Stable Results and Relative Normalization

John Glauert
School of Information Systems, UEA
Norwich NR4 7TJ England
J.Glauert@uea.ac.uk

Richard Kennaway
School of Information Systems, UEA
Norwich NR4 7TJ England
R.Kennaway@uea.ac.uk

Zurab Khasidashvili
Department of Mathematics and Computer Science
Ben-Gurion University of the Negev
Beer-Sheva 84105 Israel
Zurab@cs.bgu.ac.il [*]

Abstract

We study normalization relative to sets $\mathcal{S}$ of 'results' by generalizing Huet and Lévy's theory of normalization by *neededness*. We impose on $\mathcal{S}$ natural *stability* conditions, that are necessary and sufficient for each term not in $\mathcal{S}$ to have at least one $\mathcal{S}$-needed redex, and for repeated contraction of $\mathcal{S}$-needed redexes in a term t to lead to a term in $\mathcal{S}$ (a $\mathcal{S}$-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. This reveals a conflict between minimality and optimality: for regular stable sets of normal forms, a term need not possess a reduction that is minimal and optimal

---

[*] Part of this work was supported by the Engineering and Physical Sciences Research Council of Great Britain under grant GR/H 41300.

at the same time. All these results are obtained in the framework of orthogonal Expression Reduction Systems, a formalism for rewriting subsuming term rewriting and the λ-calculus.

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 [CF58].

Normalization by Needed Reduction

For Orthogonal Term Rewriting Systems (OTRSs), a general normalizing strategy, called the *needed* strategy, was found by Huet and Lévy in [HL91]. The needed strategy always contracts a *needed* redex – a redex with at least one *residual* which is contracted in any reduction to normal form. Huet and Lévy show 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 define the class of *strongly sequential* OTRSs where a needed redex can efficiently be found in any term.

Extending the concept of Neededness

Barendregt et al. [BKKS87] generalize the concept of neededness to the λ-calculus. They study 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. They prove correctness of the two needed strategies for computing normal forms and head-normal forms, respectively. Middeldorp [Mid97] studies normalization w.r.t. root-stable terms (which are terms that cannot be rewritten to a redex). Normalization w.r.t. another interesting set of 'normal forms', that of constructor head-normal forms in constructor OTRSs, is studied by Nöcker [Nök94].

The normalization by neededness theory has been extended in other directions too, of which we mention a few. Khasidashvili independently defined a similar normalizing strategy, called the *essential* strategy, for the λ-calculus [Kha88], OTRSs [Kha93], and *Expression Reduction Systems* (OERSs) [Kha94]. 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. Hence, essentiality makes sense for all subterms, not only for redexes. In [Mar92], Maranget introduces 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.

Sekar and Ramakrishnan [SR93] study a normalizing strategy which in each multistep contracts a *necessary* set of redexes. Khasidashvili [Kha94] shows that in Higher Order Recursive Program Schemes one can find all needed redexes in any term, implying decidability of weak and strong normalization. Khasidashvili and Piperno [KP99] designed an algorithm for statically finding all inessential subterms (i.e., the garbage) in simply typable λ-terms. 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. Kennaway et al. [KKSV95] studied needed strategies for infinitary OTRSs. Boudol [Bou85] and Melliès [Mel96] extended the needed strategy to non-orthogonal rewrite systems. A different approach to normalization is developed by Kennaway [Ken89] and by Antoy and Middeldorp [AM94].

The concept of Relative Neededness

Since in the practice of Functional Programming one is interested in terms of different shapes (normal forms, head-normal forms, etc.) as the results, it is natural to ask what properties a set of terms must possess in order for the neededness theory of Huet and Lévy still to make sense. Here we provide a comprehensive solution to that question.

We introduce the notion of *neededness* w.r.t. a set of reductions Π so that each existing notion of neededness can be given by specifying Π. Usually it is convenient to consider a set of terms $\mathcal{S}$, and the induced set of reductions $\Pi_{\mathcal{S}}$ ending at a term in $\mathcal{S}$. For example, *Huet and Lévy-neededness* [HL91] is neededness w.r.t. the set NF of normal forms, *Maranget-neededness* [Mar92] is neededness w.r.t. all fair reductions, *head-neededness* [BKKS87] is neededness w.r.t. the set of head-normal forms, *root-neededness* [Mid97] is neededness w.r.t. the set of *root-stable* forms, etc.

We impose a natural condition on $\mathcal{S}$, called *stability*, which is 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 $\mathcal{S}$.[1]

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 contracting both $\mathcal{S}$-needed and $\mathcal{S}$-unneeded redexes which never reaches a term in $\mathcal{S}$ even though $\mathcal{S}$-needed redexes are contracted infinitely many times. Therefore, *multistep $\mathcal{S}$-needed* reductions, where every multistep contracts at least one $\mathcal{S}$-needed redex, need not be $\mathcal{S}$-normalizing. This is because a 'non-standard' situation may arise in what we will call *irregular* stable sets $\mathcal{S}$, where $\mathcal{S}$-unneeded redexes may contain $\mathcal{S}$-needed ones. However, for regular stable sets $\mathcal{S}$, the $\mathcal{S}$-needed strategy is always hypernormalizing, and the multistep $\mathcal{S}$-needed strategy is normalizing.

Minimal and Optimal Relative Normalization

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 a *minimal* $\mathcal{S}$-normal form.

Minimal $\mathcal{S}$-normal forms are useful to compute since any other $\mathcal{S}$-normal form is accessible from the 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*) trees, respectively. These trees represent the values of the term according to different semantics –

---

[1] When the rewrite system is non-duplicating and non-erasing, our concept of stability coincides with the concept of stability in event structures [Win89].

Böhm semantics and 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 minimality result we are aware of was obtained by Berry and Lévy in [BL79], 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.

Here 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 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. We show also that $\mathcal{S}$-minimal reductions need not exist if $\mathcal{S}$ is stable but is not regular.

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. That is, we consider multistep reductions contracting a number of redexes in the same *family* in parallel, and consider optimality w.r.t. the number of such multisteps. This is because Barendregt et al [BBKV76] showed that no one-step optimal recursive β-reduction strategy exists for the λ-calculus. The generalization is rather straightforward, since a suitable labelling system is available, allowing us to define a family relation in OERSs. We show that complete $\mathcal{S}$-needed family-reductions, which contract all members of a family containing an $\mathcal{S}$-needed redex in a multistep, are optimal.

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 not $\mathcal{S}$-erased. 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.

Overview

In the next section, we review Expression Reduction Systems [Kha90, Kha92]. In section 3, we introduce the relative notion of neededness. In section 4, we sketch some properties of our labelling system for OERSs needed to define a family relation among redexes. We prove correctness of the $\mathcal{S}$-needed strategy for computing terms of $\mathcal{S}$, for all stable $\mathcal{S}$, in section 5, and prove hypernormalization of the $\mathcal{S}$-needed strategy w.r.t. regular stable sets $\mathcal{S}$ in section 6. In section 7, we study $\mathcal{S}$-minimal reductions for regular stable sets $\mathcal{S}$. In section 8 we establish a Relative Standardization Theorem. In section 9, we prove the Relative Optimality Theorem. Finally, in section 10, we relate relative optimal and minimal reductions. The conclusions appear in section 11.

