

Upper Bounds for Standardizations and an Application

Hongwei Xi

Department of Mathematical Sciences
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

Telephone: +1 412 268-1439 Fax: +1 412 268-6380

Email: hwxi@cs.cmu.edu

Abstract

We first present a new proof for the standardization theorem, a fundamental theorem in λ -calculus. Since our proof is largely built upon structural induction on lambda terms, we can extract some bounds for the number of β -reduction steps in the standard β -reduction sequences obtained from transforming any given β -reduction sequences. This result sharpens the standardization theorem and establishes a link between lazy and eager evaluation orders in the context of computational complexity. As an application, we establish a superexponential bound for the number of β -reduction steps in β -reduction sequences from any given simply typed λ -terms.

1 Introduction

The standardization theorem of Curry and Feys [CF58] is a very useful result, stating that if u reduces to v for λ -terms u and v , then there is a *standard* β -reduction from u to v . Using this theorem, we can readily prove the normalization theorem, i.e., a λ -term has a normal form if and only if the leftmost β -reduction sequence from the term is finite. The importance of lazy evaluation in functional programming languages largely comes from the normalization theorem. Moreover, the standardization theorem can be viewed as a syntactic version of sequentiality theorem in [Ber78]. For instance, it can be readily argued that *parallel or* is inexpressible in λ -calculus by using the standardization theorem. In fact, a syntactic proof of the sequentiality theorem can be given with the help of the standardization theorem.

There have been many proofs of the standardization theorem in the literature such as the ones in [Mit79], [Klo80], [Bar84] and [Tak95]. In the presented proof we intend to find a bound for standardizations, namely, to measure the number of steps in the standard β -reduction sequence obtained from a given β -reduction sequence. This method presents a concise and more accurate formulation of the standardization theorem. As an application, we establish a super exponential bound on the number of β -reduction steps in β -reduction sequences from any given simply typed λ -terms. This not only strengthens the strong normalization theorem in the simply typed λ -calculus λ^{\rightarrow} , but also yields more understanding on $\mu(t)$, the number of steps in a longest β -reduction sequence from a given simply typed λ -term t . Since $\mu(t)$ can often be used as an induction order, its structure plays a key role in understanding related inductive proofs.

The structure of the paper is as follows. The notations and basics are explained in Section 2. In Section 3, our proof of the standardization theorem is presented. Some upper bounds for standardizations are extracted in Section 4. In Section 5, we establish a bound for the number

of β -reduction steps in reduction sequences from any given simply typed λ -terms. Finally, some related work is mentioned and a few remarks are drawn in Section 6.

2 Preliminaries

We give a brief explanation on the notations and terminology used in this paper. Most details not included here appear in [Bar84].

Definition 1 (*λ -terms*) *The set Λ of λ -terms is defined inductively as follows.*

- (*variable*) *There are infinitely many variables x, y, z, \dots in Λ ; variables are the only subterms of themselves.*
- (*abstraction*) *If $t \in \Lambda$ then $(\lambda x.t) \in \Lambda$; u is a subterm of $(\lambda x.t)$ if u is $(\lambda x.t)$ or a subterm of t .*
- (*application*) *If $t_0, t_1 \in \Lambda$ then $t_0(t_1) \in \Lambda$; u is a subterm of $t_0(t_1)$ if u is $t_0(t_1)$ or a subterm of t_i for some $i \in \{0, 1\}$.*

The set $\text{FV}(t)$ of free variables in t is defined as follows.

$$\text{FV}(t) = \begin{cases} \{x\} & \text{if } t = x \text{ for some variable } x; \\ \text{FV}(t_0) - \{x\} & \text{if } t = (\lambda x.t_0); \\ \text{FV}(t_0) \cup \text{FV}(t_1) & \text{if } t = t_0(t_1). \end{cases}$$

The set Λ_I of λI -terms is the maximal subset of Λ such that, for every term $t \in \Lambda_I$, if $(\lambda x.t_0)$ is a subterm of t then $x \in \text{FV}(t_0)$.

$u\{x := v\}$ stands for substituting v for all free occurrences of x in u . α -conversion or renaming bounded variables may have to be performed in order to avoid name collisions. Rigorous definitions are omitted here. We assume some basic properties on substitution such as the substitution lemma (Lemma 2.1.16 [Bar84]).

Definition 2 (*β -redex, β -reduction and β -normal form*) *A term of form $(\lambda x.u)(v)$ is called a β -redex, and $u\{x := v\}$ is called the contractum of the β -redex; $t \rightarrow_\beta t'$ stands for a β -reduction step where t' is obtained from replacing some β -redex in t with its contractum; a β -normal form is a term in which there is no β -redex.*

“ β -” is often omitted if this causes no confusion or ambiguity. Let \rightarrow_β^n stand for n steps of β -reduction, and \twoheadrightarrow_β stand for some steps of β -reduction, which could be 0. Usually there are many different β -redexes in a term t ; a β -redex r_1 in t is to the left of another β -redex r_2 in t if the first symbol of r_1 is to the left of that of r_2 .

Definition 3 (*Multiplicity*) *Given a β -redex $r = (\lambda x.u)(v)$; the multiplicity $m(r)$ of r is the number of free occurrences of the variable x in u .*

Definition 4 (*β -Reduction Sequence*) *Given a β -redex r in t ; $t \xrightarrow{r}_\beta u$ stands for the β -reduction step in which β -redex r gets contracted; $[r_1] + \dots + [r_n]$ stands for a β -reduction sequence of the following form.*

$$t_0 \xrightarrow{r_1}_\beta t_1 \xrightarrow{r_2}_\beta \dots \xrightarrow{r_n}_\beta t_n$$

Conventions σ, τ, \dots range over β -reduction sequences; $\sigma : t \twoheadrightarrow_\beta t'$ or $t \xrightarrow{\sigma}_\beta t'$ stands for a β -reduction sequence from t to t' ; $|\sigma|$ is the length of σ , namely, the number of β -reduction steps in σ , which might be 0.

