The Termination of Rippling and Unblocking

The Termination of Rippling + Unblocking ⋆

Alan Bundy

Department of Artificial Intelligence, University of Edinburgh.
A.Bundy@ed.ac.uk

Abstract. Rippling is a heuristic technique for guiding rewriting of a goal with respect to one or more givens. Rewriting is restricted so that the similarities between goal and given are preserved and the movement of differences is directed and terminating. Unblocking is a technique for changing the differences to enable a subsequent ripple.
The standard definition of rippling currently prevents certain, otherwise desirable, unblocking steps. In particular, some desirable unblocking steps increase the well-founded measure on which the termination proof of rippling is based. We propose an alternative family of well-founded measures under which the combination of rippling and unblocking is terminating. These new measures extend the power of rippling.

1 Introduction

Rippling[1] is a heuristic technique for guiding rewriting of a goal with respect to one or more givens [Bundy *et al*, 1993]. Rippling and unblocking were originally developed for guiding the step cases of inductive proofs, where the given is the induction hypothesis and the goal is the induction conclusion. It has subsequently been found to be more generally useful. For instance, it has been applied to summing series, limit theorems and general equational reasoning, [Hutter, 1997].

Differences between the goal and given(s) are marked by meta-level annotations in the goal, called *wave annotations*. These annotations consist of *wave-fronts*, which mark differences, and contain *wave-holes*, which mark similarities. *Sinks* are parts of the goal which correspond to free variables in a given. The parts of the goal outside the wave-fronts and inside the wave-holes are called the *skeleton*. The skeleton is a set of formulae, each member of which is a copy of one of the givens. Rewriting is restricted so that the skeleton is preserved and the movement of wave-fronts is directed either outwards or towards *sinks*. Rippling succeeds if each member of the skeleton is wholly contained in a wave-hole, swallowed by a sink or all wave-fronts are eliminated. The givens can then be used to replace each member of the skeleton with *true*, in a process called *fertilization*. A well-founded measure, called the *wave measure*, can be defined, under which rippling terminates (but not necessarily with success). A theoretical account of rippling can be

⋆ Many thanks to David Basin , Richard Boulton and Toby Walsh for comments on an earlier draft of this paper. The research reported here was supported by EPSRC grant GR/L/11724.

[1] More information about rippling can be found in our FAQ at: `http://www.dai.ed.ac.uk/daidb/staff/personal_pages/bundy/ripple.html`

found in [Basin & Walsh, 1996]. A summary of part of this theory can be found below in appendix A.

Rippling can be implemented by annotating the rewrite rules so that they are also skeleton preserving and measure decreasing from left to right. Such annotated rewrite rules are called *wave-rules. Wave-rules only match redexes in goals if the wave annotation also matches.* This prevents some rewritings that would be legal for the unannotated rule and goal. The heuristic element of rippling is that such prevented rewritings are very unlikely to lead to a proof of the goal from the given. So preventing them reduces the combinatorial explosion. Moreover, it is usually possible to impose measure decreasing wave annotation on rewrite rules in *both orientations.* So the same rewrite rule can be used left to right and right to left without losing termination; the differences in wave annotation prevent looping. This bi-directionality makes possible proofs by rippling that would not be available by standard rewriting, *e.g.* those that require associativity in both directions.

Unblocking is a technique for rewriting the wave-fronts to enable a wave-rule application. Sometimes a wave-rule fails to apply because a wave-front in a goal is equal to but not identical to an instance of a wave-front in a wave-rule. It may be possible to rewrite the wave-front in the goal while still preserving its skeleton. Experimental testing of rippling suggests that unblocking is often required. Sometimes these unblocking steps are wave measure neutral, *i.e.* they neither increase nor decrease the wave measure. It is thus natural to think of a lexicographical combination of rippling and unblocking, *i.e.* to rewrite with respect to a measure which either decreases the wave measure or keeps the wave measure constant but decreases some reduction order applied to the wave-fronts.

Unfortunately, this lexicographical scheme is not adequate to account for some unblocking steps. Below we will exhibit unblocking steps which *increase* the standard wave measure. This makes the combination of rippling and unblocking potentially non-terminating. In this paper we propose a family of new wave measures in which unblocking is made an integral part of rippling. We show that the previously measure increasing unblocking steps are measure decreasing under some members of this family. We also show that the combination of rippling and unblocking terminate under all the new measures. Our new measures will make more powerful implementations of rippling possible, extending the abilities of automated theorem provers.

2 Examples and Notational Conventions

Wave annotation has been represented in a variety of different notations in the literature. In this paper we use grey boxes with directional arrows for wave-fronts and holes in these boxes for wave-holes. Consider the following example:

$$\text{Given: } a + b = 42$$
$$\text{Goal: } \left(\boxed{(c+d) + \underline{a}}^{\uparrow} \right) + b = \boxed{(c+d) + \underline{42}}^{\uparrow}$$

The goal is annotated with respect to the given. The expressions $(c + d)+$ represent differences between goal and given, so are put in wave-fronts annotated by grey boxes. The expressions a and 42 represent similarities between goal and given, so are put in wave-holes annotated by holes in those grey boxes. The skeleton consists of these two wave holes plus the expression $+b =$ which lies outside both wave-fronts. Sinks are annotated as $\lfloor y \rfloor$, which are intended to suggest a kitchen sink with a plug hole, but are not used in this or subsequent examples.