The concepts introduced in Section 3 are taken from [GK94], but the definitions are different in that the new definitions use only the residual concept while the original definitions are based on a concept of descendants for subterms and components. Our labelling system in Section 4 is based on an unpublished work by the second author [KS89]. The proofs in Sections 5 and 6 simplify analogous proofs published as [GK94], and the remaining results are published in [GK96a]. The results were first reported in [GK94a].

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 [Kha92, Wol93, Nip93, OR94]. We refer to van Raamsdonk [Raa96] for a survey.

Expression Reduction Systems

Here we use *Expression Reduction Systems* (ERSs), defined in [Kha92] (under the name of CRSs). The present formulation follows [KOvO99] and is simpler.

Definition 2.1 Let Σ be an *alphabet* comprising *variables* $x, y, z, \ldots$; *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 $x_1 \ldots x_m$, have 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 by using functions and quantifiers in the usual way: variables are terms, and if $t_1, \ldots, t_n$ are terms and δ is an n-ary (simple or compound) operator, then $\delta(t_1, \ldots, t_n)$ is a term too.

Metaterms are constructed similarly from *terms* and *metavariables* $A, B, \ldots$ that range over terms, but with an extra operation: if $t_0, \ldots, t_n$ are metaterms, then so is $(t_1/x_1, \ldots, t_n/x_n)t_0$, also called a *metasubstitution*, 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 right-to-left 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. The substitution operation may involve a *renaming* of bound variables to avoid collision, and we assume that the set of variables in Σ comes equipped with an equivalence relation, called renaming, such that any equivalence class of variables is infinite. We also assume that any variable can be renamed by any other variable in the corresponding equivalence class.[2] Unless otherwise specified, the default renaming relation is the total binary relation on variables (a partial renaming relation may be useful for conditional systems).

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

Definition 2.2 A *Conditional Expression Reduction System* (CERS) is a pair (Σ, R), where Σ is an *alphabet* described in Definition 2.1 and R is a

---

[2] An equivalence class of variables can, for example, be the set of variables of the same type in a typed language.

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.

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

[vcf] 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 so is any *variant* of $t\theta$ obtained from it by renaming of bound variables), and $s\theta$ is the *contractum* of $t\theta$.

If for any rule $r \in R$, $AA(r)$ is the maximal set of variable-capture free assignments, then the CERS is called an *unconditional* Expression Reduction System, or simply an Expression Reduction System (ERS).

Below we restrict ourselves to ERSs. We ignore questions relating to renaming of bound variables. As usual, a rewrite step consists of replacement of a redex by its contractum. Subterms of a redex corresponding to metavariables are *arguments* of the redex, and the rest is its *pattern*. Note that the use of metavariables in rewrite rules of ERSs is not really necessary – free variables can be used instead, as in TRSs (since (free) variables in TRS rules play the role of metavariables in ERS rules). We will indeed do so at least when giving TRS examples.

Examples

Our syntax is similar to that of Klop's CRSs [Klo80], but is simpler in that it is closer to the syntax of the λ-calculus. For example, the β-rule is written as $Ap(\lambda xA, B) \to (B/x)A$, 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 $x \not\in (A\theta)$, 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 μ-recursor rule is written as $\mu(\lambda xA) \to (\mu(\lambda xA)/x)A$. $\exists xA \to (\tau x(A)/x)A$ and $\exists !xA \to \exists xA \wedge \forall x\forall y(A \wedge (y/x)A \Rightarrow x = y)$ are rules corresponding to familiar definitions.

Notation 2.3 We use a, b, c, d for constants, t, s, e, o for terms, u, v, w for redexes, and N, P, Q for reductions. We write $s \subseteq t$ if s is a subterm of t. A one-step reduction contracting a redex $u \subseteq t$ is written as $t \xrightarrow{u} s$ or $t \to s$ or just u. We write $P : t \twoheadrightarrow s$ if P denotes a reduction of t to s. $P+Q$ denotes the concatenation of P and Q. We write $U \subseteq t$ if U is a set of redexes in t.

Orthogonality, Residuals, and Lévy-equivalence

The definition of *orthogonality* in ERSs is similar to the case of CRSs: all the rules are left-linear and in no term redex-patterns can overlap [Klo80]. The *residual* relation on redexes in ERSs is defined as a combination of the residual relations in TRSs and the λ-calculus, since any ERS step can be decomposed into a TRSs step followed by a number of substitution steps. Since the residual concept is so familiar both in TRSs and the λ-calculus we do not reintroduce the concept here for ERSs, and instead refer to [Kha92, KOvO99] for more details. *Developments* of sets of redexes $U \subseteq t$ are defined in ERSs as usual [Bar84].

As in the case of the λ-calculus [Bar84] or OCRSs, for any co-initial reductions P and Q, one can define in OERSs the notion of *residual of P under Q*, written P/Q, using Klop's method of commutative diagrams [Klo80]. Klop's method is equivalent to Lévy's original definition of the residual relation in the λ-calculus [Lév78, Lév80]; the latter is of more 'algebraic' nature and uses multisteps rather than complete developments.

We write $P \trianglelefteq Q$ if $P/Q = \emptyset$, where $\trianglelefteq$ is the *Lévy-embedding* relation. P and Q are called *Lévy-equivalent*[3], written $P \approx_L Q$, if $P \trianglelefteq Q$ and $Q \trianglelefteq P$. It follows from the definition of / that if P and Q are co-initial reductions in an OERS, then $(P + P')/Q = P/Q + P'/(Q/P)$ and $P/(Q + Q') = (P/Q)/Q'$.

We will often be interested in residuals of single redexes written u/P where u is a single-step reduction contracting u, a redex in the initial term of P. For example, $P \approx_L Q$ for finite P and Q implies that P and Q end at the same term and that $u/P = u/Q$ for all redexes, u, in the initial term.

The following *strong Church-Rosser (confluence)* property is established for OERSs in [Kha92, KOvO99]; the Finite Developments Theorem [Bar84] is proved first, from which strong confluence follows by a standard argument. Strong confluence for other higher-order rewriting formats are obtained, among others, in [Klo80, Nip93, KOR93, OR94, Oos94, Raa96, Mel96].

Theorem 2.4 (Strong Church-Rosser) For any co-initial reductions P and Q in an OERS, $P + Q/P \approx_L Q + P/Q$.

---

[3] or *strongly-equivalent*, or *permutation-equivalent*

3 Stability and 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 $\mathcal{S}$; all existing notions of neededness can be obtained by specifying Π or $\mathcal{S}$; $\mathcal{S}$-neededness is 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 5 to show that if $\mathcal{S}$ is stable, then a $\mathcal{S}$-needed strategy is $\mathcal{S}$-normalizing.

Relative Neededness

Definition 3.1

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

(2) We call $P : t \twoheadrightarrow o$ *external* to a set of redexes $U \subseteq t$ if P is external to all $u \in U$.

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

(4) We say that P *erases* u if $u/P = \emptyset$. Note that contracting a redex erases it.

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

Definition 3.2

(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 of its residuals 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. Let $\Pi_{\mathcal{S}}$ be the set of all reductions that end at a term in $\mathcal{S}$. These will be denoted *$\mathcal{S}$-normalizing* reductions. We call a redex u *$\mathcal{S}$-needed*, written $NE_{\mathcal{S}}(u,t)$, if it is $\Pi_{\mathcal{S}}$-needed, and call it $\mathcal{S}$-unneeded, written $UN_{\mathcal{S}}(u,t)$, otherwise.

