Information and Computation 141, 172–204 (1998)
Article No. IC972683

Diagram Techniques for Confluence

Marc Bezem

Department of Philosophy, Utrecht University, P.O. Box 80126, 3508 TC Utrecht, The Netherlands
E-mail: bezem@phil.ruu.nl

Jan Willem Klop

CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands; and
Department of Mathematics and Computer Science, Vrije Universiteit,
De Boelelaan 1081A, 1081 HV Amsterdam, The Netherlands
E-mail: jwk@cwi.nl

and

Vincent van Oostrom

Department of Mathematics and Computer Science, Vrije Universiteit,
De Boelelaan 1081A, 1081 HV Amsterdam, The Netherlands
E-mail: oostrom@cs.vu.nl

We develop diagram techniques for proving confluence in abstract reductions systems. The underlying theory gives a systematic and uniform framework in which a number of known results, widely scattered throughout the literature, can be understood. These results include Newman's lemma, Lemma 3.1 of Winkler and Buchberger, the Hindley–Rosen lemma, the Request lemmas of Staples, the Strong Confluence lemma of Huet, the lemma of De Bruijn.

1. INTRODUCTION

The concept of a term rewriting system (TRS) is paradigmatic for the study of computational procedures. TRSs play an important role in various areas, such as abstract data type specifications, implementations of functional programming languages, and automated deduction. Rewriting is concerned with syntactical objects like terms, strings, and term graphs, but also with equivalence classes of terms or other structured objects. Terms may be first-order or higher-order, such as λ-terms or proofs in some deduction system. Many of the basic definitions and facts can already be stated on a more abstract level, where the structure of the objects to be rewritten is not yet of relevance. To express this level of abstraction we use the neutral term "reduction" instead of "rewriting." In the next section we give the

0890-5401/98 $25.00

necessary elementary definitions and basic facts about abstract reduction systems (ARSs).

We develop diagram techniques for proving confluence in abstract reductions systems. The underlying theory gives a systematic and uniform framework in which a number of standard proof techniques for confluence, widely scattered throughout the literature, can be understood. These results include Newman's lemma (1942), Lemma 3.1 of Winkler and Buchberger (1985), the Hindley–Rosen lemma (1964), the Request lemmas of Staples (1975), the Strong Confluence lemma of Huet (1980), and the lemma of De Bruijn (1978), which served as a starting point of this research. The notions of "diagram technique" and "standard proof technique" are essentially open-ended. We certainly do not claim to cover or subsume them all. For example, Proposition 4.1 is a simple confluence result obtained by a diagram technique that is *not* a special case of the Theorem 4.28, the strongest theorem we obtain with decreasing diagrams. The completeness results at the end of Section 4 are rather weak and do not want to suggest some kind of completeness with respect to standard proof techniques. However, we do claim ease in the use of our method as compared to standard proof techniques. For all results stated above to be generalized by diagram techniques we found the diagram technique more intuitive and easier to use.

The present paper extends Van Oostrom (1994a, 1994b) in the following ways: all results about reduction diagrams are new, and the concept of a trace-decreasing diagram refining the concept of a decreasing diagram has a clearer visualization and yields a new proof of the main theorem. With over 30 figures the approach here is more geometric, as opposed to the more algebraic approach of van Oostrom (1994a).

We will assume that the reader is familiar with the terminology and notation of elementary set theory and logic, such as *sets*, *boolean operators*, *quantification*, *relations* and *functions*, *inverse*, *reflexive*, *symmetric*, *transitive* and *equivalence relations*, *closure operations* such as the *reflexive*, *symmetric*, and *transitive closure* (also simultaneously), and so on. Moreover, we assume the definitions of (*partial*) *order*, *strict order*, *totally or linearly ordered set* (*chain*), *well-founded order and well-founded induction.*

For (finite) *multisets*, denoted by $[s_0, s_1, ..., s_{n-1}]$, we refer to the appendix, where the basic definitions and the necessary results are stated. The *empty multiset* is denoted by $[\ ]$. The set of finite multisets over S is denoted by $S^{\#}$.

A *finite sequence of length* n in S is a function s: $\{0, 1, ..., n-1\} \to S (n \geqslant 0)$, also denoted by $\langle s_0, s_1, ..., s_{n-1}\rangle$ or $s_0 s_1 \cdots s_{n-1}$. The *empty sequence* is denoted by $\langle\ \rangle$. The set of finite sequences in S is denoted by S^*. *Concatenation* of finite sequences σ and τ is denoted by $\sigma \cdot \tau$.

2. ABSTRACT REDUCTION SYSTEMS

2.1. DEFINITION. An *abstract reduction system* is a structure $\mathscr{A} = (A, \{\to_\alpha \mid \alpha \in I\})$ consisting of a set A and a set of binary relations $\to_\alpha$ on A, indexed by a set I. We write $(A, \to_1, \to_2)$ instead of $(A, \{\to_\alpha \mid \alpha \in \{1, 2\}\})$. For $\alpha \in I$, the relations $\to_\alpha$ are called *reduction* or *rewrite* relations. Sometimes we will refer to $\to_\alpha$ as α. In the case of just one reduction relation, we simply write $\to$. We write $\to_I$ for the union $\bigcup\{\to_\alpha \mid \alpha \in I\}$.

In the following, we will consider several equivalence relations on the set A. One of them is identity on A. As the elements of A will often have a syntactic nature, we will use $\equiv$ to denote identity on A, conforming to the convention that $\equiv$ expresses syntactical identity. The usual symbol $=$ to denote identity now becomes available to denote another important equivalence relation, namely *convertibility*, the equivalence relation generated by $\to$. Similarly, $=_\alpha$ denotes the equivalence relation generated by $\to_\alpha$, for every $\alpha \in I$.

Let $\mathscr{A} = (A, \{\to_\alpha \mid \alpha \in I\})$ be an ARS and let $\alpha \in I$. If for $a, b \in A$ we have $(a, b) \in \to_\alpha$, we write $a \to_\alpha b$ and call b a *one-step* (α-)*reduct* of a. A *reduction sequence* with respect to $\to_\alpha$ is a (finite or infinite) sequence $a_0 \to_\alpha a_1 \to_\alpha a_2 \to_\alpha \cdots$. Every reduction sequence has a first element, and finite reduction sequences also have a last element. Whenever we want to stipulate these elements we use the following terminology: a reduction sequence starting from a is called a reduction sequence *of* a, and if such a reduction sequence ends in b, then it is called a reduction sequence *from a to b*. The element b is called an (α-)*reduct* of a in this case. Reduction sequences are also called *reduction paths*. A *reduction step* is a specific occurrence of $\to_\alpha$ in a reduction sequence. A *reduction step from a to b* is a specific occurrence of $a \to_\alpha b$. The *length* of a finite reduction sequence is the number of reduction steps occurring in this reduction sequence (one less than the number of elements!). An *indexed reduction sequence* is a finite or infinite sequence of the form $a \to_\alpha b \to_\beta c \to_\gamma \cdots$ with $a, b, c, \ldots \in A$ and $\alpha, \beta, \gamma, \ldots \in I$. Thus an indexed reduction sequence is a reduction sequence with respect to $\bigcup\{\to_\alpha \mid \alpha \in I\}$, with the reduction steps marked with the index of the reduction relation to which the step belongs. We write $a \to_I b$ if we do not wish to specify the index of the reduction step.

The transitive reflexive closure of $\to_\alpha$ is written as $\twoheadrightarrow_\alpha$ According to the definition of transitive reflexive closure, we have $a \twoheadrightarrow_\alpha b$ if and only if there is a finite reduction sequence $a \equiv a_0 \to_\alpha a_1 \to_\alpha \cdots \to_\alpha a_n \equiv b (n \geqslant 0)$. If we write $\sigma\colon a \twoheadrightarrow_\alpha b$, then σ denotes an arbitrary reduction sequence from a to b. We write $\sigma\colon a \to_\alpha a_1 \to_\alpha \cdots \to_\alpha b$ whenever we want to stipulate a specific reduction sequence σ from a to b. Similarly, finite indexed reduction sequences will be denoted by $\sigma\colon a \twoheadrightarrow_I b$.

The reflexive closure of $\to_\alpha$ is $\to_\alpha^\equiv$. The symmetric closure of $\to_\alpha$ is $\leftrightarrow_\alpha$. The transitive closure of $\to_\alpha$ is $\to_\alpha^+$. The inverse relation of $\to_\alpha$ is $\to_\alpha^{-1}$, also denoted by $\leftarrow_\alpha$. Let $\to_\beta$ also be a reduction relation on A. The union $\to_\alpha \cup \to_\beta$ is denoted by $\to_{\alpha\beta}$. The composition $\to_\alpha$ and $\to_\beta$ is denoted by $\to_\alpha \cdot \to_\beta$. We have $a \to_\alpha \cdot \to_\beta c$ if and only if $a \to_\alpha b \to_\beta c$ for some $b \in A$.

ARSs are also called *labeled transition systems* in the modal literature, see for example Popkorn (1994). An ARS with just one reduction relation $\to$ is called a *replacement system* in Staples (1975) and *transformation system* in Jantzen (1988). Below we will define a number of properties of the reduction relation $\to$. If this reduction relation has a certain property, then we will attribute this property also to the ARS in question and vice versa. Most of the properties are first defined element-wise, that is, as a property of elements of the ARS.

2.2. DEFINITION (confluence). Let $\mathcal{A} = (A, \{\to_\alpha \mid \alpha \in I\})$ be an ARS, $\alpha, \beta \in I$ and let $\to\; = \;\to_\alpha$.

(i) α *commutes weakly* with β if the diagram of Fig. 1a holds, i.e., if $\forall a, b, c \in A\ \exists d \in A\ (c \leftarrow_\beta a \to_\alpha b \Rightarrow c \twoheadrightarrow_\alpha d \twoheadleftarrow_\beta b)$. In a shorter notation: $\leftarrow_\beta \cdot \to_\alpha \subseteq \twoheadrightarrow_\alpha \cdot \twoheadleftarrow_\beta$. Furthermore, α *commutes* with β if $\twoheadrightarrow_\alpha$ and $\twoheadrightarrow_\beta$ commute weakly, or, equivalently: $\twoheadleftarrow_\beta \cdot \twoheadrightarrow_\alpha \subseteq \twoheadrightarrow_\alpha \cdot \twoheadleftarrow_\beta$.

(ii) $a \in A$ is *weakly confluent* if $\forall b, c \in A\ \exists d \in A\ (c \leftarrow a \to b \Rightarrow c \twoheadrightarrow d \twoheadleftarrow b)$. The reduction relation $\to$ is *weakly confluent* or *weakly Church–Rosser* (WCR) if every $a \in A$ is weakly confluent. Alternative characterizations are: $\leftarrow \cdot \to \subseteq \twoheadrightarrow \cdot \twoheadleftarrow$, or $\to$ is weakly self-commuting (see Fig. 1b).

(iii) $a \in A$ is *confluent* if $\forall b, c \in A\ \exists d \in A\ (c \twoheadleftarrow a \twoheadrightarrow b \Rightarrow c \twoheadrightarrow d \twoheadleftarrow b)$. The reduction relation $\to$ is *confluent* or *Church–Rosser* or has the Church–Rosser property (CR), if every $a \in A$ is confluent. Alternative characterizations are: $\twoheadleftarrow \cdot \twoheadrightarrow \subseteq \twoheadrightarrow \cdot \twoheadleftarrow$, or $\to$ is self-commuting (see Fig. 1d).

The property WCR is often called "local confluence," e.g., in Jantzen (1988). In the following we will use the terms confluent and Church–Rosser (CR) without preference. Likewise for weakly confluent and WCR. The following proposition follows immediately from the definitions. Note especially the equivalence of (i) and (v). Some authors call Definition 2.2(iii) confluent and Proposition 2.3(v) Church–Rosser.

2.3. PROPOSITION. *For every ARS* $(A, \to)$, *the following are equivalent*:

(i) $\to$ *is confluent*

(ii) $\twoheadrightarrow$ *is weakly confluent*

(iii) $\twoheadrightarrow$ *is self-commuting*

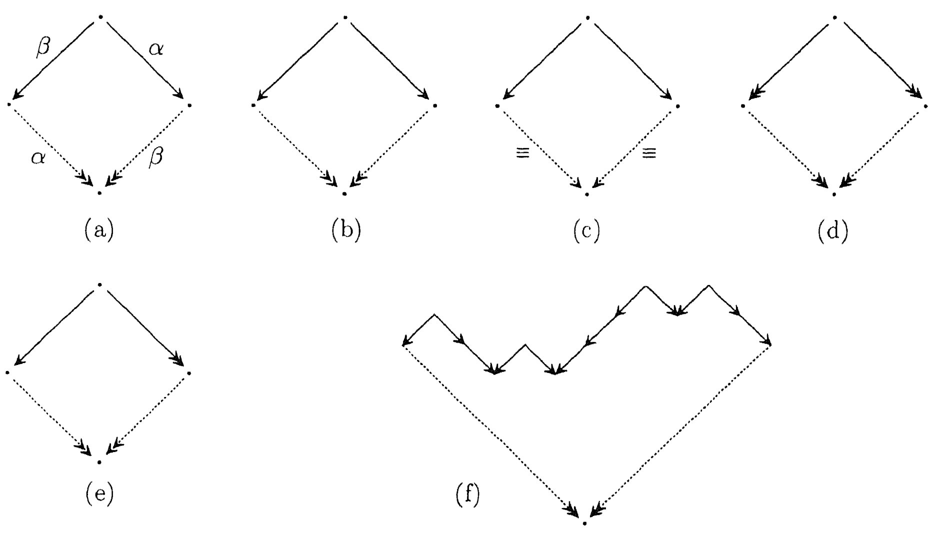


FIG. 1. Various confluence patterns.

(iv) $\leftarrow \cdot \twoheadrightarrow \;\subseteq\; \twoheadrightarrow \cdot \twoheadleftarrow$, *or* $\forall a, b, c \in A\ \exists d \in A\ (c \leftarrow a \twoheadrightarrow b \Rightarrow c \twoheadrightarrow d \twoheadleftarrow b)$, *that is, the diagram in Fig.* 1e *holds.*

(v) $= \subseteq \twoheadrightarrow \cdot \twoheadleftarrow$, *or* $\forall a, b \in A\ \exists c \in A\ (a = b \Rightarrow a \twoheadrightarrow c \twoheadleftarrow b)$, *where* $=$ *is convertibility, the equivalence relation generated by* $\rightarrow$, *or the smallest equivalence relation containing* $\rightarrow$. *See the diagram in Fig.* 1*f.*

Proof. The equivalence of (i)–(iii) follows immediately from the definitions, using that $\twoheadrightarrow$ is reflexive and transitive itself. Obviously, (v) implies (i), and (i) implies (iv) in turn. It remains to prove that (iv) implies (v). Assume (iv) and let $a = b$. Recall that $=$ is the equivalence generated by $\rightarrow$. Hence there exist $a_0, ... a_n \in A$ ($n \geqslant 0$) such that $a \equiv a_0 \leftrightarrow \cdots \leftrightarrow a_n \equiv b$, where $\leftrightarrow$ is the symmetric closure of $\rightarrow$. We argue by induction on n. If $n = 0$, then we trivially have $a \twoheadrightarrow c \twoheadleftarrow b$ by taking $c \equiv a \equiv b$. Assume (v) has been proved for n and let $a \equiv a_0 \leftrightarrow \cdots \leftrightarrow a_n \leftrightarrow a_{n+1} \equiv b$. By the induction hypothesis there exists $c \in A$ such that $a \twoheadrightarrow c \twoheadleftarrow a_n$. If $a_{n+1} \rightarrow a_n$, then we are done. If $a_n \rightarrow a_{n+1}$, then we can apply (iv) to $a_{n+1} \leftarrow a_n \twoheadrightarrow c$ and we are also done. This completes the proof of (v). ∎

2.4. Definition (normalization). Let $\mathscr{A} = (A, \rightarrow)$ be an ARS.

(i) $a \in A$ is a *normal form* if there exists no $b \in A$ such that $a \rightarrow b$.