Definition 5 (*Concatenation*) Given $\sigma : t_0 \twoheadrightarrow_{\beta} t_1$ and $\tau : t_1 \twoheadrightarrow_{\beta} t_2$; $\sigma + \tau$ stands for the concatenation of σ and τ , namely, $\sigma + \tau : t_0 \xrightarrow{\sigma} t_1 \xrightarrow{\tau} t_2$.

Conventions Let $\sigma : u \twoheadrightarrow_{\beta} v$ and $C[\]$ be a context, then σ can also be regarded as the β -reduction sequence which reduces $C[u]$ to $C[v]$ in the obvious way. In other words, we may use σ to stand for $C[\sigma]$. An immediate consequence of this is that $\sigma + \tau$ can be regarded as $C_1[\sigma] + C_2[\tau]$ for some proper contexts C_1 and C_2 .

Before moving forward, let us introduce the concept of *residuals* of β -redexes. The rigorous definition of this notion can be found in [Hue94]. Let \mathcal{R} be a set of β -redexes in a term t , $r = (\lambda x.u)(v)$ in \mathcal{R} , and $t \xrightarrow{r}_{\beta} t'$. This β -reduction step affects β -redexes r' in \mathcal{R} in the following way. We assume that all bound variables in u are distinct from the free variables in v .

- r' is r . Then r' has no residual in t' .
- r' is in v . All copies of r' in $u\{x := v\}$ are called residuals of r' in t' ;
- r' is in u . Then $r'\{x := v\}$ in $u\{x := v\}$ is the only residual of r' in t' ;
- r' contains r . Then the residual of r' is the term obtained by replacing r in r' with $u\{x := v\}$;
- Otherwise, r' is not affected and is its own residual in t' .

The residual relation is transitive.

Definition 6 (*Developments*) Given a λ -term t and a set \mathcal{R} of redexes in t ; if $\sigma : t \twoheadrightarrow_{\beta} u$ contracts only β -redexes in \mathcal{R} or their residuals, then σ is a development.

Definition 7 (*Involvement*) Given $t \twoheadrightarrow_{\beta} u$ and $\sigma : u \twoheadrightarrow_{\beta} v$; a β -redex in t is involved in σ if some of its residuals is contracted in σ .

Definition 8 (*Head β -redex*) Given t of form $\lambda x_1 \dots \lambda x_m.r(t_1) \dots (t_n)$, where r is a β -redex and $m, n \geq 0$; r is called the head β -redex in t ; a β -reduction is a head β -reduction if the contracted β -redex is a head β -redex.

Proposition 9 We have the following.

- Let r_h be the head β -redex in t ; if $t \xrightarrow{r}_{\beta} u$ for some $r \neq r_h$ then r_h has exactly one residual in u , which is the head β -redex of u .
- If $\sigma : t \twoheadrightarrow_{\beta} u$ contains a head β -reduction, then t contains a head β -redex r_h and r_h is involved in σ .

Proof Please see Section 8.3 [Bar84] for proofs. ■

3 The Proof of Standardization Theorem

Standardization theorem of Curry and Feys [CF58] states that any β -reduction sequence can be standardized in the following sense.

Definition 10 (*Standard β -Reduction Sequence*) Given a β -reduction sequence

$$\sigma : t = t_0 \xrightarrow{r_0}_{\beta} t_1 \xrightarrow{r_1}_{\beta} t_2 \xrightarrow{r_2}_{\beta} \dots;$$

σ is standard if for all $0 \leq i < j$, r_j is not a residual of some β -redex to the left of r_i .

Lemma 11 Given $t \xrightarrow{\sigma_s} u \xrightarrow{r} v$, where σ_s is standard and r is the residual of the head β -redex in t ; then we can construct a standard β -reduction sequence $t \xrightarrow{\tau} v$ with $|\tau| \leq 1 + \max\{m(r), 1\} \cdot |\sigma_s|$.

Proof Let r_h be the head β -redex in t , and we proceed by structural induction on t .

- $t = (\lambda x.t_0)$. By the induction hypothesis on t_0 , this case is trivial.
- $t = t_0(t_1)$, and r_h is in t_0 . Then we may assume $\sigma_s = \sigma_0 + \sigma_1$, where $t_i \xrightarrow{\sigma_i} u_i$ are standard for $i = 0, 1$, and $u = u_0(u_1)$. Note that r must be in u_0 by Proposition 9 (1). Hence $v = v_0(u_1)$, where $u_0 \xrightarrow{r} v_0$. By induction hypothesis, we can construct a standard β -reduction sequence $t_0 \xrightarrow{\tau_0} v_0$ with $|\tau_0| \leq 1 + \max\{m(r), 1\} \cdot |\sigma_0|$. Let $\tau = \tau_0 + \sigma_1$, then $t \xrightarrow{\tau} v$ is standard. It can be readily verified that $|\tau| \leq 1 + \max\{m(r), 1\} \cdot |\sigma_s|$.
- $t = (\lambda x.t_0)(t_1)$, and r_h is t . Then we can assume $\sigma_s = \sigma_0 + \sigma_1$, where $t_i \xrightarrow{\sigma_i} u_i$ are standard for $i = 0, 1$, and $r = u = (\lambda x.u_0)u_1$. Hence, $v = u_0\{x := u_1\}$. Let $\sigma_0 = [r_1] + \dots + [r_n]$, and $\sigma_0^* = [r_1^*] + \dots + [r_n^*]$, where $r_j^* = r_j\{x := t_1\}$ for $j = 1, \dots, n$. Then we know $\sigma_0^* : t_0\{x := t_1\} \rightarrow u_0\{x := t_1\}$ is also standard. Notice that $\sigma_0^* + \sigma_1 + \dots + \sigma_1$ is a β -reduction sequence which reduces $t_0\{x := t_1\}$ to $v = u_0\{x := u_1\}$, where σ_1 occurs $m(r)$ times and each σ_1 reduces one occurrence of t_1 in $u_0\{x := t_1\}$ to u_1 . If a β -redex contracted in some σ_1 is to the left of some r_j^* , then all β -redexes contracted in that σ_1 are to the left of that r_j^* . Hence, we can move that σ_1 to the front of r_j^* . In this way, we can construct a standard β -reduction sequence from $t_0\{x := t_1\}$ to $v = u_0\{x := u_1\}$ in the following form.

