Structure-Constrained Process Graphs for the Process Semantics of Regular Expressions

Clemens Grabmayer
Gran Sasso Science Institute
Viale F. Crispi, 7, 67100 L'Aquila AQ, Italy
clemens.grabmayer@gssi.it

Milner (1984) introduced a process semantics for regular expressions as process graphs. Unlike for the language semantics, where every regular (that is, DFA-accepted) language is the interpretation of some regular expression, there are finite process graphs that are not bisimilar to the process interpretation of any regular expression. For reasoning about graphs that are expressible by a regular expression it is desirable to have structural representations of process graphs in the image of the interpretation.

For '1-free' regular expressions, their process interpretations satisfy the structural property LEE (loop existence and elimination). But this is not always the case for general regular expressions, as we show by examples. Yet as a remedy, we describe the possibility to recover the property LEE for a close variant of the process interpretation. For this purpose we refine the process semantics of regular expressions to yield process graphs with 1-transitions, similar to silent moves for finite-state automata.

1 Introduction

Milner [7] (1984) defined a process semantics for regular expressions as process graphs: the interpretation of 0 is deadlock, of 1 is successful termination, letters a are atomic actions, the operators $+$ and $\cdot$ stand for choice and concatenation of processes, and (unary) Kleene star $(\cdot)^*$ represents iteration with the option to terminate successfully after each pass-through. In order to disambiguate the use of regular expressions for denoting processes, Milner called them 'star expressions' in this context. Unlike for the standard language semantics, where every regular language is the interpretation of some regular expression, there are finite process graphs that are not bisimilar to the process interpretation of any star expression.[1] This phenomenon led Milner to the formulation of two natural questions: (R) the problem of recognizing whether a given process graph is bisimilar to one in the image of the process interpretation of a star expression, and (A) whether a natural adaptation of Salomaa's complete proof system for language equivalence of regular expressions is complete for bisimilarity of the process interpretation of star expressions. While (R) has been shown to be decidable in principle, so far only partial solutions have been obtained for (A).

For tackling these problems it is expedient to obtain structural representations of process graphs in the image of the interpretation. The result of Baeten, Corradini, and myself [1] that the problem (R) is decidable (but not yet efficiently so) was based on the concept of 'well-behaved (recursive) specifications' that links process graphs with star expressions. Recently in [4, 5], Wan Fokkink and I obtained a partial solution for (A) in the form of a complete proof system for '1-free' star expressions, which do not contain 1, but are formed with binary Kleene star iteration $(\cdot)^{\circledast}(\cdot)$ instead of unary iteration. For this, we defined the efficiently decidable 'loop existence and elimination property (LEE)' of process graphs that holds for all process graph interpretations of 1-free star expressions, and for their bisimulation collapses.

The property LEE does unfortunately not hold for process interpretations of all star expressions. But the aim of this extended abstract is to describe how LEE can nevertheless be made applicable, by stepping

[1] E.g., the process graphs $\mathscr{C}_1^{(\mathrm{ne})}$ and $\mathscr{C}_2^{(\mathrm{ne})}$ in Ex. 2.1 on page 3 are not expressible by a star expression modulo bisimilarity.

Preliminary Report. Final version to appear in:
TERMGRAPH 2020

© C. Grabmayer
This work is licensed under the
Creative Commons Attribution License.

over to a variant of the process interpretation. In Section 3 we explain the loop existence and elimination property LEE for process graphs by means of an example, and define the concept of a 'layered LEE-witness', for short a 'LLEE-witness' for process graphs. That section is an adaptation of Section 3 in [4] from 1-free star expressions to general star expressions. LEE-witnesses arise by adding natural-number labels to transitions that are subject to suitable constraints. In Section 4 we explain examples that show that LEE does not hold in general for process interpretations of star expressions from the full class. As a remedy, in Section 5 we introduce process graphs with 1-transitions (similar to silent moves for finite-state automata), and define a variant $\underline{\mathscr{C}}(\cdot)$ of the process interpretation $\mathscr{C}(\cdot)$ that produces such graphs. We formulate a theorem that states that the process interpretation $\mathscr{C}(e)$ of a star expression e and its variant $\underline{\mathscr{C}}(e)$ are bisimilar. Finally, we explain how the definition of the variant process interpretation $\underline{\mathscr{C}}(\cdot)$ can be refined so as to also define LLEE-witnesses. In this way we obtain that process graphs with 1-transitions in the image of $\underline{\mathscr{C}}(\cdot)$ satisfy LEE.

We are hopeful that the extension of the applicability of the property LEE to the full class of star expressions can be part of a solution of problem (A), based on the partial solution in [4]. We also expect that it can lead to an efficient decision procedure for the recognition problem (R).

The idea to define structure-constrained process graphs via edge-labelings with constraints, on which LEE is based, originated from 'higher-order term graphs' that can be used for representing functional programs in a maximally compact, shared form (see [6, 3]). There, additional concepts (scope sets of vertices, or abstraction-prefix labelings) are used to constrain the form of term graphs. The common underlying idea with the situation we consider here is an enrichment of graphs that: (i) guarantees that graphs can be directly expressed by terms of some language, (ii) does not significantly hamper sharing of represented subterms, (iii) is simple enough so as to keep reasoning about graph transformations feasible.

2 Preliminaries on the process semantics of star expressions

The set $\mathit{StExps}(A)$ of *star expressions over (actions in) A* is defined by the following grammar:

$$e, e_1, e_2 \;::=\; 0 \mid 1 \mid a \mid e_1 + e_2 \mid e_1 \cdot e_2 \mid e^* \qquad (\text{where } a \in A)\,.$$

The *(syntactic) star height* $|e|_*$ of a star expression $e \in \mathit{StExps}(A)$ denotes the maximal nesting depth of stars in e via: $|0|_* := |1|_* := |a|_* := 0$, $|e_1 + e_2|_* := |e_1 \cdot e_2|_* := \max\{|e_1|_*, |e_2|_*\}$, and $|e^*|_* := 1 + |e|_*$.

