Critical Peaks Redefined, the Non-Left-Linear Case $\Phi \sqcup \Psi = \top$

Vincent van Oostrom

Department of Computer Science, University of Innsbruck, Austria
Vincent.van-Oostrom@uibk.ac.at

Abstract

In previous work we introduced clusters as terms with a number of *linear* patterns in them, gave both a geometric account of clusters as sets of positions and an inductive one as let-expressions for terms with gaps (second order variables), and showed these to be isomorphic giving rise to a lattice under the refinement order $\sqsubseteq$. This enabled us to give an alternative lattice theoretic definition of the notion of critical peak as $\Phi \sqcup \Psi = \top$, for coinitial multisteps Φ and Ψ. Here we extend this to the *non-linear* case.

1 Introduction

In [1] we gave an alternative account of critical peaks in left-linear first-order term rewriting systems. Here we extend it to the non-left-linear case. For reasons of space, we introduce our definitions as derived from those of [1], with the convention that derived notions are boldface.

2 Geometric Clusters

We formalise the notion of a ***cluster***, a term having a number of non-overlapping patterns encompassed in it, and show that these constitute a lattice under the *refinement* order $\sqsubseteq$.

Definition 1. *A (*geometric*)* ***pattern*** *for a term t is a geometric pattern [1, Definition 2] for t extended with an* ***arity****, which is a partitioning of its* fringe*: the set of positions directly below it. A (*geometric*)* ***cluster*** *is a pair of a term together with a set of disjoint* ***patterns*** *for it.*

By the boundary of a pattern comprising only vertex positions, its fringe necessarily comprises only edge positions. Given a term, we often specify a **cluster** just by its set of **patterns**.

Example 1. *The set of positions $\{\mathring{\varepsilon}, \bar{1}, \mathring{1}\}$ is a geometric pattern for the term $t = f(g(a), a)$, corresponding to the connected component comprising both f and g. Its fringe is $\{1\cdot\bar{1}, \bar{2}\}$ corresponding to both its a-subterms. It allows two* ***arities****: the first $\{\{1\cdot\bar{1}\}, \{\bar{2}\}\}$ has an equivalence class for each position separately, the second $\{\{1\cdot\bar{1}, \bar{2}\}\}$ has a single equivalence class.*

Partitionings can be represented in many ways. We will freely switch between representations as sets of blocks (as in the example), as restricted growth strings, and as equivalence relations. Because of their conciseness we will mainly employ restricted growth strings for specifying **arities** using that the positions in a fringe are totally ordered from left to right by the left-order $\prec_l$. In the example they are $[1, 2]$ and $[1, 1]$. We write $\pi : w$ to denote that pattern π has **arity** w, simply writing π if w is *natural*, i.e. of shape $[1, \ldots, n]$ for n the cardinality of the fringe. We abbreviate[1] such a natural **arity** to $[n]$. **Arities** other than the natural one are called *sharing*. In the example $[1, 2]$ is natural and $[1, 1]$ is sharing. Note that geometric clusters [1, Definition 2] are naturally embedded into geometric **clusters**, by decomposing the former into a set of disjoint geometric patterns and equipping these with their natural arity.

[1] This is unambiguous since $[n]$ itself is not a restricted growth string for $n \neq 1$ and $[1, \ldots, 1] = [1]$.

Example 2. *Some geometric **clusters** for the term $t = f(g(a), a)$ are: the* empty ***cluster*** $\emptyset$*;* $\{\{\mathring{\varepsilon}\}\}$*, the single pattern comprising only f with natural **arity** [2], i.e. [1, 2];* $\{\{\mathring{\varepsilon}, \bar{1}, \mathring{1}\} : [1, 1]\}$ *as in Example 1, having sharing **arity** [1, 1]; and* $\{\{\mathring{\varepsilon}, \bar{1}, \mathring{1}\}, \{\mathring{2}\}\}$ *comprising patterns f–g and the right a with natural **arities** [2] and [0].*

The idea of partitioning, well-known from shared (dag) representations of terms in unification and term graph rewriting, cf. e.g. [3, 4], is that **arity**-related positions indicate *equivalent* subterms, for some appropriate notion of equivalence.

