



From generic partition refinement to weighted tree automata minimization

Thorsten Wißmann^{1,2}, Hans-Peter Deifel³, Stefan Milius^{1,3}, and Lutz Schröder^{2,3}

Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

Abstract. Partition refinement is a method for minimizing automata and transition systems of various types. Recently, we have developed a partition refinement algorithm that is generic in the transition type of the given system and matches the run time of the best known algorithms for many concrete types of systems, e.g. deterministic automata as well as ordinary, weighted, and probabilistic (labelled) transition systems. Genericity is achieved by modelling transition types as functors on sets, and systems as coalgebras. In the present work, we refine the run time analysis of our algorithm to cover additional instances, notably weighted automata and, more generally, weighted tree automata. For weights in a cancellative monoid we match, and for non-cancellative monoids such as (the additive monoid of) the tropical semiring even substantially improve, the asymptotic run time of the best known algorithms. We have implemented our algorithm in a generic tool that is easily instantiated to concrete system types by implementing a simple refinement interface. Moreover, the algorithm and the tool are modular, and partition refiners for new types of systems are obtained easily by composing pre-implemented basic functors. Experiments show that even for complex system types, the tool is able to handle systems with millions of transitions.

Keywords: Partition refinement; Markov chains; Lumping; Minimization; Weighted tree automata

1. Introduction

Minimization is a basic verification task on state-based systems, concerned with reducing the number of system states as far as possible while preserving the system behaviour. This can be done by identifying states that exhibit the same behaviour. Hence, it can be used for equivalence checking of systems, and constitutes a preprocessing step in further system analysis tasks, such as model checking.

Notions of equivalent behaviour typically vary quite widely even on fixed system types [vG01]. We work with various notions of *bisimilarity*, i.e. with branching-time equivalences. Classically, bisimilarity for labelled transition systems obeys the principle “states x and y are bisimilar if for every transition $x \rightarrow x'$, there exists a transition $y \rightarrow y'$ with x' and y' bisimilar, and vice versa”. It is thus given via a fixpoint definition, to be understood as a *greatest* fixpoint, and can therefore be iteratively approximated from above.

¹ Supported by the DFG project COAX (MI 717/5-2)

² Supported by the DFG project COAX (SCHR 1118/12-2)

³ Supported by the Deutsche Forschungsgemeinschaft (DFG) as part of the Research and Training Group 2475 “Cybercrime and Forensic Computing” (393541319/GRK2475/1-2019)

This is the principle behind *partition refinement* algorithms: Initially all states are tentatively considered equivalent, and then this initial partition is iteratively refined according to observations made on the states until a fixpoint is reached. Unsurprisingly, such procedures run in polynomial time. Its comparative tractability (in contrast, e.g. trace equivalence and language equivalence of non-deterministic systems are PSPACE-complete [KS90]) makes minimization under bisimilarity interesting even in cases where the main equivalence of interest is linear-time, such as word automata.

Kanellakis and Smolka [KS90] in fact provide a minimization algorithm with run time $\mathcal{O}(m \cdot n)$ for ordinary transition systems with n states and m transitions. However, even faster partition refinement algorithms running in $\mathcal{O}((m + n) \cdot \log n)$ have been developed for various types of systems over the past 50 years. For example, Hopcroft’s algorithm minimizes deterministic automata for a fixed input alphabet A in $\mathcal{O}(n \cdot \log n)$ [Hop71]; it was later generalized to variable input alphabets, with run time $\mathcal{O}(n \cdot |A| \cdot \log n)$ [Gri73, Knu01]. The Paige-Tarjan algorithm minimizes transition systems in time $\mathcal{O}((m + n) \cdot \log n)$ [PT87], and generalizations to labelled transition systems have the same time complexity [HT92, DHS03, Val09]. Minimization of weighted systems is typically called *lumping* in the literature; Valmari and Franchescinis [VF10] exhibit a simple $\mathcal{O}((m + n) \cdot \log n)$ lumping algorithm for systems with rational weights.

In earlier work [DMSW17, WDMS20] we have developed an efficient *generic* partition refinement algorithm that can be easily instantiated to a wide range of system types, most of the time either matching or improving the previous best run time. The genericity of the algorithm is based on modelling state-based systems as coalgebras following the paradigm of universal coalgebra [Rut00], in which the branching structure of systems is encapsulated in the choice of a functor, the *type functor*. This allows us to cover not only classical relational systems and various forms of weighted systems, but also to combine existing system types in various ways, e.g. nondeterministic and probabilistic branching. Our algorithm uses a functor-specific *refinement interface* that supports a graph-based representation of coalgebras. It allows for a generic complexity analysis, and indeed the generic algorithm has the same asymptotic complexity as the above-mentioned specific algorithms. For Segala systems [Seg95] (systems that combine probabilistic and non-deterministic branching, also known as Markov decision processes), it matches the run time of a recent algorithm [GVdV18] discovered independently and almost at the same time as ours, and improves on the run time of the previously best algorithm [BEM00].

The new contributions of the present paper are twofold. On the theoretical side, we show how to instantiate our generic algorithm to weighted systems with weights in a monoid (generalizing the group-weighted case considered previously [DMSW17, WDMS20]). We then refine the complexity analysis of the algorithm, making the complexity of the implementation of the type functor a parameter $p(c)$, where c is the input coalgebra. In the new setup, the previous analysis becomes the special case where $p(c) = 1$. Under the same structural assumptions on the type functor and the refinement interface as previously, our algorithm runs in time $\mathcal{O}(m \cdot \log n \cdot p(c))$ for an input coalgebra c with n states and m transitions. Instantiated to the case of weighted systems over non-cancellative monoids (with $p(c) = \log(m)$ where m is the number of transitions in c) the run time of the generic algorithm is $\mathcal{O}(m \cdot \log^2 m)$, thus markedly improving the run time $\mathcal{O}(m \cdot n)$ of previous algorithms for weighted automata [Buc08] and, more generally, (bottom-up) weighted tree automata [HBMM07]. This includes weighted tree automata for the additive monoid $(\mathbb{N}, \max, 0)$ of the tropical semiring, which are used in natural language processing [MK06]. In addition, for cancellative monoids, we again essentially match the complexity of the previous algorithms [Buc08, HBMM07].

Our second main contribution is a generic and modular implementation of our algorithm, the *Coalgebraic Partition Refiner* (CoPaR). Instantiating CoPaR to coalgebras for a given functor requires only to implement the refinement interface. We provide such implementations for a number of basic type functors, e.g. for non-deterministic, weighted, or probabilistic branching, as well as (ranked) input and output alphabets or output weights. In addition, CoPaR is *modular*: For any type functor obtained by composing basic type functors for which a refinement interface is available, CoPaR automatically derives an implementation of the refinement interface. We explain in detail how this modularity is realized in our implementation and, extending Valmari and Franchescinis’s ideas [VF10], we explain how the necessary data structures need to be implemented so as to realize the low theoretical complexity. We thus provide a working efficient partition refiner for all the above mentioned system types. In particular, our tool is, to the best of our knowledge, the only available implementation of partition refinement for many composite system types, notably for weighted (tree) automata over non-cancellative monoids. The tool including source code and evaluation data is available at <https://git8.cs.fau.de/software/copar>.

The present paper is an extended and completely reworked version of a previous conference paper [DMSW19]. It includes full proofs, additional benchmarks, and more extensive examples and explanations. Moreover, we formally show how refinement interfaces can be combined along products of functors (Proposition 3.14 and Sect. 3.5). We have optimized the memory consumption of our implementation which has led to better performance in the benchmarks on weighted tree automata (Table 2).

Organization. The material is structured as follows. In Sect. 2 we recall the necessary technical background and the modelling of state based systems as coalgebras. In Sect. 3, we describe the tool and the underlying algorithm, discussing in particular tool usage and implementation, the generic interface, and the modularity principles that we employ. Some concrete instantiations are exhibited in Sect. 4. We then go on to elaborate the case of weighted systems in more detail, giving a refinement interface for the basic underlying functor of such systems in Sect. 5, and showing in Sect. 6 how to cover weighted tree automata—which arise by combining weighted systems and ranked alphabets—by means of our modularity principles. Benchmarks are presented in Sect. 7.

2. Preliminaries: universal coalgebra

Our algorithmic framework [WDMS20] is based on modelling state-based systems abstractly as *coalgebras* for a (set) *functor* that encapsulates the transition type, following the paradigm of *universal coalgebra* [Rut00]. We proceed to recall standard notation for sets and maps, as well as basic notions and examples in coalgebra. Occasional comments assume familiarity with basic notions of category theory (e.g. [Awo10]) but the few concepts needed for the main development are explained in full. We fix a singleton set $1 = \{*\}$; for every set X we have a unique map $! : X \rightarrow 1$. We denote composition of maps by $(-)\cdot(-)$, in applicative order. We denote the disjoint union—in categorical terms, the *coproduct*—of sets A, B by $A + B$ where we write $\text{inl} : A \rightarrow A + B$ and $\text{inr} : B \rightarrow A + B$ for the canonical injections; the disjoint union, or coproduct, of a family $(X_j)_{j \in J}$ of sets is denoted by $\coprod_{j \in J} X_j$. Similarly, we write $\prod_{j \in J} X_j$ for the (cartesian) product of a family of sets. Injection maps of disjoint unions and projection maps of products, respectively, are denoted by

$$\text{in}_i : X_i \rightarrow \coprod_{j \in J} X_j \quad \text{and} \quad \text{pr}_i : \prod_{j \in J} X_j \rightarrow X_i.$$

Given two maps $f : A \rightarrow X$ and $g : A \rightarrow Y$ we write $\langle f, g \rangle : A \rightarrow X \times Y$ for the map $a \mapsto (f(a), g(a))$. Similarly, for a family of maps $(f_i : A \rightarrow X_i)_{i \in I}$, we write $\langle f_i \rangle_{i \in I} : A \rightarrow \prod_{i \in I} X_i$ for the map $a \mapsto (f_i(a))_{i \in I}$.

We model the transition type of state based systems using *functors*. Informally, a functor F assigns to a set X a set $F X$, whose elements are thought of as structured collections over X , and an F -coalgebra is a map $c : X \rightarrow F X$ assigning to each state x in a system a structured collection $c(x) \in F X$ of successors. The most basic example is that of transition systems, where F is powerset, so a coalgebra assigns to each state a set of successors. Formal definitions are as follows.

- Definition 2.1**
1. A *functor* $F : \text{Set} \rightarrow \text{Set}$ assigns to each set X a set $F X$, and to each map $f : X \rightarrow Y$ a map $F f : F X \rightarrow F Y$, preserving identities and composition ($F \text{id}_X = \text{id}_{F X}$, $F(g \cdot f) = F g \cdot F f$).
 2. An F -*coalgebra* (C, c) consists of a set C of *states* and a *transition* structure $c : C \rightarrow F C$.
 3. A *morphism* $h : (C, c) \rightarrow (D, d)$ of F -coalgebras is a map $h : C \rightarrow D$ that preserves the transition structure, i.e. $F h \cdot c = d \cdot h$.
 4. Two states $x, y \in C$ of a coalgebra $c : C \rightarrow F C$ are *behaviourally equivalent* (notation: $x \sim y$) if there exists a coalgebra morphism h such that $h(x) = h(y)$.

As above, we usually use the letters X and Y for sets (without structure) and C or D for state sets of coalgebras.

Example 2.2 (1) The *finite powerset* functor \mathcal{P}_ω maps a set X to the set $\mathcal{P}_\omega X$ of all *finite* subsets of X , and a map $f : X \rightarrow Y$ to the map $\mathcal{P}_\omega f = f[-] : \mathcal{P}_\omega X \rightarrow \mathcal{P}_\omega Y$ taking direct images. \mathcal{P}_ω -coalgebras are finitely branching (unlabelled) transition systems and two states are behaviourally equivalent iff they are bisimilar in the sense of Milner [Mil80] and Park [Par81].

- (2) A signature Σ is a set Σ of *operation symbols* together with a map $\text{ar}: \Sigma \rightarrow \mathbb{N}$, which assigns to each operation symbol $\sigma \in \Sigma$ its *arity* $\text{ar}(\sigma)$. We write $\sigma/n \in \Sigma$ for $\sigma \in \Sigma$ with $\text{ar}(\sigma) = n$. Every signature Σ canonically defines a *polynomial functor*

$$F_{\Sigma}X = \coprod_{\sigma/n \in \Sigma} X^n.$$

We slightly abuse notation by denoting for each $\sigma/n \in \Sigma$ the corresponding injection into the coproduct by

$$\sigma: X^n \rightarrow F_{\Sigma}X.$$

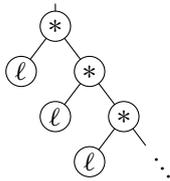
Moreover, we simply write Σ in lieu of F_{Σ} , so we have

$$\Sigma X = F_{\Sigma}X = \{\sigma(x_1, \dots, x_n) \mid \sigma/n \in \Sigma, x_1, \dots, x_n \in X\};$$

for $\sigma(x_1, \dots, x_n)$, we sometimes write $\sigma/n(x_1, \dots, x_n)$ to emphasize the arity or disambiguate overloaded symbols. This polynomial functor acts component-wise on maps $f: X \rightarrow Y$:

$$\Sigma f: \Sigma X \rightarrow \Sigma Y \quad (\Sigma f)(\sigma(x_1, \dots, x_n)) = \sigma(f(x_1), \dots, f(x_n)) \quad \text{for } \sigma/n \in \Sigma.$$

Every state in a Σ -coalgebra represents a (possibly infinite) Σ -tree, i.e. a rooted ordered tree where every node is labelled with some operation symbol $\sigma \in \Sigma$ and has precisely $\text{ar}(\sigma)$ -many children. In particular, a node is a leaf iff it is labelled with a 0-ary operation symbol. For example, for the signature $\Sigma = \{*/2, \ell/0\}$ with a binary operation symbol and a constant, we have the following example of a Σ -tree:



Given a state x in a coalgebra $c: C \rightarrow \Sigma C$, we obtain a Σ -tree t_x by unravelling the coalgebra structure at x . More precisely, t_x is uniquely defined by

$$t_x = \begin{array}{c} \sigma \\ \swarrow \quad \searrow \\ t_{x_1} \quad \dots \quad t_{x_n} \end{array} \quad \text{if } c(x) = \sigma(x_1, \dots, x_n).$$

(this equation constituting a *coinductive* definition [Rut00]). For example, the above Σ -tree is obtained by unravelling the coalgebra structure at the state x of the Σ -coalgebra

$$c: \{x, y\} \rightarrow \Sigma\{x, y\} \quad \text{with} \quad c(x) = *(y, x) \text{ and } c(y) = \ell.$$

Two states in a Σ -coalgebra are behaviourally equivalent iff they represent the same possibly infinite tree: $x \sim y$ iff $t_x = t_y$.

- (3) For a fixed finite set A , the functor given by $FX = 2 \times X^A$, where $2 = \{0, 1\}$, sends a set X to the set of pairs of boolean values and functions $A \rightarrow X$. An F -coalgebra (C, c) is a deterministic automaton (without initial state). For each state $x \in C$, the first component of $c(x)$ determines whether x is a final state, and the second component is the successor function $A \rightarrow X$ mapping each input letter $a \in A$ to the successor state of x under input letter a . States $x, y \in C$ are behaviourally equivalent iff they accept the same language in the usual sense.

This functor is (naturally isomorphic to) the polynomial functor for the signature Σ consisting of two operation symbols of arity $|A|$: $2 \times X^A \cong X^{|A|} + X^{|A|}$.

- (4) For a commutative monoid $(M, +, 0)$, the *monoid-valued* functor $M^{(-)}$ sends each set X to the set of *finitely supported* maps $f: X \rightarrow M$, i.e. $f(x) = 0$ for all but finitely many $x \in X$. In case M is even an abelian group, we also refer to $M^{(-)}$ as a *group-valued* functor.

An F -coalgebra $c: C \rightarrow M^{(C)}$ is, equivalently, a finitely branching M -weighted transition system: For a state $x \in C$, $c(x)$ maps each state $y \in C$ to the weight $c(x)(y)$ of the transition from x to y . For a map $f: X \rightarrow Y$, $M^{(f)}: M^{(X)} \rightarrow M^{(Y)}$ sends a finitely supported map $v: X \rightarrow M$ to the map $y \mapsto \sum_{x \in X, f(x)=y} v(x)$, corresponding to the standard image measure construction. As the notion of behavioural equivalence of states in $M^{(-)}$ -coalgebras, we obtain weighted bisimilarity (cf. [Buc08, KS13]), given coinductively by postulating that states $x, y \in C$ are behaviourally equivalent ($x \sim y$) iff

$$\sum_{z' \sim z} c(x)(z') = \sum_{z' \sim z} c(y)(z') \quad \text{for all } z \in C.$$

For the Boolean monoid $(2 = \{0, 1\}, \vee, 0)$, the monoid-valued functor $2^{(-)}$ is (naturally isomorphic to) the finite powerset functor \mathcal{P}_ω . For the monoid of real numbers $(\mathbb{R}, +, 0)$, the monoid-valued functor $\mathbb{R}^{(-)}$ has \mathbb{R} -weighted systems as coalgebras, e.g. Markov chains. In fact, finite Markov chains are precisely finite coalgebras of the *finite distribution functor*, i.e. the subfunctor \mathcal{D}_ω of $\mathbb{R}_{\geq 0}^{(-)}$ (and hence of $\mathbb{R}^{(-)}$) given by

$$\mathcal{D}_\omega(X) = \{\mu \in \mathbb{R}_{\geq 0}^{(X)} \mid \sum_{x \in X} \mu(x) = 1\}.$$

For the monoid $(\mathbb{N}, +, 0)$ of natural numbers, the monoid-valued functor is the bag functor \mathcal{B}_ω , which maps a set X to the set of finite multisets over X .

Notation 2.3 Note that for every commutative monoid $(M, +, 0)$, we have the canonical summation map

$$\Sigma: \mathcal{B}_\omega M \rightarrow M \quad \text{with} \quad \Sigma f = \sum_{x \in M} \underbrace{x + \dots + x}_{f(x)\text{-many}} \quad (2.1)$$

It sums up all elements of a bag f of monoid elements, where a single element of the monoid can occur multiple times.

