Minimal Relative Normalization in Orthogonal Expression Reduction Systems

John Glauert and Zurab Khasidashvili

School of Information Systems, UEA
Norwich NR4 7TJ England
jrwg@sys.uea.ac.uk, zurab@sys.uea.ac.uk *

Abstract. In previous papers, the authors studied normalization relative to desirable sets $\mathcal{S}$ of 'partial results', where it is shown that such sets must be *stable*. For example, the sets of normal forms, head-normal-forms, and weak head-normal-forms in the λ-calculus, are all stable. They showed that, for any stable $\mathcal{S}$, $\mathcal{S}$-*needed* reductions are $\mathcal{S}$-normalizing. This paper continues the investigation into the theory of relative normalization. In particular, we prove existence of *minimal* normalizing reductions for *regular* stable sets of results. All the above mentioned sets are regular. We give a sufficient and necessary criterion for a normalizing reduction (w.r.t. a regular stable $\mathcal{S}$) to be minimal. Finally, we establish a relationship between relative minimal and optimal reductions, revealing a conflict between minimality and optimality: for regular stable sets of results, a term need not possess a reduction that is minimal and optimal at the same time.

1 Introduction

The *Normalization Theorem* in the λ-calculus, due to Curry and Feys [CuFe58], states that contraction of leftmost-outermost redexes in a term t yields a normal form whenever t is normalizable, even if t has infinite reduction sequences.

Generalizing this fundamental theorem to a large class of Orthogonal Term Rewriting Systems (OTRSs), Huet and Lévy laid the foundations of a regular theory of 'normalization by neededness' in [HuLé91]. They proved that any term t not in normal form, in an OTRS, has a *needed* redex, and that contraction of needed redexes in a normalizable term results in a normal form. Here a redex u in t is needed if some residual of it is contracted in every normalizing reduction starting from t.

Barendregt et al. [BKKS87] applied the neededness notion to the λ-calculus, and studied neededness not only w.r.t. normal forms, but also w.r.t. head-normal forms. The authors proved correctness of the two needed strategies for computing corresponding normal forms. In [Mar92], Maranget also studied a strategy that computes a weak head-normal form of a term in an OTRS. Normalization w.r.t. another interesting set of 'normal forms', that of constructor head-normal forms in constructor OTRSs, is studied by Nöcker [Nök94].

In [GlKh94], the present authors studied normalization with respect to any desired set of final terms, and found the sufficient and necessary properties, called

* This work was supported by the Engineering and Physical Sciences Research Council of Great Britain under grant GR/H41300

stability, that a set $\mathcal{S}$ of terms must possess in order for the neededness theory of Huet and Lévy still to make sense. That is, they showed that, for any stable $\mathcal{S}$, each $\mathcal{S}$-normalizable term not yet in $\mathcal{S}$ (not in $\mathcal{S}$-normal form) has at least one $\mathcal{S}$-needed redex, and that repeated contraction of $\mathcal{S}$-needed redexes in a term t will lead to an $\mathcal{S}$-normal form of t whenever there is one. It is shown also that if a stable $\mathcal{S}$ is *regular*, i.e., if $\mathcal{S}$-unneeded redexes cannot duplicate $\mathcal{S}$-needed ones, then the $\mathcal{S}$-needed strategy is hypernormalizing as well. This work was performed in the context of Orthogonal *Expression Reduction Systems* (OERSs) [Kha92], a form of higher-order rewriting which subsumes TRSs and the λ-calculus and is similar to Klop's *Combinatory Reduction systems* (CRSs) [Klo80]. Most of these results were later generalized to an abstract framework of Deterministic Residual Structures [GlKh96].

Normalization theory has developed in other directions as well, of which we mention only a few. Boudol extended neededness theory to non-orthogonal TRSs [Bou85]. Khasidashvili defined a similar normalizing strategy, the *essential* strategy, in the λ-calculus, OTRSs and OERSs [Kha88, Kha93, Kha94]. Kennaway and Sleep [KeSl89] generalized the needed strategy to Klop's orthogonal CRSs [Klo80]. Sekar and Ramakrishnan [SeRa93] study a normalizing strategy which in each multi-step contracts a *necessary* set of redexes – a set at least one member of which is contracted in every normalizing reduction. A different approach to normalization in not-necessarily orthogonal rewrite systems is developed in Kennaway [Ken89] and Antoy&Middeldorp [AnMi94]. Antoy et al. [AEH94] designed a needed narrowing strategy. Gardner [Gar94] described a *complete* way of encoding neededness information using a type assignment system. Kennaway et al. [KKSV95] studied a needed strategy for infinitary OTRSs.

The contribution of this paper is to develop a theory of *minimal* reduction in the framework of relative normalization, and to establish a relationship between minimal and *optimal* [Lév78] reductions. While normal forms are unique in an OERS, a term may have many $\mathcal{S}$-normal forms. A reduction $P : t \twoheadrightarrow s$ with $t \notin \mathcal{S}$ and $s \in \mathcal{S}$ is said to be $\mathcal{S}$*-minimal* if it does no more work than any other $\mathcal{S}$-normalizing reduction $Q : t \twoheadrightarrow e$, i.e., the *residual* [Lév78] P/Q of P under Q is empty. The final term in the $\mathcal{S}$-minimal reduction is said to be an $\mathcal{S}$*-minimal* $\mathcal{S}$-normal form.

$\mathcal{S}$-minimal $\mathcal{S}$-normal forms are useful to compute since any other $\mathcal{S}$-normal form is accessible from the $\mathcal{S}$-minimal one. Further, strategies computing partial results, such as head-normal-forms (hnfs) and weak hnfs, in the λ-calculus, usually compute minimal reductions, and it is natural to ask whether optimality can be achieved while retaining minimality. The prime example is the leftmost outermost strategy computing the so called 'principal' hnf and whnf of a λ-term, and used in constructions of Böhm [Bar84] and Lévy-Longo [Lév76, Lon83] (also called *lazy* [AbOn93]) trees, respectively. These trees represent the values of the term according to different semantics – Böhm semantics and Lévy- or lazy semantics, respectively. Clearly this property of minimality is not useful for full normal forms, but full normal forms are rarely used in the practice of functional programming.