(ii) $a \in A$ is *weakly normalizing* (WN) if $a \twoheadrightarrow b$ for some normal form $b \in A$. The reduction relation $\rightarrow$ is *weakly normalizing* if every $a \in A$ is weakly normalizing.

(iii) $a \in A$ is *strongly normalizing* (SN) if every reduction sequence starting from a is finite. The reduction relation $\rightarrow$ is *strongly normalizing* if every $a \in A$ is strongly normalizing. Alternative characterization: $\leftarrow$ is well founded (WF). Obviously, SN implies WN and $\leftarrow$ is SN if and only if $\rightarrow$ is WF.

(iv) Let $a \in A$. The *reduction graph* $\mathscr{G}(a)$ of a is the ARS with all reducts of a as elements and the reduction relation $\rightarrow$ restricted to this set of reducts. Let $B \subseteq A$. Then B is *cofinal* in $\mathscr{A}$ if $\forall a \in A\ \exists b \in B\ a \twoheadrightarrow b$. We say that $\rightarrow$ has the *cofinality property* (CP) if in every reduction graph $\mathscr{G}(a)$, $a \in A$, there is a (finite or infinite) reduction sequence $a \equiv a_0 \rightarrow a_1 \rightarrow \cdots$ such that $\{a_n \mid n \geqslant 0\}$ is cofinal in $\mathscr{G}(a)$.

(v) Let $a \in A$. The *component* $\mathscr{C}(a)$ of a with respect to conversion is the ARS with $\{a' \in A \mid a = a'\}$ as set of elements and the reduction relation $\rightarrow$ restricted to this convertibility class. Now define the property $\mathrm{CP}^=$ for $\mathscr{A}$ to hold if every component $\mathscr{C}(a)$, $a \in A$, contains a reduction sequence $a \equiv a_0 \rightarrow a_1 \rightarrow \cdots$ such that $\{a_n \mid n \geqslant 0\}$ is cofinal in $\mathscr{C}(a)$.

Lemma (Klop (1980)). *For every* ARS *we have*:

(i) CP $\Rightarrow$ CR

(ii) CR $\Rightarrow$ CP, *provided the set of elements is countable.*

(iii) CP $\Leftrightarrow$ $\mathrm{CP}^=$, *due to Mano* (1993).

Proof. Let $\mathscr{A} = (A, \rightarrow)$ be an ARS.

(i) Assume CP and let $b \twoheadleftarrow a \twoheadrightarrow c$ for some $a, b, c \in A$. Let $a_0 \to a_1 \cdots$ be a reduction sequence such that $\{a_n \mid n \geqslant 0\}$ is cofinal in $\mathcal{G}(a)$. We have $b, c \in \mathcal{G}(a)$, so by the cofinality there exist $i, j \geqslant 0$ such that $b \twoheadrightarrow a_i$ and $c \twoheadrightarrow a_j$. If $i < j$, then a_j is the desired common reduct, otherwise a_i.

(ii) Assume CR, A countable, and let $a \equiv a_0, a_1, \cdots$ be an enumeration of $\mathcal{G}(a)$. Define recursively $b_0 \equiv a$ and b_{n+1} as a common reduct of b_n and a_{n+1}. Then $b_0 \twoheadrightarrow b_1 \cdots$ yields a reduction sequence and $\{b_n \mid n \geqslant 0\}$ is cofinal in $\mathcal{G}(a)$.

(iii) $\mathrm{CP}^=$ trivially implies CP since any cofinal reduction sequence $a \equiv a_0 \to a_1 \to \cdots$ in $\mathcal{C}(a)$ is also a cofinal reduction sequence in $\mathcal{G}(a)$. For the converse, assume CP. By (i) we have CR in the formulation of Proposition 2.3(v). Let $a \in A$ and $a \equiv a_0 \to a_1 \to \cdots$ be cofinal in $\mathcal{G}(a)$. If $a' \in \mathcal{C}(a)$, then $a' = a$, so by CR we have $a' \twoheadrightarrow a''$ for some $a'' \in \mathcal{G}(a)$, hence $a' \twoheadrightarrow a_i$ for suitable $i \geqslant 0$. It follows that $\{a_n \mid n \geqslant 0\}$ is cofinal in $\mathcal{C}(a)$. ∎

2.6. Lemma (Newman's lemma). *For every* ARS *we have* SN $\wedge$ WCR $\Rightarrow$ CR.

Proof. Let $\mathcal{A} = (A, \to)$ be an ARS. Short proofs of Newman's lemma can be found in Huet (1980) and in Barendregt (1984). One can also obtain proofs of Newman's lemma from more general results on reduction diagrams, see Corollary 3.9 and Example 4.20 in Section 4. We list two proofs below. The first proof is the canonical one by well-founded induction and anticipates the proof of the main theorem on trace-decreasing diagrams, Theorem 4.19. The second proof using multisets is also important, since this argument will play a role in Proposition 4.1. Assume $\mathcal{A}$ has the properties SN and WCR.

(i) As $\to$ is SN, $\leftarrow$ is WF, and hence $\leftarrow^+$ is well-founded. Thus we can apply well-founded induction, with respect to $\leftarrow^+$. We prove $\forall a \in A\ \phi(a)$, with $\phi(a)$ expressing that a is confluent. We have to show that ϕ is $\leftarrow^+$-inductive. Let $a \in A$ and assume we have proved $\phi(a')$ for all $a' \in A$ with $a' \leftarrow^+ a$. Let $c \twoheadleftarrow a \twoheadrightarrow b$. If $a \equiv b$ or $a \equiv c$, then we are done. Otherwise, $c \twoheadleftarrow c' \leftarrow a \to b' \twoheadrightarrow b$ for some c', $b' \in A$. Apply WCR to $c' \leftarrow a \to b'$ in order to find d' such that $c' \twoheadrightarrow d' \twoheadleftarrow b'$. We have $b' \leftarrow a$ and $c' \leftarrow a$, so by the induction hypothesis $\phi(b')$ and $\phi(c')$. The first gives us $e \in A$ such that $d' \twoheadrightarrow e \twoheadleftarrow b$, so $c \twoheadleftarrow c' \twoheadrightarrow d' \twoheadrightarrow e$. The second gives us $d \in A$ such that $c \twoheadrightarrow d \twoheadleftarrow e$, so $c \twoheadrightarrow d \twoheadleftarrow b$. Making a picture can now be helpful to see that we have proved $\phi(a)$.

(ii) Recall that $=$ is the equivalence generated by $\to$. Let $a = b$, then there exist $a_0, \ldots a_n \in A$ $(n \geqslant 0)$ such that $a \equiv a_0 \leftrightarrow \cdots \leftrightarrow a_n \equiv b$, where $\leftrightarrow$ is the symmetric closure of $\to$. We view $a_0 \leftrightarrow \cdots \leftrightarrow a_n$ as a *landscape* with *peaks* $a_{i-1} \leftarrow a_i \leftarrow a_{i+1}$, *valleys* $a_{i-1} \to a_i \leftarrow a_{i+1}$ and *slopes* $a_i \to \cdots \to a_{i+k}$ or $a_i \leftarrow \cdots \to a_{i+k}$, for some $k > 0$. If $a_0 \leftrightarrow \cdots \leftrightarrow a_n$ contains no peaks, then it is either one (maximal) slope, two (maximal) slopes with one valley, or one single point. In all these cases we can easily find $c \in A$ with $a \twoheadrightarrow c \twoheadleftarrow b$. If $a_0 \leftrightarrow \cdots \leftrightarrow a_n$ does contain a peak, say $a_{i-1} \leftarrow a_i \to a_{i+1}$, then we can eliminate this peak by applying WCR: for some $d \in A$ we have $a_{i-1} \twoheadrightarrow d \twoheadleftarrow a_{i+1}$. By the definition of transitive closure, there exist $c_1, \ldots, c_n$, $c'_1, \ldots, c'_{n'} \in A$ $(n, n' \geqslant 0)$ such that $a_{i-1} \to c_1 \to \cdots \to c_n \equiv d \equiv c'_{n'} \leftarrow \cdots \leftarrow c'_1 \leftarrow a_{i+1}$.

Then the landscape becomes $a_0 \leftrightarrow \cdots \leftrightarrow a_{i-1} \leftrightarrow c_1 \leftrightarrow \cdots \leftrightarrow d \leftrightarrow \cdots \leftrightarrow c'_1 \leftrightarrow a_{i+1} \leftrightarrow \cdots \leftrightarrow a_n$. This does not seem to help very much. However, we shall argue that this procedure of eliminating peaks must terminate, so that we necessarily end up with a landscape between a_0 and a_n containing no peaks. Then we will have proved CR as above. The argument uses multisets as defined in the Appendix. To a landscape $d_0 \leftrightarrow \cdots \leftrightarrow d_n$ we can associate the multiset $[d_0, ..., d_n]$. As $\rightarrow$ is SN, $\leftarrow$ is WF, and hence $\leftarrow^+$ is a well-founded order. By Lemma A.3(ii), the corresponding multiset order $\leftarrow^+_{\#}$ is also well founded. Now we observe that in the procedure of eliminating peaks the multisets associated to landscapes descend in the sense of this multiset order. For example, $[a_0, ..., a_{i-1}, c_1, ..., d, ..., c'_1, a_{i+1}, ..., a_n]$ originates from $[a_0, ..., a_n]$ by replacing a_i by the multiset $[c_1, ..., d, ..., c'_1]$ of strictly smaller elements. It follows that the procedure of eliminating peaks must terminate. ∎

3. REDUCTION DIAGRAMS

Consider the ARS $\mathscr{A} = (A, \{\rightarrow_\alpha \mid \alpha \in I\})$, which will be fixed throughout this section. A geometric tool for finding a common reduct of the end points of two diverging indexed reduction sequences is the use of reduction diagrams. Reduction diagrams are built up from so-called elementary diagrams, see the examples in Fig. 2. A completed reduction diagram contains the desired common reduct.

An elementary diagram is a scalable rectangle with vertices representing elements of A and edges representing indexed reduction sequences. The upper edge represents zero or one reduction steps from left to right. The left edge represents zero or one reduction steps downwards. The lower edge represents an indexed reduction

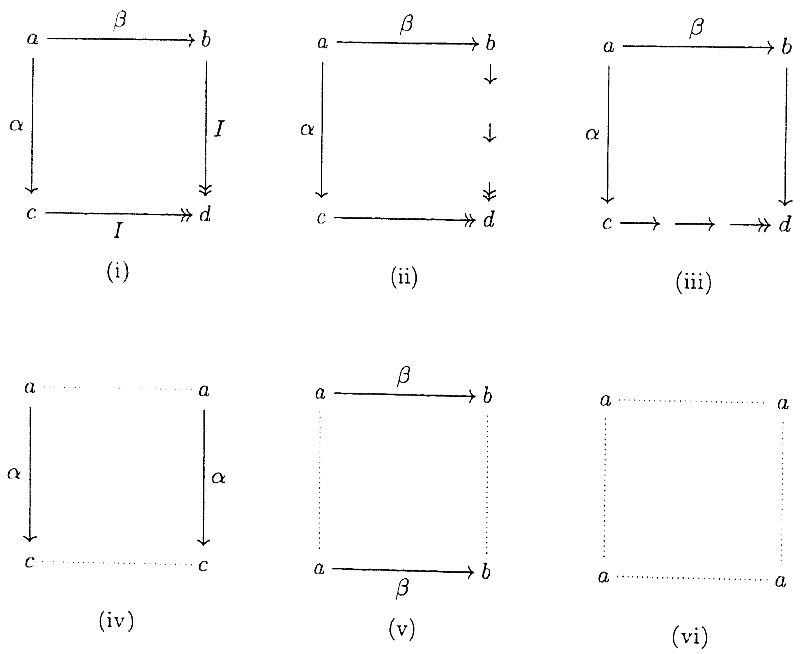


FIG. 2. Elementary diagrams.

sequence of zero or more steps from left to right. The right edge represents an indexed reduction sequence of zero or more steps downwards. Not all combinations are allowed. As a general rule we require that if the left and/or the upper edge represent zero reductions steps, then the opposite edge(s) must also represent zero reductions steps and the adjacent edges must represent at most one step.

To keep the diagrams orthogonal, opposite edges can be scaled to any convenient length, irrespective of the number of reduction steps represented by that edge. In particular edges representing zero reduction steps have positive length, but they will be depicted by dotted lines for clarity. In indexed reduction sequences, such "empty steps" will be made visible by $\rightarrow_\varepsilon$, with $\varepsilon \notin I$. Recall that finite indexed reduction sequences are denoted by $a \twoheadrightarrow_I b$. Now it is time for a formal definition.

3.1. DEFINITION (elementary diagrams).

(i) A *proper elementary diagram* is a configuration as depicted in Fig. 2 under (i), which includes the cases (ii) and (iii).

(ii) A *trivial elementary diagram* is a configuration as depicted in Fig. 2 under (iv), (v), or (vi).

(iii) A *splitting elementary diagram* is a proper elementary diagram with at least one edge representing at least two reduction steps, that is, a configuration as depicted in Fig. 2 under (ii) or (iii). The reduction steps on edges representing at least two reduction steps are called *splitting steps*, and the intermediate points are called *splitting points*.

The *origin* of an elementary diagram is the upper left corner point. The *diagonal* of an elementary diagram is the diagonal through the origin.

