Minimal and Optimal Relative Normalization in Orthogonal Expression Reduction Systems

J. Glauert and Z. Khasidashvili

Technical Report SYS-C94-06

UEA Norwich, UK

Minimal and Optimal Relative Normalization in Orthogonal Expression Reduction Systems

John Glauert and Zurab Khasidashvili [1]
jrwg@sys.uea.ac.uk, zurab@sys.uea.ac.uk
School of Information Systems
University of East Anglia
Norwich NR4 7TJ, UK

Technical Report SYS-C94-06

Abstract

We study normalization relative to desirable sets $\mathcal{S}$ of 'normal forms' by generalizing Huet&Lévy theory of 'normalization by neededness'. We impose natural conditions on $\mathcal{S}$, called *stability*, that are sufficient and necessary for each term not in $\mathcal{S}$-normal form (i.e., not in $\mathcal{S}$) to have at least one $\mathcal{S}$-needed redex, and repeated contraction of $\mathcal{S}$-needed redexes in a term t to lead to an $\mathcal{S}$-normal form of t whenever there is one. Further, we prove existence of *minimal* normalizing reductions for *regular* stable sets of normal forms. For example, the sets of normal forms, head-normal-forms, and weak head-normal-forms, in the λ-calculus, are all stable and regular. Finally, we generalize Lévy's Optimality theorem to the case of all stable sets of normal forms, and establish a relationship between relative minimal and optimal reductions, revealing a conflict between minimality and optimality of a reduction – for regular stable sets of normal forms, a term need not possess a reduction that is minimal and optimal at the same time.

All these results are obtained in the framework of orthogonal (left-linear and non-ambiguous) Expression Reduction Systems, a formalism for Term Rewriting Systems with bound variables and substitutions.

© J. Glauert & Z. Khasidashvili, UEA Norwich, 1994

[1] Supported by the Engineering and Physical Sciences Research Council of Great Britain under grant GR/H 41300

1 Introduction

Since a normalizable term, in a rewriting system, may have an infinite reduction, it is important to have a *normalizing* strategy which enables one to construct reductions to normal form. It is well known that the leftmost-outermost strategy is normalizing in the λ-calculus [CuFe58]. For Orthogonal Term Rewriting Systems (OTRSs), a general normalizing strategy, called the *needed* strategy, was found by Huet and Lévy in [HuLé91]. The needed strategy always contracts a *needed* redex – a redex whose residual is to be contracted in any reduction to normal form. Huet and Lévy showed that any term t not in normal form has a needed redex, and that repeated contraction of needed redexes in t leads to its normal form whenever there is one; we refer to it as the *Normalization Theorem*. They also defined the class of *strongly sequential* OTRSs where a needed redex can efficiently be found in any term.

Barendregt et al. [BKKS87] generalized the concept of neededness to the λ-calculus. They studied neededness not only w.r.t. normal forms, but also w.r.t. head-normal forms – a redex is *head-needed* if its residuals are contracted in each reduction to a head-normal form. The authors proved correctness of the two needed strategies for computing normal forms and head-normal forms, respectively. Khasidashvili defined a similar normalizing strategy, called the *essential* strategy, for the λ-calculus [Kha88] and OTRSs [Kha93]. The strategy contracts *essential* redexes – the redexes that have *descendants* under any reduction. The notion of descendant is a refinement of that of *residual* – the descendant of a contracted redex is its contractum, while it does not have residuals. This refined notion allows for much simpler proofs of correctness of the essential strategy in OTRSs and the λ-calculus, which generalize straightforwardly to all Orthogonal *Expression Reduction Systems* (OERSs) [Kha90] and Klop's Orthogonal *Combinatory Reduction Systems* (OCRSs) [Klo80]. Kennaway and Sleep [KeSl89] used a generalization of Lévy's labelling for the λ-calculus [Lév78] to adapt the proof from [BKKS87] to the case of Klop's OCRSs, which can also be applied to OERSs. Khasidashvili [Kha94] showed that in *Persistent* OERSs, where redex-creation is limited, one can find *all* needed redexes in any term. Gardner [Gar94] described a *complete* way of encoding neededness information, for the case of the λ-calculus, using a type assignment system in the sense that using the principal type of a term one can find all the needed redexes in it. Antoy et al. [AEH94] designed a needed narrowing strategy.

In [Mar92], Maranget introduced a different notion of neededness, where a redex u is needed if it has a residual under any reduction that does not contract the residuals of u. This neededness notion makes sense also for terms that do not have a normal form, and coincides with the notion of essentiality (essentiality makes sense for all subterms, not only for redexes). In [Mar92], Maranget studied also 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].

A question arises naturally: what are the properties that a set of terms must possess in order for the neededness theory of Huet and Lévy still to make sense? The main contribution of this paper is to provide a solution to that question. We introduce the notion of *neededness* w.r.t. a set of reductions Π or a set of terms $\mathcal{S}$ so that each existing notion of neededness can be given by specifying Π or $\mathcal{S}$. For example, *Huet&Lévy-neededness* is neededness w.r.t. the set NF of normal forms, *Maranget-neededness* is neededness w.r.t. all fair reductions, *head-neededness* is neededness w.r.t. the set of head-normal forms, etc. We impose natural conditions on $\mathcal{S}$, called *stability*, that are sufficient and necessary for each term not in $\mathcal{S}$-normal form (i.e., not in $\mathcal{S}$)

to have at least one $\mathcal{S}$-needed redex, and repeated contraction of $\mathcal{S}$-needed redexes in a term t to lead to an $\mathcal{S}$-normal form of t whenever there is one.

A set $\mathcal{S}$ of terms is stable if it is *closed under parallel moves*: for any $t \notin \mathcal{S}$, any $P : t \twoheadrightarrow o \in \mathcal{S}$, and any $Q : t \twoheadrightarrow e$ not containing terms in $\mathcal{S}$, the final term of P/Q, the *residual* of P under Q, is in $\mathcal{S}$; and is *closed under unneeded expansion*: for any $e \xrightarrow{u} o$ such that $e \notin \mathcal{S}$ and $o \in \mathcal{S}$, a residual of u is contracted in any reduction from e to a term in $\mathcal{S}$. We present a counterexample to show that the $\mathcal{S}$-needed strategy is not *hypernormalizing* for every stable $\mathcal{S}$, i.e., an $\mathcal{S}$-normalizable term may possess a reduction which never reaches a term in $\mathcal{S}$ even though $\mathcal{S}$-needed redexes are contracted infinitely many times, but $\mathcal{S}$-unneeded redexes are contracted as well. Therefore, *multistep* $\mathcal{S}$-needed reductions need not be $\mathcal{S}$-normalizing. This is because a 'non-standard' situation – when $\mathcal{S}$-unneeded redexes may contain $\mathcal{S}$-needed ones – may occur for some, call it *irregular*, stable sets $\mathcal{S}$. However, if $\mathcal{S}$ is a regular stable set, then the $\mathcal{S}$-needed strategy is again hypernormalizing, and the multistep $\mathcal{S}$-needed strategy is normalizing.

We further develop a theory of *minimal* and *optimal* reduction in the framework of relative normalization, and establish a relationship between them. 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 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. The earliest (to our knowledge) 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. Minimal and optimal reductions were studied also in arbitrary interpretations, not only for term (or Herbrand) models.

In this paper, we restrict ourselves to finite reductions only, and study only syntactic properties. 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. These redexes play the role of *standard* redexes w.r.t. $\mathcal{S}$, since they are $\mathcal{S}$-needed and cannot be duplicated. 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 not regular.

To give the characterization of $\mathcal{S}$-minimal reductions, we introduce a *relative* notion of *essentiality*. A subterm $s \subseteq t$ is said to be $\mathcal{S}$-essential if it has a descendant under at least one $\mathcal{S}$-normalizing reduction of t. Essentiality [Kha93, Kha88] of subterms in normalizable terms coincides with NF-essentiality, i.e. essentiality w.r.t. the set of normal forms. NF-essentiality of redexes in normalizable terms coincides with both Huet&Lévy and Maranget's concepts of neededness, while in the general case, for stable $\mathcal{S}$, the two concepts are different. All $\mathcal{S}$-needed redexes are $\mathcal{S}$-essential. $\mathcal{S}$-unneeded redexes may be $\mathcal{S}$-essential, in which case they contribute to the minimal $\mathcal{S}$-normal form, or can be $\mathcal{S}$-inessential in which case they will be erased under any reduction to $\mathcal{S}$-normal form. Inessentiality is related to the notion of garbage – an inessential subterm may be replaced by any term without affecting the normal form. However, for certain stable and even regular $\mathcal{S}$, replacing an $\mathcal{S}$-inessential subterm by another term may lead to an $\mathcal{S}$-essential subterm at the same position, and the $\mathcal{S}$-normal form of the term can change.

We show that an $\mathcal{S}$-normalizing reduction is $\mathcal{S}$-minimal iff it contracts $\mathcal{S}$*-erased* redexes, i.e., the redexes that do not have residuals under any $\mathcal{S}$-normalizing reduction. Further, we show that a redex u in t is $\mathcal{S}$-erased iff every one of its residuals is either $\mathcal{S}$-needed or $\mathcal{S}$-inessential.

Our study of optimal normalization w.r.t. stable sets $\mathcal{S}$ is a generalization of Lévy's optimality theory [Lév80], developed for the λ-calculus. The generalization is rather straightforward, since a suitable labelling system allowing us to define a family relation in OERSs, due to Kennaway&Sleep [KeSl89], is available. We show that complete $\mathcal{S}$-needed family-reductions, which contract all members of a *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.

It is easy to see that $\mathcal{S}$-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 could 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.

The rest of the paper is organized as follows. In the next section, we review Expression Reduction Systems, a formalism for higher order rewriting that we use here [Kha90, Kha92]. Section 3 establishes equivalence of Maranget's neededness and our essentiality for OERSs. In section 4, we introduce the relative notion of neededness. In section 5, we sketch some properties of the labelling system of Kennaway&Sleep [KeSl89] for OERSs needed to define a *family-relation* among redexes. We prove correctness of the $\mathcal{S}$-needed strategy for finding terms of $\mathcal{S}$, for all stable $\mathcal{S}$, in section 6. In section 7, we study $\mathcal{S}$-minimal reductions for regular stable sets $\mathcal{S}$. In section 8, we study the relative notion of essentiality, and in section 9, reductions standard w.r.t. regular stable sets. In section 10, we prove the Relative Optimality theorem. Finally, in section 11, we relate relative optimal and minimal reductions. The conclusions appear in section 12. The results in Sections 3 – 6 are published as [GlKh94], there we in addition study relative neededness and essentiality properties of *components* of terms in OERSs, which are occurrences of contexts. This requires the introduction of notions of *descendants* and *residuals* for components.

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] and its extensions. Restricted rewriting systems with substitutions were first studied in Pkhakadze [Pkh77] and Aczel [Acz78]. Several interesting formalisms have been introduced later [Kha90, Nip93, Wol93, OR94]. We refer to Klop et al. [KOR93] and van Oostrom [Oos94] for a survey. Here we use a system of higher order rewriting, *Expression Reduction Systems* (ERSs), defined in Khasidashvili [Kha90, Kha92] (ERSs are called CRSs in [Kha92]); the present formulation 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 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 $Ap(\lambda x\, t, s)$, 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 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*; patterns and arguments are defined similarly for the contractum of the redex if the right-hand side of the rule is a simple metaterm. 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.

