

Normalization by evaluation for typed lambda calculus with coproducts

T. Altenkirch*

P. Dybjer†

M. Hofmann‡

P. Scott§

April 10, 2001

Abstract

We solve the decision problem for simply typed lambda calculus with strong binary sums, equivalently the word problem for free cartesian closed categories with binary coproducts. Our method is based on the semantical technique known as “normalization by evaluation” and involves inverting the interpretation of the syntax into a suitable sheaf model and from this extracting appropriate unique normal forms. There is no rewriting theory involved, and the proof is completely constructive, allowing program extraction from the proof.

1 Introduction

In this paper we solve the decision problem for simply typed lambda calculus with categorical coproduct (strong disjoint sum) types. While this calculus is both natural and simple, the decision problem is a long-standing thorny issue in the subject. Our solution is based on normalization by evaluation (NBE) (also called “reduction-free normalisation”) introduced by Martin-Löf [ML75] for weak typed lambda calculus, and by Berger and Schwichtenberg [BS91] for typed lambda calculus with $\beta\eta$ conversion. The technique has been further refined by the authors and coworkers using category-theoretic methods [CD97, AHS95, CDS97]. It has also been extended to other systems, such as System F [AHS96]. As shown by Berger, Schwichtenberg, and Danvy [BS91, Da96], NBE techniques yield fast normalization algorithms, with applications in interactive proof systems [BBSSZ98] and type-directed partial evaluation [Da96, Da98, Fil01].

Here we show how to considerably extend the NBE techniques to take into account type systems with strong sums. The NBE method involves constructing a model \mathcal{M} and effectively “inverting” the evaluation of lambda

terms in \mathcal{M} and thereby extracting certain unique syntactic normal forms, from which a decision procedure easily follows (we outline the proof below). The proof uses no rewriting theory.

Typed lambda calculi with (strong) sum types arise very naturally:

- In programming language theory, coproducts model variant and enumerative types. The added categorical equation for coproducts corresponds to a kind of uniqueness for *pattern matching* or *Case* construction [AC98, Mit96, GLT89].
- In proof theory, under the Curry-Howard Isomorphism, terms correspond to natural deduction proofs in intuitionistic propositional $\{\wedge, \vee, \Rightarrow, \top\}$ logic. One then considers terms (proofs) modulo certain equations, which guarantee, for example, that the formula $A \vee B$ acts as a coproduct type (with co-pairing), as well as including the theory of commutative conversions (cf [GLT89], pp 80-81). In category theoretic terminology, such lambda theories correspond exactly to *almost bicartesian closed categories*, that is, cartesian closed categories with nonempty finite coproducts (generated by a set of atomic types) [LS86].
- As proved by Dougherty and Subrahmanyam [DS95], a Friedman completeness theorem in **Set** holds for cartesian closed categories with binary coproducts. Therefore, the equality we decide is the natural extensional equality on proofs in intuitionistic propositional logic and on terms of the typed lambda calculus with sums.

Much of traditional lambda calculus theory carries through unscathed when we add products (and even weak categorical data types) to the simply typed case. Unfortunately, the addition of coproducts is considerably more subtle. The difficulties with adding coproducts are detailed in [Do93, DS95]: for example, the analog of Statman’s 1-Section theorem fails in the presence of coproducts, confluence (of various standard rewriting presentations) fails, and the proof of Friedman’s completeness theorem for the case of coproducts uses difficult and involved syntactical arguments [DS95].

* School of Computer Science and IT, University of Nottingham, Nottingham, UK, *e-mail*: txa@cs.nott.ac.uk

† Department of Computing Science, Chalmers University of Technology, S-412 96 Göteborg, Sweden, *e-mail*: peterd@cs.chalmers.se

‡ Division of Informatics (LFCS), University of Edinburgh, Edinburgh EH9 3JZ, UK, *e-mail*: mxh@dcs.ed.ac.uk

§ Department of Mathematics, University of Ottawa, 585 King Edward, Ottawa, Ontario, K1N 6N5, Canada. Phone: 613-562-5800, ext. 3502. FAX: 613-562-5776. *e-mail*: phil@mathstat.uottawa.ca

A decision procedure for cartesian closed categories with binary coproducts has been presented in Ghani's thesis [Gh95a] (see [Gh95b] for a summary) although the proof involves intricate rewriting techniques whose details are daunting. Our method described here is quite different and we believe conceptually simpler.

An algorithm for type-directed partial evaluation for a call-by-value typed lambda calculus with sums has been given by Danvy [Da96, Da98] and Filinski [Fil01]. This algorithm uses continuations and is therefore also quite different from ours. In particular, it does not decide equality in cartesian closed categories with binary coproducts.

Like Ghani and Dougherty and Subrahmanyam, we only consider the case of finite *non-empty* coproducts, that is, **an initial object (empty type) is not part of the structure**. We conjecture that the present approach can be extended to full bicartesian closed categories including initial objects. However, this complicates the structure of our normal forms, and we have not yet completely checked that all properties hold for the extended language.

Outline of Proof

Let \mathcal{E} be a lambda theory. Our aim is to decide if

$$\Gamma \vdash_{\mathcal{E}} e_1 = e_2 : A,$$

that is, if two possibly open terms e_1 and e_2 of type A are equal wrt \mathcal{E} , where Γ is a type environment. We associate with each term e a *normal form* $\text{nf}(e)$. In this paper, these normal forms are not themselves terms, but there is a function d mapping normal forms to terms in such a way that the following two properties hold (cf. [CD97, CDS97]):

$$\text{NF1 } \Gamma \vdash_{\mathcal{E}} d(\text{nf}(e)) = e$$

$$\text{NF2 } \Gamma \vdash_{\mathcal{E}} e_1 = e_2 \text{ implies } \text{nf}(e_1) = \text{nf}(e_2).$$

This implies that $\Gamma \vdash_{\mathcal{E}} e_1 = e_2$ if and only if $\text{nf}(e_1) = \text{nf}(e_2)$, so that comparing normal forms will yield a decision procedure for \mathcal{E} .

