Axioms and Decidability for Type Isomorphism in the Presence of Sums

Danko Ilik

Axioms and Decidability for Type Isomorphism in the Presence of Sums

Danko Ilik

Inria

danko.ilik@inria.fr

Abstract

We consider the problem of characterizing isomorphisms of types, or, equivalently, constructive cardinality of sets, in the simultaneous presence of disjoint unions, Cartesian products, and exponentials. Mostly relying on results about polynomials with exponentiation that have not been used in our context, we derive: that the usual finite axiomatization known as High-School Identities (HSI) is complete for a significant subclass of types; that it is decidable for that subclass when two types are isomorphic; that, for the whole of the set of types, a recursive extension of the axioms of HSI exists that is complete; and that, for the whole of the set of types, the question as to whether two types are isomorphic is decidable when base types are to be interpreted as finite sets. We also point out certain related open problems.

Categories and Subject Descriptors [*Logic*]: Type theory; [*Logic*]: Constructive mathematics; [*Semantics and reasoning*]: Type structures

General Terms Theory

Keywords Type isomorphism, Sum types, Completeness of axioms, Decidability

1. Introduction

The class of types built from Cartesian products ($\tau \times \sigma$), disjoint unions ($\tau + \sigma$) and function spaces ($\tau \to \sigma$) lies at the core of type systems for programming languages and constructive systems of Logic. How useful could a programming language be if it did not include pairing, enumeration and functions? Likewise, how useful would a constructive logic be, if it did not have conjunction, disjunction, and implication?

It is then a strange state of affairs that we still have open questions regarding general properties of this basic class of types. In this article, we revisit the problem of when two types can be considered to be isomorphic and we show that this problem can be tackled using existing results from Mathematical Logic.

Let us be more precise. The language of *polynomials with exponentiation and with positive coefficients* is defined inductively by

$$\mathcal{E} \ni f, g ::= 1 \mid x_i \mid f + g \mid fg \mid g^f,$$

where x_i is a variable for $i \in \mathbb{N}$. This language determines the class of types that we are interested in, by the simple translation,

$$\begin{aligned} [\![1]\!] &= \mathbf{1} \\ [\![x_i]\!] &= \mathbf{x_i} \\ [\![g^f]\!] &= [\![f]\!] \to [\![g]\!] \\ [\![fg]\!] &= [\![f]\!] \times [\![g]\!] \\ [\![f + g]\!] &= [\![f]\!] + [\![g]\!], \end{aligned}$$

where $\mathbf{1}$ denotes the singleton type and $\mathbf{x_i}$ denotes a type variable (that can be instantiated during an interpretation in a concrete setting). By abuse of language, we will say that a type τ belongs to $\mathcal{E}$ ($\tau \in \mathcal{E}$) when a polynomial with exponentiation $f \in \mathcal{E}$ exists such that $\tau = [\![f]\!]$. Throughout the paper, we use the plain equality symbol "$=$" to stand for identity i.e. definitional equality.

The types of $\mathcal{E}$ are inhabited by terms of a lambda calculus, in the usual way, following the typing system shown in Figure 5. The equality between two typed terms is the relation $=_{\beta\eta}$ given by the usual axioms in Figure 7.

Two types τ and σ are called *isomorphic* (notation $\tau \cong \sigma$) when there is a pair of lambda terms $\phi^{\tau\to\sigma}$ and $\psi^{\sigma\to\tau}$ that are mutually inverse, that is, $\lambda x.\phi(\psi x) =_{\beta\eta} \lambda x.x$ and $\lambda y.\psi(\phi y) =_{\beta\eta} \lambda y.y$. The importance of this notion in "practice" is as follows.

In typed programming languages, to be able to say when two types are isomorphic amounts to being able to say when two programs implement essentially the same type signature: a program of type τ can be coerced back and forth to type σ without loss of information. One can use this, for example, to search over a library of routines for a routine of a type coercible to the type needed by the programmer [5, 22].

In propositional logic, knowing that two types are isomorphic implies that the corresponding formulas (built from $\wedge, \vee$ and $\to$) are intuitionistically equivalent.

In Constructive Mathematics, type isomorphism coincides with the notion of *constructive cardinality* [17, 20] that says that two sets (i.e., types) are isomorphic if they have indistinguishable structure, which is stronger than the classical notion of cardinality relying on "number of elements".

What is known about the isomorphism of types of $\mathcal{E}$, in general, are the following facts, both proved in [8].

Theorem 1 (Soundness of HSI)**.** *If HSI* $\vdash f \doteq g$*, then* $[\![f]\!] \cong [\![g]\!]$*. In fact, we have:*

$$\textit{HSI} \vdash f \doteq g \;\Rightarrow\; [\![f]\!] \cong [\![g]\!] \;\Rightarrow\; \mathbb{N}^+ \vDash f \equiv g.$$

Theorem 2 (Martin-Wilkie-Gurevič-Fiore-DiCosmo-Balat [8, 11, 19, 28])**.** *Isomorphism is not finitely axiomatizable, that is, for no finite set of axioms T can we show that* $[\![f]\!] \cong [\![g]\!]$ *always implies* $T \vdash f \doteq g$*.*

The notation $\mathrm{T} \vdash f \doteq g$ means that there is a formal derivation of the equation $f \doteq g$ in the derivation system shown in part (b) of Figure 3, from the axioms of the set T; therefore, $\mathrm{HSI} \vdash f \doteq g$ means that the equation is derivable from the finite set of axioms shown in part (a) of Figure 3 (HSI stands for "High-School Identities", see Subsection 1.1 below).

Having only Theorem 1 and Theorem 2 is surprisingly little if we compare to what is known for the fragments of $\mathcal{E}$ that do not mix g^f and $g+f$ *simultaneously*. For those fragments, we have, as shown in [8, 25, 26], soundness *and completeness* with respect to the suitable restriction of HSI, we have moreover equivalence with truth in the standard model of positive natural numbers $\mathbb{N}^+$ (see Subsection 1.1 for definition of truth in $\mathbb{N}^+$),

$$[\![f]\!] \cong [\![g]\!] \;\Leftrightarrow\; \mathbb{N}^+ \vDash f \equiv g,$$

and, consequently, the *decidability* of $\tau \cong \sigma$ for any τ and σ of those fragments.