To express substitution, we use the S-reduction rules

$$S^{n+1} x_1 \ldots x_n A_1 \ldots A_n A_0 \to (A_1/x_1, \ldots, A_n/x_n)A_0, \quad n = 1, 2, \ldots,$$

where S^{n+1} is the *operator sign of substitution* with arity $(n, n+1)$ and scope indicator $(n+1)$, and $x_1, \ldots, x_n$ and $A_1, \ldots, A_n, A_0$ are pairwise distinct variables and metavariables. Thus S^{n+1} binds free variables only in the last argument. The difference with β-rules is that S-reductions can only perform β-developments of λ-terms [Kha92]. Therefore, S-redexes are similar to simultaneous **let**-expressions.

Notation 2.1 We use a, b, c, d for constants, t, s, e, o for terms and metaterms, 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 in which a redex $u \subseteq t$ is contracted is written as $t \xrightarrow{u} s$ or $t \to s$ or just u. We write $P : t \twoheadrightarrow s$ if P denotes a reduction of t to s. $|P|$ denotes the length of P. $P+Q$ denotes the concatenation of P and Q. $FV(t)$ denotes the set of free variables of t.

We now give a brief but complete definition of *descendants* of subterms as introduced in [Kha88, Kha93, Kha92] for the λ-calculus, TRSs, and ERSs, respectively. [2] First, we need to split an ERS R into a *TRS-part* R_f and the *substitution-part* S. For any ERS R, which we assume does not contain symbols S^{n+1}, we define $R_{fS} =_{def} R_f =_{def} R$ if R is simple. Otherwise, we define $R_{fS} =_{def} R_f \cup S$, where R_f is the ERS obtained from R by adding symbols S^{n+1} in the alphabet and by replacing in right-hand sides of the rules all metasubstitutions of the form $(t_1/x_1, \ldots, t_n/x_n)t_0$ by $S^{n+1} x_1 \ldots x_n t_1 \ldots t_n t_0$, respectively. For example, the β_f rule would be $\beta_f : Ap(\lambda x A, B) \to S^2 xBA$. Obviously, an R-step can be *refined* into an R_f-step, followed by a number of S-steps, evaluated say in right to left order. For example, the refinement of the β-step $Ap(\lambda x t, s) \to_\beta (s/x)t$ would be $Ap(\lambda x t, s) \to_{\beta_f} S^2 xst \to_S (s/x)t$. The notion of refinement generalizes to R-reductions with 0 or more steps.

Definition 2.3 Let $t \xrightarrow{u} s$ be an R_f-step and let e be the contractum of u in s. For each argument o of u there are 0 or more arguments of e. We call them (u-)*descendants* of o. Correspondingly, subterms of o have 0 or more *descendants*. The *descendant* of each pattern-subterm (i.e., a subterm rooted within the pattern) of u is e. It is clear what is to be meant by *descendants* of a subterm of t that is not in u.

[2] Klop [Klo92] gives a slightly different notion of descendant for TRSs.

For example, in the step $t = f(g(a)) \to b = s$ according to the rule $f(g(x)) \to b$, the descendant of both pattern-subterms $f(g(a))$ and $g(a)$ of t in s is b, and a doesn't have a descendant in s.

Definition 2.4 Let $t = C[S^{n+1}x_1 \ldots x_n t_1 \ldots t_n t_0] \xrightarrow{u} C[(t_1/x_1, \ldots, t_n/x_n)t_0] = s$ be an S-step. Then the descendants of an argument t_i of u are the substituted occurrences of t_i in s; so t_i has the same number of descendants as the number of free occurrences of x_i in t_0. Subterms of t_i have the same number of *descendants*, which are the corresponding subterms occurring in the descendants of t_i. All subterms of t_0 have exactly one *descendant* (the descendants of the free occurrences of $x_1, \ldots, x_n$ in t_0 are the substituted subterms). The *descendant* of the contracted redex u itself is its contractum. It is clear what is to be meant by *descendants* of a subterm of t rooted within $C[\,]$.

For example, in the step $o = Sxaf(x) \to f(a) = e$, the descendant of $a \subseteq o$, as well as of the bound occurrence of x in $f(x)$, is the occurrence of a in e. The descendant of $f(x)$, as well as of o itself, is e.

The notion of *descendant* extends by transitivity to arbitrary R_{fS}-reductions. If P is an R-reduction, then P-*descendants* are defined to be the descendants under the refinement of P. The *ancestor* relation is converse to that of descendant. If $t \xrightarrow{u} s$ in R, then the descendants of any redex $v \neq u$ are called *residuals* of v; a redex in s is called a *new* redex or a *created* redex if it is not a residual of a redex of t. The notion of *residual* extends by transitivity to arbitrary R-reductions.

Recall that $P : t \twoheadrightarrow s$ and $Q : t \twoheadrightarrow e$ are *Hindley-equivalent*, written $P \approx_H Q$, if $s = e$ and the residuals of any redex of t under P and under Q are the same redexes in s; and are *strictly equivalent*, written $P \approx_{st} Q$, if $s = e$ and the descendants of any subterm of t under P and under Q are the same subterms in s [Kha88].

Definition 2.5 A rewrite rule $t \to s$ in an ERS R is *left-linear* if t is linear, i.e., no metavariable occurs more than once in t. R is *left-linear* if each rule in R is so. $R = \{r_i \,|\, i \in I\}$ is *non-ambiguous* or *non-overlapping* if in no term redex-patterns can overlap, i.e., if r_i-redex u contains an r_j-redex u' and $i \neq j$, then u' is in an argument of u, and the same holds if $i = j$ and u' is a proper subterm of u. R is *orthogonal* (OERS) if it is left-linear and non-overlapping.

As in the case of OTRSs [HuLé91] and 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]. $P \sqcup Q$ abbreviates $P + Q/P$. 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, P', Q and Q' are co-initial reductions in an OERS, then $(P + P')/Q \approx_L P/Q + P'/(Q/P)$ and $P/(Q + Q') \approx_L (P/Q)/Q'$. This is all well known, and we do not give more details here.

Below, we use the following *strong* and *strict* Church-Rosser properties:

Theorem 2.1 (1) [Klo80] **(Strong Church-Rosser)** For any co-initial reductions P and Q in an OERS, $P + Q/P \approx_L Q + P/Q$; and if $P \approx_L Q$, then $P \approx_H Q$.

(2) [Kha92] **(Strict Church-Rosser)** Let $P \approx_L Q$, in an OERS. Then $P \approx_{st} Q$.

3 Neededness, Essentiality and Unabsorbedness

In this section, we recall Huet&Lévy and Maranget's notions of neededness, and relate them to the notion of essentiality, in OERSs. We also prove existence of an unabsorbed, hence essential, redex in any term, in an OERS, not in normal form.

Definition 3.1 A redex u in t is *Huet&Lévy-needed* [HuLé91, Lév80] if in each reduction of t to normal form (if any) at least one residual of u is contracted; u is *Maranget-needed* [Mar92] if u has at least one residual under any reduction starting from t that does not contract residuals of u.

Definition 3.2 A subterm s in t is *essential* (written $ES(s,t)$) if it has at least one descendant under any reduction starting from t and is *inessential* (written $IE(s,t)$) otherwise [Kha93, Kha88].

Definition 3.3 A subterm s of a term t is *unabsorbed in a reduction* $P : t \twoheadrightarrow e$ if none of its descendants appear in redex-arguments of terms in P, and is *absorbed in* P otherwise; s is *unabsorbed in* t if it is unabsorbed in any reduction starting from t and *absorbed in* t otherwise [Kha93].

It is easy to see that a redex $u \subseteq t$ is unabsorbed iff u is *external* [HuLé91] in t. Clearly, unabsorbedness implies essentiality, and Huet&Levy- and Maranget-neededness coincide for normalizable terms.

Definition 3.4 (1) Let $P : t \to t_1 \to \ldots$ and $u \subseteq t_i$. Then u is *erased* in P if u does not have residuals in t_j for some $j > i$. P is *fair* if each redex in any t_i is erased in P [Klo92].

(2) We call a reduction N starting from t *strictly cofinal* if for any $Q : t \twoheadrightarrow e$ there is an initial part $N' : t \twoheadrightarrow s$ of N, and a reduction $Q' : e \twoheadrightarrow s$ such that $N' \approx_{st} Q + Q'$.

(3) We say that $P : t \to t_1 \to \ldots$ *deletes* $o \subseteq t_i$ if o doesn't have descendants in t_j for some $j > i$.

Lemma 3.1 Any strictly cofinal reduction P starting from t deletes all inessential subterms of t.

Proof Immediate from Definitions 3.2 and 3.4.

For example, fair reductions are strictly cofinal: Klop's proof of cofinality of fair reductions (Theorem 12.3 in [Klo80]) can be modified to a proof of strict cofinality of fair reductions by using Theorem 2.1.(2) instead of the CR theorem.

The following lemma follows from Theorem 2.1.(2); proofs are the same as for OTRS [Kha93].

Lemma 3.2 (1) Let $P : t \twoheadrightarrow t'$ and $s \subseteq t$. Then $IE(s,t)$ iff all P-descendants of s are inessential in t'.

(2) Let $t \xrightarrow{u} t'$ and $e \subseteq s \subseteq t$. Then any u-descendant of e is contained in some u-descendant of s.

(3) Let $e \subseteq s \subseteq t$ and $ES(e,t)$. Then $ES(s,t)$.

(4) Let s be a pattern-subterm of a redex $u \subseteq t$. Then $ES(u,t)$ iff $ES(s,t)$.

Proof (1) ($\Rightarrow$) Let $IE(s,t)$. Then some $Q : t \twoheadrightarrow e$ deletes s. Then, by Theorem 2.1.(2), Q/P deletes all P-descendants of s. ($\Leftarrow$) If all P-descendants of s are inessential in t', there is an $N : t' \twoheadrightarrow s'$ that deletes them, and $P + N$ deletes s, i.e., $IE(s,t)$. (2) is immediate from the definition of descendant relation. (3) follows from (2). The contractum of u is the unique descendant of u, as well as of s. Hence (4) follows from (1).

Notation We write $t = (t_1//s_1, \ldots, t_n//s_n)s$ if $s_1, \ldots, s_n$ are disjoint subterms in s and t is obtained from s by replacing them with $t_1, \ldots, t_n$, respectively.

Definition 3.5 Let $t = (t_1//s_1, \ldots, t_n//s_n)s$ in an OERS R and let $P : s = e_0 \stackrel{v_0}{\to} e_1 \stackrel{v_1}{\to} \ldots$ be an R_{fS}-reduction. We define the reduction $P\|(t) : t = o_0 \stackrel{u_0}{\to} o_1 \stackrel{u_1}{\to} \ldots$ as follows.

(1) If v_0 is an R_f-redex and its pattern does not overlap with $s_1, \ldots, s_n$, then u_0 is the corresponding subterm of v_0 in $t = o_0$; obviously, u_0 is an R_f-redex too.