Sometimes we will want to represent nested wave-fronts. We will emphasise the nesting by using different shades of grey for the different wave-fronts[2] , *e.g.*:

$$\left(\boxed{c + \left(\boxed{d + \underline{a}}^{\uparrow} \right)}^{\uparrow} \right) + b = \boxed{(c+d) + \underline{42}}^{\uparrow}$$

where $c+$ is in the outer wave-front (in light grey) and $d+$ is in the inner wave-front (in darker grey).

We will adopt the convention that bound variables and constants are written in lower case letters and free variables are written in upper case. We will use the single shafted arrow $\rightarrow$ for logical implication and the double shafted arrow $\Rightarrow$ for rewriting. An example wave-rule is:

$$\left(\boxed{X + \underline{Y}}^{\uparrow} \right) + Z \Rightarrow \boxed{X + (\underline{Y + Z})}^{\uparrow}$$

If the wave annotation on this wave rule and the above goal are erased then the rule applies to the goal in three distinct ways, with redexes: $(c+d)+a$, $((c+d)+a)+b$ and $(c+d)+42$. However, when the wave annotation is taken into account then the wave-rule applies to the goal in only one way, with redex $((c+d)+a)+b$, which is the only one that leads to successful fertilization.

3 Splitting and Merging of Wave-Fronts

The argument below will revolve around the fact that compound wave-fronts can be represented in merged and split form. For instance, consider the annotated term:

$$\left(\boxed{c + (d + \underline{a})}^{\uparrow} \right) + b$$

The compound wave-front $c + (d + \ldots)$ is here merged into one, but can alternatively be split into two nested wave-fronts:

$$\left(\boxed{c + \left(\boxed{d + \underline{a}}^{\uparrow} \right)}^{\uparrow} \right) + b$$

For several good reasons the theoretical account of rippling given in [Basin & Walsh, 1996] defines well annotated terms (*wats*) as being in a maximally split form, in which precisely one function symbol intervenes between wave-front and wave-hole. These reasons are:

1. *It simplifies the application of wave-rules to goals.* A wave-rule may only match part of a compound wave-front in the goal. Splitting the wave-front in the goal enables that part of it to match the wave-rule.

[2] Such shading differences are purely for presentational purposes.

2. *It removes any ambiguity about how to calculate the wave measure.* Some normal form is required for calculating the weight that each wave-front contributes to the measure. The maximally split form has the advantage of simplifying the calculation; each wave-front has a weight of 1. This gives what [Basin & Walsh, 1996] call the *width* measure.
3. *It makes the theory easier.* Various definitions and theorems are slightly simpler in maximally split form. For instance, the difference unification algorithm which inserts wave annotation most naturally returns *wats* in this form, [Basin & Walsh, 1993].

To amplify points 1 and 2, the wave measure must embody some kind of assessment of the complexity of compound wave-fronts. Suppose we want to apply the wave-rule:

$$\boxed{s(\underline{X})}^{\uparrow} + Y \Rightarrow \boxed{s(\underline{X+Y})}^{\uparrow} \qquad (1)$$

to the redex $\boxed{s(s(\underline{m}))}^{\uparrow} + n$. As it stands the wave-front in the left-hand side of the wave-rule does not match the compound wave-front in the redex. However, if the wave-front in the redex is split $\boxed{s(\underline{\boxed{s(\underline{m})}^{\uparrow}})}^{\uparrow} + n$ then the wave-front in the wave-rule *does* match the outer wave-front in the redex, with X instantiated to $\boxed{s(\underline{m})}^{\uparrow}$. By keeping wave-fronts in maximally split form it is not necessary to split and merge wave-fronts during rippling.

For this ripple to constitute a reduction in the wave-measure of the goal, the weight of $\boxed{s(s(\underline{m}))}^{\uparrow}$ must be greater than the weight of $\boxed{s(\underline{m})}^{\uparrow}$. The width measure does this by putting the wave-fronts in maximally split form and then counting the number of nested wave-fronts, *i.e.* the weight of $\boxed{s(\underline{\boxed{s(\underline{m})}^{\uparrow}})}^{\uparrow}$ is 2 and the weight of $\boxed{s(\underline{m})}^{\uparrow}$ is 1.

Note that compound terms which are wholly within the wave-front cannot be split. For instance, the compound wave-front $rev(\boxed{(h_1 + h_2) :: \underline{t}}^{\uparrow})$, where :: is the infix list constructor and rev is list reverse, is already in maximally split form. This is because the term $h_1 + h_2$ is wholly within the wave-front.

4 Examples of Unblocking

Rippling is said to be *blocked* if no wave-rule applies to any redex in the goal. Rippling can become blocked prematurely, *i.e.* before rippling succeeds. Sometimes there is a wave-rule that almost applies, *i.e.* some manipulation of non-matching wave-fronts will enable application. The manipulation required to make the wave-rule apply is called *unblocking*. Even if unblocking terminates in itself, if an unblocking step can increase the wave measure then the combination of rippling and unblocking may not terminate. Unblocking must also preserve the skeleton, so as not to prevent fertilization.

To illustrate the situation consider applying the following wave-rule:

$$(\boxed{\underline{X}+1}^{\uparrow}) + Y \Rightarrow \boxed{(\underline{X+Y})+1}^{\uparrow}$$

to the redex $(\boxed{n + (0+1)}^{\uparrow}) + m$. The application fails due to a mismatch between 1 and $0+1$. However, $0+1$ can be rewritten to 1 enabling the matching of wave-rule to redex. Note that this unblocking took place wholly within the wave-front. Unfortunately, empirical evidence suggests that unblocking is frequently required to redexes containing wave-holes, and this is where problems can arise.