Remark 2.4 For categorically-minded readers, we note that \mathcal{B}_ω is a monad on the category of sets. Moreover, commutative monoids are precisely the Eilenberg-Moore algebras (e.g. [Awo10]) for \mathcal{B}_ω . In fact, for every commutative monoid $(M, +, 0)$, the map Σ is the structure of its associated Eilenberg-Moore algebra.

3. Generic partition refinement

We recall some key aspects of our generic partition refinement algorithm [WDMS20], which *minimizes* a given coalgebra, i.e. computes its quotient modulo behavioural equivalence. We centre the presentation around the implementation and use of our tool.

The algorithm [WDMS20, Algorithm 4.5] is parametrized over a type functor F , represented by implementing a fixed *refinement interface*, which in particular allows for a representation of F -coalgebras in terms of nodes and edges (by no means implying a restriction to relational systems!). Our previous analysis has established that the algorithm minimizes an F -coalgebra $c: C \rightarrow FC$ with n nodes and m edges in time $\mathcal{O}(m \cdot \log n)$, assuming $m \geq n$ and that the operations of the refinement interface run in linear time. In the present paper, we generalize the analysis, establishing a run time in $\mathcal{O}(m \cdot \log n \cdot p(c))$, where $p(c)$ is a factor in the time complexity of the operations implementing the refinement interface which depends on the input coalgebra $c: C \rightarrow FC$. For many functors, $p(c) = 1$, reproducing the previous analysis. In some cases, $p(c)$ is not constant, and our new analysis still applies in these cases, either matching or improving the best known run time complexity in most instances, most notably weighted systems over non-cancellative monoids.

We proceed to discuss the design of the implementation, including input formats of our tool CoPaR for composite functors built from pre-implemented basic blocks and for systems to be minimized (Sect. 3.1). We then discuss the internal representation of coalgebras in the tool (Sect. 3.2). Subsequently, we recall refinement interfaces, describe their implementation (Sect. 3.4), and discuss how to combine them (Sect. 3.5). Finally, we note implementation details of our tool and, in particular, argue that it realizes the theoretical time complexity (Sect. 3.6).

3.1. Generic system specification

CoPaR accepts as input a file that represents a finite F -coalgebra $c: C \rightarrow FC$, and consists of two parts. The first part is a single line specifying the functor F . Each of the remaining lines describes one state $x \in C$ and its one-step behaviour $c(x)$. Examples of input files are shown in Fig. 1.

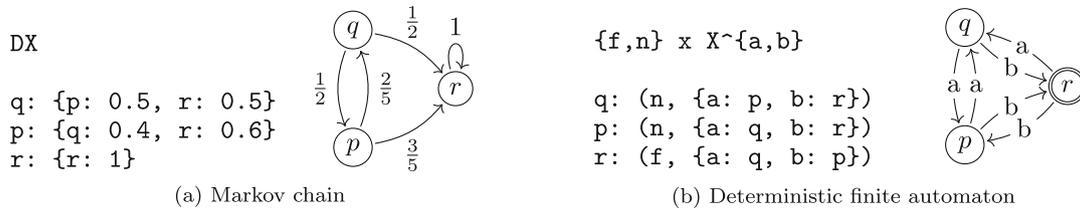


Fig. 1. Examples of input files with encoded coalgebras

3.1.1. Functor specification

Functors are specified as composites of basic building blocks; that is, the functor given in the first line of an input file is an expression determined by the grammar

$$T ::= X \mid F(T, \dots, T) \quad (F: \text{Set}^k \rightarrow \text{Set}) \in \mathcal{F}, \quad (3.1)$$

where the character X is a terminal symbol and \mathcal{F} is a set of predefined symbols called *basic functors*, representing a number of pre-implemented functors of type $F: \text{Set}^k \rightarrow \text{Set}$. Only for basic functors, a *refinement interface* needs to be implemented (Sect. 3.4); for composite functors, the tool derives an appropriate refinement interface automatically (Sect. 3.5). Basic functors currently implemented include the (finite) powerset functor \mathcal{P}_ω , the bag functor \mathcal{B}_ω , monoid-valued functors $M^{(-)}$, and polynomial functors for finite many-sorted signatures Σ , based on the description of the respective refinement interfaces given in our previous work [WDMS20] and, in the case of $M^{(-)}$ for unrestricted commutative monoids M (rather than only abelian groups), the newly developed interface described in Sect. 5.2. Since behavioural equivalence is preserved and reflected under converting G -coalgebras into F -coalgebras for a subfunctor G of F [WDMS20, Proposition 2.13], we also cover subfunctors, such as the finite distribution functor \mathcal{D}_ω as a subfunctor of $\mathbb{R}^{(-)}$. With the polynomial constructs $+$ and \times written in infix notation as usual, the currently supported grammar is effectively

$$\begin{aligned} T &::= X \mid \mathcal{P}_\omega T \mid \mathcal{B}_\omega T \mid \mathcal{D}_\omega T \mid M^{(T)} \mid \Sigma \\ \Sigma &::= C \mid T + T \mid T \times T \mid T^A \quad C ::= \mathbb{N} \mid A \quad A ::= \{s_1, \dots, s_n\} \mid n \end{aligned} \quad (3.2)$$

where $n \in \mathbb{N}$ denotes the set $\{0, \dots, n-1\}$, the s_k are strings subject to the usual conventions for variable names¹, exponents F^A are written $F^{\wedge}A$, and M is one of the monoids $(\mathbb{Z}, +, 0)$, $(\mathbb{R}, +, 0)$, $(\mathbb{C}, +, 0)$, $(\mathcal{P}_\omega(64), \cup, \emptyset)$ (i.e. the monoid of 64-bit words with bitwise or), and $(\mathbb{N}, \max, 0)$ (the additive monoid of the tropical semiring). Note that C effectively ranges over at most countable sets, and A over finite sets. A term T determines a functor $F: \text{Set} \rightarrow \text{Set}$ in the evident way, with X interpreted as the argument, i.e. $F(X) = T$. It should be noted that the implementation treats composites of polynomial (sub-)terms as a single functor in order to minimize overhead incurred by excessive decomposition, e.g. $X \mapsto \{a, b\} + \mathcal{P}_\omega(\mathbb{R}^{(X)}) + X \times X$ is composed of the basic functors \mathcal{P}_ω , $\mathbb{R}^{(-)}$ and the 3-sorted polynomial functor $\Sigma(X, Y, Z) = \{a, b\} + X + Y \times Z$.

3.1.2. Coalgebra specification

The remaining lines of an input file define a finite F -coalgebra $c: C \rightarrow FC$. Each line of the form $x: _t$ defines a state $x \in C$, where x is a variable name, and t represents the element $t = c(x) \in FC$. The syntax for t depends on the specified functor F , and follows the structure of the term T defining F ; we write $t \in T$ for a term t describing an element of FC :

- $t \in X$ is given by one of the named states specified in the file.
- $t \in T_1 \times \dots \times T_n$ is given by $t ::= (t_1, \dots, t_n)$ where $t_i \in T_i$, $i = 1, \dots, n$.
- $t \in T_1 + \dots + T_n$ is given by $t ::= \text{inj}_{_i} t_i$ where $i = 1, \dots, n$ and $t_i \in T_i$.
- $t \in \mathcal{P}_\omega T$ and $t \in \mathcal{B}_\omega T$ are given by $t ::= \{t_1, \dots, t_n\}$ with $t_1, \dots, t_n \in T$.
- $t \in M^{(T)}$ is given by $t ::= \{t_1: _m_1, \dots, t_n: _m_n\}$ with $m_1, \dots, m_n \in M$ and $t_1, \dots, t_n \in T$, denoting $\mu \in M^{(TC)}$ with $\mu(t_i) = m_i$ for $i = 1, \dots, n$, and $\mu(t) = 0$ for $t \notin \{t_1, \dots, t_n\}$.

¹a letter or an underscore character followed by alphanumeric characters or underscore

For example, the two-line declaration

$$\begin{aligned} &P(\{a, b\} \times \mathbb{R}^X) \\ x: &\{(a, \{x: 2.4\}), (a, \{\}), (b, \{x: -8\})\} \end{aligned}$$

defines an F -coalgebra for the functor $FX = \mathcal{P}_\omega(\{a, b\} \times \mathbb{R}^X)$, with a single state x , having two a -successors and one b -successor, where successors are elements of \mathbb{R}^X . One a -successor is constantly zero, and the other assigns weight 2.4 to x ; the b -successor assigns weight -8 to x . Two more examples are shown in Fig. 1.

3.1.3. Generic input file processing

After reading the functor term T , the tool builds a parser for the functor-specific input format and parses an input coalgebra specified in the above syntax into an intermediate format described in the next section. In the case of a composite functor, the parsed coalgebra then undergoes a substantial amount of preprocessing that also affects how transitions are counted; we defer the discussion of this point to Sect. 3.5, and assume for the time being that $F: \text{Set} \rightarrow \text{Set}$ is a basic functor with only one argument.

3.2. Internal representation of coalgebras

New functors are added to the framework by implementing a *refinement interface* (Definition 3.5). The interface relates to an abstract encoding of the functor and its coalgebras in terms of nodes and edges:

Definition 3.1 [WDMS20] An *encoding* of a functor F consists of a set A of *labels* and a family of maps

$$b: FX \rightarrow \mathcal{B}_\omega(A \times X),$$

one for every set X . The *encoding* of an F -coalgebra $c: C \rightarrow FC$ is given by the map

$$\langle F!, b \rangle \cdot c: C \rightarrow F1 \times \mathcal{B}_\omega(A \times C),$$

and we say that the coalgebra has $n = |C|$ states and $m = \sum_{x \in C} |b(c(x))|$ edges.

An encoding does by no means imply a reduction from F -coalgebras to $\mathcal{B}_\omega(A \times (-))$ -coalgebras, i.e. the notions of behavioural equivalence for $\mathcal{B}_\omega(A \times (-))$ and F , respectively, can be radically different. The encoding just fixes a representation format.

Remark 3.2 Categorically-minded readers will notice that b is not assumed to be a natural transformation. In fact, b fails to be natural in all encodings we have implemented except the one for polynomial functors.

Encodings typically match how one intuitively draws coalgebras of various types as certain labelled graphs. We briefly recall three examples below; see [WDMS20] for more. We note that so far, we see no general method for deriving an encoding of a functor, which therefore requires invention.

Example 3.3 (1) We have mentioned in Example 2.2(1) that finitely branching transition systems are the coalgebras for $F = \mathcal{P}_\omega$. For the encoding we choose the singleton set $A = 1$ of labels, and $b: \mathcal{P}_\omega X \rightarrow \mathcal{B}_\omega(1 \times X) \cong \mathcal{B}_\omega X$ is the obvious inclusion, i.e. $b(S)(x) = 1$ if $x \in S$ and $b(S)(x) = 0$ otherwise, for $S \in \mathcal{P}_\omega X$.

(2) For a monoid-valued functor $F = M^{(-)}$ (see Example 2.2(4)) we take $A = M_{\neq 0}$, the non-zero elements of M , and define $b: M^{(X)} \rightarrow \mathcal{B}_\omega(M_{\neq 0} \times X)$ by taking $b(f)$ to be the finite set $\{(f(x), x) \mid x \in X, f(x) \neq 0\}$, interpreted as a bag.

Special cases are the group-valued functors $G^{(-)}$ for an abelian group G , in particular $\mathbb{R}^{(-)}$ and its subfunctor \mathcal{D}_ω , whose coalgebras are Markov chains (cf. Fig. 1). In the last case, we formally inherit $A = \mathbb{R}_{\neq 0}$ from the encoding of $\mathbb{R}^{(-)}$ but can actually restrict to $A = (0, 1]$.

(3) For a polynomial functor $F = \Sigma$, the set of labels is $A = \mathbb{N}$, and the map $b: \Sigma X \rightarrow \mathcal{B}_\omega(\mathbb{N} \times X)$ is given by

$$b(\sigma(x_1, \dots, x_n)) = \{(1, x_1), \dots, (n, x_n)\}.$$

The implementation of a basic functor then consists of two ingredients: (1) a parser that transforms the syntactic specification of an input coalgebra (Sect. 3.1) into the encoded coalgebra in the above sense, and (2) an implementation of the refinement interface, which is motivated next.

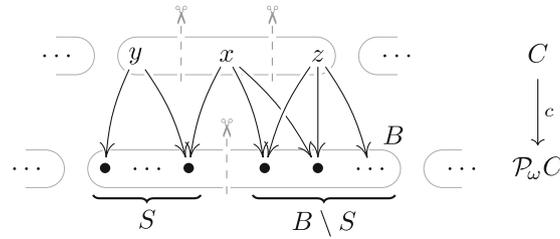


Fig. 2. Splitting of a block B into smaller blocks results in further refinement of the block $\{x, y, z\}$

3.3. Splitting blocks by $F3$

In order to understand the requirements on an interface encapsulating the functor specific parts of partition refinement, let us look at one step of the algorithm which is crucial for the overall run time complexity. Partition refinement algorithms in general maintain a *partition* of the state space, i.e. a disjoint decomposition of the state space into sets called *blocks*, adhering to the invariant that states in different blocks are behaviourally inequivalent, and ensuring upon termination that states in the same block are behaviourally equivalent. Initially, the algorithm tentatively identifies all states of a coalgebra $c: C \rightarrow FC$ in a partition consisting of only one block, C . Then, the algorithm splits this block into smaller blocks whenever states of the coalgebra turn out to be behaviourally inequivalent and successively applies this procedure to the new blocks until no further splitting is necessary. In the first iteration, the algorithm separates states $x, y \in C$ if they are distinguished by $F! \cdot c: C \rightarrow F1$, i.e. if $F!(c(x)) \neq F!(c(y))$. For example for automata, i.e. for $FX = 2 \times X^A$, we have $F1 = 2 \times 1^A \cong 2$, so this first step separates final from non-final states. In the classical Paige-Tarjan algorithm [PT87], i.e. for $FX = \mathcal{P}_\omega X$, deadlock states and states with at least one outgoing transition are separated from each other. In the subsequent steps, the representation of the coalgebra as labelled edges (i.e. $b \cdot c: C \rightarrow \mathcal{B}_\omega(A \times C)$) is used to refine the partition further. Information about the inequivalence of states is propagated from successor states to predecessor states; this is iterated until a (greatest) fixed point is reached, i.e. until no new behavioural inequivalences are discovered.

In this process of propagating inequivalences, suppose that the partition refinement has already computed a block of states $B \subseteq C$ in its partition and that states in $S \subseteq B$ have different behaviour from those in $B \setminus S$ (as illustrated in Fig. 2). From this information, the algorithm infers whether states $x, y \in C$ that are in the same block and have successors in B exhibit different behaviour and thus have to be separated. Let us explain on two concrete instances how this inference is achieved.

Example 3.4 (1) We mentioned in Example 2.2(1) that for $F = \mathcal{P}_\omega$, a coalgebra is a finitely branching transition system. A partition on the state space represents a bisimulation if it has the following property:

For every pair of states x, y in the same block and every block B :
 x has a successor in B iff y has a successor in B .

This means that when we split the block B into the two blocks S and $B \setminus S$, then x and y can stay in the same block provided that (a) x has a successor in S iff y has one and (b) x has a successor in $B \setminus S$ iff y has one. Equivalently, we can express these conditions by the equality

$$\mathcal{P}_\omega \langle \chi_S, \chi_{B \setminus S} \rangle (c(x)) = \mathcal{P}_\omega \langle \chi_S, \chi_{B \setminus S} \rangle (c(y)), \quad (3.3)$$

where $\chi_S, \chi_{B \setminus S}: C \rightarrow 2$ are the usual characteristic functions of the subsets $S, B \setminus S \subseteq C$, respectively. Indeed, the function

$$C \xrightarrow{c} \mathcal{P}_\omega C \xrightarrow{\mathcal{P}_\omega \langle \chi_S, \chi_{B \setminus S} \rangle} \mathcal{P}_\omega (2 \times 2)$$

maps a state $x \in C$ to the set

$$P := \mathcal{P}_\omega \langle \chi_S, \chi_{B \setminus S} \rangle (c(x)) \subseteq 2 \times 2$$

encoding whether x has successors in S resp. in $B \setminus S$. In fact, we see that x has a successor in S iff $(1, 0) \in P$ and x has a successor in $B \setminus S$ iff $(0, 1) \in P$. Moreover, x has a successor in $C \setminus B$ iff $(0, 0) \in P$. Since S and

$B \setminus S$ are disjoint, we have $(1, 1) \notin P$. Similar observations apply to y . Thus, in order to maintain the desired property, we need to separate x and y iff (3.3) holds.²

- (2) In the example of Markov chains, i.e. $F = \mathcal{D}_\omega$, we can make a similar observation. Here, $x, y \in C$ can stay in the same block if the weights of all transitions from x to states in S sum up to the same value as the weights of all transitions from y to states in S and similarly for $B \setminus S$; that is if

$$\sum_{x' \in S} c(x)(x') = \sum_{y' \in S} c(y)(y') \quad \text{and} \quad \sum_{x' \in B \setminus S} c(x)(x') = \sum_{y' \in B \setminus S} c(y)(y').$$

Analogously as in item (1), we can equivalently express this by stating that x and y can stay in the same block iff the following equation holds

$$\mathcal{D}_\omega \langle \chi_S, \chi_{B \setminus S} \rangle (c(x)) = \mathcal{D}_\omega \langle \chi_S, \chi_{B \setminus S} \rangle (c(y)). \quad (3.4)$$

The map

$$C \xrightarrow{c} FC \xrightarrow{\mathcal{D}_\omega \langle \chi_S, \chi_{B \setminus S} \rangle} \mathcal{D}_\omega(2 \times 2) \subseteq \mathbb{R}^{(2 \times 2)}$$

sends every state $x \in C$ to the function $t = \mathcal{D}_\omega \langle \chi_S, \chi_{B \setminus S} \rangle (c(x)): 2 \times 2 \rightarrow \mathbb{R}$ where

- $t(1, 0)$ is the accumulated weight of all transitions from x to states in S ,
- $t(0, 1)$ is the accumulated weight of all transitions from x to $B \setminus S$,
- $t(0, 0)$ is the accumulated weight of all transitions from x to states outside of B ,
- $t(1, 1) = 0$, since there is no element in $S \cap (B \setminus S)$.

Two states x, y that are in the same block before splitting B into S and $B \setminus S$ have the same accumulated weight of transitions to B and also to $C \setminus B$, so x and y can stay in the same block iff (3.4) holds.