Thus Huet and Lévy neededness coincides with neededness w.r.t. the set of normal forms. It is easy to see that Maranget's notion of neededness [Mar92], where a redex is needed if it has a residual along any reduction external to it, coincides with neededness w.r.t. the set of fair reductions [GK94a]. (Recall that a reduction is fair if all redexes in any of its terms are erased in it.) Obviously, any redex in a term that is not $\mathcal{S}$-normalizable is $\mathcal{S}$-needed; we call such redexes *trivially $\mathcal{S}$-needed.*

Stability

We will state the definition of the property of sets of terms for which it is possible to generalise the Normalization Theorem and will then give some examples showing why some intuitively simpler definitions are inadequate:

Definition 3.3 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 does not contain terms in $\mathcal{S}$, the final term of P/Q is in $\mathcal{S}$; and

(b) $\mathcal{S}$ is *closed under unneeded expansion*: for any $e \xrightarrow{u} o$ such that $e \notin \mathcal{S}$ and $o \in \mathcal{S}$, u is $\mathcal{S}$-needed.

The most appealing examples of stable sets, for an OERS, are the set of normal forms [HL91], the set of head-normal forms [BKKS87], the set of weak-head-normal forms, the set of constructor-head-normal forms for constructor TRSs [Nök94], and the set of root-stable forms (which are terms that cannot be rewritten to a redex) [Mid97]. All these sets are closed under reduction, which implies closure under parallel moves.

Closure under Parallel Moves is not sufficient

The reduction graph G_s of any 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, for $R = \{I(x) \to x\}$, 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 redex. 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 Relative Normalization Theorem.

Closure under Reduction is not necessary

Note that 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.

$$\begin{tikzcd}
f(a) \arrow[rr] \arrow[dd] & & \mathbf{g}(\mathbf{a},\mathbf{a}) \arrow[dl] \arrow[dr] & \\
& g(a,b) \arrow[dr] & & g(b,a) \arrow[dl] \\
f(b) \arrow[rr] & & \mathbf{g}(\mathbf{b},\mathbf{b}) &
\end{tikzcd}$$

Actually, a stable set need not be closed under reduction even if the rewrite system is non-duplicating (i.e., has a non-duplicating residual relation). For example, take $R = \{a \to b,\ b \to a\}$, and take $\mathcal{S} = \{a\}$. Then $\mathcal{S}$ is closed under unneeded expansion: the only step entering $\mathcal{S}$ is $b \to a$, which is $\mathcal{S}$-needed; and is trivially closed under parallel moves, but $\mathcal{S}$ is not closed under reduction.

Closure under Normalization is not sufficient: Closure under Parallel Moves is necessary

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 $R = \{f(x) \to g(x,x),\ a \to b\}$ as before, but 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.

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 3.4 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 (i.e., variable-free) term, only the terms in the following diagram are $\mathcal{S}$-normalizable.

$$\begin{tikzcd}
f(a) \arrow[rr] \arrow[dd, leftrightarrow] & & g(a,a) \arrow[dl, leftrightarrow] \arrow[dr, leftrightarrow] & \\
& \mathbf{g(a,b)} \arrow[dr, leftrightarrow] & & g(b,a) \arrow[dl, leftrightarrow] \\
f(b) \arrow[rr] & & g(b,b) &
\end{tikzcd}$$

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.

$$\begin{tikzcd}
f(a) \arrow[r] \arrow[d] & g(a,a) \arrow[r] \arrow[d] & \mathbf{g(a,b)} \arrow[d] \\
f(b) \arrow[r] & g(b,b) \arrow[r, two heads, "\emptyset"] & g(b,b)
\end{tikzcd}$$

4 A Labelling for OERSs

We now introduce a labelling system for ERSs which will be used to establish termination of certain reductions and to define the concept of a redex *family*.

Fix some OERS R. For technical reasons, we assume that R does not contain any rules whose left-hand side consists of just an operator applied to variables and metavariables. This is not a substantial restriction, since if the system includes a rule of the form $F(x, y, z) \to \ldots$, we can replace the symbol $F(\ldots)$ wherever it occurs by $G(F(\ldots))$, where G is a new unary function symbol. The new system will behave in all essential respects like the old and will satisfy the restriction.

Let there be a set $\mathcal{L}_0$ of *atomic labels*, and a set $\mathcal{C}$ of *label constructors* in $1-1$ correspondence with the set of subterms, other than metasubstitutions, of all the right-hand sides of the rules of the OERS. The arity of each constructor is the number of nodes in the left hand side of the rule which it is associated with, excluding variable and metavariable nodes and the root. The set $\mathcal{L}$ of labels is the set of terms freely generated by $\mathcal{L}_0$ and $\mathcal{C}$. The *height* of a label is its height considered as a tree.

We add all members of $\mathcal{L}$ to the OERS as new unary function symbols. By a *labelled term* we mean any term which results from this extension of the signature. Every labelled term t determines an unlabelled term $U(t)$ by simply dropping the labels — that is, by replacing every subterm of the form $l(t')$ by t', where l is any label. A labelling of a term is *initial* if all its subterms are labelled by different atomic labels.

Although for technical reasons we have introduced labels as new function symbols, in writing terms down we shall indicate labels by superscripts. A term $\alpha(\beta(\gamma(t)))$ would be written $t^{\gamma\beta\alpha}$. This suggests an alternative way of looking at labels, closer to the way Lévy and Klop use them: the labels are considered to be annotations attached to the nodes of the syntax tree of the unlabelled term.

We now construct the set of *labelled rules*. Given any rule $s \to t$ of the unlabelled system, and labellings f of s and f' of t, $s^f \to t^{f'}$ is a labelled rule provided that:

- f does not label the root of s, nor any occurrence of a variable or metavariable.
- f' attaches to each subterm of t, other than metasubstitutions, the label $c(l_1, \ldots, l_n)$, where c is the label constructor corresponding to that subterm, and $l_1, \ldots, l_n$ are the labels present in $f(s)$, listed in depth-first left-to-right order. (This tuple of labels is called the *signature* of $f(s)$.)

The choice of depth-first left-to-right ordering is not significant for the theory; it is merely a convenient standard choice. Note that f' is completely determined by f, s, and t. All nodes of $t^{f'}$ have distinct labels, and different labellings of s give different labellings of t. The reason for the restriction we made earlier is to ensure that every left-hand side has at least one place to attach a label to.

The set of labelled rules constitutes the system $\mathcal{L}(R)$. This is clearly an OERS.

As an example, consider the beta rule of lambda calculus. In OERS notation, and writing the application operator explicitly, it is:

$$Ap((\lambda x(A)), B) \to (B/x)A$$

An example of a redex is

$$Ap((\lambda y(Ap(y,y))),(\lambda z(z)))$$

A labelled version of this redex is:

$$(Ap((\lambda y(Ap(y^\alpha,y^\beta))^\gamma)^\delta,(\lambda z.z^\epsilon)^\zeta))^\eta$$

This contains a redex of the labelled rule:

$$Ap((\lambda x(A))^\delta,B)\to(B^{c(\delta)}/x)A^{d(\delta)}$$

and reduces to:

$$((Ap(((\lambda z(z^\epsilon))^\zeta)^{c(\delta)},((\lambda z(z^\epsilon))^\zeta)^{c(\delta)}))^{d(\delta)})^\eta$$

The *index* $Ind(u)$ of a redex u is the maximal height of the labels of the left-hand side of the corresponding rule. In the above example this depth is 1, since the only label in the left-hand side is δ. If the left-hand side included a label $a(b(\pi),\sigma)$, the depth would be 3. The *index* $Ind(P)$ of a reduction P is the maximal redex-index of the redexes contracted in it. For a redex u in a labelled term, we define $lab(u)$ to be the tuple of labels in the left-hand side of the labelled rule for u. We call this the label of the redex (although strictly speaking it is a tuple of labels).

