

Delimiting diagrams

Vincent van Oostrom*

Abstract

We propose a way in which a strategy may *exceed* another one. In particular, we say that a strategy *uniformly* exceeds another one, if every maximal reduction for the former is as least as long as *every* reduction for the latter. For instance, we show that the strategy F_∞ uniformly exceeds any other strategy for the ordinary λ -calculus. In order to prove these results we introduce the notion of a *delimiting* diagram. Its usefulness is shown by presenting simple proofs not only of some classical results, but also of some new ones such as maximality of F_∞ for the λ -calculus with explicit substitutions $\lambda\mathbf{x}^-$.

1 Delimiting diagrams

We assume familiarity with the basics of rewriting. Our notions and notations are from [5], For the convenience of the reader, we have recapitulated some fundamental notions from that book in Appendix A.

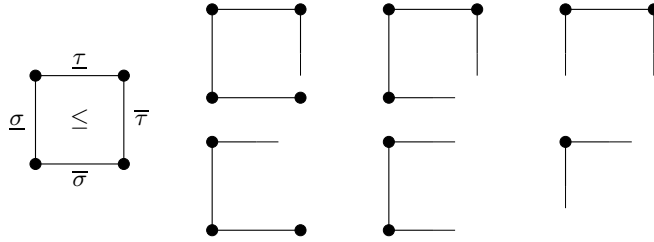
Throughout, we assume that the abstract rewrite systems considered in a given context have the same objects and normal forms. We employ $\rightarrow, \twoheadrightarrow, \mapsto$ to range over abstract rewrite systems and use σ, τ, v to denote reductions. A reduction σ may be finite or infinite. That is, its length $|\sigma|$ is either some natural number or ω . Two reductions are *co-initial* if they have the same source, *cofinal* if they have the same target, and *coextensive* if they are co-initial, and cofinal or either is infinite. The *composition* $\sigma \cdot \tau$ is only defined if σ has a target (so σ is finite) which is the source of τ . A reduction σ is a *prefix* of itself and of any $\sigma \cdot \tau$, and τ is a *suffix* of the latter reduction. We employ underlining/overlining to indicate a prefix/suffix of a reduction, implicitly assuming that whenever both a prefix $\underline{\sigma}$ and suffix $\overline{\sigma}$ of σ are specified then $\sigma = \underline{\sigma} \cdot \overline{\sigma}$. A reduction is *maximal* if it is a prefix only of itself.

Definition. A *diagram* D (for \rightarrow and \twoheadrightarrow) is a quadruple $(\underline{\sigma}, \sigma, \underline{\tau}, \tau)$, such that σ and τ are coextensive, $\underline{\sigma}$ is a \rightarrow -reduction which is a prefix of σ such that the induced suffix $\overline{\sigma}$ (if any) of σ is a \twoheadrightarrow -reduction, and vice versa $\underline{\tau}$ is a \twoheadrightarrow -reduction which is a prefix of τ such that the induced suffix $\overline{\tau}$ (if any) of τ is a \rightarrow -reduction.

*address: Universiteit Utrecht, Department of Philosophy, Heidelberglaan 8, 3584 CS Utrecht, The Netherlands. e-mail: Vincent.vanOostrom@phil.uu.nl

The diagram D is *empty/maximal* if all reductions are so, *trivial* if either prefix is empty, *local* if both prefixes consist of exactly one step, and *delimiting* if $|\sigma| \leq |\tau|$,

Note that a maximal delimiting diagram is of the shape $(\sigma, \sigma, \tau, \tau)$, hence we simply say the *pair* σ, τ is delimiting. Unlike the diagrams often found in the literature our diagrams can be both *partial* in that they may ‘lack’ some sides, and *infinite* in that sides may not have an end-point. The following figure illustrates abstractly that a delimiting diagram $D = (\underline{\sigma}, \sigma, \underline{\tau}, \tau)$ may have, apart from the ‘standard’ shape on the left, six other shapes varying in partiality and infinity.



Here edges denote reductions and vertices indicate sources/targets, as usual. Thus σ is the path from the top-left vertex going down and then right, and τ first goes right and then down. By the diagrams being delimiting, if the former path does not end in a vertex, i.e. if σ is infinite, then the latter does not end in a vertex either, i.e. τ is infinite as well.