Elementary diagrams are used as tiles to construct diagrams by adjoining elementary diagrams in inner corners of the borderline, see Fig. 3. Sometimes we will abstract from the elements labeling the corner points, from the indices of the reduction relations labeling the edges, and from the direction of the arrows: the arrows are always pointing from left to right or downward. The abstracted diagrams are also called diagrams or tilings. The rightmost diagrams in Fig. 3 and 6 are tilings as well as the three elementary tilings in Fig. 4. A beautiful example of a tiling is provided by the fractal-like Fig. 5, constructed from the elementary tilings in Fig. 4 (filling in all edges for aesthetic reasons). Interestingly, Fig. 5 can already

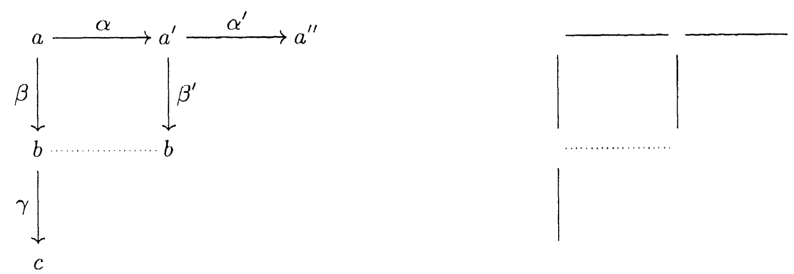


FIG. 3. A diagram with borderline $c \leftarrow_\gamma b \rightarrow_\varepsilon b \leftarrow_{\beta'} a' \rightarrow_{\alpha'} a''$ and inner corners $c \leftarrow_\gamma b \rightarrow_\varepsilon b$ and $b \leftarrow_{\beta'} a' \rightarrow_{\alpha'} a''$, and the corresponding tiling.

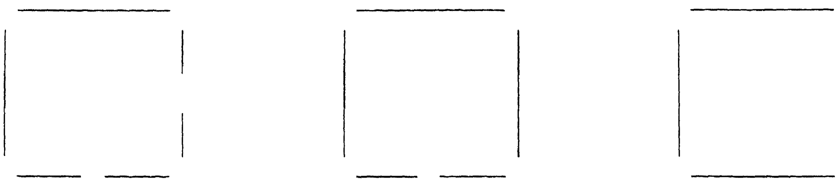


FIG. 4. Three elementary tilings used in Fig. 5.

be obtained with the leftmost elementary diagram (suitably filling in right and lower edges). The middle and rightmost elementary diagrams in Fig. 4 are sufficient for the upper (and the left) part of Fig. 5, both with one limit point, whereas the whole diagram has a borderline of limit points. The upper (and the left) part play a role in the sequel as they are the graphical representation of the classical counter-example against WCR ⇒ CR in Example 4.4. Figure 5 also serves as an example of how diagrams are scaled to accomodate adjoining. Of course, when depicting a tiling we always tacitly assume that it has been obtained by abstraction from a reduction diagram which is correct in the sense of Definition 3.2. The aim of the construction process is to obtain a completed reduction diagram as in Fig. 6.

3.2. Definition. Let $\sigma: a \twoheadrightarrow_I b$ and $\tau: a \twoheadrightarrow_I c$ be two indexed reduction sequences of a. Then σ and τ are the *spanning edges* of a class of σ, τ-reduction

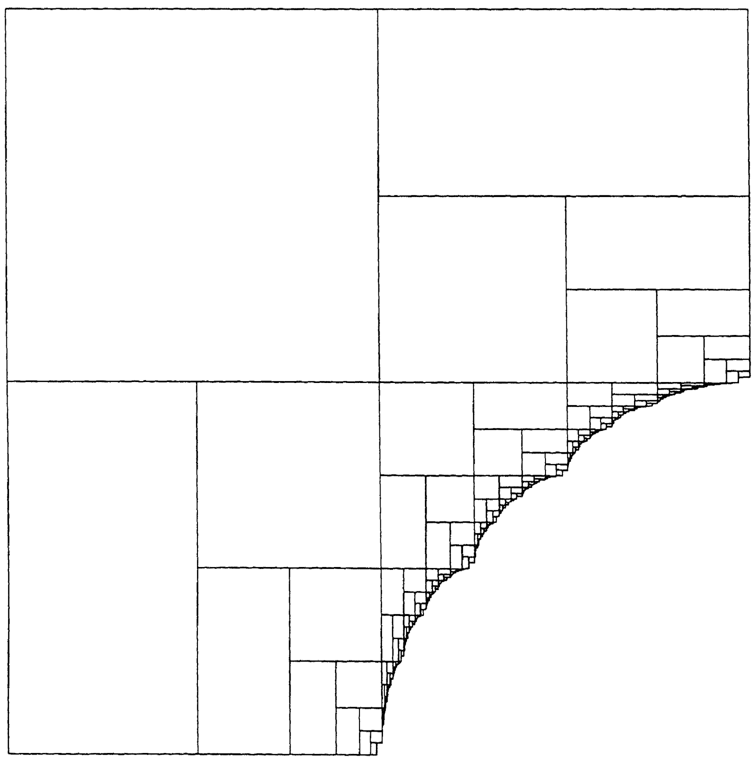


FIG. 5. Fractal- or Escher-like figure.

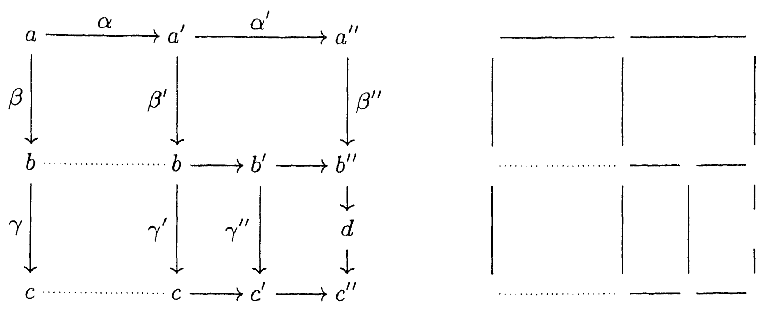


FIG. 6. A completed diagram and the corresponding tiling.

diagrams, simply called *diagrams* when σ and τ are clear from the context. The corner point a will be called the *origin* of these diagrams. The class of σ, τ-reduction diagrams will be inductively defined. Along the way we will also define the notions *borderline* and *inner corner* of a diagram, see Fig. 3.

(i) The empty diagram with spanning edges $c \twoheadleftarrow_I a \twoheadrightarrow_I b$ is a diagram with the spanning edges themselves as borderline.

(ii) If D is a diagram with borderline $c \equiv d_0 \leftrightarrow \cdots \leftrightarrow d_n \equiv b$ having inner corner $d_{i-1} \leftarrow d_i \rightarrow d_{i+1}$, then we distinguish the following four cases. Every extension of D obtained in any of the applicable cases is also a diagram. Here and below *fitting* means that the spanning edges of the elementary diagram that is adjoined to the diagram are identical to the inner corner and that the elementary diagram is scaled to the right size.

(1) If $d_{i-1} \leftarrow_\alpha d_i \rightarrow_\beta d_{i+1}$ with α, $\beta \in I$, then we extend D with a fitting elementary diagram of type (i) from Fig. 2 and change the borderline by replacing $d_{i-1} \leftarrow_\alpha d_i \rightarrow_\beta d_{i+1}$ by $d_{i-1} \twoheadrightarrow_I d_i \twoheadleftarrow_I d_{i+1}$; that is, the reduction steps represented by the left and upper edges are replaced by the indexed reduction sequences represented by the lower and right edges of the elementary diagram.

(2) If $d_{i-1} \leftarrow_\alpha d_i \rightarrow_\varepsilon d_{i+1}$ with $\alpha \in I$, then $d_{i+1} \equiv d_i$ and we extend D with a fitting elementary diagram of type (iv) from Fig. 2 and change the borderline accordingly.

(3) If $d_{i_1} \leftarrow_\varepsilon d_i \rightarrow_\beta d_{i+1}$ with $\beta \in I$, then $d_{i-1} \equiv d_i$ and we extend D with a fitting elementary diagram of type (v) from Fig. 2 and change the borderline accordingly.

(4) If $d_{i-1} \leftarrow_\varepsilon d_i \rightarrow_\varepsilon d_{i+1}$, then $d_{i-1} \equiv d_i \equiv d_{i+1}$ and we extend D with a fitting elementary diagram of type (vi) from Fig. 2 and change the borderline accordingly.

A *proper* inner corner is an inner corner of two nonempty steps, that is case (ii1) above. A diagram is *completed* if the borderline is of the form $c \twoheadrightarrow_I d \twoheadleftarrow_I b$ for suitable d; that is, without inner corners. The lower edge $c \twoheadrightarrow_I d$ and the right edge $b \twoheadrightarrow_I d$ are called the *completing edges*.

Intuitively, infinite diagrams are obtained by applying the generating rules (1)–(4) under (ii) in Definition 3.2 infinitely many times. Any inductive definition only allows for finite successive application of the generating rules, so that Definition 3.2 gives us only finite diagrams. It is well known that infinite sequences can be obtained as limits of converging sequences of finite sequences, using an appropriate metric for finite sequences: $d(s, s') = 2^{-k}$, where k is the maximum length of a common initial subsequence of s and s'. The resulting metric space of (finite or infinite) sequences is the metric completion of the metric space of finite sequences. See Sutherland (1976) for general information on metric spaces. In a similar way we can obtain infinite σ, τ-diagrams as limits of converging sequences of finite σ, τ-diagrams, the metric being $d(D, D') = 2^{-k}$, where k is the minimal length of a reduction sequence from the origin to an elementary diagram on which D and D' differ. (This is analogous to the maximum length of a common initial subsequence above.) Here we also count empty steps, and we take $d(D, D') = 0$ if D and D' coincide. As a converging sequence of finite diagrams whose limit is an infinite diagram one can take the successive stages in the construction of the infinite diagram. We stress the fact that infinite diagrams thus defined have finite spanning edges. We could have defined a larger class of infinite diagrams including those with infinite spanning edges, but the latter are less relevant for confluence. From now on we will work with finite as well as with infinite diagrams. We will always stipulate when a diagram is infinite and reserve the term "diagram" for finite diagrams. We shall now develop some theory about diagrams which seems to be of independent interest.

3.3. Definition. Let ED be a set of elementary diagrams. We will always assume that ED contains all trivial elementary diagrams. Moreover, ED is *full* if for every proper inner corner there exists a fitting elementary diagram in ED.

3.4. Example. Assume $\rightarrow_I$ is WCR. Let ED consist of all possible proper elementary diagrams and all trivial elementary diagrams. Then ED is full.

3.5. Definition. Diagrams can be ordered as follows. Let D and D' be finite or infinite diagrams. Then $D \sqsubseteq D'$ if and only if D is an initial subdiagram of D', that is, D fits on D' with coinciding origins. If $D \sqsubseteq D'$, then D' is called an extension of D. Obviously, $\sqsubseteq$ is a partial order.

3.6. Lemma. *Let ED be a set of elementary diagrams and consider finite or infinite σ, τ-diagrams built from these. Every diagram has a maximal extension with respect to $\sqsubseteq$.*

Proof. Let S be the set of finite or infinite σ, τ-diagrams and consider $D \in S$. Let $\hat{D} = \{D' \in S \mid D \sqsubseteq D'\}$ and let C be a chain in $\hat{D}$. Since C is totally ordered by $\sqsubseteq$, all diagrams in C fit when they are laid over each other with coinciding origins. Consider the figure F that arises when all diagrams from C are laid over each other with coinciding origins. If F is a finite or infinite diagram, then F is obviously an upper bound of C in $\hat{D}$. Now $\hat{D}$ contains a maximal element by Zorn's Lemma, which is the desired maximal extension of D. So it remains to prove that F is a finite or infinite diagram. Consider the empty diagram with spanning edges σ, τ. If there is one element in C with an elementary diagram in the inner corner at the origin, then this elementary diagram is unique since C is totally ordered. Adjoin this elementary diagram to the empty diagram, then the new diagram has two inner corners, which are treated in a similar way as above. This process yields a sequence of finite diagrams of which F is the limit. ∎

3.7. LEMMA. *Let* ED *be a full set of elementary diagrams and consider finite or infinite diagrams built from these. If a maximal diagram is finite, then it is completed.*

Proof. Since ED is full, any maximal finite diagram with spanning edges $c \twoheadleftarrow_I a \twoheadrightarrow_I b$ has a borderline without inner corners of the form $c \twoheadrightarrow_I d \twoheadleftarrow_I b$ for suitable d. ∎

3.8. LEMMA. *Every infinite diagram contains a reduction sequence with infinitely many horizontal splitting steps and infinitely many vertical splitting steps, in particular with infinitely many nonempty steps.*

Proof. Recall that infinite diagrams have finite spanning edges by definition. The lemma would be false if infinite spanning edges would be allowed (for example, if all steps are empty). Let D be an infinite reduction diagram with origin o. We shall prove that D contains an infinite subdiagram whose origin can be reached from o via a reduction sequence containing at least one vertical splitting step. This suffices for the lemma, since we can obtain the same result with a horizontal splitting step instead of a vertical splitting step by symmetry, and we can alternate the two versions infinitely many times to obtain the desired "meandering" reduction sequence. In the sequel we treat empty steps in the diagram just as any other nonsplitting step.

Observe first that, with the infinite diagram D as depicted in Fig. 7 (left), at least one of the points o' and o'' is origin of an infinite subdiagram of D, since the areas H and V, separated by the dotted line, cannot both be finite. More precisely, area H has finite spanning edges and is hence a (finite or infinite) subdiagram with origin o'. If the dotted line is infinite, then area H is an infinite subdiagram with origin o'. Otherwise, if the dotted line is finite, then it can be used as a spanning edge and the area V is also a (finite or infinite) subdiagram with origin o''. As H

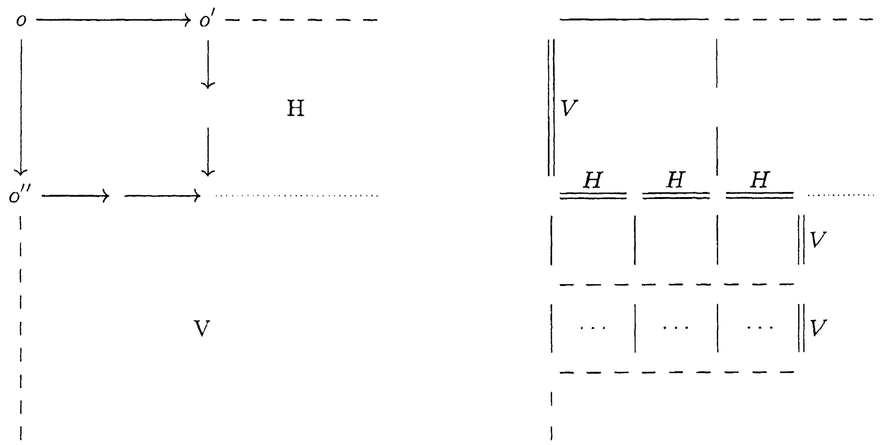


FIG. 7. Infinite subdiagrams and projection of V-steps.

and V cannot both be finite diagrams, at least one of them is an infinite subdiagram of D. Let us indicate the two possibilities (not necessarily excluding each other) of moving from o to the origin of an infinite subdiagram by H(orizontal) and V(ertical). So $o \to o'$ is an H-step if o' is the origin of an infinite subdiagram, and $o \to o''$ is a V-step if o'' is the origin of an infinite subdiagram. By the argument above sequences of H- and V-steps can always be extended and come together with sequences of nested infinite subdiagrams. See Fig. 7 (right), for an example, where we represent the successive H- and V-steps by double lines.