Definition 2. *A **cluster** $\varsigma = (t,P)$ is* f-homogenous *if for each pattern $\pi : w$ in P its w-related positions have the same f-image, for f a function on (edge) positions (that may depend on ς).*

Note that **clusters** having natural **arities** are f-homogeneous for any f, and that for a constant f-function any **cluster** is f-homogeneous. We will be particularly interested in *term*-homogeneity, obtained by the function mapping a position p to the subterm $t|_p$ of t at position p, and ***cluster***-homogeneity, obtained by the function mapping to the **subcluster** $\varsigma|_p$, defined in the obvious way. The latter, simply referred to as homogeneity, is more restrictive:

Example 3. *For the term $t = f(g(a), a)$ the **pattern** $\{\{\mathring{\varepsilon}, \bar{1}, \mathring{1}\} : [1, 1]\}$ is homogeneous having two identical empty a-**subclusters** at related fringe positions $1{\cdot}\bar{1}$ and $\bar{2}$. The same **pattern** is not even term-homogeneous for the term $f(g(a), b)$. The **cluster** $\{\{\mathring{\varepsilon}, \bar{1}, \mathring{1}\} : [1, 1], \{\mathring{2}\}\}$ for t, although term-homogeneous, is not homogeneous since its a-**subclusters** $\emptyset$ and $\{\mathring{\varepsilon}\}$ are distinct.*

Intuitively, homogeneity of a **cluster** means its singling out of patterns is compatible with their **arities**, i.e. sharing. Henceforth, we restrict attention to homogeneous **clusters**.

To extend the subset order from clusters [1, Lemma 1] to **clusters**, we first generalise the standard notion of taking a subterm $t|_p$ of t *at* a position p, from positions to patterns.

Definition 3. *The term $t|_{\pi:w}$* at ***pattern*** $\pi : w$ in t *is* $(t[\mathsf{v}_{i_1}, \ldots, \mathsf{v}_{i_n}]_{p_1,\ldots,p_n})|_p$*, for p the root of π, $w = [i_1, \ldots, i_n]$, variables $\mathsf{v}_{i_1}, \ldots, \mathsf{v}_{i_n}$, and $p_1, \ldots, p_n$ the fringe of π.*

Intuitively, this carves out the pattern π from the term t respecting the **arity** w. It naturally generalises to **clusters** via their term component, updating patterns accordingly, i.e. left-quotienting by p the **patterns** that are subsets of π and omitting the others.

Example 4. *Consider the term $t = f(g(a), a)$ and pattern $\pi = \{\mathring{\varepsilon}, \bar{1}, \mathring{1}\}$ already seen in Example 2. Then $t|_{\pi:[1,2]} = f(g(\mathsf{v}_1), \mathsf{v}_2)$ and $t|_{\pi:[1,1]} = f(g(\mathsf{v}_1), \mathsf{v}_1)$. Note that taking the subterm at the root of π has no effect in this case, since the root is the empty position.*

*Accordingly, for the **cluster** $\varsigma = \{\{\mathring{\varepsilon}, \bar{1}, \mathring{1}\} : [1, 1], \{1{\cdot}1{\cdot}\mathring{1}\}, \{\mathring{2}\}\}$, $\varsigma|_{\pi:[1,2]}$ has term $f(g(\mathsf{v}_1), \mathsf{v}_2)$ and **pattern** $\{\mathring{\varepsilon}, \bar{1}, \mathring{1}\} : [1, 1]$, and $\varsigma|_{\pi:[1,1]}$ has term $f(g(\mathsf{v}_1), \mathsf{v}_1)$ and the same **pattern**. The result of the latter is a **cluster**, but not of the former, it is not even term-homogeneous, as $\mathsf{v}_1 \neq \mathsf{v}_2$.*

Intuitively, the **arity** [1, 2] was not strong enough for carving out π to preserve homogeneity (be compatible with [1, 1]). For comparing **clusters** we demand homogeneity *is* preserved.