It is now immediate how (3.3) and (3.4) are generalized to an arbitrary functor F : States x and y stay in the same block in a refinement step iff

$$F \langle \chi_S, \chi_{B \setminus S} \rangle (c(x)) = F \langle \chi_S, \chi_{B \setminus S} \rangle (c(y)). \quad (3.5)$$

As we have seen in Example 3.4, $(1, 1)$ is never in the image of $\langle \chi_S, \chi_{B \setminus S} \rangle: C \rightarrow 2 \times 2$, because S and $B \setminus S$ are disjoint. Hence we can restrict its codomain to $3 = \{0, 1, 2\}$ by defining the map

$$\chi_S^B: C \rightarrow 3 \quad \text{for } S \subseteq B \quad \text{by} \quad \chi_S^B(x) = \begin{cases} 2 & \text{if } x \in S, \\ 1 & \text{if } x \in B \setminus S, \\ 0 & \text{if } x \in C \setminus B. \end{cases} \quad (3.6)$$

Hence, the criterion in (3.5) can be simplified: the states x and y stay in the same block in the refinement step iff

$$F \chi_S^B(c(x)) = F \chi_S^B(c(y)). \quad (3.7)$$

We conclude that the generic partition refinement algorithm needs to compute the value $F \chi_S^B(c(x)) \in F3$ for every state x . Whenever states x and y are sent to different values by $F \chi_S^B \cdot c$, we know that they are behaviourally inequivalent and need to be moved to separate blocks.

3.4. Refinement interfaces

Computing the values $F \chi_S^B(c(x))$ for states x of interest is the task of the refinement interface. We start with its formal definition and then provide an informal explanation of its ingredients.

Definition 3.5 [WDM20] Given an encoding (A, b) of the set functor F , a *refinement interface* for F consists of a set W of *weights* and functions

$$\text{init}: F1 \times \mathcal{B}_\omega A \rightarrow W \quad \text{and} \quad \text{update}: \mathcal{B}_\omega A \times W \rightarrow W \times F3 \times W$$

²Note that since the above property holds before splitting B , we have that $(0, 0)$ is either contained in both sides of (3.3) or in neither of them.

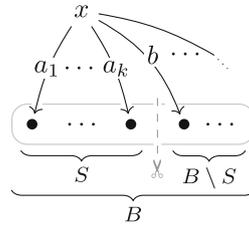


Fig. 3. Computation of the value of type $F3$ for x

satisfying the coherence condition that there exists a family of *weight maps* $w: \mathcal{P}X \rightarrow (FX \rightarrow W)$ (not themselves part of the interface to be implemented!), one for each set X , such that

$$\text{init}(F!(t), \llbracket a \mid (a, x) \in b(t) \rrbracket) = w(X)(t)$$

$$\text{update}(\llbracket a \mid (a, x) \in b(t), x \in S \rrbracket, w(B)(t)) = (w(S)(t), F\chi_S^B(t), w(B \setminus S)(t))$$

for $t \in FX$ and $S \subseteq B \subseteq X$. Here, the notation $\llbracket a \mid \dots \rrbracket$ in the arguments of *init* and *update* indicates *multiset comprehension*, i.e. multiple occurrences of a label $a \in A$ in $b(t)$ result in multiple occurrences of a in the bag $\llbracket a \mid (a, x) \in b(t), \dots \rrbracket$.

The intuition behind the refinement interface can be understood best if we consider an F -coalgebra $c: C \rightarrow FC$, put $X := C$, fix a state $x \in C$, and instantiate $t := c(x)$. As one can see from the types of *init* and *update*, the refinement interface is designed in such a way that it computes values of the functor specific type W of weights that the calling algorithm saves for subsequent calls to *update*. For every block $B \subseteq C$, the value $w(B)(c(x)) \in W$ is the accumulated weight of edges from x to (states in) B in the coalgebra (C, c) . In principle, values in W can contain whatever information about the set of edges from x to B helps the implementation of *update* to compute the result value of type $F3$, which is what the caller is actually interested in. However, while more information contained in the second argument of *update* helps this function to achieve this task more efficiently, both *init* and *update* also have to compute values of W , which may require more effort if these values carry too much information.

This trade-off is guided by the two equational axioms for *init* and *update*, which represent a contract that their implementation for a particular functor has to fulfil. The first axiom assumes that *init* receives in its first argument the output behaviour of x —e.g. whether x is final or non-final (in the case of automata), or for $F = \mathcal{P}_\omega$ whether x has any successors or is a deadlock state—and in its second argument the bag of labels of all outgoing edges of x in the graph representation of (C, c) . The axiom then requires *init* to return the accumulated weight $w(C)(c(x)) \in W$ of all edges from x to the whole state set C , which is the only block in the initial partition of C . This corresponds to the use of *init* in the actual algorithm, namely to initialize the weight value in W that is later passed to *update*.

The operation *update* is called whenever the algorithm derives that a block B in the partition of C contains behaviourally inequivalent states, i.e. when the block B has been split into smaller blocks, including, say, $S \subseteq B$, like in Fig. 3. This means that every state $x' \in S$ is behaviourally inequivalent to every state $x'' \in B \setminus S$. The first parameter of *update* is then the bag of labels of all edges from x to S , and the second parameter is the weight $w(B)(c(x)) \in W$ of all edges from x to B , which the caller has saved from return values of previous calls to *init* and *update*, respectively. From only this information (in particular, *update* does not know x , S or B explicitly), *update* computes the triple consisting of the weight $w(S)(c(x))$ of edges from x to S , the result of $F\chi_S^B \cdot c(x)$, and the weight $w(B \setminus S)(c(x))$ of edges from x to $B \setminus S$. The two weights are stored by the caller in order to supply them to *update* in the next refinement step, and $F\chi_S^B(c(x))$ is used to split the block containing x according to (3.7).

For a given functor F , it is usually easy to derive the operations *init* and *update* once appropriate choices of the set W of weights and the weight maps w are made. We now recall refinement interfaces for some functors of interest; see [WDMS20] for the verification of the axioms.

Example 3.6 [WDMS20] (1) For $F = \mathcal{P}_\omega$, we put $W = 2 \times \mathbb{N}$. For further use in the definition of the weight maps and the refinement interface routines, we define an auxiliary function

$$(-) \succ^? 0: \mathbb{N} \rightarrow 2 \quad \text{by } (n \succ^? 0) = \min(n, 1) = \begin{cases} 1 & \text{if } n > 0 \\ 0 & \text{otherwise.} \end{cases}$$

The weight maps are defined by

$$w(B)(t) = (|t \setminus B| \succ^? 0, |t \cap B|).$$

This records whether there is an edge to $X \setminus B$ and counts the numbers of edges to states in the block B . This number is crucial to be able to implement the update routine which needs to return $\mathcal{P}_\omega \chi_S^B(c(x)) \in \mathcal{P}_\omega 3$ for a coalgebra $c: C \rightarrow FC$ and a state $x \in C$. Hence, update needs to determine whether x has an edge to $B \setminus S$ —i.e. whether $1 \in \mathcal{P}_\omega \chi_S^B(c(x))$ —given only the number k of edges from x to S and the weight $w(B)(c(x))$ (cf. Fig. 3). This task can only be accomplished if $w(B)(c(x))$ holds the number n of edges from x to B : with this information, there are edges to $B \setminus S$ iff $n - k > 0$. Recall from Example 3.3(1) that the set of labels is $A = 1$. Hence, every bag of labels is just a natural number because $\mathcal{B}_\omega A = \mathcal{B}_\omega 1 \cong \mathbb{N}$. Consequently, the interface routines

$$\text{init}: \mathcal{P}_\omega 1 \times \mathbb{N} \rightarrow 2 \times \mathbb{N} \quad \text{and} \quad \text{update}: \mathbb{N} \times (2 \times \mathbb{N}) \rightarrow (2 \times \mathbb{N}) \times \mathcal{P}_\omega 3 \times (2 \times \mathbb{N})$$

are implemented as follows:

$$\begin{aligned} \text{init}(z, n) = (0, n) \quad \text{and} \quad \text{update}(n_S, (r, n_C)) = & ((r \vee (n_{C \setminus S} \succ^? 0), n_S), \\ & (r, n_{C \setminus S} \succ^? 0, n_S \succ^? 0), \\ & (r \vee (n_S \succ^? 0), n_{C \setminus S})), \end{aligned}$$

where $n_{C \setminus S} := \max(n_C - n_S, 0)$, $\vee: 2 \times 2 \rightarrow 2$ is disjunction, and the middle return value in $\mathcal{P}_\omega 3$ is written as a bit vector of length three. The axioms in Definition 3.5 ensure that $n_S, n_C, n_{C \setminus S}$ can be understood as the numbers of edges to S, C , and $C \setminus S$, respectively.

The technique of remembering the number of edges from every state to every block is already crucial in the classical algorithm by Paige and Tarjan [PT87]. In Sect. 5.2, we will generalize this trick from \mathcal{P}_ω to arbitrary monoid-valued functors.

- (2) For the group-valued functor $G^{(-)}$ for the abelian group G , we put $W = G^{(2)} \cong G \times G$, and the weight map is defined by

$$w(B) = G^{(X_B)}: G^{(X)} \rightarrow G^{(2)} \quad \text{for every subset } B \subseteq X.$$

The refinement interface routines are implemented as follows:

$$\begin{aligned} \text{init}: G^{(1)} \times \mathcal{B}_\omega(G) \rightarrow G^{(2)} \quad \text{update}: \mathcal{B}_\omega(G) \times G^{(2)} \rightarrow G^{(2)} \times G^{(3)} \times G^{(2)} \\ \text{init} \underbrace{(g, \ell)}_{G^{(1)}(c(x))} = \underbrace{(0, g)}_{G^{(X_C)}(c(x))} \quad \text{update}(\ell, \underbrace{(r, b)}_{G^{(X_B)}(c(x))}) = \underbrace{((r + (b - \Sigma \ell), \Sigma \ell))}_{G^{(X_S)}(c(x))}, \underbrace{(r, b - \Sigma \ell, \Sigma \ell)}_{G^{(X_S^B)}(c(x))}, \underbrace{(r + \Sigma \ell, b - \Sigma \ell)}_{G^{(X_{B \setminus S})}(c(x))}, \end{aligned}$$

where $\Sigma: \mathcal{B}_\omega G \rightarrow G$ is the summation map of Notation 2.3. The terms under the braces only serve as the intuition when considering a coalgebra $c: C \rightarrow FC$ and a subblock $S \subseteq B$ of a block $B \subseteq C$. The function init is called with the bag ℓ of labels of outgoing transitions of some element $x \in C$ and the sum g of all these labels. Since $w(C)(c(x)) = (0, G^{(1)}(c(x)))$ for the whole set C , the init function simply returns $(0, g)$.

In the above update routine we have used that G has inverses. In Sect. 5, we will define a refinement interface for the monoid-valued functor $M^{(-)}$ for monoids M that are not groups, using additional data structures to make up for the lack of inverses in M .

- (3) As special instances of the previous item we obtain refinement interfaces for the functors $\mathbb{R}^{(-)}$ and $\mathbb{Z}^{(-)}$.
 (4) For a polynomial functor $F = \Sigma$, we put $W = \Sigma 2$ and

$$w(B) = \Sigma \chi_B: \Sigma X \rightarrow \Sigma 2 \quad \text{for every subset } B \subseteq X.$$

For a coalgebra $c: C \rightarrow \Sigma C$ and $B \subseteq C$, this means that $w(B)(c(x))$ consists of an operation symbol $\sigma \in \Sigma$ and a bit vector of length $\text{ar}(\sigma)$. The bit vector specifies which successor states of x are in the set B . Recall

from Example 3.3(3) that the encoding of Σ uses as labels $A = \mathbb{N}$. With $1 = \{*\}$, the init routine is given by

$$\begin{aligned} \text{init} &: \Sigma 1 \times \mathcal{B}_\omega \mathbb{N} \rightarrow \Sigma 2 \\ \text{init}(\sigma(*, \dots, *), f) &= \sigma(1, \dots, 1) \end{aligned}$$

For update, we first define the map that computes the middle component of the result:

$$\begin{aligned} \text{update}' &: \mathcal{B}_\omega \mathbb{N} \times \Sigma 2 \rightarrow \Sigma 3 \\ \text{update}'(I, \sigma(b_1, \dots, b_n)) &= \sigma(b_1 + (1 \in I), \dots, b_i + (i \in I), \dots, b_n + (n \in I)) \end{aligned}$$

Here $b_i + (i \in I)$ means $b_i + 1$ if $i \in I$ and b_i otherwise. For subsets $S \subseteq B \subseteq X$ and a state $y \in X$, this sum computes $\chi_S^B(y)$ from $\chi_S(y)$ (given by $i \in I$) and $\chi_B(y)$ (given by the bit b_i). From the value in $\Sigma 3$ thus computed, we can derive the other components of the result of update. For $k \in 3$, let $(k =)$: $3 \rightarrow 2$ be the map that compares its parameter with k , i.e.

$$(k =)(k') = \begin{cases} 1 & \text{if } k = k' \\ 0 & \text{otherwise.} \end{cases}$$

The update routine now calls update' and derives the values of type $\Sigma 2$:

$$\begin{aligned} \text{update} &= (\mathcal{B}_\omega \mathbb{N} \times \Sigma 2 \xrightarrow{\text{update}'} \Sigma 3 \xrightarrow{\langle \Sigma(2=), \text{id}_{\Sigma 3}, \Sigma(1=) \rangle} \Sigma 2 \times \Sigma 3 \times \Sigma 2) \\ \text{update}(I, t) &= \langle \Sigma(2=), \text{id}_{\Sigma 3}, \Sigma(1=) \rangle (\text{update}'(I, t)) \end{aligned}$$

In order to ensure that iteratively splitting blocks using $F\chi_S^B$ in each iteration correctly computes the minimization of the given coalgebra, we need another property of the functor F :

Definition 3.7 [WDMS20, Def. 5.1] A functor F is *zippable* if for all sets X and Y the map

$$\langle F(X+!), F(!+Y) \rangle: F(X+Y) \longrightarrow F(X+1) \times F(1+Y)$$

is injective.

All functors mentioned in Example 2.2 are zippable. Moreover, zippable functors are closed under products, coproducts (both formed point-wise), and subfunctors [WDMS20, Lemma 5.4]. However, they are not closed under functor composition: for example, $\mathcal{P}_\omega \mathcal{P}_\omega$ fails to be zippable [WDMS20, Example 5.9]. We deal with this problem by a reduction discussed in Sect. 3.5 below.

For zippable set functors F with a refinement interface, we have presented a partition refinement algorithm [WDMS20, Algorithm 7.7]. The main correctness result states that for a zippable functor equipped with a refinement interface, our algorithm correctly minimizes the given coalgebra. The low time complexity of our algorithm hinges on the time complexity of the implementations of init and update . We have shown previously [WDMS20, Theorem 7.16] that if both init and update run in linear time in its input of type $\mathcal{B}_\omega A$ alone (i.e. independently of the input coalgebra size), then our generic partition refinement algorithm runs in time $\mathcal{O}((m+n) \cdot \log n)$ on coalgebras with n states and m edges (cf. Definition 3.1). In order to cover instances where the run time of init and update depends also on the input coalgebra, we make this dependence formally explicit:

Definition 3.8 The refinement interface for a functor F has *run time factor* $p(c)$ if for every map $c: X \rightarrow FY$ (in particular for every coalgebra $c: C \rightarrow FC$),

(1) the following calls to init and update run in time $\mathcal{O}(|\ell| \cdot p(c))$ for $x \in X$, $t = c(x)$, and $S \subseteq B \subseteq X$:

$$\begin{aligned} \text{init}(F!(t), \ell) & \quad \text{where } \ell = \llbracket a \mid (a, x) \in \mathfrak{b}(t) \rrbracket \\ \text{update}(\ell, w(B)(t)) & \quad \text{where } \ell = \llbracket a \mid (a, x) \in \mathfrak{b}(t), x \in S \rrbracket. \end{aligned}$$

(2) equality of values in $\{F\chi_S^B(c(q)) \mid q \in X, S \subseteq B \subseteq X\} \subseteq F3$ can be checked in time $\mathcal{O}(p(c))$.

If $p(c)$ only depends on the number of states n and number of transitions m in c , then we write $p(n, m)$ in lieu of $p(c)$. Note that the above calls to init and update are precisely those from the axioms of the refinement interface in Definition 3.5.

Example 3.9 The powerset functor \mathcal{P}_ω and the group-valued functors $G^{(-)}$ have run time factor $p(c) = 1$ [WDMS20, Examples 6.11]. For a signature Σ where the arity of operation symbols is bounded by $b \in \mathbb{N}$, the refinement interface has run time factor $p(c) = 1$. Otherwise we define the *rank* of a finite Σ -coalgebra $c: C \rightarrow \Sigma C$ to be the maximal arity that appears in c :

$$\text{rank}(c) = \max\{\text{ar}(\sigma) \mid \sigma(x_1, \dots, x_n) = c(x) \text{ for some } x_1, \dots, x_n \text{ and } x \text{ in } C\}.$$

It is easy to see that the refinement interface for Σ (see [WDMS20, Examples 6.11.3]) has run time factor $p(c) = \text{rank}(c)$.

Since we can now describe the run time of the refinement-interface in a more fine-grained way, we can lift this to the run time analysis of the overall partition refinement algorithm.

Theorem 3.10 Let F be a zippable functor equipped with a refinement interface with run time factor $p(c)$. Then the algorithm computes the behavioural equivalence relation on an input F -coalgebra $c: C \rightarrow FC$ with n states and m transitions in $\mathcal{O}((m+n) \cdot \log n \cdot p(c))$ steps.

Proof The case where $p(c) = 1$ is proved in [WDMS20, Theorem 6.22]. We reduce the general case to this one as follows. Observe that the previous complexity analysis counts the number of basic operation performed by the algorithm (e.g. comparisons of values of type $F3$) including those performed by init and update. In that analysis init and update were assumed to have run time $\mathcal{O}(|\ell|)$, and the total number of basic operations of the algorithm is then $\mathcal{O}((m+n) \cdot \log n)$.