(2) If v_0 is an R_f-redex that is not inside the subterms $s_1, \ldots, s_n$ and its pattern does overlap with some of $s_1, \ldots, s_n$, then $u_0 = \emptyset$.

(3) If v_0 is an S-redex that is outside the replaced subterms, then u_0 is the corresponding S-redex in t_0.

(4) If v_0 is in one of the subterms $s_1, \ldots, s_n$, then $u_0 = \emptyset$.

In the first case, o_1 is obtained from e_1 by replacing the descendants of $s_1, \ldots, s_n$ with the corresponding descendants of $t_1, \ldots, t_n$, respectively; in the second case, o_1 is obtained from e_1 by replacing the contractum of v_0 by the descendant of its corresponding subterm in o_0, and by replacing the descendants of the subterms s_i that do not overlap with v_0 by the corresponding descendants of t_i; in the third case, o_1 is obtained from e_1 by replacing outermost descendants of $s_1, \ldots, s_n$ with the corresponding descendants of $t_1, \ldots, t_n$; in the fourth case, o_1 is obtained from e_1 by replacing the descendants of $s_1, \ldots, s_n$ with the corresponding descendants of $t_1, \ldots, t_n$, respectively. Thus, in o_1 we can choose the redex u_1 analogously, and so on. (Note that $P\|(t)$ depends not only on P and t, but also on the choice of $s_1, \ldots, s_k$, but the notation does not give rise to ambiguity.)

Lemma 3.3 Let $s_1, \ldots, s_n$ be inessential in s, in an OERS R, and let $t = (t_1//s_1, \ldots, t_n//s_n)s$. Then $t_1, \ldots, t_n$ are inessential in t.

Proof We show by induction on $|Q|$ that if an R_{fS}-reduction $Q : s \twoheadrightarrow o$ deletes $s_1, \ldots, s_n$, then $Q\|(t)$ deletes $t_1, \ldots, t_n$; such a Q exists by Lemma 3.1. Let $Q = v + Q'$, $s \stackrel{v}{\to} e$, and $s'_1, \ldots, s'_m$ be all the v-descendants of $s_1, \ldots, s_n$. By Lemma 3.2.(1), s'_i are inessential in e. By Definition 3.5 and Lemma 3.2, if $v\|(t) : t \to o$, then o is obtained from e by replacing some inessential subterms that contain $s'_1, \ldots, s'_m$, and all the descendants of $t_1, \ldots, t_n$ also are in the replaced subterms of o. By the induction assumption, the replaced subterms in o are inessential. Hence, by Lemma 3.2.(3), all the descendants of $t_1, \ldots, t_n$ in o are inessential and, by Lemma 3.2.(1), $t_1, \ldots, t_n$ are inessential in t.

It follows immediately from Definition 3.5 and the proof of Lemma 3.3 that replacement of inessential subterms in a term does not effect its normal form.

Corollary 3.1 Any term t not in normal form, in an OERS, contains an essential redex.

Proof If all redexes in t were inessential, their replacement by fresh variables would yield a term in normal form containing inessential subterms, a contradiction.

Proposition 3.1 Let t be a term in an OERS R and let u be a redex in t. Then u is essential in t iff it is Maranget-needed in t.

Proof ($\Leftarrow$) Let $IE(u,t)$. Further, let $FV(u) = \{x_1, \ldots, x_n\}$, let f be a fresh n-ary function symbol that does not occur in the left-hand sides of rewrite rules (we can safely add such a symbol to the alphabet, if necessary), and let $s = (f(x_1, \ldots, x_n)//u)t$. By Lemma 3.3, $IE(f(x_1, \ldots, x_n), s)$, i.e., there is some reduction P starting from s such that $f(x_1, \ldots, x_n)$ does not have P-descendants. Let $r : f(x_1, \ldots, x_n) \to u$ be a rule and $Q : s \to t$ be the r-reduction step. Obviously, $R \cup \{r\}$ is orthogonal. Hence $P + Q/P \approx_{st} Q + P/Q$ and therefore u does not have P/Q-descendants. But P/Q does not contract the residuals of u. Thus u is not Maranget-needed. ($\Rightarrow$) From Definitions 3.2 and 3.1.

We now proof existence of an unabsorbed redex in any term not in normal form, in an OERS; the proof is exactly as in OTRSs [Kha93] (it does not use the proof of Corollary 3.1).

Lemma 3.4 Let $P : t_0 \stackrel{u_0}{\to} t_1 \stackrel{u_1}{\to} \ldots \to t_n$. Then for any outermost redex $v \subseteq t_n$ there is an unabsorbed in P redex $w \subseteq t_0$ that has a descendant $s \subseteq t_n$ containing a pattern-subterm e of v.

Proof By induction on $|P|$. If $|P| = 0$, then the lemma is obvious. So suppose $P : t_0 \stackrel{P'}{\twoheadrightarrow} t_{n-1} \stackrel{u_{n-1}}{\to} t_n$.

(a) Let v be a residual of a redex v' of t_{n-1}. Further, let $v^* = v'$ if $v' \not\subseteq u_{n-1}$ and $v^* = u_{n-1}$ otherwise. Since v is outermost, v^* is outermost as well. By the induction assumption, a descendant $s' \subseteq t_{n-1}$ of $w \subseteq t_0$ contains a pattern-subterm of v^* (s' is the only descendant of w since w is unabsorbed in P' by the induction assumption). Therefore, by Lemma 3.2.(2), s', and hence also w, have a descendant that contains a pattern-subterm of v, and w is unabsorbed in P.

(b) Assume now that v is created by u_{n-1}. Then u_{n-1} is outermost and, by the induction assumption, a pattern-subterm e' of u_{n-1} is contained in a descendant s' of w (the only one). Let e be the maximal common subterm of v and the contractum of u_{n-1}. Then e is a pattern-subterm of v. Since e' is a pattern-subterm of u_{n-1}, its descendant is the contractum of u_{n-1} and hence contains e. By Lemma 3.2.(2), s', and hence also w, have a descendant s that contains the pattern-subterm e of v. So w is unabsorbed in P as well.

Lemma 3.5 For any reduction $P : t \twoheadrightarrow s$ starting from a term t not in normal form there is an outermost redex u in t that is unabsorbed in P.

Proof If $|P| = 0$ or $|P| > 0$ and s is not in normal form, then the lemma follows immediately from Lemma 3.4. Otherwise, let $P : t \twoheadrightarrow s' \stackrel{v}{\to} s$. By Lemma 3.4 there is an unabsorbed in $t \twoheadrightarrow s'$ redex $u \subseteq t$ whose only descendant contains a pattern-subterm of v. Thus u has a descendant in s that is not in a redex-argument, since s is in normal form. Hence, u is unabsorbed in P.

Lemma 3.6 Let u be absorbed in $P : t \twoheadrightarrow s$ and let $Q : t \twoheadrightarrow e$. Then u is absorbed in $Q \sqcup P$.

Proof It suffices to prove the lemma for $|Q| = 1$, i.e., $Q = w$ for a redex w in t. The rest follows by induction on $|Q|$. Furthermore, we can assume that u is absorbed in the last step of P, i.e., there is a redex v created in the last step of P and containing the only descendant o of u in its argument. If o does not have a w/P-descendant, then, by Theorem 2.1.(2), u does not have $(w \sqcup P)-$descendants, so u is absorbed in $w \sqcup P$ (otherwise its descendants cannot be deleted). If o has a w/P-descendant o', then, by Lemma 3.2.(2), v also has a w/P-descendant that contains o'. Since in w/P only residuals of w are contracted and v is a new redex, v has a residual v' that contains o'. But, by Theorem 2.1.(2), o' also is a $(w \sqcup P)-$descendant of u. Hence, u is absorbed in $w \sqcup P$.

Proposition 3.2 Let t be a term not in normal form in an OERS. Then there is an unabsorbed outermost — hence essential — redex in t.

Proof If, for any outermost redex u_i of t, there is a reduction P_i that absorbs u_i $(i = 1, \ldots, k)$, then, by Lemma 3.6, $P_1 \sqcup \ldots \sqcup P_k$ absorbs all redexes u_i, contrary to Lemma 3.5.

4 Relative Notions of Neededness

In this section, we introduce notions of neededness relative to a set of reductions Π and to a set of terms $\mathcal{S}$. We show how all existing notions of neededness can be obtained by specifying Π or $\mathcal{S}$; $\mathcal{S}$-neededness is also a special case of Π-neededness. We introduce *stability* of a set of terms, in an OERS, and show that if $\mathcal{S}$ is not stable, contraction of $\mathcal{S}$-needed redexes in a term t need not terminate at a term in $\mathcal{S}$ even if t can be reduced to a term in $\mathcal{S}$. It is the aim of the Section 6 to show that if $\mathcal{S}$ is stable, then a $\mathcal{S}$-needed strategy is $\mathcal{S}$-normalizing.

Definition 4.1 (1) We call $P : t \twoheadrightarrow o$ *external* to a redex $u \subseteq t$ if P doesn't contract the residuals of u.

(2) We say that P *$\mathcal{S}$-suppresses* u if P is $\mathcal{S}$-normalizing and is external to u.

(3) We say that P *discards* u if P is external to u and $u/P = \emptyset$.

Definition 4.2 (1) Let Π be a set of reductions in an OERS R. We call a redex $u \subseteq t$ *Π-needed* if at least one residual of it is contracted in every reduction in Π starting from t, and call Π-*unneeded* otherwise.

(2) 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. (Thus u is $\mathcal{S}$-needed iff it is $\Pi_{\mathcal{S}}$-(un)needed, where $\Pi_{\mathcal{S}}$ be the set of all $\mathcal{S}$-normalizing reductions, i.e., reductions that end at a term in $\mathcal{S}$.)

Obviously, any redex in a term that is not $\mathcal{S}$-normalizable is $\mathcal{S}$-needed; we call such redexes *trivially $\mathcal{S}$-needed*.

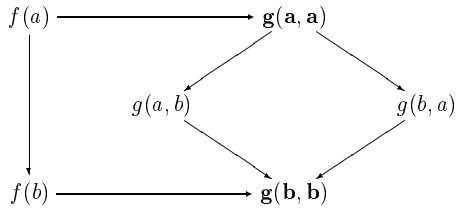
The following definition introduces the property of sets of terms for which it is possible to generalise the Normalization Theorem:

Definition 4.3 We call a set $\mathcal{S}$ of terms *stable* if:

(a) $\mathcal{S}$ is *closed under parallel moves*: for any $t \notin \mathcal{S}$, any $P : t \twoheadrightarrow o \in \mathcal{S}$, and any $Q : t \twoheadrightarrow e$ that doesn't 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.