Definition. Delimiting diagrams $D_i = (\underline{\sigma}_i, \sigma_i, \underline{\tau}_i, \tau_i)$ for $i \in \{1, 2\}$, are composed *horizontally* and *vertically* by ‘pasting’ them: E.g. if the suffix $\overline{\tau}_1$ of τ_1 exists and coincides with the prefix $\underline{\sigma}_2$ of σ_2 , then the *horizontal* composition $D_1|D_2$ of D_1 and D_2 is the diagram $(\underline{\sigma}_1, \sigma_1 \cdot \overline{\sigma}_2, \underline{\tau}_1 \cdot \underline{\tau}_2, \overline{\tau}_1 \cdot \tau_2)$ if $\overline{\sigma}_2$ exists, and $(\underline{\sigma}_1, \sigma_1, \underline{\tau}_1 \cdot \underline{\tau}_2, \overline{\tau}_1 \cdot \tau_2)$ if $\overline{\sigma}_2$ does not exist. The *vertical* composition $\frac{D_1}{D_2}$ of D_1 and D_2 is defined analogously.

Proposition. • Delimiting diagrams are closed under both compositions.

- Both compositions are associative.
- The exchange law $\frac{D_{11}}{D_{21}}| \frac{D_{12}}{D_{22}} = \frac{D_{11}|D_{12}}{D_{21}|D_{22}}$ holds, if all compositions are defined.

Proof. By routinely pasting diagrams. □

2 Uniformly exceeds

Definition. \rightarrow *uniformly exceeds* \rightarrow , denoted by $\rightarrow \gg \rightarrow$, if each pair of co-initial maximal \rightarrow, \rightarrow -reductions is delimiting.

This is a strong requirement as shown by the following examples.

Example. 1. The non-confluent ARS $b \leftarrow a \rightarrow c$ does not uniformly exceed itself as the pair of co-initial maximal reductions $a \rightarrow b, a \rightarrow c$ is not delimiting since the reductions are not co-extensive (cofinal).

2. The confluent ARS $a \xrightarrow{\curvearrowright} b \xrightarrow{\curvearrowleft} c$ does not uniformly exceed itself as the pair of co-initial maximal reductions $a \rightarrow b \rightarrow c, a \rightarrow c$ is not delimiting since the latter is not as long as the former (despite being cofinal).

The problem in the first example is that the ARS does not have unique normal forms. The problem in the second example is that the ARS is not ... uniform. It is easy to see that deterministic ARSs are uniform in the sense that on them \gg is reflexive.

Lemma. \gg is transitive.

Proof. Suppose $\mapsto \gg \rightarrow \gg \rightarrow$ and let σ, v be a co-initial maximal pair of \rightarrow, \mapsto -reductions. Choose any maximal \rightarrow -reduction τ co-initial to σ and v . By assumption both the pairs σ, τ and τ, v are delimiting, so $|\sigma| \leq |\tau| \leq |v|$. Moreover, the only way in which σ, v could fail to be coextensive is if both σ, v were finite but τ not, which is impossible since $|\tau| \leq |v|$, and we conclude. \square

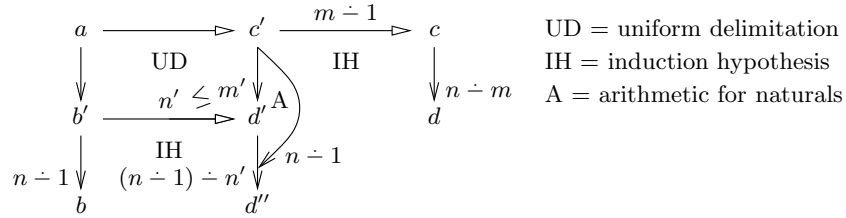
Proving that one ARS uniformly exceeds another one can be localised.

Definition. We say that \rightarrow *uniformly delimits* \mapsto if any peak $b \leftarrow_{\phi} a \rightarrow_{\psi} c$ can be completed to a delimiting diagram, i.e. $(\phi, \phi \cdot \sigma, \psi, \psi \cdot \tau)$ is a delimiting diagram for some \rightarrow -reduction σ and \mapsto -reduction τ .

Note that $|\sigma| \leq |\tau|$ holds for σ, τ as in the definition.

Lemma (Progress). If \rightarrow uniformly delimits \mapsto , and $b \leftarrow^n a \rightarrow^m c$, then $d \leftarrow^{n-m} c$ for some d .