The process semantics of star expressions is defined by the transition system specification (TSS) $\mathscr{T}$:

$$\frac{}{1{\downarrow}} \qquad \frac{e_i{\downarrow}}{(e_1+e_2){\downarrow}}\,(i\in\{1,2\}) \qquad \frac{e_1{\downarrow} \quad e_2{\downarrow}}{(e_1\cdot e_2){\downarrow}} \qquad \frac{}{(e^*){\downarrow}}$$

$$\frac{}{a \xrightarrow{a} 1} \qquad \frac{e_i \xrightarrow{a} e_i'}{e_1+e_2 \xrightarrow{a} e_i'}\,(i\in\{1,2\}) \qquad \frac{e_1 \xrightarrow{a} e_1'}{e_1\cdot e_2 \xrightarrow{a} e_1'\cdot e_2} \qquad \frac{e_1{\downarrow} \quad e_2 \xrightarrow{a} e_2'}{e_1\cdot e_2 \xrightarrow{a} e_2'} \qquad \frac{e \xrightarrow{a} e'}{e^* \xrightarrow{a} e'\cdot e^*}$$

If $e \xrightarrow{a} e'$ is derivable in $\mathscr{T}$, for $e, e' \in \mathit{StExps}(A)$, and $a \in A$, then we say that e' is a *derivative* of e. If $e{\downarrow}$ is derivable in $\mathscr{T}$, for $e \in \mathit{StExps}(A)$, then we say that *e permits immediate termination*.

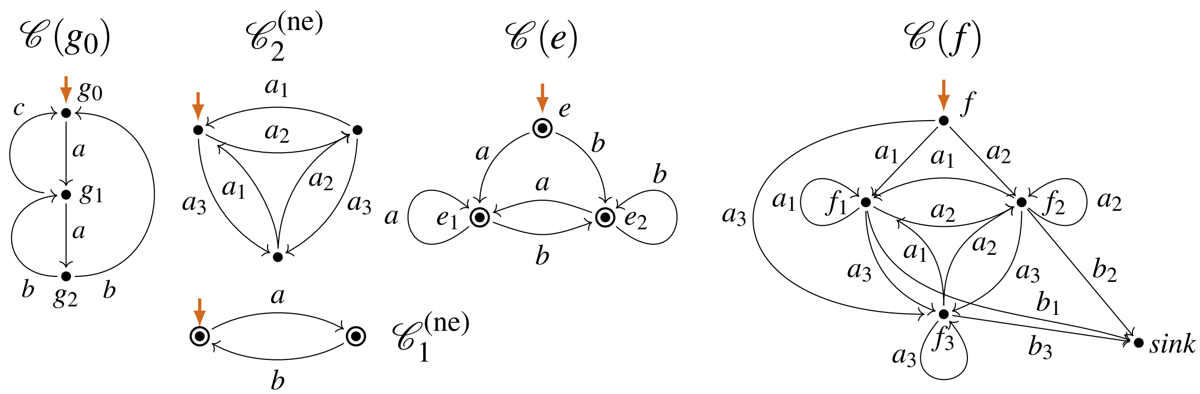
The TSS $\mathscr{T}$ defines the *labeled transition system (LTS)* $\mathscr{S}(\mathit{StExps}(A)) = \langle \mathit{StExps}(A), A, \xrightarrow{(\cdot)}, {\downarrow}\rangle$ with *transitions* $\xrightarrow{(\cdot)} \subseteq \mathit{StExps}(A) \times A \times \mathit{StExps}(A)$, and *immediate-termination property* ${\downarrow} \subseteq \mathit{StExps}(A)$ that are defined in a natural way via derivations in $\mathscr{T}$. For every set $S \subseteq \mathit{StExps}(A)$ we denote by $\mathscr{S}(S)$ the S-generated sub-LTS $\langle V_S, A, T_S, F_S\rangle$ of $\mathscr{S}(\mathit{StExps}(A))$, that is, the sub-LTS whose objects are those in S together with all star expressions that are reachable from ones in S via transitions of $\mathscr{S}(\mathit{StExps}(A))$.

A *chart* is a (rooted) LTS (with initial state) $\langle V, A, v_s, T, F\rangle$ that consists of a finite set V of *vertices*, a finite set A of *actions*, a specified *start vertex* (initial state) $v_s \in V$, a set $T \subseteq V \times A \times V$ of *labeled transitions* between vertices with action labels, and a subset $F \subseteq V$ of vertices *with immediate termination*.

The *chart interpretation* $\mathscr{C}(e) = \langle V(e), A, e, T(e), F(e)\rangle$ of a star expression $e \in \mathit{StExps}(A)$ is of the $\{e\}$-generated sub-LTS $\mathscr{S}(\{e\}) = \langle V_{\{e\}}, A, T_{\{e\}}, F_{\{e\}}\rangle$ of $\mathscr{S}(\mathit{StExps}(A))$.

Let $\mathscr{C}_i = \langle V_i, A, v_{\mathrm{s},i}, T_i, F_i\rangle$ for $i \in \{1,2\}$ be charts. By $\mathscr{C}_1 \leftrightarrows \mathscr{C}_2$ we denote that $\mathscr{C}_1$ and $\mathscr{C}_2$ are *bisimilar*, that is, that there is a *bisimulation* between $\mathscr{C}_1$ and $\mathscr{C}_2$: a relation $B \subseteq V_1 \times V_2$ that relates their start vertices, and whose elements $\langle v_1, v_2\rangle \in B$ satisfy the customary forth-, back-, and termination conditions. By $\mathscr{C}_1 \rightrightarrows \mathscr{C}_2$ we denote the statement that there is a *functional bisimulation* between $\mathscr{C}_1$ and $\mathscr{C}_2$, that is, there is a function $f : V_1 \to V_2$ such that its graph $\{\langle v, f(v)\rangle \mid v \in V_1\}$ is a bisimulation between $\mathscr{C}_1$ and $\mathscr{C}_2$.

Example 2.1. In the chart illustrations below and later, we indicate the start vertex by a brown arrow, and the property of a vertex v to permit immediate termination by highlighting the bullet that symbolizes v by an additional boldface ring. Each of the vertices of the charts $\mathscr{C}_1^{(\mathrm{ne})}$ and $\mathscr{C}(e)$ below permits immediate termination, but none of the vertices of the other charts does so.



The charts $\mathscr{C}_1^{(\mathrm{ne})}$ and $\mathscr{C}_2^{(\mathrm{ne})}$ are not expressible by the process interpretation modulo $\leftrightarrows$, as shown by Bosscher [2] and Milner [7]. That $\mathscr{C}_2^{(\mathrm{ne})}$ is not expressible, Milner proved by observing the absence of a 'loop behaviour'. That concept has inspired the stronger concept of 'loop chart' in Def. 3.1 below. The weaker result that $\mathscr{C}_1^{(\mathrm{ne})}$ and $\mathscr{C}_2^{(\mathrm{ne})}$ are not expressible by 1-free star expressions will be argued in Remark 3.3.