$$\sigma_s^* = \dots + [r_1^*] + \dots + \dots + \dots + [r_n^*] + \dots,$$

where \dots stands for a β -reduction sequence of form $\sigma_1 + \dots + \sigma_1$, which may be empty, and r_j^* may also denote their corresponding residuals. Hence $\tau = [r_h] + \sigma_s^*$ is a standard β -reduction sequence from t to v . Notice

$$|\tau| = 1 + |\sigma_s^*| = 1 + |\sigma_0^*| + m(r) * |\sigma_1| \leq 1 + \max\{m(r), 1\} \cdot |\sigma_s|.$$

■

Lemma 12 Given $t \xrightarrow{\sigma_s} u \xrightarrow{r} v$, where σ_s is standard; then we can construct a standard β -reduction sequence $t \xrightarrow{\tau} v$ with $|\tau| \leq 1 + \max\{m(r), 1\} \cdot |\sigma_s|$.

Proof The proof proceeds by induction on $|\sigma_s|$ and the structure of t , lexicographically ordered. By Corollary 8.3.8 [Bar84], t is of form $\lambda x_1 \dots \lambda x_m.t_0(t_1) \dots (t_n)$, where $m, n \geq 0$, and t_0 is either a variable or an λ -abstraction. We have two cases.

- $\sigma_s + [r]$ contains no head β -reduction. Then we may assume σ_s to be of form

$$\sigma_{0,s} + \sigma_{1,s} + \dots + \sigma_{n,s},$$

where $\sigma_{i,s}$ are standard β -reduction sequences from t_i to u_i for $i = 0, 1, \dots, n$ and $u = \lambda x_1 \dots \lambda x_m.u_0(u_1) \dots (u_n)$. Note that r must be in some u_k . Let $u_k \xrightarrow{r} v_k$. By induction hypothesis, there exists a standard β -reduction sequence τ_k from t_k to v_k with $|\tau_k| \leq 1 + \max\{m(r), 1\} \cdot |\sigma_{k,s}|$. Let $\tau = \sigma_{0,s} + \dots + \tau_k + \dots + \sigma_{n,s}$, then τ is standard.

- $\sigma_s + [r]$ contains some head β -reduction. By Proposition 9 (2), let r_h be the head β -redex in t , which is involved in $\sigma_s + [r]$. We have two subcases.
 - r is the residual of r_h . Then by Lemma 11 we are done.

- r_h is involved in σ_s . Since σ_s is standard, $\sigma_s = [r_h] + \sigma_{h,s} : t \xrightarrow{r_h} \beta t_h \xrightarrow{\sigma_{h,s}} \beta u$ for some standard β -reduction sequence $\sigma_{h,s}$. Note $|\sigma_{h,s}| < |\sigma_s|$. By induction hypothesis, we can construct a standard β -reduction sequence $\tau_h : t_h \longrightarrow \beta v$. Clearly, $\tau = [r_h] + \tau_h : t \longrightarrow \beta v$ is a standard β -reduction sequence.

In either case, it can be immediately verified that $|\tau| \leq 1 + \max\{m(r), 1\} \cdot |\sigma_s|$. ■

Theorem 13 (*Standardization*) *Every finite β -reduction sequence can be standardized.*

Proof Given $t \xrightarrow{\sigma} \beta v$, we proceed by induction on $|\sigma|$. If σ is empty then σ is standard. Now assume $\sigma = \sigma' + [r]$, where $t \xrightarrow{\sigma'} \beta u \xrightarrow{r} \beta v$. By induction hypothesis, we can construct a standard β -reduction sequence $t \xrightarrow{\sigma'_s} \beta u$. Hence, Lemma 12 yields the result. ■

4 The Upper Bounds

It is clear from the previous proofs that we actually have an algorithm to transform any β -reduction sequences into standard ones. Let $\text{std}(\sigma)$ denote the standard β -reduction sequence obtained from transforming a given β -reduction sequence σ . We are ready to give some upper bounds on the number of β -reduction steps in $\text{std}(\sigma)$.

Theorem 14 (*Standardization with bound*) *Given a β -reduction sequence $t \xrightarrow{\sigma} \beta u$, where $\sigma = [r_0] + [r_1] + \dots + [r_n]$ for some $n \geq 1$, then there exists a standard β -reduction sequence $t \xrightarrow{\sigma_s} \beta u$ with $|\sigma_s| \leq (1 + \max\{m(r_1), 1\}) \cdots (1 + \max\{m(r_n), 1\})$.*

Proof Let $\sigma_0 = [r_0]$, $\sigma_i = [r_0] + [r_1] + \dots + [r_i]$ and $l_i = |\text{std}(\sigma_i)|$ for $i = 0, 1, \dots, n$. By Lemma 12, we have $l_{i+1} \leq 1 + \max\{m(r_{i+1}), 1\} \cdot l_i$ for $i = 0, 1, \dots, n-1$ according to the proof of Theorem 13. Note $1 \leq l_i$, and thus, for $i = 0, 1, \dots, n-1$,

$$l_{i+1} \leq 1 + \max\{m(r_{i+1}), 1\} \cdot l_i \leq (1 + \max\{m(r_{i+1}), 1\}) \cdot l_i.$$

Since $l_0 = 1$, this yields $l_n \leq (1 + \max\{m(r_1), 1\}) \cdots (1 + \max\{m(r_n), 1\})$. Note $\sigma = \sigma_n$. Let $\sigma_s = \text{std}(\sigma_n)$, then we are done. ■

Clearly, this simple bound is not very tight. With a closer study, a tighter but more complex bound can be given in the same fashion. Unlike many earlier proofs in the literature, our proof of the standardization theorem does not use the *finiteness of developments theorem* (FD). In this respect, our proof is similar to the one in [Tak95]. We remark that the uses of FD in other proofs are inessential in general and can be avoided in proper ways. As a matter of fact, Theorem 14 can be modified to show that all developments are finite, following the application in the next section. We will not pursue in this direction since the work in [dV85] has produced an exact bound for finiteness of developments.