Extrapolating from this example it is tempting to combine rippling and unblocking lexicographically. Unblocking can be defined as normalisation of the wave-fronts with respect to a set of conventional rewrite rules oriented by some well-founded reduction measure. An overall measure can be defined consisting of the pair of the wave measure and this reduction measure. This overall measure can be well-ordered lexicographically. Unfortunately, the lexicographic combination of the width measure and a reduction measure[3] is not powerful enough to permit some common unblockings, as the next example illustrates.

The following rippling problem is adapted from [Hutter, 1997][p411].

$$\text{Given: } binom(x, s(0)) - x = 0$$
$$\text{Goal: } binom(\boxed{s(\underline{x})}^{\uparrow}, s(0)) - \boxed{s(\underline{x})}^{\uparrow} = 0$$

where the following wave-rules come from the definitions of *binom* and $-$:

$$binom(\boxed{s(\underline{X})}^{\uparrow}, s(Y)) \Rightarrow \boxed{\underline{binom(X, s(Y))} + binom(X, Y)}^{\uparrow} \tag{2}$$

$$\boxed{s(\underline{X})}^{\uparrow} - \boxed{s(\underline{Y})}^{\uparrow} \Rightarrow X - Y \tag{3}$$

Wave-rule (2) applies to this goal to produce:

$$\boxed{\underline{binom(x, s(0))} + binom(x, 0)}^{\uparrow} - \boxed{s(\underline{x})}^{\uparrow} = 0 \tag{4}$$

but the ripple is now blocked, as wave-rule (3) will not apply. We need to unblock the left-hand side wave-front using the rewrite rules:

$$binom(X, 0) \Rightarrow s(0)$$
$$X + s(Y) \Rightarrow s(X + Y) \tag{5}$$
$$X + 0 \Rightarrow X \tag{6}$$

which gives:

$$\boxed{\underline{binom(x, s(0))} + s(0)}^{\uparrow} - \boxed{s(\underline{x})}^{\uparrow} = 0 \tag{7}$$

$$\boxed{s(\underline{\boxed{\underline{binom(x, s(0))} + 0}^{\uparrow}})}^{\uparrow} - \boxed{s(\underline{x})}^{\uparrow} = 0 \tag{8}$$

$$\boxed{s(\underline{binom(x, s(0))})}^{\uparrow} - \boxed{s(\underline{x})}^{\uparrow} = 0 \tag{9}$$

[3] As currently implemented, for instance, in our *Clam* system.

after which wave-rule (3) applies to give:

$$binom(x, s(0)) - x = 0$$

and fertilization is possible. Note that the last two unblocking steps rewrote redexes containing wave-holes.

Unfortunately, in step (8) one wave-front is replaced by two. This will have the effect of increasing the width measure, since the weight of the wave-fronts at this point in the skeleton is increased from 1 to 2. So alternate applications of rippling and unblocking could cause looping. Many other examples follow a similar pattern, for instance the goals: $even(\boxed{s(n)}^{\uparrow} \times s(s(0)))$, $len(rev(\boxed{h :: t}^{\uparrow})) + n$ and $len(rev(\boxed{l <> (x :: (y :: nil))}^{\uparrow}))$, where $<>$ is infix list append and len returns the length of a list. One solution to these problems is to redefine the wave-measure to encompass unblocking of the kind illustrated above. This solution is described below.

5 The Main Idea of the New Measure

The main idea of the new measure is to use some reduction measure, red, well-ordered by $<$, to calculate the total weight of the wave-fronts at each node in the skeleton. In order to calculate red we need to merge all the similarly oriented wave-fronts at each point in the skeleton. This requires us to put well annotated terms in *maximally merged* normal form, rather than the maximally split normal form adopted in [Basin & Walsh, 1996]. red is then calculated for the whole wave-front. In this way the number of wave-fronts at a skeleton node does not automatically take precedence over the reduction weight of those wave-fronts, which is the problem with the width measure.

To see how this works consider again the example from §4. This required the unblocking steps (7), (8) and (9), where step (8) increased the width measure. For a suitable choice of red, each of these steps is measure decreasing under the new wave measure. The wave-fronts stay at the same position in the skeleton, but the overall measure decreases because the reduction measure of the wave-fronts decreases. We can factor the contents of the wave-hole out of this calculation, since skeleton preservation means that this is unchanged. The left-hand side wave-fronts then decrease as follows:

$$red(wh + binom(x, 0)) \ > \ red(wh + s(0)) \ > \ red(s(wh + 0)) \ > \ red(s(wh))$$

This sequence would be regarded as strictly decreasing under a number of reduction orders, for instance recursive path order (rpo), [Dershowitz, 1982]. The new measure can also handle the other counter-examples in §4.

This new measure also permits the rippling of $\boxed{s(\boxed{s(m)}^{\uparrow})}^{\uparrow} + n$ to $\boxed{s(\boxed{s(m)}^{\uparrow} + n)}^{\uparrow}$ using wave-rule (1), as discussed in §3. For the new measure to decrease it is sufficient that $red(s(s(x))) > red(s(x))$, which again is true for rpo, for instance.