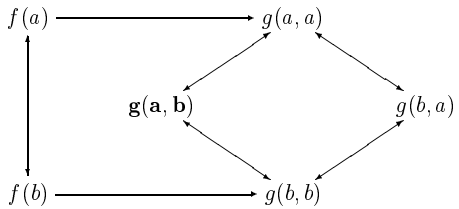
Remark 4.1 Of course, a set closed under reduction is closed under parallel moves as well. But a set closed under parallel moves, even if closed under unneeded expansion, need not be closed under reduction. Indeed, consider $R = \{f(x) \to g(x,x),\ a \to b\}$, and take $\mathcal{S} = \{g(a,a), g(b,b)\}$. The only one-step $\mathcal{S}$-normalizing reductions are $g(a,b) \to g(b,b)$, $g(b,a) \to g(b,b)$, $f(a) \to g(a,a)$, and $f(b) \to g(b,b)$. Therefore, one can check that $\mathcal{S}$ is closed under unneeded expansion. Also, $\mathcal{S}$ is closed under parallel moves, since the right-bottom term $g(b,b)$ in the diagram below, which is the only non-trivial diagram to be checked, is in $\mathcal{S}$. However, $\mathcal{S}$ is not closed under reduction, since, e.g., $g(a,a) \to g(b,a)$, $g(a,a) \in \mathcal{S}$, but $g(b,a) \notin \mathcal{S}$. Note that the second occurrence of a in $g(a,a)$ is $\mathcal{S}$-unneeded, but its residual in $g(b,a)$ is $\mathcal{S}$-needed.



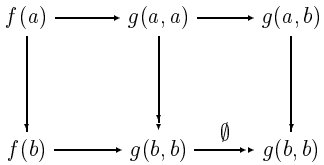
The most appealing examples of stable sets, for an OERS, are the set of normal forms [HuLé91], the set of head-normal forms [BKKS87], the set of weak-head-normal forms (a partial result is in [Mar92]), and the set of constructor-head-normal forms for constructor TRSs [Nök94]. The graph G_s of a term s (which consists of terms to which s is reducible) is closed under reduction, but need not be closed under unneeded expansion. For example, the graph $G_{I(x)} = \{I(x),\ x\}$ of $I(x)$ is closed under reduction but is not closed under unneeded expansion: $I(I(x))$ can be reduced to $I(x)$ by reducing either I-redex (according to the rule $I(x) \to x$). Hence *none* of the redexes in $I(I(x))$ are $\mathcal{S}$-needed. Thus the closure of $\mathcal{S}$ under unneeded expansion is a necessary condition for the Normalization Theorem.

We say that a set $\mathcal{S}$ of terms is *closed under ($\mathcal{S}$)-normalization* if any reduct of any $\mathcal{S}$-normalizable term is still $\mathcal{S}$-normalizable. Obviously, sets closed under parallel moves are closed under normalization as well. Even if $\mathcal{S}$ is closed under unneeded expansion, closure of $\mathcal{S}$ under normalization is also necessary for the normalization theorem to be valid for $\mathcal{S}$. Indeed, consider $R = \{f(x) \to g(x,x),\ a \to b\}$, take $\mathcal{S} = \{g(a,b)\}$, and take $t = f(a)$. Then $t \to g(a,a) \to g(a,b)$ is an $\mathcal{S}$-needed $\mathcal{S}$-normalizing reduction, while after the $\mathcal{S}$-needed step $t \to f(b)$, the term $f(b)$ is not $\mathcal{S}$-normalizable any more (the only redex in $f(b)$ is only trivially $\mathcal{S}$-needed). However, the following example shows that closure of $\mathcal{S}$ under normalization (even in combination with closure of $\mathcal{S}$ under unneeded expansion) is not enough; closure of $\mathcal{S}$ under parallel moves is necessary.

Example 4.1 Let $R = \{f(x) \to g(x,x),\ a \to b,\ b \to a\}$ and $\mathcal{S} = \{g(a,b)\}$. Since the reduction preserves the height of a term and the property to be a ground term, only the terms in the following diagram are $\mathcal{S}$-normalizable.



Therefore, it is clear from the diagram that $\mathcal{S}$ is closed under normalization. It is easy to see that, in $f(a)$ and $f(b)$, all the redexes are $\mathcal{S}$-needed; hence $f(a) \to f(b) \to f(a) \to \ldots$ is an infinite $\mathcal{S}$-needed reduction that never reaches $\mathcal{S}$ (there are many others). One can check that $\mathcal{S}$ is closed under unneeded expansion. Thus the reason for the failure of the normalization theorem is that, as it can be seen from the following diagram, $\mathcal{S}$ is not closed under parallel moves.



Proposition 4.1 A redex $u \subseteq t$ is Maranget-needed iff it is needed w.r.t. the set Π_f of all fair reductions starting from t.

Proof ($\Rightarrow$) Let u be Maranget-needed in t. Then any fair P starting from t should contract a residual of u (to erase it). ($\Leftarrow$) If u is not Maranget-needed, i.e., there is a reduction $Q : t \twoheadrightarrow e$ that discards u, then there is a reduction Q' such that $Q^* = Q + Q'$ is fair, and obviously Q^* is external to u, i.e., u is not Π_f-needed.

Proposition 4.2 A redex $u \subseteq t$ is essential iff u is needed w.r.t. the set of all fair reduction starting from t.

Proof An immediate corollary of Proposition 3.1 and Proposition 4.1.

5 A Labelling for OERSs

In Kennaway&Sleep [KeSl89] a labelling is introduced for OCRSs, based on Klop's labelling system [Klo80], which is in turn a generalization of Lévy's original labelling system for the λ-calculus [Lév78]. Each *label* of Kennaway&Sleep [KeSl89] is a tuple of labels, built up from a set of *base* labels. For any OERS R, terms in the corresponding labelled OERS R^L are those

of R where each subterm has one or more labels, represented as a *string* of labels. A labelling of a term is *initial* if all its subterms are labelled by different base labels. The *signature* of a labelled term is the tuple of all its labels, from left to right. Rules of R^L are those of R where pattern-symbols in left-hand sides are labelled by a string of labels except for the head-symbol, which has just one label (a string of length one). Each subterm (including metavariables) in the right-hand side of a rule bears the signature of the corresponding left-hand side. Further, a *redex-index* of a redex is the maximal depth of nesting in the labels of the corresponding left-hand side of the rule. The *index* $Ind(P)$ of a reduction P is the maximal redex-index of redexes contracted in it.

The crucial properties of the labelling are given by the following propositions.

Proposition 5.1 [KeSl89] If a step $t \xrightarrow{u} s$ in an OERS R creates a redex $v \subseteq s$, then, for any labelling t^l of t, the corresponding step $t^l \xrightarrow{u^{l'}} s^{l''}$ in the corresponding labelled OERS R^L creates a redex v^{l^*} whose label l^* contains the label l' of u. Thus $Ind(u^{l'}) < Ind(v^{l^*})$. If $w \subseteq s^{l''}$ is a residual of a redex $w' \subseteq t^l$, then w and w' have the same labels, thus $Ind(w) = Ind(w')$.

Proof Easy from definition of the labelling.

Corollary 5.1 [KeSl89] Let P and Q be co-initial reductions, in an OERS R, such that P creates a redex u and Q does not contract residuals of any redex of t having a residual contracted in P. Then the redexes in $u/(Q/P)$ are created by P/Q and Q/P is external to u.

Proposition 5.2 [Klo80, Lév78] Any reduction, in a labelled OERS R^L, in which only redexes with a bounded redex-index are contracted is terminating.

Proof Same as in the case of OCRSs [Klo80].

Definition 5.1 (1) For any co-initial reductions P and Q, the redex Qv in the final term of Q (read as *v with history Q*) is called a *copy* of a redex Pu if $P \trianglelefteq Q$, i.e., $P + Q/P \approx_L Q$, and v is a Q/P-residual of u; the *zig-zag* relation $\simeq_z$ is the symmetric and transitive closure of the copy relation [Lév80]. A *family* relation is an equivalence relation among redexes with histories containing the zig-zag relation.

(2) For any co-initial reductions P and Q, the redexes Qv and Pu are in the same *labelling-family* if for any initial labelling of the initial term of P and Q, they bear the same labels.

Proposition 5.1 implies that the labelling-family relation is indeed a family relation. As pointed out in [AsLa93], for OERSs in general the zig-zag and labelling family relations do not coincide. Below by family we always mean the labelling-family.

6 The Relative Normalization Theorem

In this section, we present a uniform proof of correctness of the needed strategy that works for all stable sets $\mathcal{S}$ of 'normal forms'. Our proof is different from all known proofs because properties of needed and unneeded redexes are different in the general case (the main difference is that $\mathcal{S}$-unneeded redexes may contain $\mathcal{S}$-needed ones). However, the termination argument

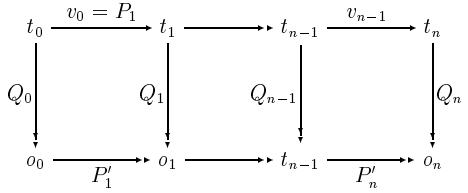


Figure 1: Preservation of External Reductions

we use is the same as in [KeSl89] and in [Mar92], and is based on Proposition 5.2. The main idea and a proof in the same spirit is already in [Lév80].

Further, we also prove a Relative Hypernormalization theorem for all *regular* stable sets of final terms, in OERSs, and demonstrate that the theorem fails for some irregular stable sets.

Below in this section $\mathcal{S}$ always denotes a stable set of terms.

Definition 6.1 We call P external to a set of redexes $F \subseteq t$ if P doesn't contract residuals of redexes in F.

Lemma 6.1 Let $P : t_0 \stackrel{v_0}{\to} t_1 \stackrel{v_1}{\to} \ldots \to t_n$ be external to $F = \{u_1, \ldots, u_n\} \subseteq t_0$, and let $Q_0 : t_0 \twoheadrightarrow o_0$. Then $P' = P/Q_0$ is external to $F' = F/Q_0$. If P is $\mathcal{S}$-normalizing, then so is P'.

Proof Let $P_i : t_0 \stackrel{v_0}{\to} t_1 \stackrel{v_1}{\to} \ldots \to t_i$, $Q_i = Q_0/P_i$, and $P'_{i+1} = v_i/Q_i$, $0 \le i < n$ (see Figure 1). Since P is external to F, we have for each i that $v_i \notin F/P_i$. Therefore, $v_i/Q_i \cap F/(P_i + Q_i) = \emptyset$ (since the residuals of different redexes are different). Thus, by Theorem 2.1, $v_i/Q_i \cap F/(Q_0 + P'_1 + \ldots + P'_i) = \emptyset$. Hence, P'_{i+1} is external to $F'/(P'_1 + \ldots + P'_i)$. This means that P' is external to F'. If P is $\mathcal{S}$-normalizing, then so is P', by the closure of $\mathcal{S}$ under parallel moves.

Corollary 6.1 For any stable $\mathcal{S}$, residuals of $\mathcal{S}$-unneeded redexes of a term t under any reduction starting from t remain $\mathcal{S}$-unneeded.

Lemma 6.2 Let $t \notin \mathcal{S}$, $t \stackrel{u}{\to} t'$, $UN_{\mathcal{S}}(u, t)$, and let $u' \subseteq t'$ be a u-new redex. Then $UN_{\mathcal{S}}(u', t')$.

Proof $UN_{\mathcal{S}}(u, t)$ implies existence of $P : t \twoheadrightarrow e$ that $\mathcal{S}$-suppresses u; thus P is external to u. By Corollary 5.1, P/u is external to u'. Also, P/u is $\mathcal{S}$-normalizing since $\mathcal{S}$ is closed under parallel moves. Hence u' is $\mathcal{S}$-unneeded.

