Available online at www.sciencedirect.com

SCIENCE DIRECT®

Annals of Pure and Applied Logic 130 (2004) 325–337

ANNALS OF PURE AND APPLIED LOGIC

www.elsevier.com/locate/apal

On the λY calculus

Rick Statman

Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Available online 24 June 2004

Abstract

The λY calculus is the simply typed λ calculus augmented with the fixed point operators. We show three results about λY: (a) the word problem is undecidable, (b) weak normalisability is decidable, and (c) higher type fixed point operators are not definable from fixed point operators at smaller types.

0. Introduction

In this note we consider three problems concerning the λY calculus obtained from the simply typed lambda calculus by the addition of fixed-point combinators

$$Y : (A \to A) \to A.$$

The "paradoxical" combinator Y was first discussed in Curry and Feys, vol. 1 [3]. It appears first in a typed context in Richard Platek's thesis [6] and then in Scott's LCF [8], and forms the basis for Milner et al.'s Edinburgh LCF system and its descendants [5,7].

In this note we shall consider

(1) The question of whether higher type fixed-point combinators are "definable" from lower type fixed-point combinators. We shall show that it is not the case in this context, sharpening a result of ours from [9]. The same result has been obtained by Werner Damm by a different method.
(2) The question of the decidability of termination. More precisely, we shall show that it is decidable whether a given term has a normal form. This extends results of Plotkin and Bercovici [2]. By similar methods we show that it is decidable whether a term has a head normal form.

E-mail address: statman@cs.cmu.edu (R. Statman).

0168-0072/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.apal.2004.04.004

(3) The question of the decidability of the word problem. This question was first put to us by Albert Meyer 20 years ago. We shall show that it is in general undecidable whether two λY terms convert. This is done by encoding the behavior of register machines. In addition we shall give a decision procedure for the special case of only Y's of type $(0 \to 0) \to 0$.

1. Notations and terminology

We adopt for the most part the notation and terminology of Barendregt [1, p. 611].

We shall denote the type $(0 \to 0) \to (0 \to 0)$ of Church numerals by $\mathbb{N}$. The Church numeral for n will be denoted $\underline{n}$. We shall need some higher type Church numerals as well and toward this end let

$$\%(n) = (0 \to (\ldots(0 \to 0)\ldots)) \to 0$$

for $n + 2$ occurrences of 0 and

$$\$(n) = (\%(n) \to \%(n)) \to (\%(n) \to \%(n)).$$

2. The λY calculus

λY is obtained from simply typed λ-calculus by the addition of constants

$$Y : (B \to B) \to B$$

for each simple type B. $\beta\eta Y$ reduction $\xrightarrow[\beta\eta Y]{}\!\!\!\!\rightarrow$ is obtained from $\beta\eta$ reduction $\xrightarrow[\beta\eta]{}\!\!\!\!\rightarrow$ by the addition of the reduction rule (Y-reduction) $Y \to \lambda x.x(Yx)$ and $\underset{\beta\eta Y}{\twoheadleftarrow}$ and $\underset{\beta\eta Y}{\twoheadleftarrow\!\!\twoheadrightarrow}$ are defined in the obvious way. Intuitively, Y is interpreted as a fixed point or "paradoxical" combinator. Let $O = \lambda yx.x(yx)$. Then $Y \underset{\beta\eta Y}{\twoheadleftarrow\!\!\twoheadrightarrow} OY$ so Y is a solution to the fixed point equation

$$y = Oy.$$

This is, of course, for $Y : (B \to B) \to B$. However for

$$Y : (((B \to B) \to B) \to ((B \to B) \to B)) \to ((B \to B) \to B)$$

another solution is

$$YO.$$

We can formally interpret λY into the untyped λ-calculus with $\underset{\beta\eta}{\twoheadleftarrow\!\!\twoheadrightarrow}$ by replacing each occurrence of Y by Turing's fixed-point combinator

$$(\lambda xy.y(xxy))\lambda xy.y(xxy).$$

Thus λY satisfies the Church–Rosser theorem, the standardization theorem, and the completeness of inside-out reductions. Since the formal interpretation is so simple it is often convenient to conflate a λY term with its untyped interpretation. Thus we can speak about the Böhm tree of a term. Additionally, we write

$$U \sqsubseteq V$$

if the Böhm tree of $[^{\Omega}/Y]U$ is less than or equal to the Böhm tree of $[^{\Omega}/Y]V$ (these Böhm trees are of course finite).

Now $\underset{Y}{\twoheadrightarrow}$ satisfies the strong diamond property. We let $\mathrm{CRY}(X)$ be a complete reduction of all the Y redexes in X.

Proposition 1. *Suppose that $X \underset{\beta\eta Y}{\twoheadleftarrow\!\twoheadrightarrow} Z$. Then there exists an integer k such that for all sufficiently large r we have*

$$\mathrm{CRY}^r(Z) \sqsubset \mathrm{CRY}^{r+k}(X)$$

Proof. Suppose that $X \underset{\beta\eta Y}{\twoheadleftarrow\!\twoheadrightarrow} Z$. By the Church–Rosser theorem there exists a U such that $Z \underset{\beta\eta Y}{\twoheadrightarrow} U \underset{\beta\eta Y}{\twoheadleftarrow} X$. By analyzing the completeness of inside-out reductions we see that there exists an integer m, and terms V, W with the property that

$$\begin{aligned} Z &\underset{Y}{\twoheadrightarrow} \mathrm{CRY}^m(Z) \underset{\beta\eta}{\twoheadrightarrow} V \underset{Y}{\twoheadleftarrow} U \\ &\underset{Y}{\twoheadrightarrow} W \underset{\beta\eta}{\twoheadleftarrow} \mathrm{CRY}^m(X) \underset{Y}{\twoheadleftarrow} X. \end{aligned}$$