This definition of labelling differs slightly from that of Klop [Klo80]. Klop combines multiple labels on a single term into a single compound label, and thus for his definitions a term t^α is not a subterm of $t^{\alpha\beta}$, whereas for us the term t^α is a subterm of $(t^\alpha)^\beta$. He also represents compound labels as flat strings, instead of structured terms, using multiple underlining where we use multiple nesting. In addition, our use of label constructor symbols provides a greater generality which allows us to simultaneously treat several variants of labelled rewrite systems uniformly. However, our definitions are sufficiently close that Klop's results for labelled systems carry over to the present setting.

The crucial properties of labelled reduction are given by the following propositions.

Proposition 4.1 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 4.2 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 4.3 Any reduction, in a labelled OERS R^L, in which only redexes with a bounded redex-index are contracted is terminating.
Proof See [Klo80, Mel96, Oos97].

5 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 differs 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 replicate $\mathcal{S}$-needed ones. However, the termination argument we use is the same as in [KS89] and in [Mar92], and is based on Proposition 4.3. The main idea and a proof in the same spirit is already in [Lév80]. As in most earlier proofs, we use the fact that residuals of unneeded redexes remain unneeded, and that unneeded steps cannot create needed redexes. However, in order to prove that every needed redex has at least one needed residual after contacting any other redex, we need first to show the existence of a needed normalizing reduction.

In the rest of this section $\mathcal{S}$ always denotes a stable set of terms.

Lemma 5.1 Let $P : t_0 \xrightarrow{v_0} t_1 \xrightarrow{v_1} \ldots \rightarrow t_n$ be external to $U = \{u_1, \ldots, u_n\} \subseteq t_0$, and let $Q_0 : t_0 \twoheadrightarrow o_0$. Then $P' = P/Q_0$ is external to $U' = U/Q_0$. If P is $\mathcal{S}$-normalizing, then so is P'.
Proof Let $P_i : t_0 \xrightarrow{v_0} t_1 \xrightarrow{v_1} \ldots \rightarrow t_i$, $Q_i = Q_0/P_i$, and $P'_{i+1} = v_i/Q_i$, $0 \leq i < n$ (see the figure below). Since P is external to U, we have for each i that $v_i \notin U/P_i$. Therefore, $v_i/Q_i \cap U/(P_i + Q_i) = \emptyset$ (since the residuals of different redexes are different). Thus, by Theorem 2.4, $v_i/Q_i \cap U/(Q_0 + P'_1 + \ldots + P'_i) = \emptyset$. Hence, P'_{i+1} is external to $U'/(P'_1 + \ldots + P'_i)$. This means that P' is external to U'. If P is $\mathcal{S}$-normalizing, then so is P', by the closure of $\mathcal{S}$ under parallel moves.

$$\begin{tikzcd}
t_0 \arrow[r, "v_0 = P_1"] \arrow[d, two heads, "Q_0"'] & t_1 \arrow[r, two heads] \arrow[d, two heads, "Q_1"'] & t_{n-1} \arrow[r, "v_{n-1}"] \arrow[d, two heads, "Q_{n-1}"'] & t_n \arrow[d, two heads, "Q_n"] \\
o_0 \arrow[r, two heads, "P'_1"'] & o_1 \arrow[r, two heads] & t_{n-1} \arrow[r, two heads, "P'_n"'] & o_n
\end{tikzcd}$$

Corollary 5.2 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 5.3 Let $t \notin \mathcal{S}$, $t \xrightarrow{u} t'$, $UN_{\mathcal{S}}(u,t)$, and let $u' \subseteq t'$ be a u-new redex. Then $UN_{\mathcal{S}}(u',t')$.

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 4.2, 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.

Relative Normalization

Theorem 5.4 (**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 $P : t \twoheadrightarrow s \xrightarrow{u} e$ be an $\mathcal{S}$-normalizing reduction that does not contain terms in $\mathcal{S}$ except for e. By the stability of $\mathcal{S}$, u is $\mathcal{S}$-needed. By Corollary 5.2 and Lemma 5.3, 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 does not contain terms in $\mathcal{S}$ except for e, and let $Q : t \xrightarrow{u_0} t_1 \xrightarrow{u_1} \ldots$ be an $\mathcal{S}$-needed reduction. Further, let $Q_i : t \xrightarrow{u_0} t_1 \xrightarrow{u_1} \ldots \xrightarrow{u_{i-1}} t_i$ and $P_i = P/Q_i$, $i \geq 1$ (see the figure). 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 4.1, $Ind(P_i) \leq Ind(P)$. 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 4.1, $Ind(u_i) \leq Ind(P_i)$. Hence $Ind(Q) \leq Ind(P)$ and Q is terminating by Proposition 4.3.

$$\begin{tikzcd}
t \arrow[r, two heads, "P"] \arrow[d, "Q_i"'] & s \arrow[d] \\
t_i \arrow[r, two heads, "P_i"'] \arrow[d, "u_i"'] & {} \arrow[d] \\
t_{i+1} \arrow[r, two heads, "P_{i+1}"'] \arrow[d] & {} \arrow[d] \\
{} \arrow[r, two heads] & {}
\end{tikzcd}$$

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

Lemma 5.5 Let $t\downarrow_{\mathcal{S}}$ and let $t\xrightarrow{u}s$. 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 5.4. 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 conclude this section by summarizing the neededness properties of residuals and created redexes, established in Corollary 5.2, Lemma 5.3, and Lemma 5.5, in the following proposition.

Proposition 5.6

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

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

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

6 The Relative Hypernormalization Theorem

In this section, we prove the Relative Hypernormalization Theorem for all *regular* stable sets of final terms in OERSs, and demonstrate that the theorem fails for some irregular stable sets. Thus this theorem refines the Relative Normalization Theorem, but is less general as it is restricted to regular stable sets of results only. However, regularity is not a restriction from a practical point of view, as all the sets of results that have been used in practical programming languages are regular. Moreover, it allows for much simpler proofs. We do not need to consider termination of reductions with bounded index (Proposition 4.3), as the Finite Developments Theorem is enough. We do not use the Standardization Theorem [HL91, Klo80] either.

Failure of Relative Hypernormalization for some stable sets

For the case of normal forms, the needed strategy is *hypernormalizing*, that is, reductions starting from a normalizable term that contract finite sequences of unneeded steps in addition to needed redexes are still normalizing.

The following example shows that this need not be the case for all stable sets $\mathcal{S}$:

Example 6.1 Consider $R = \{f(x) \to h(x, f(x)),\ a \to b\}$ and take for $\mathcal{S}$ the set of terms not containing occurrences of a. Then the reduction $f(a) \to h(a, f(a)) \to h(b, f(a)) \to h(b, h(a, f(a))) \to h(b, h(b, f(a))) \to \ldots$ contracts infinitely many $\mathcal{S}$-needed redexes, while the reduction $f(a) \to f(b)$ is $\mathcal{S}$-normalizing. This example shows also that multistep $\mathcal{S}$-needed reductions need not be $\mathcal{S}$-normalizing: group each pair of consecutive steps in the reduction above as a multistep.