The rewrite rules used for unblocking must preserve the skeleton. If the redex is wholly within the wave-front then this is automatic, *cf.* the redex in (4). However, if the redex contains a wave-hole then the rewrite rules must be annotated as wave-rules, *cf.* the redexes in (7) and (8). So the rewrite rules (5) and (6), which match these two redexes, must be annotated as wave-rules:

$$\boxed{X + s(Y)}^{\uparrow} \Rightarrow \boxed{s(\boxed{X + Y}^{\uparrow})}^{\uparrow} \tag{10}$$

$$\boxed{X + 0}^{\uparrow} \Rightarrow X$$

These are measure decreasing under our measure, provided $red(X + s(Y)) > red(s(X + Y))$ and $red(X + 0) > red(X)$, which is true, *e.g.* for rpo. In particular, wave-rule (10) is measure decreasing even though it replaces one wave-front with two. Another advantage of annotating unblocking rules is that it is then unnecessary to erase annotation before unblocking and reinstate it afterwards. Note that there are often lots of wave-rules corresponding to each rewrite rule, but it is not necessary to prestore them all; wave-rules can be annotated dynamically, by need, as proposed in [Basin & Walsh, 1996][§6.1].

Our proposed measure is similar, in spirit, to the proposal in [Basin & Walsh, 1996][§7.1] to lexicographically combine a size measure with a reduction order. Our measure is simpler than this proposal and subsumes it, since the lexicographic combination of a size and reduction order is itself a reduction order. So our new measure can also handle the example in [Basin & Walsh, 1996][§7.1]. The size measure is not developed theoretically in [Basin & Walsh, 1996]. For instance, note that, like our measure, it requires a definition of *wat* which permits maximally merged form. This paper provides that theoretical development.

Our measure also subsumes the width measure, by defining *red* to count the number of nested wave-fronts. Alternatively, *red* could be defined as the wave measure of an orthogonal wave annotation, treating the wave-front as a goal to be rippled towards some other given. This would allow nested rippling to any depth.

6 The Definition of Well Annotated Term

The formal definition of this new wave measure requires a major overhaul of the theory of rippling given in [Basin & Walsh, 1996] or appendix A. We start this with the definition of well annotated term. This was previously defined to be in maximally split normal form. We need a redefinition in which both maximally split and maximally merged normal forms are included. We need maximally merged form for defining the measure, but we still need maximally split form for applying wave-rules, in order to avoid the need for dynamic splitting and merging. So we also need normalisation procedures for switching between the two normal forms.

Below is such an inclusive redefinition of *wats*. It will be convenient for this theoretical development to follow [Basin & Walsh, 1996] by replacing the grey box notation with the meta-functions: $wf/2$, $wh/1$ and $snk/2$, which denote wave-fronts, wave-holes and sinks[4], respectively[5]. For instance, $\boxed{s(\underline{x})}^{\uparrow} + \lfloor y \rfloor$ is represented as $wf(s(wh(x)), out) + snk(y, Y)$,

[4] Actually, [Basin & Walsh, 1996] does not deal with sinks; their inclusion is a further innovation of this paper.

[5] This is essentially how wave annotation is implemented in *Clam*.

where the second argument of wf denotes the direction and the second argument of snk denotes the free variable in the given corresponding to this sink. We will add these meta-functions to some set of unannotated terms, which we will call *unat*, and then restrict their use.

Definition 1 (Well-Annotated Term:) *The annotated terms, at, are the terms generated from unat by adding the additional functions wf, wh and snk. The well-annotated terms, wat, are defined as:*

$$wat = \{t | t \in at \wedge ok(t, wat)\}$$

where $ok : at \times status \mapsto boole$ is defined as:

$$\begin{aligned}
ok(a, l) &\leftrightarrow l = unat \vee l = wat \\
ok(f(t_1, \ldots, t_n), l) &\leftrightarrow \forall i{:}[n].\ ok(t_i, l) \\
ok(wf(t, d), l) &\leftrightarrow l = wat \wedge t = f(t_1, \ldots, t_n) \wedge \exists i{:}[n].\ ok(t_i, hole(d)) \wedge \\
&\qquad \forall i{:}[n].\ (ok(t_i, hole(d)) \vee ok(t_i, unat)) && (11) \\
ok(wh(t), l) &\leftrightarrow l = hole(d) \wedge ok(t, wat) && (12) \\
ok(snk(t, A), l) &\leftrightarrow l = wat \wedge ok(t, unat)
\end{aligned}$$

where: a is a variable or constant; f is a function of unat, i.e. *not wf, wh or snk; $n \geq 1$; $[n] = \{1, \ldots, n\}$; $d \in \{out, in\}$; A is a free variable; and $status = \{unat, wat, hole(out), hole(in)\}$.*

The second argument of ok describes the status of the first argument. This is either a *unat*, a *wat* or a $hole(d)$, the latter being the part of a wave-front of direction d above a wave-hole. Note that a wave-front must contain at least one wave-hole nested somewhere within it. Those of its arguments which do not contain wave-holes must be unannotated. Wave-holes may contain nested wave-fronts. Sinks contain only unannotated terms.

To get the definition of a *wat* in maximally *split* normal form, *ms_wat*, we exchange clause (11) of this definition by:

$$\begin{aligned}
ok(wf(t, d), l) \leftrightarrow l = wat \wedge t = f(t_1, \ldots, t_n) \wedge \exists i{:}[n].\ t_i = wh(s_i) \wedge ok(s_i, wat) \wedge \\
\forall i{:}[n].\ ((t_i = wh(s_i) \wedge ok(s_i, wat)) \vee ok(t_i, unat))
\end{aligned}$$

This ensures that any wave-holes appear immediately nested within the wave-fronts.