Let π be an infinite sequence of H- and V-steps in D as above, so every point on π is the origin of an infinite subdiagram. Since infinite diagrams have finite spanning edges, π does not contain an infinite sequence of H-steps (nor of V-steps, but we do not need this). We must prove that D contains an infinite subdiagram whose origin can be reached from o via a reduction sequence containing at least one vertical splitting step. We claim that π contains such an origin, and prove this by contradiction: assume no point on π can be reached from o via a reduction sequence containing at least one vertical splitting step. By this assumption of every succession in π of an H-step immediately followed by a V-step, the H-step forms the upper and the V-step forms the right (non splitting) edge of an elementary diagram in the original infinite diagram, whose left edge is again non splitting (see Fig. 7 (right), the lower edge of this elementary diagram may contain splitting steps, but this does no harm). Starting with the first HV-succession in π, we can iterate this argument until we arrive at the vertical spanning edge of the original infinite diagram (see Fig. 7 (right)). So the V-step of the first HV-succession is, so to say, projected on the vertical spanning edge of the original infinite diagram. All V-steps can be projected in this way, in order of their appearance in π. As the original infinite diagram has finite spanning edges, π can only contain finitely many V-steps. This conflicts with π being infinite and not containing an infinite sequence of H-steps. ▮

3.9. COROLLARY. *If* $\to_I$ *is* SN, *then all reduction diagrams are finite. Moreover we have Newman's Lemma* 2.6: *if* $\to_I$ *is* SN *and* WCR, *then* $\to_I$ *is* CR.

Proof. The first statement follows by contraposition of Lemma 3.8. For the second, assume $\to_I$ is SN and WCR. By WCR the set ED of all possible elementary diagrams is full (see Example 3.4). Let D be the empty diagram with spanning edges $c \twoheadleftarrow_I a \twoheadrightarrow_I b$. By Lemma 3.6, D has a maximal extension D' which is finite by SN. By Lemma 3.7, D' is completed. So every empty diagram can be completed, or, in other words, $\to_I$ is CR. ▮

4. CONFLUENCE BY DECREASING DIAGRAMS

In this section we present a powerful criterion for confluence of ARSs. The method, developed by van Oostrom (1994a, 1994b) and called "confluence by decreasing diagrams," generalizes several well-known confluence criteria such as Newman's Lemma 2.6, Lemma 3.1 of Winkler and Buchberger, which we generalized to Lemma 4.3, the lemma of Hindley and Rosen (Lemma 4.21), the Request lemmas of Staples (Lemma 4.22), and Huet's Strong Confluence lemma (Lemma 4.29). Actually, van Oostrom's method has been prepared by an unpublished note of De Bruijn (1978), containing a slightly weaker result with a complicated inductive proof, see Lemma 4.30. The notion of decreasing diagrams was not yet present in that note.

To illustrate the use of diagrams for confluence, we start with the following proposition, which is proved by a generalization of the multiset argument (ii) for Newman's Lemma 2.6. Observe that the role of the elements there is taken over by the indices of the reduction steps here.

4.1. Proposition. *Let* $\mathscr{A} = (A, \{\rightarrow_\alpha \mid \alpha \in I\})$ *be an ARS with the index set* I *equipped with a well-founded order* $<$. *Let ED be a full set of elementary diagrams and assume that every proper elementary diagram from ED has the property that the multiset of indices along the two completing edges is smaller than the multiset consisting of the two indices along the spanning edges. Then every diagram can be completed. In particular,* $\rightarrow_I$ *is confluent.*

Proof. Consider a configuration as in Fig. 8 (left), with α, $\beta \in I$. Under the conditions of the proposition, the multiset of indices along the edges σ', τ' is smaller than $[\alpha, \beta]$. Generally, the multiset of indices along the borderline of the diagram decreases when a proper elementary diagram is adjoined. For reasons of space (surface), at most finitely many trivial elementary diagrams can be adjoined one after the other, during which the multiset stays the same. Thereafter the diagram must either be completed or a proper elementary diagram can be adjoined since ED is full. As the multiset order $<_\#$ is well founded by Lemma A.3(ii), it follows that this procedure terminates with a completed diagram. ∎

4.2. *Remark.* In fact the proposition above holds irrespective of the direction of reduction steps in the completing edges of elementary diagrams; that is, we can allow those steps to be in $\leftrightarrow_I$ instead of $\rightarrow_I$. Such diagrams can tentatively be called elementary *conversion* diagrams, depicted in Fig. 8 (right).

The proposition can be further sharpened by allowing nonsplitting elementary diagrams in ED which are *stationary*, that is, corresponding to reductions

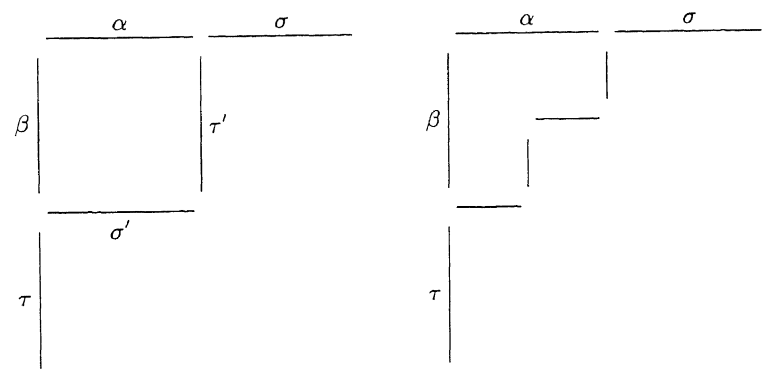


FIG. 8. Adjoining an elementary (conversion) diagram.

$d \leftarrow_\gamma c \leftarrow_\alpha a \rightarrow_\beta b \rightarrow_\delta d$ with $[\alpha, \beta] = [\gamma, \delta]$. The surface argument handling the trivial diagrams can also handle *stationary* nonsplitting elementary diagrams.

4.3. LEMMA (Lemma 3.1 of Winkler and Buchberger (1985), generalized). *Let $\mathscr{A} = (A, \rightarrow)$ be an* ARS, *B a set with well-founded order $\prec$ and $f\colon A \rightarrow B$ a function. We say that a and b are* connected below c *(w.r.t. $B, \prec, f$), denoted by $a \overset{c}{\leftrightarrow} b$, if there is a conversion $a \equiv a_0 \leftrightarrow \cdots \leftrightarrow a_n \equiv b$ such that $f(a_i) \prec f(c)$ for all $0 < i < n$. Furthermore, $\mathscr{A}$ is called* connected *if $\forall a, b, c \in A$ $(a \leftarrow c \rightarrow b \Rightarrow a \overset{c}{\leftrightarrow} b)$. We call $\mathscr{A}$* weakly connected *if, for all $a, b, c \in A$, $a \leftarrow c \rightarrow b$ implies either $a \overset{c}{\leftrightarrow} b$ or there exists $d \in A$ such that $a \rightarrow d \leftarrow b$ and $f(d) \preccurlyeq f(c)$. Then we have:*

(i) *Every connected* ARS *is confluent.*

(ii) *Every weakly connected* ARS *is confluent.*

Proof. The idea is to assign as index to any reduction step $a \rightarrow b$ the multiset $[f(a), f(b)]$ and to use Proposition 4.1, sharpened by Remark 4.2. Consider the ARS $\mathscr{A}^{\#} = (A, \{\rightarrow_\alpha \mid \alpha \in I\})$, with reduction relations $\rightarrow_\alpha$ defined by $a \rightarrow_\alpha b$ if and only if $a \rightarrow b$ and $\alpha = [f(a), f(b)]$. Then $\rightarrow = \bigcup \{\rightarrow_\alpha \mid \alpha \in I\} = \rightarrow_I$.

(i) Since $\mathscr{A}$ is connected it follows that we have a full set of elementary diagrams. Moreover, the decreasing condition in Proposition 8 in the version with completing steps in $\leftrightarrow_I$ is satisfied since the multiset associated to $a_0 \leftrightarrow \cdots \leftrightarrow a_n$,

$$[[f(a_0), f(a_1)], [f(a_1), f(a_2)], \ldots, [f(a_{n-1}), f(a_n)]],$$

is smaller than $[[f(a_0), f(c)], [f(c), f(a_n)]]$ when $f(a_i) \prec f(c)$ for all $0 < i < n$. This can be seen by either using the nested multiset order or simply by omitting the inner square brackets and using the ordinary multiset order. In both cases the order is well founded, see Lemma A.3(ii).

(ii) By Remark 4.2, in particular the second paragraph. ∎

4.4. EXAMPLE. The classical counterexample to WCR $\Rightarrow$ CR, the ARS with reductions $a \leftarrow b \leftrightarrows c \rightarrow d$, leads to a full set of elementary diagrams such that not every diagram can be completed. Let ED be the set consisting of the two proper elementary diagrams depicted in Fig. 9 plus their mirror images with respect to the diagonal. Then ED is full. However, the diagram from Fig. 10 cannot be completed, since subdiagrams continue to arise with exactly the same spanning edges as the original diagram. See also the upper (and the left) part of Fig. 5.

$$\begin{tikzcd} b \arrow[r] \arrow[dd] & c \arrow[d] \\ & b \arrow[d] \\ a \arrow[r, dotted, no head] & a \end{tikzcd} \qquad \begin{tikzcd} c \arrow[r] \arrow[dd] & b \arrow[d] \\ & c \arrow[d] \\ d \arrow[r, dotted, no head] & d \end{tikzcd}$$

FIG. 9. Elementary diagrams not giving confluence.

$$\begin{tikzcd}
b \arrow[r] \arrow[dd] & c \arrow[rr] \arrow[d] & & d \arrow[d, dotted, no head] \\
& b \arrow[r] \arrow[d] & c \arrow[r] & d \\
a \arrow[r, dotted, no head] & a & &
\end{tikzcd}$$

FIG. 10. A diagram which cannot be completed.

4.5. *Notation.* In the sequel we will consider a fixed ARS $\mathscr{A} = (A, \{\to_\alpha \mid \alpha \in I\})$, with the index set I equipped with a well-founded order $<$. In the examples, we will take for I the set of natural numbers $\mathbb{N}$ with the usual order $<$, unless explicitly stated otherwise. For $a, b \in A$ and $\alpha \in I$, let $a \twoheadrightarrow_{<\alpha} b$ express that there is an indexed reduction sequence from a to b, each reduction step having index less than α. Analogously, $a \twoheadrightarrow_{\leqslant\alpha} b$ is defined.

4.6. Discussion. The argument in the proof of Proposition 4.1 is prototypical for this section: the construction of completed diagrams is driven by a full set of elementary diagrams, and terminates since the elementary diagrams have an extra property, which ensures a decrease with respect to a well-founded order.

To be able to state what a decreasing diagram is, we need the notion of norm of a finite reduction sequence in the ARS $\mathscr{A}$ with indexed reduction relations. This will be a multiset of indices, but (surprise!) not all indices along the reduction sequence. These multisets are obtained from indexed reduction sequences by three successive operations, called *index*, *filter*, *multiset*, in order of application.

The first operation extracts the sequence of indices from a given reduction sequence. If σ is an indexed reduction sequence, $index(\sigma)$ is the sequence of the indexes of the consecutive nonempty reduction steps in σ. For example,

$$index(a \to_3 b \to_\varepsilon b \to_2 c \to_4 d \to_4 e) = \langle 3, 2, 4, 4 \rangle.$$

Recall that we use the natural numbers as a running example, but everything will be generalized to an arbitrary well-founded partial order $<$ on I. We will allow ourselves a slight abuse of notation, by denoting both finite sequences of indices and finite *reduction* sequences with σ, τ. Often, we identify a reduction step with its index. Also, $\langle \alpha \rangle \cdot \sigma$ is used for a reduction sequence starting with an α-step followed by the reduction sequence σ. If $\alpha > \alpha'$, then we say that α *majorizes* α'.

Given a finite sequence σ of natural numbers, $filter(\sigma)$ is the finite sequence obtained by processing σ from left-to-right, removing the elements from σ that are majorized by some previous element. For example,

$$filter(\langle 3, 2, 4, 4, 3, 1, 2, 6, 2, 8, 7, 8, 4, 2, 5 \rangle) = \langle 3, 4, 4, 6, 8, 8 \rangle.$$

Thus $filter(\sigma)$ is always a nondecreasing finite sequence.

The last operation on finite sequences is *multiset*, yielding the multiset corresponding to the finite sequence. For example,

$$multiset(\langle 3, 4, 4, 6, 8, 8\rangle) = [3, 4, 4, 6, 8, 8,].$$

In the following we are especially interested in $multiset(filter(index(\sigma)))$. Now we give the formal definitions of *index*, *filter*, and multiset.

4.7. DEFINITION. Let R be the set of finite indexed reduction sequences. The mapping *index*: $R \to I^*$ is defined by $index(a \to_\varepsilon a) = \langle\ \rangle$, and, for $\alpha \neq \varepsilon$,

$$index(a \to_\alpha b) = \langle \alpha \rangle, \; index(\sigma \cdot \tau) = index(\sigma) \cdot index(\tau).$$

The mapping *filter*: $I^* \to I^*$ is defined by $filter(\langle\ \rangle) = \langle\ \rangle$ and

$$filter(\sigma \cdot \langle \alpha \rangle) = \begin{cases} filter(\sigma) & \text{if } filter(\sigma) \text{ contains an element } \alpha' > \alpha \\ filter(\sigma) \cdot \langle \alpha \rangle & \text{otherwise.} \end{cases}$$

The mapping *multiset*: $I^* \to I^{\#}$ is defined by

$$multiset(\langle \alpha_0, ..., \alpha_{k-1} \rangle) = [\alpha_0, ..., \alpha_{k-1}].$$

The following observation is important: *filter* does not distribute over concatenation of finite sequences. For example, we have

$$\begin{aligned} filter(\langle 1 \rangle \cdot \langle 0, 2 \rangle) &= filter(\langle 1, 0, 2 \rangle) = \langle 1, 2 \rangle \\ &\neq \langle 1 \rangle \cdot \langle 0, 2 \rangle = filter(\langle 1 \rangle) \cdot filter(\langle 0, 2 \rangle). \end{aligned}$$

4.8. DEFINITION (norm). (i) The *norm* $|\sigma|$ of an indexed reduction sequence σ is

$$|\sigma| = multiset(filter(index(\sigma))).$$

(ii) The *norm* $|D|$ *of a diagram* D is $|D| = |\sigma| \uplus_{\#} |\tau|$, where σ and τ are the spanning edges of D.

4.9. DEFINITION (decreasing diagram). Let D be a completed diagram with spanning edges σ and τ, right edge τ', and lower edge σ' (see Fig. 11). Then D is a *decreasing diagram*, if

$$|\sigma \cdot \tau'| \leqslant_{\#} |D| \qquad \text{and} \qquad |\tau \cdot \sigma'| \leqslant_{\#} |D|.$$

Here $\leqslant_{\#}$ denotes the reflexive closure of the multiset extension $<_{\#}$ of $<$ on I.