In this paper, we will address both the questions of completeness and decidability for $\mathcal{E}$, in simultaneous presence of g^f and $g+f$. In Section 2, we will bring up the relevance of certain subclasses of types going back to Levitz, and explain how results of Henson, Rubel, Gurevič, Richardson, and Macintyre, allow to show that type isomorphism for those subclasses are complete with respect to HSI, and decidable. In Section 3, using Wilkie's positive solution of Tarski's High-School Algebra Problem (see next subsection), we will establish the same properties for the whole of $\mathcal{E}$ (decidability is proved for base types interpreted as finite sets). In Section 4, we will mention related open problems, in particular about having efficient means of deciding the type isomorphisms.

1.1 Tarski's High-School Algebra Problem

The questions that we are interested in are related to questions regarding polynomials with exponentiation from the class $\mathcal{E}$, posed by Skolem [24] and Tarski [19] in the 1960's. Especially relevant is the question known as Tarski's High-School Algebra Problem[1]: can all equations that are true in the standard model of positive natural numbers ($\mathbb{N}^+ \vDash f \equiv g$) be derived inside the derivation system of HSI from parts (a) and (b) of Figure 3 ($\mathrm{HSI} \vdash f \doteq g$)? This is a *completeness* question.

The meaning of $\mathbb{N}^+ \vDash f \equiv g$ is the standard model theoretic one: for any replacement of the variables of f and g by elements of $\mathbb{N}^+$, one computes the same positive natural number. The converse (*soundness*),

$$\mathrm{HSI} \vdash f \doteq g \;\Rightarrow\; \mathbb{N}^+ \vDash f \equiv g,$$

can easily be proved.

Martin [19] was the first to show that, if we exclude the axioms mentioning the constant 1 from HSI, the derivation system is incomplete, since it can not derive the equality

$$(x^z + x^z)^w (y^w + y^w)^z = (x^w + x^w)^z (y^z + y^z)^w.$$

Wilkie [28] generalized Martin's equality to the whole of HSI, giving the equation

$$(A^x + B^x)^y (C^y + D^y)^x = (A^y + B^y)^x (C^x + D^x)^y, \quad (1)$$

where $A = 1+x, B = 1+x+x^2, C = 1+x^3, D = 1+x^2+x^4$. He showed (1) to be non-derivable in HSI, even though it is true in $\mathbb{N}^+$. We thus have

$$\forall f, g \in \mathcal{E}(\mathrm{HSI} \vdash f \doteq g \;\Rightarrow\; \mathbb{N}^+ \vDash f \equiv g),$$

but

$$\forall f, g \in \mathcal{E}(\mathbb{N}^+ \vDash f \equiv g \;\not\Rightarrow\; \mathrm{HSI} \vdash f \doteq g),$$

which constitutes a *negative* solution to Tarski's original question.

Gurevič [11] further showed that one can not "repair" HSI by extending it with *any* finite list of axioms. He generalized Wilkie's (1) to the infinite sequence of equations

$$(A^{2^x} + B_n^{2^x})^x (C_n^x + D_n^x)^{2^x} = (A^x + B_n^x)^{2^x} (C_n^{2^x} + D_n^{2^x})^x, \; (G_n)$$

where $A = x+1, B_n = 1 + x + x^2 + \cdots + x^{n-1}, C_n = 1 + x^n, D_n = 1 + x^2 + x^4 + \cdots + x^{2(n-1)}$, and showed that for any finite extension T of HSI there is an odd $n > 3$ such that T can not prove the equality (G_n) although $\mathbb{N}^+ \vDash G_n$.

Fiore, Di Cosmo and Balat [8] showed that Gurevič's equations can be interpreted as type isomorphisms, establishing the mentioned Theorem 2.

1.2 Decidability of Arithmetic Equality for $\mathcal{E}$

A separate question of more general interest is that of the decidability of equality between polynomials with exponentiation, that is, whether there is a procedure for deciding if $\mathbb{N}^+ \vDash f \equiv g$ holds or not, for any $f, g \in \mathcal{E}$. It was first addressed by Richardson [21], who proved decidability for the univariate case (expressions of $\mathcal{E}$ in one variable). Later, Macintyre [18] showed the decidability for the multivariate case i.e. for the whole of $\mathcal{E}$.

Theorem 3 (Richardson-Macintyre [18, 21])**.** *There is a recursive procedure that decides, for any $f, g \in \mathcal{E}$, whether $\mathbb{N}^+ \vDash f \equiv g$ holds or not.*

However, we cannot use the decidability result for $\mathbb{N}^+$ to conclude decidability of type isomorphisms for $\mathcal{E}$, because, although we do have that (by Theorem 1)

$$\mathrm{HSI} \vdash f \doteq g \;\Rightarrow\; [\![f]\!] \cong [\![g]\!] \;\Rightarrow\; \mathbb{N}^+ \vDash f \equiv g,$$

a proof of

$$[\![f]\!] \cong [\![g]\!] \;\Leftarrow\; \mathbb{N}^+ \vDash f \equiv g$$

is not known, and HSI is not complete:

$$\mathrm{HSI} \vdash f \doteq g \;\not\Leftarrow\; \mathbb{N}^+ \vDash f \equiv g.$$

2. Subclasses of $\mathcal{E}$ Complete for HSI

One of the things that has not been exploited in the literature on type isomorphism is the line of research on *subclasses* of polynomials with exponentiation for which the axioms of HSI *are* complete.

In [16], while studying the relation of eventual dominance for polynomials with exponentiation, Levitz isolated the class of expressions in one variable, built by the inductive definition

$$\mathcal{S} \ni f, g ::= 1 \mid x \mid f + g \mid fg \mid x^f \mid n^f,$$

where n is a numeral. Henson and Rubel [14] extended it to the multivariate class defined by

$$\mathcal{L}(S) \ni f, g ::= s \mid x_i \mid f + g \mid fg \mid s'^f \mid x_i^f \mid (x_i^{s'})^f,$$

where S is an arbitrary set of positive real constants, $s, s' \in S$, $s' > 1$, and they proved all true equalities between expressions from $\mathcal{L}(S)$ to be derivable from HSI. They also conjectured that the result could be extended to the class defined by

$$\mathcal{R}(S) \ni f, g ::= s \mid x_i \mid f + g \mid fg \mid p^f,$$

where p is an ordinary polynomial with coefficients in S, and they remarked that Wilkie's counterexample lies "just outside" the class $\mathcal{R}(S)$.

