On the Star Height of Regular Expressions Under Bisimulation (Extended Abstract)

J.C.M. Baeten

Division of Computer Science, Technische Universiteit Eindhoven,
Eindhoven, The Netherlands, email: josb@win.tue.nl

F. Corradini

Department of Mathematics and Computer Science, Università di Camerino,
Camerino, Italy, email: flavio.corradini@unicam.it

C.A. Grabmayer

Department of Computer Science, Vrije Universiteit Amsterdam,
Amsterdam, The Netherlands, email: clemens@cs.vu.nl

Abstract

We solve the star-height problem for regular expressions under Milner's process interpretation. That is, we describe an algorithmic solution to the problem of finding, for a given regular expression e, the least natural number n such that there exists a regular expression f with star height n that is bisimilar to e. For this, we utilise our recent solution (2005), based on the concept "well-behaved recursive specification", of the problem of deciding whether or not a finite-state process is bisimilar to a regular expression. As yet another instance of reasoning with well-behaved specifications, we give an alternative proof for a result of Hirshfeld and Moller (2000) that solved a star-height question posed by Milner (1984): For every natural number n there exists a regular expression over a singleton alphabet that is not bisimilar to any regular expression of star height less than n.

In [4] Milner introduced an interpretation of regular expressions as "behaviours", bisimulation equivalence classes of processes: a regular expression e is interpreted as a finite-state process $P(e)$ that represents a "star behaviour". Different from the standard language interpretation L, the interpretation P assigns the following meaning to the regular operators: $+$ and $\cdot$ respectively denote *alternative composition* (also called *choice*) between processes, and *sequential composition* of processes; * denotes unbounded iteration of processes with the possibility of termination before each iteration; letters a stand for processes that are executions of *atomic actions*; and the constants 0 and 1 are interpreted as *deadlock*, and as the *empty process* (also called *skip*), respectively. The interpretation P gives rise to the equivalence relation $\leftrightarrows_P$, "bisimilarity of process interpretations" on regular expressions: for all regular expressions e and f, $e \leftrightarrows_P f$ is defined to hold if and only if $P(e) \leftrightarrows P(f)$, where $\leftrightarrows$ denotes bisimilarity.

Milner noticed that not every finite-state process is contained in a star behaviour, and he suggested an adaptation of Salomaa's complete axiom system for language equivalence $=_L$ as an axiomatisation for the equivalence relation $\leftrightarrows_P$. Next to the problems, how finite-state processes that represent a star behaviour can be characterised by a structural property, and whether the modified version of Salomaa's system (or some extension of it) is complete for $\leftrightarrows_P$, he put a third question: Does there exist for every natural number n a regular expression over a singleton alphabet that is not bisimilar to any regular expression of star height less than n?

This star-height question was answered affirmatively by Hirshfeld and Moller in [3]. The question concerning the axiomatisation of $\leftrightarrows_P$ suggested by Milner has frequently been studied and a number of related results have been obtained (we mention work by Sewell (1997), Fokkink (1997), Corradini, De Nicola and Labella (1998)); yet apparently, it is still unsolved. For the characterisation question concerning star behaviours, in [2] we have recently provided an answer based on the concept "well-behaved recursive specification" due to Baeten and Corradini in [1].

In formal language theory, the determination of the star height of regular language is an old problem. The *star height* $sh(e)$ of a regular expression e is defined syntactically as the maximum number of nested stars that e contains. The *star height* $sh(L)$ of a regular language L is the least natural number n such that $\mathrm{sh}(e) = n$ for some regular expression e that represents L. A large number of results concerning the star height of regular languages is known, of which we only mention three: (A) Every regular language over a single-letter alphabet is of star height one at most. (B) There are regular languages with any preassiged star height over any alphabet containing at least two letters (Eggan 1963). (C) There exists an algorithm for computing the star height of a regular language given by a regular expression (Hashiguchi, 1983); this famous statement established that "the star-height problem" (in formal language theory) is solvable.

1 Minimal Star Height under the Process Interpretation

We are interested in counterparts for regular expressions under the process interpretation of the star-height results (A)–(C) for regular languages. For this purpose we define, analogously to the star height of a regular language, the following notion: the *minimal star height* of a regular expression e (*under the process interpretation*) is the least natural number n such that there exists a regular expression e_0 with $\mathrm{sh}(e_0) = n$ and $e_0 \leftrightarrows_P e$. With this definition, the answer given by Hirshfeld and Moller to Milner's star-height question can be formulated as in Theorem 1 below, a counterpart to statement (A). In this theorem reference is made to the sequence $\{f_n\}_n$ of regular expressions over the alphabet $\{a\}$ that are defined inductively by

$$f_1 = a^* , \qquad f_{n+1} = (f_n \cdot a)^* \quad (\text{for all } n \in \mathbb{N} \setminus \{0\}) . \tag{1}$$

Milner had conjectured already in [4] that the minimal star height of f_n is actually n.

Theorem 1 (Hirshfeld and Moller). *For every $n \in \mathbb{N}\backslash\{0\}$, there exists a regular expression f_n over the single-letter alphabet $\{a\}$ such that the minimal star height of f_n under the process interpretation is n. This is witnessed by the sequence $\{f_n\}_n$ that is defined inductively by* (1).