Definition 4. $\varsigma \sqsubseteq \zeta$ *if for each $\pi : w$ in ς there is a $\rho : v$ in ζ with $\pi \subseteq \rho$ and $\varsigma|_{\rho:v}$ a **cluster**.*

The idea for that $\sqsubseteq$ is a lattice is based on [4, Section 5.3], where it is shown that a term graph G (a dag) for a term t be represented as $T/{\sim}$, the term tree T equipped with an equivalence relation $\sim$ on parallel positions (indicating which nodes are to be shared) that is homogeneous [4, Definition 5.3.14] (if a node is shared, then so are its corresponding arguments). For a given finite term, its term graphs constitute a lattice with bottom $\perp$ the term tree (for the identity relation), top $\top$ the maximal sharing graph (with pointer equality; same term iff same node), with meet and join of $T/{\sim_1}$, $T/{\sim_2}$ given by $T/({\sim_1} \cap {\sim_2})$ and $T/({\sim_1} \cup {\sim_2})^*$.[2]

Viewing our **arities** as expressing sharing, for showing that **clusters** constitute a lattice we proceed in much the same way, except that we also force connected components, i.e. patterns, first to be homogeneous and second to be viewed as 'single nodes' ('without internal sharing').

Example 5. *Consider the term* $t = f(g(h(a)), g(h(a)))$ *and* ***patterns*** $\pi = \{\mathring{\varepsilon}, \bar{1}, \mathring{1}, \bar{2}, \mathring{2}\}$ *with sharing* ***arity*** $[1,1]$*, and* $\rho = \{1{\cdot}\mathring{1}, 1{\cdot}1{\cdot}\bar{1}, 1{\cdot}1{\cdot}\mathring{1}\}$ *and* $o = \{\mathring{2}, 2{\cdot}\bar{1}, 2{\cdot}\mathring{1}\}$ *both with natural* ***arity****.*

- $\{\pi\} \sqcup \{\rho\}$ *is the* ***cluster*** $\{\pi : [1,1], \rho, \rho'\}$ *where* $\rho' = \{2{\cdot}\mathring{1}, 2{\cdot}1{\cdot}\bar{1}, 2{\cdot}1{\cdot}\mathring{1}\}$ *is a copy of* ρ *forced by homogeneity of* π*, after taking the union of the sets of (disjoint!) patterns;*
- $\{\rho\} \sqcup \{o\}$ *is* $\{\rho, o\}$*, i.e. just the union of their sets of (disjoint) patterns;*
- $\{o\} \sqcup \{\pi\}$ *is* $\{(\pi \cup \rho \cup o') : [1,1]\}$ *having one pattern since* o *and* π *have overlap and with* $o' = \{1{\cdot}\bar{1}, 1{\cdot}\mathring{1}\}$ *forced by homogeneity of* π*; the resulting pattern being larger forces to 'push the sharing' of* π *to the new fringe, in this case again resulting in* ***arity*** $[1,1]$*.*

Theorem 1. *The* ***clusters*** *for a given term ordered by* $\sqsubseteq$*, constitute a lattice.*

Proof. To compute the join $\varsigma \sqcup \zeta$ we first take the unions P of the sets of positions and $\sim$ of the **arities**, and next *homogeneously* close these to $\overline{P}$ and $\overline{\sim}$ such that:

- if $p \mathrel{\overline{\sim}} q$ and $p{\cdot}r$ in $\overline{P}$ or $\overline{\sim}$, then so is $q{\cdot}r$ and $p{\cdot}r \mathrel{\overline{\sim}} q{\cdot}r$; and
- $\overline{\sim}$ is an equivalence relation.

Finally, we (uniquely) decompose $\overline{P}$ into **patterns** with **arities** obtained by restricting $\overline{\sim}$ to their fringes. This yields a **cluster** since clusters are preserved under union, and positions in $\sim$ and hence in $\overline{\sim}$ are all edge positions, so that $\overline{P}$ will be have vertex boundaries: if $p \mathrel{\overline{\sim}} q$ and $p{\cdot}\bar{r}$ in $\overline{P}$, then also $p{\cdot}\mathring{r}$ is, so both will be adjoined. Finally, equivalence relations are preserved under restriction. That $\varsigma, \zeta \sqsubseteq \varsigma \sqcup \zeta$ and it is the $\sqsubseteq$-least such follows per construction, the former by being homogeneously closed and the latter by closure not adjoining more than needed.