Given $\sigma : t \longrightarrow \beta u$, we can also give a bound on $|\text{std}(\sigma)|$ in terms of $|\sigma|$ and the size of t defined below.

Definition 15 *The size $|t|$ of a term t is defined inductively as follows.*

$$|t| = \begin{cases} 1 & \text{if } t \text{ is a variable;} \\ 1 + |t_0| & \text{if } t = (\lambda x.t_0); \\ |t_0| + |t_1| & \text{if } t = t_0(t_1). \end{cases}$$

Proposition 16 *If $t \longrightarrow \beta u$ then $|u| < |t|^2$.*

Proof A structural induction on t yields the result. ■

Corollary 17 *Given $\sigma : t \twoheadrightarrow_{\beta} u$; then there is a standard β -reduction sequence $\sigma_s : t \twoheadrightarrow_{\beta} u$ with $|\sigma_s| < |t|^{2^{|\sigma|}}$.*

Proof This clearly holds if $|\sigma| = 1$. Now assume $\sigma = [r_0] + [r_1] + \dots + [r_{n-1}] : t \twoheadrightarrow_{\beta} u$. By Proposition 16, we have $|r_i| < |t|^{2^i}$ for $i = 1, \dots, n-1$, which yields $1 + \max\{m(r_i), 1\} \leq |t|^{2^i}$ for $i = 1, \dots, n-1$. By Theorem 14, we can construct a standard β -reduction sequence $\sigma_s : t \twoheadrightarrow_{\beta} u$ with $|\sigma_s| \leq |t|^{2^1} \dots |t|^{2^{n-1}} < |t|^{2^n}$. ■

Now we introduce a lemma which will be used in the next section.

Lemma 18 *If $\sigma : t \twoheadrightarrow_{\beta} u$ is a development, then $|u| < 2^{|t|}$.*

Proof This can be verified by a structural induction on t . ■

5 An Application

It is a well-known fact that the simply typed λ -calculus λ^{\rightarrow} enjoys strong normalization. Given a simply typed λ -term t , let $\mu(t)/\nu(t)$ be the length of a longest/shortest β -reduction sequence which reduces t to a β -normal form. In this section, as an application of our previous result, we will give an upper bound on $\mu(t)$. We first show for simply typed λ -terms t that $\nu(t)$ is not an elementary. This can bring us some feeling on how tight our upper bound on $\mu(t)$ is. Among various proofs showing the strong normalization property of λ^{\rightarrow} , some proofs such as the ones in [Gan80a] and [Sch91] can yield superexponential upper bounds on $\mu(t)$. Gandy invented a semantic approach in [Gan80a], which is called *functional interpretations* and has its traces in many following papers such as [Sch82], [Pol94] and [Kah95]. In [Sch91], Schwichtenberg adopted a syntactic approach from [How80], which bases on cut elimination in intuitionistic logic.

Compared with other related methods in the literature, our following syntactic method is not only innovative but also yields a quite intelligible and tight bound. It also exhibits a nice way in λ^{\rightarrow} to transform strong normalization into weak normalization, simplifying a much involved transformation in [Sch91]. Therefore, the new transformation has its own value in this respect.

Definition 19 (*Simple Types and λ^{\rightarrow} -terms*) *Types are formulated in the following way.*

- *Atomic types are types.*
- *If U and V are types then $U \rightarrow V$ is a type.*

λ^{\rightarrow} -terms are defined inductively as follows.

- *(variable) For each type U , there are infinitely many variables x^U, y^U, \dots of that type.*
- *(abstraction) If v is of type V and x does occur free in v then $(\lambda x^U.v)$ is of type $U \rightarrow V$.*
- *(application) If u is of type $U \rightarrow V$ and v is of type U , then $u(v)$ is of type V .*

We often omit the type superscript of a variable if this causes no confusion. On the other hand, superscripts may be used to indicate the types of λ^{\rightarrow} -terms.

5.1 $\nu(t)$ is not elementary for λ^{\rightarrow} -terms t

It is obvious that we can code λ^{\rightarrow} -terms with some elementary functions; $\nu(t)$ can then be regarded as a functions defined on the codings of λ^{\rightarrow} -terms; we omit the detailed treatment.

Definition 20 *Given an atomic type A ; let $U_0 = A$ and $U_{n+1} = U_n \rightarrow U_n$ for $n \geq 0$; let $\bar{i}_n = \lambda f^{U_{n+1}} \lambda x^{U_n}. f(\dots(f(x))\dots)$, where $i, n \geq 0$ and f occurs i times; let $s_n = \bar{2}_n(\bar{2}_{n-1}) \cdots (\bar{2}_0)$ for $n \geq 0$; let function $\text{tower}(n, m)$ be defined as follows.*

$$\text{tower}(n, m) = \begin{cases} m & \text{if } n = 0; \\ 2^{\text{tower}(n-1, m)} & \text{if } n > 0. \end{cases}$$

The following properties follow immediate.

Proposition 21 *For every $n \geq 0$,*

- $|s_n| = 5(n + 1)$;
- $\bar{i}_{n+1} \bar{j}_n \twoheadrightarrow_{\beta} \bar{k}_n$ for $k = j^i$;
- $s_n \twoheadrightarrow_{\beta} \bar{k}_0$ for $k = \text{tower}(n + 1, 1)$.

Theorem 22 *$\nu(t)$ is not elementary.*

Proof Let σ be a shortest β -reduction sequence which reduces s_n to its β -normal form, which is \bar{k}_0 for $k = \text{tower}(n + 1, 1)$ by the Church-Rosser theorem. Hence, it follows from Proposition 16 that

$$\text{tower}(n + 1, 1) < |\bar{k}_0| < |s_n|^{2^{\nu(s_n)}} = (5n + 1)^{2^{\nu(s_n)}}$$

for $n \geq 0$. Since $\text{tower}(n + 1, 1)$ is not elementary, $\nu(\cdot)$ cannot be elementary, either. ■

5.2 A bound on $\mu(t)$ for $\lambda^{\rightarrow}I$ -terms t

Since the leftmost β -reduction sequence from any λI -term t is a longest one among all β -reduction sequences from t , it goes straightforward to establish a bound on $\mu(t)$ for $\lambda^{\rightarrow}I$ -terms t if we can find any normalization sequences for them. In order to get a tighter bound, the key is to find shorter normalization sequences. We start with a weak normalization proof due to Turing according to [Gan80], which can also be found in many other literatures such as [And71] and [GLT89].