The chart $\mathscr{C}(g_0)$ on the left above is the interpretation of the star expression $g_0 = ((1 \cdot a) \cdot g) \cdot 0$ where $g = (c \cdot a + a \cdot (b + b \cdot a))^*$, and with $g_1 = (1 \cdot g) \cdot 0$, and $g_2 = ((1 \cdot (b + b \cdot a)) \cdot g) \cdot 0$ as remaining vertices. The chart $\mathscr{C}(e)$ is the interpretation of $e = (a^* \cdot b^*)^*$ with $e_1 = ((1 \cdot a^*) \cdot b^*) \cdot e$, and $e_2 = (1 \cdot b^*) \cdot e$. With $f_0 = a_1 \cdot (1 + b_1 \cdot 0) + a_2 \cdot (1 + b_2 \cdot 0)) + a_3 \cdot (1 + b_3 \cdot 0))$, the chart $\mathscr{C}(f)$ is the interpretation of $f = f_0^* \cdot 0$ with $f_i = (1 \cdot (1 + b_i \cdot 0) \cdot f_0^*) \cdot 0$ (for $i \in \{1,2,3\}$), and $\mathit{sink} = ((1 \cdot 0) \cdot f_0^*) \cdot 0$. The chart interpretations $\mathscr{C}(e)$ and $\mathscr{C}(f)$, which will be used later, have been constructed as expressible variants of the not expressible charts $\mathscr{C}_1^{(\mathrm{ne})}$ and $\mathscr{C}_2^{(\mathrm{ne})}$. In particular, $\mathscr{C}(e)$ contains $\mathscr{C}_1^{(\mathrm{ne})}$ as a subchart, and $\mathscr{C}(f)$ contains $\mathscr{C}_2^{(\mathrm{ne})}$ as a subchart (where a 'subchart' arises by taking a part of a chart, and picking a start vertex). We finally note that all of these charts with the exception of $\mathscr{C}(e)$ are bisimulation collapses.

3 Loop existence and elimination

The chart translation $\mathscr{C}(g_0)$ of g_0 as in Ex. 2.1 satisfies the 'loop existence and elimination' property LEE that we will explain in this section. For this purpose we summarize Section 3 in [4], and in doing so we adapt the concepts defined there from 1-free star expressions to the full class of star expressions as defined in Section 2. The property LEE is defined by a dynamic elimination procedure that analyses the structure of a chart by peeling off 'loop subcharts'. Such subcharts capture, within the chart interpretation of a star expression e, the behavior of the iteration of f within innermost subterms f^* in e.

Definition 3.1. A chart $\mathscr{L} = \langle V, A, v_s, T, F \rangle$ is a *loop chart* if:

(L1) There is an infinite path from the start vertex v_s.

(L2) Every infinite path from v_s returns to v_s after a positive number of transitions (and so visits v_s infinitely often).

(L3) Immediate termination is only permitted at the start vertex, that is, $F \subseteq \{v_s\}$.

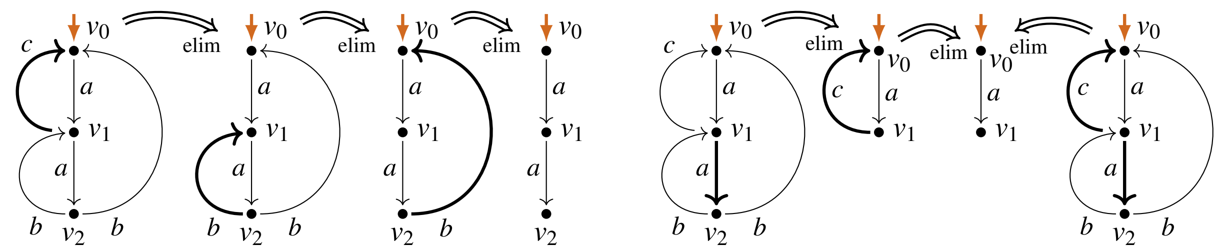
We call the transitions from v_s *loop-entry transitions*, and all other transitions *loop-body transitions*. A loop chart $\mathscr{L}$ is a *loop subchart of* a chart $\mathscr{C}$ if it is the subchart of $\mathscr{C}$ rooted at some vertex $v \in V$ that is generated, for a nonempty set U of transitions of $\mathscr{C}$ from v, by all paths that start with a transition in U and continue onward until v is reached again (so the transitions in U are the loop-entry transitions of $\mathscr{L}$).

Both of the not expressible charts $\mathscr{C}_1^{(\mathrm{ne})}$ and $\mathscr{C}_2^{(\mathrm{ne})}$ in Ex. 2.1 are not loop charts: $\mathscr{C}_1^{(\mathrm{ne})}$ violates (L3), and $\mathscr{C}_2^{(\mathrm{ne})}$ violates (L2). Moreover, none of these charts contains a loop subchart. The chart $\mathscr{C}(g_0)$ in Ex. 2.1 is not a loop chart either, as it violates (L2). But we will see that $\mathscr{C}(g_0)$ has loop subcharts.

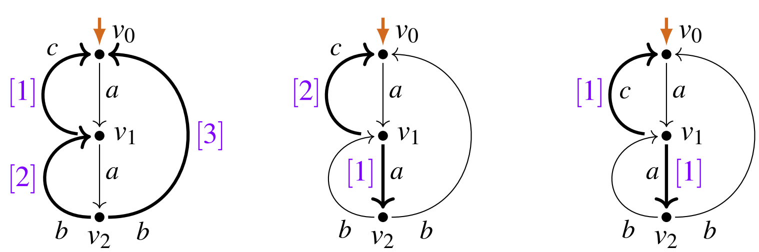
Let $\mathscr{L}$ be a loop subchart of a chart $\mathscr{C}$. Then the result of *eliminating* $\mathscr{L}$ *from* $\mathscr{C}$ arises by removing all loop-entry transitions of $\mathscr{L}$ from $\mathscr{C}$, and then removing all vertices and transitions that become unreachable. We say that a chart $\mathscr{C}$ has the *loop existence and elimination property (LEE)* if the procedure, started on $\mathscr{C}$, of repeated eliminations of loop subcharts results in a chart that does not have an infinite path.