[1] For various results around this problem, please look at the survey articles [2] and [3].

$$f \doteq f$$
$$f + g \doteq g + f$$
$$(f + g) + h \doteq f + (g + h)$$
$$fg \doteq gf$$
$$(fg)h \doteq f(gh)$$
$$f(g + h) \doteq fg + fh$$
$$1f \doteq f$$
$$f^1 \doteq f$$
$$1^f \doteq 1$$
$$f^{g+h} \doteq f^g f^h$$
$$(fg)^h \doteq f^h g^h$$
$$(f^g)^h \doteq f^{gh}$$

(a) Axioms of HSI

$$\frac{f \doteq g}{g \doteq f} \qquad \frac{f \doteq g \quad g \doteq h}{f \doteq h} \qquad \frac{f_1 \doteq g_1 \quad f_2 \doteq g_2}{f_1^{f_2} \doteq g_1^{g_2}}$$

$$\frac{f_1 \doteq g_1 \quad f_2 \doteq g_2}{f_1 + f_2 \doteq g_1 + g_2} \qquad \frac{f_1 \doteq g_1 \quad f_2 \doteq g_2}{f_1 f_2 \doteq g_1 g_2}$$

(a) Equality and congruence rules

$$t_1 \doteq 1$$
$$t_{x_i} \doteq x_i$$
$$t_{zu} \doteq t_z t_u$$
$$t_{z+u} \doteq t_z + t_u$$
$$t_z \doteq t_u \qquad (\text{when } \mathbb{N}^+ \vDash z \equiv u)$$

(b) Additional axioms of HSI*

Notation: f, g, h denote polynomials with exponentiation (possibly containing t_z terms in the congruence and equality rules), while z, u denote ordinary polynomials (i.e., without non-constant exponents) with possibly negative monomial coefficients.

Figure 3: The derivation system of (Extended) High-School Identities

Finally, Gurevič [12] showed that HSI are complete for the proper extension $\mathcal{L}$ of $\mathcal{R}(S)$ defined by[2]

$$\mathcal{L} \ni f, g ::= s \mid x_i \mid f + g \mid fg \mid l^f,$$

where $l \in \Lambda$ is defined by

$$\Lambda \ni f, g ::= s \mid x_i \mid f + g \mid fg \mid l_0^f,$$

and $l_0 \in \Lambda$ has no variables.

Theorem 4 (Levitz-Henson-Rubel-Gurevič [12, 14, 16])**.** *For all $f, g \in \mathcal{L}$,*

$$\mathbb{N}^+ \vDash f \equiv g \;\Rightarrow\; \mathit{HSI} \vdash f \doteq g.$$

For our purposes, it suffices to take $S = \{1\}$, and we will henceforth use $\mathcal{L}$ specialized to this S.

Example 1. *Wilkie's equation (1) deals with terms that do not belong to the class $\mathcal{L}$. Although $A, B, C, D \in \Lambda \subset \mathcal{L}$, and hence $A^x, B^x, C^x, D^x, A^x + B^x, C^x + D^x \in \mathcal{L} \setminus \Lambda$, we have $(A^x + B^x)^y, (C^x + D^x)^y \notin \mathcal{L}$, because bases of exponentiation are not allowed to contain bases of exponentiation that contain variables.*

Example 2. *The term*[3]

$$(y + z)^{x(y+z)(y+z)^{x(y+z)}} \in \mathcal{L}, \tag{2}$$

but

$$\left(((y + z)^x)^{((y+z)^{y+z})^x}\right)^{y+z} \notin \mathcal{L}, \tag{3}$$

although the two terms are inter-derivable using the HSI axioms. This means that, even though HSI is complete for $\mathcal{L}$, there is room for extension of $\mathcal{L}$ to subclasses that are still finitely axiomatizable by HSI. In other words, $\mathcal{L}$ is not closed under HSI-derivability.

[2] We stick to the original notations $\mathcal{L}(S)$ and $\mathcal{L}$, although the latter also depends on the set S and we have $\mathcal{L}(S) \subsetneq \mathcal{R}(S) \subsetneq \mathcal{L}$.

[3] These terms correspond to simply typed versions of an induction axiom for decidable predicates, in "curried" and "uncurried" variant.

Theorems 1, 3 and 4 allow us to conclude the following.

Corollary 1 (Completeness for $\mathcal{L}$)**.** *For all $f, g \in \mathcal{L}$, if $[\![f]\!] \cong [\![g]\!]$ then HSI $\vdash f \doteq g$.*

Corollary 2 (Decidability for $\mathcal{L}$)**.** *There is an algorithm that decides, for all $f, g \in \mathcal{L}$, whether $[\![f]\!] \cong [\![g]\!]$ or not.*

3. The Extended High-School Identities

As explained in Subsection 1.1, Wilkie's negative solution of Tarski's problem, together with the generalization of Gurevič, was used by Fiore, Di Cosmo, and Balat, to show the incompleteness of HSI, and the impossibility of a finite axiomatization, for type isomorphism over $\mathcal{E}$.

However, in the paper [28], a *positive* solution to Tarski's problem was also given. Namely, Wilkie showed that HSI is almost complete: by extending it with all equations that hold between ordinary positive polynomials[4] — that is, positive polynomials without exponents containing variables, but possibly with negative monomial coefficients — one obtains an axiomatization (HSI* from Figure 3) which is complete for all true equations in $\mathbb{N}^+$ between expressions of $\mathcal{E}$. Since equality of ordinary (positive) polynomials is decidable, we have a recursive procedure for determining if an equation belongs to the set of axioms, and therefore equality between terms of $\mathcal{E}$ — although not finitely axiomatizable — is recursively axiomatizable.