4.10. *Discussion.* The main line of the argumentation will be as follows. First prove that completed diagrams that are built from decreasing elementary diagrams are decreasing. Then prove that, given an empty diagram, the procedure of completing the diagram terminates, using the fact that all completed diagrams

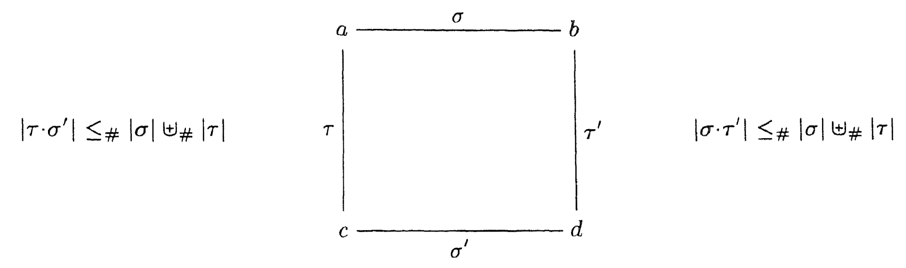


FIG. 11. A decreasing diagram.

involved are decreasing. It may come as a surprise that the inequalities in Fig. 11 are not strict and yet termination is guaranteed. The reason is that the norm according to Proposition 4.18 nevertheless strictly decreases when a decreasing elementary diagram is adjoined.

Decreasing diagrams are a useful technical device, but they have some properties that make them hard to understand. For example, the property of being decreasing is not preserved under extension of the order. Figure 12 presents an example of a diagram which is decreasing with respect to the empty order (then $|D| = [0, 2, 1] = |\langle 2, 1, 0\rangle| = |\langle 0, 1, 2\rangle|$), but not decreasing with respect to the usual order $<$ on natural numbers (then $|D| = [0, 2]$ and $|\langle 0, 1, 2\rangle| = [0, 1, 2]$, so $|D| < [0, 1, 2]$). Also, this diagram cannot be built up from elementary diagrams that are decreasing with respect to the empty order.

We therefore define a slightly stronger notion of decreasing, which we will call trace-decreasing. Lemma 4.14 states that trace-decreasing implies decreasing, but not vice versa. The property of being trace-decreasing is preserved under extension of the order. Moreover, if the order is total, then the two notions coincide, see Lemma 4.15. Also, the two notions coincide for elementary diagrams, see Lemma 4.16. The notion of trace-decreasing may be more cumbersome to formulate than that of decreasing, but it has a clearer visualization (see for example Fig. 13, with the usual order on $\mathbb{N}$ and $n < 8$).

4.11. Definition (trace-decreasing). Let D be a completed diagram with spanning edges σ and τ and right edge τ' and lower edge σ' (see Fig. 11). The edges τ' and σ' are dealt with symmetrically, so we restrict attention to the first, only indicating the symmetrical case between parentheses. D is a *trace-decreasing diagram* if there exists a tracing map t (tracing map s) mapping every nonempty step in τ' (σ') to a nonempty step in σ or τ such that the conditions (i)–(iii) below hold. Steps that are related by the tracing maps are connected by *traces*. A *trace* is a full line or a dashed line. Traces from τ' to τ (from σ' to σ) are called *horizontal* (*vertical*).

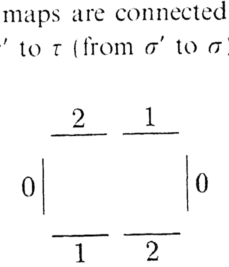


FIG. 12. A diagram which is not decreasing when $0 < 1 < 2$.

Traces from τ' to σ (or from σ' to τ) are called *diagonal*. We say that every step α' in τ' *traces back* to the step $\alpha = t(\alpha')$. In that case α is called the *ancestor* of α'. If a step α' in τ' traces back to a step $\alpha = t(\alpha')$ in σ or in τ, then it is required that $index(\alpha') \leqslant index(\alpha)$. If $index(\alpha') < index(\alpha)$, then α' and α are connected by a dashed line. If $index(\alpha') = index(\alpha)$, then α' and α are connected by a full line. Symmetrically for s.

(i) Any horizontal or vertical trace can be either a full or a dashed line. Diagonal traces can only be dashed lines.

(ii) Horizontal full lines do not cross or split. In other words, if α' and β' are steps in τ' tracing back to $\alpha = t(\alpha')$ and $\beta = t(\beta')$ in τ by full lines, and α' occurs (strictly) before β' in τ', then α occurs (strictly) before β in τ. Symmetrically for the vertical case.

(iii) Horizontal full lines do not cross or split from left below to right up by a dashed line. In other words, if α' and β' are steps in τ' tracing back to $\alpha = t(\alpha')$ and $\beta = t(\beta')$ in τ by a dashed line and a full line, respectively, and α' occurs (strictly) before β' in τ', then α occurs (strictly) before β in τ. Symmetrically for the vertical case.

See Fig. 13 for an example of a trace-decreasing diagram. Omitting the trivial parallel cases, the notion of a trace-decreasing diagram can conveniently be described by distinguishing between allowed and forbidden configurations of the traces, namely those in Fig. 14 and 15, respectively.

4.12. *Remarks.* The intuition behind the forbidden and allowed configurations can be guided by the multiset inequalities in Fig. 11, providing necessary but not sufficient conditions according to Lemma 4.14. Splitting full lines is forbidden since one step on the left edge cannot be used to cancel two steps on the right edge. A splitting full line and a dashed line as in Fig. 15c is also forbidden since the point on the left edge would become overloaded. Similarly for diagonal full lines. The forbidden crossing situations can be explained by observing that they could give rise to forbidden splitting configurations when diagrams are adjoined. Note that

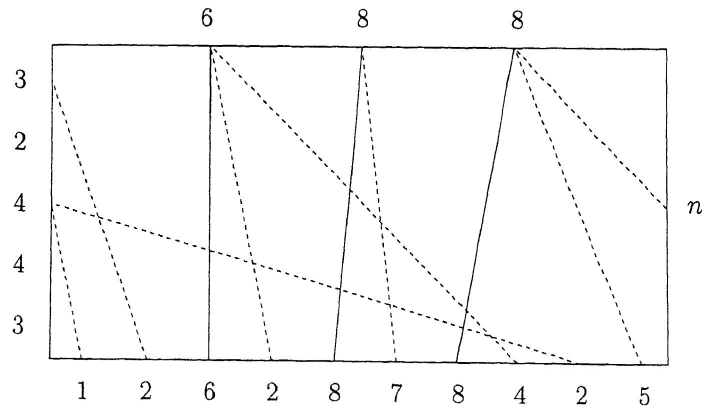


FIG. 13. Example of a trace-decreasing diagram ($n < 8$).

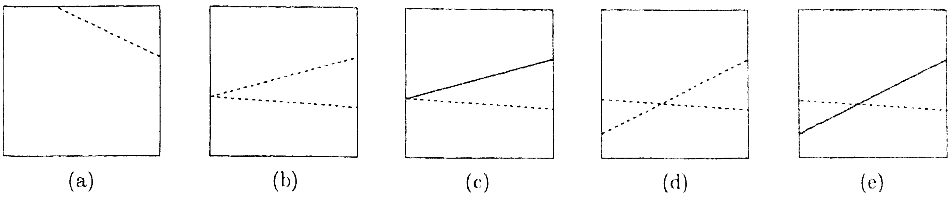


FIG. 14. Allowed configurations (one symmetrical half).

configurations as in Fig. 14a, c (and also their vertical variants) are allowed. The reason is that the majorized element on the right edge is filtered out.

Decreasing diagrams could also be visualized along the above lines. Crossing configurations are allowed in the case of decreasing diagrams. Forbidden splitting configurations that could arise when diagrams are adjoined are avoided by requiring that all steps on the left (upper) edge that are endpoints of traces belong to $|\tau|$ ($|\sigma|$) instead of $\tau(\sigma)$. This is exactly the reason why the diagram in Fig. 12 is not decreasing when $0 < 1 < 2$: the 1 on the upper edge is filtered out and therefore can not be used.

One easily verifies in the definition above that trace-decreasing is preserved under extension of the order, since *filter* is avoided.

4.13. Examples. In Figs. 16 and 17, where $0 < 1$, we give examples of decreasing and nondecreasing elementary diagrams, respectively. We leave it as an exercise to the reader to reconstruct in the diagrams of Fig. 16 tracing maps according to Definition 4.11; there is often more than one such tracing map. The general form of a trace-decreasing elementary diagram is depicted in Fig. 18, whose justification can be drawn from the proof of Lemma 4.16.

4.14. Lemma. *Every trace-decreasing diagram is decreasing, but not conversely.*

Proof. Let D be a trace-decreasing diagram. We have to prove that the edges of D satisfy the multiset inequalities as given Fig. 11. The edges τ' and σ' are dealt with symmetrically, so we restrict our attention to the first. We have $|\sigma \cdot \tau'| = |\sigma| \uplus_{\#} |\tau'| -_{\#} M$, where M consists of all elements from $|\tau'|$ that are majorized by some element from σ. To prove that $|\sigma \cdot \tau'| \leqslant_{\#} |\sigma| \uplus_{\#} |\tau|$, we first cancel left and right $|\sigma|$ using Lemma A.3(v). It remains to prove that $|\tau'| -_{\#} M \leqslant_{\#} |\tau|$. This will be done in the proof of Proposition 4.18, in the form of the inequality $M_{\tau'}^{\ast\sigma} \leqslant_{\#} |\tau|$.

A counterexample to the converse is provided in Fig. 12. As argued in Remark 4.12, this diagram is decreasing with the empty order. It is not trace-decreasing since the traces must cross. ∎

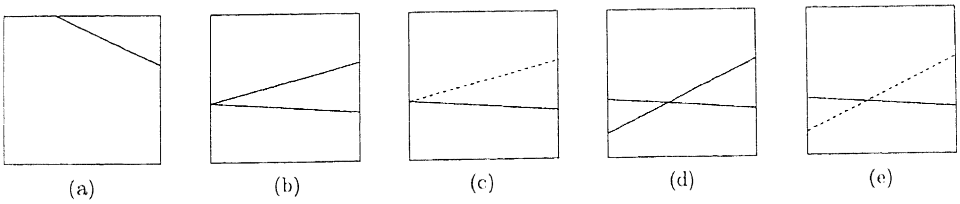


FIG. 15. Forbidden configurations (one symmetrical half).

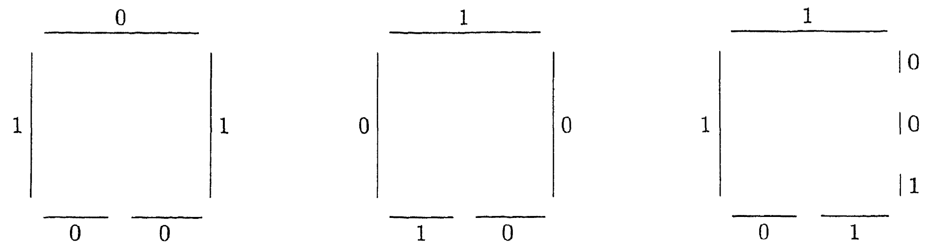


FIG. 16. (Trace-)decreasing elementary diagrams.

4.15. LEMMA. *Every decreasing diagram is trace-decreasing if the order is total.*

Proof. Assume $<$ is total and let D be a decreasing diagram as depicted in Fig. 11. The edges τ' and σ' are dealt with symmetrically, so we restrict attention to the first. We have $|\sigma \cdot \tau'| = |\sigma| \uplus_{\#} |\tau'| -_{\#} M$, where M consists of all elements from $|\tau'|$ that are majorized by some element from σ. From $|\sigma \cdot \tau'| \leqslant_{\#} |\sigma| \uplus_{\#} |\tau|$ then follows $|\tau'| -_{\#} M \leqslant_{\#} |\tau|$ by canceling left and right $|\sigma|$ using Lemma A.3(v). In the rest of this proof we will use that *filter* yields weakly increasing sequences as the order is total. We assume that this sorting is maintained in the multisets $|\tau|$ and $|\tau'|$. Then M is an initial segment of $|\tau'|$. If $M = |\tau'|$, then we are done by connecting all steps in τ' to majorizing steps in σ by dashed lines. Otherwise, let F be the largest common final segment of $|\tau|$ and $|\tau'| -_{\#} M$. Then either $|\tau'| -_{\#} M = |\tau| = F$, or there exist multisets I, I' and elements m, m' with $m' < m$ such that $|\tau| = I \uplus_{\#} [m] \uplus_{\#} F$ and $|\tau'| -_{\#} M = I' \uplus_{\#} [m'] \uplus_{\#} F$, where the right hand sides are again assumed to be sorted as weakly increasing sequences. In both cases we give the traces in the obvious way: connect corresponding elements of F by horizontal full lines (of course avoiding crossings) and, in the second case, let every element of $I' \uplus_{\#} [m']$ trace back by a dashed line to m. Finally we must take care of the steps in τ' that do not occur in $|\tau'| -_{\#} M$. These steps fall apart into steps that are majorized by some step in σ (and connected accordingly by a dashed line) and steps that are not majorized by some step in σ, but are majorized by some previous step in $|\tau'|$. The latter steps are connected by a dashed line to the step in σ or τ to which the nearest previous majorizing step in $|\tau'|$ traces back. One easily checks that no forbidden configurations are introduced. This completes the proof that D is trace-decreasing.

4.16. LEMMA. *Every decreasing elementary diagram is trace-decreasing.*

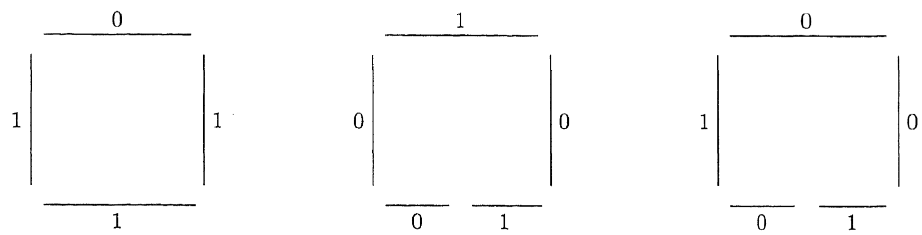


FIG. 17. Elementary diagrams which are not (trace-)decreasing.