We call $P : t_0 \to t_1 \to \ldots$ *$\mathcal{S}$-(un)needed* if it contracts only $\mathcal{S}$-(un)needed redexes.

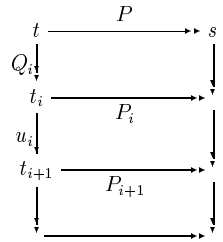
Theorem 6.1 (**Relative Normalization**) Let $\mathcal{S}$ be a stable set of terms in an OERS R.

(1) Any $\mathcal{S}$-normalizable term $t \notin \mathcal{S}$ in R contains an $\mathcal{S}$-needed redex.

(2) If $t \notin \mathcal{S}$ is $\mathcal{S}$-normalizable, then any $\mathcal{S}$-needed reduction starting from t eventually reaches a term in $\mathcal{S}$.

Proof (1) Let $P : t \twoheadrightarrow s \stackrel{u}{\to} e$ be an $\mathcal{S}$-normalizing reduction that doesn't contain terms in $\mathcal{S}$ except for e. By the stability of $\mathcal{S}$, u is $\mathcal{S}$-needed. By Corollary 6.1 and Lemma 6.2, it is either created by or is a residual of an $\mathcal{S}$-needed redex in s, and (1) follows by repeating the argument.

(2) Let $P : t \twoheadrightarrow s$ be an $\mathcal{S}$-normalizing reduction that doesn't contain terms in $\mathcal{S}$ except for e, and let $Q : t \stackrel{u_0}{\to} t_1 \stackrel{u_1}{\to} \ldots$ be an $\mathcal{S}$-needed reduction. Further, let $Q_i : t \stackrel{u_0}{\to} t_1 \stackrel{u_1}{\to} \ldots \stackrel{u_{i-1}}{\to} t_i$ and $P_i = P/Q_i$ $(i \ge 1)$. Let R^L be the corresponding labelled OERS of R, let t have an initial labelling, and assume that P and Q are reductions in R^L. By Proposition 5.1, $Ind(P_i) \le Ind(P)$. Since Q is $\mathcal{S}$-needed and P_i is $\mathcal{S}$-normalizing (by the closure of $\mathcal{S}$ under parallel moves), at least one residual of u_i is contracted in P_i. Therefore, again by Proposition 5.1, $Ind(u_i) \le Ind(P_i)$. Hence $Ind(Q) \le Ind(P)$ and Q is terminating by Proposition 5.2.



Notation Below $t{\downarrow_{\mathcal{S}}}$ will denote that t is $\mathcal{S}$-normalizable, i.e., reducible to a term in $\mathcal{S}$.

Lemma 6.3 Let $t{\downarrow_{\mathcal{S}}}$ and let $t \stackrel{u}{\to} s$. Then any $\mathcal{S}$-needed redex $v \subseteq t$ different from u has an $\mathcal{S}$-needed residual.

Proof Let $P : s \twoheadrightarrow o$ be an $\mathcal{S}$-needed $\mathcal{S}$-normalizing reduction; there is one by Theorem 6.1. Then if all u-residuals of v were $\mathcal{S}$-unneeded, P would $\mathcal{S}$-suppress them, and $u + P$ would $\mathcal{S}$-suppress v, a contradiction.

We summarize the neededness properties of residuals and created redexes, established in Corollary 6.1, Lemma 6.2, and Lemma 6.3, in the following proposition:

Proposition 6.1 (1) Residuals of $\mathcal{S}$-unneeded redexes in a term $t \notin \mathcal{S}$ remain $\mathcal{S}$-unneeded.

(2) Let $t \notin \mathcal{S}$, $t \stackrel{u}{\to} t'$, $UN_{\mathcal{S}}(u, t)$, and let $u' \subseteq t'$ be a u-new redex. Then $UN_{\mathcal{S}}(u', t')$.

(3) Let $t{\downarrow_{\mathcal{S}}}$, $t \stackrel{u}{\to} s$, $v \subseteq t$ be $\mathcal{S}$-needed, and $v \ne u$. Then v has an $\mathcal{S}$-needed residual in s.

As mentioned above, for some $\mathcal{S}$, $\mathcal{S}$-unneeded redexes may contain $\mathcal{S}$-needed ones. Take for example $R = \{f(x) \to g(x),\ a \to b\}$, and take for $\mathcal{S}$ the set of terms not containing occurrences of a. Then $\mathcal{S}$ is stable, a is $\mathcal{S}$-needed in $f(a)$, but $f(a)$ is not. Hence the following definition is justified.

Definition 6.2 We call a stable set $\mathcal{S}$ *regular* if $\mathcal{S}$-unneeded redexes cannot duplicate $\mathcal{S}$-needed ones.

Definition 6.3 We call P *$\mathcal{S}$-quasi-needed* if $NE_{\mathcal{S}}|P| = \infty$, where $NE_{\mathcal{S}}|P|$ denotes the number of $\mathcal{S}$-needed steps in P. We call P *$\mathcal{S}$-semi-needed* if it can be expressed as $P = P_1 + P_2$, where P_1 is $\mathcal{S}$-needed and P_2 is $\mathcal{S}$-unneeded. In the latter case, we call P_1 the *$\mathcal{S}$-needed part* of P, and call P_2 the *$\mathcal{S}$-unneeded part* of P.

In the following definition, we describe an algorithm that, for a *regular* stable $\mathcal{R}$ in an OERS, transforms any finite reduction P into an $\mathcal{R}$-semi-needed reduction Lévy-equivalent to P, and any $\mathcal{R}$-quasi-needed reduction Q into an infinite $\mathcal{R}$-needed reduction. The regularity of $\mathcal{R}$ is essential for termination of the transformation process, and for preservation of $\mathcal{R}$-quasi-neededness of the reduction under transformation. The latter property is crucial for our method of proving the Relative Hypernormalization theorem.

Definition 6.4 Let Red be the set of all reductions in an OERS R, let $\mathcal{R}$ be a regular stable set of terms in R, and let Red^* be the set of $\mathcal{R}$-semi-needed reductions. We define a function $K : Red \to Red^*$ as follows:

(a) Let P be a finite reduction. $K(P)$ is an $\mathcal{R}$-semi-needed reduction obtained from P in the following way: find in P the first from the left subreduction $P_0 : t \xrightarrow{u} s \xrightarrow{v} o$ such that $UN_{\mathcal{R}}(u,t)$ and $NE_{\mathcal{R}}(v,s)$. Let $P = P_1 + P_0 + P_2$. By Proposition 6.1.(2), v is a residual of a redex $v' \subseteq t$, which is $\mathcal{R}$-needed by Proposition 6.1.(1). Since $\mathcal{R}$ is regular, v is the only residual of v', hence P_0 and $P_0' = v' + u/v'$ are both complete developments of the set $u, v' \subseteq t$. Now replace P_0 by P_0' in P. Transform the obtained reduction P' in the same way, and so on, as long as possible. Obviously, the number of $\mathcal{R}$-unneeded steps in P' preceding v' is less than the number preceding v in P, and the number of $\mathcal{R}$-needed steps in P and P' coincide. Therefore, the procedure terminates eventually. The result is $K(P)$. Obviously, $P \approx_L K(P)$.

(b) Let $NE_{\mathcal{R}}|P| < \infty$ ($|P| = \infty$ is possible). P can be expressed as $P = P_1 + P_2$, where P_2 is $\mathcal{R}$-unneeded and the last step in P_1 is $\mathcal{R}$-needed. Then we take $K(P) = K(P_1) + P_2$ (this is possible since $P_1 \approx_L K(P_1)$). Obviously, $K(P) \in N^*$ and $P \approx_L K(P)$.

(c) Let $NE_{\mathcal{R}}|P| = \infty$. Suppose that P_i is the initial part of P with a length i and Q_i is the $\mathcal{R}$-needed part of $K(P_i)$. Let $|Q_i| = m_i$. By definition, $K(P)$ is the reduction whose initial part of the length m_i coincides with Q_i ($i = 0, 1, \ldots$). It follows from (a)-(b) that if $i < j$, then Q_i is an initial part of Q_j, so $K(P)$ is defined correctly. Obviously, $K(P)$ is $\mathcal{R}$-needed and hence $K(P) \in Red^*$.

Lemma 6.4 Let P be a finite or infinite reduction in an OERS, and let $\mathcal{R}$ be regular.

(1) If P ends at a term in $\mathcal{R}$, then the $\mathcal{R}$-needed part of $K(P)$ ends at a term in $\mathcal{R}$ as well.

(2) If $NE_{\mathcal{R}}|P| = \infty$, then $K(P)$ is $\mathcal{R}$-needed, and is infinite.

Proof (1) follows from the stability of $\mathcal{R}$ — $K(P)$ is $\mathcal{R}$-semi-needed, it ends at $\mathcal{R}$, and the step of $K(P)$ entering $\mathcal{R}$ is $\mathcal{R}$-needed. (2) follow immediately from Definition 6.4.

Lemma 6.5 Let t_0 have an $\mathcal{R}$-quasi-needed reduction and $t_0 \xrightarrow{u} s_0$. Then s_0 also has an $\mathcal{R}$-quasi-needed reduction.

Proof By Lemma 6.4, t_0 has an infinite $\mathcal{R}$-needed reduction $P : t_0 \xrightarrow{u_0} t_1 \xrightarrow{u_1} \ldots$. Let $F_i = u/(u_0 + \ldots + u_{i-1})$, $i = 0, 1, \ldots$. We can construct the diagram in Figure 2. It follows from

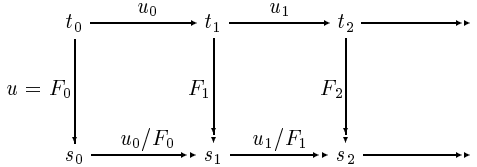


Figure 2: Preservation of Quasi-needed Reductions

finiteness of developments that there are infinitely many numbers k such that $u_k \notin F_k$ (otherwise there should be a number m such that $t_m \xrightarrow{u_m} t_{m+1} \xrightarrow{u_{m+1}} \ldots$ is an infinite F_m-development). By Proposition 6.1.(3), $u_k \notin F_k$ and $NE_{\mathcal{R}}(u_k, t_k)$ imply that u_k has at least one $\mathcal{R}$-needed F_k-residual in s_k, i.e. u_k/F_k contains at least one $\mathcal{R}$-needed step. Hence P/u is $\mathcal{R}$-quasi-needed.

Theorem 6.2 (Relative Hypernormalization) Let $\mathcal{R}$ be a regular stable set of terms in an OERS R, and let $t \notin \mathcal{R}$ be a term in R. Then t has an $\mathcal{R}$-normal form iff it does not possess a reduction in which infinitely many times $\mathcal{R}$-needed redexes are contracted.