To be more precise, let $\mathcal{E}^*$ be the language extending $\mathcal{E}$ with a constant t_z for every ordinary positive polynomial z:

$$\mathcal{E}^* \ni f, g ::= t_z \mid 1 \mid x_i \mid g^f \mid fg \mid f + g.$$

The system of axioms of HSI* is the extension of the system HSI — that applies only to expressions of the original language $\mathcal{E}$ — with the axioms given in part (c) of Figure 3 — that apply to a strict subset of the extended language $\mathcal{E}^*$. The two types of axioms

[4] A polynomial is positive if it computes to a positive natural number for every replacement of variables by positive natural numbers.

can "interact" through the derivation system of part (b) of Figure 3, thus making possible equalities between expressions of the full language $\mathcal{E}^*$.

The left-hand side of the new axioms is always an expression of form t_z, while the right hand side can either be an expression t_u — whenever z and u are equal polynomials (which can be decided by bringing them into canonical form) — or an expression reflecting the structure of z when possible: when z has no negative coefficients (we use in that case letters p, q instead of z, u), we can fully reflect it into the language, that is, one can prove HSI* $\vdash t_p \doteq p$. This representation of the axioms is inspired from the one of Asatryan [1]. Note that, for any t_z, there are infinitely many axioms having t_z on the left hand side that can be used.

We can now state Wilkie's result.

Theorem 5 (Completeness of HSI* (Wilkie [28]))**.** *For all $f, g \in \mathcal{E}$ (that is, all f, g of $\mathcal{E}^*$ that do **not** contain t_z-symbols), we have that $\mathbb{N}^+ \vDash f \equiv g$ implies HSI* $\vdash f \doteq g$.*

This is a statement concerning terms of $\mathcal{E}$ (the original language), in the proof of which terms of $\mathcal{E}^*$ (the extended language) are used. In this respect, it is reminiscent of meta-mathematical statements like Hilbert's ϵ-elimination theorems [15] or Henkin's version of Gödel's completeness theorem [13].

Using theorems 1 and 5, we immediately obtain the completeness of HSI* for type isomorphism over $\mathcal{E}$.

Corollary 3. *Given $f, g \in \mathcal{E}$ such that $[\![f]\!] \cong [\![g]\!]$, we have that HSI* $\vdash f \doteq g$.*

We now move on to the decidability question. We will show that derivations of HSI* can be interpreted as type isomorphisms, which suffices, since then the circuit

$$[\![f]\!] \cong [\![g]\!] \;\Rightarrow\; \mathbb{N}^+ \vDash f \equiv g \;\Rightarrow\; \text{HSI*} \vdash f \doteq g \;\Rightarrow\; [\![f]\!] \cong [\![g]\!]$$

allows one to use Macintyre's decidability results for $\mathbb{N}^+$ (Theorem 3) to conclude decidability of type isomorphism over $\mathcal{E}$.

At first thought, interpreting the new t_z symbols might seem problematic, since negative monomial coefficients in z would imply the use of some kind of negative types. However, we also have the additional property that z is positive, which means that if we instantiate its variables with positive natural numbers (positive types), we will obtain a positive natural number (positive type).

In this paper, we will work with the restriction that base types are finite sets. Although the method does work for base types isomorphic to ordinals in Cantor normal form[5], that requires a careful constructive treatment of ordinals beyond the scope of this paper.

For HSI, one can keep the interpretation of base types implicit: soundness of HSI equations as type isomorphisms is proved uniformly, regardless of the actual interpretations of base types. For HSI*, we will need to be explicit about interpretation, that is, we will prove the soundness theorem *point-wise*. We will thus introduce an explicit environment ρ mapping variables to types and extend the interpretation $[\![\cdot]\!]$ for the extra t_z-terms.

$$\begin{aligned}
[\![1]\!]_\rho &= \mathbf{1}\\
[\![x_i]\!]_\rho &= \rho(x_i)\\
[\![g^f]\!]_\rho &= [\![f]\!]_\rho \to [\![g]\!]_\rho\\
[\![fg]\!]_\rho &= [\![f]\!]_\rho \times [\![g]\!]_\rho\\
[\![f+g]\!]_\rho &= [\![f]\!]_\rho + [\![g]\!]_\rho\\
[\![t_z]\!]_\rho &= \underbrace{1+1+\cdots+1}_{k\text{-times}} = \mathbf{k} \quad \text{where } k = \mathrm{eval}(t_z, \rho)
\end{aligned}$$

The number $\mathrm{eval}(t_z, \rho)$ is the result of evaluating z for the variables interpreted in ρ by positive natural numbers. We also denote the type

$$\underbrace{1+1+\cdots+1}_{k\text{-times}}$$

with bold-face $\mathbf{k}$.

Theorem 6. *Let $f, g \in \mathcal{E}^*$. If HSI* $\vdash f \doteq g$ then $[\![f]\!]_\rho \cong [\![g]\!]_\rho$ for any ρ that interprets variables by types of form $\mathbf{k}$.*

Proof. The proof is by induction on the derivation. We first give explicit isomorphisms for the axioms of HSI:

- $f \doteq f$ is interpreted with the identity lambda term $\lambda x.x$ in both directions;
- $f + g \doteq g + f$ is interpreted by $\lambda x.\delta[x|x_1.\iota_2 x_1|x_2.\iota_2 x_2]$ in both directions;
- $(f + g) + h \doteq f + (g + h)$ is interpreted by

  $$\delta[x|x_1.\iota_1\iota_1 x|x_2.\delta[x_2|x_{21}.\iota_1\iota_2 x_2|x_{22}.\iota_2 x_{22}]]$$

  and

  $$\delta[x|x_1.\delta[x_1|x_{11}.\iota_1 x_{11}|x_{12}.\iota_2\iota_1 x_{12}]|x_2.\iota_2\iota_2 x_2];$$
- $fg \doteq gf$ is interpreted by $\lambda x.\langle \pi_2 x, \pi_1 x\rangle$ in both directions;
- $(fg)h \doteq f(gh)$ is interpreted by

  $$\lambda x.\langle\langle \pi_1 x, \pi_1\pi_2 x\rangle, \pi_2\pi_2 x\rangle$$

  and

  $$\lambda x.\langle \pi_1\pi_1 x, \langle \pi_2\pi_1 x, \pi_2 x\rangle\rangle;$$
- $f(g + h) \doteq fg + fh$ is interpreted by