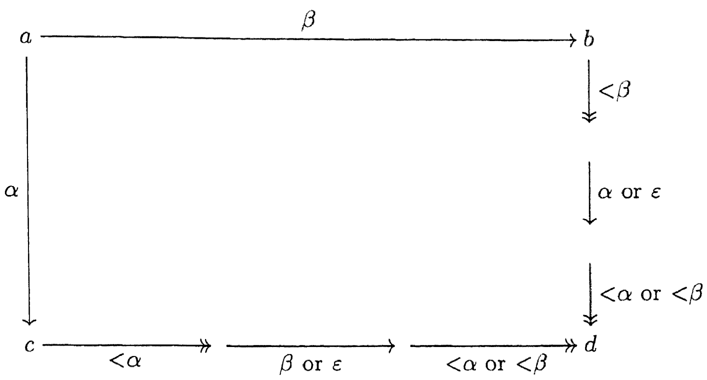
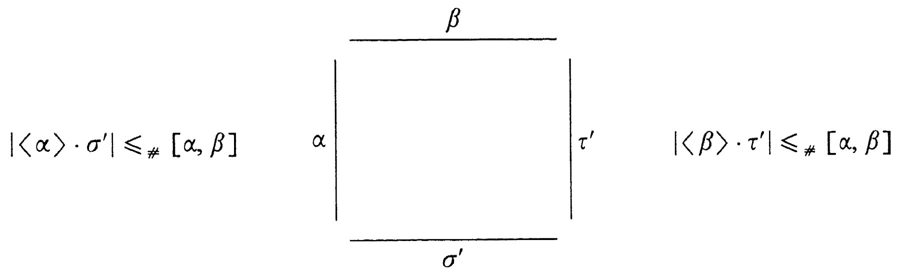


FIG. 18. General form of a (trace-)decreasing elementary diagram.

Proof. Consider a decreasing elementary diagram D as depicted below. Note that $|D| = |\alpha| \uplus_{\#} |\beta| = [\alpha, \beta]$.



The edges τ' and σ' are dealt with symmetrically, so we restrict attention to the first. The multiset $|\langle\beta\rangle \cdot \tau'|$ extends the multiset $[\beta]$ with elements at most α, as $|\langle\beta\rangle \cdot \tau'| \leqslant_{\#} [\beta, \alpha]$. If one of the elements of $|\langle\beta\rangle \cdot \tau'| -_{\#} [\beta]$ equals α, then τ' contains exactly one α-step. In this case the steps in τ' before α are majorized by β (and connected by a dashed line with β), the step α in τ' is connected by a full line with the left edge α, and the steps in τ' after the step α are majorized by either β or α (and are connected by a dashed line with either β or α). In the case that all elements of $|\langle\beta\rangle \cdot \tau'| -_{\#} [\beta]$ are less than α, the steps in τ' are majorized by either α or β (and are connected by a dashed line with either α or β). In both cases we have proved that the elementary diagram is trace-decreasing. ∎

Now we will establish the two important properties of trace-decreasing diagrams that give confluence. The first states that trace-decreasing is preserved under adjoining along fitting edges. The second ensures that adjoining of trace-decreasing diagrams terminates.

4.17. PROPOSITION. *Let D_1, D_2, D_3 be three trace-decreasing diagrams as in Fig. 19. Then the diagrams which result from adjoining D_1 and D_2 along the fitting edge τ', and from adjoining D_1 and D_3 along the fitting edge σ', are trace-decreasing.*

Proof. The proof is simply by checking that no forbidden trace configurations arise by adjoining two trace-decreasing diagrams as indicated. The traces are concatenated in the obvious way: two full lines combine into a full line, two dashed lines

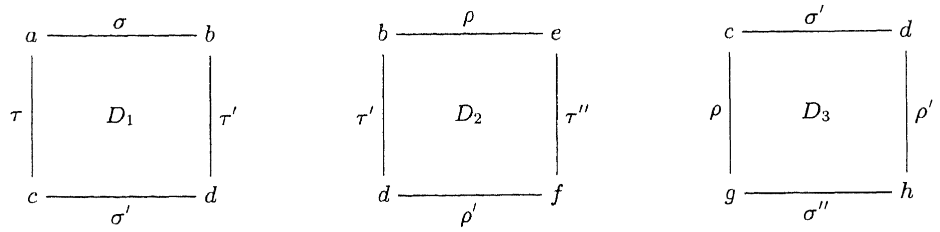


FIG. 19. Diagrams with fitting edges.

as well as a full line and a dashed line combine into a dashed line. In this way, allowed configurations can only yield allowed configurations; see Fig 14. ∎

The second important property is indicated in Fig. 20: adjoining a decreasing diagram to an empty diagram with spanning edges $\sigma \cdot \rho$ and $\tau \cdot \rho'$ yields diagrams with spanning edges τ', ρ and σ', ρ', respectively, having smaller norms.

4.18. PROPOSITION. *Let the trace-decreasing diagram D with spanning edges σ, τ and completing edges σ', τ' be adjoined to an empty diagram as in Fig. 20. Assume that σ and τ contain both at least one nonempty step. Let $D'_\varnothing$ be the empty diagram with spanning edges $\sigma \cdot \rho$ and $\tau \cdot \rho'$. Then $|\rho| \uplus_\# |\tau'| <_\# |D'_\varnothing|$ and $|\sigma'| \uplus_\# |\rho'| <_\# |D'_\varnothing|$.*

Proof. Both cases are dealt with symmetrically, so we restrict our attention to one. Since $|\tau| \leqslant_\# |\tau \cdot \rho'|$, it suffices to prove $|\tau'| \uplus_\# |\rho| <_\# |\tau| \uplus_\# |\sigma \cdot \rho|$. Observe that elements of $|\sigma|$ may majorize elements in $|\rho|$ as well as elements in $|\tau'|$. It is convenient to single out these elements. We write $|\rho| = M_\rho^{\not<\sigma} \uplus_\# M_\rho^{<\sigma}$ and $|\tau'| = M_{\tau'}^{\not<\sigma} \uplus_\# M_{\tau'}^{<\sigma}$, where $M_\rho^{<\sigma}$ (resp. $M_{\tau'}^{<\sigma}$) is the multiset consisting of all occurrences of elements from $|\rho|$ (resp. $|\tau'|$) that are majorized by some element from $|\sigma|$. It follows that $|\sigma \cdot \rho| = |\sigma| \uplus_\# M_\rho^{\not<\sigma}$. Hence we have to prove

$$M_{\tau'}^{\not<\sigma} \uplus_\# M_{\tau'}^{<\sigma} \uplus_\# M_\rho^{\not<\sigma} \uplus_\# M_\rho^{<\rho} <_\# |\tau| \uplus_\# |\sigma| \uplus_\# M_\rho^{\not<\sigma}.$$

We obviously have $M_{\tau'}^{<\sigma} \uplus_\# M_\rho^{<\sigma} <_\# |\sigma|$. Using Lemma A.3, it suffices to prove that $M_{\tau'}^{\not<\sigma} \leqslant_\# |\tau|$.

In order to prove $M_{\tau'}^{\not<\sigma} \leqslant_\# |\tau|$, we take into account the traces in D. Each step in τ' traces back either to σ (a diagonal trace) or to the opposite edge τ (a horizontal trace). Diagonal traces are by definition dashed lines, which expresses that the step in τ' traces back to a majorizing step in σ. It follows that all steps from $M_{\tau'}^{\not<\sigma}$ trace back to τ; in other words, all traces with endpoints in $M_{\tau'}^{\not<\sigma}$ are horizontal.

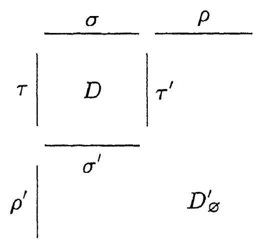


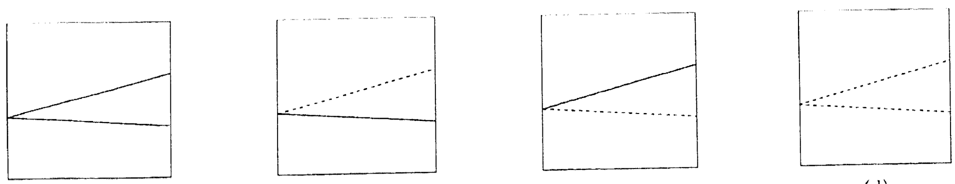
FIG. 20. Adjoining a trace-decreasing diagram.

Now consider the horizontal traces between steps in $M_{\tau'}^{\ddagger\sigma}$ and steps in τ. We compare the endpoints in $M_{\tau'}^{\ddagger\sigma}$ with the endpoints in τ as multisets. For this comparison we consider all possibilities for splitting traces in Fig. 21. Configurations (a) and (b) can be excluded by the definition of trace-decreasing. Configuration (c) does not occur in $|\tau'|$ due to *filter*. Configuration (d) is unproblematic. However, we cannot yet conclude to $M_{\tau'}^{\ddagger\sigma} \leqslant_{\#} |\tau|$ with Lemma A.3(iv), since the endpoints in τ may be filtered out in $|\tau|$. Fortunately, if such an endpoint in τ is filtered out, then there is always a majorizing previous step in τ which is not filtered out. For the purpose of the multiset inequality $M_{\tau'}^{\ddagger\sigma} \leqslant_{\#} |\tau|$, we can redirect all horizontal traces that have an endpoint in τ which is filtered out in $|\tau|$ to the nearest previous majorizing step which is not filtered out. We must check that there are no problematic splitting configurations introduced by this redirection (crossing is irrelevant for multiset inequality). We first argue that majorizing steps in $|\tau|$ that were already used as endpoints of a *line* do not occur. In Fig. 22 we list all possible configurations, assuming always that α is the majorizing step, so $index(\alpha) > index(\beta)$. Now configurations (a) and (b) can be excluded by the definition of trace-decreasing, and configurations (c) and (d) do not occur in $|\tau'|$ due to *filter*. Hence all majorizing steps in $|\tau|$ are either used as endpoints of *dashed lines*, or were not used as endpoints before redirection. Hence redirection can only give rise to unproblematic splitting configurations of type (d) in Fig. 21. This completes the proof of $M_{\tau'}^{\ddagger\sigma} \leqslant_{\#} |\tau|$ and hence of the proposition. ∎

Finally, we can combine the two previous properties of trace-decreasing diagrams to prove the main theorem.

4.19. THEOREM (Main theorem on trace-decreasing diagrams). *Let $\mathscr{A}$ be the ARS $(A, \{\to_\alpha \mid \alpha \in I\})$, with the index set I equipped with a well-founded order $<$. Let* ED *be a full set of (trace-)decreasing elementary diagrams. Then every diagram built from elements of* ED *can be completed into a trace-decreasing diagram. As a consequence we have that $\to_I$ is confluent.*

Proof. It suffices to prove the theorem for empty diagrams. We use well-founded induction with respect to the multiset order $<_{\#}$, which is well-founded according to Lemma A.3(ii). The proof follows the pattern of proof (i) of Newman's Lemma 2.6. Let $D_\varnothing$ be an empty diagram. Assume the theorem has been proved for all empty diagrams with norm smaller than $|D_\varnothing|$. If one of the spanning edges of $D_\varnothing$ is empty, then we are done. Otherwise, we may assume without loss of generality that both spanning edges start with a nonempty step since possible initial empty steps can be dealt with by trivial elementary diagrams. So let the spanning edges of $D_\varnothing$



(a) (b) (c) (d)

FIG. 21. Splitting configurations.

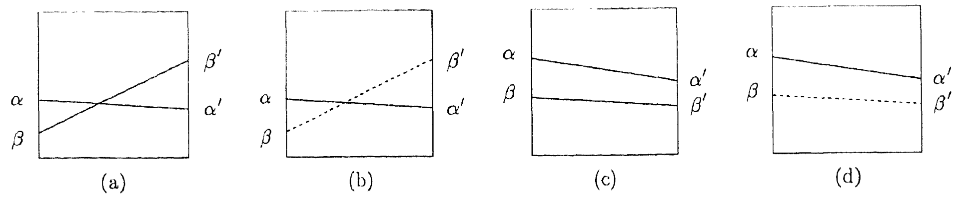


FIG. 22. Configurations to be excluded for redirection.

be of the form $\langle\alpha\rangle\cdot\sigma$ and $\langle\beta\rangle\cdot\tau$ for suitable α, $\beta\in I$ and indexed reduction sequences σ, τ, see Fig. 23(i). Since ED is a full set of trace-decreasing elementary diagrams, there exists a proper trace-decreasing elementary diagram D with spanning edges α and β. We adjoin this elementary diagram to the origin a of $D_\varnothing$ and arrive at a situation as in Fig. 23(ii). By Proposition 4.18, $|D'_\varnothing| <_{\#} |D_\varnothing|$. Hence $D'_\varnothing$ can be completed according to the induction hypothesis, say by a diagram D'. By Proposition 4.17, adjoining D' to D yields a trace-decreasing diagram, see Fig. 24(iii). Again by Proposition 4.18, $|D''_\varnothing| <_{\#} |D_\varnothing|$. Now $D_\varnothing$ can be completed by applying the induction hypothesis to $D''_\varnothing$, see Fig. 24(iv). ∎

4.20. Example (Alternative proof of Newman's Lemma 2.6). Let $\mathscr{A}=(A,\rightarrow)$ satisfy WCR and SN. We can recast $\mathscr{A}$ as the ARS $(A,\{\rightarrow_a \mid a\in A\})$, with $\rightarrow_a=\{(a,b)\mid a\rightarrow b\}$. By SN we have that $\leftarrow^+$ is a well-founded order on A. The set of elementary diagrams is full by WCR, and all elementary diagrams are trace-decreasing by the definition of the order. By Theorem 4.19 it follows that $\rightarrow$ is CR.

4.21. Lemma (Hindley (1964)). *Let* $(A,\{\rightarrow_\alpha \mid \alpha\in I\})$ *be an* ARS *such that for all* α, $\beta\in I$, $\rightarrow_\alpha$ *commutes with* $\rightarrow_\beta$. (*In particular*, $\rightarrow_\alpha$ *commutes with itself.*) *Then the union* $\rightarrow\; = \bigcup\{\rightarrow_\alpha \mid \alpha\in I\}$ *is confluent.* (*This proposition is usually referred to as the lemma of Hindley–Rosen*; *see, e.g., Barendregt* (1984), *Proposition* 3.3.5.)