Regular Stable Sets

For some stable $\mathcal{S}$, $\mathcal{S}$-unneeded redexes may contain $\mathcal{S}$-needed ones. Consider the simpler example $R = \{f(x) \to g(x),\ a \to b\}$, where $\mathcal{S}$ is the set of terms not containing occurrences of a: we observe that a is $\mathcal{S}$-needed in $f(a)$, but $f(a)$ is not.

Nesting of $\mathcal{S}$-needed redexes within $\mathcal{S}$-unneeded ones need not cause problems; the failure of hypernormalization arises from duplication of $\mathcal{S}$-needed redexes by $\mathcal{S}$-unneeded ones. Hence we introduce the following definition:

Definition 6.2 We call a stable set $\mathcal{R}$ *regular* if, for any $t \not\in \mathcal{R}$, $\mathcal{R}$-unneeded redexes cannot duplicate $\mathcal{R}$-needed ones.

Recall from [Lév80] that multistep needed reductions are normalizing in the λ-calculus. The same holds for all regular stable $\mathcal{R}$; this follows immediately from hypernormalization of the $\mathcal{R}$-needed strategy, which we will prove in the rest of the section.

Canonical forms for Relative Reductions

The proof uses the ability to transform any reduction into an equivalent form where $\mathcal{R}$-needed steps precede $\mathcal{R}$-unneeded steps, while conserving the number of $\mathcal{R}$-needed steps.

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 $P = P_N + P_U$, where P_N is $\mathcal{S}$-needed and P_U is $\mathcal{S}$-unneeded. In the latter case, we call P_N the *$\mathcal{S}$-needed part* of P, and call P_U 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 will define a function $K : Red \to Red^*$. For any reduction P, with finite $\mathcal{R}$-needed steps, we will denote the $\mathcal{R}$-needed part $K_N(P)$ and the $\mathcal{R}$-unneeded part $K_U(P)$ so that $K(P) = K_N(P) + K_U(P)$. K is defined as follows:

(a) Let P be a finite reduction. $K(P)$ is the $\mathcal{R}$-semi-needed reduction obtained from P as follows: find in P the leftmost subreduction $P_1 : t \xrightarrow{u} s \xrightarrow{v} o$ such that $UN_{\mathcal{R}}(u, t)$ and $NE_{\mathcal{R}}(v, s)$. Let $P = P_0 + P_1 + P_2$. By Proposition 5.6.(2), v is a residual of a redex $v' \subseteq t$, which is $\mathcal{R}$-needed by Proposition 5.6.(1). Since $\mathcal{R}$ is regular, v is the only residual of v', hence P_1 and $P_1' = v' + u/v'$ are both complete developments of the set $u, v' \subseteq t$. Now replace P_1 by P_1' 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, $P \approx_L K(P)$, and $K(P) \in \mathit{Red}^*$.

(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 $K(P) = K(P_1) + P_2$, which is possible since $P_1 \approx_L K(P_1)$. Obviously, $K(P) \in \mathit{Red}^*$ 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 $Q_i = K_N(P_i)$. Let $|Q_i| = m_i$. We define $K(P)$ as the reduction whose prefix of length m_i is given by Q_i ($i = 0, 1, \ldots$). It follows from (a)-(b) that if $i < j$, then Q_i is a prefix of Q_j, so $K(P)$ is defined consistently. Obviously, $K(P)$ is $\mathcal{R}$-needed and hence $K(P) \in \mathit{Red}^*$.

Lemma 6.5 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 $K_N(P)$ 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) follows immediately from Definition 6.4.

Lemma 6.6 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 6.5, t_0 has an infinite $\mathcal{R}$-needed reduction $P : t_0 \xrightarrow{u_0} t_1 \xrightarrow{u_1} \ldots$. Let $U_i = u/(u_0 + \ldots + u_{i-1})$, $i = 0, 1, \ldots$. We can construct the diagram below. It follows from finiteness of developments that there are infinitely many numbers k such that $u_k \not\in U_k$ (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 U_m-development). By Proposition 5.6.(3), $u_k \not\in U_k$ and $NE_{\mathcal{R}}(u_k, t_k)$ imply that u_k has at least one $\mathcal{R}$-needed U_k-residual in s_k, i.e. u_k/U_k contains at least one $\mathcal{R}$-needed step. Hence P/u is $\mathcal{R}$-quasi-needed.

$$\begin{tikzcd}
t_0 \arrow[r, "u_0"] \arrow[d, "u = U_0"'] & t_1 \arrow[r, "u_1"] \arrow[d, two heads, "U_1"'] & t_2 \arrow[r, two heads] \arrow[d, two heads, "U_2"'] & {} \\
s_0 \arrow[r, two heads, "u_0/U_0"'] & s_1 \arrow[r, two heads, "u_1/U_1"'] & s_2 \arrow[r, two heads] & {}
\end{tikzcd}$$

Relative Hypernormalization

Theorem 6.7 (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 $t \overset{P}{\twoheadrightarrow} s \in \mathcal{R}$. Suppose on the contrary that there is an $\mathcal{R}$-quasi-needed Q starting from t. Then by Lemma 6.6 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 6.6 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 5.4.(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 does not have an infinite $\mathcal{R}$-needed reduction.

7 Minimal Relative Normalization

The theory of *minimal* reduction in the framework of relative normalization is based on reducing no more redexes than necessary when computing $\mathcal{S}$-normal forms. A reduction $P : t \twoheadrightarrow s$ with $t \notin \mathcal{S}$ and $s \in \mathcal{S}$ is $\mathcal{S}$-minimal if $P \trianglelefteq Q$ for all $\mathcal{S}$-normalizing $Q : t \twoheadrightarrow e$. Minimal $\mathcal{S}$-normal forms, such as s above, are useful to compute since any other $\mathcal{S}$-normal form is accessible from them.

We define *$\mathcal{S}$-unabsorbed*, *persistently $\mathcal{S}$-needed*, and *$\mathcal{S}$-erased* redexes. We show that each class is a strict subset of the next when $\mathcal{S}$ is regular, and use such redexes to build $\mathcal{S}$-minimal reductions for regular $\mathcal{S}$. We show that $\mathcal{S}$-minimal reductions need not exist for irregular stable $\mathcal{S}$.

Persistently $\mathcal{S}$-needed and $\mathcal{S}$-erased redexes

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 does not have a residual under any $\mathcal{S}$-normalizing reduction.

(3) We call a reduction *$\mathcal{S}$-erased* if it only contracts $\mathcal{S}$-erased redexes.

