If y happens to be a variable ranging over the real numbers in the interval $[0, 1]$ then (given that Φ is monotone on the second argument) the functional $\Psi(x) := \Phi(x, 1) \geq \Phi(x, y)$ can also be taken as a bound on z. Moreover, given that the formula $A_0(x, y, z)$ is monotone in z (in the sense that $A_0(x, y, z_1) \wedge z_1 \leq z_2 \to A_0(x, y, z_2)$) then $\Psi(x)$ realizes z. If we use that same reasoning through the whole proof (and not only after obtaining the final term) we obtain monotone functional interpretation.

A number of metatheorems have been proved in [Koh93a] of the form,

Theorem 2.15 ([Koh93a], theorem 2.2(a)) *The following rule holds:*

$$\begin{cases} \textbf{\textit{WE–PA}}^\omega + AC_{qf}^{1,0} + WKL \vdash \forall n^0; x^1; y \leq_1 sx \; \exists k^0 \, A_1(n, x, y, k) \\ \textit{then one can extract a } \textbf{\textit{WE–PA}}^\omega\textit{-definable functional } \Phi \textit{ s.t.} \\ \textbf{\textit{WE–HA}}^\omega \vdash \forall n^0; x^1; y \leq_1 sx \; \exists k^0 \leq \Phi(n, x) \, A_1(n, x, y, k). \end{cases}$$

where A_1 is purely existential and s is some closed term in the language of **WE–PA**$^\omega$*.*

In order to apply such metatheorems to proofs in Analysis where the objects range over real numbers, polynomials, continuous functions and Polish spaces in general, we have to fix the representation of these objects. As it happens, all these objects when suitably represented can be viewed as a number theoretic function (i.e. function between natural numbers) and the theorem above applies. We can even restate the theorem in a more applied perspective,

Theorem 2.16 ([Koh93a], theorem 4.1) *Let X, K be **WE–PA**$^\omega$-definable Polish spaces, K compact and consider a sentence which can be written (when formalized in the language of **WE–PA**$^\omega$) in the form*

$$A := \forall n \in \mathbb{N}; x \in X; y \in K\ \exists k \in \mathbb{N}\, A_1(n, x, y, k),$$

where A_1 is a purely existential. Then the following rule holds:

$$\begin{cases} \textbf{\textit{WE–PA}}^\omega + AC^{1,0}_{qf} + WKL \vdash \forall n \in \mathbb{N}; x \in X; y \in K\ \exists k \in \mathbb{N}\, A_1(n, x, y, k) \\ \textit{then one can extract a } \textbf{\textit{WE–PA}}^\omega\textit{-definable functional } \Phi \textit{ s.t.} \\ \textbf{\textit{WE–HA}}^\omega \vdash \forall n \in \mathbb{N}; x \in X; y \in K \exists k \leq \Phi(n, x)\, A_1(n, x, y, k). \end{cases}$$

It is important to notice that the functional Φ depends on the representation of the Polish space X. For instance, in the case where X is the space of continuous functions in the interval $[0, 1]$, the elements $f : [0, 1] \to \mathbb{R}$ are represented as a pair (f_r, ω_f) of functions, where $f_r : \mathbb{Q} \to \mathbb{Q}$ is the restriction of f to the rational numbers and $\omega_f : \mathbb{N} \to \mathbb{N}$ is some uniform modulus of continuity of f, i.e.

$$\forall x, y \in [0, 1]; k \in \mathbb{N}(|x - y| < 2^{-\omega_f(k)} \to |f(x) - f(y)| < 2^{-k}).$$

For applications of monotone functional interpretation to Approximation Theory and Fixed Point theory see [Koh93b], [Kohar], [Kohb] and section 3 of this report. [Koh96] describes the general program of applying monotone functional interpretation to Analysis.

2.5.2 No counterexample interpretation

Suppose **PA** proves the following formula,

$$A := \forall y_0 \exists x_1, \forall y_1 \ldots \exists x_n \forall y_n A_0(y_0, x_1, y_1, \ldots, x_n, y_n),$$

where A_0 is quantifier free. By the Soundness of N-translation **HA** proves,

$$(A)^N := \neg\neg\forall y_0 \neg\neg\exists x_1 \forall y_1 \ldots \exists x_n \forall y_n \neg\neg A_0(y_0, x_1, y_1, \ldots, x_n, y_n),$$

which intuitionistically implies,

$$\neg\exists y_0 \forall x_1 \ldots \exists y_n \forall x_n \neg A_0(y_0, x_1, y_1, \ldots, x_n, y_n).$$

Moreover, the above formula is (relative to the D-interpretation) equivalent to,

$$(*)\ \forall f_0, \ldots, f_n \exists x_1, \ldots, x_n \neg\neg A_0(f_0, x_1, f_1(x_1), \ldots, x_n, f_n(x_1, \ldots, x_n)).$$

By the Soundness of D-interpretation there exist terms,

$$t_1 := F_1(f_0, \ldots, f_n) \quad \ldots \quad t_n := F_n(f_0, \ldots, f_n)$$

such that, $\mathbf{HA}^\omega \vdash A_0(f_0, t_1, f_1(t_1), \ldots, t_n, f_n(t_1, \ldots, t_n))$. Kreisel noticed that the functionals $F_1, \ldots, F_n$ can be taken to be recursive in their arguments. The formula $(*)$ above came to be called the *no counterexample interpretation* of A in the sense that whenever $f_1, \ldots, f_n$ are given as an attempt to falsify $(*)$ the functionals $F_1, \ldots, F_n$ should provide witnesses for the $y_1, \ldots, y_n$ blocking the counterexample.

Besides the explicit connection with functional interpretation seen above, the no counterexample interpretation is quite related to Herbrand's disjunction as it can be seen in the following example.

Example: In section 2.1 we saw that,

$$\mathbf{PA} \vdash \forall x \exists y \forall z (T(x, x, y) \vee \neg T(x, x, z)).$$

By the no counterexample interpretation we have, there exists a term $t := F(f_1, f_2)$ such that, $\mathbf{HA}^\omega \vdash T(f_0, f_0, t) \vee \neg T(f_0, f_0, f_1(t))$. By taking,

$$F(f_0, f_1) := \begin{cases} f_0 & \neg T(f_0, f_0, f_1(f_0)) \\ f_1(f_0) & otherwise, \end{cases}$$

we see how the choice of terms in the Herbrand disjunction can be used to form the functionals F for the no counterexample interpretation via the use of definition by cases.

Remark 2.17 *Note that the no counterexample interpretation is not exactly a tool for Proof Mining since it does not give a procedure for building the terms $t_1, \ldots, t_n$. The no counterexample interpretation rather tells which information should be obtained so that using some other technique (as ϵ-substitution in the case of Kreisel) we can perform the actual extraction. Moreover, it has been shown in [Koh99] that (contrary to the functional interpretation) the no counterexample interpretation behaves poorly with respect to modus pones.*

3 Case study: Uniqueness for L_1-approximation

The content of this section is based on the article [KO01] which for convenience can be found in the appendix of this report. The article (a joint work with U. Kohlenbach) describes in full details how the first effective uniform modulus of uniqueness for the L_1-approximation was extracted from a classical proof of the uniqueness statement. Here we present the motivations for the analysis and the results obtained. First we introduce some basic notions.