Proof. Consider the ARS $\mathscr{A}=(A,\{\twoheadrightarrow_\alpha \mid \alpha\in I\})$, that is, with reduction relations $\twoheadrightarrow_\alpha$ instead of $\rightarrow_\alpha$. Put $\rightarrow_I=\bigcup\{\twoheadrightarrow_\alpha \mid \alpha\in I\}$. By Proposition 2.3 we have that $\rightarrow(\rightarrow_I)$ is confluent if and, only if $\twoheadrightarrow(\twoheadrightarrow_I)$ is confluent. As $\twoheadrightarrow\;=\;\twoheadrightarrow_I$, it

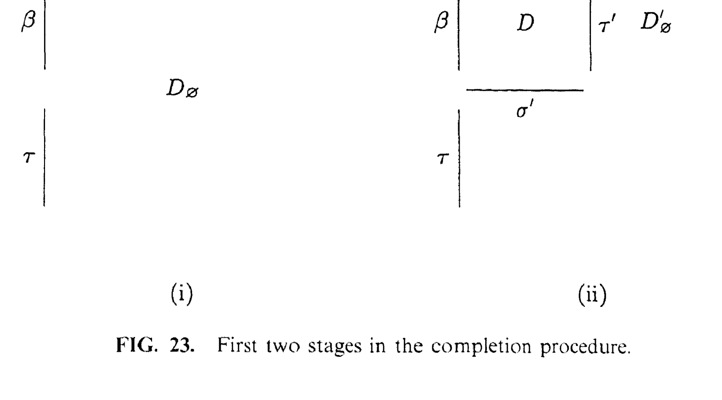
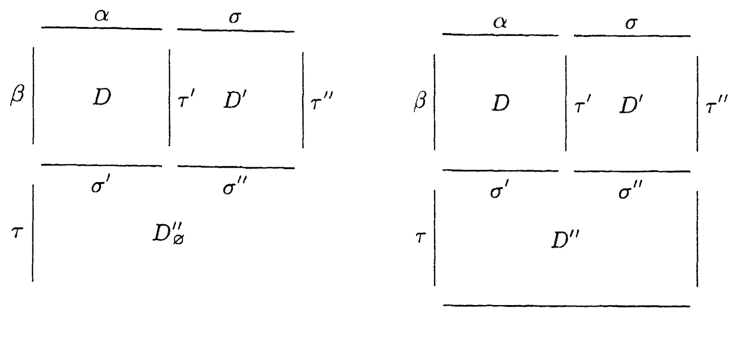


FIG. 23. First two stages in the completion procedure.



(iii) (iv)

FIG. 24. Last two stages in the completion procedure.

suffices to prove that $\to_I$ is confluent. Since $\to_\alpha$ and $\to_\beta$ commute for all $\alpha, \beta \in I$, we immediately get a full set of decreasing elementary diagrams of the form Fig. 25a for $\mathcal{A}$: opposite edges have identical indices. Here the order on I is irrelevant. The confluence follows from Theorem 4.19. ∎

4.22. LEMMA (Rosen (1973), Staples (1975)). *Let* $(A, \to_1, \to_2)$ *be an* ARS. *Define* $\to_1$ *requests* $\to_2$ *if* $\forall a, b, c \in A\ \exists d, e \in A\ (b \twoheadleftarrow_1 a \twoheadrightarrow_2 c \Rightarrow b \twoheadrightarrow_2 d \twoheadleftarrow_2 e \twoheadleftarrow_1 c)$.

(i) *Suppose* $\to_1$ *requests* $\to_2$ *and* $\to_2$ *is confluent. Suppose moreover that* $\forall a, b, c \in A\ \ \exists d \in A\ \ (b \twoheadleftarrow_1 a \twoheadrightarrow_1 c \Rightarrow b \to_3 d \leftarrow_3 c)$, *where* $\to_3 = \twoheadrightarrow_1 \cdot \twoheadrightarrow_2$ *is the composition of* $\twoheadrightarrow_1$ *and* $\twoheadrightarrow_2$. *Then* $\to_{12}$ *is confluent.*

(ii) *If* $\to_1$, $\to_2$ *are confluent and* $\to_1$ *requests* $\to_2$, *then* $\to_{12}$ *is confluent.*

Proof. As in the previous proof we shift to the ARS with reduction relations $\twoheadrightarrow_1$ and $\twoheadrightarrow_2$.

(i) The confluence of $\to_2$ yields elementary diagrams with all edges consisting of one reduction step $\twoheadrightarrow_2$, hence obviously decreasing. The request property gives elementary diagrams of the form Fig. 25b and their mirror images with respect to the diagonal, which are decreasing if we take $1 > 2$. The other given property gives elementary diagrams of the form Fig. 25c, which are also decreasing when $1 > 2$. The total set of decreasing diagrams is full. The confluence follows from Theorem 4.19.

(ii) By the previous case, since $\to_1$ is confluent and $\twoheadrightarrow_1 \subseteq \to_3$. ∎

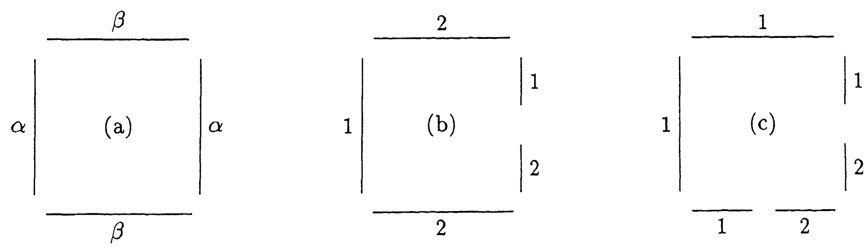


FIG. 25. Elementary diagrams.

4.23. Lemma (Barthe). *Let* $(A, \to_1, \to_2)$ *be an* ARS *such that* $\to_1$ *is confluent and* $\to_1$ *and* $\to_2$ *commute. Assume, moreover, that* $\forall a, b, c \in A\ \exists b', c', d \in A\ (b \twoheadleftarrow_2 a \twoheadrightarrow_2 c \Rightarrow b \twoheadrightarrow_2 b' \twoheadrightarrow_1 d \twoheadleftarrow_1 c' \twoheadleftarrow_2 c)$. *Alternatively,* $\twoheadleftarrow_2 \cdot \twoheadrightarrow_2 \subseteq \twoheadrightarrow_2 \cdot \twoheadrightarrow_1 \cdot \twoheadleftarrow_1 \cdot \twoheadleftarrow_2$. *Then* $\to_{12}$ *is confluent.*

Proof. As in the previous lemmas we shift to the ARS with reduction relations $\twoheadrightarrow_1$ and $\twoheadrightarrow_2$. From the fact that $\to_1$ is confluent and $\to_1$ and $\to_2$ commute we get decreasing diagrams with any order. The third property gives us elementary diagrams of the form in Fig. 25c, with 1 and 2 interchanged, thus decreasing with $1 < 2$. The total set of decreasing diagrams is full. The confluence follows from Theorem 4.19. ∎

For some applications, we need a slightly stronger version of Theorem 4.19. It aims at commutation of reduction relations rather than at confluence. It is stronger since confluence is equivalent to self-commuting, see Definition 2.2.

4.24. Definition. Let $\mathscr{A}$ be the ARS $(A, \{\to_\alpha \mid \alpha \in I\})$. Let $I = I_0 \cup I_1$ and let ED be a set of elementary diagrams. We say that ED is *commuting-full*, if for every proper inner corner $c \leftarrow_{\alpha_0} a \to_{\alpha_1} b$ with $\alpha_0 \in I_0$, $\alpha_1 \in I_1$, there exists a fitting elementary diagram in ED with completing edges $c \twoheadrightarrow_{I_1} d \twoheadleftarrow_{I_0} b$. In other words, opposite edges in the elementary diagram represent reduction steps with indices from the same index subset.

4.25. Theorem (Commuting version of Theorem 4.19). *Let* $\mathscr{A}$ *be then* ARS $(A, \{\to_\alpha \mid \alpha \in I\})$, *with the index set* I *equipped with a well-founded order* $<$. *Let* $I = I_0 \cup I_1$ *and let* ED *be a commuting-full set of (trace-)decreasing elementary diagrams. Then every diagram with spanning edges* $\sigma\colon a \twoheadrightarrow_{I_0} c$ *and* $\tau\colon a \twoheadrightarrow_{I_1} b$ *and built from elements of* ED, *can be completed into a trace-decreasing diagram with completing edges* $c \twoheadrightarrow_{I_1} d \twoheadleftarrow_{I_0} b$. *As a consequence use have that* $\to_{I_0}$ *and* $\to_{I_1}$, *commute.*

Proof. Analogous to the proof of Theorem 4.19, loading the induction as follows: every empty diagram with spanning edges $\sigma\colon a \twoheadrightarrow_{I_0} c$ and $\tau\colon a \twoheadrightarrow_{I_1} b$ can be completed with completing edges $\tau'\colon c \twoheadrightarrow_{I_1} d$ and $\sigma'\colon b \twoheadrightarrow_{I_0} d$. ∎

4.26. Lemma. *Let* $(A, \to_1, \to_2)$ *be an* ARS *such that* $\to_1$ *and* $\to_2$ *commute weakly and* $\to_{12}$ *is* SN, *then* $\to_1$ *and* $\to_2$ *commute. (The condition* $\to_{12}$ *is* SN *cannot be weakened to* $\to_1$ *is* SN *and* $\to_2$ *is* SN.)

Proof. Analogous to the proof of Newman's lemma in Example 4.20, but with Theorem 4.25 instead of 4.19. ∎

4.27. Lemma (Hindley (1964)). *Let* $(A, \to_1, \to_2)$ *be an* ARS *and suppose* $\forall a, b, c \in A\ \exists d \in A\ (b \leftarrow_1 a \to_2 c \Rightarrow b \twoheadrightarrow_2 d \leftarrow_1^{=} c)$. *Then* $\to_1$ *and* $\to_2$ *commute.*

Proof. Write $I = \{1\} \cup \{2\}$ and put $1 > 2$. The elementary diagrams are decreasing since $|\langle 1, 2, ..., 2\rangle| = [1] <_\# [1, 2]$ and $|\langle 2, 1\rangle| = [1, 2]$, $|\langle 2\rangle| = [2] <_\# [1, 2]$. The set of elementary diagrams is commuting-full, so we can apply Theorem 4.25. ∎

4.28. THEOREM (Strong confluence theorem). *Let* $\mathscr{A} = (A, \{\to_\alpha \mid \alpha \in I\})$ *be an* ARS, *with the index set* I *equipped with a well-founded order* $<$. *Assume that* $\mathscr{A}$ *is* strongly confluent, *that is, there exists a full set of elementary diagrams as specified in Fig.* 26. *Then* $\to_I$ *is confluent.*

Proof. Comparing Fig. 26 with Fig. 18 one observes that the elementary diagrams in Fig. 26 are not trace-decreasing due to two times $\leqslant\beta$ instead of $<\beta$ on the right edge. However, we can exploit the fact that this happens on the right edge only and not on both the right and the lower edge. Consider the ARS $\mathscr{A}^{01} = (A, \{\to_{\alpha_0} \mid \alpha \in I\} \cup \{\to_{\alpha_1} \mid \alpha \in I\})$, with $\to_{\alpha_0} = \to_{\alpha_1} = \to_\alpha$ for every $\alpha \in I$. Put $I_0 = \{\alpha_0 \mid \alpha \in I\}$, $I_1 = \{\alpha_1 \mid \alpha \in I\}$. Order $I_0 \cup I_1$ lexicographically, that is, $\alpha_i < \alpha'_j$ if and only if $\alpha < \alpha'$ or $\alpha = \alpha' \wedge i = 0 \wedge j = 1$. The order $<$ on $I_0 \cup I_1$ is well founded since $<$ on I is so. Every elementary diagram of $\mathscr{A}$ is transformed into an elementary diagram of $\mathscr{A}^{01}$ by giving all indices of horizontal steps subscript 1 and all indices of vertical steps subscript 0. For example, take an elementary diagram as in Fig. 26. Every index $\beta' \leqslant \beta$ on the right edge becomes $\beta'_0 \leqslant \beta_0 < \beta_1$, α becomes α_0, and $\alpha' < \alpha$ becomes $\alpha'_0 < \alpha_0$. On the lower edge, $\alpha' < \alpha$ becomes $\alpha'_1 < \alpha_0$, β becomes β_1, and $\beta' < \beta$ becomes $\beta'_1 < \beta_1$. The result is a trace-decreasing diagram as depicted in Fig. 27. Since the set of elementary diagrams of $\mathscr{A}$ is full, it follows that the set of elementary diagrams of $\mathscr{A}^{01}$ is commuting-full. By Theorem 4.25 we have that $\to_{I_0}$ and $\to_{I_1}$ commute. Since $\to_{\alpha_0} = \to_{\alpha_1} = \to_\alpha$ for every $\alpha \in I$, it follows that $\to_I$ is confluent. ∎

4.29. LEMMA (Huet (1980)). *Let* $(A, \to)$ *be an* ARS. *If* $\to$ *is strongly confluent, that is, if* $\forall a, b, c \in A\ \exists d \in A\ (b \leftarrow a \to c \Rightarrow b \twoheadrightarrow d \leftarrow^{\equiv} c)$, *then* $\to$ *is confluent.*

Proof. Take the set I to be a singleton. Interchange b and c to comply with the format of possibly many equal step on the right edge and at most one equal step on the lower edge. Now apply Theorem 4.28. ∎

4.30. LEMMA (De Bruijn (1978), 4). *Let* $(A, \{\to_\alpha \mid \alpha \in I\})$ *be an* ARS, *with* $<$ *a well-founded order on* I. *Recall that* $\to_{<\alpha}$ $(\to_{\leqslant\alpha})$ *is the union of the reduction relations with index* $<\alpha$ $(\leqslant\alpha)$, *with reflexive transitive closure* $\twoheadrightarrow_{<\alpha}$ $(\twoheadrightarrow_{\leqslant\alpha})$.

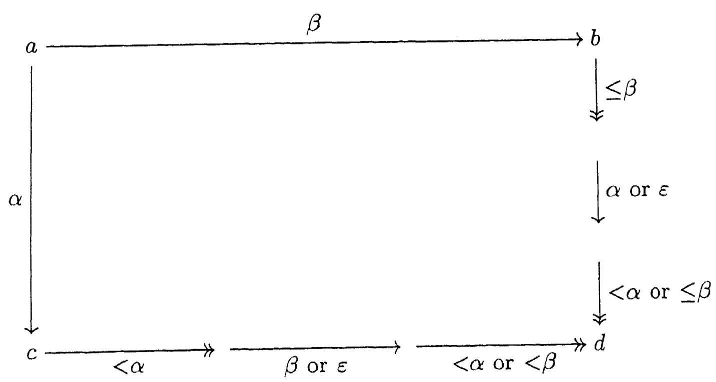


FIG. 26. Elementary diagrams for strong confluence.

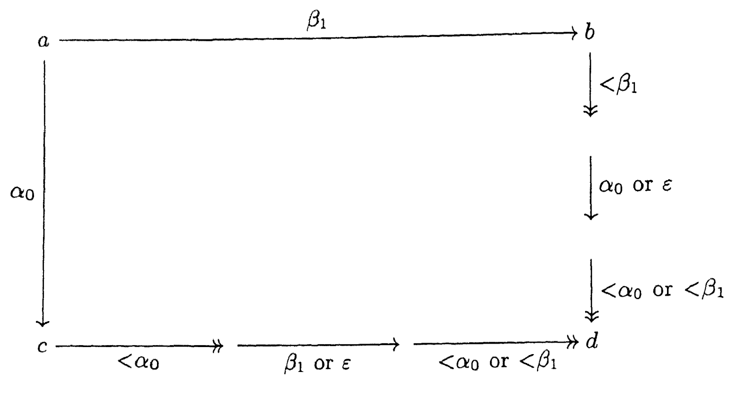


FIG. 27. Trace-decreasing elementary diagram for commuting.