  $$\lambda x.\delta[\pi_2 x|x_1.\iota_1\langle \pi_1 x, x_1\rangle|x_2.\iota_2\langle \pi_1 x_1, \iota_2\pi_2 x_1\rangle]$$

  and

  $$\lambda x.\delta[x|x_1.\langle \pi_1 x_1, \iota_1\pi_2 x_1\rangle|x_2.\langle \pi_1 x_2, \iota_2\pi_2 x_2\rangle];$$
- $f1 \doteq f$ is interpreted by $\lambda x.\pi_1 x$ and $\lambda x.\langle x, \star\rangle$;
- $f^1 \doteq f$ is interpreted by $\lambda x.x\star$ and $\lambda xy.x$;
- $1^f \doteq 1$ is interpreted by $\lambda x.\star$ and $\lambda xy.\star$;
- $f^{g+h} \doteq f^g f^h$ is interpreted by

  $$\lambda x.\langle \lambda y.x(\iota_1 y), \lambda y.x(\iota_2 y)\rangle$$

  and

  $$\lambda xy.\delta[y|y_1.(\pi_1 x)y_1|y_2.(\pi_2 x)y_2];$$
- $(fg)^h \doteq f^h g^h$ is interpreted by $\lambda x.\langle \lambda y.\pi_1 xy, \lambda y.\pi_2 xy\rangle$ and $\lambda xy.\langle(\pi_1 x)y, (\pi_2 x)y\rangle$;
- $(f^g)^h \doteq f^{gh}$ is interpreted by $\lambda xy.x(\pi_1 y)(\pi_2 y)$ and $\lambda xyz.x\langle z, y\rangle$.

The congruence and equality rules are handled using the induction hypotheses:

[5] Subtraction $\alpha - \beta$ between two such ordinals can be defined when we know that $\alpha < \beta$. Since we can always rewrite t_z as $t_p - t_q$, and we know $p < q$, an ordinal in Cantor normal form can always be computed for t_z whenever ones for p and q are given.

$$\Lambda^+ \ni M, N, P ::= x \mid \star \mid \lambda x.M \mid MN \mid \iota_1 M \mid \iota_2 M \mid \delta[M|x.N|x.P] \mid \langle M, N\rangle \mid \pi_1 M \mid \pi_2 M$$

(a) Raw language of lambda terms

$$\frac{x^\tau \in \Gamma}{\Gamma \Vdash x^\tau} \qquad \frac{}{\Gamma \Vdash \star^{\mathbf{1}}} \qquad \frac{\Gamma \cup x^\tau \Vdash M^\sigma}{\Gamma \Vdash (\lambda x.M)^{\tau\to\sigma}} \qquad \frac{\Gamma \Vdash M^{\tau\to\sigma} \quad \Gamma \Vdash N^\tau}{\Gamma \Vdash (MN)^\sigma}$$

$$\frac{\Gamma \Vdash M^\tau}{\Gamma \Vdash (\iota_1 M)^{\tau+\sigma}} \qquad \frac{\Gamma \Vdash M^\sigma}{\Gamma \Vdash (\iota_2 M)^{\tau+\sigma}} \qquad \frac{\Gamma \Vdash M^{\tau+\sigma} \quad \Gamma \cup x^\tau \Vdash N^\rho \quad \Gamma \cup x^\sigma \Vdash P^\rho}{\Gamma \Vdash \delta[M|x.N|x.P]^\rho}$$

$$\frac{\Gamma \Vdash M^\sigma \quad \Gamma \Vdash N^\tau}{\Gamma \Vdash \langle M, N\rangle^{\sigma\times\tau}} \qquad \frac{\Gamma \Vdash M^{\sigma\times\tau}}{\Gamma \Vdash (\pi_1 M)^\sigma} \qquad \frac{\Gamma \Vdash M^{\sigma\times\tau}}{\Gamma \Vdash (\pi_2 M)^\tau}$$

(b) Typing system (well-formed lambda terms)

Figure 5: Inhabitation of types of $\mathcal{E}$ with lambda terms

$$\begin{aligned}
M &=_{\beta\eta} \star && \text{for any term } M \text{ of type } \mathbf{1}\\
(\lambda x.M)N &=_{\beta\eta} M\{N/x\} && \text{where } x \text{ is not a variable of } N\\
M &=_{\beta\eta} \lambda x.Mx && \text{where } x \text{ is not a variable of } M\\
\delta[\iota_1 M|x.N|y.P] &=_{\beta\eta} N\{M/x\} && \text{where } x \text{ is not a variable of } M\\
\delta[\iota_2 M|x.N|y.P] &=_{\beta\eta} P\{M/y\} && \text{where } y \text{ is not a variable of } M\\
N\{M/z\} &=_{\beta\eta} \delta[M|x.N\{\iota_1 x/z\}|y.N\{\iota_2 y/z\}] && \text{where } z \text{ is not a variable of } M\\
\pi_1\langle M, N\rangle &=_{\beta\eta} M\\
\pi_2\langle M, N\rangle &=_{\beta\eta} N\\
M &=_{\beta\eta} \langle \pi_1 M, \pi_2 M\rangle
\end{aligned}$$

The full relation $=_{\beta\eta}$ is the reflexive, symmetric, transitive, and congruent closure of the above.