For any given continuous function f in the interval $[0,1]$ (we write $f \in C[0,1]$) the L_1-*norm of* f is defined as,

$$||f||_1 := \int_0^1 |f(x)|\, dx.$$

The space of polynomials up to degree n (including n) is denoted by P_n. Given a function $f \in C[0,1]$ we define the L_1-*distance of* f from P_n by,

$$dist_1(f, P_n) := \inf_{p \in P_n} \|f - p\|_1.$$

L_1-approximation (or 'approximation in the mean') has been extensively studied in numerical mathematics since 1859 when was first considered by Chebycheff (see [Pin89] for a comprehensive survey). In 1921, Jackson [Jac21] proved that for any given function $f \in C[0,1]$ and for a fixed $n \in \mathbb{N}$ there exists a unique polynomial $p_n \in P_n$ such that

$$\|f - p_n\|_1 = dist_1(f, P_n).$$

In 1966, Cheney [Che66] simplified (in the sense that he eliminated the use of measure theory) and generalized Jackson's proof to arbitrary Haar subspaces of $C[0,1]$.[12] This is the proof we analyzed.

The uniqueness of L_1-approximation from the space P_n can be written as,

$$(*)\ \forall f \in C[0,1]; n \in \mathbb{N}; p_1, p_2 \in P_n(\bigwedge_{i=1}^{2} \|f - p_i\|_1 =_{\mathbb{R}} dist_1(f, P_n) \to \|p_1 - p_2\|_1 =_{\mathbb{R}} 0).$$

If $(*)$ has been proved in $\mathbf{WE\text{–}PA}^\omega$ by Soundness of N-translation $\mathbf{WE\text{–}HA}^\omega$ proves,[13]

$$\neg\neg\forall f \in C[0,1]; n \in \mathbb{N}; p_1, p_2 \in P_n \neg\neg(\bigwedge_{i=1}^{2}(\|f - p_i\|_1 =_{\mathbb{R}} dist_1(f, P_n) \to \|p_1 - p_2\|_1 =_{\mathbb{R}} 0),$$

where the double negations inside the real equalities $=_{\mathbb{R}}$ can be eliminated. The formula above is intuitionistically equivalent to $(*)$ itself (see i), ii) and iii) in section 2.3) and, therefore, $(*)$ can be proved in $\mathbf{WE\text{–}HA}^\omega$. Presenting the hidden quantifiers we have,

[12] P is a n-dimensional Haar subspace of $C[0,1]$ if 0 is the only element with n roots. Note that P_n is a Haar subspace of $C[0,1]$ of dimension $n+1$.

[13] It is worth noticing that, according to the representation of real numbers by Cauchy sequences with fixed rate of convergence, equality between real numbers is an abbreviation for a $\forall$-statement (see remark 3.2 in the appendix). Those we call *hidden quantifiers*.

$(**)\ \forall f \in C[0,1]; n \in \mathbb{N}; p_1, p_2 \in P_n$
$$(\bigwedge_{i=1}^{2} \forall d \in \mathbb{N}(\|f - p_i\|_1 - dist_1(f, P_n) \leq 2^{-d}) \to \forall k \in \mathbb{N}\ \|p_1 - p_2\|_1 < 2^{-k}),$$

which has D-interpretation,

$\exists \Phi \forall f \in C[0,1]; n \in \mathbb{N}; p_1, p_2 \in P_n; k \in \mathbb{N}$
$$(\bigwedge_{i=1}^{2} \|f - p_i\|_1 - dist_1(f, P_n) \leq 2^{-\Phi(f,n,p_1,p_2,k)} \to \|p_1 - p_2\|_1 < 2^{-k}),$$

where Φ is a functional of appropriate type. Note that (as already mentioned in the example given in section 2.5) the modified realizability translation of $(*)$, in contrast to that of D-interpretation, does not ask for any realizing terms.

The existence of a functional $\Phi(f, n, p_1, p_2, k)$ of the form $\Phi'(f, n) + k$ (i.e. independent of p_1 and p_2 and linear in k) is called strong uniqueness[14] and $\Phi'(f, n)$ is called the 'constant of strong uniqueness'. In [Koh93a] this notion was generalized to consider functionals with arbitrary k dependency, i.e. $\Phi(f, n, k)$. Those Φ were named *uniform modulus of uniqueness* and should satisfy,.

$(1)\ \forall f \in C[0,1]; n \in \mathbb{N}; p_1, p_2 \in P_n; k \in \mathbb{N}$
$$(\bigwedge_{i=1}^{2} \|f - p_i\|_1 - dist_1(f, P_n) \leq 2^{-\Phi(f,n,k)} \to \|p_1 - p_2\|_1 < 2^{-k}).$$

It is important to note that the best L_1-approximation of any given continuous functions $f \in C[0,1]$ from P_n lives in a compact subspace of $C[0,1]$, for instance $K_{f,n} := \{p \in P_n : \|p\|_1 \leq 2\|f\|_1\}$, i.e. for a fixed $f \in C[0,1]$ its best L_1-approximation p_n belongs to $K_{f,n}$ for if not we would have $\|p_n\|_1 > 2\|f\|_1$ which, by the triangle inequality for the uniform norm, implies $\|f - p_n\|_1 > \|f\|_1$, and 0 would be better approximant than p_n, a contradiction. Therefore, $(**)$ can be written as,

$\forall f \in C[0,1]; n \in \mathbb{N}; p_1, p_2 \in K_{f,n}; k \in \mathbb{N} \exists d \in \mathbb{N}$
$$(\bigwedge_{i=1}^{2} \|f - p_i\|_1 - dist_1(f, P_n) \leq 2^{-d} \to \|p_1 - p_2\|_1 < 2^{-k}).$$

Since Cheney's proof can be formalized in $\mathbf{PA}^{\omega} + AC_{qf}^{1,0} + WKL$ (which was established in [Koh90]), by metatheorem 2.16, we are guaranteed the existence of a uniform modulus of uniqueness.

Notice that, although the first proof of uniqueness of best L_1-approximation was given already in 1921, no effective uniform modulus of uniqueness had ever been presented before.[15]

Now we present one of the applications of the uniform modulus of uniqueness, but first we define yet another norm. Let $a := (a_0, a_1, \ldots, a_n)$ be an $(n+1)$-dimensional tuple of real numbers. The *max-norm of* a is defined as,

$$\|a\|_{max} := \max\{|a_0|, \ldots, |a_n|\}.$$

[14]The term 'strong unicity' was introduced by Newman and Shapiro [NS63] and has been studied extensively in approximation theory. See e.g. the introduction in [BL95] and the references given there for a discussion of the crucial importance of estimates of strong unicity for the convergence analysis of iterative algorithms and for stability analysis.

[15]The state of the art in the study of uniform rates of strong uniqueness for L_1-approximation was partially supplied to us by Professor Kroo.

Since the polynomials of P_n can be viewed as a $(n+1)$-tuples of real numbers it makes sense to speak of the max-norm of $p \in P_n$.