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

Example 7.2 Consider an OTRS $R = \{f(x) \to g(x, h(x)),\ h(x) \to c,\ a \to b\}$ and a term (redex) $u = f(a)$:

$$\begin{tikzcd}
f(a) \arrow[rr, "f"] \arrow[dd, "a"'] & & g(a,h(a)) \arrow[rr, "h"] \arrow[dl, "a"'] \arrow[dr, "a"] & & g(a,c) \arrow[dd, "a"] \\
& g(b,h(a)) \arrow[drrr, "h"] \arrow[dr, "a"'] & & g(a,h(b)) \arrow[ur, "h"] \arrow[dl, "a"] & \\
f(b) \arrow[rr, "f"'] & & g(b,h(b)) \arrow[rr, "h"'] & & g(b,c)
\end{tikzcd}$$

Consider the following sets of terms in R: the set $\mathcal{S}_1 = \{g(b,c)\}$ of normal forms; the set $\mathcal{S}_2 = \{g(b,h(a)), g(b,h(b)), g(b,c)\}$ of terms not containing a redex on the left-spine, i.e. not containing a redex with its head symbol on the left-spine, when the term is considered as a tree; the set $\mathcal{S}_3 = \{f(b), g(b,h(b)), g(b,c)\}$ of terms not containing occurrences of a; and the set $\mathcal{S}_4 = \{g(a,h(b)), g(a,c), f(b), g(b,h(b)), g(b,c)\}$ 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 $g(a,h(a))$ is $\mathcal{S}_1$-unneeded. Nevertheless, 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. a is not $\mathcal{S}_2$-erased as it 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 7.3 Every persistently $\mathcal{S}$-needed redex is $\mathcal{S}$-erased, but an $\mathcal{S}$-erased redex, even if $\mathcal{S}$-needed, need not be persistently $\mathcal{S}$-needed.

Proof ($\Rightarrow$) Let $u \subseteq t$ be persistently $\mathcal{S}$-needed, and let $P : t \twoheadrightarrow s$ be $\mathcal{S}$-normalizing. If u/P was not empty, then every $u' \in u/P$ (the set of P-residuals of u) would be $\mathcal{S}$-needed, which is not possible since $s \in \mathcal{S}$. ($\Leftarrow$) From Example 7.2 (cases 1 and 3).

Minimal Relative Reduction

Definition 7.4 We call $P : t \twoheadrightarrow s$ *$\mathcal{S}$-minimal*[4] 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 a *minimal* $\mathcal{S}$-normal form of t.

It follows immediately from Definition 7.4 that if $t\!\downarrow_{\mathcal{S}}$ and $t \notin \mathcal{S}$, then t has no more than one 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 minimal $\mathcal{S}$-normal forms cannot be taken as the definition, because in that case an $\mathcal{S}$-normalizable term could have many 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 regular stable set $\mathcal{S} = \{a, b\}$. The term $t = f(a)$ has two $\mathcal{S}$-normal forms, a and b, and each is accessible from the other. However, any reduction from t to b must contract the $\mathcal{S}$-unneeded redex a and therefore no reduction from t to b can be considered as $\mathcal{S}$-minimal.

Lemma 7.5 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.

Existence of Minimal $\mathcal{S}$-normal forms

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 7.6 Consider $R = \{f(x) \to g(x, x),\ a \to b\}$, with term $t = f(a)$, and let $\mathcal{S} = \{g(b, a), f(b), g(b, b)\}$, the set of terms not containing a on the

---

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

left spine.

$$\begin{tikzcd}
f(a) \arrow[rr] \arrow[dd] & & g(a,a) \arrow[dl] \arrow[dr] & \\
& g(a,b) \arrow[dr] & & \mathbf{g}(\mathbf{b},\mathbf{a}) \arrow[dl] \\
\mathbf{f}(\mathbf{b}) \arrow[rr] & & \mathbf{g}(\mathbf{b},\mathbf{b}) &
\end{tikzcd}$$

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, but duplicates the innermost one which *is* $\mathcal{S}$-needed.

There are three candidates for $\mathcal{S}$-minimal reductions starting from t: $P : f(a) \to f(b)$ and $Q : f(a) \to g(a,a) \to g(b,a)$ and $N : f(a) \to g(a,a) \to g(a,b) \to g(b,b)$. There are two more reductions that continue P and Q, but they clearly cannot be $\mathcal{S}$-minimal. We have $P \not\trianglelefteq Q$ as $P/Q = g(b,a) \to g(b,b)$, $Q \not\trianglelefteq P$ as $Q/P = f(b) \to g(b,b)$, and $N \not\trianglelefteq P$ as, again, $N/P = f(b) \to g(b,b)$. Hence none of the reductions is $\mathcal{S}$-minimal.

$\mathcal{S}$-unabsorbed redexes

We go on to 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. However, $\mathcal{S}$-erased reductions need not be $\mathcal{S}$-needed, and hence need not be $\mathcal{S}$-normalizing, and again for regular $\mathcal{S}$, we show existence of $\mathcal{S}$-unabsorbed $\mathcal{S}$-normalizing reductions, which are $\mathcal{S}$-needed $\mathcal{S}$-minimal reductions.

Definition 7.7 Let $U \subseteq t$. We call P an *U-reduction* if it contracts only residuals of redexes from U and created redexes; we call such redexes *U-redexes*. Below $U(t)$ will denote the set of all redexes of t, and $U_{\mathcal{S}}(t)$ will denote the set of $\mathcal{S}$-needed redexes of t.

Definition 7.8

(1) Let $U \subseteq t$. We call a redex $u \subseteq t$ *U-unabsorbed* (in t) if $u \in U$ and, for any U-reduction P, none of the residuals of u along P appear in

arguments of U-redexes; we call u *U-absorbed* in t if $u \in U$ and it is not U-unabsorbed.

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

Example 7.9 Consider an OTRS $R = \{a \to c,\ b \to b',\ f(c,x) \to c'\}$, and take a term $t = g(f(a,b),a)$. Then both occurrences of a in t are $U(t)$-unabsorbed in t, while b is $U(t)$-absorbed in t: we have $t \to g(f(c,b),a) = s$, and the residual of b in s is in an argument of the created redex $f(c,b)$. If $U \subseteq t$ contains two redexes – the first occurrence of a in t and the redex $b \subseteq t$, then only the first $a \subseteq t$ is U-unabsorbed in t. If the set of terms not having a left-spine redex is taken for $\mathcal{S}$, then the first a is the only $\mathcal{S}$-unabsorbed redex in t (it is the only $\mathcal{S}$-needed redex too).

It is shown in [HL91, Kha93, GK94] that any term t not in normal form contains an $U(t)$-unabsorbed redex (such redexes are called *external* in [HL91]). Now, if one ignores all redexes in t except those in $U \subseteq t$, it follows that, for any $U \neq \emptyset$, U contains an U-unabsorbed redex. And by taking $U_{\mathcal{S}}(t)$ for U ($U_{\mathcal{S}}(t) \neq \emptyset$ by Theorem 5.4), we obtain the following proposition:

Proposition 7.10 Every term $t \notin \mathcal{S}$ contains an ($\mathcal{S}$-needed) $\mathcal{S}$-unabsorbed redex.

Lemma 7.11 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 $U_{\mathcal{R}}(t)$-reduction. By Proposition 5.6.(3), it is enough to show that if a residual u' of u can appear inside an $\mathcal{R}$-needed redex $w' \neq u'$, then w' cannot replicate u'; therefore u has at most one residual in any term of P. Suppose, on the contrary, that there is $P : t \twoheadrightarrow o$ such that a residual u' of u is inside an $\mathcal{R}$-needed redex w' such that w' replicates u'; and assume that P is a shortest such a reduction, i.e., u has exactly one residual in every term in P. By Lemma 6.5, 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 5.6.(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 2.4,

$w'' + P''/w''$ replicates u'', since w' replicates u'. Thus P''/w'' replicates u^* – a contradiction, since P''/w'' is $\mathcal{R}$-unneeded by Proposition 5.6.(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 7.6. Then a in t is $\mathcal{S}$-needed, so is its leftmost residual in $g(a, a)$, but the rightmost residual is $\mathcal{S}$-unneeded, and $a/Q = \emptyset$. Hence $a \subseteq t$ is not persistently $\mathcal{S}$-needed or $\mathcal{S}$-erased. But $a \subseteq t$ is $\mathcal{S}$-unabsorbed, since the only $U_{\mathcal{S}}(t)$-reduction is $N : f(a) \to f(b)$, and a is $U_{\mathcal{S}}(t)$-unabsorbed in N.