For the not expressible charts $\mathscr{C}_1^{(\mathrm{ne})}$ and $\mathscr{C}_2^{(\mathrm{ne})}$ in Ex. 2.1 the procedure stops immediately, as they do not contain loop subcharts. Since both of them have infinite paths, it follows that they do not satisfy LEE.

Now we consider three runs of the elimination procedure for the chart $\mathscr{C}(g_0)$ in Ex. 2.1. The loop-entry transitions of loop subcharts that are removed in each step are marked in bold.



Each run witnesses that $\mathscr{C}(g_0)$ satisfies LEE. Note that loop elimination does not yield a unique result.[2] Runs can be recorded, in the original chart, by attaching a marking label to transitions that get removed in the elimination procedure. That label is the sequence number of the corresponding elimination step. For the three runs of loop elimination above we get the following marking labeled versions of $\mathscr{C}$, respectively:



Since all three runs were successful (as they yield charts without infinite paths), these recordings (marking-labeled charts) can be viewed as 'LEE-witnesses'. We now will define the concept of a 'layered LEE-witness' (LLEE-witness), i.e., a LEE-witness with the added constraint that in the recorded run of the loop

[2]Confluence can be shown if a pruning operation is added that permits to drop transitions to deadlocking vertices.

elimination procedure it never happens that a loop-entry transition is removed from within the body of a previously removed loop subchart. This refined concept has simpler properties, but is equally powerful.

By an *entry/body-labeling of a chart* $\mathscr{C} = \langle V, A, v_s, T, F \rangle$ we mean a chart $\widehat{\mathscr{C}} = \langle V, A \times \mathbb{N}, v_s, \widehat{T}, F \rangle$ that arises from $\mathscr{C}$ by adding, for each transition $\tau = \langle v_1, a, v_2 \rangle \in T$, to the action label a of τ a *marking label* $\alpha \in \mathbb{N}$, yielding $\widehat{\tau} = \langle v_1, \langle a, \alpha \rangle, v_2 \rangle \in \widehat{T}$. In such an entry/body-labeling we call transitions with marking label 0 *body transitions*, and transitions with marking labels in $\mathbb{N}^+$ *entry-transitions*.

Let $\widehat{\mathscr{C}}$ be an entry/body-labeling of $\mathscr{C}$, and let v and w be vertices of $\mathscr{C}$ and $\widehat{\mathscr{C}}$. We denote by $v \to_{\text{bo}} w$ that there is a body-transition $v \xrightarrow{\langle a,0 \rangle} w$ in $\widehat{\mathscr{C}}$ for some $a \in A$, and by $v \to_{[\alpha]} w$, for $\alpha \in \mathbb{N}^+$ that there is an entry-transition $v \xrightarrow{\langle a,\alpha \rangle} w$ in $\widehat{\mathscr{C}}$ for some $a \in A$. By the set $E(\widehat{\mathscr{C}})$ of *entry-transition identifiers* we denote the set of pairs $\langle v, \alpha \rangle \in V \times \mathbb{N}^+$ such that an entry-transition $\to_{[\alpha]}$ departs from v in $\widehat{\mathscr{C}}$. For $\langle v, \alpha \rangle \in E(\widehat{\mathscr{C}})$, we define by $\mathscr{C}_{\widehat{\mathscr{C}}}(v, \alpha)$ the subchart of $\mathscr{C}$ with start vertex v_s that consists of the vertices and transitions which occur on paths in $\mathscr{C}$ as follows: they start with a $\to_{[\alpha]}$ entry-transition from v, continue with body transitions only, and halt immediately if v is revisited.

The three recordings obtained above of the loop elimination procedure for the chart $\mathscr{C}(g_0)$ in Ex. 2.1 indicate entry/body-labelings by signaling the entry-transitions but neglecting body-step labels 0.

Definition 3.2. Let chart $\mathscr{C} = \langle V, A, v_s, T, F \rangle$ be a chart. A LLEE-*witness* (a *layered* LEE-*witness*) *of* $\mathscr{C}$ is an entry/body-labeling $\widehat{\mathscr{C}}$ of $\mathscr{C}$ that satisfies the following two properties:

(W1) There is no infinite path of $\to_{\text{bo}}$ transitions from the start vertex v_s of $\mathscr{C}$.

(W2) For all $\langle v, \alpha \rangle \in E(\widehat{\mathscr{C}})$,
  (a) (*loop condition*) $\mathscr{C}_{\widehat{\mathscr{C}}}(v, \alpha)$ is a loop chart, and
  (b) (*layeredness*) if an entry-transition $\to_{[\beta]}$ departs from a vertex $w \neq v$ of $\mathscr{C}_{\widehat{\mathscr{C}}}(v, \alpha)$, then its marking label satisfies $\beta < \alpha$.

The condition (W2)(a) justifies to call an entry-transition in a LLEE-witness a *loop-entry transition*. For a loop-entry transition $\to_{[\beta]}$ with $\beta \in \mathbb{N}^+$, we call β its *loop level*.

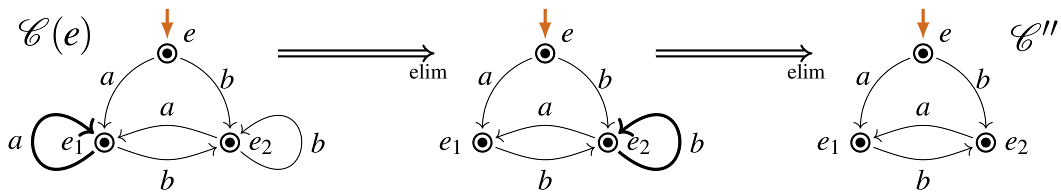
The condition (W2)(b) on a LLEE-witness $\widehat{\mathscr{C}}$ of a chart $\mathscr{C}$ requires the loop structure defined by $\widehat{\mathscr{C}}$ to be hierarchical. This permits to extract a star expression $\tilde{e}$ from $\widehat{\mathscr{C}}$ (defined in [4, 5]) that expresses $\mathscr{C}$ in the sense that $\mathscr{C}(\tilde{e}) \;\underline{\leftrightarrow}\; \mathscr{C}$ holds, intuitively by unfolding the underlying chart $\mathscr{C}$ to the syntax tree of $\tilde{e}$.