Now, let $f \in C[0,1]$ be fixed and assume p_n is the best L_1-approximation of f from P_n. The uniform modulus of uniqueness Φ can be used to compute the Cauchy sequence $(p_{n,k})_{k\in\mathbb{N}}$ of approximants of p_n, i.e.,

$$\|p_n - p_{n,k}\|_{max} \le 2^{-k}.$$

That can be done once we have a functional $\Psi(f,n,k)$ such that,

(2) $\forall f \in C[0,1]; n,k \in \mathbb{N}(\Psi(f,n,k) \in P_n \wedge \|f - \Psi(f,n,k)\|_1 < dist_1(f,P_n) + 2^{-k})$,

and a function $\Theta(n,k)$ such that,[16]

$$(3)\ \forall n \in \mathbb{N}; p \in P_n; k \in \mathbb{N}(\|p\|_1 < 2^{-\Theta(n,k)} \to \|p\|_{max} < 2^{-k})$$

Now, by (1), (2) and (3), it is clear that the defined functional

$$B_f(k,n) := \Psi(f,n,\Phi(f,n,\Theta(n,k)))$$

generates the double sequence $(p_{n,k})_{n,k\in\mathbb{N}}$. Moreover, the complexity of the sequence depends only on the complexity of the functionals Θ, Φ, Ψ and on the complexity of integration (see section 4). This important application of moduli of uniqueness for the computation of unique solutions of existence statements was first investigated and used in [Koh93a].

Remark 3.1 *Note that the sequence of best approximating polynomials for a given function f can be viewed in two ways. The sequence $(p_n)_{n\in\mathbb{N}}$ is the sequence of actual best approximants of f from each space P_n. On the other hand, the sequence $(p_{n,k})_{k\in\mathbb{N}}$ is the sequence of polynomials converging to p_n. In this sense, we can also talk about the double sequence $(p_{n,k})_{n,k\in\mathbb{N}}$.*

In [Bjö75] (see also [Bjö79]) Björnestal proved that a uniform modulus of uniqueness having the form $(k+c)+\omega_f(k+c)$ exists, where the constant c depended at most on f and n. Kroo ([Kro78] and [Kro81]) improved this result by showing that c needed only to depend on the modulus of continuity of f (but not on any particular value of f). Moreover, he showed that the k-dependency, i.e. $k + \omega_f(k)$, was optimal.

By applying the techniques of monotone functional interpretation to Cheney's uniqueness proof we obtained the explicit uniform modulus of uniqueness,

Theorem 3.2 *Let*

$$\begin{aligned} \Phi(\omega,n,k) &:= \max\{k+1+c_n+\omega(k+3+c_n), \\ &\qquad 2k+8+2c_n+4\log(n+1)+\log(N)\}, \ \textit{where} \\ c_n &:= (2n^2+5n+4)\log(n+3), \end{aligned}$$

and N is an upper bound on $\sup_{x\in[0,1]}|f(x)-f(0)|$ (for instance $2^{\omega(0)}$). The functional Φ is a uniform modulus of uniqueness for the best L_1-approximation of any function f in $C[0,1]$ (having modulus of uniform continuity ω) from P_n.

[16] The functions Ψ and Θ are going to be presented in section 4.2.

As we mentioned, before doing the actual extraction of the uniform modulus of uniqueness presented in theorem 3.2, just by looking at the logical form of Jackson's theorem, and by the fact the Cheney's proof could be formalized in $\mathbf{PA}^\omega + AC^{1,0}_{qf} + WKL$, we knew already based on the metatheorem 2.16, that such a functional could be extracted. Moreover, since p_1, p_2 range over a compact space the extracted functional had to be independent of these polynomials, i.e. we could extract the uniform modulus of uniqueness. Those we call the *a priori* information, i.e. information that can be inferred just by the formalization of the proof.

As a corollary of theorem 3.2 we obtained the following explicit version of Kroo's and Björnestal's result,[17]

Corollary 3.3 *Let $f \in C[0,1]$, ω_f be some modulus of uniform continuity of f, and $p \in P_n$. Then,*

$$\forall k \in \mathbb{N}(\|f-p\|_1 < dist_1(f,P_n) + 2^{-(k+c)-\omega_{f-p_n}(k+c)} \rightarrow \|p-p_n\|_1 \leq 2^{-k}),$$

where $c := (2n^2+7n+7)\log(n+3)$, p_n is the best L_1-approximant of f from P_n, and ω_{f-p_n} is an estimate on the modulus of continuity of the function $f-p_n$ (see section 3.5 of the appendix).

Note that, as *a posteriori* information, we see that the constant c in corollary 3.3 depends only on n. In contrast to the *a priori* information mentioned above, the fact that c is independent of f could only be seen after extracting the functional. Now we present some special cases of the uniform modulus of uniqueness.

A function $f \in C[0,1]$ is said to *be Lipschitz continuous with Lipschitz constant* λ if $|f(x)-f(y)| \leq \lambda|x-y|$ (i.e. $k+\log(\lambda)$ is a modulus of continuity for f) and is *Lipschitz-α continuous with constant* λ, $0<\alpha\leq 1$, if $|f(x)-f(y)| \leq \lambda|x-y|^\alpha$ (equivalently, $\frac{k+\log(\lambda)}{\alpha}$ is a modulus of continuity for f).[18] In this way, if a function f is Lipschitz continuous (or Lipschitz-α continuous) with constant λ then $\sup_{x\in[0,1]}|f(x)-f(0)| \leq \lambda$ (and we can take $N := \lambda$ in theorem 3.2).

Corollary 3.4 *For any $f \in C[0,1]$,*

i) Let $\Phi_L(\lambda,n,k) := 2k+8+2c_n+4\log(n+1)+\log(\lambda)$. If f is Lipschitz continuous with constant λ then the functional $\Phi_L(\lambda,n,k)$ is a modulus of uniqueness for f.

ii) Let $\Phi_{L_\alpha}(\lambda,\alpha,n,k) := k+5+c_n+4\log(n+1)+\frac{k+3+c_n+\log(\lambda)}{\alpha}$. If f is Lipschitz-α continuous with constant λ then the functional $\Phi_{L_\alpha}(\lambda,\alpha,n,k)$ is a modulus of uniqueness for f.

Let $\mathcal{P}(f,n)$ denote the projection operator which assigns to any given function $f \in C[0,1]$ and any $n \in \mathbb{N}$ the best L_1-approximation of $f \in C[0,1]$ from P_n, i.e,

$$\mathcal{P}(f,n) := p_n, \text{ such that } \|f-p_n\|_1 = dist_1(f,P_n).$$

As a corollary of proposition 5.4 from [Koh93a] and theorem 3.2 above we get,

[17]The results of Björnestal and Kroo cannot be directly compared to ours since they use different notions of 'modulus of uniqueness' and 'uniform modulus of continuity'. The results, however, can still be compared in an indirect manner, see section 4.1 of the appendix.

[18]In Approximation Theory the condition '$|f(x)-f(y)| \leq \lambda|x-y|^\alpha$ for some λ' is called *Hölder condition with exponent α.*

Theorem 3.5 *Let $\Phi_P(\omega_f, n, k) := \Phi(\omega_f, n, k) + 1$, Φ as defined in Theorem 3.2. Then, Φ_P is a modulus of pointwise continuity for the operator $\mathcal{P}(f, n)$, i.e.,*

$$\forall f, \tilde{f} \in C[0,1]; n \in \mathbb{N}; k \in \mathbb{N}(\|f - \tilde{f}\|_1 < 2^{-\Phi_P(\omega_f, n, k)} \to \|\mathcal{P}(f, n) - \mathcal{P}(\tilde{f}, n)\|_1 \leq 2^{-k}).$$

Note that our case study is twofold: we test the techniques of Proof Theory (monotone functional interpretation in the present case) with respect to its applicability and feasibility, and at the same time we obtain new results in Analysis.