Proof ($\Rightarrow$) Let $t \xrightarrow{P} s \in \mathcal{R}$. Suppose on the contrary that there is an $\mathcal{R}$-quasi-needed Q starting from t. Then by Lemma 6.5 Q/P is $\mathcal{R}$-quasi-needed as well. By closure of $\mathcal{R}$ under parallel moves, infinitely many terms of Q/P are in $\mathcal{R}$, and it follows from the proof of Lemma 6.5 that infinitely many of them contain $\mathcal{R}$-needed redexes – a contradiction, since terms in $\mathcal{R}$ do not contain $\mathcal{R}$-needed redexes. ($\Leftarrow$) By Theorem 6.1.(1), one can repeatedly contract $\mathcal{R}$-needed redexes in t, unless a term in $\mathcal{R}$ is reached; the latter is inevitable since t doesn't have an infinite $\mathcal{R}$-needed reduction.

The following example shows that the above Hypernormalization Theorem fails for irregular stable sets.

Example 6.1 Consider $R = \{f(x) \to h(f(x), f(x)), a \to b\}$ and take for $\mathcal{S}$ the set of terms not containing occurrences of a. Then the reduction $f(a) \to h(f(a), f(a)) \to h(f(b), f(a)) \to h(f(b), h(f(a), f(a))) \to h(f(b), h(f(b), f(a))) \to \ldots$ contracts infinitely many $\mathcal{S}$-needed redexes, while $f(a) \to f(b)$ is $\mathcal{S}$-normalizing. This example shows also that multistep $\mathcal{S}$-needed reductions need not be $\mathcal{S}$-normalizing — just omit in the above reduction the initial step and group each pair of consecutive steps as a single multistep. Recall that multistep needed reductions are normalizing in the λ-calculus [Lév80]. The same holds for all regular stable $\mathcal{R}$; this follows immediately from hypernormalization of the $\mathcal{R}$-needed strategy.

7 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, 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 $\mathcal{S}$.

Definition 7.1 (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 (it may have descendants). We call a reduction *$\mathcal{S}$-erased* if it only contracts $\mathcal{S}$-erased redexes.

Note that every persistently $\mathcal{S}$-needed redex is $\mathcal{S}$-needed, but $\mathcal{S}$-erased redexes need not be (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 TRS.

Example 7.1 Consider an orthogonal TRS $R = \{f(x) \to g(x, h(x)),\ h(x) \to c,\ a \to b\}$, 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 7.1 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.

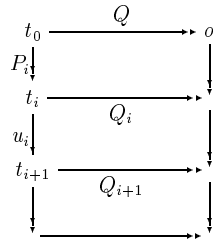
Proof ($\Rightarrow$) Let $u \subseteq t$ be persistently $\mathcal{S}$-needed, and let $P : t \twoheadrightarrow s$ be $\mathcal{S}$-normalizing. If u/P was not empty, then every $u' \in u/P$ (the set of P-residuals of u) would be $\mathcal{S}$-needed, which is not possible since $s \in \mathcal{S}$. ($\Leftarrow$) Take $R = \{f(x) \to g(x, x),\ g(x, y) \to x,\ a \to b\}$, take for $\mathcal{S}$ the set of terms not containing left-spine redexes, and consider $P : f(a) \to g(a, a) \to a \to b$. Then a in $f(a)$ is $\mathcal{S}$-needed, so is the first a in $g(a, a)$, but the second a in $g(a, a)$ is $\mathcal{S}$-unneeded (and $\mathcal{S}$-erased), and therefore a in $f(a)$ is not persistently $\mathcal{S}$-needed.

Definition 7.2 We call $P : t \twoheadrightarrow s$ *$\mathcal{S}$-minimal* if it is $\mathcal{S}$-normalizing and $P \trianglelefteq Q$ for any $\mathcal{S}$-normalizing $Q : t \twoheadrightarrow o$. When P is $\mathcal{S}$-minimal, we call s an *$\mathcal{S}$-minimal* $\mathcal{S}$-normal form of t.

It follows immediately from Definition 7.2 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 7.2 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 7.3 Let $F \subseteq t$ (i.e., F is 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*.

Definition 7.4 (1) Let $F \subseteq t$. We call a redex $u \subseteq t$ *F-unabsorbed in an F-reduction* $P : t \twoheadrightarrow e$ if $u \in F$ and none of the descendants of u along P appear in arguments of F-redexes; we call u *F-absorbed in P* otherwise. We call u *F-unabsorbed in t* if it is F-unabsorbed in any F-reduction and *F-absorbed in t* otherwise.

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

Obviously, a redex $u \subseteq t$ is unabsorbed in t, according to Definition 3.3, iff it is $F(t)$-unabsorbed, where $F(t)$ is the set of all redexes in t. By Proposition 3.2, any term t not in normal form contains an unabsorbed redex. Now, if one ignores all redexes in t except those in F, it follows that:

Lemma 7.3 Let $F \neq \emptyset$ be a set of redexes in a term t. Then F contains an F-unabsorbed redex.

Proposition 7.1 Every term t not in $\mathcal{S}$-normal form contains an ($\mathcal{S}$-needed) $\mathcal{S}$-unabsorbed redex.

Proof It is enough to take $F^n_{\mathcal{S}}(t)$ for F in Lemma 7.3; $F^n_{\mathcal{S}}(t) \neq \emptyset$ by Theorem 6.1. Note that any $\mathcal{S}$-unabsorbed redex is necessarily $\mathcal{S}$-needed.

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 7.2 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 : $P : t \to f(b)$; $Q : t \to g(a,a) \to g(b,a)$, and $N : t \to g(a,a) \to g(a,b) \to g(b,b)$. (There are two more reductions that continue Q and N, 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 7.4 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 $F^n_{\mathcal{R}}$-reduction. By Proposition 6.1.(3), it is enough to show that if a residual u' of u can be absorbed inside an $\mathcal{R}$-needed redex w', 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 absorbed inside an $\mathcal{R}$-needed redex w' such that w' can replicate 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 6.4, 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 Proposition 6.1.(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.(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 Proposition 6.1.(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 7.2. 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 $a/Q \neq \emptyset$. 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 $F^n_{\mathcal{S}}(t)$-reduction is $N : f(a) \to f(b)$, and a is $F^n_{\mathcal{S}}(t)$-unabsorbed in N.

Proposition 7.2 An $\mathcal{R}$-normalizing reduction is $\mathcal{R}$-minimal iff it is $\mathcal{R}$-erased.

Proof ($\Leftarrow$) From Lemma 7.2. ($\Rightarrow$) 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 7.4 and 7.1, $\mathcal{R}$-normalizing reduction; Q exists by Proposition 7.1. Further, let $P_i : t_0 \overset{u_0}{\to} \ldots \to t_i$ and let $Q_i = Q/P_i$ (see figure for Lemma 7.2). 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 7.2. 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.

Now we are ready to prove the main result of this section.

Theorem 7.1 (**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 7.1, 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 7.4 and Lemma 7.1. It remains to apply Theorem 6.1 and Proposition 7.2.

8 Relative Essentiality

In this section, we introduce the relative notion of essentiality, relate it to relative neededness, and give a characterization of $\mathcal{R}$-erased redexes in terms of relative neededness and essentiality.

First we generalize some properties of essential subterms [Kha93, GlKh94] to the relative case.

Definition 8.1 We call $s \subseteq t$ $\mathcal{S}$*-essential*, written $ES_{\mathcal{S}}(s,t)$, if it has a P-descendant for at least one $\mathcal{S}$-normalizing P starting from t, and call it $\mathcal{S}$*-inessential*, written $IE_{\mathcal{S}}(s,t)$, otherwise.

Lemma 8.1 (1) Let $e \subseteq s \subseteq t$ and $ES_{\mathcal{S}}(e,t)$. Then $ES_{\mathcal{S}}(s,t)$.

(2) Let s be a pattern-subterm of a redex $u \subseteq t$. Then $ES_{\mathcal{S}}(u,t)$ iff $ES_{\mathcal{S}}(s,t)$.

Proof From the definition of the descendant relation for subterms and Definition 8.1.

Lemma 8.2 Let $P : t \twoheadrightarrow o$ be $\mathcal{S}$-minimal. Then $ES_{\mathcal{S}}(s,t)$ iff s has a P-descendant.

Proof ($\Rightarrow$) Since $ES_{\mathcal{S}}(s,t)$, s has a Q-descendant for some $Q : t \twoheadrightarrow o \in \mathcal{S}$. Since P is $\mathcal{S}$-minimal, we have $P \trianglelefteq Q$, and hence $Q \approx_L P + P'$ for some P'. Therefore, by Theorem 2.1.(2), $Q \approx_{st} P + P'$, and s has a P-descendant as well. ($\Leftarrow$) is immediate.

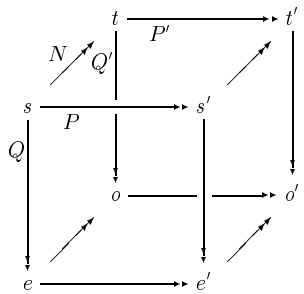
Lemma 8.3 Let $N : s \twoheadrightarrow t$, $P : s \twoheadrightarrow s'$, and $P' = P/N$. Then

(1) If P is $\mathcal{R}$-minimal, then so is P'.

(2) If P' is $\mathcal{R}$-minimal and N is $\mathcal{R}$-inessential, then P is $\mathcal{R}$-minimal as well.

Proof (1) By Proposition 7.2, P is $\mathcal{R}$-erased. Hence, so is P', and again by Proposition 7.2, P' is $\mathcal{R}$-minimal.

(2) We can construct the following diagram, where Q and Q' are $\mathcal{R}$-normalizing:



We have to show that $P \trianglelefteq Q$. We have that $N/Q = N/P = \emptyset$, since N is $\mathcal{R}$-inessential and $\mathcal{R}$ is closed under parallel moves. Hence $P/Q = (P + N/P)/Q \approx_L (N + P')/Q \approx_L N/Q + P'/Q' = P'/Q' = \emptyset$, where $Q' = Q/N$. Hence P is $\mathcal{R}$-minimal.

Lemma 8.4 (1) Descendants of $\mathcal{R}$-inessential subterms of any $s\!\downarrow_{\mathcal{R}} \notin \mathcal{R}$ remain $\mathcal{R}$-inessential.

(2) Let $N : s \twoheadrightarrow o$ be $\mathcal{R}$-inessential and let $ES_{\mathcal{R}}(s', s)$. Then s' has exactly one N-descendant $o' \subseteq o$, which is $\mathcal{R}$-essential; the $\mathcal{R}$-minimal $\mathcal{R}$-normal forms of s and o coincide.

Proof (1) Let $IE_{\mathcal{R}}(s', s)$, let $P : s \twoheadrightarrow e$ be $\mathcal{R}$-minimal, let $N : s \twoheadrightarrow o$, and let o' be an N-descendant of s'. By Lemma 8.3.(1), P/N is $\mathcal{R}$-minimal. By Lemma 8.2, s' does not have P-descendants. Further, $P + N/P \approx_{st} N + P/N$ by Theorem 2.1.(2). Hence o' does not have P/N-descendants, and $IE_{\mathcal{R}}(o', o)$ by Lemma 8.2.