For the three entry/body-labelings of the chart $\mathscr{C}(g_0)$ in Ex. 2.1 that we have obtained above as recordings of runs of the loop elimination procedure it is easy to verify that they are LLEE-witnesses of $\mathscr{C}(g_0)$.

Remark 3.3. In [4, 5] we established a connection between charts that have a LLEE-witness (and hence satisfy LEE) and charts that are expressible by 1-free star expressions (that is, star expressions without 1, and with binary star iteration instead of unary star iteration). By saying that a chart $\mathscr{C}$ 'is expressible' we mean here that $\mathscr{C}$ is bisimilar to the chart interpretation $\mathscr{C}(\tilde{e})$ of some star expression $\tilde{e}$. Now Corollary 6.10 in [4, 5] states that if a chart is expressible by a 1-free star expression then its bisimulation collapse has a LLEE-witness, and thus satisfies LEE. This statement entails that neither of the charts $\mathscr{C}_1^{(\text{ne})}$ and $\mathscr{C}_2^{(\text{ne})}$ in Ex. 2.1 is expressible by a 1-free star expression, because both are bisimulation collapses, and neither of them satisfies LEE, as we have already observed above.

4 LEE may fail for process interpretations of star expressions

The chart interpretations $\mathscr{C}(e)$ of e, and $\mathscr{C}(f)$ of f in Ex. 2.1 do not satisfy LEE, contrasting with $\mathscr{C}(g_0)$. For $\mathscr{C}(e)$ we find the following run of the loop elimination procedure that successively eliminates the two loop subcharts induced by the cycling transitions at e_1, and at e_2:



The resulting chart $\mathscr{C}''$ does not contain loop subcharts any more, because taking, for example, a transition from e_1 to e_2 as an entry-transition does not yield a loop subchart, because in the induced subchart immediate termination is not only possible at the start vertex e_1 but also in the body vertex e_2, in contradiction to (L3). But while $\mathscr{C}''$ does not contain a loop subchart any more, it still has an infinite trace. Therefore it follows that $\mathscr{C}(e)$ does not satisfy LEE.

In order to see that $\mathscr{C}(f)$ does not satisfy LEE, we can consider a run of the loop elimination procedure that successively removes the cyclic transitions at f_1, f_2, and f_3. After these removals a variant of the not expressible chart $\mathscr{C}_2^{(\mathrm{ne})}$ is obtained that still describes an infinite behavior, but that does not contain any loop subchart. The latter can be argued analogously as for $\mathscr{C}_2^{(\mathrm{ne})}$, namely that for all choices of entry-transitions between f_1, f_2, and f_3 the loop condition (L2) fails. We conclude that $\mathscr{C}(f)$ does not satisfy LEE.

The reason for this failure of LEE is that, while the syntax trees of star expressions provide a nested-loop like structure, this is not guaranteed by the specific form of the TSS $\mathscr{T}$. Execution of an iteration g^* in an expression $g^* \cdot h$ leads eventually, in case that termination is reachable in g, to an iterated derivative $(1 \cdot g^*) \cdot h$. Also, as in the examples above, an iterated derivative $(\tilde{g} \cdot g^*) \cdot h$ with $\tilde{g}{\downarrow}$ may be reached. In these cases, continued execution will bypass the initial term $g^* \cdot h$, and either proceed with another execution of the iteration to $(g' \cdot g^*) \cdot h$, where g' is a derivative of g, or take a step into the exit to h', where h' is a derivative of h. In both cases the execution does not return to the initial term g^* of the execution, as would be required for a loop subchart at g^* to arise in accordance with loop condition (L2).

5 Recovering LEE for a variant definition of the process semantics

A remedy for the frequent failure of LEE for the chart translation of star expressions can consist in the use of '1-transitions'. Such transitions may be used to create a back-link to an expression $g^* \cdot h$ from an iterated derivative $(\tilde{g} \cdot g^*) \cdot h$ with $\tilde{g}{\downarrow}$ (where $\tilde{g}$ is an iterated derivative of g) that is reached by a descent of the execution into the body of g. This requires an adapted refinement of the TSS $\mathscr{T}$ from page 2.

In particular we want to create transition rules that facilitate a back-link to an expression g^* after the execution has descended into g reaching $\tilde{g} \cdot g^*$ with $\tilde{g}{\downarrow}$. In order to distinguish a concatenation expression $\tilde{g} \cdot g^*$ that arises from the descent of the execution into an iteration g^* from other concatenation expressions we introduce a variant operation $*$. The rules of the refined TSS should guarantee that in the example the reached iterated derivative of g^* is a 'stacked star expression' $G * g^*$ where G is itself a stacked star expression that denotes an iterated derivative of g. If now G is also a star expression $\tilde{g}$ with $\tilde{g}{\downarrow}$, then the expression $G * g^*$ of the form $\tilde{g} * g^*$ should permit a 1-transition that returns to g^*.

This intuition guided the definition of the rules of the TSS $\underline{\mathscr{T}}^{(*)}$ in Def. 5.2 below, starting from the adaptation of the rule for steps from iterations e^*, and the rule that creates 1-transition backlinks to iterations e_2^* from stacked expressions $e_1 * e_2^*$ with $e_1{\downarrow}$. The 'stacked product' $*$ has the following features: $E_1 * E_2$ never permits immediate termination; for defining transitions it behaves similarly as concatenation $\cdot$ except that a transition from $E_1 * E_2$ into E_2 when E_1 permits immediate termination now requires a 1-transition to E_2 first. The formulation of these rules of $\underline{\mathscr{T}}^{(*)}$ led to the tailor-made set of stacked star expressions as defined below.

Definition 5.1. The set $StExps^{(*)}(A)$ of *stacked star expressions over (actions in) A* is defined by:

$$E ::= e \mid E \cdot e \mid E * e^* \qquad (\text{where } e \in StExps(A)).$$

The *projection function* $\pi : StExps^{(*)}(A) \to StExps(A)$ is defined by interpreting $*$ as $\cdot$ by the clauses: $\pi(E \cdot e) := \pi(E) \cdot e$, $\pi(E * e^*) := \pi(E) \cdot e^*$, and $\pi(e) := e$, for all $E \in StExps^{(*)}(A)$, and $e \in StExps(A)$.