Meets $\varsigma \sqcap \zeta$ are computed by taking the intersection of the sets of positions with the resulting **patterns** having as **arity** the intersection of those of the respective **patterns** involved. $\square$

Note that for natural **clusters**, having natural **arities**, meet and join coincide with those for clusters [1, Lemma 1], since then closing is vacuous. However, for non-linear terms the respective tops differ, since unlike **clusters**, clusters cannot capture sharing:

Example 6. *For* $t = f(\mathsf{v}_3, \mathsf{v}_3)$ *the cluster top is* $\{\{\mathring{\varepsilon}\} : [1,2]\}$ *but the* ***cluster*** *top is* $\{\{\mathring{\varepsilon}\} : [1,1]\}$*.*

We call a term t or a **cluster** for it *standard*, cf. [2], if $t = t|_\top$, capturing that variables are in accord with the restricted growth sequence of its top $\top$. For instance, $t = f(\mathsf{v}_3, \mathsf{v}_1)$ is not standard, since its top has **arity** $[1,2]$ so $t|_\top = f(\mathsf{v}_1, \mathsf{v}_2)$.[3]

In contrast to cluster lattices [1, Theorem 1], **cluster** lattices need not be[4] distributive: $\{\rho\} \sqcap (\{\pi\} \sqcup \{o\}) = \{\{1{\cdot}\mathring{1}\}\} \neq \emptyset = (\{\rho\} \sqcap \{\pi\}) \sqcup (\{\rho\} \sqcap \{o\})$ in Example 5. Note that the **arity** $[1,1]$ of π is sharing here. In a sense, non-distributivity of the lattice reflects non-left-linear rules being problematic for confluence analysis.

[2] The unification algorithm of [3] can be viewed as computing the join in time linear in the size of the *dags*.
[3] A check for standardness is part of the COPS database infrastructure to avoid duplicate problems.
[4] But sometimes they are, e.g. for terms not allowing any sharing.

3 Inductive Clusters

We adapt inductive clusters [1, Definition 3] to inductive **clusters** to allow for non-linear patterns, in such a way that they again [1, Theorem 1] are isomorphic to geometric **clusters**. To that end we first extend the notion of signature to be able to express non-linearity constraints.

Definition 5. *A* ***signature*** *is a set of symbols having restricted growth sequences as* ***arities****.*

We will employ the conventions for **arities** introduced in the previous section. Usual signatures are naturally embedded into **signatures** by embedding a symbol having arity n as one having natural arity $[n]$. We henceforth assume a signature with symbols of four types: *function* symbols f, g, ... having natural **arities**;[5] *variables* v_1, v_2, ... having **arity** $[]$; *gap* symbols X, Y, ... infinitely many of each **arity**; and *rule* symbols ϱ, θ, ... Terms are built according to the arities associated to (being the length of) the **arities**. A term *has* arity n if variables in it have indices $\leqslant n$. The arity *of* a term is its list w of variable indices, if that is a restricted growth sequence and undefined otherwise, and *has* **arity** v if w refines v (as partitions).

Example 7. *The idea of, say,* $t = f(\mathsf{v}_1, \mathsf{v}_1, \mathsf{v}_2)$ *having* ***arities*** *both* $[1,1,2]$ *and* $[1,1,1]$ *is that it is 'safe' to substitute* t *for a symbol having either of those* ***arities****, since they 'guarantee' at least as much as the* ***arity*** $[1,1,2]$ *of* t*. This would not hold for, e.g., a symbol of* ***arity*** $[1,2,2]$.

Definition 6. *An assignment (called* 2*nd order substitution in [1, Definition 3] and denoted there by* $[\![\vec{f} := \vec{t}]\!]$*)* $\mathsf{let}\ \vec{f} = \vec{t}$ *is an* ***assignment*** *if each* t_i *is homogeneous and has the* ***arity*** *of* f_i.