Our research on minimal $\mathcal{S}$-normalizing reductions was inspired by a result of Maranget [Mar92], stating that *standard* reductions are minimal among reductions computing a 'stable prefix' of a given term. However, we will show that standard reductions are not always minimal in the relative case, and a different concept of standard reduction is required.

The earliest minimality result was obtained by Berry and Lévy in [BeLé79], where existence of minimal reductions was shown for any (finite or infinite) approximation of a possibly infinite value of a term, for Recursive Program Schemes. Minimal reductions were used to design optimal reductions, both finite and infinite, and minimality and optimality of *outermost complete family-reductions* were shown.

In this paper, we restrict ourselves to finite reductions only. We show that, for any stable and regular $\mathcal{S}$, any $\mathcal{S}$-normalizable term not yet in $\mathcal{S}$ possesses an $\mathcal{S}$-needed $\mathcal{S}$-*unabsorbed* redex, and repeated contraction of such redexes gives $\mathcal{S}$-minimal $\mathcal{S}$-normalizing reductions. We further give a sufficient and necessary criterion for an $\mathcal{S}$-normalizing reduction to be $\mathcal{S}$-minimal. We show also that $\mathcal{S}$-minimal reductions need not exist if $\mathcal{S}$ is stable but is irregular.

It has been shown in [GlKh96] that complete $\mathcal{S}$-needed family-reductions, which contract all members of a *redex-family* containing an $\mathcal{S}$-needed redex in a multi-step, are *optimal* in the sense that they reach $\mathcal{S}$ in the least number of family-reduction steps. $\mathcal{S}$-needed complete family reductions, though optimal, need not be $\mathcal{S}$-minimal, because they may contract $\mathcal{S}$-unneeded redexes that are $\mathcal{S}$-essential. It is tempting to think that contracting only the $\mathcal{S}$-needed redexes of $\mathcal{S}$-needed families would yield $\mathcal{S}$-optimal reductions that are $\mathcal{S}$-minimal at the same time. We show however that this is not the case either in the λ-calculus or in OTRSs. As a consequence, a term need not have a reduction that is both minimal and optimal at the same time.

The paper is organized as follows. In section 2, we introduce higher order rewriting through *Expression Reduction Systems*. In section 3, we review the theory of relative normalization. In section 4, we study $\mathcal{S}$-minimal reductions for regular stable sets $\mathcal{S}$, and in section 5, we relate relative optimal and minimal reductions. Conclusions appear in section 6. More details can be found in [GlKh94a].

2 Orthogonal Expression Reduction Systems

Klop introduced *Combinatory Reduction Systems* (CRSs) in [Klo80] to provide a uniform framework for reductions with substitutions (also referred to as higher order rewriting) as in the λ-calculus [Bar84]. Several interesting formalisms have been introduced later [Kha92, Nip93, OR94]. We refer to van Raamsdonk [Raa96] for a survey. Here we use a system of higher order rewriting, *Expression Reduction Systems* (ERSs), defined in [Kha92] (under the name of CRSs); the present formulation follows [GlKh94] and is simpler.

Definition 2.1 Let Σ be an *alphabet*, comprising *variables*, denoted by $x, y, z, \ldots$; *function symbols*, also called *simple operators*; and *operator signs* or *quantifier signs*. Each function symbol has an *arity* $k \in N$, and each operator sign σ has an *arity* (m, n) with $m, n \neq 0$ such that, for any sequence $x_1, \ldots, x_m$ of pairwise distinct variables, $\sigma x_1 \ldots x_m$ is a *compound operator* or a *quantifier* with *arity* n. Occurrences of $x_1, \ldots, x_m$ in $\sigma x_1 \ldots x_m$ are called *binding variables*. Each quantifier sign σ, as well as any corresponding quantifier $\sigma x_1 \ldots x_m$ and binding variables $x_1 \ldots x_m$, has a *scope indicator* $(k_1, \ldots, k_l)$ to specify the arguments in which $\sigma x_1 \ldots x_m$ binds all free occurrences of $x_1, \ldots, x_m$. *Terms* are constructed from variables using functions and quantifiers in the usual way.

Metaterms are constructed similarly from *terms* and *metavariables* $A, B, \ldots$, which range over terms. In addition, *metasubstitutions*, expressions of the form $(t_1/x_1, \ldots, t_n/x_n)t_0$, with t_j as arbitrary metaterms, are allowed, where the *scope* of each x_i is t_0. Metaterms without metasubstitutions are *simple metaterms.* An *assignment* maps each metavariable to a term over Σ. If t is a metaterm and θ is an assignment, then the θ*-instance* $t\theta$ of t is the term obtained from t by replacing metavariables with their values under θ, and by replacing metasubstitutions $(t_1/x_1, \ldots, t_n/x_n)t_0$, in the left to right order, with the result of substitution of terms $t_1, \ldots, t_n$ for free occurrences of $x_1, \ldots, x_n$ in t_0.

For example, a β-redex in the λ-calculus appears as $Ap(\lambda x\, t, s)$ in our notation, where Ap is a function symbol of arity 2, and λ is an operator sign of arity (1,1) and scope indicator (1). Integrals such as $\int_s^t f(x)\, dx$ can be represented as $\int x\, s\, t\, f(x)$ using an operator sign $\int$ of arity (1,3) and scope indicator (3).

Definition 2.2 An *Expression Reduction System* (ERS) is a pair (Σ, R), where Σ is an *alphabet*, described in Definition 2.1, and R is a set of *rewrite rules* $r : t \to s$, where t and s are closed metaterms (i.e., no free variables) such that t is a simple metaterm and is not a metavariable, and each metavariable that occurs in s occurs also in t.

Further, each rule r has a set of *admissible assignments* $AA(r)$ which, in order to prevent undesirable confusion of variable bindings, must satisfy the condition that:

(a) for any assignment $\theta \in AA(r)$, any metavariable A occurring in t or s, and any variable $x \in FV(A\theta)$, either every occurrence of A in r is in the scope of some binding occurrence of x in r, or no occurrence is.

For any $\theta \in AA(r)$, $t\theta$ is an *r-redex* or an *R-redex*, and $s\theta$ is the *contractum* of $t\theta$. We call R *simple* if right-hand sides of R-rules are simple metaterms.