Figure 7: Beta-eta equality between lambda terms of the same type

- Given an interpretation of $f \doteq g$, i.e. $\Phi : [\![f]\!]_\rho \to [\![g]\!]_\rho$ and $\Psi : [\![g]\!]_\rho \to [\![f]\!]_\rho$ such that
$$\lambda x.\Phi(\Psi x) =_{\beta\eta} \lambda x.x$$
$$\lambda y.\Psi(\Phi y) =_{\beta\eta} \lambda y.y,$$
we just swap the order of the two equations in order to interpret $g \doteq f$.
- Given interpretations of $f \doteq g$ and $g \doteq h$ by four terms
$$\Phi_1 : [\![f]\!]_\rho \to [\![g]\!]_\rho \qquad \Phi_2 : [\![g]\!]_\rho \to [\![h]\!]_\rho$$
$$\Psi_1 : [\![g]\!]_\rho \to [\![f]\!]_\rho \qquad \Psi_2 : [\![h]\!]_\rho \to [\![g]\!]_\rho,$$
we interpret $f \doteq h$ by composing Φ_1 and Φ_2, and Ψ_1 and Ψ_2.
- Given interpretations of $f_1 \doteq g_1$ and $f_2 \doteq g_2$ by four terms
$$\Phi_1 : [\![f_1]\!]_\rho \to [\![g_1]\!]_\rho \qquad \Phi_2 : [\![f_2]\!]_\rho \to [\![g_2]\!]_\rho$$
$$\Psi_1 : [\![g_1]\!]_\rho \to [\![f_1]\!]_\rho \qquad \Psi_2 : [\![g_2]\!]_\rho \to [\![f_2]\!]_\rho,$$
we interpret $f_1^{f_2} \doteq g_1^{g_2}$ by using the terms
$$\Phi = \lambda x.\lambda y.\Phi_1(x(\Psi_2 y))$$
$$\Psi = \lambda x.\lambda y.\Psi_1(x(\Phi_2 y)).$$
The fact that Φ and Ψ are mutually inverse w.r.t. $=_{\beta\eta}$ is proved by using the η-axiom
$$\lambda x.\lambda y.xy =_{\beta\eta} \lambda x.x.$$
- Given interpretations of $f_1 \doteq g_1$ and $f_2 \doteq g_2$ by four terms
$$\Phi_1 : [\![f_1]\!]_\rho \to [\![g_1]\!]_\rho \qquad \Phi_2 : [\![f_2]\!]_\rho \to [\![g_2]\!]_\rho$$
$$\Psi_1 : [\![g_1]\!]_\rho \to [\![f_1]\!]_\rho \qquad \Psi_2 : [\![g_2]\!]_\rho \to [\![f_2]\!]_\rho,$$
we interpret $f_1 f_2 \doteq g_1 g_2$ by using the terms
$$\Phi = \lambda x.\langle \Phi_1(\pi_1 x), \Phi_2(\pi_2 x)\rangle$$
$$\Psi = \lambda y.\langle \Psi_1(\pi_1 y), \Psi_2(\pi_2 y)\rangle.$$
The fact that Φ and Ψ are mutually inverse w.r.t. $=_{\beta\eta}$ is proved by using the η-axiom
$$\lambda y.\langle \pi_1 y, \pi_2 y\rangle =_{\beta\eta} \lambda y.y.$$
- Given interpretations of $f_1 \doteq g_1$ and $f_2 \doteq g_2$ by four terms
$$\Phi_1 : [\![f_1]\!]_\rho \to [\![g_1]\!]_\rho \qquad \Phi_2 : [\![f_2]\!]_\rho \to [\![g_2]\!]_\rho$$
$$\Psi_1 : [\![g_1]\!]_\rho \to [\![f_1]\!]_\rho \qquad \Psi_2 : [\![g_2]\!]_\rho \to [\![f_2]\!]_\rho,$$
we interpret $f_1 + f_2 \doteq g_1 + g_2$ by using the terms
$$\Phi = \lambda x.\delta[x|x_1.\iota_1(\Phi_1 x_1)|x_2.\iota_2(\Phi_2 x_2)]$$
$$\Psi = \lambda y.\delta[y|y_1.\iota_1(\Psi_1 y_1)|y_2.\iota_2(\Psi_2 y_2)].$$

The fact that Φ and Ψ are mutually inverse w.r.t. $=_{\beta\eta}$ is proved by using the η-axiom for sums twice. Once we use it with

$$\begin{aligned} M :=& y \\ N :=& \delta[\big(\delta[z|y_1.\iota_1(\Psi_1 y_1)|y_2.\iota_2(\Psi_2 y_2)]\big) \\ & |x_1.\iota_1(\Phi_1 x_1) \\ & |x_2.\iota_2(\Phi_2 x_2)], \end{aligned}$$

and the second time with $M := y$, $N := z$.

It remains to interpret the rest of the axioms of HSI*, those that involve t_z-terms.

- $t_1 \doteq 1$ is interpreted as the isomorphism $\mathbf{1} \cong [\![1]\!]_\rho$ i.e. $\mathbf{1} \cong \mathbf{1}$ using the lambda term $\lambda x.x$ in both directions;
- $t_{x_i} \doteq x_i$ is interpreted as $\mathbf{k} \cong \mathbf{k}$, for $k = \mathrm{eval}(x_i, \rho)$, by the lambda term $\lambda x.x$ in both directions;
- $t_{zu} \doteq t_z t_u$ is interpreted as $\mathbf{k} \cong \mathbf{k_1} \times \mathbf{k_2}$, for $k = \mathrm{eval}(t_{zu}, \rho)$, $k_1 = \mathrm{eval}(t_z, \rho)$, and $k_2 = \mathrm{eval}(t_u, \rho)$, by the lambda term of Lemma 1;
- $t_{z+u} \doteq t_z + t_u$ is interpreted as $\mathbf{k} \cong \mathbf{k_1} + \mathbf{k_2}$, for $k = \mathrm{eval}(t_{zu}, \rho)$, $k_1 = \mathrm{eval}(t_z, \rho)$, and $k_2 = \mathrm{eval}(t_u, \rho)$, by the lambda term of Lemma 1;
- $t_z \doteq t_u$ is interpreted as $\mathbf{k} \cong \mathbf{k}$, for $k = \mathrm{eval}(z, \rho) = \mathrm{eval}(u, \rho)$, by the lambda term $\lambda x.x$ in both directions.

□

Lemma 1. *Let $p \in \mathcal{E}^*$ be an ordinary polynomial with positive coefficients (possibly containing t_z-terms), $k \in \mathbb{N}^+$, and ρ be an interpretation such that $k = \mathrm{eval}(p, \rho)$. Then, $\mathbf{k} \cong [\![p]\!]_\rho$.*

Proof. We do induction on p.