After this generalisation of assignments (note that if all **arities** are natural the condition holds vacuously, so is not restrictive), all other notions carry, mutatis mutandis (switching from regular to boldface), over from Definitions 3 and 4 of [1], in particular the inductive definition of the refinement order $\sqsubseteq$, In accord with the above we let (*inductive*) ***clusters*** be denoted by (second order) let-expessions: $\mathsf{let}\ \vec{f} = \vec{t}\ \mathsf{in}\ t$.

Theorem 2. *For a given term, geometric* ***clusters*** *ordered by* $\subseteq$ *are isomorphic to inductive* ***clusters****, up to renaming of gaps, ordered by* $\sqsubseteq$*. The order is a finite lattice.*

Proof. The proof extends that of [1, Theorem 1], verifying that homogeneity is preserved by the transformation in either direction. ☐

Lemma 1. *A* ***cluster*** ς *for non-variable* t *with* $\varsigma \neq \top$ *can be written as* $\mathsf{let}\ \vec{\mathsf{v}} = \vec{\varsigma}\ \mathsf{in}\ \varsigma_0$*, with* ς_0 *standard and non-variable, and non-empty vectors.*

Proof. Let $t = f(\vec{t}^n)$ and let $[i_1, \dots, i_n]$ be the **arity** of f. By homogeneity of ς we may write t as $\mathsf{let}\ \vec{\mathsf{v}} = \overrightarrow{t_m}\ \mathsf{in}\ f(\mathsf{v}_{i_1}, \dots, \mathsf{v}_{i_n})$ with $\overrightarrow{t_m} \subseteq \vec{t}$. Decomposing ς accordingly yields the result. ☐

4 Critical Pairs

Having introduced the new set up, we redefine fundamental notions such as overlap for **clusters** (without reference to rules yet!), then *inducing* the same for **multisteps**, **clusters** assigning single rule symbols to terms.

Definition 7. *A pair* (ς,ζ) *for* t *is* critical *if* $\varsigma \sqcup \zeta = \top$*,* t *is standard,* $\varsigma \neq \bot$ *and* $\zeta \neq \bot$*,* overlapping *if* $\varsigma \sqcap \zeta \neq \bot$*, and* empty *if* $\varsigma = \bot$ *or* $\zeta = \bot$.

[5] It could be interesting to consider sharing **arities** for function symbols as well.

Lemma 2. *Every critical pair is overlapping.*

Proof. A pair (ς,ζ) for t being critical means $\varsigma \sqcup \zeta$ can be written as $\mathsf{let}\ X = t\ \mathsf{in}\ X(\mathsf{v}_{i_1},\dots,\mathsf{v}_{i_n})$ for $[i_1,\dots,i_n]$ the arity of t. Per construction of $\sqcup$ in the proof of Theorem 1, this means that the head of t belongs to a **pattern** π in ς or ζ. But if it belongs to only one of them, then some **pattern** in the other must overlap it, otherwise π would by the construction be a **pattern** of $\varsigma \sqcup \zeta$, its only one by criticality, as well. □

Notions for clusters extend to **multisteps**, via the map lhs mapping the rule symbols in its **assignment** to their left-hand side; geometrically, expanding the vertices in the set of positions labelled by rule symbols, to their left-hand sides. In particular, a critical peak is a peak, i.e. a pair (ϕ,ψ) of co-initial steps, that is critical, i.e. such that $(\mathsf{lhs}(\phi),\mathsf{lhs}(\psi))$ is critical. A peak is *trivial* if it is parallel, i.e. if also its targets are the same.