For the reduction, we consider ‘macro’ operations that run in $\mathcal{O}(p(c))$ time. In particular, every constant-time operation that is performed in the algorithm can be viewed as a macro performing precisely this single operation. Then we can view the generalized run time assumptions on the refinement-interface as follows:

- (1) all calls to init and update on $\ell \in \mathcal{B}_\omega A$ perform $\mathcal{O}(|\ell|)$ macro operations (each of which takes $\mathcal{O}(p(c))$ time).
- (2) all values of type $F3$ that arise during the execution of the algorithm are in the set in Definition 3.8(2). Hence, every comparison of such values is done in one macro operation, which takes $\mathcal{O}(p(c))$ steps.

By the previous complexity analysis for $p(c) = 1$, the partition refinement for a coalgebra with n states and m edges performs $\mathcal{O}((m+n) \cdot \log n)$ macro calls. Thus, the overall run time lies in

$$\mathcal{O}((m+n) \cdot \log n \cdot p(c))$$

as desired. □

Obviously, for $p(c) \in \mathcal{O}(1)$, we obtain the previous complexity.

Example 3.11 For a (possibly infinite) signature Σ , the coalgebraic partition refinement runs, by Theorem 3.10 and Example 3.9, in time $\mathcal{O}(r \cdot (m+n) \cdot \log n)$ for an input coalgebra $c: C \rightarrow \Sigma C$ with rank r , n states, and m edges. Note that every state $x \in C$ has at most r many outgoing edges in the graph representation. Hence, we have $m \leq r \cdot n$ so that the time complexity may be simplified to $\mathcal{O}(r^2 \cdot n \cdot \log n)$.

In Sect. 5 we will discuss how Theorem 3.10 instantiates to the example that mainly motivated the above generalization of the complexity analysis: weighted systems with weights from an unrestricted commutative monoid.

3.5. Combining refinement interfaces

In addition to supporting genericity via direct implementation of the refinement interface for basic functors, our tool is *modular* in the sense that it automatically derives a refinement interface for functors built from the basic ones according to the grammar (3.1). In other words, for such a combined functor the user does not need to write a single line of new code. Moreover, when the user implements a refinement interface for a new basic functor, this automatically extends the effective grammar. For example, our tool can minimize systems of type

$$FX = \mathcal{D}_\omega(\mathbb{N} \times \mathcal{P}_\omega X \times \mathcal{B}_\omega X).$$

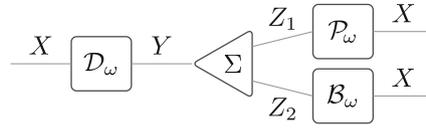


Fig. 4. Visualization of $F X = \mathcal{D}_\omega(\Sigma(\mathcal{P}_\omega X, \mathcal{B}_\omega X))$ for $\Sigma(Z_1, Z_2) = \mathbb{N} \times Z_1 \times Z_2$

However, while all basic functors from which F is formed are zippable (see Definition 3.7), there is no guarantee that F is so because zippable functors are not closed under functor composition in general. In order to circumvent this problem, a given F -coalgebra is transformed into one for the functor

$$F' X = \mathcal{D}_\omega X + (\mathbb{N} \times X \times X) + \mathcal{P}_\omega X + \mathcal{B}_\omega X.$$

This functor is obtained as the sum of all basic functors involved in F , i.e. of all the nodes in the visualization of the functor term F (Fig. 4). Then the components of the refinement interfaces of the four functors involved, viz. \mathcal{D}_ω , Σ , \mathcal{P}_ω , and \mathcal{B}_ω , are combined by disjoint union $+$. The transformation of a finite coalgebra $c: C \rightarrow FC$ into a finite F' -coalgebra introduces a set of intermediate states for each edge in the visualization of the term F ; we have labelled the edges in Fig. 4 by these sets. The construction starts with $X := C$ and constructs a finite F' -coalgebra on the set $C' := X + Y + Z_1 + Z_2$ as follows. The set Y contains an intermediate state for every \mathcal{D}_ω -edge out of a state $x \in X$, i.e.

$$Y = \{y \mid c(x)(y) \neq 0\} \subseteq \mathbb{N} \times \mathcal{P}_\omega X \times \mathcal{B}_\omega X.$$

This also yields a map $c_X: X \rightarrow \mathcal{D}_\omega Y$. Furthermore, intermediate states in Y have successors in $\mathbb{N} \times \mathcal{P}_\omega X \times \mathcal{B}_\omega X$, and by a similar definition as for Y , we obtain finite sets

$$Z_1 \subseteq \mathcal{P}_\omega X \quad Z_2 \subseteq \mathcal{B}_\omega X$$

and a map $c_Y: Y \rightarrow \mathbb{N} \times Z_1 \times Z_2$. Finally, intermediate states in Z_1 and Z_2 have successors in $\mathcal{P}_\omega X$ and $\mathcal{B}_\omega X$, respectively, which yields (inclusion) maps $c_{Z_1}: Z_1 \rightarrow \mathcal{P}_\omega X$ and $c_{Z_2}: Z_2 \rightarrow \mathcal{B}_\omega X$. Putting these maps together we obtain a finite F' -coalgebra

$$\underbrace{X + Y + Z_1 + Z_2}_{C'} \xrightarrow{c_X + c_Y + c_{Z_1} + c_{Z_2}} \mathcal{D}_\omega Y + \mathbb{N} \times Z_1 \times Z_2 + \mathcal{P}_\omega X + \mathcal{B}_\omega X \xrightarrow{\text{can}} \underbrace{F'(X + Y + Z_1 + Z_2)}_{FC'}$$

where can is the canonical inclusion map. The minimization of this F' -coalgebra yields the minimization of the given F -coalgebra (C, c) . The details of the construction in full generality and its correctness are established in [WDMS20, Section 8].

Remark 3.12 We show in the cited work that we can derive a refinement interface for F' from the refinement interfaces of the basic functors used. It is easy to see that the run time factor of the refinement interface of the above F' is

$$p(c) = p_{\mathcal{D}_\omega}(c_X) + p_\Sigma(c_Y) + p_{\mathcal{P}_\omega}(c_{Z_1}) + p_{\mathcal{B}_\omega}(c_{Z_2})$$

where the summands on the right-hand side are the run time factors of the respective refinement interfaces of the building blocks. Note that here we use the full generality of Definition 3.8, i.e. that c_X , c_Y , c_{Z_1} and c_{Z_2} are not required to be coalgebras but only maps of the shape $X \rightarrow HY$ for the relevant functor H .

3.5.1. Combination by product

CoPaR moreover implements a further optimization of this procedure that leads to fewer intermediate states in the case of polynomial functors Σ : Instead of putting the refinement interface of Σ side by side with those of its arguments, CoPaR includes a systematic procedure to combine the refinement interfaces of the arguments of Σ into a single refinement interface. For instance, starting from $F X = \mathcal{D}_\omega(\mathbb{N} \times \mathcal{P}_\omega X \times \mathcal{B}_\omega X)$ as above, a given F -coalgebra is transformed into a coalgebra for the functor

$$F'' X = \mathcal{D}_\omega X + \mathbb{N} \times \mathcal{P}_\omega X \times \mathcal{B}_\omega X,$$

effectively inducing intermediate states in Y as above but avoiding Z_1 and Z_2 . In order to run the generic partition refinement algorithm for F'' -coalgebras, we need a refinement interface for F'' . CoPaR derives a refinement interface for F'' by first combining the refinement interfaces of \mathcal{P}_ω , \mathcal{B}_ω , and that of the constant functor $X \mapsto \mathbb{N}$,

yielding a refinement interface for $X \mapsto \mathbb{N} \times \mathcal{P}_\omega X \times \mathcal{B}_\omega X$. Then, this refinement interface is combined with that of \mathcal{D}_ω , finally yielding one for F'' . The combination of refinement interfaces along coproducts of functors is already described in [WDMS20, Sec. 8.3]; in the following, we describe how refinement interfaces are combined along cartesian product \times .

Construction 3.13 Suppose we are given a finite family of functors

$$F_i : \text{Set} \rightarrow \text{Set}, \quad i \in I,$$

such that each F_i has the encoding $b_i : F_i X \rightarrow \mathcal{B}_\omega(A_i \times X)$ with label set A_i and is equipped with the refinement interface

$$\text{init}_i : F_i 1 \times \mathcal{B}_\omega A_i \rightarrow W_i, \quad \text{update}_i : \mathcal{B}_\omega A_i \times W_i \rightarrow W_i \times F_i 3 \times W_i, \quad w_i : \mathcal{P}X \rightarrow (F_i X \rightarrow W_i).$$

We construct an encoding and a refinement interface for $F X = \prod_{i \in I} F_i X$ as follows. The encoding of F is given by taking the disjoint union of the label sets A_i and the obvious component-wise definition of b :

$$A = \coprod_{i \in I} A_i \quad b : \prod_{i \in I} F_i X \longrightarrow \mathcal{B}_\omega(A \times X) \quad \underbrace{b(t)}_{A \times X \rightarrow \mathbb{N}} (\text{in}_i(a), x) = \underbrace{b_i(\text{pr}_i(t))}_{(A_i \times X) \rightarrow \mathbb{N}}(a, x) \quad \text{for every set } X. \quad (3.8)$$

The set W of weights consists of tuples of weights in W_i , and the weight function simply applies the map $w_i(C) : F_i X \rightarrow W_i$ in the i -th component:

$$W = \prod_{i \in I} W_i \quad w(C) : \prod_{i \in I} F_i X \xrightarrow{\prod_{i \in I} w_i(C)} \prod_{i \in I} W_i \quad \text{for all sets } X \text{ and } C \subseteq X.$$

The refinement interface routines of F now have the following types:

$$\begin{aligned} \text{init} : (\prod_{i \in I} F_i 1) \times \mathcal{B}_\omega(\prod_{i \in I} A_i) &\rightarrow \prod_{i \in I} W_i \\ \text{update} : \mathcal{B}_\omega(\prod_{i \in I} A_i) \times \prod_{i \in I} W_i &\rightarrow \prod_{i \in I} W_i \times \prod_{i \in I} F_i 3 \times \prod_{i \in I} W_i. \end{aligned}$$

For their definition, we introduce the following auxiliary function π_i , which restricts bags of labels to only those labels that come from A_i :

$$\pi_i : \mathcal{B}_\omega(\prod_{j \in I} A_j) \rightarrow \mathcal{B}_\omega A_i \quad \underbrace{\pi_i(f)}_{A_i \rightarrow \mathbb{N}}(a) = \underbrace{f}_{A \rightarrow \mathbb{N}}(\text{in}_i(a)) \quad \text{for every } i \in I.$$

Then we define init by

$$\text{init} : \prod_{i \in I} F_i 1 \times \mathcal{B}_\omega(\prod_{i \in I} A_i) \xrightarrow{(\text{init}_i \cdot (\text{pr}_i \times \pi_i))_{i \in I}} \prod_{i \in I} W_i,$$

i.e.

$$(\text{init}((t_j)_{j \in I}, \ell))_i = \text{init}_i(t_i, \pi_i(\ell)) \quad \text{for every } i \in I,$$

and we define update as the composite

$$\text{update} = (\mathcal{B}_\omega(\prod_{i \in I} A_i) \times \prod_{i \in I} W_i \xrightarrow{(\text{update}_i \cdot (\pi_i \times \text{pr}_i))_i} \prod_{i \in I} (W_i \times F_i 3 \times W_i) \xrightarrow{\phi} W \times F 3 \times W)$$

where ϕ is the obvious bijection reordering tuples in the evident way.

Proposition 3.14 Let $F_i : \text{Set} \rightarrow \text{Set}$, $i \in I$, be equipped with refinement interfaces with run time factors $p_i(c)$. Then Construction 3.13 defines a refinement interface for $F = \prod_{i \in I} F_i$ with run time factor

$$\max\{p_i(c) \mid i \in I\}.$$

In particular, if the refinement interface of every F_i has run time factor $p(c) = 1$, then so does the refinement interface for $F = \prod_{i \in I} F_i$.

Proof To simplify notation in the composition of maps, we define the following filter map for every $S \subseteq X$ and $i \in I$:

$$f_S : \mathcal{B}_\omega(A_i \times X) \rightarrow \mathcal{B}_\omega A_i \quad f_S(\underbrace{g}_{A_i \times X \rightarrow \mathbb{N}}) = (a \mapsto \sum_{x \in S} g(a, x)) = \{\{ a \mid (a, x) \in g, x \in S \}\}.$$

Using the filter maps f_S , we can rephrase the axioms of the refinement interface in Definition 3.5 as the commutativity of the following diagrams

$$\begin{array}{ccc} FX & & FX \\ \langle F!, f_X \cdot b \rangle \downarrow & \searrow^{w(X)} & \downarrow \langle f_S \cdot b, w(B) \rangle \\ F1 \times \mathcal{B}_\omega A & \xrightarrow{\text{init}} & W \end{array} \quad \text{and} \quad \begin{array}{ccc} FX & & FX \\ \langle f_S \cdot b, w(B) \rangle \downarrow & \searrow^{\langle w(S), F\chi_S^B, w(B \setminus S) \rangle} & \downarrow \\ \mathcal{B}_\omega A \times W & \xrightarrow{\text{update}} & W \times F3 \times W \end{array} \quad \text{for all } S \subseteq B \subseteq X. \quad (3.9)$$

In the proof that these diagrams commute, we will use the equalities

$$\begin{array}{ccc} \prod_{j \in I} F_j X & \xrightarrow{\text{pr}_i} & F_i X \\ \downarrow b & & \downarrow b_i \\ \mathcal{B}_\omega(\prod_{j \in I} A_j \times X) & & \mathcal{B}_\omega(A_i \times X) \\ \downarrow f_S & & \downarrow f_S \\ \mathcal{B}_\omega(\prod_{j \in I} A_j) & \xrightarrow{\pi_i} & \mathcal{B}_\omega A_i \end{array} \quad \text{for } i \in I \text{ and } S \subseteq X. \quad (3.10)$$

This diagram commutes because for $t \in \prod_{j \in I} F_j X$ and $a \in A_i$, we have

$$\begin{aligned} (\pi_i \cdot f_S \cdot b)(t)(a) &= \pi_i(f_S(b(t)))(a) = f_S(b(t))(in_i(a)) && \text{(def. of } \pi_i) \\ &= \sum_{x \in S} b(t)(in_i(a), x) && \text{(def. of } f_S) \\ &= \sum_{x \in S} b_i(\text{pr}_i(t))(a, x) && \text{(def. of } b, \text{ see (3.8))} \\ &= f_S(b_i(\text{pr}_i(t)))(a) = (f_S \cdot b_i \cdot \text{pr}_i)(t)(a) && \text{(def. of } f_S) \end{aligned}$$

Moreover, we will use that for every family $(Y_j)_{j \in I}$ of sets, every map $b: B \rightarrow B'$, and every $i \in I$, the diagram

$$\begin{array}{ccccc} & & \text{pr}_i \times b & & \\ & & \downarrow & & \\ (\prod_{j \in I} Y_j) \times B & \xrightarrow{\text{pr}_i \times \text{id}_B} & Y_i \times B & \xrightarrow{\text{id}_{Y_i} \times b} & Y_i \times B' \\ \langle \text{pr}_j \times \text{id}_B \rangle_{j \in I} \downarrow & \nearrow \text{pr}_i & \text{naturality} & & \uparrow \text{pr}_i \\ & & \text{of } \text{pr}_i & & \\ \prod_{j \in I} (Y_j \times B) & \xrightarrow{\prod_{j \in I} (\text{id}_{Y_j} \times b)} & \prod_{j \in I} (Y_j \times B'). & & \end{array} \quad (3.11)$$

commutes, as verified by straightforward evaluation of the maps involved. The categorically-minded reader will notice that commutation of the right-hand inner quadrangle is simply naturality of the projection maps pr_i , as indicated in the diagram; we will use this property of pr_i again later, referring to it as “naturality of pr_i ” (without requiring further understanding of the concept of natural transformation). Furthermore, we clearly have commutative squares

$$\begin{array}{ccc} B \times (\prod_{j \in I} Y_j) & \xrightarrow{b \times \text{pr}_i} & B' \times Y_i \\ \langle \text{id}_B \times \text{pr}_j \rangle_{j \in I} \downarrow & & \uparrow \text{pr}_i \\ \prod_{j \in I} (B \times Y_j) & \xrightarrow{\prod_{j \in I} (b \times \text{id}_{Y_j})} & \prod_{j \in I} (B' \times Y_j) \end{array} \quad \text{for every } i \in I. \quad (3.12)$$

We are ready to verify the axioms of the refinement interface. To this end, we use that the product projections pr_i , $i \in I$, form a jointly injective family. This means that for every pair of maps $f, g: Z \rightarrow \prod_{j \in I} Y_j$ we have that

$$\text{pr}_i \cdot f = \text{pr}_i \cdot g \text{ for all } i \in I \text{ implies } f = g. \quad (3.13)$$

Axiom for init. For every $i \in I$, the outside of the following diagram commutes because all its inner parts commute, for the respective indicated reasons:

$$\begin{array}{ccccc}
 & & w(X) = \prod_{j \in I} w_j(X) & & \\
 & & \downarrow & & \\
 \prod_{j \in I} F_j X & \xrightarrow{\text{pr}_i} & F_i X & \xrightarrow{w_i(X)} & \prod_{j \in I} W_j \\
 \downarrow \langle F^!, f_X \cdot b \rangle & \text{(3.10)} & \downarrow \langle F^!, f_X \cdot b_i \rangle & \text{Axiom for init}_i \text{ (3.9)} & \downarrow \text{pr}_i \\
 \prod_{j \in I} F_j 1 \times \mathcal{B}_\omega(\prod_{j \in I} A_j) & \xrightarrow{\text{pr}_i \times \pi_i} & F_i 1 \times \mathcal{B}_\omega A_i & \xrightarrow{\text{init}_i} & W_i \\
 \downarrow \langle \text{pr}_j \times \text{id} \rangle_{j \in I} & \text{(3.11) for } Y_j = F_j 1, b = \pi_i & \uparrow \text{pr}_i & \text{Naturality of pr}_i & \uparrow \text{pr}_i \\
 \prod_{j \in I} (F_j 1 \times \mathcal{B}_\omega(\prod_{k \in I} A_k)) & \xrightarrow{\prod_{j \in I} (\text{id} \times \pi_j)} & \prod_{j \in I} (F_j 1 \times \mathcal{B}_\omega A_j) & \xrightarrow{\prod_{j \in I} \text{init}_j} & \prod_{j \in I} W_j \\
 & & \text{init} & &
 \end{array}$$

By (3.13), we obtain commutation of the left-hand triangle in (3.9) as desired.