Definition 23 *The rank $\rho(T)$ of a simple type T is defined as follows.*

$$\rho(T) = \begin{cases} 0 & \text{if } T \text{ is atomic;} \\ 1 + \max\{\rho(T_0), \rho(T_1)\} & \text{if } T = T_0 \rightarrow T_1. \end{cases}$$

The rank $\rho(r)$ of a β -redex $r = (\lambda x^U. v^V)u^U$ is $\rho(U \rightarrow V)$, and the rank of a term t is

$$\hat{\rho}(t) = \begin{cases} \langle 0, 0 \rangle & \text{if } t \text{ is in } \beta\text{-normal form;} \text{ or} \\ \langle k = \max\{\rho(r) : r \text{ is a } \beta\text{-redex in } t\}, \text{ the number of } \beta\text{-redexes } r \text{ in } t \text{ with } \rho(r) = k \rangle. \end{cases}$$

The ranks of terms are lexically ordered.

Notice that a β -redex has a redex rank, which is a number, and also has a term rank, which is a pair of numbers.

Observations Now let us observe the following.

- If $t \rightarrow_{\beta} t'$ and β -redex r' in t' is a residual of some β -redex r in t , then $\rho(r') = \rho(r)$.
- Given $t = t[r]$ with $\hat{\rho}(t) = \langle k, n \rangle$, where $r = (\lambda x^V.u^U)v^V$ is a β -redex with $\rho(r) = k$ and no β -redexes in r have rank k . Then $\hat{\rho}(t') < \hat{\rho}(t)$ for $t = C[r] \rightarrow_{\beta} t' = C[u\{x := v\}]$. This can be verified by counting the number of β -redexes in t' with rank k . It is easy to see that any β -redex in t' which is not a residual must have rank $\rho(U)$ or $\rho(V)$, which is less than k . Hence, a β -redex in r' with rank k must be a residual of some β -redex r_1 in t with rank k . Note r_1 has only one residual in t' since r_1 is not in r . This yields $\hat{\rho}(t') < \hat{\rho}(t)$ since $\hat{\rho}(t')$ is either $\langle k, n-1 \rangle$ or $\langle k', n' \rangle$ for some $k' < k$.

Lemma 24 *Given t with $\hat{\rho}(t) = \langle k, n \rangle$ for some $k > 0$; then we can construct a development $\sigma : t \twoheadrightarrow_{\beta} u$ such that $|\sigma| = n$ and $\hat{\rho}(u) = \langle k', n' \rangle$ for some $k' < k$.*

Proof Following the observations, we can always reduce innermost β -redexes with rank k until there exist no β -redexes with rank k . This takes n steps and reaches a term with a less rank. ■

Definition 25 *We define*

$$m(\sigma) = \begin{cases} 1 & \text{if } |\sigma| = 0; \\ (1 + \max\{m(r_1), 1\}) \cdots (1 + \max\{m(r_n), 1\}) & \text{if } \sigma = [r_1] + \cdots + [r_n]. \end{cases}$$

Clearly, $m(\sigma_1 + \sigma_2) = m(\sigma_1)m(\sigma_2)$.

Theorem 26 *If t is a λ^{\rightarrow} -term with $\hat{\rho}(t) = \langle k, n \rangle$ for some $k > 0$, then there exists $\sigma : t \twoheadrightarrow_{\beta} u$ such that u is in β -normal form and $m(\sigma) < \text{tower}(1, \sum_{i=1}^k (\text{tower}(i-1, |t|))^2)$.*

Proof By Lemma 24, there exists a development $\sigma' : t \twoheadrightarrow_{\beta} t'$ with $|\sigma'| = n$ and $\hat{\rho}(t') = \langle k', n' \rangle$ for some $k' < k$. Let $\sigma' = [r_1] + \cdots + [r_n]$, then $1 + m(r_n) < 2^{|t|}$ by Lemma 18. Hence, $m(\sigma') < 2^{n|t|} < 2^{t^2}$ since $n < |t|$ clearly holds. Now let us proceed by induction on k .

- $k = 1$. Since t' is in β -normal form, let $\sigma = \sigma'$ and we are done.
- $k > 1$. By induction hypothesis, there exists $\sigma'' : t' \twoheadrightarrow_{\beta} u$ such that u is in β -normal form and $m(\sigma'') < \text{tower}(1, \sum_{i=1}^{k-1} (\text{tower}(i-1, |t'|))^2)$. Let $\sigma = \sigma' + \sigma''$, then

$$\begin{aligned} m(\sigma) &= m(\sigma')m(\sigma'') \\ &< \text{tower}(1, |t|^2) \text{tower}(1, \sum_{i=1}^{k-1} (\text{tower}(i-1, |t'|))^2) \\ &< \text{tower}(1, \sum_{i=1}^k (\text{tower}(i-1, |t|))^2) \end{aligned}$$

since $|t'| < 2^{|t|}$ by Lemma 18. ■

It is a well-known fact that the leftmost β -reduction sequence from a λI -term t is a longest one if t has a β -normal form.