When \mathcal{E} = the typed lambda calculus with $\beta\eta$ -conversion, the authors and coworkers showed in [AHS95, CDS97] how to obtain a function nf by inverting the presheaf interpretation of \mathcal{E} . One defines two natural transformations $q^A : \llbracket A \rrbracket \rightarrow NF(A)$ and $u^A : NE(A) \rightarrow \llbracket A \rrbracket$, where $NF(A)$ is the presheaf of normal forms and $NE(A)$ is the presheaf of neutral terms of type A from \mathcal{E} . Given a typing judgement $\Gamma \vdash_{\mathcal{E}} e : A$, where $\Gamma = x_1 : A_1, \dots, x_n : A_n$, we define

$$\text{nf}(e) = q(\llbracket e \rrbracket(u(1_{\Gamma})))$$

where 1_{Γ} is the sequence (x_1, \dots, x_n) . Since $\llbracket - \rrbracket$ is an interpretation, we have immediately that $\Gamma \vdash_{\mathcal{E}} e_1 = e_2$ implies $\llbracket e_1 \rrbracket = \llbracket e_2 \rrbracket$, and hence NF2 follows and NF1 is proved by induction on e , using for example logical relations.

How do we obtain a function nf when we add strong sums to \mathcal{E} ? The problem is that although the category of presheaves has coproducts, a difficulty arises when we try to invert the interpretation of coproducts. The maps q and u are defined by induction on types, so in particular we need to define $u^{A_0 + A_1}$ in terms of u^{A_0} and u^{A_1} . But coproducts in presheaves are calculated pointwise so how do we for example define $u^{A_0 + A_1}(s) \in \llbracket A_0 \rrbracket_{\Gamma} + \llbracket A_1 \rrbracket_{\Gamma}$ for a neutral term $\Gamma \vdash s : A_0 + A_1$? Since variables are neutral terms, we must in particular define $u^{A_0 + A_1}(x)$, but there is no sensible way to decide whether this should be in the first or the second disjunct?

As we shall show, the solution of this problem is to introduce an appropriate topology and consider the sheaves for that topology. This will give us a way to “amalgamate” the contributions of u^{A_0} and u^{A_1} in the definition of $u^{A_0 + A_1}$.

Plan of the paper

In Section 2 we formally define the typed lambda calculus with strong sums and show how it yields a free cartesian closed category with binary coproducts. In Section 3 we introduce our normal forms, and the auxiliary notions of pure normal forms and neutral terms. The main idea is to introduce a parallel case statement, and impose variable conditions and a condition of redundancy-freeness to obtain uniqueness of normal forms. In Section 4 we introduce the category of constrained environments, where objects are environments (type assignments) equipped with equational constraints. This will serve as the underlying category of our Grothendieck topology which is defined in Section 5. There we also introduce the category of sheaves for this topology and its bicartesian closed structure. This yields a canonical interpretation of the syntax in the category of sheaves and in Section 6 we show how to invert this interpretation and obtain normal forms.

2 Syntax

We follow the treatment of sums in natural deduction, as in [GLT89, pp 80-81]. For ease of presentation, we restrict ourselves to one base type.

Types are given by the grammar

$$A ::= o \mid A \Rightarrow A \mid A \times A \mid \top \mid A + A$$

Terms are given by

$$e ::= x \mid \lambda x. e \mid e e \mid \langle e, e \rangle \mid \pi_0(e) \mid \pi_1(e) \mid \langle \rangle \mid \iota_0(e) \mid \iota_1(e) \mid \delta(x.e)(x.e)e$$

The *Case* term $\delta(x_0.e_0)(x_1.e_1)e_2$ simultaneously binds x_0 in e_0 and x_1 in e_1 .

A type environment Γ is a finite function from variables to types. The typing judgement $\Gamma \vdash e : A$ meaning e has

type A in type environment Γ is defined in the obvious way. For example, the rule for `Case` is:

$$\frac{(\Gamma, x_i : A_i \vdash e_i : C)_{i \in \{0,1\}} \quad \Gamma \vdash e : A_0 + A_1}{\Gamma \vdash \delta(x_0.e_0)(x_1.e_1)e : C}$$

Definition 2.1 Equality between terms $\Gamma \vdash - = - : A$ is the least (typed) congruence generated by the following rules (omitting types to improve readability):

$$\begin{array}{ll} (\beta) & (\lambda x.e_0)e_1 = e_0[e_1/x] \\ (\eta) & e = \lambda x.e x, \text{ if } x \notin \text{FV}(e) \\ \text{Proj}_i & \pi_i(\langle e_0, e_1 \rangle) = e_i \\ \text{SP} & e = \langle \pi_0(e), \pi_1(e) \rangle \\ \text{Unit} & e = \langle \rangle \\ \text{In}_i & \delta(x_0.e_0)(x_1.e_1)\iota_i(e_2) = e_i[e_2/x_i] \\ \text{Coproduct} & \delta(x_0.\iota_0(x_0))(x_1.\iota_1(x_1))e = e \\ \text{Distrib} & e(\delta(x_0.e_0)(x_1.e_1)e_2) = \\ & \delta(x_0.e e_0)(x_1.e e_1)e_2 \\ & \text{if } x_0, x_1 \notin \text{FV}(e) \end{array}$$

We will refer to this equational theory as BiCCC. The key categorical axiom (Coproduct) is dual to (SP) and guarantees uniqueness of the co-pairing arrow out of a coproduct. BiCCC entails all the usual commutative conversions for sums, [GLT89], pp. 80-81.

It can be shown (cf. [LS86, CDS97]) that the free almost bicartesian closed category \mathcal{B}_0 over one base object o can be obtained as the category whose objects are type environments and where a morphism from $\Gamma = x_1 : A_1, \dots, x_m : A_m$ to $y_1 : B_1, \dots, y_n : B_n$ is a sequence of terms (e_1, \dots, e_n) , modulo BiCCC equality, where $\Gamma \vdash e_i : B_i$. *Freeness* means that for each BiCCC \mathcal{B} and object $\llbracket o \rrbracket \in \mathcal{B}$ we have a unique structure- and equation-preserving interpretation functor $\llbracket - \rrbracket : \mathcal{B}_0 \rightarrow \mathcal{B}$.