4 Computable Analysis

We sketched in section 3 how to use the uniform modulus of uniqueness to compute the sequence of best L_1-approximations of a given function $f \in C[0,1]$. We mentioned also that the complexity of generating such a sequence depends on the complexity of the functionals Θ, Ψ and Φ. Now we introduce some notions from Computable Analysis which are going to be useful in defining the computational complexity of the sequence $(p_{n,k})_{n,k\in\mathbb{N}}$. After that, we present possible functionals Θ and Ψ, and we fully describe how such sequence can be computed using the presented functionals. In section 5.1, we point to future work on the estimation of the complexity of the algorithm generating $(p_{n,k})_{n,k\in\mathbb{N}}$.

4.1 Introduction

While classical complexity theory deals with countable sets (e.g. $\mathbb{N}$, Σ^* for a finite alphabet Σ, etc.) Computable (or Effective) Analysis deals with uncountable sets (e.g. $\mathbb{R}$, $C[0,1]$, Σ^ω, etc.). In the following we present Ko's approach to effective analysis as defined in [Ko91] and [Ko98]. Therefore, all the definitions, theorem 4.3 and corollary 4.4 in this subsection are taken from [Ko91] with small changes on the notation.[19] In Ko's approach (whose concepts can be traced back to [Grz55]) a real number is viewed as a converging sequence of rational approximations[20] with fixed rate of convergence. Computability is then defined in this framework,

Definition 4.1 *A real number x is computable if there exists a Turing machine M_x which for each input $n \in \mathbb{N}$ outputs a rational number q_n such that $|x - q_n| \leq 2^{-n}$.*

In other words, $x \in \mathbb{R}$ is computable if there exists a Turing machine M_x generating on each input n some rational q_n such that $q_0, q_1, \ldots$ is a Cauchy sequence converging to x with fixed rate 2^{-n}. Suppose the input n is given in unary. If the machine M_x works in polynomial time we say that x is a *poly-time computable real number*, and the same holds for any other complexity classes: **EXP**, **NP**, etc..

Now we can go one step higher and ask what is the complexity of a function $f : \mathbb{R} \to \mathbb{R}$. In this case we are interested in estimating the time required to compute $f(x)$ for any given $x \in \mathbb{R}$. Since we are only interested in the complexity of f we abstract the complexity of the input x. We do this by assuming that x is given through an *oracle machine O_x*, where on input m the oracle returns in constant time a rational number q_m such that $|x - q_m| \leq 2^{-m}$.

Definition 4.2 *A function $f : \mathbb{R} \to \mathbb{R}$ is said to be computable if there exist a Turing machines M_f which on input n (and oracle O_x) outputs a rational number q_n such that $|f(x) - q_n| \leq 2^{-n}$.*

[19] For other essentially equivalent approaches to computable analysis see for instance [Wei00] and [PER89].

[20] In fact, for simplicity in the complexity analysis, Ko uses only rationals which are dyadic (i.e. have finite binary representation) as approximants. In our case here, where we do not endeavor to give any explicit complexity results, the choice is irrelevant.

It is easy to see that any computable function must be continuous. Moreover, it can also be proved that any computable f, at a fixed compact interval $[a,b]$, has a computable uniform modulus of continuity, i.e., there exists a computable $w_f : \mathbb{N} \to \mathbb{N}$ s.t.,

$$\forall n \in \mathbb{N}; x, y \in [a,b](|x-y| \leq 2^{-w_f(n)} \to |f(x)-f(y)| \leq 2^{-n}).$$

Theorem 4.3 *If $f : [a,b] \to \mathbb{R}$ is computable on $[a,b]$ then,*

- *f is continuous on $[a,b]$ and,*
- *f has a computable uniform modulus of continuity.*

As a corollary of this fact we can get a complete characterization of the computable functions in terms of two number theoretic functions,

Corollary 4.4 *A function $f : [a,b] \to \mathbb{R}$ is computable iff there exists two computable functions $f_r : \mathbb{Q} \cap [a,b] \times \mathbb{N} \to \mathbb{Q}$ and $\omega_f : \mathbb{N} \to \mathbb{N}$ such that,*

- $\forall q \in \mathbb{Q} \cap [a,b]; n \in \mathbb{N}(|f_r(q,n) - f(q)| \leq 2^{-n})$,
- *ω_f is a uniform modulus of continuity for f.*

The restriction to a compact interval here is essential since a continuous function on $\mathbb{R}$ need not to be uniformly continuous on $\mathbb{R}$. Once we have this characterization of computable functions f on compact intervals via a pair of computable number theoretic functions (f_r, ω_f) we can define very easily the complexity of real functions on $[a,b]$. A function $f : [a,b] \to \mathbb{R}$ is *poly-time computable* if f_r is poly-time computable and w_f is a polynomial. Similar definitions can be given regarding other complexity classes: **EXP**, **NP**, etc..

Again, it is worth noticing that without the restriction of f to a compact set $[a,b]$ we could not talk about the complexity of f in general. For instance, if f happens to be unbounded the size of the outputs q_n can be arbitrarily large, and the time spent to compute q_n cannot be distinguished from the time spent to write down these long outputs.

Now we have defined what it means to compute real numbers and functions on the reals. Since the elements of P_n can be viewed as $(n+1)$-tuples of real numbers, and therefore as real numbers, the sequence $(p_n)_{n\in\mathbb{N}}$ can be viewed as a sequence of real numbers, i.e. $\mathbb{N} \to (\mathbb{N} \to \mathbb{Q})$, or equivalently $\mathbb{N} \times \mathbb{N} \to \mathbb{Q}$. Let us define the notion of computability for such sequences,

Definition 4.5 *A sequence $(x_n)_{n\in\mathbb{N}}$ of real numbers is said to be computable if there exists a Turing machine M which on input n, k outputs a rational $q_{n,k}$ such that $|x_n - q_{n,k}| \leq 2^{-n}$.*

In this last case, the sequence is said to be *poly-time computable* if on inputs n and k (in unary) the machine M outputs $q_{n,k}$ in polynomial time.

4.2 Using the Uniform Modulus of Uniqueness

Here we present an algorithm that receives as input $f \in C[0,1]$ and $n, k \in \mathbb{N}$, and outputs $p_{n,k} \in \mathbb{Q}$ such that $\|p_n - p_{n,k}\|_{max} < 2^{-k}$, where p_n is the best L_1-approximation of f from P_n. More precisely,

Definition 4.6 *An operator $B_f : \mathbb{N} \times \mathbb{N} \to P_n$ computes the sequence of unique best L_1-approximations, $(p_n)_{n \in \mathbb{N}}$, of a function $f \in C[0,1]$ if for any given $n, k \in \mathbb{N}$ it generates $B_f(n,k) \in P_n$ such that $\|p_n - B_f(n,k)\|_{max} < 2^{-k}$.*

We show explicitly how $B_f(n,k)$ can be defined using the uniform modulus of uniqueness $\Phi(f,n,k)$ presented in the previous section. First we need to define the two auxiliary functionals Ψ and Θ mentioned in section 3.

Let $\|\cdot\|_\infty$ denote the uniform norm on $C[0,1]$, i.e. $\|f\|_\infty := \max_{x \in [0,1]} |f(x)|$. In the following we will use Markov's inequality which states that, for any given polynomial p of degree n, $\|p'\|_\infty \leq 2n^2 \|p\|_\infty$, where p' denotes the first derivative of p.