Axiom for update. For all $S \subseteq B \subseteq X$, the outside of the diagram below commutes because all its inner parts commute for the respective indicated reasons:

$$\begin{array}{ccccc}
 & & \prod_{j \in I} \langle w_j(S), F_j \chi_S^B, w_j(B \setminus S) \rangle & & \\
 & & \downarrow & & \\
 \prod_{j \in I} F_i X & \xrightarrow{\text{pr}_i} & F_i X & \xrightarrow{\langle w_i(S), F_i \chi_S^B, w_i(B \setminus S) \rangle} & \prod_{j \in I} (W_j \times F_j 3 \times W_j) \\
 \downarrow \langle f_S \cdot b, w(B) \rangle & \text{(3.10)} & \downarrow \langle f_S \cdot b_i, w_i(B) \rangle & \text{Axiom for update}_i \text{ (3.9)} & \downarrow \text{pr}_i \\
 \mathcal{B}_\omega(\prod_{j \in I} A_j) \times \prod_{j \in I} W_j & \xrightarrow{\pi_i \times \text{pr}_i} & \mathcal{B}_\omega A_i \times W_i & \xrightarrow{\text{update}_i} & W_i \times F_i 3 \times W_i \\
 \downarrow \langle \text{id} \times \text{pr}_j \rangle_{j \in I} & \text{(3.12) for } Y_j = W_j, b = \pi_i & \uparrow \text{pr}_i & \text{Naturality of pr}_i & \uparrow \text{pr}_i \\
 \prod_{j \in I} (\mathcal{B}_\omega(\prod_{k \in I} A_k) \times W_j) & \xrightarrow{\prod_{j \in I} \pi_j \times \text{id}} & \prod_{j \in I} (\mathcal{B}_\omega A_j \times W_j) & \xrightarrow{\prod_{j \in I} \text{update}_j} & \prod_{j \in I} (W_j \times F_j 3 \times W_j) \\
 & & \phi^{-1} \cdot \text{update} & &
 \end{array}$$

Using (3.13) and then composing with ϕ on the left, we obtain

$$\begin{aligned}
 \text{update} \cdot \langle f_S \cdot b, w(B) \rangle &= \phi \cdot \prod_{j \in I} \langle w_j(S), F_j \chi_S^B, w_j(B \setminus S) \rangle \\
 &= \langle \prod_{j \in I} w_j(S), \prod_{j \in I} F_j \chi_S^B, \prod_{j \in I} w_j(B \setminus S) \rangle = \langle w(S), F \chi_S^B, w(B \setminus S) \rangle,
 \end{aligned}$$

which is the desired right-hand triangle in (3.9).

Run time factor. Both `init` and `update` preprocess their parameters in linear time (via π_i and by accessing elements of tuples), before calling the `initi` and `updatei` routines of all F_i . Since $|I|$ is constant, this results in a run time of $\mathcal{O}(|\ell| \cdot \max_{i \in I} p_i(c))$ for $\ell \in \mathcal{B}_\omega(\prod A_i)$.

We order $F 3$ lexicographically by assuming a total order on the index set I :

$$x < y \text{ in } \prod_{i \in I} F_i 3 \quad \text{iff} \quad \text{there is } i \in I \text{ with } \text{pr}_i(x) < \text{pr}_i(y) \text{ and } \text{pr}_j(x) = \text{pr}_j(y) \text{ for all } j < i.$$

This comparison takes time $\mathcal{O}(\max_{i \in I} p_i(c))$, again because $|I|$ is constant.

Remark 3.15 In summary we obtain that the run time factor $p(c)$ for a composite functor F is the maximum of the respective run time factors of the refinement interfaces of the basic functors from which F is built. Specifically, suppose that F is built from basic functors G_1, \dots, G_n using composition, product, and coproduct, and that

G_1, \dots, G_n have refinement interfaces with respective run time factors $p_1(c), \dots, p_n(c)$. Then the modularity mechanism decomposes an F -coalgebra $c: C \rightarrow FC$ into maps $f_i: X_i \rightarrow G_i Y_i$, for $1 \leq i \leq n$ (i.e. one map per block in the illustration in Fig. 4). The run time factor for the refinement interface arising by modular construction is given by

$$p(c) = \max_{1 \leq i \leq n} p_i(f_i). \quad (3.14)$$

This is because the refinement interface for a particular functor F_i only sees the labels and weights for the map f_i and never those from other sorts.

3.6. Implementation details

Our implementation is geared towards realizing both the level of genericity and the efficiency afforded by the abstract algorithm, while staying as close as possible at the theory as presented in the present and preceding work [WDMS20]. Regarding genericity, each basic functor is defined (in its own source file) as a single Haskell data type that implements two type classes:

- (1) the class `RefinementInterface` with functions `init` and `update`, which directly corresponds to the mathematical notion (Definition 3.5), and
- (2) the class `ParseMorphism`, which provides a parser that defines the coalgebra syntax for the functor.

This means that new basic functors can be implemented without modifying any of the existing code, except for registering the new type in a list of functors (the existing functor implementations are in `src/Copar/Functors`). The type class modelling refinement interfaces is defined as follows in `CoPaR`:

```
class (Ord (F1 f), Ord (F3 f)) => RefinementInterface f where
  init :: F1 f -> [Label f] -> Weight f
  update :: [Label f] -> Weight f -> (Weight f, F3 f, Weight f)
```

Here, the type `f` serves as the name of the functor F of interest, and `Label f` is the type representing the label set A from the encoding of the functor. Similarly, the type `Weight f` represents W and the types `F1 f` and `F3 f` represents the sets $F1$ and $F3$. For example, if we want to implement the refinement interface for $FX = \mathbb{R}^{(X)}$ explicitly we can write the following:

```
data R x = R x
type instance Label R = Double
type instance Weight R = (Double, Double)
type instance F1 R = Double
type instance F3 R = (Double, Double, Double)

instance RefinementInterface R where
  init g e = (0, g)
  update l (r,b) = ((r + b - sum l, sum l),
                  (r, b - sum l, sum l),
                  (r + sum l, b - sum l))
```

The first line defines a parametrized type `R` (with one constructor of the same name) representing the functor $(\mathbb{R}^{(-)})$ in this case), with the parameter `x` representing the functor argument. The next lines define the types representing the sets that appear in the refinement interface, and the instance of `RefinementInterface` for `R` implements the `init` and `update` routines for $\mathbb{R}^{(-)}$ as we have defined them before (Example 3.6.(2)).

Concerning efficiency, `CoPaR` faithfully implements our imperative algorithm [WDMS20] in the functional language Haskell. We have made sure that this implementation actually realizes the good theoretical complexity of the algorithm. This is achieved by ample use of the `ST` monad [LP94] and by disabling lazy evaluation for the core parts of the algorithm using `GHC`'s `Strict` pragma. The `ST` monad also enables the use of efficient data structures like mutable vectors where possible.

One such data structure, which is central to the efficient implementation of the generic algorithm, is a *refinable partition*, which stores the blocks of the current partition of the state set C of the input coalgebra during the execution of the algorithm. This data structure has to provide constant time operations for finding the size of

a block, marking a state and counting the marked states in a block. Splitting a block in marked and unmarked states must only take linear time in the number of marked states of this block. Valmari and Franceschinis [VF10] have described a data structure (for use in Markov chain lumping) fulfilling all these requirements, and this is what we use in CoPaR.

Our abstract algorithm maintains two partitions P, Q of C , where P is one transition step finer than Q ; i.e. P is the partition of C induced by the map $Fq \cdot c: C \rightarrow FQ$, where $q: C \rightarrow Q$ is the canonical quotient map assigning to every state the block which contains it. The key to the low time complexity is to choose in each iteration a *subblock* S in P whose surrounding *compound block* B in Q (with $S \subseteq B$) satisfies $2 \cdot |S| \leq |B|$, and then refine Q (and P) as explained in Sect. 3.4 (see Fig. 3). This idea goes back to Hopcroft [Hop71], and is also used in all other partition refinement algorithms mentioned in the introduction. Our implementation maintains a queue of subblocks S satisfying the above property, and the termination condition $P = Q$ of the main loop then translates to this queue being empty.

One optimization that is new in CoPaR in relation to previous work [VF10, WDMS20] is that weights for blocks of exactly one state are not computed, because such blocks cannot be split any further. This has drastic performance benefits for inputs where the algorithm produces many single-element blocks early on, e.g. for nearly minimal systems or fine-grained initial partitions, see [Dei19] for details and measurements.

4. Instances

Many systems are coalgebras for functors formed according to the grammar (3.2). In Table 1, we list various system types that can be handled by our algorithm, taken from [WDMS20] except for Markov chains with weights in a monoid and weighted tree automata, which are new in the present paper. In all cases, m is the number of edges and n is the number of states of the input coalgebra, and we compare the run time of our generic algorithm with that of specifically designed algorithms from the literature. In most instances, we match the complexity of the best known algorithm. In the one case where our generic algorithm is asymptotically slower (LTS with unbounded alphabet), this is due to assuming a potentially very large number of alphabet letters—as soon as the number of alphabet letters is assumed to be polynomially bounded in the number n of states, the number m of transitions is also polynomially bounded in n , so $\log m \in \mathcal{O}(\log n)$. This argument also explains why ‘<’ and ‘=’, respectively, hold in the last two rows of Table 1, as we assume Σ to be (fixed and) finite; the case where Σ is infinite and unranked is more complicated. Details on the instantiation to weighted tree automata are discussed in Sect. 6.

Our algorithm and tool can handle further system types that arise by combining functors in various ways. For instance, so-called simple Segala systems are coalgebras for the functor $\mathcal{P}_\omega(A \times \mathcal{D}_\omega(-))$, and are minimized by our algorithm in time $\mathcal{O}((m + n) \cdot \log n)$, improving on the best previous algorithm [BEM00] and matching the complexity of the algorithm by Groote et al. [GVdV18]. Other type functors for various species of probabilistic systems are listed in [BSdV03], including the ones for general Segala systems, reactive systems, generative systems, stratified systems, alternating systems, bundle systems, and Pnueli-Zuck systems. Hence, CoPaR provides an off-the-shelf minimization tool for all these types of systems.

Remark 4.1 (Initial partitions). Note that in the classical Paige-Tarjan algorithm [PT87], the input includes an initial partition. Initial partitions as input parameters are covered via the genericity of our algorithm. In fact, initial partitions on F -coalgebras are accommodated by moving to the functor $F'X = \mathbb{N} \times FX$, where the first component of a coalgebra

$$C \xrightarrow{(c_1, c_2)} \mathbb{N} \times FX$$

assigns to each state the number of its block in the initial partition. Under the optimized treatment of the polynomial functor $\mathbb{N} \times (-)$ (Sect. 3.5), this transformation does not enlarge the state space and also leaves the run time factor $p(c)$ unchanged [WDMS20]; that is, the asymptotic run time of the algorithm remains unchanged under adding initial partitions.

5. Weighted transition systems

We have seen in Example 3.6(2) that weighted systems with weights in a group easily fit into our framework of generic partition refinement, since we can use inverses to implement the refinement interface efficiently.

Table 1. Asymptotic complexity of the generic algorithm (2017) compared to specific algorithms, for systems with n states and m transitions, respectively $m_{\mathcal{P}_\omega}$ nondeterministic and $m_{\mathcal{D}_\omega}$ probabilistic transitions for Segala systems. For simplicity, we assume that $m \geq n$ and that A and Σ are finite and fixed. M is a possibly infinite and possibly non-cancellative commutative monoid

System	Functor	Run Time	Specific Algorithm	Year	Reference
DFA	$2 \times (-)^A$	$n \cdot \log n$	= $n \cdot \log n$	1971	[Hop71]
Transition Systems	\mathcal{P}_ω	$m \cdot \log n$	= $m \cdot \log n$	1987	[PT87]
Labelled Transition Systems	$\mathcal{P}_\omega(\mathbb{N} \times -)$	$m \cdot \log m$	= $m \cdot \log m$ > $m \cdot \log n$	2004 2009	[DPP04] [Val09]
Markov Chains	$\mathbb{R}^{(-)}$	$m \cdot \log n$	= $m \cdot \log n$	2010	[VF10]
Weighted Systems	$M^{(-)}$	$m \cdot \log n \cdot \log(\min(m, M))$			
Simple Segala Systems	$\mathcal{P}_\omega(A \times -) \cdot \mathcal{D}_\omega$	$m_{\mathcal{D}_\omega} \cdot \log m_{\mathcal{P}_\omega}$	< $m \cdot \log n$ = $m_{\mathcal{D}_\omega} \cdot \log m_{\mathcal{P}_\omega}$	2000 2018	[BEM00] [GVdV18]
Colour Refinement	\mathcal{E}_ω	$m \cdot \log n$	= $m \cdot \log n$	2017	[BBG17]
Weighted Tree Automata	$M \times M^{(\Sigma^{(-)})}$ $M \times M^{(\Sigma^{(-)})}$ (M cancellative)	$m \cdot \log^2 m$ $m \cdot \log m$	< $m \cdot n$ = $m \cdot \log n$	2007 2007	[HBMM07] [HBMM07]

In the following we generalize this by allowing the weighted systems to be weighted in an arbitrary commutative monoid M that does not necessarily have inverses.

Systems with weights in a monoid are studied by Klin and Sassone [KS13], and they show that behavioural equivalence of $M^{(-)}$ -coalgebras is precisely weighted bisimilarity (cf. Example 2.2.(4)). Weighted transition systems with weights in a monoid also serve as a base for—and are in fact a special case of—weighted tree automata as studied by Högberg et al. [HBMM09], which we will discuss in the next section.

In the following we distinguish between cancellative and non-cancellative monoids because the respective refinement interfaces for $M^{(-)}$ are implemented differently, with the interface for a cancellative monoid allowing for a lower time complexity.

5.1. Cancellative monoids

Recall that a commutative monoid $(M, +, 0)$ is *cancellative* if $a + b = a + c$ implies $b = c$. Clearly, every submonoid of a group is cancellative, for example $(\mathbb{N}, +, 0)$ and $(\mathbb{Z} \setminus \{0\}, \cdot, 1)$. It is well-known that every cancellative commutative monoid M embeds into an abelian group G via the standard Grothendieck construction. Explicitly,

$$G = (M \times M) / \equiv \quad (a_+, a_-) \equiv (b_+, b_-) \text{ iff } a_+ + b_- = b_+ + a_-,$$

and the group structure is given by the usual component-wise addition on the product:

$$[(a_+, a_-)] + [(b_+, b_-)] = [(a_+ + b_+, a_- + b_-)] \quad \text{and} \quad -[(a_+, a_-)] = [(a_-, a_+)].$$

The embedding of M into G is given by the monoid homomorphism

$$\iota: M \rightarrow G \quad \iota(m) = [(m, 0)].$$

Hence, we have in total:

Corollary 5.1 A monoid is cancellative iff it is a submonoid of a group.

Informally speaking, the inverses of elements of a cancellative monoid M exist, albeit not within M itself. The embedding $M \rightarrow G$ extends to a component-wise injective natural transformation $\alpha: M^{(-)} \rightarrow G^{(-)}$, and therefore, computing behavioural equivalence for $M^{(-)}$ reduces to that of $G^{(-)}$ [WDMS20, Proposition 2.13]. Hence, we can convert every $M^{(-)}$ -coalgebra $c: C \rightarrow M^{(C)}$ into the $G^{(-)}$ -coalgebra

$$C \xrightarrow{c} M^{(C)} \xrightarrow{\alpha c} G^{(C)}$$

and use the refinement interface for $G^{(-)}$ from Example 3.6(2), obtaining:

Corollary 5.2 Let M be a cancellative monoid. Then partition refinement on a weighted transition system $c: C \rightarrow M^{(C)}$ with n states and m transitions runs in time $\mathcal{O}((m+n) \cdot \log n)$.

Indeed, this is immediate from Theorem 3.10 for $p(c) = 1$.

5.2. Non-cancellative monoids

There are monoids that are used in practice that fail to be cancellative, for example the additive monoid $(\mathbb{N}, \max, 0)$ of the tropical semiring. Assume given a non-cancellative commutative monoid $(M, +, 0)$. Then M does not embed into a group, so we need a new refinement interface for the type functor $M^{(-)}$ of M -weighted transition systems in our algorithm, rather than being able to reuse the one for group-valued functors as in the case of cancellative commutative monoids (Sect. 5.1). The basic idea in the construction of a refinement interface for $M^{(-)}$ is to incorporate bags of monoid elements into the weights, and consider subtraction of bags. We implement this idea as follows.

We use the same encoding of $M^{(-)}$ as for group-valued functors:

$$A = M_{\neq 0} = M \setminus \{0\}, \text{ and } \mathfrak{b}(f) = \{(f(x), x) \mid x \in X, f(x) \neq 0\} \text{ for } f \in M^{(X)}.$$

The refinement interface for $M^{(-)}$ has weights $W = M \times \mathcal{B}_\omega(M_{\neq 0})$ and uses weight functions

$$w(B)(f) = \left(\sum_{x \in X \setminus B} f(x), (m \mapsto |\{x \in B \mid f(x) = m\}|) \right) \in M \times \mathcal{B}_\omega(M_{\neq 0}); \quad (5.1)$$

that is, $w(B)(f)$ returns the total weight of $X \setminus B$ under f and the bag of non-zero elements of M occurring in f . The interface functions $\text{init}: M^{(1)} \times \mathcal{B}_\omega M_{\neq 0} \rightarrow W$, $\text{update}: \mathcal{B}_\omega M_{\neq 0} \times W \rightarrow W \times M^{(3)} \times W$ use the summation map $\Sigma: \mathcal{B}_\omega M \rightarrow M$ from (2.1) and are given by

$$\text{init}(f, \ell) = (0, \ell)$$

$$\text{update}(\ell, (r, c)) = ((r + \Sigma(c - \ell), \ell), (r, \Sigma(c - \ell), \Sigma(\ell)), (r + \Sigma(\ell), c - \ell)),$$

where for $a, b \in \mathcal{B}_\omega Y$, the bag $a - b$ is defined by

$$(a - b)(y) = \max(0, a(y) - b(y)).$$

As for groups, we denote elements of $M^{(3)}$ as triples of elements from M .