Corollary 27 *Given any simply typed $\lambda^{\rightarrow}I$ -term t with $\hat{\rho}(t) = \langle k, n \rangle$; every β -reduction sequence from t is of length less than $\text{tower}(k+1, |t|)$.*

Proof It can be verified that the result holds if $|t| \leq 3$. For $|t| > 3$, we have

$$\text{tower}(1, \sum_{i=1}^k (\text{tower}(i-1, |t|))^2) \leq \text{tower}(k+1, |t|).$$

By Theorem 26, there exists $t \xrightarrow{\sigma} u$ such that $m(\sigma) < \text{tower}(k+1, |t|)$ and u is in β -normal form. This yields that $\text{std}(\sigma) < \text{tower}(k+1, |t|)$ by Theorem 14. Since t is a $\lambda^{\rightarrow}I$ -term, the leftmost β -reduction sequence from t is a longest one. This concludes the proof. ■

Notice that the leftmost β -reduction sequence from t may not yield a longest one if t is not a $\lambda^{\rightarrow}I$ -term. Therefore, the proof of Corollary 27 cannot go through directly for all λ^{\rightarrow} -terms.

5.3 An upper bound on $\mu(t)$ for λ^\rightarrow -terms t

Our following method is to transform a λ^\rightarrow -term t into a $\lambda^\rightarrow I$ -term $\llbracket t \rrbracket$ such that $\mu(t) \leq \mu(\llbracket t \rrbracket)$. Since we have already established a bound for $\mu(\llbracket t \rrbracket)$, this bound certainly works for $\mu(t)$.

Lemma 28 *Given $t = r(u_1) \dots (u_n)$ and $t_0 = u\{x := v\}(u_1) \dots (u_n)$, where $r = (\lambda x.u)(v)$; if t_0 and v are strongly normalizable, then t is strongly normalizable and $\mu(t) \leq 1 + \mu(t_0) + \mu(v)$.*

Proof Let $\sigma : t \twoheadrightarrow_\beta t^*$ be a β -reduction sequence, and we verify that $|\sigma| \leq 1 + \mu(t_0) + \mu(v)$. Clearly, we can assume that β -redex r is involved in σ . Then $\sigma = \sigma_1 + [r'] + \sigma_2$ is of the following form.

$$t \xrightarrow{\sigma_1}_\beta (\lambda x.u')(v')(u'_1) \dots (u'_n) \xrightarrow{r'}_\beta u'\{x := v'\}(u'_1) \dots (u'_n) \xrightarrow{\sigma_2}_\beta t^*,$$

where $\sigma_1 = \sigma_u + \sigma_v + \sigma_{u_1} + \dots + \sigma_{u_n}$ for $u \xrightarrow{\sigma_u}_\beta u'$, $v \xrightarrow{\sigma_v}_\beta v'$, $u_1 \xrightarrow{\sigma_{u_1}}_\beta u'_1, \dots$, and $u_n \xrightarrow{\sigma_{u_n}}_\beta u'_n$. Let $\tau_1 : u\{x := v\} \twoheadrightarrow_\beta u\{x := v'\}$ be the β -reduction sequence which reduces each occurrence of v in $u\{x := v\}$ to v' by following σ_v , and let $\tau_2 : u\{x := v'\} \twoheadrightarrow_\beta u'\{x := v'\}$ be the β -reduction sequence which reduces $u\{x := v'\}$ to $u'\{x := v'\}$ by following σ_u . Clearly, $|\tau_1| = m(r)|\sigma_v|$ and $|\tau_2| = |\sigma_u|$. Also let $\tau : u\{x := v\}(u_1) \dots (u_n) \twoheadrightarrow_\beta t^*$ be $\tau_1 + \tau_2 + \sigma_{u_1} + \dots + \sigma_{u_n} + \sigma_2$, then $|\tau| \leq \mu(t_0)$ by definition. Note

$$|\sigma| = |\sigma_1 + [r'] + \sigma_2| = |\sigma_u| + |\sigma_v| + |\sigma_{u_1}| + \dots + |\sigma_{u_n}| + 1 + |\sigma_2| \leq 1 + |\tau| + |\sigma_v|.$$

By definition, $|\sigma_v| \leq \mu(v)$. Hence, $|\sigma| \leq 1 + \mu(t_0) + \mu(v)$. ■

Definition 29 (Transformation) *To facilitate the presentation, we assume that there exist constants \langle, \rangle of type $U \rightarrow (V \rightarrow U)$ for all types U and V . Let $\langle u, v \rangle$ denote $\langle, \rangle(u)(v)$.*

$$\llbracket t \rrbracket = \begin{cases} t & t \text{ is a variable;} \\ \lambda x \lambda y_1^{U_1} \dots \lambda y_m^{U_m} . \langle \llbracket t_0 \rrbracket (y_1) \dots (y_m), x \rangle & t = (\lambda x.t_0), \text{ where } t_0 \text{ has type} \\ & U_1 \rightarrow \dots \rightarrow U_m \rightarrow V \text{ and } V \text{ is atomic.} \\ \llbracket t_0 \rrbracket (\llbracket t_1 \rrbracket) & t = t_0(t_1). \end{cases}$$

We will see clearly that \langle, \rangle can always be replaced by a free variable of the same type without altering our following argument..

Proposition 30 *For every λ^\rightarrow -term t of type T , we have the following.*

1. $\llbracket t \rrbracket$ is a $\lambda^\rightarrow I$ -term of type T ;
2. $\llbracket t\{x^U := u\} \rrbracket = \llbracket t \rrbracket \{x^U := \llbracket u \rrbracket\}$ for any λ^\rightarrow -term u of type U ;
3. $\mu(t) \leq \mu(\llbracket t \rrbracket)$.

Proof (1) and (2) can be readily proven by structural induction on t . By (1) and Corollary 27, we know $\mu(\llbracket t \rrbracket)$ exists for every λ^\rightarrow -term t . We now proceed to show (3) by induction on $\mu(\llbracket t \rrbracket)$ and the structure of $\llbracket t \rrbracket$, lexicographically ordered.

- $t = \lambda x.u$. By induction, $\mu(t) = \mu(u) \leq \mu(\llbracket u \rrbracket) \leq \mu(\llbracket t \rrbracket)$.
- $t = x(u_1) \dots (u_n)$, where x is some variable. Note $\mu(\llbracket t \rrbracket) = x(\llbracket u_1 \rrbracket) \dots (\llbracket u_n \rrbracket)$. By induction hypothesis, $\mu(t) = \mu(u_1) + \dots + \mu(u_n) \leq \mu(\llbracket u_1 \rrbracket) + \dots + \mu(\llbracket u_n \rrbracket) = \mu(\llbracket t \rrbracket)$.