Characterizing $\mathcal{R}$-minimal reductions

Proposition 7.12 An $\mathcal{R}$-normalizing reduction is $\mathcal{R}$-minimal iff it is $\mathcal{R}$-erased.
Proof ($\Leftarrow$) From Lemma 7.5. ($\Rightarrow$) Let $P : t_0 \xrightarrow{u_0} t_1 \to \ldots \to t_n$ be $\mathcal{R}$-minimal, and let $Q : t_0 \twoheadrightarrow o$ be $\mathcal{R}$-unabsorbed, hence $\mathcal{R}$-erased by Lemmas 7.11 and 7.3, $\mathcal{R}$-normalizing reduction; Q exists by Proposition 7.10. Further, let $P_i : t_0 \xrightarrow{u_0} \ldots \to t_i$ and let $Q_i = Q/P_i$. Since Q is $\mathcal{R}$-erased, so is Q_i, and Q_i is $\mathcal{R}$-normalizing by the closure of $\mathcal{R}$ under parallel moves. Hence Q_i is $\mathcal{R}$-minimal by Lemma 7.5. 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.

Minimal Relative Normalization

Theorem 7.13 (**Minimal Relative Normalization**) Let $\mathcal{R}$ be a regular stable set of terms in an OERS, and let $t \downarrow_{\mathcal{R}}$ where $t \notin \mathcal{R}$. 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 $t \downarrow_{\mathcal{R}}$ where $t \notin \mathcal{R}$ has an $\mathcal{R}$-unabsorbed $\mathcal{R}$-minimal reduction, which is $\mathcal{R}$-needed.
Proof By Proposition 7.10, any $t \downarrow_{\mathcal{R}}$ where $t \notin \mathcal{R}$ has an $\mathcal{R}$-unabsorbed redex, which is $\mathcal{R}$-needed and $\mathcal{R}$-erased by Lemma 7.11 and Lemma 7.3. It remains to apply Theorem 5.4 and Proposition 7.12.

8 The Relative Standardization Theorem

Recall that a reduction is *standard* if redexes in it are contracted in left-to-right outside-in order [Bar84, HL91, Klo80]. Maranget proved in [Mar92] that standard reductions are minimal among reductions computing a 'stable prefix' of a term, in an OTRS. Note that for regular $\mathcal{R}$, in general, $\mathcal{R}$-normalizing standard reductions need not be $\mathcal{R}$-needed, according to the definitions of [Bar84, HL91, Klo80], or according to [GLM92], where left-to-right order of contracted redexes is not required.

Take for example $R = \{f(x) \to g(x,x), a \to b\}$, and take for $\mathcal{R}$ the set of terms not containing a redex on the right-spine; then $\mathcal{R}$ is regular, $f(a) \to g(a,a) \to g(b,a) \to g(b,b)$ is standard and $\mathcal{R}$-normalizing, but the second step is $\mathcal{R}$-unneeded and the final term is not a minimal $\mathcal{R}$-normal form.

Therefore, we should require that relative standard reductions are $\mathcal{R}$-minimal as in the following definition:

Definition 8.1 We call an $\mathcal{R}$-normalizing reduction $\mathcal{R}$*-standard* if it is outside-in and $\mathcal{R}$-minimal.

It is not difficult to check that $\mathcal{R}$-unabsorbed $\mathcal{R}$-normalizing reductions are then $\mathcal{R}$-standard (see [GK94a]), and the left-to-right order of contraction of $\mathcal{R}$-unabsorbed redexes can also be arranged. Hence we have the following Relative Standardization Theorem:

Relative Standardization

Theorem 8.2 (Relative Standardization) Let $\mathcal{R}$ be a regular stable set of terms in an OERS, and let $t{\downarrow_{\mathcal{R}}}$ where $t \notin \mathcal{R}$. Then t has an $\mathcal{R}$-standard $\mathcal{R}$-normalizing reduction. In particular, $\mathcal{R}$-unabsorbed $\mathcal{R}$-normalizing reductions are $\mathcal{R}$-standard.

9 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. The family concept we use is based on the labelling system of section 4. Other labelling systems for for orthogonal (first and higher order) rewrite systems are studied in [Mar92, AL93, Oos96].

Definition 9.1 For co-initial reductions $P : t \twoheadrightarrow s$ and $Q : t \twoheadrightarrow o$, redexes $u \in s$ and $v \in o$ with *histories* P and Q, written Pu and Qv, are in the same *(labelling-)family* if for any initial labelling of t, they bear the same labels.

Definition 9.2 A multistep reduction $P : t_0 \overset{U_0}{\twoheadrightarrow} t_1 \overset{U_1}{\twoheadrightarrow} \dots \twoheadrightarrow t_n$ is called a *family-reduction* if each U_i is a set of redexes belonging to the same family ($t_i \overset{U_i}{\twoheadrightarrow} t_{i+1}$ is a complete development of U_i). $\|P\|$ will denote the number of multisteps in P. The family-reduction P is *complete* if each U_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 U_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 9.3 Every family is contracted at most once in any complete family-reduction.

Proof Let $P_n : t_0 \overset{U_0}{\twoheadrightarrow} t_1 \overset{U_1}{\twoheadrightarrow} \dots \overset{U_{n-1}}{\twoheadrightarrow} t_n$ 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 label of any new redex in t_n must contain at least one of $lab(U_{n-1})$ as the arguments to a label constructor. A redex with the same label cannot occur in $t_0, \dots, t_{n-1}$ since t_0 has an initial labelling and, by the induction assumption, $lab(U_{n-1}) \neq lab(U_i)$, for all $i < n - 1$. Thus $(b)_n$ is also valid.

Relative Optimality

Theorem 9.4 (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 :

$t \overset{U_0}{\twoheadrightarrow} t_1 \overset{U_1}{\to} \dots$ be an $\mathcal{S}$-needed complete family-reduction. Further, let $Q_i : t \overset{U_0}{\twoheadrightarrow} t_1 \overset{U_1}{\to} \dots \twoheadrightarrow t_i$ and $P_i = P/Q_i$. By Proposition 4.1, $FAM(P_i) \subseteq FAM(P)$. Since Q is $\mathcal{S}$-needed, at least one residual of some $u_i \in U_i$ is contracted in P_i (P_i is $\mathcal{S}$-normalizing by the closure of $\mathcal{S}$ under parallel moves). Hence, again by Proposition 4.1, $FAM(Q) \subseteq FAM(P)$. Hence, by Lemma 9.3, $\|Q\| = Card(FAM(Q)) \leq Card(FAM(P)) \leq \|P\|$.

10 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 7.12 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, consider $R = \{g(x) \to f(x,x),\ a \to b\}$, where $\mathcal{R}$ is the set of terms not containing left-spine redexes. Then $g(a) \to f(a,a) \to f(b,a)$ is both $\mathcal{R}$-minimal and a $\mathcal{R}$-optimal semi-complete family-reduction, whereas complete family reductions would lead to $f(b,b)$ which is not the minimal $\mathcal{R}$-normal form.