Definition 5.2. The transition system specification $\underline{\mathscr{T}}^{(*)}(A)$ has the following axioms and rules, where $1 \notin A$ is an additional label (for representing empty steps), $a \in A$, $\underline{a} \in \underline{A} := A \cup \{1\}$, stacked star expressions $E_1, E_2, E_1', E_2', E' \in \underline{\mathscr{T}}^{(*)}(A)$, and star expressions $e_1, e_2, e_2^*, e^* \in StExps(A)$ (here and below we highlight in red transitions that may involve 1-transitions):

$$\frac{}{1\downarrow} \qquad \frac{e_i\downarrow}{(e_1+e_2)\downarrow}\,(i\in\{1,2\}) \qquad \frac{e_1\downarrow \quad e_2\downarrow}{(e_1\cdot e_2)\downarrow} \qquad \frac{}{(e^*)\downarrow}$$

$$\frac{}{a \xrightarrow{a} 1} \qquad \frac{e_i \xrightarrow{a} E_i'}{e_1+e_2 \xrightarrow{a} E_i'}\,(i\in\{1,2\}) \qquad \frac{E_1 \xrightarrow{\underline{a}} E_1'}{E_1\cdot e_2 \xrightarrow{\underline{a}} E_1'\cdot e_2} \qquad \frac{e_1\downarrow \quad e_2 \xrightarrow{a} E_2'}{e_1\cdot e_2 \xrightarrow{a} E_2'} \qquad \frac{e \xrightarrow{a} E'}{e^* \xrightarrow{a} E' * e^*}$$

$$\frac{E_1 \xrightarrow{\underline{a}} E_1'}{E_1 * e_2^* \xrightarrow{\underline{a}} E_1' * e_2^*} \qquad \frac{e_1\downarrow}{e_1 * e_2^* \xrightarrow{1} e_2^*}$$

Via its derivations the TSS $\underline{\mathscr{T}}^{(*)}$ defines the 1-LTS $\underline{\mathscr{L}}(StExps^{(*)}(A)) = \langle StExps^{(*)}(A), A, 1, \xrightarrow{\cdot}, \downarrow\rangle$ with separate 1-transitions. For $S \subseteq StExps^{(*)}(A)$, S-generated sub-LTSs are defined similarly as for $\mathscr{L}(StExps(A))$.

A 1-*chart* is a (rooted) 1-LTS $\langle V, A, 1, v_s, T, F\rangle$ that consists of a finite set V of *vertices*, a finite set A of *proper actions*, a specified symbol 1 with $1 \notin A$, a *start vertex* $v_s \in V$, a set $T \subseteq V \times \underline{A} \times V$ of *labeled transitions* with $\underline{A} := A \cup \{1\}$, and a set $F \subseteq V$ of vertices *with immediate termination*. For such a 1-chart we understand by a *proper transition* a transition in $T \cap (V \times A \times V)$ (labeled by a *proper action*), and by a 1-*transition* a transition in $T \cap (V \times \{1\} \times V)$ (labeled by the *empty-step symbol* 1).

The 1-*chart interpretation* $\underline{\mathscr{C}}(e) = \langle V(e), A, 1, e, T(e), F(e)\rangle$ of a star expression $e \in StExps(A)$ is the e-rooted version of the $\{e\}$-generated sub-1-LTS $\underline{\mathscr{L}}(\{e\}) = \langle V_{\{e\}}, A, T_{\{e\}}, F_{\{e\}}\rangle$ of $\underline{\mathscr{L}}(StExps(A))$.

In order to link the 1-LTS $\mathscr{L}(StExps^{(*)}(A))$ to the LTS $\mathscr{L}(StExps(A))$, we need to take account of the semantics of 1-transitions as empty steps (see Vrancken [8]). For this, we introduce the 'induced LTS' $(\mathscr{L}(StExps^{(*)}(A))]$ of $\mathscr{L}(StExps^{(*)}(A))$ with *induced transitions* $\xrightarrow{(\cdot]}$, and *induced termination* $\downarrow_{(1)}$ that are defined as follows: $E \xrightarrow{(a]} E'$ holds if there is a sequence of 1-transitions from E to some $\tilde{E}$ from which there is an a-transition to E', and $E\downarrow_{(1)}$ holds if there is a sequence of 1-transitions from E to some $\tilde{E}$ with $\tilde{E}\downarrow$. The asymmetric notation $\xrightarrow{(\cdot]}$ is intended to reflect the asymmetry that an induced transition consists of an arbitrary number of leading 1-transitions that is trailed by a single proper transition.

Definition 5.3. The LTS $(\underline{\mathscr{L}}(StExps^{(*)}(A))] = \langle StExps^{(*)}(A), A, \downarrow_{(1)}, \xrightarrow{(\cdot]}\rangle$ is defined via derivations in the TSS $(\underline{\mathscr{T}}^{(*)}]$ that in addition to the axioms and rules of $\underline{\mathscr{T}}^{(*)}$ also contains the following rules:

$$\frac{e\downarrow}{e\downarrow_{(1)}} \qquad \frac{E \xrightarrow{1} \tilde{E} \quad \tilde{E}\downarrow_{(1)}}{E\downarrow_{(1)}} \qquad \frac{E \xrightarrow{a} E'}{E \xrightarrow{(a]} E'} \qquad \frac{E \xrightarrow{1} \tilde{E} \quad \tilde{E} \xrightarrow{(a]} E'}{E \xrightarrow{(a]} E'}$$

Now the following lemma explains the relationship between, on the one hand, the TSS $\mathscr{T}(A)$ and its appertaining LTS $\mathscr{L}(StExps(A))$, and on the other hand, the TSS $(\underline{\mathscr{T}}^{(*)}]$ and its LTS $(\mathscr{L}(StExps^{(*)}(A))]$.