3 Normal Forms

Normal forms are defined simultaneously with **pure normal forms** and **neutral terms**. Normal (and pure normal) forms are not genuine terms, but defined inductively by the clauses below. If Γ is a type environment we write $\Gamma \vdash_{\text{NF}} t : A$, resp. $\Gamma \vdash_{\text{PNF}} t : A$, resp. $\Gamma \vdash_{\text{NE}} t : A$ to mean that expression t is a normal form, resp. pure normal form, resp. neutral term of type A . We write $\text{FV}(t)$ for the set of free variables occurring in t . We write $\text{Guards}(t)$ for the set of *guards* of a normal form t : this will be defined below as part of the rule for forming normal forms.

$$\frac{x \in \text{dom}(\Gamma)}{\Gamma \vdash_{\text{NE}} x : \Gamma(x)} \quad \frac{\Gamma \vdash_{\text{NE}} s : o}{\Gamma \vdash_{\text{PNF}} s : o}$$

$$\Gamma \vdash_{\text{PNF}} \langle \rangle : \top$$

$$\frac{\Gamma \vdash_{\text{PNF}} t_0 : A_0 \quad \Gamma \vdash_{\text{PNF}} t_1 : A_1}{\Gamma \vdash_{\text{PNF}} \langle t_0, t_1 \rangle : A_0 \times A_1}$$

$$\frac{\Gamma \vdash_{\text{NE}} t : A_0 \times A_1}{\Gamma \vdash_{\text{NE}} \pi_i(t) : A_i} \quad i \in \{0, 1\}$$

$$\frac{\Gamma \vdash_{\text{PNF}} t : A_i}{\Gamma \vdash_{\text{PNF}} \iota_i(t) : A_0 + A_1} \quad i \in \{0, 1\}$$

$$\frac{\Gamma \vdash_{\text{NE}} s : A \Rightarrow B \quad \Gamma \vdash_{\text{PNF}} t : A}{\Gamma \vdash_{\text{NE}} st : B}$$

$$\frac{\Gamma, x:A \vdash_{\text{NF}} t : B}{\Gamma \vdash_{\text{PNF}} \lambda x.t : A \Rightarrow B}$$

where in the last rule we have the variable condition that $x \in \text{FV}(s)$ for each $s \in \text{Guards}(t)$.

We have two rules for forming normal forms:

$$(a) \quad \frac{\Gamma \vdash_{\text{PNF}} t : A}{\Gamma \vdash_{\text{NF}} t : A} \quad \text{and} \quad \text{Guards}(t) = \emptyset$$

- (b) Let $M = \{s_1, \dots, s_n\}$ be a nonempty finite set of neutral terms (so we assume the s_i are pairwise distinct). For each $f : M \rightarrow \{0, 1\}$ we use the abbreviation $\Gamma_f = \Gamma, x_1 : A_{f(s_1)}^1, \dots, x_n : A_{f(s_n)}^n$:

$$\frac{(\Gamma \vdash_{\text{NE}} s_i : A_0^i + A_1^i)_{i \in \{1, \dots, n\}} \quad (\Gamma_f \vdash_{\text{NF}} t_f : C)_{f : M \rightarrow \{0, 1\}}}{\Gamma \vdash_{\text{NF}} \mathcal{C}(M, (x_1 \cdots x_n.t_f)_f) : C}$$

and

$$\text{Guards}(\mathcal{C}(M, (x_1 \cdots x_n.t_f)_f)) = M$$

where $(t_f)_{f : M \rightarrow \{0, 1\}}$ is a family of normal forms satisfying the following two side conditions:

Variable-condition: for each $s \in \text{Guards}(t_f)$ we have $\{x_1, \dots, x_n\} \cap \text{FV}(s) \neq \emptyset$.

Redundancy-freeness: The family $(t_f)_f$ is not redundant at any $s_i \in M$, where $(t_f)_f$ is redundant at s_i whenever for all $g : M \setminus \{s_i\} \rightarrow \{0, 1\}$ $t_{g[s_i \mapsto 0]}$ and $t_{g[s_i \mapsto 1]}$ are equal and neither contains the variable x_i .

The variables x_1, \dots, x_n become bound in the \mathcal{C} -construct. For brevity we shall often use the alternative notation $\mathcal{C}(M, (t'_f)_f)$, where the t'_f range over abstractions $x_1, \dots, x_n.t_f$.

The idea is that \mathcal{C} performs a simultaneous case split over all the “guards”. For example, $t_{f[s \mapsto 0]}$ corresponds to a branch to be taken when s is of the form $\iota_0(x)$.

Example 3.1 The following examples show how the side-conditions ensure uniqueness of normal forms as computed by nf in Section 1.1. Let for simplicity the variables z (possibly with indices) in the examples below have type o , so that they are normal terms.

1. The normal form of $\lambda w. \delta (x_1.z_0) (x_1.z_1) y$ will be $\mathcal{C}(\{y\}, (x_1.t_f)_f)$ where $t_{[y \mapsto i]} = \lambda w. z_i$. Note that the expression $\lambda w. \mathcal{C}(\{y\}, (x_1.t_f)_f)$, where $t_{[y \mapsto i]} = z_i$, violates the side condition for (pure) normal forms of λ -form.

2. The normal form of the term

$$\begin{aligned} & \delta (x_1. \delta (x_2.z_{00}) (x_2.z_{01}) y_2) \\ & (x_1. \delta (x_2.z_{10}) (x_2.z_{11}) y_2) \\ & y_1 \end{aligned}$$

will be

$$\mathcal{C}(\{y_1, y_2\}, (x_1 x_2. t_f)_f)$$

where $t_{[y_1 \mapsto i, y_2 \mapsto j]} = z_{ij}$. Note that the expression $\mathcal{C}(\{y_1\}, (x_1. \mathcal{C}(\{y_2\}, (x_2. t_{f_1 \cup f_2})_{f_2})_{f_1}))$ is not a normal form since it violates the variable-condition: x_1 is not free in the guard y_2 of the normal form $\mathcal{C}(\{y_2\}, (x_2. t_{f_1 \cup f_2})_{f_2})$.

3. The normal form of $\delta (x.z) (x.z) y$ will be z . Note that $\mathcal{C}(\{y\}, (x.z)_f)$ is not a normal form as $(x.z)_f$ is redundant at y .
4. Note however, that the normal form of $\delta (z.z) (z.z) y$ will be $\mathcal{C}(\{y\}, (z.z)_f)$ which is not redundant at y because of the variable condition in the definition of redundancy.

Definition 3.2 The function d mapping $\Gamma \vdash_X t : A$ with $X \in \{\text{NF}, \text{PNF}, \text{NE}\}$ to terms $\Gamma \vdash d(t) : A$ is defined in the following way:

- d commutes with all the term formers except \mathcal{C} .
- $d(\mathcal{C}(M \cup \{s\}, (t_f)_f)) = \delta (x_0.e_0) (x_1.e_1) d(s)$, where $e_i = d(\mathcal{C}(M, (t_{g[s \mapsto i]})_g))$.