Proof. By induction on n .¹ In case $n \leq m$, setting $d = c$ works, and in case $m = 0$, $d = b$ does. Supposing $n > m \geq 1$ (see the figure below), there exist b' and c' such that $b \leftarrow^{n-1} b' \leftarrow a \rightarrow c' \rightarrow^{m-1} c$. By uniform delimitation, either there is an infinite \rightarrow -reduction from c' or there are $n' \leq m'$, d' such that $b' \rightarrow^{n'} d' \leftarrow^{m'} c'$. We claim that in either case there is a \rightarrow -reduction of length $n - 1$ from c' . If the claim holds true, then by the IH for that reduction and $c' \rightarrow^{m-1} c$, there is a d such that $d \leftarrow^{(n-1)-(m-1)} c$ and we conclude.



¹M. Bezem has shown that formalising the induction step in Geometric Logic [1] its proof is automatable.

To prove the claim note that it holds trivially if c' allows an infinite \rightarrow -reduction. Otherwise (see the figure above), by the IH for $b \leftarrow^{n-1} b' \rightarrow^{n'} d'$, there is a d'' such that $d'' \leftarrow^{(n-1)-n'} d'$, hence $d'' \leftarrow^{((n-1)-n')+m'} c'$. By calculating $((n-1)-n') + m' \geq ((n-1)-m') + m' = \max(n-1, m') \geq n-1$ we find a reduction of length $n-1$ from c' as desired. \square

Lemma (Extension). If \rightarrow uniformly delimits \rightarrow , then for c normal, $b \leftarrow^n a \rightarrow^m c$ is completable into a delimiting diagram, i.e. $b \rightarrow^k c$ with $n+k \leq m$.

Proof. By induction on m . In case $m=0$, then $n=0$ by our global assumption that \rightarrow, \rightarrow -normal forms coincide. In case $n=0$, setting $k=m$ works. Supposing $n, m \geq 1$, there exist b' and c' such that $b \leftarrow^{n-1} b' \leftarrow a \rightarrow c' \rightarrow^{m-1} c$. By uniform delimitation, either there is an infinite \rightarrow -reduction from c' , or there are $n' \leq m'$, d' such that $b' \rightarrow^{n'} d' \leftarrow^{m'} c'$. The former case is impossible, since this would yield by the Progress Lemma some non-empty reduction from c contradicting it being in normal form. Thus the latter case holds and by the IH for $d' \leftarrow^{m'} c' \rightarrow^{m-1} c$, $d' \rightarrow^{k'} c$ with $m'+k' \leq m-1$, hence $b' \rightarrow^{n'+k'} c$. By $n'+k' \leq m'+k' \leq m-1$ the IH may be applied to $b \leftarrow^{n-1} b' \rightarrow^{n'+k'} c$, yielding $b \rightarrow^k c$ with $(n-1)+k \leq n'+k'$, which implies $n+k \leq m$ as desired. \square

In the following, we first assume \rightarrow to be a strategy for \rightarrow , next, dually, \rightarrow to be a strategy for \rightarrow , and finally both, i.e. that \rightarrow, \rightarrow coincide. By the definition of strategy the normal forms of \rightarrow, \rightarrow coincide each time, in concordance with our global assumption. So, first, let \rightarrow be a strategy for \rightarrow . Then we leave \rightarrow implicit, and say \rightarrow is *uniformly minimal* instead of \rightarrow uniformly delimits \rightarrow .

Lemma (Internal Needed). The internal needed strategy is uniformly minimal for orthogonal TRSs.

Proof. ² An internal needed redex is an innermost redex among the needed ones. We distinguish cases on the relative positions of the redexes contracted at p, q in a peak $s \leftarrow_p t \rightarrow_q u$:

(=) Then $s = u$, so the \rightarrow -step is uniformly minimal.

(||) Then $s \rightarrow_q v \leftarrow_p u$, for some term v . Since at least one of the residuals of a needed redex must be needed, and here the needed redex has a unique residual, it must therefore be needed, so in fact $v \leftarrow_p u$ and we are ok.

(<) Then q is non-needed. Now consider a maximal \rightarrow -reduction from t via s . Since this is an internal needed reduction, its projection over the non-needed q -step yields an internal needed reduction again from u , of exactly the same length. If the former reduction is finite, then the latter reduction being its projection has the same target, from which one easily concludes.