- When $p = 1$, we have $\mathrm{eval}(1, \rho) = 1 = k$, so $\mathbf{k} = \mathbf{1} \cong \mathbf{1} = [\![1]\!]_\rho$ is established by using $\lambda x.x$ in both directions.
- When $p = x_i$, we have $\mathrm{eval}(x_i, \rho) = \rho(x_i) = k$, so $\mathbf{k} \cong [\![x_i]\!]_\rho$ by $\lambda x.x$ in both directions.
- When $p = t_z$, we have $\mathrm{eval}(t_z, \rho) = k$, so $\mathbf{k} \cong [\![t_z]\!]_\rho$ by $\lambda x.x$ in both directions.
- When $p = p_1 + p_2$, we have

  $$\begin{aligned} k = \mathrm{eval}(p, \rho) = \mathrm{eval}(p_1 + p_2, \rho) = \\ = \mathrm{eval}(p_1, \rho) + \mathrm{eval}(p_2, \rho) = k_1 + k_2 \end{aligned}$$

  for some $k_1, k_2 \in \mathbb{N}^+$. By applying the induction hypothesis twice, we obtain $\mathbf{k_1} \cong [\![p_1]\!]_\rho$ and $\mathbf{k_2} \cong [\![p_2]\!]_\rho$, therefore,

  $$\mathbf{k} \cong \mathbf{k_1} + \mathbf{k_2} \cong [\![p_1]\!]_\rho + [\![p_2]\!]_\rho = [\![p]\!]_\rho$$

  using the obvious isomorphism $\mathbf{k} \cong \mathbf{k_1} + \mathbf{k_2}$, that holds when $k = k_1 + k_2$ and the lambda terms interpreting congurence for "+" from the proof of the previous theorem.
- When $p = p_1 p_2$, we have

  $$\begin{aligned} k = \mathrm{eval}(p, \rho) = \mathrm{eval}(p_1 p_2, \rho) = \\ = \mathrm{eval}(p_1, \rho)\,\mathrm{eval}(p_2, \rho) = k_1 k_2 \end{aligned}$$

  for some $k_1, k_2 \in \mathbb{N}^+$. By applying the induction hypothesis twice, we obtain $\mathbf{k_1} \cong [\![p_1]\!]_\rho$ and $\mathbf{k_2} \cong [\![p_2]\!]_\rho$, therefore,

  $$\mathbf{k} \cong \mathbf{k_1} \times \mathbf{k_2} \cong [\![p_1]\!]_\rho \times [\![p_2]\!]_\rho = [\![p]\!]_\rho$$

  using the obvious isomorphism $\mathbf{k} \cong \mathbf{k_1} + \mathbf{k_2}$ that holds when $k = k_1 k_2$, and the lambda terms interpreting congurence for "$\times$" from the proof of the previous theorem.

□

Corollary 4. *Given two types $f, g \in \mathcal{E}$, one can decide whether $[\![f]\!]_\rho \cong [\![g]\!]_\rho$ or not, and this holds whenever ρ interprets variable by types of form* $\mathbf{k}$.

4. Conclusion

We showed that existing results from Mathematical Logic allow us to conclude that type isomorphism over $\mathcal{E}$ is recursively axiomatizable, and that a subclass $\mathcal{L}$ of types can be isolated for which type isomorphism is even finitely axiomatizable by the well known High-School Identities and decidable. Our Theorem 6 allows us to conclude decidability for the whole of $\mathcal{E}$ when base types are finite sets.

These results also apply to questions of cardinality of sets in Constructive Mathematics, and to isomorphism of objects in the corresponding category. However, further work is needed to understand fully their implications in practice.

4.1 Future Work (Open Questions)

4.1.1 Extensions and Practical Importance of the Levitz Class

We saw that the class $\mathcal{L}$ of Gurevič is a generalization of the classes $\mathcal{R}(S)$ and $\mathcal{L}(S)$ of Henson and Rubel, which are in turn generalizations of Levitz's class $\mathcal{S}$. We also saw in Example 1 that there are two HSI-equal types, one of which is in $\mathcal{L}$ while the other is not.

Therefore, it does not seem unlikely that the class $\mathcal{L}$ can be further extended. For example, cannot we allow the bases of exponentiation to contain variables in their bases of exponentiation up to a fixed (but arbitrary) height n? Would not such a theory also be finitely axiomatizable by HSI?

Another interesting thing to investigate would be what the practical interest of these subclasses is. For example, how many programs of a standard library for functional programming or theorem proving would fall outside (extensions of) $\mathcal{L}$?

4.1.2 Simpler Completeness Proof for HSI*

Wilkie's proof of completeness of HSI* relies on two components.

In Theorem 2.8 of [28], it is shown that each polynomial with exponentiation f can be proved to be equal in HSI* to a positive polynomial with positive coefficients, but with extra variables, some of which are instantiated with witnessing terms τ_i of the form $p_i^{q_i}$. The proof of this theorem proceeds by induction on the construction of f, the difficult case being when $f = f_1^{f_2}$; here, the induction hypothesis is used together with the fact that each positive polynomial can be factored as a monomial and irreducible polynomial. In fact, an enumeration of all possible pairs $\langle$ irreducible , monomial $\rangle$ with the right properties is used, and then, when constructing the representative for f one need only look up in the enumeration. A large number of extra variables will generally be added to the representing "polynomial".

The second component of Wilkie's proof uses Differential Algebra to show that the representation from Theorem 2.8 is unique.

If we could obtain a simpler version of Wilkie's proof of Theorem 2.8, that avoids an ad hoc enumeration, it would be easier to interpret Corollary 3 as a program that given a concrete proof of $[\![f]\!] \cong [\![g]\!]$ builds a concrete derivation of HSI* $\vdash f \doteq g$.

4.1.3 Efficient Decision Procedures

It is not clear what is the computational complexity of Macintyre's decision procedute [18], although some authors suggests that it is exponential [6].

On the other hand, we know that equality between ordinary polynomials can be decided in $O(n \log(n))$ [23], and in the context

of type isomorphisms decision algorithms with similar complexity have been given by Considine, Gil and Zibin [4, 9, 10].

Is to possible to obtain a practical decision algorithm for type isomorphism over $\mathcal{E}$? Is it at least possible to do so for some subclass of terms like $\mathcal{L}$?