It is easy to see that up to BiCCC equality this does not depend on the choice of the witnessing term e_Γ and on the order of the guards.

4 Neutral constrained environments

Like Dougherty and Subrahmanyam [DS95] and Fiore and Simpson [FS99] we need to supply our type environments with constraints. These will be the objects of a

category of constrained environments \mathcal{N} , where the morphisms will be injective renamings. The constraints are of the form $s = \iota_i(x_i)$ and express which branch a certain guard s takes. This is the idea behind our Grothendieck topology on \mathcal{N} : a “covering” expresses case-splitting. This use of Grothendieck topologies is closely related to [FS99] where they were used for proving a definability result for a language with coproducts.

Definition 4.1 A *neutral constrained environment*, environment for short, is a pair $\Gamma \mid \Xi$ where Γ is a type environment and Ξ is a set of constraints of the form $s = \iota_0(x_0)$ or $s = \iota_1(x_1)$ where $\Gamma \vdash_{\text{NE}} s : A_0 + A_1$ and $x_0 : A_0$ (resp. $x_1 : A_1$) is contained in Γ and moreover,

- no two distinct constraints involve the same neutral term, for example, Ξ cannot contain $s = \iota_0(x_0)$ and $s = \iota_1(x_1)$
- no two distinct constraints refer to the same variable, for example, Ξ cannot contain $s = \iota_0(x_0)$ and $s' = \iota_0(x_0)$ unless s and s' are identical.

A *morphism* from environment $\Delta \mid \Psi$ to environment $\Gamma \mid \Xi$ is given by an *injective* function $\sigma : \text{dom}(\Gamma) \rightarrow \text{dom}(\Delta)$ satisfying $\Delta(\sigma(x)) = \Gamma(x)$ and $\sigma(s) = \iota_i(\sigma(x))$ is in Ψ for each constraint $s = \iota_i(x)$ in Ξ . In this way the environments form a category \mathcal{N} in which composition is composition of functions.

If Δ extends Γ and Ψ extends Ξ then the inclusion $\sigma : \text{dom}(\Gamma) \hookrightarrow \text{dom}(\Delta)$ defines a morphism from $\Delta \mid \Psi$ to $\Gamma \mid \Xi$ which we call a *projection*.

We are interested in studying equality of terms relative to a neutral constrained environment. The following definition is due to [DS95].

Definition 4.2 Let $\Gamma \mid \Xi$ be an environment and \vec{d} be a list of dummy terms of the same length as Ξ and of appropriate (to be explained) type. A (variable-binding) type environment $C_{\vec{d}}^{\Gamma \mid \Xi}[\]$ is defined as follows.

$$C_{\vec{d}}^{\Gamma \mid \emptyset}[\] = [\]$$

$$\begin{aligned} C_{\vec{d}, d_1}^{\Gamma, x_0 : A_0 \mid \Xi, s = \iota_0(x_0)}[\] = \\ \delta (x_0. C_{\vec{d}}^{\Gamma \mid \Xi}[\]) (x_1. d_1 x_1) d(s) \end{aligned}$$

$$\begin{aligned} C_{\vec{d}, d_0}^{\Gamma, x_1 : A_1 \mid \Xi, s = \iota_1(x_1)}[\] = \\ \delta (x_0. d_0 x_0) (x_1. C_{\vec{d}}^{\Gamma \mid \Xi}[\]) d(s) \end{aligned}$$

Note that $C_{\vec{d}}^{\Gamma \mid \Xi}[e]$ binds all variables mentioned in Ξ .

Given two terms $\Gamma \vdash e_1 : C$ and $\Gamma \vdash e_2 : C$ we write $\Gamma \mid \Xi \vdash e_1 = e_2 : C$ to mean that

$$\Gamma' \vdash C_{\vec{d}}^{\Gamma \mid \Xi}[e_1] = C_{\vec{d}}^{\Gamma \mid \Xi}[e_2] : C$$

in the theory BiCCC for all appropriate Γ' and \vec{d} . Here \vec{d} must be chosen such that the terms $C_{\vec{d}}^{\Gamma|\Xi}[e_i]$ are type correct and Γ' is obtained from Γ by removing the variables mentioned in Ξ and possibly adding any extra free variables occurring in the dummy terms \vec{d} .

Note that ordinary type environments have no constraints but it follows immediately from the above definition that $\Gamma|\emptyset \vdash e_1 = e_2$ implies $\Gamma \vdash e_1 = e_2$.

5 Sheaves over environments

We consider the functor category $\hat{\mathcal{N}} \stackrel{\text{def}}{=} \text{Sets}^{\mathcal{N}^{\text{op}}}$ of presheaves and natural transformations between them. We recall the following definitions of the structure of $\hat{\mathcal{N}}$. A *presheaf* is given by a family of sets $F_{\Gamma|\Xi}$ indexed by environments and for each morphism $\sigma : \Delta|\Psi \rightarrow \Gamma|\Xi$ a function $F_\sigma : F_{\Gamma|\Xi} \rightarrow F_{\Delta|\Psi}$ such that $F_1 = 1$ and $F_{\sigma \circ \tau} = F_\tau \circ F_\sigma$. If $a \in F_{\Gamma|\Xi}$ we may write $a|_{\Delta|\Psi}$ for $F_\sigma(a)$ in case σ is clear from the context. This notation will in particular be used when σ is a projection.

A *natural transformation* from presheaf F to presheaf G is given by a family $g_{\Gamma|\Xi}$ of maps $g_{\Gamma|\Xi} : F_{\Gamma|\Xi} \rightarrow G_{\Gamma|\Xi}$ such that $G_{\sigma \circ \tau} g_{\Gamma|\Xi} = g_{\Delta|\Psi} F_\sigma$ (*naturality*). If $a \in F_{\Gamma|\Xi}$ we may write $g(a)$ for $g_{\Gamma|\Xi}(a)$. Naturality then reads $g(a)|_{\Delta|\Psi} = g(a|_{\Gamma|\Xi})$.