- $t = r(u_1) \dots (u_n)$, where $r = (\lambda x.u)(v)$. By definition, $\llbracket t \rrbracket = \llbracket r \rrbracket(\llbracket u_1 \rrbracket) \dots (\llbracket u_n \rrbracket)$, and $\llbracket r \rrbracket = (\lambda x \lambda y_1 \dots \lambda y_m. \langle \llbracket u \rrbracket(y_1) \dots (y_m), x \rangle)(\llbracket v \rrbracket)$. Hence,

$$\llbracket r \rrbracket \longrightarrow_{\beta} \lambda y_1 \dots \lambda y_m. \langle \llbracket u \rrbracket\{x := \llbracket v \rrbracket\}(y_1) \dots (y_m), \llbracket v \rrbracket \rangle.$$

Since $\llbracket u \rrbracket(y_1) \dots (y_m)$ is of atomic type, $m \geq n$. This yields

$$\llbracket t \rrbracket \longrightarrow_{\beta} \lambda y_{m-n+1} \dots \lambda y_m. \langle \llbracket u \rrbracket\{x := \llbracket v \rrbracket\}(\llbracket u_1 \rrbracket) \dots (\llbracket u_n \rrbracket)(y_{m-n+1}) \dots (y_m), \llbracket v \rrbracket \rangle.$$

By (2), $\llbracket u \rrbracket\{x := \llbracket v \rrbracket\} = \llbracket u \rrbracket\{x := v\}$, and thus,

$$\llbracket u \rrbracket\{x := \llbracket v \rrbracket\}(\llbracket u_1 \rrbracket) \dots (\llbracket u_n \rrbracket) = \llbracket u \rrbracket\{x := v\}(\llbracket u_1 \rrbracket) \dots (\llbracket u_n \rrbracket) = \llbracket u \rrbracket\{x := v\}(u_1) \dots (u_n).$$

By induction hypothesis, $\mu(u\{x := v\}(u_1) \dots (u_n)) \leq \mu(\llbracket u \rrbracket\{x := v\}(u_1) \dots (u_n))$. Therefore, by Lemma 28,

$$\begin{aligned} \mu(t) &\leq 1 + \mu(u\{x := v\}(u_1) \dots (u_n)) + \mu(v) \\ &\leq 1 + \mu(\llbracket u \rrbracket\{x := v\}(u_1) \dots (u_n)) + \mu(\llbracket v \rrbracket) \leq \mu(\llbracket t \rrbracket). \end{aligned}$$

■

Corollary 31 *Given any simply typed λ^{\rightarrow} -term t with $\hat{\rho}(t) = \langle k, n \rangle$; every reduction sequence from t is of length less than $\text{tower}(k+1, (2k+3)|t|)$.*

Proof Given a subterm $\lambda x.u$ of type $U = U_1 \rightarrow \dots \rightarrow U_m \rightarrow V$ in t , where V is atomic, we can simply transform $\lambda x.u$ into $\lambda x.\llbracket u \rrbracket$ if $k < \rho(U)$ since no β -redexes with rank greater than k can occur in any β -reduction sequence of t ; if $\rho(U) \leq k$, we have

$$\|\llbracket \lambda x.u \rrbracket\| = |\lambda x \lambda y_1 \dots \lambda y_m. \langle \llbracket u \rrbracket(y_1) \dots (y_m), x \rangle| = \|\llbracket u \rrbracket\| + 2m + 3 \leq \|\llbracket u \rrbracket\| + 2k + 3.$$

Thus, it can be readily shown that $\|\llbracket t \rrbracket\| \leq (2k+3)|t|$. Also it can be immediately verified by the definition that if $\rho(t) = \langle k, n \rangle$ for some k and n then $\rho(\llbracket t \rrbracket) = \langle k, n \rangle$. By Corollary 27, we have

$$\mu(\llbracket t \rrbracket) < \text{tower}(k+1, \|\llbracket t \rrbracket\|) \leq \text{tower}(k+1, (2k+3)|t|).$$

This yields $\mu(t) < \text{tower}(k+1, (2k+3)|t|)$ by Proposition 30 (3). ■

6 Related Work and Conclusion

For those who know the strong equivalence relation \cong on β -reductions in [Bar84], originally due to Berry and Lévy, it can be verified that $\sigma \cong \text{std}(\sigma)$ for all β -reduction sequences σ .

There is a short proof of the standardization theorem due to Mitschke [Mit79], which analyses the relation between head and internal β -reductions. It shows any β -reduction sequence can be transformed into one which starts with head β -reductions followed by internal β -reductions. In this formulation, it is not easy to extract a bound directly from the proof. Our proof is a variant of Mitschke's proof. Lemma 11 simplifies the process which commutes head β -reduction with internal β -reductions, illuminating on why this process halts eventually. In this respect, our proof resembles a proof in [Tak95], where Takahashi exploited the notion of parallel β -reduction to show the termination of the commutation process.

There are also two proofs due to Klop [Klo80], to which the present proof bears some connection. Though all these proofs aim at commuting the contracted leftmost β -redexes to the front, our proof uses a different strategy to show the termination of such commutations. While Klop focuses on the strong equivalence relation \cong , we establish Lemma 11 by a structural induction

without using the finiteness developments theorem. This naturally yields an upper bound for standardizations.

In our application, an upper bound is given for the lengths of β -reduction sequences in λ^{\rightarrow} . This is a desirable result since $\mu(t)$, the length of a longest β -reduction sequence from t , can often be used as an induction order in many proofs. Gandy mentions a similar bound in [Gan80a] but details were left out. His semantic method, which aims at giving strong normalization proofs, is quite different from ours. Schwichtenberg presents a similar bound in [Sch91] using an approach adopted from [How80]. His method of transforming λ^{\rightarrow} -terms into $\lambda^{\rightarrow}I$ -terms closely relates to our presented method but is very much involved. It seems – in the author’s opinion – that such involvedness is not only unnecessary but also obscures the merits in Schwichtenberg’s proof. In addition, the proof of *finiteness of developments* theorem by de Vrijer [dV85] yields an exact bound for the lengths of developments, and thus, is casually related to our proof of the standardization theorem with bound.