Our syntax is similar to that of Klop's CRSs [Klo80], but is simpler and is closer to the syntax of the λ-calculus and of First Order Logic. For example, the β-rule is written as $\beta : Ap(\lambda x A, B) \to (B/x)A$, where A and B can be instantiated by any terms; the η-rule is written as $\lambda x(Ax) \to A$ which requires that an assignment θ is admissible iff $x \notin (A\theta)$, otherwise an x occurring in $A\theta$ and therefore bound in $\lambda x(A\theta x)$ would become free. A rule like $f(A) \to \exists x(A)$ is also allowed, but an assignment θ with $x \in A\theta$ is not. The recursor rule is written as $\mu(\lambda x A) \to (\mu(\lambda x A)/x)A$.

Below we restrict ourselves to the case of non-conditional ERSs, i.e., ERSs where an assignment is admissible iff the condition (a) of Definition 2.2 is satisfied. We ignore questions relating to renaming of bound variables. As usual, a rewrite step consists of replacement of a redex by its contractum. Subterms of a redex corresponding to metavariables are *arguments* of the redex, and the rest is its *pattern.* Note that the use of metavariables in rewrite rules of ERSs is not really necessary – free variables can be used instead, as in TRSs. We will indeed do so at least when giving TRS examples.

Notation 2.1 We use a, b, c, d for constants, t, s, e, o for terms, u, v, w for redexes, and N, P, Q for reductions. We write $s \subseteq t$ if s is a subterm of t. A one-step reduction contracting a redex $u \subseteq t$ is written as $t \stackrel{u}{\to} s$ or $t \to s$ or just u. We write $P : t \twoheadrightarrow s$ if P denotes a reduction of t to s. $P + Q$ denotes the concatenation of P and Q.

The definition of *orthogonality* in ERSs is similar to the case of CRSs: all the rules are left-linear and in no term redex-patterns can overlap [Klo80]. As in the case of the λ-calculus [Bar84], for any co-initial reductions P and Q, one can define in OERSs the notion of *residual of* P *under* Q, written P/Q, due to Lévy [Lév78]. We write $P \trianglelefteq Q$ if $P/Q = \emptyset$ ($\trianglelefteq$ is the *Lévy-embedding* relation); P and Q are called *Lévy-equivalent*, *strongly-equivalent*, or *permutation-equivalent* (written $P \approx_L Q$) if $P \trianglelefteq Q$ and $Q \trianglelefteq P$. It follows immediately from the definition of / that if P and Q are co-initial reductions in an OERS, then $(P + P')/Q \approx_L P/Q + P'/(Q/P)$ and $P/(Q + Q') \approx_L (P/Q)/Q'$.

The following *strong Church-Rosser (confluence)* property is proved for ERSs in [Kha92]; the same result for other higher-order rewriting formats are obtained, among others, in [Klo80, Nip93, KOR93, OR94, Oos94, KvO95, Raa96].

Theorem 2.1 (Strong Church-Rosser) For any co-initial reductions P and Q in an OERS, $P + Q/P \approx_L Q + P/Q$.

3 Relative Normalization

In this section, we review some notions and results concerning relative normalization from [GlKh94].

Definition 3.1 Let $\mathcal{S}$ be a set of terms in an OERS R. We call a redex $u \subseteq t$ $\mathcal{S}$-*needed*, written $NE_{\mathcal{S}}(u, t)$, if at least one residual of it is contracted in any reduction from t to a term in $\mathcal{S}$, and call it $\mathcal{S}$-unneeded, written $UN_{\mathcal{S}}(u, t)$, otherwise.

Definition 3.2 (1) We call a set $\mathcal{S}$ of terms *stable* iff (a) $\mathcal{S}$ is *closed under parallel moves*: for any $t \notin \mathcal{S}$, any $P : t \twoheadrightarrow o \in \mathcal{S}$, and any $Q : t \twoheadrightarrow e$ which does not contain terms in $\mathcal{S}$, the final term of P/Q is in $\mathcal{S}$; and (b) $\mathcal{S}$ is *closed under unneeded expansion*: for any $e \xrightarrow{u} o$ such that $e \notin \mathcal{S}$ and $o \in \mathcal{S}$, u is $\mathcal{S}$-needed. (2) We call a stable $\mathcal{S}$ *regular* iff $\mathcal{S}$-unneeded redexes cannot duplicate $\mathcal{S}$-needed ones.

Below $\mathcal{S}$, resp. $\mathcal{R}$, will denote a stable, resp. regular stable, set of terms in an OERS. $t{\downarrow_{\mathcal{S}}}$ will denote that t is $\mathcal{S}$-normalizable, i.e., reducible to a term in $\mathcal{S}$, and similarly for $t{\downarrow_{\mathcal{R}}}$.

Lemma 3.1 (1) Residuals of $\mathcal{S}$-unneeded redexes in a term $t \notin \mathcal{S}$ are $\mathcal{S}$-unneeded.
(2) Let $t \notin \mathcal{S}, t \xrightarrow{u} t', UN_{\mathcal{S}}(u, t)$, and let $u' \subseteq t'$ be a u-new redex. Then $UN_{\mathcal{S}}(u', t')$.
(3) Let $t{\downarrow_{\mathcal{S}}}$, $t \xrightarrow{u} s$, $NE_{\mathcal{S}}(v, t)$, and $v \neq u$. Then v has an $\mathcal{S}$-needed residual in s.

Theorem 3.1 (**Relative Normalization**) Let $\mathcal{S}$ be a stable set of terms in an OERS R. Then any $\mathcal{S}$-normalizable term t in R not in $\mathcal{S}$-normal form contains an $\mathcal{S}$-needed redex; and any $\mathcal{S}$-needed reduction starting from t eventually terminates at a term in $\mathcal{S}$. If $\mathcal{S}$ is moreover regular, then $\mathcal{S}$-needed reductions starting from t eventually reach $\mathcal{S}$ even if finite sequences of consecutive $\mathcal{S}$-unneeded steps are also allowed.