As any category of presheaves, the category $\hat{\mathcal{N}}$ is bicartesian closed, that is, supports the interpretation of the type formers $\top, \times, \Rightarrow, +$, (and \perp). If we denote the interpreting presheaves with the same symbols thus writing e.g. $F \Rightarrow G$ for the function space of presheaves, we have the following explicit constructions of the type formers in $\text{Sets}^{\mathcal{N}^{\text{op}}}$:

$$\begin{aligned} \top_{\Gamma|\Xi} &= \{\langle \rangle\} \\ (F \times G)_{\Gamma|\Xi} &= F_{\Gamma|\Xi} \times G_{\Gamma|\Xi} \\ (F + G)_{\Gamma|\Xi} &= F_{\Gamma|\Xi} + G_{\Gamma|\Xi} \\ (F \Rightarrow G)_{\Gamma|\Xi} &= \hat{\mathcal{N}}(\mathcal{N}(-, \Gamma|\Xi) \times F, G) \end{aligned}$$

However, as we mentioned in the introduction, we are not able to obtain normal forms by inverting this presheaf interpretation. Instead we shall consider the interpretation of terms in the category of sheaves over a certain topology, and show that this can be inverted.

Recall that the basis of a *Grothendieck topology* is a collection of *basic coverings*, satisfying the Sxioms of identity, transitivity, and stability [MM92, p111]. A covering of an object $\Gamma|\Xi$ in \mathcal{N} is here a family of arrows with codomain $\Gamma|\Xi$. Since, the category \mathcal{N} does not have pullbacks in general, we use a modified axiom of stability [MM92, p156]. Moreover, like [FS99] we only require that the identity is a singleton covering, not that all isomorphisms are coverings.

Definition 5.1 The basis K for a Grothendieck topology on \mathcal{N} is inductively generated by the following clauses:

- The identity covering containing only the arrow $1_{\Gamma|\Xi}$ is a basic covering of $\Gamma|\Xi$.
- If $\Gamma \vdash_{\text{NE}} s : A + B$ and s is not mentioned in Ξ , and if the family of projections from $(\Gamma_i|\Xi_i)_i$ forms a basic covering of $\Gamma, x : A|\Xi, s = \iota_0(x)$ and the family of projections from $(\Gamma_j|\Xi_j)_j$ forms a basic covering of $\Gamma, y : B|\Xi, s = \iota_1(y)$, then the disjoint union of the projections from $(\Gamma_i|\Xi_i)_i$ and $(\Gamma_j|\Xi_j)_j$ forms a basic covering of $\Gamma|\Xi$.

The general concept of sheaves for Grothendieck topologies need not be presented, since it here specialises to the following rather digestible definition:

Proposition 5.2 A presheaf F is a sheaf for K iff whenever $\Gamma|\Xi$ is covered by $\Gamma, x_0:A_0|\Xi, s=\iota_0(x_0)$ and $\Gamma, x_1:A_1|\Xi, s=\iota_1(x_1)$, that is, $\Gamma \vdash_{\text{NE}} s : A_0 + A_1$ and

$$\begin{aligned} f_0 &\in F_{\Gamma, x_0:A_0|\Xi, s=\iota_0(x_0)} \\ f_1 &\in F_{\Gamma, x_1:A_1|\Xi, s=\iota_1(x_1)} \end{aligned}$$

then there exists a unique $f \in F_{\Gamma|\Xi}$ (called *pasting*) such that

$$\begin{aligned} f|_{\Gamma, x_0:A_0|\Xi, s=\iota_0(x_0)} &= f_0 \\ f|_{\Gamma, x_1:A_1|\Xi, s=\iota_1(x_1)} &= f_1 \end{aligned}$$

The following result follows from general properties of Grothendieck topologies and will therefore not be proved, see [MM92] for an exposition.

Proposition 5.3

1. The terminal object in $\hat{\mathcal{N}}$ is a sheaf,
2. if F, G are sheaves so is $F \times G$ (cartesian product),
3. if G is a sheaf and F is a presheaf then $F \Rightarrow G$ is a sheaf (function space)
4. for each presheaf F there exists a sheaf aF (the associated sheaf or sheafification) and a natural transformation $\eta : F \rightarrow aF$ such that whenever G is a sheaf and $f : F \rightarrow G$ then there exists a unique $f^\# : aF \rightarrow G$ with $f^\# \circ \eta = f$. In other words, the sheaves form a reflective subcategory of \mathcal{N} ,
5. The sheafification functor a preserves binary products.
6. if F, G are sheaves the coproduct $F + G$ is in general not a sheaf, but $a(F + G)$ is the coproduct of F and G in the subcategory of sheaves.
7. if $u, v : F \rightarrow G$ and F, G are sheaves then the equaliser of u and v is a sheaf.¹

¹PD: remove?

We write $\Gamma \mid \Xi \vdash_{\text{NF}} t : A$ to mean that $\Gamma \vdash_{\text{NF}} t : A$ and, moreover, none of the neutral terms mentioned in Ξ is contained in $\text{Guards}(t)$. Intuitively, this is because no case split is ever needed for a guard whose value is already known through the environment. Note that there is no need to define $\Gamma \mid \Xi \vdash_{\text{NE}} t : A$ and $\Gamma \mid \Xi \vdash_{\text{PNF}} t : A$, since all guards inside neutral and pure normal terms include variables bound by λ s. Hence the constraints in Ξ cannot affect t .

For a type A we define the presheaves $\text{NF}(A)$, $\text{PNF}(A)$, $\text{NE}(A)$, $\text{VAR}(A)$, $\text{Term}(A)$ as follows:

$$\begin{aligned} \text{NF}(A)_{\Gamma \mid \Xi} &= \{t \mid \Gamma \mid \Xi \vdash_{\text{NF}} t : A\} \\ \text{PNF}(A)_{\Gamma \mid \Xi} &= \{t \mid \Gamma \vdash_{\text{PNF}} t : A\} \\ \text{NE}(A)_{\Gamma \mid \Xi} &= \{t \mid \Gamma \vdash_{\text{NE}} s : A\} \\ \text{VAR}(A)_{\Gamma \mid \Xi} &= \text{dom}(\Gamma) \\ \text{Term}(A)_{\Gamma \mid \Xi} &= \{t \mid \Gamma \mid \Xi \vdash t : A\} / \sim_{\vdash} \end{aligned}$$

where $t \sim_{\vdash} t'$ stands for $\Gamma \mid \Xi \vdash t = t' : A$.