In Gentzen’s sequent calculus, there exists a similar bound for the sizes of cut-free proofs obtained from cut elimination. Mints [Min79] (of which I have only learned the abstract) showed a way of computing the maximum length of a β -reduction from the length of a standard β -reduction sequence. In this respect, our work can be combined with his to show the maximum length of a β -reduction sequence from the length of an *arbitrary* one. This also motivates our planning on establishing a similar bound for the first-order λ -calculus with dependent types. On the other hand, Theorem 22 suggests that a lower bound for $\mu(t)$ have a similar superexponential form, and this makes it a challenging task to sharpen our presented bound for $\mu(t)$, although it seems to be somewhat exaggerated. Also Statman proved that λ^{\rightarrow} is not elementary [Sta79].

7 Acknowledgement

I gratefully acknowledge Richard Statman’s efforts on reading the paper and providing his comments. I also thank Frank Pfenning and Peter Andrews for their support and for providing me a nice work environment.

References

- [And71] P.B. Andrews (1971), Resolution in type theory, *J. Symbolic Logic* 36, pp. 414-432.
- [Bar76] H.P. Barendregt et al. (1976), Some notes on lambda reduction, *Preprint No. 22, University of Utrecht, Department of mathematics*, pp. 13-53.
- [Bar84] H.P. Barendregt (1984), The Lambda Calculus: Its Syntax And Semantics, *North-Holland publishing company, Amsterdam*.
- [Ber78] G. Berry (1978), Séquentialité de l’évaluation formelle des λ -expressions, *Proc. 3-e Colloque International sur la Programmation, Paris*.
- [Chu41] A. Church, (1941), The calculi of lambda conversion, *Princeton University Press, Princeton*.
- [CF58] H.B. Curry and R. Feys (1958), Combinatory Logic, *North-Holland Publishing Company, Amsterdam*.
- [dV85] R. de Vrijer (1985), A direct proof of the finite developments theorem, *Journal of Symbolic Logic*, 50:339-343.
- [Gan80] R.O. Gandy (1980), An early proof of normalization by A.M. Turing, *To: H.B. Curry: Essays on combinatory logic, lambda calculus and formalism*, edited by J.P. Seldin and J.R. Hindley, Academic press, pp. 453-456.
- [Gan80a] R.O. Gandy (1980), Proofs of Strong Normalization, *To: H.B. Curry: Essays on Combinatory logic, lambda calculus and formalism*, edited by J.P. Seldin and J.R. Hindley, Academic press, pp. 457-478.

- [GLM92] G. Gonthier, J.J. Lévy and P.-A. Mellès (1992), An abstract standardization theorem, in *Proceedings of Logic in Computer Science*, pp. 72–81.
- [GLT89] J.-Y. Girard et al. (1989), Proofs and types, *Cambridge Press*, 176 pp.
- [Hue94] Gérard Huet (1994), Residual Theory in λ -Calculus: A Formal Development, *Journal of Functional Programming Vol. 4*, pp. 371–394.
- [Hin78] J.R. Hindley (1978), Reductions of residuals are finite, *Trans. Amer. Math. Soc.* 240, pp. 345–361.
- [How80] W. Howard (1980), Ordinal analysis of terms of finite type, *Journal of Symbolic Logic*, 45(3):493–504.
- [Hyl73] J.M.E. Hyland (1973), A simple proof of the Church-Rosser theorem, *Typescript, Oxford University*, 7 pp.
- [Kah95] Stefan Kahrs (1995), Towards a Domain Theory for Termination Proofs, *Laboratory for Foundation of Computer Science*, 95-314, Department of Computer Science, The University of Edinburgh.
- [Klo80] J.W. Klop (1980), Combinatory reduction systems, *Ph.D. thesis, CWI, Amsterdam, Mathematical center tracts, No. 127*.
- [Lév78] J.-J. Lévy (1978), Réductions correctes et optimales dans le lambda calcul, *Thèse de doctorat d'état, Université Paris VII*.
- [Min79] G.E. Mints (1979), A primitive recursive bound of strong normalization for predicate calculus (in Russian with English abstract), *Zapiski Nauchnyh Seminarov Leningradskogo Otdelenija Matematicheskogo Instituta im V.A. Steklova Akademii Nauk SSSR (LOMI) 88*, pp. 131–135.
- [Mit79] G. Mitschke (1979), The standardization theorem for the λ -calculus, *Z. Math. Logik Grundlag. Math.* 25, pp. 29–31.
- [Pol94] J. van de Pol (1994), Strict functionals for termination proofs, *Lecture Notes in Computer Science 902*, edited by J. Heering, pp. 350–364.
- [Sta79] Richard Statman (1979), The typed λ -calculus is not elementary, *Theoretical Computer Science* 9, pp. 73–81.
- [Sch82] H. Schwichtenberg (1982), Complexity of normalization in the pure typed lambda-calculus, *The L.E.J. Brouwer Centenary Symposium*, edited by A.S. Troelstra and D. van Dalen, North-Holland publishing company, pp. 453–457.
- [Sch91] H. Schwichtenberg (1991), An upper bound for reduction sequences in the typed lambda-calculus, *Archive for Mathematical Logic*, 30:405–408.
- [Tak95] Masako Takahashi (1995), Parallel Reductions in λ -Calculus, *Information and Computation* 118, pp. 120–127.
- [Wad76] C.P. Wadsworth (1976), The relation between computational and denotational properties for Scott's D_∞ -models of λ -calculus, *SIAM Journal of Computing*, 5(3):488–521.
- [Xi96a] H. Xi (1996), An induction measure on λ -terms and its applications, *Technical report 96-192, Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh*.
- [Xi96b] H. Xi (1996), Separating developments in λ -calculus, *Manuscripts*.