By the Church–Rosser theorem for $\underset{Y}{\twoheadrightarrow}$ there exists an integer k and a term T such that

$$V \underset{Y}{\twoheadrightarrow} T \underset{Y}{\twoheadleftarrow} W$$

and

$$T \underset{Y}{\twoheadrightarrow} \mathrm{CRY}^k(W) \underset{Y}{\twoheadleftarrow} W.$$

Now suppose that $r > m$ is given. Then

$$\begin{aligned} \mathrm{CRY}^r(Z) &\underset{\beta\eta}{\twoheadrightarrow} \mathrm{CRY}^{r-m}(V) \\ &\underset{Y}{\twoheadrightarrow} \mathrm{CRY}^{k+r-m}(W) \\ &\underset{\beta\eta}{\twoheadleftarrow} \mathrm{CRY}^{k+r}(X). \end{aligned}$$

Thus the Böhm tree of $[^{\Omega}/Y]\mathrm{CRY}^r(Z)$ is contained in the Böhm tree of $[^{\Omega}/Y]\mathrm{CRY}^{k+r}(X)$ as desired. □

Recall that the depth of a type is defined by $\mathrm{depth}(0) = 0$ and

$$\mathrm{depth}(A \to B) = \max\{\mathrm{depth}(A), \mathrm{depth}(B)\} + 1.$$

Define the stack of 2's function $s(m, n)$ by $s(0, n) = n$ and $s(m + 1, n) = 2^{s(m,n)}$. In the β redex $(\lambda x.X)Z$ the "cut type" is the type of $\lambda x.X$.

Recall the following observation, which goes back to Bill Tait [11].

Proposition 2. *If the maximum depth of a cut type in X is m and X has length n then the $\beta\eta$-normal form of X has length at most $s(m, n)$.*

3. The undefinability of higher type *Y*

We shall need to treat certain variables of low type in a special way. We shall call these variables "parameters" and we shall denote them by lower case Roman letters at the beginning of the alphabet. The parameters are $a : 0 \to (0 \to 0)$, and $b, c, d, \ldots : 0$. Given a term X we can lift the type of X by applying the substitution $[0 \to 0/0]$ to the types of the variables and bound parameters in X. If X contains free parameters then instead of applying the substitution to the types of these parameters we replace them as follows:

$$a \mapsto \lambda uvw.a(uw)(vw) \qquad \text{where } w : 0 \text{ and } u, v : 0 \to 0$$
$$b \mapsto \lambda z.b,$$
$$c \mapsto \lambda z.c,$$
$$\ldots$$

Otherwise parameters function like variables and can be both free and bound.

We shall denote the result of this operation by

$$\mathrm{LIFT}(X).$$

We shall sometimes wish to iterate the operation LIFT an unspecified number of times and we shall denote the result by

$$\mathrm{LIFT}^*(X).$$

An applicative combination of parameters which is of type 0 is a binary tree. If $x : 0 \to 0$ then an applicative combination of parameters and x of type 0 is a tree with both binary and unary branching. In a tree, the height of a path (to a leaf) is the number of occurrences of x on the path, and the height of the tree is the maximum height of a path to some occurrence of b. If there is no occurrence of b then the height of the tree is undefined.

A term $U : (0 \to 0) \to (0 \to 0)$ is said to be *well put together* if it is in long $\beta\eta$-normal form, and has the shape

$$\lambda xb.X$$

where X is a tree. The following can be proved by induction:

Lemma 1. *Let U be well put together.*
Then $\mathrm{LIFT}(U)\ \underline{2}$ *has a long $\beta\eta$-normal form V satisfying*

(1) *V is well put together (after α-conversion between variables and parameters)*
(2) *The height of V is $2^{\mathrm{height}(U)}$ provided U has a height.*

Corollary 1. *Let U be well put together with a height then*

$$\mathrm{LIFT}(\ldots\mathrm{LIFT}(\mathrm{LIFT}(U)\underline{2})\underline{2}\ldots)\underline{2}$$

(with m occurrences of $\underline{2}$) has a long $\beta\eta$-normal form V satisfying

(1) *V is well put together (after α-conversion), and*
(2) *V has height $s(m, \mathrm{height}(U))$.*

Corollary 2. *If U is well put together with a height then*

$$\mathrm{LIFT}(\ldots \mathrm{LIFT}(\mathrm{LIFT}(U)\underline{2})\underline{2}\ldots)\underline{2}\ (\lambda z.azz)b$$

(with m occurrences of $\underline{2}$) has a long $\beta\eta$-normal form which is a binary tree with at least $s(m+1, \mathrm{height}(U))$ leaves which are occurrences of b.

A typical such U is

$$\lambda xb.\ ab(x(ab(x\ldots(x(abc))\ldots)))$$

(with n occurrences of x) which has height n.

Now define the type n, for $n > 0$, by $n+1 = n \to n$.

Theorem 1. *The fixed point equation $y = Oy$ has no solution*

$$M : (n \to n) \to n$$

containing only Y's of type of depth $\leq n+1$.