To get the definition of a *wat* in maximally *merged* normal form, *mm_wat*, we exchange clause (12) of this definition by:

$$ok(wh(t), l) \leftrightarrow l = hole(d) \wedge ok(t, wat) \wedge \neg \exists t'.\ t = wf(t', d)$$

This ensures that no wave-front is nested immediately within a wave-hole dominated by a wave-front with the same direction. Note how the argument to *hole* is used to store the direction of the dominating wave-front.

To define the wave measure we also need to redefine *simply well annotated terms* (*swat*s), in which wave-fronts have exactly one wave-hole nested within them. To get the definition of *swat* we exchange clause (11) by:

$$\begin{aligned}
ok(wf(t, d), l) \leftrightarrow l = wat \wedge t = f(t_1, \ldots, t_n) \wedge \exists i{:}[n].\ [ok(t_i, hole) \wedge \\
\forall j{:}[n].\ j \neq i \rightarrow ok(t_j, unat)]
\end{aligned}$$

This ensures that precisely one argument of the wave-front has a wave-hole and all the others are *unat*s.

7 Conversion to Normal Form

We need two normalisation functions: one to convert a *wat* to maximally split normal form and one to convert to maximally merged normal form. We will call these *split*/1 and *merge*/2, respectively.

The function *split* needs an auxiliary function *awf* which adds additional wave-holes and wave-fronts on appropriate arguments. The only interesting case is when *split* encounters a wave-front. Otherwise, it recurses on the arguments of its inputs.

Definition 2 (The *split* **Function:)** *The function* $split : wat \mapsto ms\_wat$, *is defined by:*

$$\begin{aligned}
split(a) &= a \\
split(f(t_1, \ldots, t_n)) &= f(split(t_1), \ldots, split(t_n)) \\
split(wf(f(t_1, \ldots, t_n), d)) &= wf(f(awf(t_1, d), \ldots, awf(t_n, d)), d) \\
split(wh(t)) &= wh(split(t)) \\
split(snk(t, A)) &= snk(t, A)
\end{aligned}$$

where $awf : wh\_at \times dir \mapsto wh\_at$ *is defined by:*

$$\begin{aligned}
ok(t, unat) \leftrightarrow awf(t, d) &= t \\
ok(t, hole) \leftrightarrow awf(t, d) &= wh(split(wf(t, d)))
\end{aligned}$$

where $dir = \{out, in\}$, $wh\_at = \{t | ok(t, unat) \vee ok(t, hole)\}$ *and other conventions are as for* $ok/2$.

There is only one case of interest in the definition of *merge*: when a wave-front is immediately nested within a wave-hole. If the nested wave-front is in the same direction as the wave-front the wave-hole is descended from then they both need to be dropped. Otherwise, *merge* just recurses on the arguments of its input. *merge* needs a second argument which records the direction of any wave-front it is nested within. This second argument starts off as *undef*.

Definition 3 (The *merge* **Function:)** *The function* $merge : wat \times dir' \mapsto mm\_wat$, *is defined by:*

$$\begin{aligned}
merge(a, d) &= a \\
merge(f(t_1, \ldots, t_n), d) &= f(merge(t_1, d), \ldots, merge(t_n, d)) \\
merge(wf(t, d_1), d_2) &= wf(merge(t, d_1), d_1) \\
merge(wh(wf(t, d)), d) &= merge(t, d) \\
d_1 \neq d_2 \leftrightarrow merge(wh(wf(t, d_1)), d_2) &= wh(wf(merge(t, d_1), d_1)) \\
merge(snk(t, A), d) &= snk(t, A)
\end{aligned}$$

where $dir' = \{out, in, undef\}$ *and other conventions are as for* $ok/2$.

8 The New Wave Measure Family

This section adapts the theory of the width measure which is summarised in appendix A, and should be read in conjunction with that appendix. This appendix first defines outwards and inwards wave measures for *swats*, then, uses multi-sets to combine them into outwards and inwards measures for *wats* and lexicographic combination to make an overall measure. It is only necessary to change the definitions of the outwards and inwards width measures for *swats*, namely $wom/1$ and $wim/1$ and the types of the various measures. We will overload our notation by retaining the old names for the revised definitions. In particular, the new measure will still be called m and the new order $\prec$.

The new measure is parameterised by a well-founded, stable and monotonic reduction order, $<$ on the set ord, and a measure $red : wat \mapsto ord$ on wave-fronts. Most orders used to orient rewrite rules have these properties, *e.g.* rpo.

Note that both wom and wim are defined in terms of $wt/3$, where $wt_n^d(t)$ gives the total weight at depth n of the skeleton of t of wave-fronts of direction d. It is only necessary to change the definition of wt. The old definition is:

$$wt_i^d(t) = |\{s|dp(wf(s,d),t,i)\}_m|$$

where $\{x|P(x)\}_m$ is the multi-set of elements x for which $P(x)$ and $|s|$ is the number of elements in multi-set s. $dp(s,t,i)$ means that term s is nested at depth i in the skeleton of term t. Note that the weight at each depth in the skeleton is defined to be the number of wave-fronts at that depth.

Our new weight function will form the multi-set of the reduction measures of wave-fronts at each depth. To measure this properly it is necessary to merge adjacent wave-fronts. The new definition of wt is:

Definition 4 (Weight Function:) *The weight function of wave-fronts at each depth,* $wt : wat \times dir \times nat \mapsto \mathcal{P}_m(ord)$ *is defined by:*

$$wt_i^d(t) = \{red(s)|dp(wf(s,d), merge(t, undef), i)\}_m$$

where $\mathcal{P}_m$ is the power function on multi-sets. For reduction orders, like rpo, which apply to $unat$, red must first erase all wave annotation. Some versions of red may exploit wave annotation, *e.g.* the width measure.