Proposition 5.3 For every commutative monoid M , the above functions init and update define a refinement interface for the functor $M^{(-)}$.

Proof We define the weight functions $w: \mathcal{P}X \rightarrow (M^{(X)} \rightarrow M \times \mathcal{B}_\omega(M_{\neq 0}))$ as in (5.1), and show that init and update then satisfy the two axioms in Definition 3.5. Let $t \in FX$. For the first axiom we compute as follows:

$$\begin{aligned} w(X)(t) &= \left(\sum_{x \in X \setminus X} t(x), \overbrace{(m \mapsto |\{x \in X \mid t(x) = m\}|)}^{\text{in } M_{\neq 0} \rightarrow \mathbb{N}} \right) \\ &= (0, \llbracket m \mid x \in X, t(x) = m, m \neq 0 \rrbracket) \\ &= (0, \llbracket m \mid (m, x) \in \{(t(x), x) \mid x \in X, t(x) \neq 0\} \rrbracket) \\ &= \text{init}(F!(t), \llbracket m \mid (m, x) \in \mathfrak{b}(t) \rrbracket). \end{aligned}$$

In the first of the above multiset comprehensions, different $x \in X$ with $t(x) = m$, $m \neq 0$ lead to multiple occurrences of m in the multiset, and similarly in the second multiset comprehension. Let us now verify the second axiom concerning update. In order to simplify the notation, we define

$$(t \downarrow B) \in \mathcal{B}(M_{\neq 0}), \quad (t \downarrow B)(m) = |\{x \in B \mid t(x) = m\}| \quad \text{for } B \subseteq X, t \in M^{(X)}.$$

The accumulated weight of edges into $B \subseteq X$ in $t \in M^{(X)}$ is then denoted by

$$\sum_B t := \Sigma(t \downarrow B) = \sum_{x \in B} t(x).$$

In this notation, we have

$$\begin{aligned} w(B)(t) &= \left(\sum_{x \in X \setminus B} t(x), m \mapsto |\{x \in B \mid t(x) = m\}| \right) \\ &= \left(\sum_{X \setminus B} t, (t \downarrow B) \right). \end{aligned}$$

With subtraction of bags defined as above by $(a - b)(y) = \max(0, a(y) - b(y))$ for $a, b \in \mathcal{B}_\omega Y$, we have

$$(t \downarrow B) - (t \downarrow S) = (t \downarrow (B \setminus S)) \quad \text{for } S \subseteq B \subseteq X.$$

Then we compute as follows, for $S \subseteq B \subseteq X$:

$$\begin{aligned} &\langle w(S), F\chi_S^B, w(B \setminus S) \rangle(t) \\ &= (w(S)(t), F\chi_S^B(t), w(B \setminus S)(t)) \\ &= \left(\left(\sum_{X \setminus S} t, (t \downarrow S) \right), \left(\sum_{X \setminus B} t, \sum_{B \setminus S} t, \sum_S t \right), \left(\sum_{X \setminus (B \setminus S)} t, (t \downarrow (B \setminus S)) \right) \right) \\ &= \left(\left(\sum_{X \setminus B} t + \underbrace{\Sigma(t \downarrow (B \setminus S))}_{\sum_{B \setminus S} t}, (t \downarrow S) \right), \left(\sum_{X \setminus B} t, \Sigma(t \downarrow (B \setminus S)), \Sigma(t \downarrow S) \right), \right. \\ &\quad \left. \left(\sum_{X \setminus B} t + \sum_S t, ((t \downarrow B) - (t \downarrow S)) \right) \right) \\ &= \left(\left(\sum_{X \setminus B} t + \Sigma((t \downarrow B) - (t \downarrow S)), (t \downarrow S) \right), \right. \\ &\quad \left. \left(\sum_{X \setminus B} t, \Sigma((t \downarrow B) - (t \downarrow S)), \Sigma(t \downarrow S) \right), \right. \\ &\quad \left. \left(\sum_{X \setminus B} t + \Sigma(t \downarrow S), ((t \downarrow B) - (t \downarrow S)) \right) \right) \\ &\stackrel{(*)}{=} \text{update}((t \downarrow S), \left(\sum_{X \setminus B} t, (t \downarrow B) \right)) \\ &= \text{update}(\llbracket m \in M_{\neq 0} \mapsto |\{x \in S \mid t(x) = m\}| \rrbracket, w(B)(t)) \\ &= \text{update}(\llbracket m \in M_{\neq 0} \mid (m, x) \in \mathfrak{b}(t), x \in S \rrbracket, w(B)(t)) \\ &= \text{update}(\llbracket a \in A \mid (a, x) \in \mathfrak{b}(t), x \in S \rrbracket, w(B)(t)), \end{aligned}$$

where the step labelled (*) uses the definition of update:

$$\text{update}(\ell, (r, c)) = ((r + \Sigma(c - \ell), \ell), (r, \Sigma(c - \ell), \Sigma(\ell)), (r + \Sigma(\ell), c - \ell)).$$

In order to determine the run time factor of the above refinement interface, we need to describe how we handle elements of $W = M \times \mathcal{B}_\omega(M_{\neq 0})$ in the routines of the refinement interface. We implement a bag in $\mathcal{B}_\omega(M_{\neq 0})$ as a balanced search tree with keys $M_{\neq 0}$ and values \mathbb{N} . In addition to the standard structure of a balanced search tree, we store in every node the value $\Sigma(b)$, where b is the bag encoded by the subtree rooted at that node. Hence, for every bag $b \in \mathcal{B}_\omega(M_{\neq 0})$, the value $\Sigma(b)$ is immediately available at the root node of the search tree for b . For CoPaR, we have implemented the basic operations on balanced search trees following Adams [Ada93]. For the complexity analysis, we prove that maintaining the values $\Sigma(b)$ in the nodes only adds constant overhead to the operations on search trees, so that we obtain

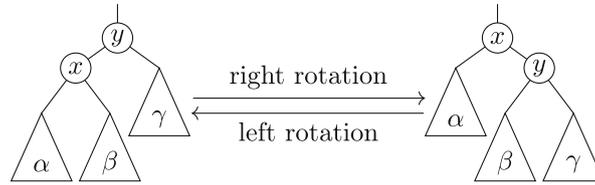


Fig. 5. Rotation operations in binary search trees

Proposition 5.4 The above refinement interface for $M^{(-)}$ has run time factor

$$p(n, m) = \log \min(|M|, m).$$

More precisely, the above functions $\text{init}(f, \ell)$ and $\text{update}(\ell, (r, c))$ can be computed in time $\mathcal{O}(|\ell| \cdot \log \min(|M|, m))$, where m is the number of edges in the input coalgebra, and values in $M^{(3)}$ can be compared in constant time.

Proof For a node x in a binary search tree encoding a bag in $\mathcal{B}_\omega(M_{\neq 0})$ as described above, we write $\Sigma(x)$ for the value in M stored at that node.

Note that our search trees cannot have more nodes than the size $|M|$ of their index set, and the number of nodes is also not greater than the number m of all edges. Hence, their size is bounded by $\min(|M|, m)$.

Recall from algorithm textbooks (e.g. [CLR90, Section 14]) that the key operations insert, delete and search have logarithmic time complexity in the size of a given balanced binary search tree.

We need to argue that maintaining the values $\Sigma(x)$ in the nodes does not increase this complexity. This is obvious for the search operation as it does not change its argument search tree. For insert and delete, these operations essentially trace down one path starting at the root to a node (at worst, a leaf) of the given search tree (as described e.g. by [CLR90]). Additionally, we need to rebalance the search tree after inserting or deleting a node. This is done by tracing back the same path to the root and (possibly) performing *rotations* on the nodes occurring on that path. Rotations are local operations changing the structure of a search tree but preserving the inorder key ordering of subtrees (see Fig. 5).

Clearly, in order to maintain the correct summation values in a search tree under a rotation, we only need to adjust those values in the nodes x and y . This is achieved as follows:

$$\begin{aligned} \Sigma(x) &= \Sigma(\alpha) + \Sigma(\beta); & \Sigma(y) &= \Sigma(x) + \Sigma(\gamma) & \text{for left rotation,} \\ \Sigma(y) &= \Sigma(\beta) + \Sigma(\gamma); & \Sigma(x) &= \Sigma(\alpha) + \Sigma(y) & \text{for right rotation.} \end{aligned}$$

In addition, when inserting or deleting a node x we must recompute the $\Sigma(y)$ of all nodes y along the path from the root to x when we trace that path back to the root. This can clearly be performed in constant time for each node y since $\Sigma(y)$ is the sum (in M) of $\Sigma(y_1)$ and $\Sigma(y_2)$, which are stored at the child nodes y_1 and y_2 of y , respectively.

In summary, we see that maintaining the desired summation values only requires an additional constant overhead in the backtracing step. Consequently, the operations of our balanced binary search trees run in time $\mathcal{O}(\log \min(|M|, m))$. Subtraction $c - \ell$ of bags c, ℓ performs $|\ell|$ -many calls to delete on c , and computing the sum $\Sigma(\ell)$ takes time $\mathcal{O}(|\ell|)$, since ℓ , the bag of labels passed to update, is not represented as a search tree. Hence, we obtain the desired overall time complexity $\mathcal{O}(|\ell| \cdot \log \min(|M|, m))$ of update.

Similarly, for $\text{init}(f, \ell)$, we need $|\ell|$ -many calls to insert in order to initialize the search tree representing $\mathcal{B}_\omega(M_{\neq 0})$. □

Remark 5.5 It is no coincidence that we use $\mathcal{B}_\omega(M_{\neq 0})$ in the refinement interface. In fact, for every set X , the set $\mathcal{B}_\omega X$, with union of bags as addition, is the free commutative monoid on X . Moreover, $\mathcal{B}_\omega X$ is cancellative, so that we may use a form of subtraction on bags. Thus, we see that $\mathcal{B}_\omega(M_{\neq 0})$ is a canonical cancellative monoid containing $M_{\neq 0}$ (via the identification of elements of $M_{\neq 0}$ with singleton bags). Moreover, the summation map

$$\Sigma : \mathcal{B}_\omega M \rightarrow M$$

is the canonical unique monoid homomorphism freely extending the identity map on M . Thus, this map allows us to go back from bags to monoid elements. This is essentially the point of Eilenberg-Moore algebras in general (cf. Remark 2.4).

Corollary 5.6 Let M be any commutative monoid. Then partition refinement on weighted transition systems $c: C \rightarrow M^{(C)}$ with n states and m transitions runs in time $\mathcal{O}((m+n) \cdot \log n \cdot \log \min(|M|, m))$.

Indeed, this is immediate by Proposition 5.4 and Theorem 3.10.

6. Weighted tree automata

We proceed to take a closer look at weighted tree automata as a worked example. It is this example that mainly motivates the discussion of non-cancellative monoids in the last section, since in this case the generic algorithm improves on the run time of the best known specific algorithms in the literature.

Weighted tree automata simultaneously generalize tree automata and weighted (word) automata. A partition refinement construction for weighted automata (w.r.t. weighted bisimilarity) was first considered by Buchholz [Buc08, Theorem 3.7]. Högberg et al. first provided an efficient partition refinement algorithm for tree automata [HBMM09], and moreover for weighted tree automata [HBMM07]. Generally, tree automata differ from word automata in replacing the input alphabet, which may be seen as sets of unary operations, with an algebraic signature Σ :

Definition 6.1 Let $(M, +, 0)$ be a commutative monoid. A (bottom-up) *weighted tree automaton* (WTA) (over M) consists of a finite set X of states, a finite signature Σ , an output map $f: X \rightarrow M$, and for each $k \geq 0$, a transition map $\mu_k: \Sigma_k \rightarrow M^{X^k \times X}$, where Σ_k denotes the set of k -ary input symbols in Σ ; the maximum arity of symbols in Σ is called the *rank*.

Given a weighted tree automaton $(X, f, (\mu_k)_{k \in \mathbb{N}})$ as in Definition 6.1 we see that it is, equivalently, a finite coalgebra for the functor $FX = M \times M^{(\Sigma X)}$, where we identify the signature Σ with its corresponding polynomial functor $\Sigma X = \coprod_{\sigma/k \in \Sigma} X^k$. Indeed, $(\mu_k)_{k \geq 0}$ is equivalently expressed by a map

$$\bar{\mu}: X \rightarrow M^{(\Sigma X)} \quad \text{with} \quad \bar{\mu}(x)(\sigma(x_1, \dots, x_k)) := \mu_k(\sigma)((x_1, \dots, x_k), x). \quad (6.1)$$

Note that $\bar{\mu}$ is finitely supported because X and ΣX are finite. Thus we obtain a coalgebra

$$g: X \rightarrow M \times M^{(\Sigma X)} \quad \text{with} \quad g(x) = (f(x), \bar{\mu}(x)). \quad (6.2)$$

Example 6.2 Let $\Sigma = \{*/2\}$ be a signature with a single binary symbol. Consider the WTA (X, f, μ_2) for Σ over the monoid $(\mathbb{Z}, \max, -\infty)$ (simply (\mathbb{Z}, \max) in the following) with $X = \{a, b, c, d\}$ and $f: X \rightarrow \mathbb{Z}$ being the constant function $x \mapsto 1$ and μ_2 given by

$$\begin{array}{llll} \mu_2(*) (b, a, a) = 3, & \mu_2(*) (a, a, a) = 5, & \mu_2(*) (a, b, b) = 5, & \mu_2(*) (b, b, b) = 2, \\ \mu_2(*) (b, a, c) = 5, & \mu_2(*) (c, a, c) = 7, & \mu_2(*) (b, a, d) = 5, & \mu_2(*) (a, c, d) = 7 \quad \text{and} \\ \mu_2(*) (x, y, z) = -\infty & \text{otherwise.} & & \end{array}$$

This WTA is equivalently expressed as the coalgebra $g: X \rightarrow \mathbb{Z} \times (\mathbb{Z}, \max)^{(X \times X)}$ given by $g(x) = (f(x), \bar{\mu}(x))$ with $\bar{\mu}: X \rightarrow (\mathbb{Z}, \max)^{(X \times X)}$ defined as

$$\begin{array}{llll} \bar{\mu}(a)(b, a) = 3, & \bar{\mu}(a)(a, a) = 5, & \bar{\mu}(b)(a, b) = 5, & \bar{\mu}(b)(b, b) = 2, \\ \bar{\mu}(c)(b, a) = 5, & \bar{\mu}(c)(c, a) = 7, & \bar{\mu}(d)(b, a) = 5, & \bar{\mu}(d)(a, c) = 7, \end{array}$$

where again $\bar{\mu}$ is $-\infty$ in all other cases. In the syntax of our tool, the map $\bar{\mu}: X \rightarrow \mathbb{Z}^{(X \times X)}$ can be written as:

```
(Z, max)^(X x X)
a: {(b, a): 3, (a, a): 5}
b: {(a, b): 5, (b, b): 2}
c: {(b, a): 5, (c, a): 7}
d: {(b, a): 5, (a, c): 7}
```

For the minimization of weighted tree automata, two bisimulation notions are considered in the literature [Buc08, HBMM07]: *forward* and *backward* bisimulation. Here, we treat backward bisimulation, as it corresponds to coalgebraic behavioural equivalence.

Definition 6.3 (Högberg et al. [HBMM07, Def. 16]). A *backward bisimulation* on a weighted tree automaton $(X, f, (\mu_k)_{k \in \mathbb{N}})$ is an equivalence relation $R \subseteq X \times X$ such that for every $(p, q) \in R$, $\sigma/k \in \Sigma$, and every $L \in \{D_1 \times \cdots \times D_k \mid D_1, \dots, D_k \in X/R\}$ the following equation holds:

$$\sum_{w \in L} \mu_k(\sigma)(w, p) = \sum_{w \in L} \mu_k(\sigma)(w, q). \quad (6.3)$$

Remark 6.4 Note that L consists of the $w \in X^k$ such that $e^k(w) = L$, where $e^k: X^k \rightarrow (X/R)^k$ is the k -fold power of the canonical quotient map $e: X \rightarrow X/R$.

Example 6.5 The equivalence relation $R = \{a, b\}^2 \cup \{c\}^2 \cup \{d\}^2$ is a backward bisimulation for the automaton defined in Example 6.2.

We can regard the output map f as a transition map for a constant symbol, so it suffices to consider the functor $FX = M^{(\Sigma X)}$ (and in fact the output map is ignored in the definition of backward bisimulation given above and in [HBMM07]). Then, we obtain the following result:

Proposition 6.6 Backward bisimulation of weighted tree automata coincides with behavioural equivalence of $M^{(\Sigma(-))}$ -coalgebras.