Proof. Now suppose that M is a solution to the fixed point equation

$$y = Oy$$

for $y : (n \to n) \to n$ and M contains Y's of types with depth at most $n+1$. Let F be the term

$$\lambda uxb.\ ab(u(\lambda v.x(abv))).$$

Note that $\mathrm{LIFT}^*(F)\ \mathrm{CRY}^n(Y) \underset{\beta\eta}{\twoheadrightarrow} \mathrm{LIFT}^*(U)$ of a typical U, where some terms beginning with Y are substituted for the parameter c. By Proposition 1 there exists an integer k such that for all sufficiently large r and all t

$$O(\ldots(O\ \mathrm{CRY}^r(M))\ldots) \sqsubset \mathrm{CRY}^{r+kt}(M)$$

where O occurs t times. Now the term

$$G = (\ldots((\mathrm{LIFT}^*(F)M\ \mathrm{LIFT}^*(\underline{2}))\mathrm{LIFT}^*(\underline{2}))\ldots\mathrm{LIFT}^*\underline{2})\underline{2}(\lambda z.azz)b$$

with $n+1$ occurrences of $\underline{2}$ has a long $\beta\eta$-normal form P, say of length p. Since the depth of the type of any Y redex in P is at most $n+1$ we have that the length of

$$\mathrm{CRY}^{r+kt}(P)$$

is $q(r+kt)$ for some constant q and the length of the long $\beta\eta$-normal form of $\mathrm{CRY}^{r+kt}(P)$ is at most $s(n+1, qr+qkt)$. Now the term

$$H = (\ldots(\mathrm{LIFT}^*(F)(O(\ldots O(\mathrm{CRY}^r(M))\ldots\mathrm{LIFT}^*(\underline{2})))$$
$$\mathrm{LIFT}^*(\underline{2}))\ldots\mathrm{LIFT}^*(\underline{2}))\underline{2}(\lambda z.azz)b$$

with $n+1$ occurrences of $\underline{2}$ and t occurrences of O has a long $\beta\eta$-normal form which is a binary tree with at least $s(n+2, t)$ leaves which are occurrences of b and leaves c replaced by other terms. However $H \sqsubset G$, but this is impossible once t is large enough so that

$$qr + qkt < 2^t.$$

This completes the proof. $\square$

4. Terms convertible to Y-free terms

Proposition 1 implies that if X $\beta\eta Y$-converts to a Y free term then for some positive integer r the $\beta\eta$ normal form of

$$[x/Y]\,\mathrm{CRY}^r(X),$$

where x is a new variable, does not contain x. Clearly x can be replaced here by $\lambda w\lambda z_1 \ldots \lambda z_k.\ z$ for $z : 0$. For each $Y : ((A \to A) \to A)$ select a new variable $y : (A \to A) \to (A \to A)$ and replace each occurrence of Y in X by $\lambda w.yw(\lambda z_1 \ldots z_k.z)$. Let the result be $X\hat{}$. Then we have the

Fact 1. X $\beta\eta$-converts to a Y-free term $\Leftrightarrow$ there exist Church numerals $\ldots \underline{n} \ldots$ of types $\ldots (A \to A) \to (A \to A) \ldots$ such that $[\ldots, \underline{n}/y, \ldots](\lambda z.X\hat{})$ $\underset{\beta\eta}{\twoheadleftarrow\!\!\twoheadrightarrow}$ to a term beginning with a dummy λ (vacuous abstraction).

Now let $\mathcal{M}$ be the full type structure over the two element domains $\{0, 1\}$. We observe that:

Fact 2. Let $M : A$ be closed. Then

$$M \underset{\beta\eta}{\twoheadleftarrow\!\!\twoheadrightarrow} \text{ to a term beginning with a dummy } \lambda$$

if and only if there exists a term N beginning with a dummy λ such that $\mathcal{M} \models M = N$.

This is true because in the long $\beta\eta$-normal form of M all the type 0 variables in the prefix besides the first can be interpreted as 1, the first can be interpreted as 0, and all the higher type variables as multivariate multiplication. This interpretation sends all normal terms with dummy first λ to 1 and all the other normal terms to 0. More precisely, for our application, working in $\mathcal{M}$, define functionals F_A recursively as follows. First, $F_0 = 1$. Next, for a type $A = A_1 \to (\ldots(A_t \to 0)\ldots)$, such that for $i = 1, \ldots, t$ $A_i = A_{i,1} \to (\ldots(A_{i,s_i} \to 0)\ldots)$, and $x_1 : A_1, \ldots, x_t : A_t$ set $F_A = \lambda x_1 \ldots x_t(x_1 F_{A_{1,1}} \ldots F_{A_{1,s_1}}) \bullet \cdots \bullet (x_t F_{A_{t,1}} \ldots F_{A_{t,s_t}})$ where $\bullet$ is multiplication mod 2. Now given $M : 0 \to A$ in normal form, and Y-free, M begins with a dummy λ if and only if $[\![M]\!]_{\mathcal{M}} 0 F_{A_1} \ldots F_{A_t} = 1$. Thus we have the following:

Corollary 3. *X $\underset{\beta\eta}{\twoheadleftarrow\!\!\twoheadrightarrow}$ a Y-free term $\Leftrightarrow$ there exist Church numerals $\ldots n \ldots$ of types $\ldots (A \to A) \to (A \to A) \ldots$ such that $[\ldots, n/y, \ldots]F(\lambda z.X\hat{})$ is equal in $\mathcal{M}$ to a term beginning with a dummy λ.*

Theorem 2. *The problem of determining whether X has a normal form is decidable.*

Proof. For any type A the finite set of denotations in $\mathcal{M}$ of the Church numerals of type $(A \to A) \to (A \to A)$ is easily computed. $\square$