(>) Then $s \rightarrow_q v \leftarrow u$ for some v , where the \rightarrow -reduction in fact contracts the residuals of the redex at position ϕ which all are at disjoint positions. We may partition this set into the non-needed and the internal needed residuals, and accordingly partition the reduction $v \leftarrow u$ as $v \leftarrow v' \leftarrow u$. Note that since

²Apart from separating out uniform minimality, this proof also corrects the flawed proof of Theorem 9.4.7 in [5].

p is needed, the latter partition is non-empty. Since taking the residual of any needed reduction from v' along $v \leftarrow v'$ yields a reduction from v of exactly the same length, the result follows from confluence (or non-termination). \square

Minimality and normalisation of the internal needed strategy follow:

Theorem (Minimality). If \rightarrow is uniformly minimal, then it is minimal and normalising.

Proof. Suppose $a \rightarrow^m b$ with b a normal form. Then, by the Extension Lemma, m exceeds the length of any \rightarrow -reduction from a (normalisation), in particular of any \rightarrow -reduction from a to b (minimality). \square

Another application of the Minimality Theorem is the well-known fact that the Gross-Knuth strategy \rightarrow_{GK} , contracting all redexes in a term in one go, is minimal and normalising for orthogonal rewrite systems, be it TRSs, the $\lambda\beta$ -calculus, or a rewrite system in any higher-order format such as Nipkow's higher-order pattern rewrite systems (PRSs), Klop's combinatory reduction systems (CRSs), or Khasidashvili's expression reduction systems (ERSs).

Lemma (Gross-Knuth). \rightarrow_{GK} is uniformly minimal for orthogonal PRSs.

Proof. If $s \leftarrow_{GK} t \rightarrow u$, then either $s = u$ or $s \leftarrow_{GK} u$ or there exists some v such that $s \rightarrow v \leftarrow_{GK} u$, by the standard theory for projecting multi-steps. \square

Next, let \rightarrow be a strategy for \rightarrow . We leave \rightarrow implicit and say \rightarrow is *uniformly maximal* instead of \rightarrow uniformly delimits \rightarrow . How to find such a \rightarrow ? For orthogonal second-order rewrite systems³ external steps would fit the bill, but for their failure to deal with erasure: e.g. for the λ -calculus leftmost-outermost (lmo) steps are external but uniform maximality fails for $(\lambda x.y)N' \leftarrow (\lambda x.y)N \rightarrow y$ since $(\lambda x.y)N' \rightarrow y$. The limit strategy⁴ solves this by calling itself on erased arguments, instead of taking the external step in such cases.

Lemma (Limit). The limit strategy is uniformly maximal for orthogonal second-order PRSs (e.g. for CRSs).⁵

Proof. Distinguish cases on the relative positions of the redexes contracted at p, q in $s \leftarrow_p t \rightarrow_q u$:

(=) Then $s = u$, so uniform maximality holds.

(||) Then $s \rightarrow_q v \leftarrow_p u$, for some term v , and we are ok again.

(<) Then by definition of \rightarrow and orthogonality, p must be a redex erasing q , so $s \leftarrow_p u$.

(>) Then $s \rightarrow_q v \leftarrow_{\ominus} u$ for some v , where the latter reduction contracts the

³The rewrite systems should also be *fully-extended* in the sense of [4, Def. 10], meaning that rules such as the η -rule in the λ -calculus testing for the *absence* of a bound variable are forbidden.

⁴The strategy F_∞ for the λ -calculus is a special case of a limit strategy.

⁵Also here, full-extendedness is required (to guarantee existence of external steps).

residuals of the redex at position p , which holds by the diamond property for multi-steps in orthogonal PRSs [5, Theorem 11.6.29]. If $s \rightarrow_q v$ is non-erasing, then $s \rightarrow_q v$. Otherwise,⁶ \rightarrow -reduce each erased argument in turn to its normal form, before contracting the (then limit) redex at q to v . By [5, Theorem 11.6.29] again, u reduces to v by performing for each erased argument of s , these same steps on the (non-empty) set of descendants of each argument in u .⁷ To guarantee that this reduction has at least the same length as that from s , it suffices to perform the reduction on the descendants according to the inside-out order of the descendants. The only exception to this is when the \rightarrow -reduction from s be infinite, but then u allows an infinite reduction as well. \square

The λ -calculus with explicit substitutions $\lambda\mathbf{x}^-$ is a left-linear and left-normal second-order PRS [4, Def. 13]. So lmo-redexes are external, inducing a limit strategy known as F_∞ .

Lemma ($\lambda\mathbf{x}^-$ -limit). F_∞ is uniformly maximal for $\lambda\mathbf{x}^-$.