As a consequence, the predicate "minimal star height is less or equal to n", where n is a natural number, defines a proper hierarchy on the set of regular expressions over every non-empty alphabet. This contrasts with the star-height result (A), while not with the result (B), for the language interpretation of regular expressions. Turning to the question of a possible counterpart for statement (C) with respect to the process interpretation, we formulate the following problem.

The Star-Height Problem for the Process Interpretation
Instance: A regular expression e over alphabet A.
Question: What is the minimal star height of e under the process interpretation?

We show that the star-height problem for the process interpretation is solvable, and give an alternative proof for Theorem 1. More details of the proofs sketched in Section 2 and Section 3 can be found in a technical report that is being assembled at `http://www.cs.vu.nl/~clemens/sthps.pdf`.

2 Solution of the Star-Height Problem

The solution of the star-height problem for the process interpretation employs the concept "well-behaved specification" that was introduced in [1], and has been used in [2] for showing that the problem of determining whether a finite-state process represents a star-behaviour is decidable. First, we give a syntactical definition, for every well-behaved specification $\mathcal{E}$, of the "star height" $\mathrm{sh}(\mathcal{E})$ of $\mathcal{E}$, in such a way that two facts can be proved: (D) For every regular expression e a well-behaved recursive specification "in restricted form" (cf. [2]) with the same star height can be constructed that has $P(e)$ as a solution. And conversely, (E) for every well-behaved recursive specification $\mathcal{E}$ a regular expression e with the same star height as $\mathcal{E}$ can be built such that $P(e)$ solves $\mathcal{E}$. With these preparations, we then define an algorithm CSH for computing the minimal star height of a regular expression e that proceeds by the following three steps:

(CSH1) Produce, by a procedure guaranteed by (D) (the one in Section 4 of [2] can be chosen) a well-behaved specification $\mathcal{E}$ in restr. form with $\mathrm{sh}(\mathcal{E}) = \mathrm{sh}(e)$ and with $P(e)$ as a solution.

(CSH2) Determine n, the number of states of the finite-state process $P(e)$, and k, the maximal number of successor states of a state in the process $P(e)$.

(CSH3) Generate successively all of the finite number of well-behaved recursive specifications in restricted form over the action set of the letters occurring in e, with depth less or equal to $(n+1)^3.2^k$, and with less or equal to k summands in each defining equation. For each of these specifications check whether it has $P(e)$ as a solution. By Proposition 14 in [2] it follows that at least one well-behaved specification satisfying all of the mentioned conditions is encountered during the search, namely, the result of simplifying the specification $\mathcal{E}$ by applying the transformation steps described in the proof of Proposition 14 in [2]. Determine the minimum n_0 of the star heights of the well-behaved specifications fulfilling all conditions. Finally, output n_0 as the minimal star height of e under the process interpretation.

It is not difficult to establish the correctness of the algorithm CSH, which justifies Theorem 2 below, as a consequence of the statements (D) and (E), and of the observation that simplification steps for well-behaved specifications explained in the proof of Proposition 14 in [2] do not increase (but may possibly decrease) the star height of a well-behaved specification.

Theorem 2. *The star-height problem for the process interpretation is solvable: there is an algorithm that, for all regular expressions e, on input e computes the minimal star height of e.*

3 Alternative Solution of Milner's Star-Height Question

Our proof of Theorem 1 again employs the relationship mentioned in Section 2 between regular expressions e and well-behaved specifications of the same star height with $P(e)$ as a solution.

It suffices to show that the regular expressions f_n defined in (1) have minimal star height at least n (because f_n has star height n). The process $P(f_n)$ has n non-bisimilar states s_1, $\ldots$, s_n with the property that for $i \leq j$ the process p_i determined by s_i is a subprocess of the process p_j determined by s_j. Using this as well as a careful analysis of the situations in which a subprocess p_i of $P(f_n)$ can be the solution of a sub-specification of a well-behaved specification in restricted form, we obtain our main lemma: (F) For every well-behaved specification $\mathcal{E}$ in restricted form that has $P(f_n)$ as a solution, $\mathrm{sh}(\mathcal{E}) \geq n$ holds. Now let $n \in \mathbb{N} \setminus \{0\}$, and a regular expression e over the alphabet $\{a\}$ be arbitrary such that $e \leftrightarrows_P f_n$. Then by (E) a well-behaved specification $\mathcal{E}$ in restricted form with $\mathrm{sh}(\mathcal{E}) = \mathrm{sh}(e)$ can be found such that $P(e)$, and hence also $P(f_n)$, is a solution of $\mathcal{E}$. By (F), $\mathrm{sh}(\mathcal{E}) \geq n$ follows, which entails $\mathrm{sh}(e) \geq n$. This argument shows that the minimal star height of f_n is at least n, and therefore must be n.

References

[1] J. C. M. Baeten and F. Corradini. Regular expressions in process algebra. In *Proceedings of LICS 2005*, 2005.

[2] J. C. M. Baeten, F. Corradini, and C. A. Grabmayer. A characterisation of regular expressions under bisimulation. Technical report, Technische Universiteit Eindhoven, October 2005. Submitted to a journal. Available at `http://www.cs.vu.nl/~clemens/acoreub.pdf`.

[3] Y. Hirshfeld and F. Moller. On the star height of unary regular behaviours. In *Proof, Language, Interaction: Essays in honour of Robin Milner*, pages 497–509. MIT Press, 2000.

[4] R. Milner. A complete inference system for a class of regular behaviours. *Journal of Computer and System Sciences*, 28:439–466, 1984.