The redefinition of wt has a knock-on effect on the types of all the wave measure functions and orders. These need to be changed by replacing all occurrences of nat by $\mathcal{P}_m(ord)$.

9 The Termination of Rippling+Unblocking

To show that the combination of rippling and unblocking is terminating under the new measure we need to show that the order, $\prec$, of the overall measure m is well-founded, monotonic and stable. The well-foundedness is easily established since $\prec$ is formed only from lexicographic and multi-set extensions of the well-founded order $<$.

If, in addition, $\prec$ is monotonic and stable then the application of wave-rules is terminating. Wave-rules are measure decreasing by definition, *i.e.* $m(r) \prec m(l)$ for each wave-rule $l \Rightarrow r$ (see appendix A). The properties of monotonicity and stability ensure that the goals they rewrite are also measure decreasing. The proofs of these two properties are essentially the same as those given for the standard wave measure order in [Basin & Walsh, 1996]. We do not repeat those proofs here but merely remark on any changes required to them[6].

Both proofs require a multi-part supplementary lemma about splicing together lists of numbers. The well-ordered sets returned by *wom* and *wim* are lists of numbers. Under our new measure *wom* and *wim* return lists of multi-sets of elements from *ord*. It is necessary to check that the supplementary lemma still holds for these lists.

The lemma concerns a function[7] $+ : list(nat) \times list(nat) \times nat \mapsto list(nat)$ which splices together lists of numbers. $l +_d r$ means list r is spliced into list l at depth d, *i.e.* from the d^{th} element onwards[8], *e.g.*

$$[l_1, l_2, l_3, l_4] +_3 [r_1, r_2, r_3] = [l_1, l_2, l_3 + r_1, l_4 + r_2, r_3]$$

We need to prove the equivalent lemma for splicing lists of multi-sets of elements of *ord*. We must adapt the definition of $+/3$ to pairwise multi-set union, *e.g.*

$$[l_1, l_2, l_3, l_4] +_3 [r_1, r_2, r_3] = [l_1, l_2, l_3 \cup_m r_1, l_4 \cup_m r_2, r_3]$$

We also need a dual function $-/3$ based on pairwise multi-set subtraction.

$$[l_1, l_2, l_3, l_4] -_3 [r_1, r_2, r_3] = [l_1, l_2, l_3 \setminus_m r_1, l_4 \setminus_m r_2, \{\}]$$

However, the revised definition of $-/3$ means that lemma 1 part 1 only holds conditionally. For the condition we will need a pairwise multi-set subset relation $\subset_d$. The lemma also uses the lexicographic order on lists, $<_l$, defined in appendix A.

Lemma 1 (Supplementary Lemma:) *This three part lemma is adapted from lemma 4 in [Basin & Walsh, 1996][p168].*

Let l and l' be lists of length i and $l' <_l l$. Let $r, r_1, \ldots, r_k$ be lists of length j then:

1. $\forall d{:}[j].\ l \subset_d r \rightarrow (r -_d l) +_d l' <_l r$
2. $\forall d{:}[i].\ l' +_d r <_l l +_d r$
3. $\forall d_1, \ldots, d_k{:}[i].\ (\ldots((l' +_{d_1} r_1) +_{d_2} r_2) \ldots +_{d_k} r_k) <_l (\ldots((l +_{d_1} r_1) +_{d_2} r_2) \ldots +_{d_k} r_k)$

Proof Remarks: *Part 1 is adapted from the original, which is $\forall d{:}[j].\ r >_l r +_d (l' - l)$. The new condition is necessary to ensure that each element in l causes a deletion from r. The remaining two lemmas are unproblematic.*

---

[6] As a matter of interest, note that both the monotonicity and stability lemmas are rippling problems and that their proofs use rippling.

[7] For a definition of this function see [Basin & Walsh, 1996][p167-8].

[8] $+/3$ has been slightly modified to apply to lists numbered from 1 instead of 0.

Lemma 2 (Monotonicity of $\prec$ with respect to *wats*:) *This lemma is based on lemma 5 in [Basin & Walsh, 1996][p168-9].*

$$m(l) \prec m(r) \rightarrow m(s[l]) \prec m(s[r])$$

$s[r]$ means $s[l]$ with the distinguished subterm l replaced by r.

Proof Remarks: *The proof is identical to that in [Basin & Walsh, 1996][p168-9], except for the appeal to lemma 1 part 1. This is used to show, in the simply annotated case, that:*

$$wom(s[r]) = (wom(s[l]) -_d wom(l)) +_d wom(r) <_l wom(s[l])$$
$$wim(s[r]) = (wim(s[l]) -_d wim(l)) +_d wim(r) <_l wim(s[l])$$

In our proof it is necessary to prove the conditions $wom(l) \subset_d wom(s[l])$ and $wim(l) \subset_d wim(s[l])$. These conditions are true by the definition of wom and wim.