If $\sigma : \Delta \mid \Psi \rightarrow \Gamma \mid \Xi$ and $\Gamma \mid \Xi \vdash_{\text{NE}} t : A$ then $\text{NE}(A)_{\sigma}(t) \in \text{NE}(A)_{\Delta \mid \Psi}$ is defined by replacing each free variable x in t by $\sigma(x)$. The morphism parts Term_{σ} and PNF_{σ} are defined analogously.

If $x \in \text{VAR}(A)_{\Gamma \mid \Xi}$ then $\text{VAR}(A)_{\sigma}(x) = \sigma(x)$.

If $t \in \text{NF}_{\Gamma \mid \Xi}(A)$ then $\text{NF}_{\sigma}(t)$ is defined by first replacing each free variable x in t by $\sigma(x)$ and then plugging in all the constraints mentioned in Ψ by repeatedly performing the following atomic restriction operation (an analogous operation appears in Ghani's thesis [Gh95a] under the name ‘‘first and second residue’’).

Definition 5.4 Let $t \in \text{NF}(C)_{\Gamma \mid \Xi}$ and $\Gamma \vdash_{\text{NE}} s : A_0 + A_1$. Then we define the restriction $t[s := \iota_i(x_i)]$ of t to $\Gamma, x_i : A_i \mid \Xi, s = \iota_i(x_i)$ (along the projections) as follows.

$$\begin{aligned} t[s := \iota_i(x)] &= t, \text{ if } s \notin \text{Guards}(t) \\ \mathcal{C}(M \cup \{s\}, (t_f)_f)[s := \iota_i(x_i)] &= \mathcal{C}^{\text{nf}}(M, (t_{g[s \mapsto i]})_g) \end{aligned}$$

where \mathcal{C}^{nf} computes a normal form to be defined below. Note that we cannot simply replace \mathcal{C}^{nf} by \mathcal{C} because the set of guards can become empty upon plugging in a constraint, new redundancies may be created, and the variable conditions may be violated. We define $\mathcal{C}^{\text{nf}}(\emptyset, \{t\})$ to be t and $\mathcal{C}(M \cup \{s\}, (t_f)_f)$ to be $\delta^{\text{nf}}(x_0. \mathcal{C}^{\text{nf}}(M, (t_{f[s \mapsto 0]})_f)) (x_1. \mathcal{C}^{\text{nf}}(M, (t_{f[s \mapsto 1]})_f)) s$.

To compute $\delta^{\text{nf}}(x_0.t_0)(x_1.t_1)s$ we first check whether t_i depend on x_i and are different (see the definition of redundancy). If not, we return $t_0 (= t_1)$, or otherwise, we return $\mathcal{C}(\{s\} \cup M_0 \cup M_1, t_g)$, where

$$M_i = \{s_i \in \text{Guards}(t_i) \mid x_i \notin \text{FV}(s_i)\}$$

for $i = 0, 1$, and the family t_g is adjusted accordingly.

Proposition 5.5 *d defines natural transformations*
 $\text{NF}(A) \rightarrow \text{Term}(A)$, $\text{PNF}(A) \rightarrow \text{Term}(A)$,
 $\text{NE}(A) \rightarrow \text{Term}(A)$, $\text{VAR}(A) \rightarrow \text{Term}(A)$.

If $f : \mathcal{B}(\Delta, \Gamma)$ is a morphism in the free BiCCC \mathcal{B} , that is, a sequence of terms in type environment Δ , then $[t] \mapsto [ft]$ defines a natural transformation $\text{Term}(f) : \text{Term}(\Delta) \rightarrow \text{Term}(\Gamma)$. This makes $\text{Term}(-)$ a functor from \mathcal{B} to $\hat{\mathcal{N}}$ preserving \top and cartesian products.

Proposition 5.6 *The presheaf $\text{Term}(C)$ is a sheaf.*

Proposition 5.7 *Let A, B be types. There is a natural transformation*

$$\lambda^{\text{nf}} : (\text{VAR}(A) \Rightarrow \text{NF}(B)) \rightarrow \text{NF}(A \Rightarrow B)$$

When applying λ^{nf} to a normal form $\mathcal{C}(M, (x_1 \dots x_n.t_f)_f)$, depending naturally on a variable x of type A , we divide M into two sets M_0 , which contains the guards which do not depend on x , and M_1 , which contains the guards which do. Then we return

$$\mathcal{C}(M_0, (x_1 \dots x_{n_0}. \lambda x. \mathcal{C}^{\text{nf}}(M_1, (x_1 \dots x_{n_1}. t_{f_0 \cup f_1})_{f_1}))_{f_0})$$

Compare also example 1 in 3.1.

Proposition 5.8 *The presheaf $\text{NF}(A)$ is a sheaf and is isomorphic to the sheafification $a(\text{PNF}(A))$ of $\text{PNF}(A)$ with the embedding $\eta : \text{PNF}(A) \rightarrow \text{NF}(A)$ given by $\eta_{\Gamma \mid \Xi}(t) = t$.*

If $\Gamma \vdash s : A_0 + A_1$, then the pasting of two normal forms $t_i \in \text{NF}(A)_{\Gamma, x_i : A_i \mid \Xi, s = \iota_i(x_i)}$ is the normal form $\delta^{\text{nf}}(x_0.t_0)(x_1.t_1)s \in \text{NF}(A)_{\Gamma \mid \Xi}$.

Let us write $Sh(\mathcal{N})$ for the full subcategory of $\hat{\mathcal{N}}$ consisting of the sheaves. We know from Prop. 5.3 that $Sh(\mathcal{N})$ is a BiCCC. Since the category \mathcal{B}_0 of sequences of types and terms is a free BiCCC there is a unique interpretation functor $\llbracket - \rrbracket : \mathcal{B}_0 \rightarrow Sh(\mathcal{N})$, determined by

$$\llbracket o \rrbracket = \text{NF}(o)$$

Concretely, this functor is given by defining a canonical BiCCC structure on $Sh(\mathcal{N})$.

6 Inverting the interpretation function

We will now define natural transformations

$$\begin{aligned} q^A : \llbracket A \rrbracket &\rightarrow \text{NF}(A) \\ u^A : \text{NE}(A) &\rightarrow \llbracket A \rrbracket \end{aligned}$$

in such a way that for a term $\Gamma \vdash e : A$,

$$\text{nf}(e) \stackrel{\text{def}}{=} q_{\Gamma}^A(\llbracket e \rrbracket(u_{\Gamma}^{\top}(1_{\Gamma})))$$

will satisfy $\Gamma \vdash e = \text{nf}(e)$ (NF1):

- $q^o : \text{NF}(o) \rightarrow \text{NF}(o)$ is the identity function.
- $u^o : \text{NE}(o) \rightarrow \text{NF}(o)$ is the injection from neutral terms to normal terms given by the obvious term-formation rules.