Proof. It suffices to extend the case analysis in the proof of the Limit Lemma with overlap:

(#) $C[M\langle x:=N \rangle \langle y:=P \rangle] \leftarrow_p C[(\lambda x.M)N\langle y:=P \rangle] \rightarrow_q C[(\lambda x.M)\langle y:=P \rangle N\langle y:=P \rangle]$.

As in the ($>$)-case we simulate a \rightarrow -reduction from s , by a reduction from u which is at least as long, giving the desired result by confluence of $\lambda\mathbf{x}^-$. Start simulating with $u \rightarrow C[M\langle y:=P \rangle \langle x:=N \rangle \langle y:=P \rangle] = u'$. Then the idea is that the $\langle x:=N \rangle \langle y:=P \rangle$ -closure occurring in s as indicated, is only ever replicated in its entirety, in two consecutive \rightarrow -steps, which are matched on u' each time by replicating the corresponding $\langle y:=P \rangle \langle x:=N \rangle \langle y:=P \rangle$ -closure. Only if a variable is reached, a further case distinction is needed:

(x) Then conclude from $x\langle x:=N \rangle \langle y:=P \rangle \rightarrow N\langle y:=P \rangle \leftarrow^2 x\langle y:=P \rangle \langle x:=N \rangle \langle y:=P \rangle$.

(y) Then \rightarrow recurs on N and we go via $y\langle x:=N' \rangle \langle y:=P \rangle \rightarrow y\langle y:=P \rangle \rightarrow P$ to P' , with N' , P' the normal forms of N , P . This is matched by a reduction from $y\langle y:=P \rangle \langle x:=N \rangle \langle y:=P \rangle$ via $y\langle y:=P \rangle \langle x:=N' \rangle \langle y:=P \rangle \rightarrow P\langle x:=N' \rangle \langle y:=P \rangle$ to $P'\langle x:=N' \rangle \langle y:=P \rangle$ to P' (in at least one step), using that x occurs in P nor P' , and that P' is in normal form so does not contain closures.

(z) As in the previous case but simpler as we end up just in z (if \rightarrow terminates at all). \square

These strategies being uniformly maximal implies their maximality and perpetuality, by substituting the latters for minimality and normalisation in (the proof of) the Minimality Theorem:

Theorem (Maximality). If \rightarrow is uniformly maximal, then it is maximal and perpetual.

⁶In the λ -calculus this is witnessed by e.g. $(\lambda x.y)N \leftarrow (\lambda x.(\lambda z.y)x)N \rightarrow (\lambda z.y)N$. In that case, we should have first \rightarrow -reduced N to normal form, say N' , before to proceed with $(\lambda x,y)N' \rightarrow y$

⁷For third order rewrite systems this might not be possible, and the theorem fails.

This answers the open question whether F_∞ is maximal [2, Rem. 3.18] for $\lambda\mathbf{x}^-$, in the affirmative.⁸

Finally, we let \rightarrow coincide with \rightarrow , in which case we speak of \rightarrow being *self-delimiting*. Joining minimality and maximality into *equidistance*, i.e. all reductions from a given object a to a normal form b have the same length, and normalisation and perpetuality into uniform normalisation [4, Def. 2], i.e. for all objects a , $\text{WN}(a)$ implies $\text{SN}(a)$, we have:

Theorem (Self-delimitation). if \rightarrow is *self-delimiting*, then it is equidistant and uniformly normalising.

This covers many notions and results in the literature: Balanced WCR [6] requires $k = l$ (in the definition of \rightarrow uniform delimitation \rightarrow), and its generalisation balanced SCR [3] requires the natural numbers and object chosen for a given a peak and its symmetric version to be identical. Linear biclosed [4] requires $k = 0$ or $k = 1 = l$, and its generalisation $\text{SCR}^{\geq 1}$ [3] $k = 0 = l$ or $k \leq 1 \leq l$. Our generalisation is proper in all these cases as witnessed by $a \rightarrow b \rightarrow e \curvearrowright f$ and $a \rightarrow c \rightarrow d \rightarrow e$. Beware of symmetry (cf. Exc. 1.3.11): for any ARS if $b \leftarrow a \rightarrow c$ then also $c \leftarrow a \rightarrow b$. Hence self-delimitation requires that in such a case we not only have $b \rightarrow^n d \leftarrow^m c$, for some d and some $n \leq m$, but also $b \rightarrow^{n'} d' \leftarrow^{m'} c$, for some d' and some $n' \geq m'$

Acknowledgments We would like to thank F.-J. de Vries for the initial question and him, M. Bezem, E. Bonelli, B. Gramlich, Z. Khasidashvili, A. Middeldorp, Y. Toyama, and the referees for useful comments.