4 Minimal Relative Normalization

In this section, we define $\mathcal{S}$*-unabsorbed*, *persistently* $\mathcal{S}$*-needed*, and $\mathcal{S}$*-erased* redexes, and show that each class is a strict subset of the next when $\mathcal{S}$ is regular. Further, we define $\mathcal{S}$*-minimal* reductions as minimal w.r.t. Lévy-embedding $\trianglelefteq$ among co-initial $\mathcal{S}$-normalizing reductions, and show that, when $\mathcal{S}$ is regular, an $\mathcal{S}$-normalizing reduction is $\mathcal{S}$-minimal iff it is $\mathcal{S}$-erased, i.e., contracts only $\mathcal{S}$-erased redexes. But $\mathcal{S}$-erased reductions need not be $\mathcal{S}$-needed, hence need not be $\mathcal{S}$-normalizing, and again for regular $\mathcal{S}$, we show existence of $\mathcal{S}$-unabsorbed $\mathcal{S}$-normalizing reductions, which are $\mathcal{S}$-needed $\mathcal{S}$-minimal reductions. We show that $\mathcal{S}$-minimal reductions need not exist for irregular stable $\mathcal{S}$. Below we always consider reductions in OERSs.

Definition 4.1 (1) We call $u \subseteq t$ *persistently* $\mathcal{S}$*-needed* if all residuals of u are $\mathcal{S}$-needed. (2) We call $u \subseteq t$ $\mathcal{S}$*-erased* if u doesn't have a residual under any $\mathcal{S}$-normalizing reduction. We call a reduction $\mathcal{S}$*-erased* if it only contracts $\mathcal{S}$-erased redexes.

Note that $\mathcal{S}$-erased redexes need not be $\mathcal{S}$-needed (e.g., when $\mathcal{S}$ is the set of normal forms and the OERS has an erasing rule, say $f(x) \to a$). The following example illustrates the introduced concepts using a simple OTRS.

Example 4.1 Consider an OTRS $R = \{f(x) \to g(x, h(x)),\ h(x) \to c,\ a \to b\}$, consider a term (redex) $u = f(a)$, and the following sets of terms in R: the set $\mathcal{S}_1$ of normal forms; the set $\mathcal{S}_2$ of terms not containing a redex on the left-spine (i.e., not containing a redex with the top symbol on the left-spine, when the term is considered as a tree); the set $\mathcal{S}_3$ of terms not containing occurrences of a; and the set $\mathcal{S}_4$ of terms not containing a on the right-spine. Then, for the two redexes u and a in $u = f(a)$, we have the following:

1. u is $\mathcal{S}_1$-needed, persistently $\mathcal{S}_1$-needed, and $\mathcal{S}_1$-erased. $a \subseteq u$ is $\mathcal{S}_1$-needed but not persistently $\mathcal{S}_1$-needed (since the second residual of a in $g(x, h(a))$ is $\mathcal{S}_1$-unneeded); still, a is $\mathcal{S}_1$-erased.
2. u is $\mathcal{S}_2$-needed, persistently $\mathcal{S}_2$-needed, and $\mathcal{S}_2$-erased. $a \subseteq u$ is $\mathcal{S}_2$-needed but not persistently $\mathcal{S}_2$-needed; and a is not $\mathcal{S}_2$-erased – a has a residual along the $\mathcal{S}_2$-normalizing reduction $u \to g(a, h(a)) \to g(b, h(a))$.
3. u is neither (persistently) $\mathcal{S}_3$-needed nor $\mathcal{S}_3$-erased. $a \subseteq u$ is $\mathcal{S}_3$-needed but not persistently $\mathcal{S}_3$-needed (since the second residual of a in $g(a, h(a))$ is $\mathcal{S}_3$-unneeded); still, a is $\mathcal{S}_3$-erased.
4. both u and a are neither (persistently) $\mathcal{S}_4$-needed nor $\mathcal{S}_4$-erased.

Note that $\mathcal{S}_1$ and $\mathcal{S}_2$ are regular stable sets; $\mathcal{S}_3$ is stable but not regular, since $\mathcal{S}_3$-unneeded redex u duplicates the $\mathcal{S}_3$-needed redex a; and $\mathcal{S}_4$ is not stable (therefore, u does not contain an $\mathcal{S}_4$-needed redex).

Lemma 4.1 Every persistently $\mathcal{S}$-needed redex is $\mathcal{S}$-erased, but an $\mathcal{S}$-erased redex, even if $\mathcal{S}$-needed, need not be persistently $\mathcal{S}$-needed.

Proof. ($\Rightarrow$) Let $u \subseteq t$ be persistently $\mathcal{S}$-needed, and let $P : t \twoheadrightarrow s$ be $\mathcal{S}$-normalizing. If u/P was not empty, then every $u' \in u/P$ (the set of P-residuals of u) would be $\mathcal{S}$-needed, which is not possible since $s \in \mathcal{S}$. ($\Leftarrow$) From Example 4.1 (cases 1 and 3).

Definition 4.2 We call $P : t \twoheadrightarrow s$ $\mathcal{S}$-*minimal* if it is $\mathcal{S}$-normalizing and $P \trianglelefteq Q$ for any $\mathcal{S}$-normalizing $Q : t \twoheadrightarrow o$.[2] When P is $\mathcal{S}$-minimal, we call s an $\mathcal{S}$-*minimal* $\mathcal{S}$-normal form of t.

It follows immediately from Definition 4.2 that if $t{\downarrow_{\mathcal{S}}} \notin \mathcal{S}$ (i.e., $t{\downarrow_{\mathcal{S}}}$ and $t \notin \mathcal{S}$), then t has no more than one $\mathcal{S}$-minimal $\mathcal{S}$-normal form s. For any other $\mathcal{S}$-normal form e of t, it holds that $s \twoheadrightarrow e$. Note that the latter property of $\mathcal{S}$-minimal $\mathcal{S}$-normal forms cannot be taken as the definition, because in that case an $\mathcal{S}$-normalizable term could have many $\mathcal{S}$-minimal $\mathcal{S}$-normal forms, due for example to a cycle in $\mathcal{S}$, and some of them may require more reduction to be reached than others. For example, take $R = \{a \to b,\ b \to a,\ f(x) \to x\}$ and $\mathcal{S} = \{a, b\}$. Then $\mathcal{S}$ is stable and regular, $t = f(a)$ has two $\mathcal{S}$-normal forms from which any other one is accessible – a and b, but any reduction from t to b should contract the $\mathcal{S}$-unneeded redex a in t; therefore, no reduction from t to b can be considered as $\mathcal{S}$-minimal.