Lemma 3. *Critical (one–one) peaks correspond to the usual ones.*

Proof. This follows from Lemma 2, and observing that usual critical pairs are induced by most general (w.r.t. $\sqsubseteq$) overlaps between left-hand sides; cf. [1, Lemma 3]. □

Lemma 4. *For a peak (Φ,Ψ) for t that is neither critical nor empty, $\varsigma = \Phi \sqcup \Psi$ can be written as $\mathsf{let}\ \vec{\mathsf{v}} = \vec{\varsigma}\ \mathsf{in}\ \varsigma_0$ such that for every step $\chi : t \to s$ in Φ or Ψ there is a step χ' in some ς_i that* extends *χ in the sense that the target of χ' (as a step from t) is reachable from that of χ.*

Example 8. *For the term $f(a,a)$ and rules $a \to b$, $a \to c$, $f(\mathsf{v}_1,\mathsf{v}_1) \to \dots$, let Φ contract the f-redex and Ψ contract both a-redexes but to b and c respectively. Contracting a to b in the decomposed term $\mathsf{let}\ \mathsf{v}_1 = a\ \mathsf{in}\ f(\mathsf{v}_1,\mathsf{v}_1)$ yields $\mathsf{let}\ \mathsf{v}_1 = b\ \mathsf{in}\ f(\mathsf{v}_1,\mathsf{v}_1)$ i.e. $f(b,b)$ which is distinct but reachable from the result $f(b,a)$ of contracting the left a to b.*

As the example shows, due to sharing we do not have in general that steps all belonging to either Φ or Ψ can be done in parallel on the decomposition; *either* of the Ψ-redexes can be contracted but not *both*, on the decomposition. This is as intended: it witnesses that there *is* overlap, albeit only after enforcing the sharing due to Φ.

Lemma 5 (Huet)**.** *A term rewrite system is locally confluent iff its critical peaks are joinable.*

Proof. Let (ϕ,ψ) be a critical peak. If the pair is trivial, then we conclude trivially, if critical, then by assumption, and if empty, then we conclude by swapping the steps. Otherwise, ϕ, ψ are in distinct components of the decomposition of $\phi \sqcup \psi$, and a common reduct is found by contracting them *simultaneously*, which is reachable from the targets of ϕ, ψ by Lemma 4. □

Cleanness and constructiveness is the main goal of this approach. For example, Okui's confluence criterion was implemented using the same definition of criticality, as in the appendix.

References

[1] N. Hirokawa, J. Nagele, V. van Oostrom, and M. Oyamaguchi. Critical peaks redefined, $\Phi \sqcup \Psi = \top$. In B. Accattoli and B. Felgenhauer, editors, *Proc. 6th IWC*, pages 33–37, 2017. https://arxiv.org/pdf/1708.07877v1.pdf.

[2] Yves Métivier. About the rewriting systems produced by the Knuth–Bendix completion algorithm. *Inf. Process. Lett.*, 16(1):31–34, 1983.

[3] M.S. Paterson and M.N. Wegman. Linear unification. *Journal of Computer and System Sciences*, 16(2):158–167, 1978.

[4] J.E.W. Smetsers. *Graph Rewriting and Functional Languages.* PhD thesis, Katholieke Universiteit Nijmegen, 1993.

A Appendix

```
okuiMultiOne :: Signature a => SemiClosure a -> LocalPeak a -> SemiValley a
okuiMultiOne d (LocalPeak z dl dr) = if (lpeakSize p) <= 1 then emptyorcritical else recurse where
  p = LocalPeak z l r
  l = bb dl -- normalise input
  r = bb dr
  emptyorcritical = if (lpeakCritical p) then (if (lpeakTrivial p) then (sValleyTrivial z (tgt z (lmstp p))) else critical) else empty where
    critical = varSValleyCompose (sclosure d (fst dp)) (snd dp) where -- search for -o->*;<-o- closure in list of closures, solve, and var inst
      dp = varLPeakDecompose p -- decomposition into critical peak and variable instantiation
    empty = (SemiValley z [r'] l') where
      r' = if (isGap (trmHead (skeleton l))) then InduCluster (tgt z l) [] else r
      l' = if (isGap (trmHead (skeleton r))) then InduCluster (tgt z r) [] else l
  recurse = aux (lpeakDecompose p) where
    aux (t,v,b) = sValleyCompose v1 v v2 where
      v1 = okuiMultiOne d t
      v2 = okuiMultiOne d b
```