(2) Let $P : s \twoheadrightarrow e$ be $\mathcal{R}$-minimal. Then by Lemma 8.2, s' has a P-descendant. Since N is $\mathcal{R}$-inessential, we have that $N/P = \emptyset$, and hence $P \approx_{st} N + P/N$ by Theorem 2.1.(2). Further, s' has exactly one N-descendant o' by Lemma 8.1.(1), and o' has P/N-descendant since o has a P-descendant. By Lemma 8.3.(1), P/N is $\mathcal{R}$-minimal. Hence o' is $\mathcal{R}$-essential by Lemma 8.2. The $\mathcal{R}$-minimal $\mathcal{R}$-normal forms of s and o coincide because $N/P = \emptyset$.

As shown above (Lemma 3.3), if $s_1, \ldots, s_n$ are inessential in s and t is obtained from s by replacing them with $t_1, \ldots, t_n$, then $t_1, \ldots, t_n$ are inessential in t. The following example shows that, surprisingly, this is not the case for every stable, even regular, set.

Example 8.1 Consider the OERS $R = \{f(x) \to a\}$ with the alphabet $\Sigma = \{f, a, x, y, \ldots\}$, and take $\mathcal{R} = \{t \,\|\, t = a \text{ or } t \text{ contains a variable}\}$. Then any $\mathcal{R}$-normalizable term not in $\mathcal{R}$-normal form has the shape $f(\ldots(f(a))\ldots)$, only the outermost f-redexes are $\mathcal{R}$-needed, and their contraction yield the $\mathcal{R}$-normal form a. Thus $\mathcal{R}$ is regular and stable. Now take $s = f(a)$ and $t = f(x)$. Then a is $\mathcal{R}$-inessential in s – it does not have descendants in the only $\mathcal{R}$-normal form a of s; but x is $\mathcal{R}$-essential in t, since $t \in \mathcal{R}$.

Next we establish the relationship between relative neededness and essentiality and prove a characterization of $\mathcal{R}$-erased redexes. We need two lemmas first.

Lemma 8.5 Every $\mathcal{R}$-unabsorbed redex $u \subseteq t\!\downarrow_{\mathcal{R}}$ is $\mathcal{R}$-essential.

Proof Let $P : t \twoheadrightarrow e$ be $\mathcal{R}$-unabsorbed $\mathcal{R}$-minimal reduction, which exists by Theorem 7.1. Since u is $\mathcal{R}$-unabsorbed, it has (exactly one) P-descendant by Definition 7.4. Hence u is $\mathcal{R}$-essential by Lemma 8.2.

Lemma 8.6 Let $P : t \twoheadrightarrow s \in \mathcal{R}$ be $\mathcal{R}$-unabsorbed. Then $IE_{\mathcal{R}}(u, t)$ iff P discards u.

Proof First note that, by Theorem 7.1, P is $\mathcal{R}$-minimal. ($\Rightarrow$) By Lemma 8.5, P is $\mathcal{R}$-essential. By Lemma 8.4.(1), all descendants of u along P remain $\mathcal{R}$-inessential. Hence P is external to u, and every descendant of u along P is its residual. Since $IE_{\mathcal{R}}(u, t)$, u does not have P-descendants by Lemma 8.2. Hence $u/P = \emptyset$, and P discards u. ($\Leftarrow$) If P discards u, then u does not have P-descendants, hence is $\mathcal{R}$-inessential by Lemma 8.2.

Proposition 8.1 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.

Proof ($\Rightarrow$) From Lemma 8.6. ($\Leftarrow$) Let $P : t \twoheadrightarrow s \in \mathcal{R}$ be $\mathcal{R}$-unabsorbed; P exists by Theorem 7.1. Since $UN_{\mathcal{R}}(u, t)$, P is external to u by Proposition 6.1.(1), and since u is $\mathcal{R}$-erased, $u/P = \emptyset$. Thus P discards u. Now it remains to apply Lemma 8.6.

Proposition 8.2 A redex $u \subseteq t \notin \mathcal{R}$ is $\mathcal{R}$-erased iff (α): every residual of u (in particular, u itself) along any reductions starting from t is either $\mathcal{R}$-needed or $\mathcal{R}$-inessential.

Proof ($\Rightarrow$) Let u be $\mathcal{R}$-erased, and suppose on the contrary that u doesn't satisfy (α), i.e., there are $P' : t \twoheadrightarrow s$ and a residual $u' \in u/P'$ such that u' is $\mathcal{R}$-unneeded and $\mathcal{R}$-essential. Let $P'' : s \twoheadrightarrow o \in \mathcal{R}$ be an $\mathcal{R}$-needed $\mathcal{R}$-minimal reduction; P'' exists by Theorem 7.1. Then P'' is external to u' by Proposition 6.1.(1), and $u'/P'' \neq \emptyset$ by Lemma 8.2, since every descendant of u' along P'' is its residual (P'' is external to u'). (Alternatively, $u'/P'' \neq \emptyset$ can be shown as follows. Since u' is $\mathcal{R}$-unneeded and $\mathcal{R}$-essential, it is not $\mathcal{R}$-erased by Proposition 8.1, and since P'' is $\mathcal{R}$-minimal, $u'/P'' \neq \emptyset$.) Thus, $u/(P' + P'') = (u/P')/P'' \supseteq u'/P'' \neq \emptyset$, i.e., u is not $\mathcal{R}$-erased in t – a contradiction. ($\Leftarrow$) We show that $u/P = \emptyset$ for any $\mathcal{R}$-normalizing reduction P by induction on $|P|$. Indeed, let $P = v + P'$. Since u satisfies (α), so does every v-residual of it (by the property (α) itself). By the induction assumption, $(u/v)/P' = \emptyset$. Hence $u/P = \emptyset$.

9 Relative Standardization Theorem

In this section, we define $\mathcal{R}$-*standard* $\mathcal{R}$-*normalizing* reductions, and show that $\mathcal{R}$-unabsorbed $\mathcal{R}$-normalizing reductions are $\mathcal{R}$-standard.

In [HuLé91], for any reduction $P : t \twoheadrightarrow s$, *external* and *outermost* redexes of P are defined; we call them P-*external* and P-*outermost*, respectively. (The Huet&Lévy concept of a P-external redex must not be confused with our earlier concept of a reduction being external to a redex, as in Definition 4.1.) It is easy to see that u is P-external iff a residual of u is contracted in P and no descendant of u is absorbed in a redex contracted in P (i.e., no descendant of u appears in an argument of a redex contracted in P); and u is P-outermost iff a residual of u is contracted in P and no residual of u is absorbed in a redex contracted in P.

In [GLM92], a reduction P is called *standard* if $P = \emptyset$ or $P = u + P'$, where u is Q-outermost for any $Q \approx_L P$, and P' is standard. The following *Standardization Theorem* is proved: For

any finite P, there is a standard Q such that $Q \approx_L P$. This result is obtained for *Abstract Reduction Systems* in which the residual relation satisfy several axioms. It is not difficult to check that the standardization theorem applies to OERSs as well (and similarly to OCRSs, as suggested already in [GLM92]).

The following lemma, relating the above concepts, was proved in [HuLé91] for OTRSs, but the proofs remain valid for OERSs too.

Lemma 9.1 ([HuLé91]) Let $P : t \twoheadrightarrow s$. A redex $u \subseteq t$ is P-external iff it is Q-outermost for any $Q \approx_L P$, and the latter holds iff u is N-outermost for a standard $N \approx_L P$. Therefore, u is P-external iff it is Q-external for any $Q \approx_L P$.

Hence a reduction P is standard if $P = \emptyset$ or $P = u + P'$, where u is P-external and P' is standard. Thus a normalizing reduction is standard iff it contracts unabsorbed redexes. One of the most important properties of a standard reduction, not difficult to prove, is that it is (persistently) needed w.r.t. the class of reductions Lévy-equivalent to it. Since $\mathcal{R}$-normalizing reductions need not be Lévy-equivalent, $\mathcal{R}$-normalizing standard reductions 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 stable and regular, $f(a) \to g(a,a) \to g(b,a) \to g(b,b)$ is a standard $\mathcal{R}$-normalizing reduction, 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. Corollary 9.2 below justifies our choice.

Definition 9.1 We call an $\mathcal{R}$-normalizing reduction *$\mathcal{R}$-standard* if it is standard and $\mathcal{R}$-minimal.

Now, we have from Theorem 7.1 and the Standardization Theorem [GLM92] the following corollary:

Corollary 9.1 (Relative Standardization Theorem) Let $\mathcal{R}$ be a regular stable set of terms in an OERS, and let $t{\downarrow_{\mathcal{R}}} \notin \mathcal{R}$. Then t has an $\mathcal{R}$-standard $\mathcal{R}$-normalizing reduction.

Lemma 9.2 Let $P : t \twoheadrightarrow s$ be $\mathcal{R}$-minimal and $u \subseteq t$ be P-external and $\mathcal{R}$-erased. Then $NE_{\mathcal{R}}(u,t)$.

Proof Obviously, u has a P-descendant, hence is $\mathcal{R}$-essential by Lemma 8.2. Therefore, by Proposition 8.2, u is $\mathcal{R}$-needed.

Lemma 9.3 Let $u \subseteq t{\downarrow_{\mathcal{R}}}$ be $\mathcal{R}$-unabsorbed and $P : t \twoheadrightarrow s$ be $\mathcal{R}$-minimal. Then u is P-external.

Proof Suppose on the contrary that u is not P-external. By Lemma 9.1, u is not N-outermost for a standard $N \approx_L P$, i.e., there is a decomposition $N = N' + v + N''$ such that a residual of u is in an argument of v. Since N is $\mathcal{R}$-erased by Proposition 7.2, it is $\mathcal{R}$-needed by Lemma 9.2, hence an $F^n_{\mathcal{R}}$-reduction – a contradiction, since u is $\mathcal{R}$-unabsorbed.

Corollary 9.2 $\mathcal{R}$-unabsorbed $\mathcal{R}$-normalizing reductions are $\mathcal{R}$-standard.

10 The Relative Optimality Theorem

In this section, we generalize Lévy's Optimality theorem to the case of all stable sets of normal forms, in OERSs.

Definition 10.1 A multistep reduction $P : t_0 \overset{F_0}{\twoheadrightarrow} t_1 \overset{F_1}{\twoheadrightarrow} \ldots \twoheadrightarrow t_n$ is called a *family-reduction* if each F_i is a set of redexes belonging to the same family ($t_i \overset{F_i}{\twoheadrightarrow} t_{i+1}$ is a complete development of F_i). $\|P\|$ will denote the number of multisteps in P. The family-reduction P is *complete* if each F_i is a complete development of a maximal set of redexes of t_i belonging to the same family. A family-reduction P is called *$\mathcal{S}$-needed* if each F_i contains at least one $\mathcal{S}$-needed redex.

Notation Below $FAM(P)$ will denote the set of families (whose member redexes are) contracted in P. $Card(FAM(P))$ will denote the number of families in $FAM(P)$.

Lemma 10.1 Every family is contracted at most once in any complete family-reduction.