Claim 4.7 *Let $p \in P_n$. Then, $\|p\|_\infty \leq 2(n+1)^2 \|p\|_1$.*

Proof: Let $p = a_0 + a_1 x + \ldots + a_n x^n$. Define $p^*(x) := a_0 x + \frac{a_1}{2} x^2 + \ldots + \frac{a_n}{n+1} x^{n+1}$. It is clear that for any $x \in [0,1]$,

$$|p^*(x)| = |\int_0^x p(x)dx| \leq \int_0^x |p(x)|dx \leq \|p\|_1,$$

therefore, $\|p^*\|_\infty \leq \|p\|_1$. Since the derivative of p^* equals p, by Markov's inequality, we have $\|p\|_\infty \leq 2(n+1)^2 \|p\|_1$.

□

The function $\Theta(n,k)$: This function tells us how small the L_1-norm of p should be so that the coefficients of p are small.

Claim 4.8 *Let $\Theta(n,k) := n + k + 1 + 2(n+1)\log(n+1)$. Then,*

$$\forall n \in \mathbb{N}; p \in P_n; k \in \mathbb{N}(\|p\|_1 \leq 2^{-\Theta(n,k)} \to \|p\|_{max} \leq 2^{-k}).$$

Proof: Assume $\|p\|_1 \leq \frac{2^{-n-k}}{2(n+1)^{2(n+1)}}$. By claim 4.7 we have, $\|p\|_\infty \leq \frac{2^{-n-k}}{(n+1)^{2n}}$. Let $p = a_0 + a_1 x + \ldots + a_n x^n$ and $p^{(i)}$ denote the i-th derivative of p. It is clear that $a_i = \frac{p^{(i)}(0)}{i!}$. By applying Markov's inequality i times we have,

$$||p^{(i)}||_\infty \leq \frac{2^i (n!)^2}{((n-i)!)^2} \|p\|_\infty \leq \frac{2^{i-n-k}(n!)^2}{((n-i)!)^2 (n+1)^{2n}}.$$

Therefore,

$$|a_i| \leq \frac{2^{i-n-k}(n!)^2}{i!((n-i)!)^2(n+1)^{2n}} = \frac{2^{i-1}}{i!} \cdot \frac{(n!)^2}{((n-i)!)^2 2^{n-1}(n+1)^{2n}} \cdot 2^{-k} \leq 2^{-k}.$$

□

The functional $\Psi(f, n, k)$**:** Now we show how to find a polynomial $p \in P_n$ such that $\|f-p\|_1 < dist_1(f, P_n) + 2^{-k}$. As it was pointed out in section 3, such p lives in the compact space $K_{f,n}$. The idea is to build a finite net of elements from $K_{f,n}$ (called k-net) such that at least one of them (say p') is $(k+1)$-*close* to p_n, i.e. $\|p_n - p'\|_1 \leq 2^{-k-1}$, which by triangle inequality ($\|f-p'\|_1 \leq \|f-p_n\|_1 + \|p_n - p'\|_1$) yields (1) $\|f-p'\|_1 \leq dist_1(f, P_n) + 2^{-k-1}$. Once we have the k-net we compute $\|f-p\|_1$ for any element p in the net with $(k+1)$-precision and we take a $\tilde{p}$ which gives the minimum value (e.g. the one with smallest code). Since the L_1-norm was computed with $(k+1)$-precision, by (1), we have, $\|f-\tilde{p}\|_1 \leq dist_1(f, P_n) + 2^{-k}$. We present now such a k-net, for this we use C_n, the upper bound on the coefficients of p_n which can be derived from $\|p_n\|_1 \leq 2\|f\|_1$ (as seen in claims 4.7 and 4.8).

Claim 4.9 *Let* $C_{n,k,i} := \{\frac{2^{-k}(i+1)}{n+1} j : -C_n \leq j \leq C_n\}$*. The space of polynomials* $N_{n,k} = \{a_0 + \ldots + a_n x^n : a_i \in C_{n,k,i}, 0 \leq i \leq n\}$ *is a* k*-net for* P_n*.*

Proof: Take an arbitrary element of P_n, say $p = b_0 + \ldots + b_n x^n$. In the way we chose the coefficients of the elements of $N_{n,k}$ we are sure to find $p^* = a_0 + \ldots + a_n x^n \in N_{n,k}$ such that $|a_i - b_i| \leq \frac{2^{-k-1}(i+1)}{n+1}$, $0 \leq i \leq n$, i.e.,

$$\begin{aligned}
||p - p^*||_1 &= ||(a_0 - b_0) + \ldots + (a_n - b_n)x^n||_1 \\
&= \int_0^1 |(a_0 - b_0) + \ldots + (a_n - b_n)x^n| dx \\
&\leq |a_0 - b_0| + \ldots + \frac{|a_{n-1} - b_{n-1}|}{n} + \frac{|a_n - b_n|}{n+1} \\
&\leq \frac{2^{-k-1}}{n+1} + \ldots + \frac{2^{-k-1}n}{n(n+1)} + \frac{2^{-k-1}(n+1)}{(n+1)(n+1)} = 2^{-k-1}
\end{aligned}$$

□

Let $N_{n,k}$ be as above, we define $\Psi(f, n, k)$ as one of the elements $\tilde{p} \in N_{n,k}$ s.t. $\|f - \tilde{p}\|_{1,k+1} = \min\{\|f-p\|_{1,k+1} : p \in N_{n,k}\}$, where $\|\cdot\|_{1,k+1}$ denotes the value of the L_1-norm with $(k+1)$-precision. Then, as shown above,

$$\|f - \Psi(f, n, k)\|_1 < dist_1(f, P_n) + 2^{-k}.$$

Now we spell out the whole algorithm again. Let $f \in C[0,1]$ and $n, k \in \mathbb{N}$ be fixed. We want to generate a polynomial $p_{n,k}$ in P_n such that, $\|p_n - p_{n,k}\|_{max} < 2^{-k}$. By claim 4.8 it is sufficient to find a $p_{n,k}$ such that, $\|p_n - p_{n,k}\|_1 \leq 2^{-\Theta(k,n)}$. And by theorem 3.2 it is enough to find a $p_{n,k}$ such that,

$$\|f - p_{n,k}\|_1 \leq dist_1(f, P_n) + 2^{-\Phi(f,n,\Theta(n,k))}.$$

Finally, by building a $\Phi(f, n, \Theta(n,k))$-net as above the element $p_{n,k}$ can be found, i.e. $B_f(n,k) := \Psi(f, n, \Phi(f, n, \Theta(n,k)))$.

5 Future Work

5.1 Complexity of finding the best L_1-approximation

Once we abstract the complexity of integration, the most costly step in the procedure described in section 4.2 is the search through the k-net. Therefore, in order to analyze the complexity of generating the sequence of best L_1-approximation of a given function $f \in C[0,1]$ one has (in the case of the algorithm above) to give an estimate on the size of the $\Phi(f, n, \Theta(f, n, k))$-net. In a subsequent paper we present that estimate together with a full analysis of the complexity of the algorithm. We also intend to consider special cases where the complexity of integration can be taken into account to provide absolute results for the complexity of the sequence. A similar complexity analysis for the case of best Chebycheff approximation (where the *uniform* norm is considered) was carried out in [Ko86].