- $q^\top : \top \rightarrow \text{NF}(\top)$ is the constant function returning the normal form $\langle \rangle$.

$u^\top : \text{NE}(\top) \rightarrow \top$ is the constant function returning the element $\langle \rangle \in \top$. (As before we use the same signs for corresponding syntactic and semantic notions.)

- $q^{A_0 \times A_1} = \text{pair}^{\text{nf}}_o(q^{A_0} \times q^{A_1})$ where $\text{pair}^{\text{nf}} : \text{NF}(A_0) \times \text{NF}(A_1) \rightarrow \text{NF}(A_0 \times A_1)$ is the unique map satisfying $\text{pair}^{\text{nf}}(t_1, t_2) = \langle t_1, t_2 \rangle$ for pure normal forms t_1, t_2 . This map exists by Proposition 5.8 and the fact that a preserves products.

$$u_{\Gamma|\Xi}^{A_0 \times A_1}(s) = \langle u_{\Gamma|\Xi}^{A_0}(\pi_0(s)), u_{\Gamma|\Xi}^{A_1}(\pi_1(s)) \rangle$$

- Let $\theta \in \llbracket A \Rightarrow B \rrbracket_{\Gamma|\Xi} = \widehat{\mathcal{N}}(\mathcal{N}(-, \Gamma|\Xi) \times \llbracket A \rrbracket, \llbracket B \rrbracket)$. Then

$$q_{\Gamma|\Xi}^{A \Rightarrow B}(\theta) = \lambda_{\Gamma|\Xi}^{\text{nf}}(g) \in \text{NF}(A \Rightarrow B)_{\Gamma|\Xi},$$

where λ^{nf} refers to the natural transformation in Prop. 5.7 and $g \in (\text{VAR}(A) \Rightarrow \text{NF}(B))_{\Gamma|\Xi} = \widehat{\mathcal{N}}(\mathcal{N}(-, \Gamma|\Xi) \times \text{VAR}(A), \text{NF}(B))$ is defined by

$$g_{\Delta|\Psi}(\sigma, x) =$$

$$q_{\Delta|\Psi}^B(\theta_{\Delta|\Psi}(\sigma, u_{\Delta|\Psi}^A(x))) \in \text{NF}(B)_{\Delta|\Psi},$$

where $\sigma \in \mathcal{N}(\Delta|\Psi, \Gamma|\Xi)$ and $x \in \text{VAR}(A)_{\Delta|\Psi}$.

Let $s \in \text{NE}(A \Rightarrow B)_{\Gamma|\Xi}$. Then $u_{\Gamma|\Xi}^{A \Rightarrow B}(s) \in \llbracket A \Rightarrow B \rrbracket_{\Gamma|\Xi}$ is defined by

$$(u_{\Gamma|\Xi}^{A \Rightarrow B}(s))_{\Delta|\Psi}(\sigma, a) =$$

$$u_{\Delta|\Psi}^B(\text{NE}_\sigma(s)(q_{\Delta|\Psi}^A(a))) \in \llbracket B \rrbracket_{\Delta|\Psi}$$

where $\sigma \in \mathcal{N}(\Delta|\Psi, \Gamma|\Xi)$ and $a \in \llbracket A \rrbracket_{\Delta|\Psi}$.

- $q^{A_0 + A_1}$ is the unique map (arising from the coproduct property of $\llbracket A_0 + A_1 \rrbracket$) satisfying

$$\begin{aligned} q^{A_0 + A_1}(\iota_0^{\text{sh}}(a)) &= \iota_0^{\text{nf}}(q^{A_0}(a)) \\ q^{A_0 + A_1}(\iota_1^{\text{sh}}(b)) &= \iota_1^{\text{nf}}(q^{A_1}(b)) \end{aligned}$$

Here $\iota_0^{\text{sh}}, \iota_1^{\text{sh}}$ are the coproduct injections in $\text{Sh}(\mathcal{N})$ and $\iota_0^{\text{nf}} : \text{NF}(A_0) \rightarrow \text{NF}(A_0 + A_1)$ is the unique map satisfying $\iota_0^{\text{nf}}(t) = \iota_0(t)$ for pure normal form $t : A_0$. Similarly for ι_1^{nf} .

To construct $u^{A_0 + A_1} \in \text{NE}(A_0 + A_1) \rightarrow \llbracket A_0 + A_1 \rrbracket$ consider $s \in \text{NE}(A_0 + A_1)_{\Gamma|\Xi}$: either $s = \iota_0(x) \in \Xi$ in which case we put $f_{\Gamma|\Xi}(s) = \iota_0^{\text{sh}}(u_{\Gamma|\Xi}^{A_0}(x))$, or $s = \iota_1(y) \in \Xi$ and we put $f_{\Gamma|\Xi}(s) = \iota_1^{\text{sh}}(u_{\Gamma|\Xi}^{A_1}(y))$, or s is not mentioned in Ξ in which case we define $f_{\Gamma|\Xi}(s)$ as the unique pasting of

$$\begin{aligned} a_0 &\stackrel{\text{def}}{=} \iota_0^{\text{sh}}(u_{\Gamma, x:A_0|\Xi, s=\iota_0(x)}^{A_0}(x)) \\ a_1 &\stackrel{\text{def}}{=} \iota_1^{\text{sh}}(u_{\Gamma, x:A_1|\Xi, s=\iota_1(x)}^{A_1}(x)) \end{aligned}$$

It follows by straightforward calculations that all these are indeed natural transformations.