References

- [1] M. Bezem and T. Coquand. Newman’s lemma – a case study in proof automation and geometric logic. BEATCS 79, 2004.
- [2] E. Bonelli. *Substitutions explicites et réécriture de termes*. PhD thesis, Paris XI, 2001.
- [3] B. Gramlich. On some abstract termination criteria. Workshop on Termination ’99.
- [4] Z. Khasidashvili, M. Ogawa, and van Oostrom. Uniform normalisation beyond orthogonality. RTA01.
- [5] Terese. *Term Rewriting Systems*. Cambridge University Press, 2003.
- [6] Y. Toyama. Strong sequentiality of left-linear overlapping term rewriting systems. LICS92.

⁸As our proof is a more precise version of the proof [4, Lem. 6] of preservation of strong normalisation (PSN) for $\lambda\mathbf{x}^-$, which only used perpetuality (not maximality), we expect it works for PSN λ -calculi with explicit substitutions.

A Basic notions and facts

For the convenience of the reader we recapitulate some basic notions and facts from [5] as used in this paper. For each, we put a reference to the corresponding item-number in [5] between parentheses to facilitate locating the motivation for them as given there.

Definition (8.2.2). An *abstract rewriting system*, ARS for short, is a quadruple $\langle A, \Phi, \text{src}, \text{tgt} \rangle$ with A a set of *objects*, Φ a set of *steps* and $\text{src}, \text{tgt} : \Phi \rightarrow A$ the *source* and *target* functions, respectively.

Definition (8.2.5(i)). Let $\rightarrow_i = \langle A_i, \Phi_i, \text{src}_i, \text{tgt}_i \rangle$ be abstract rewriting systems, for $i \in \{1, 2\}$. We say \rightarrow_1 is a *sub-abstract-rewriting-system* (sub-ARS) of \rightarrow_2 if $A_1 \subseteq A_2$, $\Phi_1 \subseteq \Phi_2$, and $\text{src}_1, \text{tgt}_1$ are the restrictions of $\text{src}_2, \text{tgt}_2$ to Φ_1 .

Definition (8.5.59). A term rewrite rule is *left-normal* if in its left-hand side no function symbols occur to the right of a variable. A term rewriting system is *left-normal* if all its rules are.

Definition (9.1.1). A *strategy* for an abstract rewriting system \rightarrow is a sub-ARS of \rightarrow having the same objects and normal forms.

Definition (9.1.4). An object of an abstract rewriting system is *deterministic* if it is the source of at most one step. An abstract rewriting system is *deterministic* if all its objects are so.

Definition (9.1.12). A strategy for an ARS \rightarrow is *normalizing* if it is terminating.

Definition (9.2.1). A step ϕ from a term t is said to be *needed* if some residual of ϕ must be eliminated by overlap in any reduction from t to normal form.

Theorem (9.2.9). The needed strategy is normalizing for orthogonal term rewriting systems.

Definition (9.2.31). A step from a term is *external to* a reduction from that term if its residuals are not nested by other (not necessarily contracted) redexes in the course of the reduction. The step is *external* if it is external to any reduction.

Definition (9.4.1). A strategy is *minimal* if for any term t , the length of any reduction from t to a normal form s according to the strategy, is minimal among all possible reductions from t to s .

Definition (9.4.4). A step from a term is an *internal* step if its residuals do not nest needed redexes in the course of any reduction. The *internal needed* strategy contracts internal needed steps.

Lemma (9.4.5). Innermost needed redexes are internal needed for orthogonal term rewriting systems.

Definition (9.5.1). A strategy is *maximal* if the minimal number of steps according to the strategy, needed to reach a normal form is maximal among all reductions to normal form, for any given term.

A step is *perpetual* if it preserves non-termination. A strategy is *perpetual* if it performs perpetual steps.

Definition (9.5.5). The *limit* strategy is inductively defined by:

- any non-redex-erasing external step is a limit step.
- if $\phi:t \rightarrow s$ is a limit step, then $C[\phi]:C[t] \rightarrow C[s]$ is a limit step, if t is an erased argument of an external redex in $C[t]$.