5.2 Extend result to other Haar subspaces of $C[0,1]$

Although Jackson's original proof deals only with the problem of approximating $f \in C[0,1]$ by polynomials of bounded degree (i.e. P_n), Cheney's proof (which we analyzed in [KO01], as presented in section 3) treats arbitrary Haar subspaces H of $C[0,1]$,

$$\forall f \in C[0,1]; h_1, h_2 \in H(\bigwedge_{i=1}^{2} \|f - h_i\|_1 = dist(f, H) \to \|h_1 - h_2\|_1 = 0).$$

This means we can also extract from Cheney's proof a general uniform modulus of uniqueness parametrized with some information α about the Haar subspace H under consideration,

$$\forall f \in C[0,1]; h_1, h_2 \in H; k \in \mathbb{N}(\bigwedge_{i=1}^{2} \|f - h_i\|_1 \leq dist(f, H) + 2^{-\Phi(f,\alpha,k)} \to \|h_1 - h_2\|_1 \leq 2^{-k}).$$

The information α is normally a constructive proof of the fact that H is a Haar subspace of $C[0,1]$ or, if H is given by a Chebycheff system $\underline{\phi} = \{\phi_1, \ldots, \phi_n\}$, a constructive proof that $\underline{\phi}$ a Chebycheff system. From α we can extract e.g. a function $\delta_H : \mathbb{Q} \to \mathbb{Q}$ such that,

$$\forall h \in H; x_1, \ldots, x_n \in [0,1]; q, r \in \mathbb{Q}_+^*(\bigwedge_{i=1}^{n-1} x_{i-1} + q \leq x_i \wedge \bigwedge_{i=1}^{n} |h(x_i)| \leq \delta_H(q, r) \to \|h\|_1 \leq r).$$

Those functions are used in the construction of the uniform modulus of continuity. Therefore, alternatively we can take the information α to be those functions, i.e. the uniform modulus of uniqueness for L_1-approximation of $C[0,1]$ from arbitrary Haar subspaces H has the form $\Phi(f, \delta_H, \ldots, k)$.

In this sense, by considering effective representations of Chebycheff systems as the one presented in [Bri80] and [Bri82], we intend to carry out the extraction of uniform moduli of uniqueness for arbitrary Haar subspaces H, as an extension of the result obtained in [KO01].

5.3 Relation between the different approaches to Proof Mining

As we have seen, some approaches to Proof Mining (e.g. (N, A, mr)-translation) are more suitable and have been implemented in real computer systems (e.g. [HN88] [BS95] [BSS01] [BSBar]) where the extraction of information from the proof can be performed automatically in an iterative process. Unfortunately, just simple case studies have been carried out due to the complexity of formalizing the proofs in these systems. On the other hand, through the use of proof interpretation in an informal way, where shortcuts can be taken and trivial steps ignored, quite a few interesting case studies have been treated (e.g. [Ber93], [Luc89], [Koh93a], [Koh93b], [Kohar], [Kohb], [KO01]) even yielding new results in functional programming, number theory, approximation theory and fixed point theory.

We intend to investigate the relation between the different approaches to Proof Mining with respect to (among others),

- Applicability: are there proofs which can be analyzed in such framework yielding interesting information?
- Implementation: can the process be easily implemented in a real system running in feasible time?

5.4 Computable Analysis via Domain Theory

Let C denote the set of bounded and closed intervals in $\mathbb{R}$. The poset $(C, \supseteq)$ was first considered as a data type for real numbers in [Sco70]. In fact, $(C, \supseteq)$ is a Scott domain with countable basis given by the set of all closed intervals with rational endpoints. The maximal elements of this domain are the singleton sets $\{x\}$, $x \in \mathbb{R}$. The lub of any directed subset of this domain is the intersection of all intervals in the subset. Therefore, a real number is approximated by a shrinking sequence of rational nested intervals. In this setting, a *computable real number* is the lub of a shrinking sequence of rational intervals which can be generated by some Turing machine. Similar domain-theoretic definitions of computability have been given with respect to functions on real numbers, Banach spaces, complete metric spaces, etc. (see [SHT88], [Eda97], [ES98] and [ES99]).

Notice that, the domain-theoretic approach of computability is weaker (with respect to complexity analysis) than the one used in section 4.1. For instance, considering real numbers, for a given sequence $q_0, q_1, \ldots$ of rational approximations of $x \in \mathbb{R}$ with $|x - q_i| \leq 2^{-k}$ we can generate very efficiently the sequence of shrinking intervals with rational endpoints $I_0, I_1, \ldots$, where e.g. $I_i := [q_i - 2^{-i}, q_i + 2^{-i}]$; but the other way around requires *unbounded search*. Therefore, the domain-theoretic approach is a priori not suitable for complexity analysis.

In order to cope with this problem very recently the idea of Quantitative Domain theory (where the partial order in the domains are equipped with qualitative information) started to be developed (cf. [Sün00]). We intend to study in this new framework the relation between Domain Theory and Proof Mining by applying Quantitative Domain theory to analyze the complexity of the extracted terms from given proofs.

5.5 Weak fragments of Analysis

The weakest second-order system considered in reverse mathematics (as it appears in [Sim99]) is the system RCA_0 whose provably total functions are the primitive recursive functions. By further restrictions one obtains for instance the system BTFA (Base Theory for Feasible Analysis, [Fer94]) with only polynomials as provably total functions. In [FFar] it has been shown that basic notions of analysis can be formalized in BTFA. This system is defined in a language over strings and the standard structure has domain $\{0,1\}^*$. The main point is that induction in BTFA is restricted to induction on notation exclusively for Σ_1^b-formulas (formulas which define exactly the NP predicates in the standard model), i.e.

$$A(\varepsilon) \wedge \forall x(A(x) \rightarrow A(x0) \wedge A(x1)) \rightarrow \forall x A(x).$$

In BTFA not even the exponential functional can be proven total. A benefit is that from any proof *which can be formalized in this system* one can extract realizing terms denoting polynomials. The drawback is that very few and mostly trivial proofs have been shown to be formalized in BTFA. Natural questions involve,

- look for proofs which can be formalized in those weak systems,
- extract interesting/new information from these proofs, and
- look for a balance between effectiveness and applicability.

It is worth noting that is these systems full classical logic is used. For a survey on weak Analysis using BTFA see [FFar]. See also [CU93] for a D-interpretation of such systems.

5.6 Semi-classical systems

Let $A := \forall x \exists y A_0(x,y)$, where A_0 is quantifier free and A has no free variables. As it was shown in theorem 2.15 from a proof of A in

$$\mathcal{A}_1 := \mathbf{PA}^\omega + AC_{qf} + WKL$$

one can always extract a term tx such that $\mathcal{A}_1 \vdash \forall x \exists y \leq tx A_0(x,y)$.

From the system $\mathcal{A}_1$ we can define the system $\mathcal{A}_2$ by moving down to intuitionistic logic and at the same time strengthening the principles added, i.e.

$$\mathcal{A}_2 := \mathbf{HA}^\omega + AC + CA_\neg,$$

where

$$CA_\neg : \exists \Phi^{\rho \rightarrow 0} \leq 1 \forall x^\rho (\Phi(x) = 0 \leftrightarrow \neg A(x)),$$

and AC denotes full axiom of choice. The semi-classical system $\mathcal{A}_2$ has the nice property that, whenever $\mathcal{A}_2 \vdash \forall x \exists y A(x,y)$ for *arbitrary* A, a bounding term can be extracted for y.