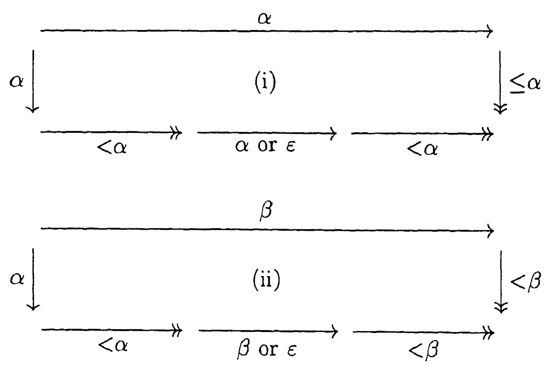
Assume

(i) $\forall \alpha \in I(\leftarrow_{\alpha} \cdot \rightarrow_{\alpha} \subseteq \twoheadrightarrow_{<\alpha} \cdot \rightarrow_{\alpha}^{\equiv} \cdot \twoheadrightarrow_{<\alpha} \cdot \twoheadleftarrow_{\leqslant \alpha})$, *and*

(ii) $\forall \alpha, \beta \in I(\alpha < \beta \Rightarrow \leftarrow_{\alpha} \cdot \rightarrow_{\beta} \subseteq \twoheadrightarrow_{<\alpha} \cdot \rightarrow_{\beta}^{\equiv} \cdot \twoheadrightarrow_{<\beta} \cdot \twoheadleftarrow_{<\beta})$.

Then $\rightarrow_I$ is confluent.

Proof. The elementary diagrams corresponding to (i) and (ii) are as follows. It is important to note that (ii) also allows the mirror image with respect to the diagonal of the diagram.



Elementary diagram (i) complies to the strong confluence format specified in Fig. 26. Elementary diagram (ii) and its mirror image are even trace-decreasing and hence also comply to the strong confluence format. It follows from the linearity of the order that the set of elementary diagrams is full: in each of the three cases, $\alpha = \beta$, $\alpha < \beta$, $\alpha > \beta$, there exists a fitting elementary diagram to the inner corner $c \leftarrow_{\alpha} a \rightarrow_{\beta} b$. Now apply Theorem 4.28 to conclude that $\rightarrow_I$ is confluent. ∎

As shown in van Oostrom (1994b), the result of Geser (1990) [p. 77] can be obtained from Theorem 4.28 along the lines of the proof of Lemma 4.30. We finish this section with some results on the completeness of the method.

4.31. DEFINITION. Let $\mathscr{A} = (A, \to)$ and $\mathscr{A}^I = (A, \{\to_\alpha \mid \alpha \in I\})$ be ARSs such that $\to\ = \ \to_I$. Then $\mathscr{A}^I$ is called an *indexed version* of $\mathscr{A}$. We say that $\mathscr{A}$ has the property DCR (decreasing Church–Rosser) if there exists an indexed $\mathscr{A}^I$ version $\mathscr{A}$ and a full set of (trace-)decreasing elementary diagrams for $\mathscr{A}^I$.

We have proved in Theorem 4.19 that DCR $\Rightarrow$ CR; the obvious question is whether the converse also holds. Van Oostrom (1994b) conjectured that CR $\Rightarrow$ DCR does *not* hold in general, but gives a proof in the countable case. This result can be viewed as the "completeness" of the method of decreasing diagrams with respect to establishing confluence in the countable case, which is a satisfactory state of affairs. Recall the cofinality properties CP and CP$^=$ in Definition 2.4 and Lemma 2.5.

4.32. PROPOSITION. *For every* ARS *we have* CP $\Rightarrow$ DCR.

Proof. Let $\mathscr{A} = (A, \to)$ be an ARS. We first define the *rewrite distance* $d(a, b)$ of $a \in A$, $b \in \mathscr{G}(a)$ as the minimal length of a reduction sequence from a to b. For $a \in A$ and $X \cap \mathscr{G}(a)$ nonempty we define the *distance* $d(a, X) = \min\{d(a, x) \mid x \in X \cap \mathscr{G}(a)\}$. Recall Lemma 2.5 for the definition of *component* and the equivalence CP$^=$ $\Leftrightarrow$ CP. To prove DCR for $\mathscr{A}$, it suffices to prove this property for the (disjoint) components of $\mathscr{A}$. Let $\mathscr{C}(a)$ be a component of $\mathscr{A}$. By CP we have CP$^=$, so we have a (finite or infinite) reduction sequence σ: $a \equiv a_0 \to a_1 \to \cdots$ which is cofinal in $\mathscr{C}(a)$. It is an easy exercise that we may suppose that σ is acyclic. We index reduction steps in $\mathscr{C}(a)$ with natural numbers as:

(i) $b \to_0 c$ if $b \to c$ occurs in σ; i.e. $b \equiv a_i$ and $c \equiv a_{i+1}$ for some $i \geqslant 0$.

(ii) $b \to_{n+1} c$ if $b \to c$ and $n = d(c, \{a_i \mid i \geqslant 0\})$.

Obviously, $\mathscr{A}^I = (\mathscr{C}(a), \{\to_n \mid n \in \mathbb{N}\})$ is an indexed version of $\mathscr{A}$. We will show that $\mathscr{A}^I$ satisfies DCR.

Consider $c \leftarrow_{m'} d \to_m b$. If $m' = m = 0$, then the steps $d \to c$ and $d \to b$ occur in σ and hence coincide since σ is acyclic. So $c \equiv b$ and we can complete the two diverging steps by two empty steps into a decreasing diagram. If $m, m' > 0$, then we have the situation in Fig. 28, clearly constituting a decreasing diagram. If $m = 0$, $m' > 0$, then we have the situation in Fig. 29, also giving a decreasing diagram. ∎

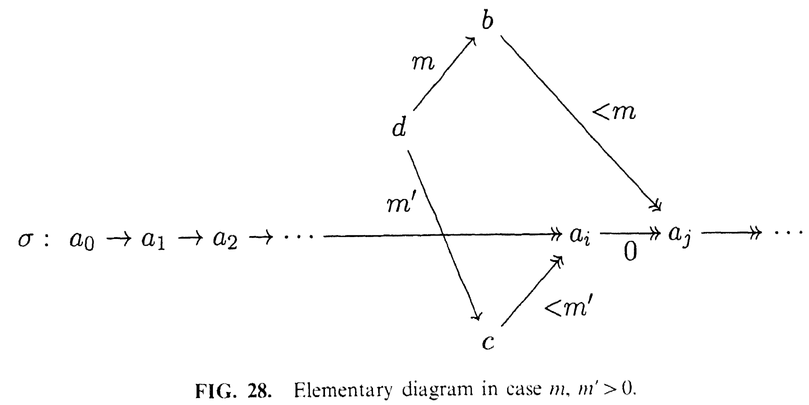


FIG. 28. Elementary diagram in case $m, m' > 0$.

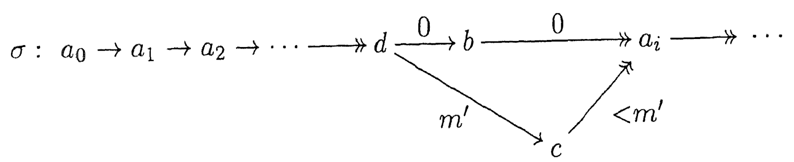


FIG. 29. Elementary diagram in case $m = 0$, $m' > 0$.

4.33. *Remark*. The uncountable ARS $(\aleph_1, <)$, with reduction relation $\to\, = \,<$, provides a counterexample to DCR $\Rightarrow$ CP. The property DCR holds in $(\aleph_1, <)$, since we have a full set of trivial elementary diagrams, by taking the maximum of b and c as common reduct in $c \leftarrow a \rightarrow b$. However, CP fails, as $\aleph_1$ is a regular cardinal, i.e. a cardinal without a cofinal countable subset.

By combining the previous results we obtain the following theorem.

4.34. THEOREM. *For countable* ARSs *we have* CP $\Leftrightarrow$ DCR $\Leftrightarrow$ CR.

Proof. Follows by combining the results DCR $\Rightarrow$ CR (Theorem 4.19), CR $\Rightarrow$ CP for countable ARSs (Lemma 2.5(ii)), and CP $\Rightarrow$ DCR (Proposition 4.32). ∎

APPENDIX: MULTISETS

A.1. DEFINITION. Let S be a set. A *multiset* M with elements from S is a function $M: S \to \mathbb{N}$ such that $\{s \in S \mid M(s) > 0\}$, the *set of elements* of M, is finite. Such M is also called a multiset *over* S and can be described explicitly by

$$M = [\underbrace{s_1, ..., s_1}_{n_1}, ..., \underbrace{s_k, ..., s_k}_{n_k}],$$

where the $n_i = M(s_i)(1 \leqslant i \leqslant k)$ give the *multiplicities* of the elements. In this notation it is tacitly assumed that there are no other elements of M than those explicitly shown. Moreover, permutations of the occurrences of elements in the multiset are allowed and we will often leave the multiplicities implicit or will express them by using M as a function.

The set of multisets over S will be denoted by $S^{\#}$. We define *membership* for multisets by $s \in_{\#} M \Leftrightarrow M(s) > 0$. Multiset *inclusion* is defined by $M \subseteq_{\#} M'$ if and only if $M(s) \leqslant M'(s)$ for all $s \in S$. As usual, $\subset_{\#}$ is the strict (irreflexive) version of $\subseteq_{\#}$. The *size* $|M|$ of a multiset M is the natural number defined by $|M| = \sum_{s \in S} M(s)$.

Let M and M' be multisets over S. We shall define the *union*, *difference*, and *intersection* of the multisets M and M'. Actually, there are two notions of union for multisets: one where the two multiplicities of any element are *added* and one where for any element the *maximum* of the two multiplicities is taken. The first is dual to multiset difference, the second is dual to multiset intersection. Given the importance of multiplicities when dealing with multisets, we only give the first notion of union, which can also be viewed as disjoint union. Let $n \ominus m = n - m$ if $n \geqslant m$ and 0 otherwise $(n, m \in \mathbb{N})$. Define

$$(M \uplus_{\#} M')(s) = M(s) + M'(s)$$

$$(M -_{\#} M')(s) = m(s) \ominus M'(s)$$

$$(M \cap_{\#} M')(s) = \min\{M(s), M'(s)\}.$$

Multiset union and intersection are associative and commutative.

A.2. DEFINITION. Let $\prec$ be a strict partial order on a set S. We will extend $\prec$ to a strict partial order $\prec_{\#}$ on $S^{\#}$, the set of multisets over S, as follows: $\prec_{\#}$ is the smallest transitive relation satisfying

$$\text{if } \forall x \in_{\#} M'\ x \prec s, \text{ then } M \uplus_{\#} M' \prec_{\#} M \uplus_{\#} [s]$$

for all $s \in S$, and $M, M' \in S^{\#}$. The intuition is that a multiset becomes smaller in the sense of $\prec_{\#}$ by replacing one or more of its elements by an arbitrary number of smaller elements. In particular we can have $M' = [\ ]$, in which case the element s is simply deleted. The reflexive closure of $\prec_{\#}$ will be denoted by $\preccurlyeq_{\#}$ (and not by $\preccurlyeq_{\#}$).

Without proof we mention the following results concerning the multiset order.

A.3. LEMMA. *Let $\prec$ be a strict partial order on a set S. Then*:

(i) $\prec_{\#}$ *is a strict order on* $S^{\#}$.

(ii) $\prec_{\#}$ *is well founded if and only if* $\prec$ *is well founded.*

(iii) *For all* $M, M' \in S^{\#}$ *we have* $M \subseteq_{\#} M' \Leftrightarrow M' = (M' -_{\#} M) \uplus_{\#} M$.

(iv) *Let* $M, M' \in S^{\#}$ *and* $C = M \cap_{\#} M'$. *Then we have* $M \preccurlyeq_{\#} M' \Leftrightarrow \forall x \in_{\#} M -_{\#} C\ \exists y \in_{\#} M' -_{\#} C x \prec y$.

(v) *Cancellation for multisets: for all* $X, Y, Z \in S^{\#}$ *we have* $X \uplus_{\#} Y \prec_{\#} X \uplus_{\#} Z \Leftrightarrow Y \prec_{\#} Z$ *(also for* $\preccurlyeq_{\#}$*).*

ACKNOWLEDGMENTS

This work was partially supported by ESPRIT Working Group 6345 Semagraph and by ESPRIT BRA 6454 CONFER. We thank Yde Venema for discussions from which the current proof of Lemma 3.8 arose.

Received September 25, 1997; final manuscript received August 22, 1997

REFERENCES

Barendregt, H. P. (1984), "The Lambda Calculus: Its Syntax and Semantics," second, revised edition, North-Holland, Amsterdam.

de Bruijn, N. G. (1978), "A Note on Weak Diamond Properties," Memoradum, 78-08, Eindhoven University of Technology.

Geser, A. (1990), "Relative Termination," Dissertation, Ph.D. Thesis, University of Passau.

Hindley, J. R. (1964), "The Church-Rosser Property and a Result in Combinatory Logic," Dissertation, University of Newcasle-upon-Tyne.

Huet, G. (1980), Confluent reductions: Abstract properties and applications to term rewriting systems, *Assoc. Comput. Mach.* **27**(4), 797–821.

Jantzen, M. (1988), "Confluent String Rewriting and Congruences", EATCS (European Association for Theoretical Computer Science) Monographs on Theoretical Computer Science, Vol. 14, Springer-Verlag, Berlin.

Klop, J. W. (1980), "Combinatory Reduction Systems," Mathematical Centre Tracts, Vol. 127, Mathematisch Centrum, Amsterdam.

Mano, K. (1993), Personal communication.

Newman, M. H. A. (1942), On theories with a combinatorial definition of "equivalence," *Ann. Math.* **43**(2), 223–243.

Oostrom, V. van (1994a), Confluence by decreasing diagrams, *Theoret. Comput. Sci.* **126**, 259–280.

Oostrom, V. van (1994b), "Confluence for Abstract and Higher-Order Rewriting," Dissertation, Vrije Universiteit, Amsterdam.

Popkorn, S. (1994), "First Steps in Modal Logic," Cambridge Univ. Press, Cambridge, UK.

Rosen, B. K. (1973), Tree-manipulating systems and Church–Rosser theorems, *J. Assoc. Comput. Mach.* **20**(1), 160–187.

Staples, J. (1975), Church–Rosser theorems for replacement systems, *in* "Algebra and Logic" (J. Crossley, Ed.), Lecture Notes in Mathematics, Vol. 450, pp. 291–307, Springer-Verlag, Berlin/New York.

Sutherland, W. A. (1976), "Introduction to Metric and Topological Spaces," Oxford Uni. Press, Oxford.

Winkler, F., and Buchberger, B. (1985), A criterion for eliminating unnecessary reductions in the Knuth–Bendix algorithm, *in* "Proceedings of the Colloquium on Algebra, Cominatorics and Logic in Computer Science, Vol. II," pp. 849–869, Györ, Hungary.