Proof We show that for every $M^{(\Sigma(-))}$ -coalgebra $\bar{\mu}: X \rightarrow M^{(\Sigma X)}$ defined as in (6.1), an equivalence relation $R \subseteq X \times X$ is a backward bisimulation on the corresponding WTA iff the canonical quotient map $e: X \rightarrow X/R$ is an $M^{(\Sigma(-))}$ -coalgebra homomorphism with domain $(X, \bar{\mu})$. First, let $x \in X$, $\sigma/k \in \Sigma$, $D_1, \dots, D_k \in X/R$, and $L = D_1 \times \cdots \times D_k$ for some equivalence relation R . Then we have $\sigma(D_1, \dots, D_k) \in \Sigma(X/R)$, and recalling Remark 6.4, we can rewrite the sums in (6.3) in terms of the map $M^{(\Sigma e)}: M^{(\Sigma X)} \rightarrow M^{(\Sigma(X/R))}$ as follows:

$$\begin{aligned} \sum_{w \in L} \mu_k(\sigma)(w, x) &= \sum_{\substack{w \in X^k \\ e^k(w) = L}} \mu_k(\sigma)(w, x) = \sum_{\substack{w \in X^k \\ \tau \in \Sigma_k \\ e^k(w) = L \\ \tau = \sigma}} \mu_k(\tau)(w, x) = \sum_{\substack{w \in X^k \\ \tau \in \Sigma_k \\ \Sigma e(\tau(w)) = \sigma(L)}} \mu_k(\tau)(w, x) = \sum_{\substack{\tau(w) \in \Sigma X \\ \Sigma e(\tau(w)) = \sigma(L)}} \mu_k(\tau)(w, x) \\ &\stackrel{(6.1)}{=} \sum_{\substack{\tau(w) \in \Sigma X \\ \Sigma e(\tau(w)) = \sigma(L)}} \bar{\mu}(x)(\tau(w)) = \sum_{\substack{t \in \Sigma X \\ \Sigma e(t) = \sigma(L)}} \bar{\mu}(x)(t) = M^{(\Sigma e)}(\bar{\mu}(x))(\sigma(L)). \end{aligned}$$

Hence, for every equivalence relation $R \subseteq X \times X$ we have the following chain of equivalences:

$$\begin{aligned} &R \text{ is a backward bisimulation} \\ \Leftrightarrow &\forall (p, q) \in R, \sigma/k \in \Sigma, L \in (X/R)^k: \quad \sum_{w \in L} \mu_k(\sigma)(w, p) = \sum_{w \in L} \mu_k(\sigma)(w, q) \\ \Leftrightarrow &\forall (p, q) \in R, \sigma/k \in \Sigma, L \in (X/R)^k: \quad M^{(\Sigma e)}(\bar{\mu}(p))(\sigma(L)) = M^{(\Sigma e)}(\bar{\mu}(q))(\sigma(L)) \\ \Leftrightarrow &\forall (p, q) \in R, t \in \Sigma(X/R): \quad M^{(\Sigma e)}(\bar{\mu}(p))(t) = M^{(\Sigma e)}(\bar{\mu}(q))(t) \\ \Leftrightarrow &\forall (p, q) \in R: \quad M^{(\Sigma e)}(\bar{\mu}(p)) = M^{(\Sigma e)}(\bar{\mu}(q)) \\ \Leftrightarrow &\forall (p, q) \in R: \quad (M^{(\Sigma e)} \cdot \bar{\mu})(p) = (M^{(\Sigma e)} \cdot \bar{\mu})(q) \end{aligned}$$

The last equation holds iff there exists a map $r: X/R \rightarrow M^{(\Sigma(X/R))}$ such that $r \cdot e = M^{(\Sigma e)} \cdot \bar{\mu}$, that is, iff e is a coalgebra homomorphism:

$$\begin{array}{ccc} X & \xrightarrow{\bar{\mu}} & M^{(\Sigma X)} \\ e \downarrow & & \downarrow M^{(\Sigma e)} \\ X/R & \xrightarrow{r} & M^{(\Sigma(X/R))} \end{array}$$

Example 6.7 As we mentioned already, weighted tree automata subsume two other notions:

- (1) For M being the Boolean monoid (w.r.t. or), we obtain (bottom-up) non-deterministic tree automata and their bisimilarity.

(2) If all operations $\sigma \in \Sigma$ have arity one, then we obtain ordinary weighted automata and weighted bisimilarity.

Since $M^{\Sigma(-)}$ is composed of $M^{(-)}$ and a polynomial functor Σ , we have all the refinement interfaces defined already in previous work and in Sect. 5, distinguishing like Högberg et al. [HBMM07] between cancellative and non-cancellative monoids and obtaining run time factors of $p(n, m) = 1$ and $p(n, m) = \log \min(|M|, m)$ respectively. Both cases appear in the applications of weighted tree automata to natural language processing, for example the cancellative monoid of real numbers with addition $(\mathbb{R}, +, 0)$ [PBTK06, PK07], or the non-cancellative monoid $(\mathbb{N}, \max, 0)$ from the tropical semiring [MK06].

In the following, we additionally distinguish between finite and infinite monoids.

Theorem 6.8 Let M be a commutative monoid. On weighted tree automata with n states, k transitions, and rank r , our algorithm runs in time

- (1) $\mathcal{O}((r^2k + rm) \cdot \log(k + n))$ if M is cancellative or finite, and
- (2) $\mathcal{O}((rk + n) \cdot \log(k + n) \cdot (\log k + r))$ otherwise.

Proof The functor $FX = M^{\Sigma X}$ is first transformed into $F'X = M^{(X)} + \Sigma X$ according to Sect. 3.5. Given a coalgebra $c: C \rightarrow M^{\Sigma C}$ with $n = |K|$ states, this transformation introduces a set K of intermediate states, one for every outgoing transition from every $x \in C$:

$$K = \{(x, t) \in C \times \Sigma C \mid c(x)(t) \neq 0\},$$

hence $|K| = k$. The given coalgebra structure yields the two evident maps $c_1: C \rightarrow M^{(K)}$ and $c_2: K \rightarrow \Sigma C$ given by

$$c_1(x)(x', t) = \begin{cases} c(x)(t) & \text{if } x' = x \\ 0 & \text{otherwise,} \end{cases} \quad c_2(x, t) = t.$$

It takes k edges to encode c_1 and at most $k \cdot r$ edges to encode c_2 . Partition refinement is now performed on the following F' -coalgebra:

$$C + K \xrightarrow{c_1 + c_2} M^{(K)} + \Sigma C \xrightarrow{M^{(\text{inl})} + \Sigma \text{inl}} M^{(C+K)} + \Sigma(C + K),$$

where $\text{inl}: C \rightarrow C + K$ and $\text{inr}: K \rightarrow C + K$ are the canonical injections. This coalgebra on $C + K$ has $n' := |C + K| = n + k$ states and at most $m' = (r + 1) \cdot k$ edges. Since the refinement interface for F' is a combination of those of $M^{(-)}$ and Σ , its run time factor is given by the maximum of the run time factors $p_M(c_1)$ and $p_\Sigma(c_2)$, respectively, of those two refinement interfaces (cf. Remark 3.15). We can further simplify this maximum to the asymptotically equivalent sum

$$p(c) = p_M(c_1) + p_\Sigma(c_2).$$

Since $p_\Sigma(c_2) = r$ and the number of edges in c_1 is bounded by k , we obtain, by Theorem 3.10, an overall time complexity of

$$\mathcal{O}((m' + n') \cdot \log n' \cdot (p_M(c_1) + p_\Sigma(c_2))) = \mathcal{O}((r \cdot k + n) \cdot \log(n + k) \cdot (p_M(c_1) + r)).$$

We proceed by distinguishing the following cases:

- (a) If M is cancellative, then we can use the refinement interface for groups (see Example 3.6(2)) with $p_M(c_1) = 1$ as explained in Sect. 5.1. Thus, the overall time complexity simplifies to

$$\mathcal{O}((r \cdot k + n) \cdot \log(n + k) \cdot r) = \mathcal{O}((r^2 \cdot k + r \cdot n) \cdot \log(n + k)).$$

- (b) Otherwise, we have $p_M(c_1) = \log \min(|M|, k)$ by Proposition 5.4, since c_1 has at most k edges.

(b1) If M is finite, then we have $p_M(c_1) \leq \log |M| \in \mathcal{O}(1)$, and thus obtain the same overall run time complexity as in the previous case. We have thus proved item (1) in the statement of the theorem.

(b2) If M is infinite, then $\log \min(|M|, k) = \log k$. Thus, the overall run time is in

$$\mathcal{O}((r \cdot k + n) \cdot \log(n + k) \cdot (\log k + r)).$$

Note that the number m of edges of the input coalgebra satisfies $m \leq rk$. Thus, for a fixed input signature Σ , we see that m and k are asymptotically equivalent. If we further assume that $m \geq n$, which means that there are no isolated states, then we obtain the bound in Table 1:

Corollary 6.9 Let M be a commutative monoid. For a fixed input signature and on input coalgebras with n states and $m \geq n$ edges, our algorithm runs in time

- (1) $\mathcal{O}(m \cdot \log(m))$, if M is cancellative, and
- (2) $\mathcal{O}(m \cdot \log(m)^2)$, otherwise.

Remark 6.10 We now provide a comparison of the complexity of Högberg et al.’s algorithm with the instances of our algorithm for weighted tree automata.

- (1) For arbitrary (non-cancellative) monoids, Högberg et al. establish a complexity of $\mathcal{O}(r \cdot k \cdot n)$ [HBMM07, Theorem 27]. Under the assumptions of Corollary 6.9, we see that our bound indeed improves the complexity $\mathcal{O}(r \cdot k \cdot n) = \mathcal{O}(m \cdot n)$ of Högberg et al.’s algorithm. To see this, note first that the number m of edges is in $\mathcal{O}(n^{r+1})$, so that we obtain

$$\mathcal{O}(m \cdot \log(m)^2) \subsetneq \mathcal{O}(m \cdot \sqrt[r+1]{m}) \subseteq \mathcal{O}(m \cdot \sqrt[r+1]{n^{r+1}}) = \mathcal{O}(m \cdot n)$$

using in the first step that $\mathcal{O}(\log(m)^d) \subsetneq \mathcal{O}(m^c)$ for every $d \geq 1$ and $0 < c < 1$.

- (2) For cancellative monoids, the time bound given by Högberg et al. is $\mathcal{O}(r^2 \cdot k \cdot \log n)$ [HBMM07, Theorem 29]. Assuming again that $m \geq n$, and recalling that $rk \geq m$, the complexity of our algorithm according to Theorem 6.8 is $\mathcal{O}(r^2 \cdot k \cdot \log(k + n))$, i.e. only slightly worse for non-constant signatures.

In addition to guaranteeing a good theoretical complexity, our tool immediately yields an efficient implementation. For the case of non-cancellative monoids, this is, to the best of our knowledge, the only available implementation of partition refinement for weighted tree automata.

7. Evaluation and benchmarking

We report on a number of benchmarks³ that illustrate the practical scalability of our tool CoPaR and hence our generic algorithm. These benchmarks cover a selection of different system types and include randomly generated inputs as well as real world examples. We also compare CoPaR with two other minimization tools, where applicable. Details and results of further benchmarks, in particular for the optimizations described in Sect. 3.5.1 and at the end of Sect. 3.6, are reported in [Dei19]. All benchmarks were run and measured on the same Intel® Core™ i5-6500 processor with 3.20GHz clock rate running a Linux system. We report the timing results of our tool CoPaR (compiled with GHC 8.4.4) separately for the three phases parsing, initialization and the actual refinement loop.

Recall from Remark 4.1, that the input coalgebra implicitly defines an initial partition according to the output behaviour of states. We have taken care to ensure that in all the following benchmarks, this initial partition is still coarse, i.e. the algorithm has to perform some actual refinement steps after initialization.

7.1. Maximal feasible weighted tree automata

We first focus on the instantiation of our algorithm for weighted tree automata as described in Sect. 6. Previous studies on the practical performance of partition refinement on large labelled transition systems [Val10, Val09] show that memory rather than run time seems to be the limiting factor. Since labelled transition systems are a special case of weighted tree automata, we expect to see similar phenomena. Hence, we evaluate the maximal automata sizes that can be processed on a typical current computer setup: We randomly generate weighted tree automata for various signatures and monoids, looking for the maximal size of WTAs that can be handled with 16 GB of RAM, and measure the respective run times of our tool.

³The full set of benchmarks and their results can be found at <https://git8.cs.fau.de/software/copar-benchmarks>

Table 2. Processing times for partition refinement on maximal weighted tree automata (i.e. coalgebras for $M \times M^{(\Sigma(-))}$) in 16 GB of memory with n states and 50 transitions per state, leading to n' states and m edges in total. The column ‘Init’ provides the time needed to compute the initial partition P_1 (on all n' states), and ‘Refine’ the time to compute the final partitions P'_f (on n' states) and P_f (omitting intermediate states)

Functor Parameters		Input File			Partition Sizes				Time (s) to		
Monoid M	$\Sigma X =$	n	n'	m	Size (MB)	P'_1	P'_f	P_f	Parse	Init	Refine
$(2, \vee, 0)$	$4 \times X$	154863	7898013	15486300	101	6	774313	154863	53	36	183
	$4 \times X^2$	138000	7038000	20700000	143	6	7037670	138000	54	31	311
	$4 \times X^3$	134207	6844557	26841400	191	6	6844557	134207	62	27	285
	$4 \times X^4$	92491	4717041	23122750	163	6	4717041	92491	51	28	175
	$4 \times X^5$	86852	4429452	26055600	186	6	4429452	86852	54	26	165
$(\mathbb{N}, \max, 0)$	$4 \times X$	156913	8002563	15691300	131	437	784564	156913	62	31	93
	$4 \times X^2$	118084	6022284	17712600	143	416	6021960	118084	58	30	119
	$4 \times X^3$	100799	5140749	20159800	158	414	5140749	100799	57	28	105
	$4 \times X^4$	92879	4736829	23219750	181	409	4736829	92879	60	28	105
	$4 \times X^5$	94451	4817001	28335300	219	417	4817001	94451	63	29	108
$(\mathcal{P}_\omega(64), \cup, \emptyset)$	$4 \times X$	152107	7757457	15210700	141	54	760534	152107	65	27	149
	$4 \times X^2$	134082	6838182	20112300	176	54	6837891	134082	66	29	229
	$4 \times X^3$	94425	4815675	18885000	157	54	4815675	94425	57	25	185
	$4 \times X^4$	83431	4254981	20857750	170	54	4254981	83431	56	24	175
	$4 \times X^5$	92615	4723365	27784500	223	54	4723365	92615	64	21	194

To this end, we minimize randomly generated coalgebras for the functors

$$FX = M \times M^{(\Sigma X)}, \quad \text{with } \Sigma X = 4 \times X^r, \quad (7.1)$$

where r ranges over $\{1, \dots, 5\}$ and weight monoids M range over

- $(2, \vee, 0)$ (functor available as powerset $\mathbb{P}(X)$ in CoPaR)
- $(\mathbb{N}, \max, 0)$ (syntactically: $(\mathbb{Z}, \max)^\wedge(X)$)
- $(2, \vee, 0)^{64} \cong (\mathcal{P}_\omega(64), \cup, \emptyset)$ (syntactically: $(\text{Word}, \text{or})^\wedge(X)$)

We write n for the number of states, k for the number of transitions, and m for the number of edges in the coalgebra encoding. When generating a coalgebra with n states, we randomly create 50 outgoing transitions per state, obtaining $k = 50 \cdot n$ transitions in total. The transformation described in Sect. 3.5 additionally introduces one intermediate state per transition, leading to an actual number $n' = 51 \cdot n$ of states. Every transition of rank r has one incoming edge and r outgoing edges, hence $m = (r + 1) \cdot k = 50 \cdot (r + 1) \cdot n$.

Table 2 lists the maximal sizes of weighted tree automata that CoPaR is able to process in the mentioned 16 GB of RAM, along with the associated run times. Since our implementation of the refinement interface for $\mathcal{P}_\omega \cong (2, \vee, 0)^{(-)}$ is optimized for its specific functor, the tool needs less memory in this case, allowing for higher values of n , an effect that decreases with increasing rank r .

When generating weighted tree automata for partition refinement, one needs to be careful to avoid systems for which the partition refinement is trivial—either because the ‘init’ step already distinguishes all states or because ‘refine’ does not split any blocks. To this end, we restrict to generating at most 50 different elements of M in each automaton, and we generate 50 transitions per state. As one can tell from the columns P'_1 and P'_f in Table 2, this strategy excludes the above-mentioned trivial cases: the initial partition P'_1 is small compared to the total number of states n' whereas the size P'_f of the final partition nearly matches n' . This implies that the actual partition refinement process has to do the main work of distinguishing nearly all of the n' states; hence, the test cases present the worst case when it comes to the run time of the algorithm.

The first refinement step produces in the order of $|\Sigma| \cdot \min(50, |M|)^r$ subblocks, implying earlier termination for high values of $|M|$ and r and explaining the slightly longer run time for $M = (2, \vee, 0)$ on small r . We note in summary that WTAs with well over 15 million edges are processed in less than five minutes, and in fact the run time of minimization is of the same order of magnitude as that of input parsing.

Table 3. Performance on weighted tree automata from the `berkeleyparser` project, after transforming them to $\mathbb{R}^{\Sigma(-)}$ -coalgebras. The column ‘Init’ provides the time needed to compute the initial partition P'_1 (on all n' states) and ‘Refine’ the time to compute the final partitions P'_f (on n' states) and P_f (omitting intermediate states)

Filename	Input			Partition Sizes			Time (s) to		
	n	n'	m	P'_1	P'_f	P_f	Parse	Init	Refine
<code>arb_sm5</code>	1026	946509	2788134	315	26427	404	9.41	2.47	3.63
<code>bul_sm5</code>	2545	834777	2173611	743	37660	1327	8.80	2.07	2.57
<code>chn_sm5</code>	945	1110445	3236715	426	31386	579	11.31	3.28	4.43
<code>eng_sm6</code>	1133	1843351	5410006	449	30692	548	19.71	5.25	7.59
<code>fra_sm5</code>	737	451171	1343591	178	18770	319	4.21	0.98	1.57
<code>ger_sm5</code>	986	617762	1848751	273	20324	484	6.22	1.48	2.03

Since the publication of the conference paper [DMSW19], we have optimized the memory consumption in CoPaR, especially in the refinement interface for the functor Σ . With the optimizations, CoPaR can now handle coalgebras with 20% more states within the same memory limit of 16GB of RAM for the cases with $\Sigma X = 4 \times X$ and even 75% more states for the cases with $\Sigma X = 4 \times X^5$.

7.2. Weighted tree automata from grammars

One widespread practical use of weighted tree automata is to learn natural language grammars [PBTK06, PK07]. Methods of this type have been implemented in the `berkeleyparser` project⁴. The project also makes six language grammars available, which we use to demonstrate that CoPaR is capable of handling inputs that arise in practice.

We have parsed⁵ the weighted tree automata given in the `berkeleyparser`-specific format, obtaining coalgebras for the functor

$$FX = \mathbb{R}^{\Sigma X}$$

where Σ is (the polynomial functor for) a ranked alphabet containing, in the case at hand, only operation symbols of arities 1 and 2. The ranked alphabet is implicit in the original grammar file. In the grammar file, every state (i.e. symbol of the grammar) is of the shape S_i where $i \in \mathbb{N}$ is an index and S indicates that there are operation symbols $S/1$ and $S/2$ in Σ . The rules in the grammar files are of the shape

$$S_i \xrightarrow{w} T_j R_k \quad \text{or} \quad S_i \xrightarrow{w} T_j$$

where S_i, T_j, R_k are states and $w \in [0, 1]$ is the weight of the transition. In the induced coalgebra $\mu: X \rightarrow (\mathbb{R}, +, 0)^{\Sigma X}$, the state set X is the set of symbols, and the above rules correspond to the transitions

$$\mu(S_i)(S/2(T_j, R_k)) = w \quad \text{resp.} \quad \mu(S_i)(S/1(T_j)) = w.$$

Hence, by definition of the `berkeleyparser`-format, every symbol (i.e. state) mentions at most two different operation symbols in its outgoing transitions.