Lemma 5.4. *The LTS $\mathscr{L}(StExps(A))$ is bisimilar to the LTS $(\underline{\mathscr{L}}(StExps^{(*)}(A))]$ via the bisimulation that is defined by the projection function π, that is, for all $E, E' \in StExps^{(*)}(A)$, $e' \in StExps(A)$, and $a \in A$:*

$$\vdash_{\mathscr{T}} \pi(E)\downarrow \iff_{\Leftarrow} \vdash_{(\underline{\mathscr{T}}^{(*)}]} E\downarrow_{(1)},$$

$$\vdash_{\mathscr{T}} \pi(E) \xrightarrow{a} \pi(E') \Longleftarrow \vdash_{(\underline{\mathscr{T}}^{(*)}]} E \xrightarrow{(a]} E',$$

$$\vdash_{\mathscr{T}} \pi(E)\downarrow \Longrightarrow \vdash_{(\underline{\mathscr{T}}^{(*)}]} E\downarrow_{(1)},$$

$$\vdash_{\mathscr{T}} \pi(E) \xrightarrow{a} e' \Longrightarrow \exists E' \in StExps^{(*)}\big[\,\pi(E') = e' \;\wedge\; \vdash_{(\underline{\mathscr{T}}^{(*)}]} E \xrightarrow{(a]} E'\,\big].$$

The *induced chart* $(\underline{\mathscr{C}}(e)] = \langle V, A, v_s, (T], (F) \rangle$ of a 1-chart $\underline{\mathscr{C}} = \langle V, A, 1, v_s, T, F \rangle$ is defined analogously as the induced LTS of a 1-LTS with transitions in $(T]$ that are induced by those in T, and vertices in (F) with induced termination with respect to vertices with immediate termination in F. With this, we now obtain the following connection between the chart, and the 1-chart, interpretation from Lemma 5.4.

Theorem 5.1. *$(\underline{\mathscr{C}}(e)] \rightrightarrows \mathscr{C}(e)$ holds for all $e \in StExps(A)$, that is, there is a functional bisimulation from the induced chart of the* 1*-chart interpretation of a star expression e to the chart interpretation of e.*

Now we define, similarly as we have done so for 1-free star expressions in [4, 5], a refinement of the TSS $\underline{\mathscr{T}}^{(*)}$ into a TSS that will supply entry/body-labelings for LLEE-witnesses, by adding marking labels to the rules of $\underline{\mathscr{T}}^{(*)}$. In particular, body labels are added to transitions that cannot return to their source expression. The rule for transitions from an iteration e^* is split into the case in which e is normed or not (e is called *normed* if it enables an execution to an expression with immediate termination). Only if e is normed can e^* return to itself, and then a (loop-) entry-transition with the star height $|e^*|_*$ of e^* as its level is created; otherwise a body label is introduced.

Definition 5.5. The TSS $\widehat{\underline{\mathscr{T}}}^{(*)}$ has the following rules, where $l \in \{\mathrm{bo}\} \cup \{[\alpha] \mid \alpha \in \mathbb{N}^+\}$:

$$\frac{}{1{\downarrow}} \qquad \frac{e_i{\downarrow}}{(e_1 + e_2){\downarrow}}\,(i \in \{1,2\}) \qquad \frac{e_1{\downarrow} \quad e_2{\downarrow}}{(e_1 \cdot e_2){\downarrow}} \qquad \frac{}{(e^*){\downarrow}}$$

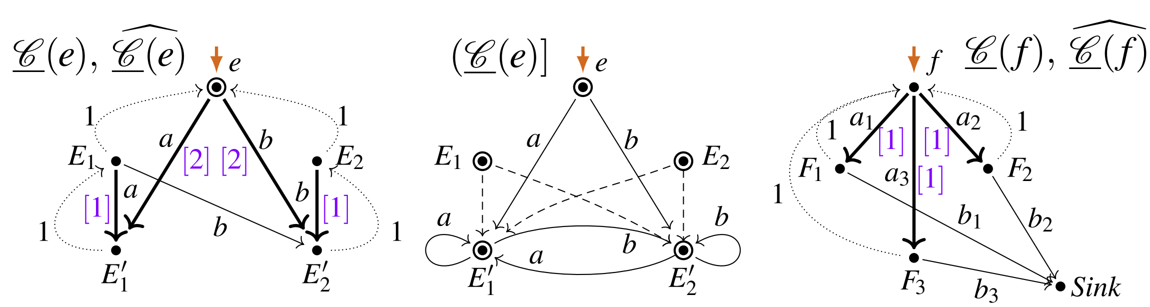
$$\frac{}{a \xrightarrow{a}_{\mathrm{bo}} 1} \qquad \frac{e_i \xrightarrow{a}_{l} E_i'}{e_1 + e_2 \xrightarrow{a}_{\mathrm{bo}} E_i'}\,(i \in \{1,2\}) \qquad \frac{E_1 \xrightarrow{a}_{l} E_1'}{E_1 \cdot e_2 \xrightarrow{a}_{l} E_1' \cdot e_2} \qquad \frac{e_1{\downarrow} \quad e_2 \xrightarrow{a}_{l} E_2'}{e_1 \cdot e_2 \xrightarrow{a}_{\mathrm{bo}} E_2'}$$

$$\frac{e \xrightarrow{a}_{l} E' \quad (e \text{ not normed})}{e^* \xrightarrow{a}_{\mathrm{bo}} E' * e^*} \qquad \frac{e \xrightarrow{a}_{l} E' \quad (e \text{ normed})}{e^* \xrightarrow{a}_{[|e^*|_*]} E' * e^*} \qquad \frac{E_1 \xrightarrow{a}_{l} E_1'}{E_1 * e_2^* \xrightarrow{a}_{l} E_1' * e_2^*} \qquad \frac{e_1{\downarrow}}{e_1 * e_2^* \xrightarrow{1}_{\mathrm{bo}} e_2^*}$$

For every star expression $e \in StExps(A)$ we denote by $\widehat{\underline{\mathscr{C}}(e)}$ the entry/body-labeling of the 1-chart interpretation $\underline{\mathscr{C}}(e)$ that is defined according to the TSS $\widehat{\underline{\mathscr{T}}}^{(*)}$. For this entry/body-labeling we can show the following theorem that recovers the property LEE for the 1-chart interpretation of star expressions.