5. The word problem

The general word problem is to determine, given closed terms M and N, whether $M \underset{\beta\eta Y}{\twoheadleftarrow\!\!\twoheadrightarrow} N$. It will be proved that this problem is undecidable. For an interesting variation the reader should consult Damm [4]. Here we shall solve the word problem for the special case where the only fixed-point combinator is $Y : (0 \to 0) \to 0$. Let $X : 0$. We define a modified Y reduction by

$$Y(\lambda x.X) \rightarrowtail [Y(\lambda x.X)/x]X.$$

This is a three step Y reduction. We also add the special case of Y reduction

$$Y(yU \dots V) \rightarrowtail yU \dots V(Y(yU \dots V))$$

where $yU \dots V : 0 \to 0$. Note that, since $Y : (0 \to 0) \to 0$, $\rightarrowtail$ does not create any new β redexes although η redexes can be created.

Lemma 2. *Suppose that $U : 0$ is in long $\beta\eta$-normal form with only $Y : (0 \to 0) \to 0$ and $U \underset{\beta Y}{\twoheadrightarrow} V$. Then there exists Z such that $U \rightarrowtail\!\!\!\twoheadrightarrow Z \underset{\beta}{\twoheadleftarrow} V$.*

Proof. By induction on the length of a standard reduction of U to V with a subsidiary induction on the depth of U. □

Proposition 3. *Suppose that $U, V : 0$ are long $\beta\eta$-normal forms with only $Y : (0 \to 0) \to 0$ and*

$$U \underset{\beta\eta Y}{\twoheadleftarrow\!\!\twoheadrightarrow} V.$$

Then there exists Z such that

$$U \rightarrowtail\!\!\!\twoheadrightarrow Z \twoheadleftarrow\!\!\!\leftarrowtail V.$$

Proof. First observe that if $U, V : 0$ are in long $\beta\eta$ normal form then

$$U \underset{\beta\eta Y}{\twoheadleftarrow\!\!\twoheadrightarrow} V \text{ if and only if } U \underset{\beta Y}{\twoheadleftarrow\!\!\twoheadrightarrow} V.$$

By the Church–Rosser theorem there exists Z s.t.

$$U \underset{\beta Y}{\twoheadrightarrow} Z \underset{\beta Y}{\twoheadleftarrow} V.$$

W.l.o.g. we can assume Z is β normal. Thus by the lemma $U \rightarrowtail\!\!\!\twoheadrightarrow Z \twoheadleftarrow\!\!\!\leftarrowtail V$. □

We shall now give a procedure for deciding whether $U \underset{\beta\eta Y}{\twoheadleftarrow\!\!\twoheadrightarrow} V$. We shall assume that both U, V are terms of type 0 in long $\beta\eta$-normal form. We shall also assume that in both U and V no variable is both free and bound and no variable is bound more than once (this property will not be preserved but it is useful at the start). We shall also assume that no variable is bound in both U and V. For each variable bound in U or V we introduce a new variable of the same type called the associated parameter. The free variables in U and V are also called parameters. If $U \underset{\beta\eta Y}{\twoheadleftarrow\!\!\twoheadrightarrow} V$ then by the above proposition there exists a W such that $U \rightarrowtail\!\!\!\twoheadrightarrow W \twoheadleftarrow\!\!\!\leftarrowtail V$. Now consider standard reductions from U to W and V to W. We have the following cases

(1) $U = uU' \dots U''$
(2) $U = Y(\lambda x X)$ and
   (a) $V = vV' \dots V''$
   (b) $V = Y(\lambda z Z)$

which give rise to 4 possible comparison cases:

(1a) Clearly we must have $u = v$ and each component of $U \underset{\beta\eta Y}{\twoheadleftarrow\!\!\twoheadrightarrow}$ to the corresponding component of V. For each such pair of components do the following which we illustrate on U' and V': write $U' = \lambda x_1 \ldots x_t.P$ and $V' = \lambda y_1 \ldots y_t.Q$ for $P, Q : 0$ and let $z_1, \ldots, z_t$ be the parameters corresponding to $x_1, \ldots, x_t$; repeat the procedure on $[z_1/x_1, \ldots, z_t/x_t]P$ and $[z_1/y_1, \ldots, z_t/y_t]Q$.

(1b) Repeat the procedure on U and $[V/z]Z$ (note that after the substitution the same variable may be bound in two different places, but no α-conversions are necessary).

(2a) Similar to (1b).

(2b) In this case we branch. The first branch repeats the procedure on X and Z, the second branch repeats the procedure on $[Y(\lambda xX)/x]X$ and $Y(\lambda zZ)$, the third branch repeats on $Y(\lambda xX)$ and $[Y(\lambda zZ)/z]Z$ and the fourth branch on $[Y(\lambda xX)/x]X$ and $[Y(\lambda zZ)/z]Z$.

Now it is clear that only finitely many terms can appear on each branch and so by Konig's lemma each branch terminates or repeats a term. This completes the proof.

6. Undecidability

We have the following fact:

Lemma 3. *If $m < n + 1$ then there exist $F : \$(n) \to (0 \to (0 \to 0))$ and $G : \$(n) \to \mathbb{N}$ such that*

(1) $F\, \underline{k} \underset{\beta\eta}{\twoheadleftarrow\!\!\twoheadrightarrow} K \Leftrightarrow k$ *is divisible by* m

$F\underline{k} \underset{\beta\eta}{\twoheadleftarrow\!\!\twoheadrightarrow} K^* \Leftrightarrow k$ *is not divisible by* m.