Lemma 3 (Stability of $\prec$ with respect to *wats*:) *This lemma is based on lemma 6 in [Basin & Walsh, 1996][p169-70].*

$$m(l) \prec m(r) \rightarrow m(l\sigma) \prec m(r\sigma)$$

Proof Remarks: *The proof is almost identical to that in [Basin & Walsh, 1996][p168-9]. In the simply annotated case of that proof there is an appeal to the fact that instantiations of a variable in a wave-front (*not *in a wave-hole) have no effect on the weight of that wave-front. This is not true of our measure. However, we can appeal to the assumed monotonicity and stability of $<$ on ord to show that the effect on $r\sigma$ will be less than that on $l\sigma$, as required.*

Theorem 1 (Termination of Rippling+Unblocking:) *There is no infinite descending chain of rippling and unblocking steps.*

Proof Remarks: *This follows from lemmas 2, 3 and the well-foundedness of $\prec$.*

10 Conclusion

We have defined a new family of wave measures which ensure the terminating combination of rippling and unblocking. These measures require *wats* to be in maximally merged form. We have given a general definition of *wats* which includes maximally merged and maximally split *wats*. We have defined normalisation procedures to put *wats* into each of these normal forms. We have redefined the wave measure so that it forms the multi-set of reduction orders of maximally merged wave-fronts at each depth in the skeleton. A member of our new family of wave measures can handle the counter-examples to combined rippling and unblocking termination given in §4. We have shown rippling plus unblocking to be terminating. Our definitions and proofs generalise and realise a proposal made in [Basin & Walsh, 1996][§7.1]. Our wave measure family includes all the wave measures proposed in [Basin & Walsh, 1996] and provides powerful new kinds of measure.

wats need to be put into maximally merged form for the calculation of the wave measure and maximally split form for the application of wave-rules. In practice, *wats* can be kept in maximally split form, as now, except for wave-rule parsing. There is no need to keep recalculating the measure after each ripple or unblock since the monotonicity and stability of wave-rules guarantees that it will be inherited.

Empirical work is now required to explore this new family of wave measures and discover which members of it are most powerful in practice. It may be that no one member is best in all circumstances, but that the choice can be made dynamically as wave-rules are constructed. We have provided a more powerful theoretical framework for: the combination of standard reduction orders like rpo with rippling; and nested rippling towards different givens. Hutter has shown that such nested rippling is necessary for applying rippling to general equational reasoning, [Hutter, 1997].

A The Standard Theory of Rippling

Rippling is a form of rewriting performed on the well-annotated terms of some underlying logic, usually use first-order predicate calculus. Let *unat* (unannotated terms) be the set of first-order expressions. The following definitions are adapted from [Basin & Walsh, 1996]. Conventions are as for the definition of $ok/2$.

Definition 5 (Well-Annotated Term:) *The annotated terms, at, are the terms generated from unat by adding the additional functions $wf/2$, $wh/1$ and $snk/2$.*

- $wf : wat \times dir \mapsto wat$ *is the* wave-front *function, where* $dir = \{in, out\}$*;*
- $wh : wat \mapsto wat$ *is the* wave-hole *function; and*
- $snk : unat \times var \mapsto wat$ *is the* sink function*, where var is the type of free variables.*

wat is the subset of at in which wf and wh occur only in terms of the form $wf(f(t_1, \ldots, t_n), d)$ *where:*

- $\exists i{:}[n].\ t_i = wh(t_i')$*;*
- $\forall i{:}[n].$ *if* $t_i = wh(t_i')$ *then* $t_i'{:}wat$*; and*
- $\forall i{:}[n].$ *if* $t_i \neq wh(t_i')$*, then* $t_i{:}unat$*.*

Definition 6 (Erasure of a Well-Annotated Term:) *The erasure function,* $erase : wat \mapsto unat$*, is defined by:*

- $\forall t{:}unat.\ erase(t) = t$*;*
- $erase(wf(t, d)) = erase(t)$*;*
- $erase(wh(t)) = erase(t)$*;*
- $erase(snk(t, A)) = t$*;*
- $erase(f(t_1, \ldots, t_n)) = f(erase(t_1), \ldots, erase(t_n))$

Definition 7 (Skeleton of a Well-Annotated Term:) *The skeleton function,* $skel : wat \mapsto \mathcal{P}(unat)$*, is defined by:*

- $\forall t{:}unat.\ skel(t) = \{t\}$*;*
- $skel(snk(t, A)) = \{A\}$*;*

- $skel(wf(f(t_1,\dots,t_n),d)) = \{s | \exists i{:}[n].\ t_i = wh(t'_i) \wedge s \in skel(t'_i)\}$
- $skel(f(t_1,\dots,t_n)) = \{f(s_1,\dots,s_n) | \forall i{:}[n].\ s_i \in skel(t_i)\}$

The next four definitions are auxiliary to the definition of a the wave measure.

Definition 8 (Simply Annotated Terms:) *The simply annotated terms, swat, are defined as the subset of wat in which* wf *and* wh *occur only in terms of the form* $wf(f(t_1,\dots,t_n),d)$ *where:*

$$\exists i{:}[n].\ t_i = wh(t'_i) \wedge \forall j{:}[n].\ j \neq i \rightarrow t_j{:}unat$$

Note that the skeleton of a *swat* is a singleton.