Proposition 6.1 *In order to establish NF1, that is, $e = d(q^A(\llbracket e \rrbracket(u(1_\Gamma))))$ for $\Gamma \vdash e : A$ we define a family of subsheaves $R_{\Gamma|\Xi}^A \subseteq \llbracket A \rrbracket_{\Gamma|\Xi} \times \text{Term}(A)_{\Gamma|\Xi}$, such that*

- (i) *For all $a \in \llbracket A \rrbracket_{\Gamma|\Xi}$ and $\Gamma \vdash e : A$:*

$$a R_{\Gamma|\Xi}^A e \Rightarrow \Gamma|\Xi \vdash d(q_{\Gamma|\Xi}^A(a)) = e$$

- (ii) *For all $s \in \text{NE}(A)_{\Gamma|\Xi}$*

$$u_{\Gamma|\Xi}^A(s) R_{\Gamma|\Xi}^A d(s)$$

We can extend R to type environments $\Gamma = x_1 : A_1, \dots, x_n : A_n$ by letting $(a_1, \dots, a_n) R_{\Gamma|\Xi}^\Gamma (f_1, \dots, f_n)$ iff $a_i R_{\Gamma|\Xi}^{A_i} f_i$ for $1 \leq i \leq n$. Similarly, we can extend q and u to type environments as well.

Proposition 6.2 (Logical Relations Lemma) *If $\Gamma \vdash e : C$ and $\vec{a} R_{\Gamma|\Xi}^\Gamma \vec{f}$ then*

$$\llbracket e \rrbracket(\vec{a}) R_{\Gamma|\Xi}^C e[\vec{f}/\vec{x}],$$

where \vec{x} are the variables in Γ .

Theorem 6.3 *The equational theory BiCCC is decidable.*

Proof. The above shows that the normalisation function nf satisfies NF1, because by (ii) and $d(1_\Gamma) = 1_\Gamma$, we know that

$$u_\Gamma^\Gamma(1_\Gamma) R_\Gamma^\Gamma 1_\Gamma$$

Hence by proposition 6.2, we know that

$$\llbracket e \rrbracket(u_\Gamma^\Gamma(1_\Gamma)) R_\Gamma^A e$$

Hence, by (i)

$$\Gamma \vdash d(\text{nf}(e)) = d(q_\Gamma^A(\llbracket e \rrbracket(u_\Gamma^\Gamma(1_\Gamma)))) = e$$

As we pointed out in the introduction NF2 holds automatically, and hence it follows that

$$\Gamma \vdash e_1 = e_2 \iff \text{nf}(e_1) = \text{nf}(e_2)$$

This yields a decision procedure since equality of normal forms is decidable. (Note that when writing the algorithm for we represent the finite set of guards as a list or a tree, so that normal forms are only unique up to the ordering of the guards.) Furthermore, the interpretation in $\text{Sh}(\mathcal{N})$ as well as the definition of q, u are clearly algorithmic. In fact, the whole development can be formalised in extensional Martin-Löf type theory using standard methods for formalizing category theory in Martin-Löf type theory. This would be one way of demonstrating explicitly that all functions we construct by abstract mathematical means are computable. \square

References

- [AHS95] T. Altenkirch, M. Hofmann, T. Streicher, Categorical reconstruction of a reduction-free normalisation proof, *Proc. CTCS '95 Springer LNCS* **953**, 182–199.
- [AHS96] T. Altenkirch, M. Hofmann, T. Streicher, Reduction-free normalisation for a polymorphic system, *11th Annual IEEE LICS Symposium*, 1996, 98–106.
- [AC98] R. M. Amadio, P.-L. Curien, *Selected Domains and Lambda Calculi*, Camb. Univ. Press, 1998.
- [BBSSZ98] H. Benl, U. Berger, H. Schwichtenberg, M. Seisenberger and W. Zuber, Proof theory at work: Program development in the Minlog system, in: *Automated Deduction*, W. Bibel and P.H. Schmitt, eds., Vol. II, Kluwer 1998).
- [BS91] U. Berger and H. Schwichtenberg, An inverse to the evaluation functional for typed λ -calculus, *6th Annual IEEE LICS Symposium*, 1991, 203–211.
- [CD97] T. Coquand and P. Dybjer, Intuitionistic Model Constructions and Normalization Proofs, *Math. Structures in Computer Science* **7**, 1997, 75–94.
- [CDS97] D. Čubrić, P. Dybjer and P.J.Scott. Normalization and the Yoneda Embedding, *Math. Structures in Computer Science* **8**, No.2, 1997, 153–192.
- [Da96] O. Danvy. Type-directed partial evaluation, POPL'96, ACM Press, 242–257.
- [Da98] O. Danvy. Type-directed partial evaluation, Partial evaluation, Practice and Theory, *Proceedings of the 1998 DIKU Summer School, Springer LNCS* **1706**, 367–411.
- [DiCo95] R. Di Cosmo. *Isomorphism of Types: from λ -calculus to information retrieval and language design*, Birkhäuser, 1995.
- [Do93] D. Dougherty. Some lambda calculi with categorical sums and products, RTA5, *Springer LNCS* **690**, 1993, 135–151.
- [DS95] D. Dougherty and R. Subrahmanyam, Equality between Functionals in the Presence of Coproducts, *Inf. & Comp.* (to appear). Prelim. version in: *10th Annual IEEE LICS Symposium*, 1995, 282–291.
- [Fil01] A. Filinski, Normalization by Evaluation for the Computational Lambda-Calculus, to appear in the proceedings of TLCA 2001.
- [FS99] M. Fiore and A. Simpson. Lambda Definability with Sums via Grothendieck Logical Relations, TLCA'99, *Springer LNCS* **1581**.
- [Gh95a] N. Ghani, *Adjoint Rewriting*, PhD thesis, LFCS, Univ. of Edinburgh, Nov. 1995.
- [Gh95b] N. Ghani, $\beta\eta$ -equality for coproducts. *TLCA '95 Springer LNCS* **902**, 1995, 171–185.
- [GLT89] J.-Y. Girard, Y. Lafont, P. Taylor. *Proofs and Types*, Cambridge Tracts in Theoretical Computer Science **7**, 1989.
- [JGh95] C. B. Jay and N. Ghani, The virtues of eta expansion, *Journal of Functional Programming*, **5**, no.2, 1995, 135–154.
- [LS86] J. Lambek and P. J. Scott. *Introduction to Higher Order Categorical Logic*, Cambridge Studies in Advanced Mathematics **7**, Cambridge University Press, 1986.
- [MM92] S. Mac Lane, I. Moerdijk. *Sheaves in Geometry and Logic*, Springer-Verlag, 1992.
- [ML75] P. Martin-Löf. An Intuitionistic Theory of Types: Predicative Part, *Logic Colloquium '73*, North-Holland, 1975.
- [Mit96] J. C. Mitchell. *Foundations for Programming Languages*, MIT Press, 1996.