Lemma 4.2 Every $\mathcal{S}$-erased $\mathcal{S}$-normalizing reduction is $\mathcal{S}$-minimal.

Proof. Let $P : t_0 \xrightarrow{u_0} t_1 \to \ldots \to t_n$ be an $\mathcal{S}$-erased $\mathcal{S}$-normalizing reduction, let $P_i : t_0 \xrightarrow{u_0} \ldots \to t_i$, and let $Q : t_0 \twoheadrightarrow o \in \mathcal{S}$. By stability of $\mathcal{S}$, $Q_i = Q/P_i$ is $\mathcal{S}$-normalizing. Since u_i is $\mathcal{S}$-erased and Q_i is $\mathcal{S}$-normalizing, $u_i/Q_i = \emptyset$. Hence $P/Q = \emptyset$, i.e., P is $\mathcal{S}$-minimal.

Definition 4.3 Let F be a set of redexes in t. We call P an F-*reduction* if it contracts only residuals of redexes from F and created redexes; we call such redexes F-*redexes*. Below $F \subseteq t$ will denote that F is a set of redexes in a term t, and $F(t)$ will denote the set of all redexes of t.

Definition 4.4 (1) Let $F \subseteq t$. We call a redex $u \subseteq t$ F-*unabsorbed* (in t) if $u \in F$ and, for any F-reduction P, none of the residuals of u along P appear in arguments of F-redexes; we call u F-*absorbed in* if $u \in F$ and it is not F-unabsorbed.

(2) We call $u \subseteq t$ $\mathcal{S}$-*(un)absorbed* if it is $F_{\mathcal{S}}(t)$-(un)absorbed, where $F_{\mathcal{S}}(t)$ is the set of $\mathcal{S}$-needed redexes of t. (Thus any $\mathcal{S}$-unabsorbed redex is necessarily $\mathcal{S}$-needed.) We call a reduction P $\mathcal{S}$-*unabsorbed* if each redex contracted in it is.

Example 4.2 Consider an OTRS $R = \{a \to c,\ b \to b',\ f(c, x) \to c'\}$, and take a term $t = g(f(a, b), a)$. Then both occurrences of a in t are $F(t)$-unabsorbed in t, while b is $F(t)$-absorbed in t: we have $t \to g(f(c, b), a) = s$, and the residual of b in s is in an argument of the created redex $f(c, b)$. If $F \subseteq t$ contains two redexes – the first occurrence of a in t and the redex $b \subseteq t$, then only the first $a \subseteq t$ is F-unabsorbed in t. If the set of terms not having a left-spine redex is taken for $\mathcal{S}$, then the first a is the only $\mathcal{S}$-unabsorbed redex in t (it is the only $\mathcal{S}$-needed redex too).

It is shown in [HuLé91, Kha93, GlKh94] that any term t not in normal form contains an $F(t)$-unabsorbed redex (such redexes are called *external* in [HuLé91]). Now, if one ignores all redexes in t except those in $F \subseteq t$, it follows that, for any $F \neq \emptyset$, F contains an F-unabsorbed redex. And by taking $F_{\mathcal{S}}(t)$ for F ($F_{\mathcal{S}}(t) \neq \emptyset$ by Theorem 3.1), we obtain the following proposition:

[2] We prefer to use minimal rather than *least* or *smallest*.

Proposition 4.1 Every term $t \notin \mathcal{S}$ contains an ($\mathcal{S}$-needed) $\mathcal{S}$-unabsorbed redex.

Below, in the study of $\mathcal{S}$-minimal reductions, we will restrict ourselves to *regular* stable $\mathcal{S}$. The reason is that, as shown by the following example, an $\mathcal{S}$-normalizable term need not have an $\mathcal{S}$-minimal reduction when $\mathcal{S}$ is irregular.

Example 4.3 Consider $R = \{f(x) \to g(x, x),\ a \to b\}$, take for $\mathcal{S}$ the set of terms not containing a as the leftmost innermost node, and take $t = f(a)$. Obviously, $\mathcal{S}$ is closed under unneeded expansion, because the only $\mathcal{S}$-needed redex in a term $s \notin \mathcal{S}$ is the leftmost occurrence of a in it, and $\mathcal{S}$ is closed under reduction. $\mathcal{S}$ is not regular, because the outermost redex in t is $\mathcal{S}$-unneeded, while the innermost one is $\mathcal{S}$-needed. Further, there are three $\mathcal{S}$-normalizing reductions starting from t : $P : f(a) \to f(b)$; $Q : f(a) \to g(a, a) \to g(b, a)$, and $N : f(a) \to g(a, a) \to g(a, b) \to g(b, b)$. (There are two more reductions that continue Q and P, but we do not need to consider them because they cannot be $\mathcal{S}$-minimal.) We have $P \not\trianglelefteq Q$, $Q \not\trianglelefteq P$, and $N \not\trianglelefteq P$. Hence none of the reductions is $\mathcal{S}$-minimal.

Lemma 4.3 For any $P : t \twoheadrightarrow s$ with $t \notin \mathcal{R}$, there is an $\mathcal{R}$-needed Q, containing the same number of steps as that of $\mathcal{R}$-needed steps in P, and an $\mathcal{R}$-unneeded N, such that $P \approx_L Q + N$; and if P is $\mathcal{R}$-normalizing or contains infinitely many $\mathcal{R}$-needed steps, then $N = \emptyset$.

Proof. The lemma was proved in [Kha88, Kha93] for the case of essentiality in place of $\mathcal{R}$-neededness. The same proof applies in this case.

Lemma 4.4 If a redex $u \subseteq t$ is $\mathcal{R}$-unabsorbed, then it need not be unabsorbed in t, but it cannot be replicated and is persistently $\mathcal{R}$-needed.