Proof Let $P_n : t_0 \overset{F_0}{\twoheadrightarrow} t_1 \overset{F_1}{\twoheadrightarrow} \ldots \overset{F_{n-1}}{\twoheadrightarrow} t_n$ be a complete family-reduction. We show by induction on $n = \|P\|$ that $(a)_n$: all families contracted in P_n are different; and $(b)_n$: there is no redex in t_n whose family has been contracted in P_n. So let us assume that P_n is a labelled reduction with t_0 having an initial labelling. The case $n = 0$ is clear. Further, $(a)_n$ follows immediately from $(a)_{n-1}$ and $(b)_{n-1}$. Again by $(a)_{n-1}$ and $(b)_{n-1}$, and by completeness of P_n, all redexes in t_n that are residuals of redexes of t_{n-1} are in families that have not been contracted before. The labels of new redexes in t_n contain the label $lab(F_{n-1})$ of (redexes in the family) F_{n-1}; more precisely, the label of a new redex in t_n is a tuple that has $lab(F_{n-1})$ as a component. A redex with the same label cannot occur in $t_0, \ldots, t_{n-1}$ since t_0 has an initial labelling and, by the induction assumption, $lab(F_{n-1}) \neq lab(F_i)$, for all $i < n-1$. Thus $(b)_n$ is also valid.

Theorem 10.1 (Relative Optimality) Let $\mathcal{S}$ be a stable set of terms in an OERS R, and let $t{\downarrow_{\mathcal{S}}}$. Then any $\mathcal{S}$-needed complete family-reduction starting from t is $\mathcal{S}$-optimal in the sense that it reaches an $\mathcal{S}$-normal form of t in a minimal number of family-reduction steps.

Proof Similar to the proof of the optimality theorem for the λ-calculus in [Lév80]. Let $P : t \twoheadrightarrow s$ be an $\mathcal{S}$-normalizing family-reduction and $Q : t \overset{F_0}{\twoheadrightarrow} t_1 \overset{F_1}{\twoheadrightarrow} \ldots$ be an $\mathcal{S}$-needed complete family-reduction. Further, let $Q_i : t \overset{F_0}{\twoheadrightarrow} t_1 \overset{F_1}{\twoheadrightarrow} \ldots \twoheadrightarrow t_i$ and $P_i = P/Q_i$. Since $\simeq_z \subseteq \simeq$, we have $FAM(P_i) \subseteq FAM(P)$. Since Q is $\mathcal{S}$-needed, at least one residual of some $u_i \in F_i$ is contracted in P_i (P_i is $\mathcal{S}$-normalizing by the closure of $\mathcal{S}$ under parallel moves). Hence, again by $\simeq_z \subseteq \simeq$, $FAM(Q) \subseteq FAM(P)$. Hence, by Lemma 10.1, $\|Q\| = Card(FAM(Q)) \leq Card(FAM(P)) \leq \|P\|$.

11 Relative optimal versus minimal reductions

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 this an $\mathcal{R}$-needed *semi-complete* family-reduction. It follows from Proposition 7.2 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 example shows that a term in an OERS need not possess an $\mathcal{R}$-minimal $\mathcal{R}$-optimal family-reduction.

Example 11.1 Consider the OERS $R = \{\sigma xAB \to \delta x(((A/x)B/x)A),\ f(A) \to g(A,A)\}$, and let $\mathcal{R}$ be the set of terms not containing left-spine redexes. Then $\mathcal{R}$ is closed under unneeded expansion because for any $t \notin \mathcal{R}$ such that $t \xrightarrow{u} s \in \mathcal{R}$, u must be the outermost left-spine redex. Also, $\mathcal{R}$ is closed under reduction – no redex can be created or put on the left-spine without contracting a left-spine redex, which do not exist in terms from $\mathcal{R}$. Thus $\mathcal{R}$ is stable. $\mathcal{R}$ is moreover regular, since if there is u such that $UN_{\mathcal{R}}(u,t)$, then the reduction $P : t \twoheadrightarrow s \in \mathcal{R}$ that contracts outermost left-spine redexes is $\mathcal{R}$-needed, and is external to u, and each residual of u along P that is placed on the left-spine is discarded by contraction of a left-spine redex above it. But so do the residuals of redexes that are in u, and hence they cannot be $\mathcal{R}$-needed.

Now let us consider the term $t = \sigma x(f(x),x)$. There are two $\mathcal{R}$-needed semi-complete family-reductions starting from t: $P : t \to \delta x(f(f(x))) \twoheadrightarrow \delta x(g(g(x,x),g(x,x)))$ (since both occurrences of f in $\delta x(f(f(x)))$ are $\mathcal{R}$-needed) and $Q : t \to \sigma x(g(x,x),x) \to \delta x(g(g(x,x),g(x,x)))$, but neither reaches the $\mathcal{R}$-minimal $\mathcal{R}$-normal form $\delta x(g(g(x,x),f(x)))$ of t, obtainable by the reduction $t \to \delta x(f(f(x))) \to \delta x(g(f(x),f(x))) \to \delta x(g(g(x,x),f(x)))$.

The reason why the above counterexample works is that redexes of the same family are nested. However the following examples show that, even if there is a 'hierarchy' of nesting of redexes of different families, the term still need not have $\mathcal{R}$-minimal $\mathcal{R}$-optimal family-reductions.

Example 11.2 Consider the OTRS $R = \{f(x) \to g(x,x),\ g(b,x) \to h(x,x),\ a \to b\}$, and again take for $\mathcal{R}$ the set of terms not containing left-spine redexes. By the same argument as in Example 11.1, we can show that $\mathcal{R}$ is a regular stable set. 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 11.3 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 $(\lambda z.zvz)(\lambda z.zvz)$.

Obviously, if reductions are non-duplicating, then every $\mathcal{R}$-optimal reduction is $\mathcal{R}$-minimal too, and every $\mathcal{R}$-needed $\mathcal{R}$-normalizing reduction is both $\mathcal{R}$-minimal and $\mathcal{R}$-optimal at the same time. This suggests that there may not be conflict between minimality and optimality then graph rewriting is concerned. But this goes beyond the scope of this paper.

12 Conclusions and Future Work

We have investigated properties which a set $\mathcal{S}$ of 'final terms' or '(partial) results' should possess in order for the normalization-by-neededness theory still to make sense. We introduced appropriate notions of neededness, defined stability and regularity of sets of terms, and proved a Normalization Theorem relative to stable sets of 'normal forms', and a Hypernormalization Theorem relative to regular stable sets of 'normal forms'. We have also studied minimal and optimal 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 any $\mathcal{R}$-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 as well, and we expect that they remain valid for other systems of higher-order rewriting too.

Analogous questions arise for other strategies. For example, how can one construct reductions that avoid head-normal forms? Besides strong sequentiality for normal forms studied by Huet&Lévy in [HuLé91], strong sequentiality is studied w.r.t. head-normal forms in Kennaway [Ken94]. Investigation of relative strong sequentiality and related to it strictness analysis (see e.g., [Nök94]) seems also as an interesting topic for future research.

Study of minimal and optimal relative normalization makes sense also for infinite reductions. The starting point here is the work [BeLé79] by Berry and Lévy, where minimality and optimality of outermost complete family-reductions is shown to coincide. An infinite complete family-reduction P computing an infinite tree T is called optimal if every family contracted in P is contracted in any complete family-reduction computing T (and co-initial with P). Minimality is defined w.r.t. $\trianglelefteq$ again. For the general case, stability and regularity of sets of finite and infinite reductions must be defined first.

Acknowledgements

We thank J. R. Kennaway, J.-J. Lévy, L. Maranget, P.-A. Melliès, F. van Raamsdonk, and M. R. Sleep for useful comments and discussions. The diagrams were drawn using P. Taylor's diagram package.

References

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

[Acz78] Aczel P. A general Church-Rosser theorem. Preprint, University of Manchester, 1978.

[AEH94] Antoy S., Echahed R., Hanus M. A needed narrowing strategy. In: Proc. of the 21st ACM Symposium on Principles of Programming Languages, POPL'94, Portland, Oregon, 1994.

[AsLa93] Asperti A., Laneve C. Interaction Systems I: The theory of optimal reductions. Mathematical Structures in Computer Science, vol. 11, Cambridge Univ. Press, 1993, p. 1-48.

[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. Information and Computation, v. 75, no. 3, 1987, p. 191-231.

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

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

[DeJo90] Dershowitz N., Jouannaud J.-P. Rewrite Systems. In: J. van Leeuwen ed. Handbook of Theoretical Computer Science, Chapter 6, vol. B, 1990, p. 243-320.

[GlKh94] Glauert J.R.W., Khasidashvili Z. On Relative Normalization in Orthogonal Expression Reduction Systems. In: Proc. of the 4^{th} International workshop on Conditional (and Typed) Term Rewriting Systems, CTRS'94, Springer LNCS, vol. 968, N. Dershowitz, N. Lindenstrauss, eds. Jerusalem, 1994, pp. 144-165.

[Gar94] Gardner P. Discovering needed reductions using type theory. In: Proc. of the 2^{nd} International Symposium on Theoretical Aspects of Computer Software, TACS'94, Springer LNCS, v. 789, M. Hagiya, J. C. Mitchell, eds. Sendai, 1994, p. 555-574.

[GLM92] Gonthier G., Lévy J.-J., Melliès P.-A. An abstract Standardisation theorem. In: Proc. of the 7^{th} Annual IEEE Symposium on Logic in Computer Science, LICS'92, Santa Cruz, California, IEEE Computer Society Press, 1992, 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.

[Ken94] Kennaway J. R. A conflict between call-by-need computation and parallelism. Workshop on conditional (and typed) term rewriting systems, CTRS'94, Springer LNCS, vol. 968, N. Dershowitz, ed. Jerusalem, 1994, p. 247-261.

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

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

[Kha90] Khasidashvili Z. Expression Reduction Systems. Proceedings of I. Vekua Institute of Applied Mathematics of Tbilisi State University, vol. 36, 1990, p. 200-220.

[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 the 5^{th} International Conference on Rewriting Techniques and Applications, 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 the 19^{th} International Colloquium on Trees in Algebra and Programming, CAAP'94, Springer LNCS, vol. 787, S. Tison, ed. Edinburgh, 1994, p. 172-186.

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

[Klo92] Klop J. W. Term Rewriting Systems. In: S. Abramsky, D. Gabbay, and T. Maibaum eds. Handbook of Logic in Computer Science, vol. II, Oxford Univ. Press, 1992, p. 1-116.

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

[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, pp. 159-192.

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

[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 the 1^{st} International Conference on Typed Lambda Calculus and Applications, TLCA'93, Springer LNCS, vol. 664, Bazem M., Groote J.F., eds. Utrecht, 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.

[OR94] Van Oostrom V., van Raamsdonk F. Weak orthogonality implies confluence: the higher-order case. In: Proc. of the 3^{rd} International Conference on Logical Foundations of Computer Science, 'Logic at St. Petersburg', LFCS'94, Springer LNCS, vol. 813, Narode A., Matiyasevich Yu. V. eds. St. Petersburg, 1994. p. 379-392.

[Pkh77] Pkhakadze Sh. Some problems of the Notation Theory (in Russian). Proceedings of I. Vekua Institute of Applied Mathematics of Tbilisi State University, Tbilisi, 1977.

[Wol93] Wolfram D.A. The Clausal Theory of Types. Cambridge Tracts in Theoretical Computer Science, vol. 21, Cambridge University Press, 1993.