(2) $G\underline{k} \underset{\beta\eta}{\twoheadleftarrow\!\!\twoheadrightarrow} \underline{[k/m]}$.

Proof. Since $m \leq n$, $\%(n)$ can be used for m-tupling, [1, p. 130], and $\underline{k} : \$(n)$ can be used to iterate cyclic rotation. The construction is routine. □

We shall show that the general word problem is unsolvable by giving a simulation of register machines. We shall take a straightforward encoding of instantaneous descriptions d of register machines T as integers $\#d$. For each such T we choose a type $\$(n)$ depending on the number of states and registers in T, and we consider Church numerals $\underline{m}$ in $\$(n)$. We construct a closed term F such that T started in $d1$ and T started in $d2$ eventually go into the same instantaneous description if and only if

$$F\underline{\#d1} \underset{\beta\eta Y}{\twoheadleftarrow\!\!\twoheadrightarrow} F\underline{\#d2}.$$

So now let T be a register machine with k states and l registers.

Put $t = k + l$ and let $p(1), \ldots, p(t)$ be the first t primes. If the machine has $n(i)$ in register i and is in state j this instantaneous description d is encoded by the integer $\#d =$

$$p(1)^{n(1)} * \ldots * p(l)^{n(l)} * p(l + j).$$

Lemma 4. *There exists a closed term $G : \$(p(t)) \to \mathbb{N}$ depending only on T such that $G\underline{\#d} \underset{\beta\eta}{\twoheadleftarrow\!\!\twoheadrightarrow} \underline{\#\text{ the next instantaneous description of } T \text{ after } d}$.*

Note that because $G : \$(p(t)) \to \mathbb{N}$ *it is not immediate how to iterate* G *and this leads to the more difficult part of the construction. Note also that there is a closed term* $H : \$(p(t)) \to \mathbb{N}$ *such that* $H\underline{m} \underset{\beta\eta}{\twoheadleftarrow\!\!\twoheadrightarrow} \underline{m}$.

Proposition 4. *There exists a closed term* $J : \$(p(t)) \to ((\mathbb{N} \to \$(p(t))) \to (\$(p(t)) \to \$(p(t))))$ *such that* T *started in* $d1$ *and* T *started in* $d2$ *eventually go into the same instantaneous description if and only if*

$$J\underline{\#d1} \underset{\beta\eta Y}{\twoheadleftarrow\!\!\twoheadrightarrow} J\underline{\#d2}$$

Remark. Undecidability is an immediate consequence of Proposition 4.

Proof. The proof uses a typed Plotkin term construction similar to the one in [10]. Let G and H be as above; working in the context

$$\begin{aligned}
&u : \%(p(t)) \to \%(p(t)),\ v : \%(p(t)),\ x : \$(p(t)),\ y : \mathbb{N},\ z : \$(p(t)),\\
&f : D = \$(p(t)) \to (\mathbb{N} \to (\$(p(t)) \to (\$(p(t)) \to \$(p(t))))),\\
&g : \$(p(t)) \to \$(p(t)),
\end{aligned}$$

define

$$\begin{aligned}
S &= \lambda xuv.u(xuv) : \$(p(t)) \to \$(p(t))\\
M &= \lambda fgxyz.fx(Hx)[f(Sx)y(g(Sx))z] : D \to ((\$(p(t)) \to \$(p(t))) \to D)\\
N &= \lambda fgx.\ f(Sx)(Gx)(g(Sx))(gx) : D \to ((\$(p(t))\\
&\quad \to \$(p(t))) \to (\$(p(t)) \to \$(p(t))))\\
L &= Y(\lambda g.N(Y(\lambda f.Mfg))g) : \$(p(t)) \to \$(p(t))\\
P &= Y(\lambda f.MfL) : D
\end{aligned}$$

and, set

$$F = \lambda x\,P\underline{0}x(L\underline{0}) \qquad \text{and} \qquad J = F \circ G.$$

The proof that J is as desired is similar to the morphism representation theorem of [10]. First we calculate for any m, if $G\underline{m} \underset{\beta\eta Y}{\twoheadleftarrow\!\!\twoheadrightarrow} H\underline{k}$ then

$$\begin{aligned}
F(G\underline{m}) &\underset{\beta\eta Y}{\twoheadleftarrow\!\!\twoheadrightarrow} F(H\underline{k})\\
&\to P\underline{0}(H\underline{k})(L\underline{0})\\
&\underset{\beta\eta Y}{\twoheadrightarrow} P\underline{0}(H\underline{0})[P\underline{1}(H\underline{k})(L\underline{1})(L\underline{0})]\\
&\underset{\beta\eta Y}{\twoheadrightarrow} P\underline{0}(H\underline{0})\big(P\underline{1}(H\underline{1})\big(\ldots P\underline{k-1}(H\underline{k-1})\\
&\qquad [P\underline{k}(H\underline{k})(L\underline{k})(L\underline{k-1})]\ldots\big)(L\underline{1})\big)(L\underline{0})\\
&\underset{\beta\eta Y}{\twoheadrightarrow} P\underline{0}(H\underline{0})(P\underline{1}(H\underline{1})\big(\ldots P\underline{k-1}(H\underline{k-1})\big[P\underline{k}(H\underline{k})\\
&\qquad [P\underline{k+1}(G\underline{k})(L\underline{k+1})(L\underline{k})](L\underline{k-1})\big]\ldots\big)(L\underline{1}))(L\underline{0})\\
&\underset{\beta\eta Y}{\twoheadrightarrow} P\underline{0}(H\underline{0})\big(P\underline{1}(H\underline{1})(\ldots P\underline{k-1}(H\underline{k-1}))\big[P\underline{k}(H\underline{k})\\
&\qquad [P\underline{k+1}(G\underline{k})(L\underline{k+1})(L\underline{k})](L\underline{k-1})\big]\ldots\big)(L\underline{1})(L\underline{0})\\
&\underset{\beta\eta Y}{\twoheadleftarrow} F(G\underline{k}).
\end{aligned}$$