Table 3 shows the performance of CoPaR on the input files obtained. Most of the running time is spent during parsing of our generic input format (Sect. 3.1) whereas the actual partition refinement runs in under 13 seconds on each of the files. The number of states mentioned explicitly in the files is denoted by n , and n' denotes the total number of states obtained after introducing intermediate states via our modularity mechanism (Sect. 3.5). As one can tell from the partition sizes P'_1 and P'_f , a considerable number of refinement steps is needed before the final partition is reached.

⁴Available at <https://github.com/slavpetrov/berkeleyparser>

⁵See <https://git8.cs.fau.de/software/copar-benchmarks/-/tree/master/berkeleyparser> for details

Table 4. Performance on PRISM benchmarks

PRISM Model	Functor	Input		Time (s) to			Time (s) of	
		States	Edges	Parse	Init	Refine	Valmari	mCRL2
<code>fms (n=4)</code>	$\mathbb{R}^{(-)}$	35910	237120	0.48	0.12	0.16	0.21	–
<code>fms (n=5)</code>	$\mathbb{R}^{(-)}$	152712	1111482	2.46	0.68	1.1	1.21	–
<code>fms (n=6)</code>	$\mathbb{R}^{(-)}$	537768	4205670	9.94	2.91	5.56	5.84	–
<code>wlan2_collide(COL=2, TRANS_TIME_MAX=10)</code>	$\mathbb{N} \times \mathcal{P}(\mathbb{N} \times \mathcal{D}_\omega(-))$	65718	94452	0.5	0.3	0.58	0.12	0.41
<code>wlan0_time_bounded (TRANS_TIME_MAX=10, DEADLINE=100)</code>	$\mathbb{N} \times \mathcal{P}(\mathbb{N} \times \mathcal{D}_\omega(-))$	582327	771088	5.19	3.13	5.5	0.88	3.18
<code>wlan1_time_bounded (TRANS_TIME_MAX=10, DEADLINE=100)</code>	$\mathbb{N} \times \mathcal{P}(\mathbb{N} \times \mathcal{D}_\omega(-))$	1408676	1963522	13.37	6.18	16.18	2.44	8.44

7.3. Benchmarks for PRISM models

In order to see how CoPaR performs on models of other system types that arise in practice, we have taken two kinds of models from the benchmark suite [KNP12] of the probabilistic model checker PRISM [KNP11]. We derived coalgebras for the functors

- $FX = \mathbb{R}^{(X)}$ from continuous time Markov chains (CTMC), and
- $FX = \mathbb{N} \times \mathcal{P}(\mathbb{N} \times (\mathcal{D}_\omega X))$ from Markov decision processes (MDP).

This translation deliberately ignores the variable valuations present in the original benchmark models to avoid situations where all states are already distinguished after the first refinement step. For MDPs, the translation instead generates a coarse initial partition for each model (the outer $\mathbb{N} \times (-)$). For the CTMCs considered, the functor $\mathbb{R}^{(-)}$ is already sufficient since the initial partition distinguishes states by the accumulated weight of their outgoing transitions.

Like in the case of WTAs, the functor for MDPs is a composite of several basic functors and thus requires use of the construction described in Sect. 3.5. Two of the benchmarks are shown in Table 4 with different parameters, resulting in three differently sized coalgebras each. The *fms* family of systems model a flexible manufacturing system [CT93] as CTMCs (without initial partition), and we minimize them under the usual weighted bisimilarity, i.e. as $\mathbb{R}^{(-)}$ -coalgebras. The *wlan* benchmarks [KNS02] model various aspects of the IEEE 802.11 Wireless LAN protocol as MDPs.

Table 4 also includes the total run time of two additional partition refinement tools: Valmari’s C++ implementation⁶ of algorithms described by Valmari and Franceschinis [VF10, Val10], which can minimize MDPs as well as CTMCs, and the tool `ltspbisim` from the mCRL2 toolset⁷ [BGK⁺19] version 201808.0, which implements a recently discovered refinement algorithm for MDPs [GVdV18] (but does not support CTMCs directly, hence there is no data in the first three lines).

The results in Table 4 show that refinement for the *fms* benchmarks is faster than for the respective *wlan* ones, even though the first group has more edges. This is due to (a) the fact that the functor for MDPs is more complex and thus introduces more indirection into our algorithms, as explained in Sect. 3.5, and (b) that our optimization for one-element blocks fires much more often for *fms*.

It is also apparent that CoPaR is slower than both of the other tools in our comparison for the presented examples in Table 4: CoPaR takes slightly more than 30 seconds whereas Valmari’s optimized implementation only takes 2.44. To some extent, this performance difference can be attributed to the fact that our implementation is written in Haskell and the other tools in C++. In addition, CoPaR incurs a certain amount of overhead for genericity and modularity. Moreover, CoPaR’s input format as described in Sect. 3.1 is much more complex to parse than Valmari’s format, which is essentially a whitespace separated list of integers.

⁶ Available at <https://git8.cs.fau.de/hpd/mdpmin-valmari>

⁷ One should note that mCRL2 offers a whole suite of reasoning services besides partition refinement.

8. Conclusion and future work

We have instantiated a generic and efficient partition refinement algorithm that we introduced in previous work [WDMS20] to weighted (tree) automata, and we have refined the generic complexity analysis of the algorithm to cover this case. Moreover, we have described an implementation of the generic algorithm in the form of the tool CoPaR, which supports the modular combination of basic system types without requiring any additional implementation effort, and allows for easy incorporation of new basic system types by implementing a generic refinement interface. CoPaR is currently concerned entirely with partition refinement, and does not implement other algorithmic tasks (such as simulation, visualization, or model checking), which for specific system types such as labelled transition systems or Markov chains are covered by existing well-developed tool suites (Sect. 7). The salient feature of CoPaR is its genericity, which allows for instantiation of the partition refinement algorithm to new system types with minimal effort.

In future work, we will further broaden the range of system types that our algorithm and tool can accommodate, and provide support for base categories beyond sets, e.g. nominal sets, which underlie nominal automata [BKL14, SKMW17], or algebraic categories [EM11, SBBR13].

Concerning genericity, there is an orthogonal approach by Ranzato and Tapparo [RT08], which is generic over notions of process equivalence but fixes the system type to standard labelled transition systems; see also [GJKW17]. Similarly, Blom and Orzan [BO03, BO05] present *signature refinement*, which covers, e.g. strong and branching bisimulation as well as Markov chain lumping, but requires adapting the algorithm for each instance. These algorithms have also been improved using symbolic techniques (e.g. [vDvdP18]). Moreover, many of the mentioned approaches and others [BDJM05, BO03, BO05, GH02, vDvdP18] focus on parallelization. We will explore in future work whether symbolic and distributed methods can be lifted to coalgebraic generality.

In a recent alternative approach to bisimilarity minimization called *partition aggregation* [BC20], the behavioural equivalence relation is approximated from below, rather than from above as in partition refinement. Partition aggregation has worse run-time complexity than partition refinement on the global task of minimizing entire (reachable) systems, but it can be executed partially and thus may be more efficient on the local task of checking equivalence of two given states in a labelled transition system. In future work, we aim at a coalgebraic generalization of partition aggregation, which might help circumvent the bottleneck of (linear) memory consumption that is intrinsic to partition refinement [Val10, Val09].

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- [Ada93] Adams S (1993) Efficient Sets: A Balancing Act. *J Funct Program* 3(4):553–561
- [Awo10] Awodey S (2010) *Category Theory*, volume 52 of *Oxford Logic Guides*. Oxford University Press, Oxford, 2 edition
- [BBG17] Berkholz C, Bonsma PS, Grohe M (2017) Tight Lower and Upper Bounds for the Complexity of Canonical Colour Refinement. *Theory Comput Syst* 60(4):581–614
- [BC20] Björklund J, Cleophas L (2020) Aggregation-Based Minimization of Finite State Automata. *Acta Informatica*, January 2020
- [BDJM05] Bergamini D, Descoubes N, Joubert C, Mateescu R (2005) BISIMULATOR: A Modular Tool for on-the-Fly Equivalence Checking. In: *Tools and Algorithms for the Construction and Analysis of Systems, TACAS 2005*, volume 3440 of *LNCIS*. Springer, pp 581–585
- [BEM00] Baier C, Engelen B, Majster-Cederbaum M (2000) Deciding bisimilarity and similarity for probabilistic processes. *J Comput Syst Sci* 60:187–231

- [BGK⁺19] Bunte O, Groote JF, Keiren Jeroen JA, Laveaux M, Neele T, de Vink Erik P, Wesselink W, Wijs A, Willemsse Tim AC (2019) The mCRL2 Toolset for Analysing Concurrent Systems: Improvements in Expressivity and Usability. In: Tools and Algorithms for the Construction and Analysis of Systems, TACAS 2019, pp. 21–39
- [BKL14] Bojańczyk M, Klin B, Lasota S (2014) Automata Theory in Nominal Sets. *Log Methods Comput Sci* 10(3)
- [BO03] Blom S, Orzan S (2006) Distributed Branching Bisimulation Reduction of State Spaces. In: Parallel and Distributed Model Checking, PDMC 2003, volume 89 of ENTCS. Elsevier, pp. 99–113
- [BO05] Blom S, Orzan S (2005) A Distributed Algorithm for Strong Bisimulation Reduction of State Spaces. *J Softw Tools Technol Transfer* 7(1):74–86
- [BSdV03] Bartels F, Sokolova A, de Vink E (2003) A Hierarchy of Probabilistic System Types. In: Coalgebraic Methods in Computer Science, CMCS 2003, volume 82 of ENTCS. Elsevier, pp. 57–75
- [Buc08] Buchholz P (2008) Bisimulation Relations for Weighted Automata. *Theor Comput Sci* 393:109–123
- [CLR90] Cormen T, Leiserson C, Rivest R (1990) Introduction to Algorithms. MIT Press, Cambridge
- [CT93] Ciardo G, Trivedi KS (1993) A Decomposition Approach for Stochastic Reward Net Models. *Perform Evaluat* 18(1):37–59
- [Dei19] Deifel H-P (2019) Implementation and Evaluation of Efficient Partition Refinement Algorithms. Master's thesis, Friedrich-Alexander Universität Erlangen-Nürnberg, <https://hpdeifel.de/master-thesis-deifel.pdf>
- [DHS03] Derisavi S, Hermanns H, Sanders W (2003) Optimal state-space lumping in Markov chains. *Inf Process Lett* 87(6):309–315
- [DMSW17] Dorsch U, Milius S, Schröder L, Wißmann T (2017) Efficient Coalgebraic Partition Refinement. In: Concurrency Theory, CONCUR 2017, volume 85 of LIPIcs, pp. 32:1–32:16. Schloss Dagstuhl – Leibniz-Zentrum für Informatik
- [DMSW19] Deifel H-P, Milius S, Schröder L, Wißmann T (2019) Generic Partition Refinement and Weighted Tree Automata. In: ter Beek Maurice H, McIver Annabelle, Oliveira José N editors, Formal Methods – The Next 30 Years, Cham, 10. Springer, pp. 280–297
- [DPP04] Dovier A, Piazza C, Policriti A (2004) An efficient algorithm for computing bisimulation equivalence. *Theor Comput Sci* 311(1-3):221–256
- [EM11] Esik Z, Maletti A (2011) The Category of Simulations for Weighted Tree Automata. *Int J Found Comput Sci* 22:1845–1859, 12
- [GH02] Garavel H, Hermanns H (2002) On Combining Functional Verification and Performance Evaluation Using CADP. In: Formal Methods Europe, FME 2002, volume 2391 of LNCS. Springer, pp. 410–429
- [GJKW17] Groote JF, Jansen DN, Keiren Jeroen JA, Wijs Anton (2017) An $O(m \log n)$ Algorithm for Computing Stuttering Equivalence and Branching Bisimulation. *ACM Trans Comput Log* 18(2):13:1–13:34
- [Gri73] Gries D (1973) Describing an Algorithm by Hopcroft. *Acta Informatica* 2:97–109
- [GVdV18] Groote JF, Verduzco JR, de Vink Erik P (2018) An Efficient Algorithm to Determine Probabilistic Bisimulation. *Algorithms* 11(9):131
- [HBMM07] Högberg J, Maletti A, May J (2007) Bisimulation Minimisation for Weighted Tree Automata. In: Developments in Language Theory, DLT 2007, volume 4588 of LNCS. Springer, pp. 229–241
- [HBMM09] Högberg J, Maletti A, May J (2009) Backward and Forward Bisimulation Minimization of Tree Automata. *Theor Comput Sci* 410:3539–3552
- [Hop71] Hopcroft John (1971) An $n \log n$ Algorithm for Minimizing States in a Finite Automaton. In: Theory of Machines and Computations. Academic Press, pp. 189–196
- [HT92] Huynh D, Tian L (1992) On Some Equivalence Relations for Probabilistic Processes. *Fund Inform* 17:211–234
- [KNP11] Kwiatkowska M, Norman G, Parker D (2011) PRISM 4.0: Verification of Probabilistic Real-Time Systems. In: Computer Aided Verification, CAV 2011, volume 6806 of LNCS. Springer, pp. 585–591
- [KNP12] Kwiatkowska MZ, Norman G, Parker D (2012) The PRISM Benchmark Suite. In: Ninth International Conference on Quantitative Evaluation of Systems, QEST 2012, London, United Kingdom, September 17–20, 2012. IEEE Computer Society, pp. 203–204
- [KNS02] Kwiatkowska MZ, Norman G, Sproston J (2002) Probabilistic Model Checking of the IEEE 802.11 Wireless Local Area Network Protocol. In: Holger Hermanns and Roberto Segala, editors, Process Algebra and Probabilistic Methods, Performance Modeling and Verification, Second Joint International Workshop PAPM-PROBMIV 2002, Copenhagen, Denmark, July 25–26, 2002. Proceedings, volume 2399 of Lecture Notes in Computer Science. Springer, pp. 169–187
- [Knu01] Knuutila T (2001) Re-describing an algorithm by Hopcroft. *Theor Comput Sci* 250:333–363
- [KS90] Kanellakis P, Smolka S (1990) CCS expressions, finite state processes, and three problems of equivalence. *Inf Comput* 86(1):43–68
- [KS13] Klin B, Sassone V (2013) Structural Operational Semantics for Stochastic and Weighted Transition Systems. *Inf Comput* 227:58–83
- [LP94] Launchbury J, Peyton J, Simon L (1994) Lazy functional state threads. In Sarkar Vivek, Ryder Barbara G, Soffa Mary Lou editors, Proceedings of the ACM SIGPLAN'94 Conference on Programming Language Design and Implementation (PLDI), Orlando, Florida, USA, June 20–24, 1994. ACM, pp. 24–35
- [Mil80] Milner R (1980) A Calculus of Communicating Systems, volume 92 of LNCS. Springer
- [MK06] May J, Knight K (2006) Tiburon: A Weighted Tree Automata Toolkit. In: Ibarra OH, Yen H-C editors, Implementation and Application of Automata, Berlin, Heidelberg, pp. 102–113
- [Par81] Park D (1981) Concurrency and Automata on Infinite Sequences. In: Theoretical Computer Science, 5th GI-Conference, volume 104 of LNCS. Springer, pp. 167–183
- [PBTK06] Petrov S, Barrett L, Thibaux R, Klein D (2006) Learning Accurate, Compact, and Interpretable Tree Annotation. In: Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, Sydney, Australia, July 2006. Association for Computational Linguistics, pp. 433–440
- [PK07] Petrov S, Klein D (2007) Improved Inference for Unlexicalized Parsing. In Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference, Rochester, New York, April 2007. Association for Computational Linguistics, pp. 404–411
- [PT87] Paige R, Tarjan R (1987) Three Partition Refinement Algorithms. *SIAM J Comput* 16(6):973–989
- [RT08] Ranzato F, Tapparo F (2008) Generalizing the Paige-Tarjan Algorithm by Abstract Interpretation. *Inf Comput* 206:620–651
- [Rut00] Rutten Jan (2000) Universal Coalgebra: A Theory of Systems. *Theor Comput Sci* 249:3–80

- [SBBR13] Silva A, Bonchi F, Bonsangue MM, Rutten Jan JMM (2013) Generalizing determinization from automata to coalgebras. *Log Methods Comput Sci* 9(1)
- [Seg95] Segala R (1995) Modelling and Verification of Randomized Distributed Real-Time Systems. PhD thesis, MIT
- [SKMW17] Schröder L, Kozen D, Milius S, Wißmann T (2017) Nominal automata with name binding. In: *Foundations of Software Science and Computation Structures, FOSSACS 2017*, volume 10203 of LNCS, pp. 124–142
- [Val09] Valmari A (2009) Bisimilarity minimization in $\mathcal{O}(m \log n)$ time. In: *Applications and Theory of Petri Nets, PETRI NETS 2009*, volume 5606 of LNCS. Springer, pp. 123–142
- [Val10] Valmari A (2010) Simple bisimilarity minimization in $\mathcal{O}(m \log n)$ time. *Fund Inform*, 105(3):319–339
- [vDvdP18] van Dijk T, van de Pol J (2018) Multi-Core Symbolic Bisimulation Minimization. *J Softw Tools Technol Transfer* 20(2):157–177
- [VF10] Valmari A, Franceschinis G (2010) Simple $\mathcal{O}(m \log n)$ time Markov chain lumping. In: *Tools and Algorithms for the Construction and Analysis of Systems, TACAS 2010*, volume 6015 of LNCS. Springer, pp. 38–52
- [vG01] van Glabbeek R (2001) The Linear Time—Branching Time Spectrum I; the Semantics of Concrete, Sequential Processes. In: Bergstra J, Ponse A, Smolka S (2001) editors, *Handbook of Process Algebra*. Elsevier, pp. 3–99
- [WDMS20] Wißmann T, Dorsch U, Milius S, Schröder L (2020) Efficient and Modular Coalgebraic Partition Refinement. *Logical Methods in Computer Science*, volume 16, Issue 1, January 2020

Received 28 February 2020

Accepted in revised form 11 November 2020 by Annabelle McIver, Maurice ter Beek and Cliff Jones

Published online 23 March 2021