Theorem 5.2. *For every $e \in StExps(A)$, the entry/body-labeling $\widehat{\underline{\mathscr{C}}(e)}$ of $\underline{\mathscr{C}}(e)$ is a* LLEE*-witness of $\underline{\mathscr{C}}(e)$. Consequently the* 1*-chart interpretation $\underline{\mathscr{C}}(e)$ of a star expression e satisfies the property* LEE*.*

We consider again the chart interpretations $\mathscr{C}(e)$ and $\mathscr{C}(f)$ for the star expressions e and f in Ex. 2.1 for which we saw in Section 4 that LEE fails. The picture below shows the 1-chart interpretations $\underline{\mathscr{C}}(e)$ of e (left) and $\underline{\mathscr{C}}(f)$ of f (right), together with the entry/body-labelings $\widehat{\underline{\mathscr{C}}(e)}$ of $\underline{\mathscr{C}}(e)$ and $\widehat{\underline{\mathscr{C}}(f)}$ of $\underline{\mathscr{C}}(f)$:



The dotted transitions indicate 1-transitions. The non-initial vertices in the 1-chart interpretations are derivatives obtained via the TSS $\underline{\mathscr{T}}^{(*)}$, for $\underline{\mathscr{C}}(e)$: $E_1' = ((1 * a^*) \cdot b^*) * e$, $E_1 = (a^* \cdot b^*) * e$, $E_2 = b^* * e$, and $E_2' = (1 * b^*) * e$, and for $\underline{\mathscr{C}}(f)$: $F_i = (1 \cdot (1 + b_i \cdot 0)) * f$ for $i \in \{1,2,3\}$, and $Sink = (1 \cdot 0) * f$.

The chart interpretations $\mathscr{C}(e)$ of e, and $\mathscr{C}(f)$ of f arise as images of functional bisimulations from the induced charts $(\underline{\mathscr{C}}(e)]$ of $\underline{\mathscr{C}}(e)$, and $(\underline{\mathscr{C}}(f)]$ of $\underline{\mathscr{C}}(f)$, as stated by Thm. 5.1. We argue this as follows.

The transitions of the induced chart $(\underline{\mathscr{C}}(e)]$ (above in the middle) of the 1-chart interpretation $\underline{\mathscr{C}}(e)$ of e correspond to paths in $\underline{\mathscr{C}}(e)$ that start with a (potentially empty) 1-transition path and have a final proper action transition, which also provides the label of the induced transition. For example the b-transition from E_1' to E_2' in $(\underline{\mathscr{C}}(e)]$ arises as the induced transition in $\underline{\mathscr{C}}(e)$ that is the path that consists of the 1-transitions from E_1' to E_1, and from E_1 to e, followed by the final b-transition from e to E_2'. The vertices with immediate termination in $(\underline{\mathscr{C}}(e)]$ are all those that permit 1-transition paths in $\underline{\mathscr{C}}(e)$ to vertices with immediate termination. Therefore in $\underline{\mathscr{C}}(e)$ only e needs to permit immediate termination in order to get induced termination in $(\underline{\mathscr{C}}(e)]$ also at all other vertices (like in $\mathscr{C}(e)$). Now clearly the function that maps $e \mapsto e$, and $E_i, E_i' \mapsto e_i$ for $i \in \{1,2\}$ defines a bisimulation from $(\underline{\mathscr{C}}(e)]$ to $\mathscr{C}(e)$. Indeed this function defines an isomorphism, if the unreachable vertices E_1 and E_2 are removed by garbage collection.

In the second example, the function that maps $f \mapsto f$, $\mathit{Sink} \mapsto \mathit{sink}$, and $F_i \mapsto f_i$ for $i \in \{1,2,3\}$ defines a functional bisimulation from $(\underline{\mathscr{C}}(f)]$ to $\mathscr{C}(f)$, which in fact is an isomorphism.

Furthermore, the entry/body-labelings $\widehat{\underline{\mathscr{C}}(e)}$ of $\underline{\mathscr{C}}(e)$, and $\widehat{\underline{\mathscr{C}}(f)}$ of $\underline{\mathscr{C}}(f)$ as illustrated above can be readily checked to be LLEE-witnesses of $\underline{\mathscr{C}}(e)$ and $\underline{\mathscr{C}}(f)$, respectively.

In this way we have verified the joint claims of Thm. 5.1 and of Thm. 5.2 for the two examples e and f of star expressions from Ex. 2.1 for which, as we saw in Section 4, LEE fails for their chart interpretations $\mathscr{C}(e)$ and $\mathscr{C}(f)$: the property LEE can be recovered for the 1-chart interpretations $\underline{\mathscr{C}}(e)$ of e, and $\underline{\mathscr{C}}(f)$ of f (by Thm. 5.2), and also, the induced charts $(\underline{\mathscr{C}}(e)]$ of $\underline{\mathscr{C}}(e)$, and $(\underline{\mathscr{C}}(f)]$ of $\underline{\mathscr{C}}(f)$ map to the original process interpretations $\mathscr{C}(e)$ of e, and $\mathscr{C}(f)$ of f, respectively, via a functional bisimulation (by Thm. 5.1).

Acknowledgement. I want to thank my co-author Wan Fokkink for suggesting to define LLEE-witnesses directly from the transition system specification, which we worked out for 1-free star expressions in [4, 5]. That inspired me for the generalization here to the full class of star expressions. Also, I want to thank Luca Aceto for his comments on a draft, and the reviewers for their observations, questions, and suggestions.

References

[1] J.C.M. Baeten, F. Corradini & C.A. Grabmayer (2007): *A Characterization of Regular Expressions Under Bisimulation*. Journal of the ACM 54(2), doi:10.1145/1219092.1219094.

[2] Doeko Bosscher (1997): *Grammars Modulo Bisimulation*. Ph.D. thesis, University of Amsterdam.

[3] Clemens Grabmayer (2019): *Modeling Terms by Graphs with Structure-Constraints (Two Illustrations)*. In: TERMGRAPH 2018 post-proceedings, 288, pp. 1–13, doi:10.4204/EPTCS.288.

[4] Clemens Grabmayer & Wan Fokkink (2020): *A Complete Proof System for 1-Free Regular Expressions Modulo Bisimilarity*. In: Proceedings of LICS 2020, ACM, New York, NY, p. To appear. Extended report [5] on arXiv.

[5] Clemens Grabmayer & Wan Fokkink (2020): *A Complete Proof System for 1-Free Regular Expressions Modulo Bisimilarity*. Technical Report `arXiv:2004.12740`, `arxiv.org`.

[6] Clemens Grabmayer & Jan Rochel (2014): *Maximal Sharing in the Lambda Calculus with Letrec*. ACM SIGPLAN Notices 49(9), p. 67–80, doi:10.1145/2692915.2628148.

[7] Robin Milner (1984): *A Complete Inference System for a Class of Regular Behaviours*. Journal of Computer and System Sciences 28, pp. 439–466.

[8] J.L.M. Vrancken (1997): *The Algebra of Communicating Processes with Empty Process*. Theoretical Computer Science 177(2), pp. 287 – 328, doi:https://doi.org/10.1016/S0304-3975(96)00250-2.