Thus we conclude that if T started in $d1$ goes into $d2$ then $F(G\underline{\#d1}) \underset{\beta\eta Y}{\twoheadleftarrow\!\!\twoheadrightarrow} F(G\underline{\#d2})$. Conversely, suppose that

$$\#d1 = m$$
$$\#d2 = n,$$

and that

$$P\underline{k}(G\underline{m})(L\underline{k}) \underset{\beta\eta Y}{\twoheadleftarrow\!\!\twoheadrightarrow} P\underline{k}(G\underline{n})(L\underline{k}).$$

Then by the Church–Rosser and standardization theorems there exists a common reduct Z and standard reductions

$$R_1 : P\underline{k}(G\underline{m})(L\underline{k}) \underset{\beta\eta Y}{\twoheadrightarrow} Z$$

and

$$R_2 : P\underline{k}(G\underline{n})(L\underline{k}) \underset{\beta\eta Y}{\twoheadrightarrow} Z.$$

More generally, suppose we have standard reduction sequences

$$R_1 : PU(G\underline{m})(LU) \underset{\beta\eta Y}{\twoheadrightarrow} Z$$
$$R_2 : PU(G\underline{n})(LU) \underset{\beta\eta Y}{\twoheadrightarrow} Z$$

where $U \underset{\beta\eta}{\twoheadrightarrow} \underline{k}$ for some k.

We proceed by induction on the sum of the lengths of these standard reduction sequences. We shall show that T started in $d1$ and T started in $d2$ eventually wind up in the same instantaneous description. In this case we shall say that m and n are equivalent.

We intend to take certain liberties below by replacing the above terms U and other similar terms by ambiguous notation $\boxed{\underline{\mathrm{k}}}$ indicating their Church numeral normal forms. This does not effect the argument and makes notation much simpler.

Basis: The sum of the lengths of the reductions is 0. In this case it is clear that $G\underline{m} \underset{\beta\eta Y}{\twoheadleftarrow\!\!\twoheadrightarrow} G\underline{n}$ and hence, by the definition of G, m and n are equivalent.

Induction step: The sum of the lengths is say $r > 0$. For any term X the head reduction of $P\boxed{\underline{\mathrm{k}}}\,X(L\boxed{\underline{\mathrm{k}}})$ cycles through segments which are 8 terms long viz:

$$\begin{array}{rcl}
P\boxed{\underline{\mathrm{k}}}X(L\boxed{\underline{\mathrm{k}}}) & = & Y(\lambda f.MfL)\boxed{\underline{\mathrm{k}}}X(L\boxed{\underline{\mathrm{k}}}) \\
& \underset{\beta\eta Y}{\twoheadrightarrow} & (\lambda x.x(Yx))(\lambda f.MfL)\boxed{\underline{\mathrm{k}}}X(L\boxed{\underline{\mathrm{k}}}) \\
& \underset{\beta\eta Y}{\twoheadrightarrow} & (\lambda f.MfL)P\boxed{\underline{\mathrm{k}}}X(L\boxed{\underline{\mathrm{k}}}) \\
& \underset{\beta\eta Y}{\twoheadrightarrow} & MPL\boxed{\underline{\mathrm{k}}}X(L\boxed{\underline{\mathrm{k}}}) \\
& \underset{\beta\eta Y}{\twoheadrightarrow} & (\lambda gxyz.Px\ (Hx)[P(Sx)y(g(Sx))z])L\boxed{\underline{\mathrm{k}}}X(L\boxed{\underline{\mathrm{k}}}) \\
& \underset{\beta\eta Y}{\twoheadrightarrow} & (\lambda xyz.Px(Hx)[P(Sx)y(L(Sx))z])\boxed{\underline{\mathrm{k}}}X(L\boxed{\underline{\mathrm{k}}}) \\
& \underset{\beta\eta Y}{\twoheadrightarrow} & (\lambda yz.P\boxed{\underline{\mathrm{k}}}(H\boxed{\underline{\mathrm{k}}})[P(S\boxed{\underline{\mathrm{k}}})y(L(S\boxed{\underline{\mathrm{k}}}))z])X(L\boxed{\underline{\mathrm{k}}}) \\
& \underset{\beta\eta Y}{\twoheadrightarrow} & (\lambda z.P\boxed{\underline{\mathrm{k}}}(H\boxed{\underline{\mathrm{k}}})[P(S\boxed{\underline{\mathrm{k}}})X(L(S\boxed{\underline{\mathrm{k}}}))z])(L\boxed{\underline{\mathrm{k}}}) \\
& \underset{\beta\eta Y}{\twoheadrightarrow} & P\boxed{\underline{\mathrm{k}}}(H\boxed{\underline{\mathrm{k}}})[P(S\boxed{\underline{\mathrm{k}}})X(L(S\boxed{\underline{\mathrm{k}}}))(L\boxed{\underline{\mathrm{k}}})]
\end{array}$$