4.2 Other Related Work

Although there is a rich literature on type isomorphisms in the absence of sum types (see [5, 6] for a survey), the only work covering sums types that we are aware of is the mentioned article of Fiore, Di Cosmo, and Balat, [8].

Di Cosmo and Dufour consider the extension of Tarski's original question to the structure $\mathbb{N}^+ \cup \{0\}$, for which they manage to show that decidability and non-finite-axiomatizability still hold. It is not clear what the implications for type isomorphism are in the presence of both sums and the empty type, since it is known that there are true arithmetical equalities concerning 0 that do not hold as type isomorphisms when 0 is interpreted as the empty type (an observation attributed to Alex Simpson in [8]).

Soloviev [27], and later Došen and Petrić [7], show type isomorphism in the context of symmetric monoidal closed categories is finitely axiomatizable and decidable.

Acknowledgments

I would like to thank Olivier Danvy for being an excellent host at Aarhus University, where this project finally took off. I would also like to thank Alex Simpson for noticing an error at a workshop talk that I gave on the subject, and Gurgen Asatryan for sending me a copy of a hard-to-find paper.

This work was mainly supported by a Kurt Gödel Research Prize Fellowship 2011. Final changes were funded by the ERC Advanced Grant ProofCert.

References

[1] G. Asatryan. On models of exponentiation. identities in the HSI-algebra of posets. *Mathematical Logic Quarterly*, 54(3):280–287, 2008.

[2] S. Burris and S. Lee. Tarski's high school identities. *American Mathematical Monthly*, 100:231–236, 1993.

[3] S. N. Burris and K. A. Yeats. The saga of the high school identities. *Algebra Universalis*, 52:325–342, 2004.

[4] J. Considine. Deciding isomorphisms of simple types in polynomial time. Technical report, Boston University Computer Science Department, 2000.

[5] R. D. Cosmo. A short survey of isomorphisms of types. *Mathematical Structures in Computer Science*, 15:825–838, 2005.

[6] R. D. Cosmo and T. Dufour. The equational theory of $\langle n, 0, 1, +, \times, \uparrow \rangle$ is decidable, but not finitely axiomatisable. In *Logic for Programming, Artificial Intelligence, and Reasoning*, pages 240–256. Springer-Verlag Berlin Heildelberg, 2005.

[7] K. Došen and Z. Petrić. Isomorphic objects in symmetric monoidal closed categories. *Mathematical Structures in Computer Science*, 7 (6):639–662, Dec. 1997. ISSN 0960-1295.

[8] M. Fiore, R. D. Cosmo, and V. Balat. Remarks on isomorphisms in typed lambda calculi with empty and sum types. *Annals of Pure and Applied Logic*, 141:35–50, 2006.

[9] J. Gil and Y. Zibin. Efficient algorithms for isomorphisms of simple types. *Mathematical Structures in Computer Science*, 15(5):917–957, 2005.

[10] J. Gil and Y. Zibin. Randomised algorithms for isomorphisms of simple types. *Mathematical Structures in Computer Science*, 17(3): 565–584, 2007.

[11] R. H. Gurevič. Equational theory of positive numbers with exponentiation is not finitely axiomatizable. *Annals of Pure and Applied Logic*, 49:1–30, 1990.

[12] R. H. Gurevič. Detecting algebraic (in)dependence of explicitly presented functions (Some applications of Nevalinna theory to mathematical logic. *Transactions of the American Mathematical Society*, 336 (1):1–67, 1993.

[13] L. Henkin. The completeness of the first-order functional calculus. *The Journal of Symbolic Logic*, 14(3):159–166, 1949.

[14] C. W. Henson and L. A. Rubel. Some application of Nevanlinna theory to mathematical logic: Identities of exponential functions. *Transactions of the American Mathematical Society*, 282(1), March 1984.

[15] D. Hilbert and P. Bernays. *Fondements des mathmatiques*. Paris ; Budapest ; Turin : l'Harmattan, 2001. French translation of Grundlagen der Mathematik, 2nd edition, 1970.

[16] H. Levitz. An ordered set of arithmetic functions representing the least ϵ number. *Zeitschrift fr mathematische Logik und Grundlagen der Mathematik*, 21:115–120, 1975.

[17] H. Lombardi and C. Quitt. *Algbre commutative – Mthodes constructives*. Calvage & Mounet, Paris, 2011.

[18] A. Macintyre. *Model Theory and Arithmetic*, volume 890 of *Lecture Notes in Mathematics*, chapter The laws of exponentiation, pages 185–197. Springer Berlin Heidelberg, 1981.

[19] C. F. Martin. *Equational theories of natural numbers and transfinite ordinals*. PhD thesis, University of California, Berkeley, 1973.

[20] R. Mines and F. Richman. *A course in constructive algebra*. Springer, 1988.

[21] D. Richardson. Solution of the identity problem for integral exponential functions. *Zeitschrift fr mathematische Logik und Grundlagen der Mathematik*, 15:333–340, 1969.

[22] M. Rittri. Using types as search keys in function libraries. *Journal of Functional Programming*, 1(1):71–89, 1991.

[23] J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *Journal of the ACM*, 27(4):701–717, Oct. 1980.

[24] T. Skolem. An ordered set of arithmetic functions representing the least ϵ-number. *Det Konigelige Norske Videnskabers Selskab Fordhandlinger*, 29:54–59, 1956.

[25] S. Soloviev. The category of finite sets and Cartesian closed categories. In *Theoretical application of methods of mathematical logic. Part III*, volume 105 of *Zap. Nauchn. Sem. LOMI*, pages 174–194. "Nauka", Leningrad. Otdel., 1981.

[26] S. Soloviev. The category of finite sets and Cartesian closed categories. *Journal of Soviet Mathematics*, 22(3):1387–1400, 1983. English translation of [25].

[27] S. Soloviev. A complete axiom system for isomorphism of types in closed categories. In A. Voronkov, editor, *Logic Programming and Automated Reasoning*, volume 698 of *Lecture Notes in Computer Science*, pages 360–371. Springer Berlin Heidelberg, 1993.

[28] A. Wilkie. On exponentiation – a solution to Tarski's high school algebra problem. *Quaderni di Matematica*, 6, 2000.