Proof. Let $P : t \twoheadrightarrow o$, not necessarily an $F_{\mathfrak{R}}(t)$-reduction. By Lemma 3.1.(3), it is enough to show that if a residual u' of u can appear inside an $\mathcal{R}$-needed redex $w' \neq u'$, then w' cannot replicate u'; therefore u has at most one residual in any term of P. Suppose, on the contrary, that there is $P : t \twoheadrightarrow s$ such that a residual u' of u is inside an $\mathcal{R}$-needed redex w' such that w' replicates u'; and assume that P is a shortest such a reduction, i.e., u has exactly one residual in every term in P. By Lemma 4.3, there are $\mathcal{R}$-needed P' and $\mathcal{R}$-unneeded P'' such that $P \approx_L P' + P''$. Since u' and w' are $\mathcal{R}$-needed and P'' is $\mathcal{R}$-unneeded, it follows from Lemma 3.1.(2) that there are $\mathcal{R}$-needed u'' and w'' in the final term of P' such that u' and w' are the only residuals of u'' and w'', respectively. Since u is $\mathcal{R}$-unabsorbed, $u'' \notin w''$. Hence u'' has exactly one w''-residual, say u^*. By Theorem 2.1, $w'' + P''/w''$ replicates u'', since w' replicates u'. Thus P''/w'' replicates u^* – a contradiction, since P''/w'' is $\mathcal{R}$-unneeded by Lemma 3.1.(1), and $\mathcal{R}$ is regular.

Note that if $\mathcal{S}$ is irregular, then an $\mathcal{S}$-unabsorbed redex $u \subseteq t$ need not be persistently $\mathcal{S}$-needed or $\mathcal{S}$-erased. Indeed, take R, $\mathcal{S}$, and Q as in Example 4.3. Then a in t is $\mathcal{S}$-needed, so is its leftmost residual in $g(a, a)$, but the rightmost residual is $\mathcal{S}$-unneeded, and $a/Q = \emptyset$. Hence $a \subseteq t$ is not persistently $\mathcal{S}$-needed or $\mathcal{S}$-erased. But $a \subseteq t$ is $\mathcal{S}$-unabsorbed, since the only $F_{\mathcal{S}}(t)$-reduction is $N : f(a) \to f(b)$, and a is $F_{\mathcal{S}}(t)$-unabsorbed in N.

Proposition 4.2 An $\mathcal{R}$-normalizing reduction is $\mathcal{R}$-minimal iff it is $\mathcal{R}$-erased.

Proof. ($\Leftarrow$) From Lemma 4.2. ($\Rightarrow$) Let $P : t_0 \overset{u_0}{\to} t_1 \to \ldots \to t_n$ be $\mathcal{R}$-minimal, and let $Q : t_0 \twoheadrightarrow o$ be $\mathcal{R}$-unabsorbed, hence $\mathcal{R}$-erased by Lemmas 4.4 and 4.1, $\mathcal{R}$-normalizing reduction; Q exists by Proposition 4.1. Further, let $P_i : t_0 \overset{u_0}{\to} \ldots \to t_i$ and let $Q_i = Q/P_i$. Since Q is $\mathcal{R}$-erased, so is Q_i, and Q_i is $\mathcal{R}$-normalizing by the closure of $\mathcal{R}$ under parallel moves. Hence Q_i is $\mathcal{R}$-minimal by Lemma 4.2. Since P is $\mathcal{R}$-minimal too, $u_i/Q_i = \emptyset$ for every i. But for every $\mathcal{R}$-normalizing reduction $Q'_i : t_i \twoheadrightarrow o_i$, it holds that $Q_i \trianglelefteq Q'_i$ (since Q_i is $\mathcal{R}$-minimal). Hence $u_i/Q'_i = \emptyset$, i.e., u_i is $\mathcal{R}$-erased, and P is $\mathcal{R}$-erased too.

Remark 4.1 It can be shown that a redex $u \subseteq t \notin \mathcal{R}$ is $\mathcal{R}$-erased iff every residual of u (in particular, u itself) along any reductions starting from t is either $\mathcal{R}$-needed or $\mathcal{R}$-*inessential*. Here a subterm $s \subseteq t$ is $\mathcal{S}$-inessential iff there is no $\mathcal{S}$-normalizing P starting from t such that s has a P-*descendant*. The latter notion is a refinement of that of residual, allowing tracing of contracted redexes – the descendant of a contracted redex is its contractum, while it does not have residuals [Kha92]. One can show also that a redex $u \subseteq t \downarrow_{\mathcal{R}} \notin \mathcal{R}$ is $\mathcal{R}$-inessential iff it is $\mathcal{R}$-unneeded and $\mathcal{R}$-erased. Note that the latter proposition can be taken as the definition of $\mathcal{S}$-(in)essentiality, thus avoiding the use of the descendant concept, and the above characterization of $\mathcal{R}$-erased redexes follows logically. See [GlKh94a] for details.

Now we are ready to prove the main result of the paper.

Theorem 4.1 (**Minimal Relative Normalization**) Let $\mathcal{R}$ be a regular stable set of terms in an OERS, and let $t \downarrow_{\mathcal{R}} \notin \mathcal{R}$. Then repeated contraction of $\mathcal{R}$-needed $\mathcal{R}$-erased redexes in t yields an $\mathcal{R}$-minimal $\mathcal{R}$-normalizing reduction, even if a finite number of $\mathcal{R}$-unneeded $\mathcal{R}$-erased, and only such, redexes are also contracted. In particular, any $t \downarrow_{\mathcal{R}} \notin \mathcal{R}$ has an $\mathcal{R}$-unabsorbed $\mathcal{R}$-minimal reduction, which is $\mathcal{R}$-needed.

Proof. By Proposition 4.1, any $t \downarrow_{\mathcal{R}} \notin \mathcal{R}$ has an $\mathcal{R}$-unabsorbed redex, which is $\mathcal{R}$-needed and $\mathcal{R}$-erased by Lemma 4.4 and Lemma 4.1. It remains to apply Theorem 3.1 and Proposition 4.2.

Remark 4.2 (**Relative Standardization**) Note that $\mathcal{R}$-normalizing standard reductions (in the sense of [Bar84, Klo80], or in the sense of [GLM92], where left-to-right order of contracted redexes is not required) need not be $\mathcal{R}$-needed. Indeed, take for example $R = \{f(x) \to g(x,x),\ a \to b\}$, and take for $\mathcal{R}$ the set of terms not containing a redex on the right-spine; then $\mathcal{R}$ is regular, $f(a) \to g(a,a) \to g(b,a) \to g(b,b)$ is standard and $\mathcal{R}$-normalizing, but the second step is $\mathcal{R}$-unneeded. Therefore, we should take standard $\mathcal{R}$-minimal reductions for the $\mathcal{R}$-*standard* $\mathcal{R}$-*normalizing* reductions. It is not difficult to see that $\mathcal{R}$-unabsorbed $\mathcal{R}$-normalizing reductions are then $\mathcal{R}$-standard in the sense of [GLM92], and the left-to-right order of contraction of $\mathcal{R}$-unabsorbed redexes can also be achieved by Klop's standardization theorem [Klo80], which is valid for OERSs as well.