where $H\boxed{\underline{k}}$ becomes the new X and $P(S\boxed{\underline{k}})X(L(S\boxed{\underline{k}}))(L\underline{k})$ becomes the new $L\boxed{\underline{k}}$. First observe that the head reduction parts of R1 and R2 cannot terminate at different spots in the 8 term cycle. This can be seen by computing the cut type of the head redex. If neither R1 nor R2 completes the first full 8 term cycle then clearly $G\underline{m} \twoheadleftarrow\!\!\twoheadrightarrow_{\beta\eta Y} G\underline{n}$ except possibly in the 8th case. In the 8th case we have standard reductions from

$$P\boxed{\underline{k}}(H\boxed{\underline{k}})[P(S\boxed{\underline{k}})(G\underline{m})(L(S\boxed{\underline{k}}))z]$$

and

$$P\boxed{\underline{k}}(H\boxed{\underline{k}})[P(S\boxed{\underline{k}})(G\underline{n})(L(S\boxed{\underline{k}}))z]$$

to a common reduct. The sum of the lengths of these reductions is $<r$. By the above analysis of the 8 term cycle the induction hypothesis applies to $P(S\boxed{\underline{k}})(G\underline{m})(L(S\boxed{\underline{k}}))$ and $P(S\boxed{\underline{k}})(G\underline{n})(L(S\boxed{\underline{k}}))$. Thus m and n are equivalent. Thus we may assume that $P\boxed{\underline{k}}(G\underline{m})(L\boxed{\underline{k}})$ completes the full 8 term cycle at least once and its head reduction part ends in a term

$$U(P(S\boxed{\underline{k}})(H\boxed{\underline{k}})(L(S\boxed{\underline{k}}))(\ldots(P(S\boxed{\underline{k}})(H\boxed{\underline{k}})(L(S\boxed{\underline{k}})))$$
$$[P(S\boxed{\underline{k}})(G\underline{m})(L(S\boxed{\underline{k}}))(L\boxed{\underline{k}})]\ldots))$$

where $P(S\boxed{\underline{k}})(H\boxed{\underline{k}})(L(S\boxed{\underline{k}}))$ appears $t \geq 0$ times and $P\boxed{\underline{k}}(H\boxed{\underline{k}})$ head reduces to U in ≤ 7 steps. We distinguish two cases.

Case 1: $P\boxed{\underline{k}}(G\underline{n})(L\boxed{\underline{k}})$ does not complete the full cycle of the first 8 head reductions. In this case the head reduction part of $P\boxed{\underline{k}}(G\underline{n})(L\boxed{\underline{k}})$ ends in a term $V(L\boxed{\underline{k}})$ where $P\boxed{\underline{k}}(G\underline{n})$ head reduces to V in the same ≤ 7 steps. If the number of steps is indeed <7 then clearly $G\underline{n} \twoheadleftarrow\!\!\twoheadrightarrow_{\beta\eta Y} H\boxed{\underline{k}}$.

Suppose then that

$$U = \lambda z.P\boxed{\underline{k}}(H\boxed{\underline{k}})[P(S\boxed{\underline{k}})(H\boxed{\underline{k}})(L(S\boxed{\underline{k}}))z]$$

and

$$V = \lambda z.P\boxed{\underline{k}}(H\boxed{\underline{k}})[P(S\boxed{\underline{k}})(G\underline{n})(L(S\boxed{\underline{k}}))z].$$

Then $Z = W''W'''$ with standard reductions of U and V to W'' of total length $<r$. By our previous analysis of the 8 term cycle the induction hypothesis applies to the pair $P(S\boxed{\underline{k}})(H\boxed{\underline{k}})(L(S\boxed{\underline{k}}))$ and $P(S\boxed{\underline{k}})(G\underline{n})(L(S\boxed{\underline{k}}))$. Thus in any case $G\underline{n} \twoheadleftarrow\!\!\twoheadrightarrow_{\beta\eta Y} H\boxed{\underline{k}}$. In addition, the arguments of U and $L\boxed{\underline{k}}$ have a common reduct by standard reductions whose lengths are less than R1 and R2, respectively. Since the $t+1$ components of the argument of U are of order 0, we have $Z = Z_0(Z_1(\ldots(Z_t(WW'))\ldots))$ and $P(S\boxed{\underline{k}})(H\boxed{\underline{k}})(L(S\boxed{\underline{k}})) \twoheadrightarrow_{\beta\eta Y} Z_i$ and $P(S\boxed{\underline{k}})(G\underline{m})(L(S\boxed{\underline{k}})) \twoheadrightarrow_{\beta\eta Y} W$ for $i = 1, \ldots, t$

all by standard reductions of length less than R1. In addition, the head reduction of $L\boxed{\underline{k}}$ begins

$$
\begin{aligned}
L\boxed{\underline{k}} &= Y(\lambda u.N(Y(\lambda v.Mvu))u)\boxed{\underline{k}}\\
&\underset{\beta\eta Y}{\twoheadrightarrow} (\lambda x.x(Yx))(\lambda u.N(Y(\lambda v.Mvu))u)\boxed{\underline{k}}\\
&\underset{\beta\eta Y}{\twoheadrightarrow} (\lambda u.N(Y(\lambda v.Mvu))u)L\boxed{\underline{k}}\\
&\underset{\beta\eta Y}{\twoheadrightarrow} NPL\boxed{\underline{k}}\\
&\underset{\beta\eta Y}{\twoheadrightarrow} (\lambda gx.P(Sx)(Gx)(g(Sx))(gx))L\boxed{\underline{k}}\\
&\underset{\beta\eta Y}{\twoheadrightarrow} (\lambda x.P(Sx)(Gx)(L(Sx))(Lx))\boxed{\underline{k}}\\
&\underset{\beta\eta Y}{\twoheadrightarrow} P(S\boxed{\underline{k}})(G\boxed{\underline{k}})(L(S\boxed{\underline{k}}))(L\boxed{\underline{k}})
\end{aligned}
$$

and none of the heads of any of these terms except the last is of order 0. Thus the head reduction part of the standard reduction of $L\boxed{\underline{k}}$ goes at least this far. Indeed, we may assume that $t = 0$ and the induction hypothesis applies to the pair $P(S\boxed{\underline{k}})(G\underline{m})(L(S\boxed{\underline{k}}))$ and $P(S\boxed{\underline{k}})(G\boxed{\underline{k}})(L(S\boxed{\underline{k}}))$. This completes the proof for Case 1.