Definition 9 (Weakest:) *The weakest relation,* $weak : wat \times swat \mapsto boole$*, is defined by:*

- $\forall t{:}unat.\ weak(t,t)$;
- $weak(snk(t,A), snk(t,A))$;
- $weak(wf(f(t_1,\dots,t_n),d), wf(f(s_1,\dots,s_n),d))$ *iff* $\exists i{:}[n].\ t_i = wh(t'_i) \wedge s_i = wh(s'_i) \wedge weak(t'_i, s'_i)$ *and* $\forall j \neq i{:}[n].$ *if* $t_j = wh(t'_j)$ *then* $s_j = erase(t'_j)$ *else* $t_j = s_j$.
- $weak(f(t_1,\dots,t_n), f(s_1,\dots,s_n))$ *iff* $\forall i{:}[n].\ weak(t_i,s_i)$

Definition 10 (Depth in Skeleton of Simply Annotated Terms:) *The relation for depth of a subterm in the skeleton of a swat,* $dp : swat \times swat \times nat \mapsto boole$ *is defined by:*

- $dp(t,t,1)$;
- $dp(s, snk(t,A), d)$ *iff* $dp(s,A,d)$
- $dp(s, wf(f(t_1,\dots,wh(t_i),\dots,t_n),d),k)$ *iff* $dp(s,t_i,k)$;
- $dp(s, f(t_1,\dots,t_n), k+1)$ *iff* $\exists i{:}[n].\ dp(s,t_i,k)$

Note that a subterm s may occur several times in a term t at the same or different depths.

Definition 11 (Height of the Skeleton of Simply Annotated Terms:) *The function for height of the skeleton,* $ht : swat \mapsto nat$*, is defined by:*

$$ht(t) = max(\{k | \exists s.\ dp(s,t,k)\})$$

Definition 12 (Wave Measure:) *The weight function of wave-fronts at each depth,* $wt : wat \times dir \times nat \mapsto nat$ *is defined by:*

$$wt^d_i(t) = |\{s | dp(wf(s,d),t,i)\}_m|$$

where $\{x|P(x)\}_m$ *is the multi-set of elements* x*, for which* $P(x)$*, and* $|s|$ *is the number of elements in multi-set* s*. Note that nat is well-ordered by* $<: nat \times nat \mapsto boole$.

The weak outwards measure function, $wom : swat \mapsto list(nat)$ *and the weak inwards measure function,* $wim : swat \mapsto list(nat)$ *are defined by:*

$$wom(t) = [wt^{out}_{ht(t)}(t),\dots,wt^{out}_1(t)]$$
$$wim(t) = [wt^{in}_1(t),\dots,wt^{in}_{ht(t)}(t)]$$

The well-founded order, $<_l$: $list(nat) \times list(nat) \mapsto boole$ *is defined as the lexicographic extension of* $<$.

These two functions can be extended to wats as follows. The outwards measure function, $om : wat \mapsto \mathcal{P}_m(list(nat))$ *and the inwards measure function,* $im : wat \mapsto \mathcal{P}_m(list(nat))$ *are defined by:*

$$om(t) = \{wom(s)|weak(t, s)\}_m$$
$$im(t) = \{wim(s)|weak(t, s)\}_m$$

The well-founded order, $\ll_l$: $\mathcal{P}_m(list(nat)) \times \mathcal{P}_m(list(nat)) \mapsto boole$, *is defined as the multi-set extension of* $<_l$.

Finally, an overall measure, $m : wat \mapsto \mathcal{P}_m(list(nat)) \times \mathcal{P}_m(list(nat))$, *is defined by:*

$$m(t) = \langle om(t), im(t)\rangle$$

The well-founded order, $\prec$: $(\mathcal{P}_m(list(nat))\times\mathcal{P}_m(list(nat)))\times(\mathcal{P}_m(list(nat))\times\mathcal{P}_m(list(nat))) \mapsto boole$, *is defined as the lexicographic extension of* $\ll_l$.

Definition 13 (Wave-Rule:) *The rewrite rule,* $l \Rightarrow r$, *is defined to be a* wave-rule *iff: Both* l *and* r *are wats;* $skel(r) \subseteq skel(l)$; *and* $m(r) \prec m(l)$.

A wave-rule, $l \Rightarrow r$ *is said to be an* annotation *of a rewrite rule* $l' \Rightarrow r'$ *iff* l', r':*unat and* $erase(l) = l'$ *and* $erase(r) = r'$.

Rewrite rules on *unat* can be automatically annotated as wave-rules using ground difference unification, [Basin & Walsh, 1993].

Rippling is defined as standard rewriting on *wat*s but with a non-standard definition of term replacement. Rewriting may result in a term which is not a *wat*, *i.e.* a wf may occur nested within another wf without an intervening wh. In such cases term replacement is defined to erase the inner occurrence of wf. The order, $\prec$, can be shown to be well-founded, stable and monotonic for *wat*s under this non-standard term replacement. Hence, rippling can be shown to be sound and terminating, [Basin & Walsh, 1996][§5.3].

References

[Basin & Walsh, 1993] Basin, David and Walsh, Toby. (1993). Difference unification. In Bajcsy, R., (ed.), *Proc. 13th Intern. Joint Conference on Artificial Intelligence (IJCAI '93)*, volume 1, pages 116–122, San Mateo, CA. Morgan Kaufmann. Also available as Technical Report MPI-I-92-247, Max-Planck-Institute für Informatik.

[Basin & Walsh, 1996] Basin, David and Walsh, Toby. (1996). A calculus for and termination of rippling. *Journal of Automated Reasoning*, 16(1–2):147–180.

[Bundy *et al*, 1993] Bundy, A., Stevens, A., van Harmelen, F., Ireland, A. and Smaill, A. (1993). Rippling: A heuristic for guiding inductive proofs. *Artificial Intelligence*, 62:185–253. Also available from Edinburgh as DAI Research Paper No. 567.

[Dershowitz, 1982] Dershowitz, N. (1982). Orderings for term rewriting systems. *Journal of Theoretical Computer Science*, 17(3).

[Hutter, 1997] Hutter, D. (1997). Coloring terms to control equational reasoning. *Journal of Automated Reasoning*, 18(3):399–442.