5 Relative optimal versus minimal reductions

Lévy introduced the notion of *redex family* in the λ-calculus, and showed that any multi-step reduction that in each multi-step contracts all redexes in a needed family (i.e., a family containing a needed redex) is optimal in the sense that it reaches a normal form (when it exists) in a minimal number of family-reduction steps [Lév78, Lév80]. This theory has been generalized to OTRSs, Interaction Systems, and higher-order rewrite systems [Mar91, AsLa93, Oos96], and to the case of relative normalization, to all Deterministic Family Structures [GlKh96]. The latter are abstract rewrite systems with axiomatized residual and family relations, and model family concepts in all orthogonal rewrite systems, OERSs among them. Redex families consist of 'redexes with the same origin', and here we only need to know that, in particular, all residuals of the same redex are in the same family.

It is easy to see that any $\mathcal{R}$-needed family-reduction that in each step contracts all the $\mathcal{R}$-needed redexes of some family, but does not necessarily contract its $\mathcal{R}$-unneeded members, is still optimal. We will call such reductions $\mathcal{R}$-needed *semi-complete* family-reductions. It follows from Proposition 4.2 that such a reduction is $\mathcal{R}$-minimal as well iff every $\mathcal{R}$-needed redex contracted in it is $\mathcal{R}$-erased. For example, $g(a) \to f(a,a) \to f(b,a)$ is both $\mathcal{R}$-minimal and $\mathcal{R}$-optimal semi-complete family-reduction in $R = \{g(x) \to f(x,x),\ a \to b\}$, where $\mathcal{R}$ is the set of terms not containing left-spine redexes. However, the following examples show that a term either in an OTRS or in the λ-calculus need not possess an $\mathcal{R}$-minimal $\mathcal{R}$-optimal family-reduction.

Example 5.1 Consider the OTRS $R = \{f(x) \to g(x,x),\ g(b,x) \to h(x,x),\ a \to b\}$, and take for $\mathcal{R}$ the set of terms not containing left-spine redexes. One can show that $\mathcal{R}$ is regular. Now $P : f(a) \to g(a,a) \to g(b,a) \to h(a,a) \to h(b,a)$ is an $\mathcal{R}$-minimal reduction, but $h(b,a)$ is not reachable by an $\mathcal{R}$-needed semi-complete family reduction. If the first step reduces a then we reach the $\mathcal{R}$-normal form $h(b,b)$ which is not $\mathcal{R}$-minimal. Hence, in order to reduce $f(a)$ to $h(b,a)$, one should delay contraction of the $\mathcal{R}$-needed occurrences of a (which all belong to the same family). So $f(a) \to g(a,a)$ must be the first step. In $g(a,a)$, both occurrences of a are $\mathcal{R}$-needed, but their contraction makes $h(b,a)$ unreachable. Thus there is no $\mathcal{R}$-minimal reduction that is $\mathcal{R}$-optimal at the same time.

Example 5.2 Take for $\mathcal{R}$ the set of λ-terms in head-normal form, which is regular, and take $t = (\lambda x.xx)u$, where $u = (\lambda y.\lambda z.zvz)w$, and y, z, v and w are different variables. Then $P : t \to uu \to (\lambda z.zvz)u \to uvu \to (\lambda z.zvz)vu \to vvvu = e$ is an $\mathcal{R}$-minimal reduction. In order to reach e from t by a semi-complete $\mathcal{R}$-needed family reduction, one should delay contraction of $\mathcal{R}$-needed redexes in the family of u. So the outermost redex in t must be contracted first. In the obtained term $o = uu$, both occurrences of u are $\mathcal{R}$-needed, and their contraction would make e unreachable – there is no occurrence of w in $(\lambda z.zvz)(\lambda z.zvz)$.

6 Conclusions and Future Work

We have studied minimal normalization relative to regular stable sets $\mathcal{R}$ of final terms, and have shown that $\mathcal{R}$-normalizing reductions that are both minimal and optimal need not exist for an $\mathcal{R}$-normalizable term t, despite the fact that t possesses minimal as well as optimal $\mathcal{R}$-normalizing reductions. These results were obtained for orthogonal ERSs, but are valid for Klop's CRSs and for context-sensitive conditional OERSs [KvO95], and therefore apply to numerous typed λ-calculi as well. We expect that the results remain valid for other systems of higher-order rewriting too.

Similar questions arise for infinite reductions. Stability and regularity of sets of finite and infinite reductions must be defined first, and we expect a strong connection between this concept and the concept of stability in interpretations [BeLé79].

As already mentioned in [GlKh94], it would be interesting to investigate strong sequentiality and strictness analyses for arbitrary stable sets of normal forms. Investigation of minimal relative normalization in an abstract setting seems also feasible and is useful.

Acknowledgments We thank J. R. Kennaway, J.-J. Lévy, L. Maranget, V. van Oostrom and F. van Raamsdonk for useful comments.

References

[AbOn93] Abramsky S., Ong C.-H. L. Full abstraction in the lazy lambda calculus. Inf.& Comp., 105:159-267, 1993.

[AEH94] Antoy S., Echahed R., Hanus M. A needed narrowing strategy. In: Proc. of POPL'94, Portland, Oregon, 1994.

[AnMi94] Antoy S. and Middeldorp A. A Sequential Reduction Strategy. In Proc. of ALP'94, Springer LNCS, vol. 850, p. 168-185, 1994.

[AsLa93] Asperti A., Laneve C. Interaction Systems I: the theory of optimal reductions. MSCS, 11:1-48, Cambridge University Press, 1993.

[Bar84] Barendregt H. P. The Lambda Calculus, its Syntax and Semantics. North-Holland, 1984.

[BKKS87] Barendregt H. P., Kennaway J. R., Klop J. W., Sleep M. R. Needed Reduction and spine strategies for the lambda calculus. Inf.& Comp., 75(3):191-231, 1987.