Case 2: Both R1 and R2 complete the full cycle of the first 8 head reductions. W.l.o.g. we may assume that the head reduction part of R2 ends in

$$
\begin{aligned}
&V((P(S\boxed{\underline{k}})(H\boxed{\underline{k}})(L(S\boxed{\underline{k}}))\\
&(\ldots(P(S\boxed{\underline{k}})(H\boxed{\underline{k}})(L(S\boxed{\underline{k}})))[P(S\boxed{\underline{k}})(G\underline{n})(L(S\boxed{\underline{k}}))(L\boxed{\underline{k}})]\ldots)),
\end{aligned}
$$

where $P(S\boxed{\underline{k}})(H\boxed{\underline{k}})(L(S\boxed{\underline{k}}))$ appears $s \le t$ times, and $P(S\boxed{\underline{k}})(H\boxed{\underline{k}})$ head reduces to V in ≤ 7 steps. We distinguish two subcases.

Subcase 1: $s = t$. In this case, the induction hypothesis applies to the pair $P(S\boxed{\underline{k}})(G\boxed{\underline{m}})(L(S\boxed{\underline{k}}))$ and $P(S\boxed{\underline{k}})(G\underline{n})(L(S\boxed{\underline{k}}))$.

Subcase 2: $s < t$. In this case, by Fact 1, the induction hypothesis applies to the pair

$$P(S\boxed{\underline{k}})(G\boxed{\underline{n}})(L(S\boxed{\underline{k}}))$$

and

$$P(S\boxed{\underline{k}})(H\boxed{\underline{k}})(L(S\boxed{\underline{k}})).$$

In addition, $L\boxed{\underline{k}}$ and

$$
\begin{aligned}
&P(S\boxed{\underline{k}})(H\boxed{\underline{k}})(L(S\boxed{\underline{k}}))(\ldots(P(S\boxed{\underline{k}})(H\boxed{\underline{k}})(L(S\boxed{\underline{k}})))\\
&\Big[P(S\boxed{\underline{k}})(G\underline{m})(L(S\boxed{\underline{k}}))(L\boxed{\underline{k}})\Big]\ldots),
\end{aligned}
$$

with $t - s$ occurrences of $P(S\boxed{\underline{k}})(H\boxed{\underline{k}})(L(S\boxed{\underline{k}}))$, have a common reduct by standard reductions whose total length is $<r$. Here the argument of case 1 applies and this completes the proof. $\square$

Remark. Proposition 4 could be generalized as follows:

Let $G : C \to \mathbb{N}$ be any closed term and let $\sim$ be the equivalence relation generated by $G\underline{m} \underset{\beta\eta}{\twoheadleftarrow\!\!\twoheadrightarrow} \underline{n} \Rightarrow m \sim n$. Then there exists a term $J : C \to (C \to C)$ such that

$$m \sim n \Leftrightarrow J\underline{m} \underset{\beta\eta Y}{\twoheadleftarrow\!\!\twoheadrightarrow} J\underline{n}.$$

Acknowledgements

The author would like to thank the referees for several helpful additions, corrections and simplifications.

References

[1] H.P. Barendregt, The Lambda Calculus, North-Holland, Amsterdam, 1981.
[2] I. Bercovici, Unsolvable terms in typed λ-calculus with fixed-point operators, LNCS, vol. 193, 1985, pp. 16–22.
[3] H.B. Curry, R. Feys, Combinatory Logic, vol. 1, North-Holland, Amsterdam, 1958.
[4] W. Damm, The *IO* and *OI* hierachies, Theoret. Comput. Sci. 20 (1982) 95–207.
[5] M.J. Gordon, A.J. Milner, C.P. Wadsworth, Edinburgh LCF, Lecture Notes in Computer Science, vol. 78, Springer-Verlag, Berlin, 1979.
[6] R. Platek, Dissertation, Stanford University, Department of Mathematics, 1963.
[7] G. Plotkin, C. Stirling, M. Tofte (Eds.), Proof, Language, and Interaction: Essays in Honour of Robin Milner, MIT Press, Cambridge, 2000.
[8] D. Scott, A type-theoretic equivalent to ISWIM, CUCH, OWHY, in: Dezani-Ciancaglini, Ronci della Rocca, Venturini Zilli (Eds.), A Collection of Contributions in Honor of Corrado Böhm, Elsevier, 1993, pp. 411–440.
[9] R. Statman, Completeness, invariance and lambda definability, J. Symbolic Logic 47 (1982) 17–26.
[10] R. Statman, Morphisms and partitions of V-sets, CSL 98, LNCS, vol. 1584, 1998, pp. 313–322.
[11] W.W. Tait, Constructive reasoning, LMPS, vol. III, North-Holland, Amsterdam, 1968, pp. 185–199.