A third system can be defined from $\mathcal{A}_2$ by adding Markov principle and weakening comprehension to WKL, i.e.

$$\mathcal{A}_3 := \mathbf{HA}^\omega + MP + AC + WKL.$$

The system $\mathcal{A}_3$ enjoys the same nice feature of $\mathcal{A}_2$. The WKL in the setting of $\mathcal{A}_3$ corresponds to the *lesser limited principle of omniscience*,

$$LLPO : \forall x \in \mathbb{R}(x \geq 0 \vee x \leq 0),$$

as shown in [Ish90] (see also [Koha], [Man88], [HN] and [CP00])

The semi-classical systems $\mathcal{A}_2$ and $\mathcal{A}_3$ have been analysed in e.g. [Koha] and [Koh98b]. Future work in this area involves studying those systems and specially looking for proofs in Analysis which can be captured precisely by such systems. If such proofs are found, we intend to carry out extraction of hopefully interesting information.

Acknowledgement: I am grateful to Ulrich Kohlenbach for stimulating discussions and helpful suggestions which led to various improvements of this report. I am also most indebted to Kohlenbach and my colleague Frank Valencia for proofreading the final draft.

References

[AF98] J. Avigad and S. Feferman. Gödel's functional ("Dialectica") interpretation. In S.R. Buss, editor, *Handbook of proof theory*, volume 137, pages 337–405. Elsevier, North-Holland, Amsterdam, 1998.

[Ber93] U. Berger. Program extraction from normalization proofs. In M. Bezem, editor, *Typed Lambda Calculi and Applications (TLCA'93), Lecture Notes in Computer Science*, volume 664, pages 91–106. Springer Verlag, 1993.

[Bjö75] B.O. Björnestal. Continuity of the metric projection operator i-iii. *The preprint series of Department of Mathematics. Royal Institute of Technology. Stockholm, TRITA-MAT*, 17, 1975.

[Bjö79] B.O. Björnestal. Local Lipschitz continuity of the metric projection operator. *Approximation theory. In: Papers, VIth Semester, Stefan Banach Internat. Math. Center, Warsaw, 1975)*, pages 43–53, 1979.

[BL95] M. Bartelt and W. Li. Error estimates and Lipschitz constants for best approximation in continuous function spaces. *Computers and Mathematics with Application*, 30(3-6):255–268, 1995.

[Bri80] D.S. Bridges. A constructive development of Chebychev approximation theory. *Journal of Approximation Theory*, 30:99–120, 1980.

[Bri82] D.S. Bridges. Lipschitz constants and moduli of continuity for the Chebychev projection. *Proc. Amer. Math. Soc.*, 85:557–561, 1982.

[BS95] U. Berger and H. Schwichtenberg. Program extraction from classical proofs. In D. Leivant, editor, *Logic and Computational Complexity Workshop (LCC'94), Lecture Notes in Computer Science*, volume 960, pages 77–97. Springer Verlag, 1995.

[BSBar] U. Berger, H. Schwichtenberg, and W. Buchholz. Refined program extraction from classical proofs. *Annals of Pure and Applied Logic*, to appear.

[BSS01] U. Berger, H. Schwichtenberg, and M. Seisenberger. The Warshall algorithm and Dickson's lemma: Two examples of realistic program extraction. *Journal of Automated Reasoning*, 26, 2001.

[Bus95] S. R. Buss. On Herbrand's theorem. In D. Leivant, editor, *Logic and Computational Complexity, Lecture Notes in Computer Science*, volume 960, pages 195–209. Springer Verlag, 1995.

[Bus98] S.R. Buss. *Handbook of proof theory*, volume 137. Elsevier, 1998.

[CH99] T. Coquand and M. Hofmann. A new method for establishing conservativity of classical systems over their intuitionistic version. *Math. Struct. in Comp. Science*, 9:323–333, 1999.

[Che66] E.W. Cheney. *Approximation Theory*. AMS Chelsea Publishing, 1966.

[CP00] T. Coquand and E. Palmgren. Intuitionistic choice and classical logic. *Arch. Math. Logic*, 39:53–74, 2000.

[CU93] S. Cook and A. Urquhart. Functional interpretations of feasibly constructive arithmetic. *Annals of Pure and Applied Logic*, 63:103–200, 1993.

[Del96] C. Delzell. Kreisel's unwinding of Artin's proof. In P. Odifreddi, editor, *Kreiseliana: About and around George Kreisel*, pages 113–246. A. K. Peters, Wellesley, MA, 1996.

[Dra79] A. Dragalin. New kinds of realizability. In *Abstracts of the 6th International Congress of Logic, Methodology and Philosophy of Sciences, Hannover, Germany*, pages 20–24, 1979.

[Eda97] A. Edalat. Domains for computation in mathematics, physics and exact real arithmetic. *Bulletin of Symbolic Logic*, 3(4):401–452, 1997.

[ES98] A. Edalat and P. Sünderhauf. A domain-theoretic approach to real number computation. *Theoretical Computer Science*, 210:73–98, 1998.

[ES99] A. Edalat and P. Sünderhauf. Computable Banach spaces via domain theory. *Theoretical Computer Science*, 219:169–184, 1999.

[Fef93] S. Feferman. Gödel's Dialectica interpretation and its two-way stretch. In G. et al. Gottlob, editor, *Computational Logic and Proof Theory, Lecture Notes in Computer Science*, pages 23–40. Springer Verlag, 1993.

[Fer94] F. Ferreira. A feasible theory for analysis. *Journal of Symbolic Logic*, pages 1001–1011, 1994.

[FFar] F. Ferreira and A. Fernandes. Groundwork for weak Analysis. *Journal of Symbolic Logic*, (to appear).

[Fri78] H. Friedman. Classically and intuitionistically provably recursive functions. In D. Scott and G. Müller, editors, *Higher Set Theory, Lecture Notes in Mathematics*, volume 669, pages 21–28. Springer Verlag, 1978.

[Göd58] K. Gödel. Über eine bisher noch nicht benützte erweiternung des finiten standpunktes. *Dialectica*, 12:280–287, 1958.

[Göd90] K. Gödel. Collected works, vol. ii. pages 1938–1974. Oxford University Press, Oxford, 1990.

[Grz55] A. Grzegorczyk. Computable functions. *Fund. Math.*, 42:168–202, 1955.

[HN] S. Hayashi and M. Nakata. Realizability interpretation for limit computable mathematics. preprint.

[HN88] S. Hayashi and H. Nakano. *PX: A Computational Logic.* MIT Press, 1988.

[Ish90] H. Ishihara. An omniscience principle, the König lemma and the Hahn-Banach theorem. *Zeitschr. f. math. Logik und Grundlagen d. Math.*, pages 237–240, 1990.

[Jac21] D. Jackson. Note on a class of polynomials of approximation. *Transactions of the American Mathematical Society*, 22:320–326, 1921.

[Kle45] S. C. Kleene. On the interpretation of intuitionistic number theory. *Journal of Symbolic Logic*, 10:109–124, 1945.

[Kle52] S. C. Kleene. *Introduction to Metamathematics*. North-Holland Publishing Co., 1952.