However, the following example shows that a term in an OERS need not possess an $\mathcal{R}$-minimal $\mathcal{R}$-optimal semi-complete family-reduction.

Example 10.1 Consider the OERS $R = \{\sigma x A B \to \delta x(((A/x)B/x)A),\ f(A) \to g(A,A)\}$, 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 $t \overset{u}{\to} s \in \mathcal{R}$, 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

$$Q : t \rightarrow \sigma x(g(x,x),x) \rightarrow \delta x(g(g(x,x),g(x,x))),$$

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 $t \rightarrow \delta x(f(f(x))) \rightarrow \delta x(g(f(x),f(x))) \rightarrow \delta x(g(g(x,x),f(x)))$.

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 10.2 Consider the OTRS $R = \{f(x) \rightarrow g(x,x),\ g(b,x) \rightarrow h(x,x),\ a \rightarrow b\}$, and again take for $\mathcal{R}$ the set of terms not containing left-spine redexes (e.g. $\{h(b,a), h(b,b)\}$).

$$\begin{tikzcd}
f(a) \arrow[rr] \arrow[dd] & & g(a,a) \arrow[dl] \arrow[dr] & & \\
& g(a,b) \arrow[dr] & & g(b,a) \arrow[r] \arrow[dl] & h(a,a) \arrow[dl] \arrow[d] \\
f(b) \arrow[rr] & & g(b,b) \arrow[dr] & \mathbf{h}(\mathbf{b},\mathbf{a}) \arrow[d] & h(a,b) \arrow[dl] \\
& & & \mathbf{h}(\mathbf{b},\mathbf{b}) &
\end{tikzcd}$$

By the same argument as in Example 10.1, we can show that $\mathcal{R}$ is a regular stable set. Now $P : f(a) \rightarrow g(a,a) \rightarrow g(b,a) \rightarrow h(a,a) \rightarrow h(b,a)$ 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 $f(a) \rightarrow g(a,a)$ 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 10.3 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.

11 Conclusions

We have investigated properties which a set $\mathcal{S}$ of 'final terms' or '(partial) results' should posses 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 showed 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 and for context-sensitive conditional orthogonal ERSs [KOvO99], and therefore apply to numerous typed λ-calculi as well. Actually, the normalization and optimality results have already been generalized to abstract orthogonal rewrite systems with axiomatized residual relation, more precisely, to *stable Deterministic Residual Structures (SDRSs)* and to *Deterministic*

Family Structures, respectively [GK96]; the latter structure is a refinement of the former in that it formalizes the concept of redex-family. And the above minimality results hold in such SDRSs where every set U of redexes in a term has a *non-duplicable* element – one that has at most one residual under any U-reduction (for example, U-unabsorbed redexes in OERSs are non-duplicable w.r.t. U). Since this work was first presented, Melliès [Mel98] has also established minimality results for his axiomatic rewrite systems.

Acknowledgements

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

References

[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 21^{st} ACM Symposium on Principles of Programming Languages, POPL'94, 1994, p. 268-279.

[AM94] Antoy S. and Middeldorp A. A Sequential Reduction Strategy. In Proc. of the 4^{th} International Conference on Algebraic and Logic Programming, ALP'94, Springer LNCS, vol. 850, p. 168-185, 1994.

[AL93] 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.

[BBKV76] Barendregt H.P., Bergstra J., Klop J.W., Volken H. Degrees, reductions and representability in the λ-calculus. Report 22, University of Utrecht, 1976.

[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.

[BL79] Berry G., Lévy J.-J. Minimal and optimal computations of recursive programs. JACM 26, 1979, p. 148-175.

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

[CF58] Curry H. B., Feys R. Combinatory Logic. vol. 1, North-Holland, 1958.

[DJ90] 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.

[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. 1994, p. 555-574.

[GK94] 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. 1994, p. 144-165.

[GK94a] Glauert J.R.W., Khasidashvili Z. Minimal and optimal relative normalization in Expression Reduction Systems. Report SYS-C94-06, UEA, Norwich, 1994.

[GK96] Glauert J.R.W., Khasidashvili Z. Relative normalization in deterministic residual structures. In: Proc. of the 19^{th} International Colloquium on Trees in Algebra and Programming, CAAP'96, Springer LNCS, vol. 1059, H. Kirchner, ed. 1996, p. 180-195.

[GK96a] Glauert J.R.W, Khasidashvili Z. Minimal relative normalization in orthogonal expression reduction systems. In Proc. of the 16^{th} International Conference on Foundations of Software Technology and Theoretical Computer Science, FST&TCS'96, Springer LNCS, vol. 1180, V. Chandru, ed. 1996, p. 238-249.

[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, IEEE Computer Society Press, 1992, p. 72-81.

[HL91] Huet G., Lévy J.-J. Computations in Orthogonal Rewriting Systems. In: Computational Logic, Essays in Honor of Alan Robinson, J.-L. Lassez and G. Plotkin, eds. MIT Press, 1991.

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

[KS89] 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.

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

[Kha88] Khasidashvili Z. β-reductions and β-developments of λ-terms with the least number of steps. In: Proc. of the International Conference on Computer Logic COLOG'88, 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.

[KOvO99] Khasidashvili Z., Ogawa M., van Oostrom V. Perpetuality and uniform normalization in orthogonal rewrite systems. Submitted to Inf.& Comp. Available from http://www.brl.ntt.co.jp/people/mizuhito/papers/TRS.html.

[KP99] Khasidashvili Z., Piperno A. Normalization of typable terms by superdevelopments. In. Proc. of CSL'99, Springer LNCS, vol. 1584, 1999, p. 260-282.

[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, TCS vol. 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 vol. 2, 1976, p. 97-114.

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

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

[Lon83] Longo G. Set theoretic models of lambda calculus: theories, expansions and isomorphisms. Annals of Pure and Applied Logic, vol. 24, 1983, p. 53-188.

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

[Mel96] Melliès P.-A. Description Abstraite des Systèmes de Réécriture. Thèse de l'Université Paris 7, 1996.

[Mel98] Melliès P.-A. A stability theorem for in rewriting theory. In. Proc. of LICS'98, 1998.

[Mid97] Middeldorp A. Call by need computations to root-stable form. In Proc. of the 24th ACM Symposium on Principles of Programming Languages, POPL'97, 1997, p. 94-105.

[Nip93] Nipkow T. Orthogonal higher-order rewrite systems are confluent. In: Proc. of the 1st International Conference on Typed Lambda Calculus and Applications, TLCA'93, Springer LNCS, vol. 664, Bezem 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, Amsterdam, 1994.

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

[Oos96] Van Oostrom V. Higher order families. In proc. of the 7th International Conference on Rewriting Techniques and Applications, RTA'96, Springer LNCS, vol. 1103, Ganzinger, H., ed. 1996, p. 392–407.

[Oos97] Van Oostrom V. Finite family developments. In proc. of the 8th International Conference on Rewriting Techniques and Applications, RTA'97, Springer LNCS, vol. 1232, Common H., ed. 1997, p. 308-322.

[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.

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

[SR93] Sekar R.C., Ramakrishnan I.V. Programming in Equational Logic: Beyond Strong Sequentiality. Information and Computation, vol. 104, n. 1, 1993, p. 78-109.

[Win89] G. Winskel, An introduction to event structures, in: Proc. Linear time, branching time and partial order in logics and models of concurrency, Springer LNCS, vol. 354, 1989, p. 364-397.

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