[BeLé79] Berry G., Lévy J.-J. Minimal and optimal computations of recursive programs. JACM 26, 1979, p. 148-175.

[Bou85] Boudol G. Computational semantics of term rewriting systems. In: Algebraic methods in semantics. Nivat M., Reynolds J.C., eds. Cambr. Univ. Press, 1985, p. 169-236.

[CuFe58] Curry H. B., Feys R. Combinatory Logic. vol. 1, North-Holland, 1958.

[Gar94] Gardner P. Discovering needed reductions using type theory. In: Proc. of TACS'94, Springer LNCS, v. 789, M. Hagiya, J. C. Mitchell, eds. Sendai, 1994, p. 555-574.

[GlKh94] Glauert J.R.W., Khasidashvili Z. Relative Normalization in Orthogonal Expression Reduction Systems. In: Proc. of CTRS'94, Springer LNCS, vol. 968, N. Dershowitz, N. Lindenstrauss, eds. Jerusalem, 1994, p. 144-165.

[GlKh94a] Glauert J.R.W., Khasidashvili Z. Minimal and optimal relative normalization in Expression Reduction Systems. Report SYS-C94-06, UEA, Norwich, 1994.

[GlKh96] Glauert J.R.W., Khasidashvili Z. Relative normalization in deterministic residual structures. In: Proc. of CAAP'96, Springer LNCS, vol. 1059, H. Kirchner, ed. 1996, p. 180-195.

[GLM92] Gonthier G., Lévy J.-J., Melliès P.-A. An abstract Standardisation theorem. In: LICS'92, p. 72-81.

[HuLé91] Huet G., Lévy J.-J. Computations in Orthogonal Rewriting Systems. In: Computational Logic, Essays in Honor of Alan Robinson, J.-L. Lassez and G. Plotkin, eds. MIT Press, 1991.

[Ken89] Kennaway J.R. Sequential evaluation strategy for parallel-or and related reduction systems. Annals of Pure and Applied Logic, 43:31-56, 1989.

[KeSl89] Kennaway J. R., Sleep M. R. Neededness is hypernormalizing in regular combinatory reduction systems. Preprint, University of East Anglia, Norwich, 1989.

[KKSV95] Kennaway J. R., Klop J. W., Sleep M. R, de Vries F.-J. Transfinite reductions in orthogonal term rewriting. Information and Computation, 119(1):18-38, 1995.

[Kha88] Khasidashvili Z. β-reductions and β-developments of λ-terms with the least number of steps. In: Proc. of the Int. Conference on Computer Logic COLOG'88, Tallinn 1988, Springer LNCS, v. 417, P. Martin-Löf and G. Mints, eds. 1990, p. 105-111.

[Kha92] Khasidashvili Z. The Church-Rosser theorem in Orthogonal Combinatory Reduction Systems. Report 1825, INRIA Rocquencourt, 1992.

[Kha93] Khasidashvili Z. Optimal normalization in orthogonal term rewriting systems. In: Proc. of RTA'93, Springer LNCS, vol. 690, C. Kirchner, ed. Montreal, 1993, p. 243-258.

[Kha94] Khasidashvili Z. On higher order recursive program schemes. In: Proc. of CAAP'94, Springer LNCS, vol. 787, S. Tison, ed. Edinburgh, 1994, p. 172-186.

[KvO95] Khasidashvili Z., van Oostrom V. Context-sensitive conditional expression reduction systems. Electronic Notes in Theoretical Computer Science, vol. 2, A. Corradini, U. Montanari, eds. Elsevier, 1995, p. 141-150.

[Klo80] Klop J. W. Combinatory Reduction Systems. Mathematical Centre Tracts n. 127, CWI, Amsterdam, 1980.

[KOR93] Klop J. W., van Oostrom V., van Raamsdonk F. Combinatory reduction systems: introduction and survey. In: To Corrado Böhm, TCS 121:279-308, 1993.

[Lév76] Lévy J.-J. An algebraic interpretation of the $\lambda\beta K$-calculus; and an application of a labelled λ-calculus. TCS 2(1):97-114, 1976.

[Lév78] Lévy J.-J. Réductions correctes et optimales dans le lambda-calcul, Thèse de l'Université de Paris VII, 1978.

[Lév80] Lévy J.-J. Optimal reductions in the Lambda-calculus. In: To H. B. Curry: Essays on Combinatory Logic, Lambda-calculus and Formalism, Hindley J. R., Seldin J. P. eds, Academic Press, 1980, p. 159-192.

[Lon83] Longo G. Set theoretic models of lambda calculus: theories, expansions and isomorphisms. Annals of pure and applied logic, 24:153-188, 1983.

[Mar91] Maranget L. Optimal derivations in weak λ-calculi and in orthogonal Term Rewriting Systems. In: Proc. POPL'91, p. 255-269.

[Mar92] Maranget L. La stratégie paresseuse. Thèse de l'Université de Paris VII, 1992.

[Nip93] Nipkow T. Orthogonal higher-order rewrite systems are confluent. In: Proc. of TLCA'93, Springer LNCS, vol. 664, Bazem M., Groote J.F., eds., 1993, p. 306-317.

[Nök94] Nöcker E. Efficient Functional Programming. Compilation and Programming Techniques. Ph. D. Thesis, Katholic University of Nijmegen, 1994.

[Oos94] Van Oostrom V. Confluence for Abstract and Higher-Order Rewriting. Ph.D. Thesis, Free University of Amsterdam, 1994.

[Oos96] Van Oostrom V. Higher order families. In proc. of RTA'96, Springer LNCS, vol. 1103, Ganzinger, H., ed., 1996, p. 392–407.

[OR94] Van Oostrom V., van Raamsdonk F. Weak orthogonality implies confluence: the higher-order case. In: Proc. of LFCS'94, Springer LNCS, vol. 813, Narode A., Matiyasevich Yu. V. eds. St. Petersburg, 1994. p. 379-392.

[Raa96] Van Raamsdonk F. Confluence and normalisation for higher-order rewriting. Ph.D. Thesis, CWI Amsterdam, 1996.

[SeRa93] Sekar R.C., Ramakrishnan I.V. Programming in Equational Logic: Beyond Strong Sequentiality. Information and Computation, 104(1):78-109, 1993.

This article was processed using the LaTeX macro package with LLNCS style