[Kle65] S. C. Kleene. *The Foundations of Intuitionistic Mathematics*. North-Holland Publishing Co., 1965.

[Kle67] S. C. Kleene. *Mathematical Logic*. John Wiley and Sons, 1967.

[Ko86] K.-I. Ko. On the computational complexity of best Chebycheff approximation. *Journal of Complexity*, 2:95–120, 1986.

[Ko91] K.-I. Ko. *Complexity theory of real functions*. Birkhäuser, Boston-Basel-Berlin, 1991.

[Ko98] K. Ko. *Polynomial-Time Computability in Analysis*, volume 2, pages 1271–1317. North-Holland, Amsterdam, 1998.

[KO01] U. Kohlenbach and P. Oliva. Effective bounds on strong unicity in L_1-approximation. BRICS Report Series RS-01-14, BRICS, 2001.

[Koha] U. Kohlenbach. Intuitionistic choice and restricted classical logic. To appear in: Math. Logic Quaterly, 9 pages.

[Kohb] U. Kohlenbach. A quantitative version of a theorem due to Borwein-Reich-Shafrir. To appear in: Numerical Functional Analysis and Optimization, 15 pages.

[Koh90] U. Kohlenbach. *Theory of majorizable and continuous functionals and their use for the extraction of bounds from non-constructive proofs: effective moduli of uniqueness for best approximations from ineffective proofs of uniqueness (german).* PhD thesis, Frankfurt, pp. xxii+278, 1990.

[Koh93a] U. Kohlenbach. Effective moduli from ineffective uniqueness proofs. an unwinding of de La Vallée Poussin's proof for Chebycheff approximation. *Annals of Pure and Applied Logic*, 64:27–94, 1993.

[Koh93b] U. Kohlenbach. New effective moduli of uniqueness and uniform a–priori estimates for constants of strong unicity by logical analysis of known proofs in best approximation theory. *Numerical Functional Analysis and Optimization*, 14:581–606, 1993.

[Koh96] U. Kohlenbach. Analysing proofs in Analysis. In W. Hodges, M. Hyland, C. Steinhorn, and J. Truss, editors, *Logic: from Foundations to Applications*, pages 225–260. European Logic Colloquium (Keele, 1993), Oxford University Press, 1996.

[Koh98a] U. Kohlenbach. Proof interpretations. BRICS lecture series, Basic Research in Computer Science, 1998.

[Koh98b] U. Kohlenbach. Relative constructivity. *Journal of Symbolic Logic*, 63:1218–1238, 1998.

[Koh99] U. Kohlenbach. On the no-counterexample interpretation. *Journal of Symbolic Logic*, 64:1491–1511, 1999.

[Kohar] U. Kohlenbach. On the computational content of the Krasnoselski and Ishikawa fixed point theorems. In J. Blanck, V. Brattka, P. Hertling, and K. Weihrauch, editors, *Computability and Complexity in Analysis, (CCA'2000), Lecture Notes in Computer Science*. Springer LNCS, to appear.

[Kre51] G. Kreisel. On the interpretation of non-finitist proofs, part I. *Journal of Symbolic Logic*, 16:241–267, 1951.

[Kre52] G. Kreisel. On the interpretation of non-finitist proofs, part II: Interpretation of number theory. *Journal of Symbolic Logic*, 17:43–58, 1952.

[Kre58] G. Kreisel. Mathematical significance of consistency proof. *Journal of Symbolic Logic*, 23(2):155–182, 1958.

[Kre62] G. Kreisel. On weak completeness of intuitionistic predicate logic. *Journal of Symbolic Logic*, 27:139–158, 1962.

[Kro78] A. Kroo. On the continuity of best approximations in the space of integrable functions. *Acta Mathematica Academiae Scientiarum Hungaricae*, 32:331–348, 1978.

[Kro81] A. Kroo. On strong unicity of L_1-approximation. *Proceedings of the American Mathematical Society*, 83(4), 1981.

[Kur51] S. Kuroda. Intuitionistische Untersuchungen der formalistischen Logik. *Nagoya Math.*, 3:35–47, 1951.

[Lei85] D. Leivant. Syntactic translations and provably recursive functions. *Journal of Symbolic Logic*, 50(3):682–688, Sep., 1985.

[Luc89] H. Luckhardt. Herbrand-Analysen zweier Beweise des Satzes von Roth: Polynomiale Anzahlschranken. *Journal of Symbolic Logic*, 54:234–263, 1989.

[Luc96] H. Luckhardt. Bounds extracted by Kreisel from ineffective proofs. In P. Odifreddi, editor, *Kreiseliana: About and around George Kreisel*, pages 289–300. A. K. Peters, Wellesley, MA, 1996.

[Man88] M. Mandelkern. Limited omniscience and the Bolzano-Weierstrass principle. *Bull. London Math. Soc.*, 20:319–320, 1988.

[Mur90] C. Murthy. *Extracting Constructive Content from Classical Proofs*. PhD thesis, Cornell University, 1990.

[NS63] D. J. Newman and H. S Shapiro. Some theorems on Chebycheff approximation. *Duke Mathematical Journal*, 30:673–681, 1963.

[PER89] M. Pour-El and J. Ian Richards. *Computability in Analysis and Physics*. Perspectives in Mathematical Logic. Springer-Verlag, 1989.

[Pin89] A. Pinkus. *On L_1-Approximation*, volume 93 of *Cambridge Trats in Mathematics*. Cambridge University Press, 1989.

[Sco70] D. S. Scott. Outline of a mathematical theory of computation. In *4th Annual Princeton Conference on Information Sciences and Systems*, pages 169–176, 1970.

[Sho67] J. R. Shoenfield. *Mathematical Logic*. Addison-Wesley Publishing Company, 1967.

[SHT88] V. Stoltenberg-Hansen and J. V. Tucker. Complete local rings as domains. *Journal of Symbolic Logic*, 53(2):603–624, 1988.

[Sim99] S.G. Simpson. *Subsystems of Second Order Arithmetic*. Perspectives in Mathematical Logic. Springer-Verlag, 1999.

[Sün00] P. Sünderhauf. Tensor products and powerspaces in quantitative domain theory. In S. Brookes and M. Mislove, editors, *Electronic Notes in Theoretical Computer Science*, volume 6. Elsevier Science Publishers, 2000.

[Tro73] A.S. Troelstra. *Metamathematical investigation of intuitionistic Arithmetic and Analysis, Lecture Notes in Mathematics*, volume 344, pages 1–323. Springer Verlag, 1973.

[Tro98] A. S. Troelstra. Realizability. In S.R. Buss, editor, *Handbook of proof theory*, volume 137, pages 408–473. North Holland, Amsterdam, 1998.

[TS96] A. S. Troelstra and H. Schwichtenberg. *Basic Proof Theory*. Cambridge University Press, 1996.

[Tv88] A. S. Troelstra and Dalen D. van. *Constructivism in Mathematics. An Introduction*, volume 121. North Holland, Amsterdam, 1988.

[vD00] D. van Dalen. Intuitionism – counting its blessings. In G. Jean-Yves, M. Okada, and A. Scedrov, editors, *Electronic Notes in Theoretical Computer Science*, volume 3. Elsevier Science Publishers, 2000.

[Wei00] K. Weihrauch. *Computable Analysis*. Springer-Verlag, 2000.

6 Appendix

6.1 Effective bounds on strong unicity in L_1-approximation.