Mathematical Structures in Computer Science (2020), **30**, pp. 752–832
doi:10.1017/S0960129520000195

CAMBRIDGE UNIVERSITY PRESS

PAPER

Forward analysis for WSTS, part I: completions

Alain Finkel[1] and Jean Goubault-Larrecq[2,*,†]

[1]Université Paris-Saclay, ENS Paris-Saclay, CNRS, LSV, Institut Universitaire de France, 91190 Gif-sur-Yvette, France and
[2]Université Paris-Saclay, ENS Paris-Saclay, CNRS, LSV, 91190 Gif-sur-Yvette, France
*Corresponding author. Email: goubault@ens-paris-saclay.fr

(Received 29 November 2016; revised 13 July 2020; accepted 18 August 2020; first published online 19 October 2020)

Abstract

We define representations for downward-closed subsets of a rich family of well-quasi-orders, and more generally for closed subsets of an even richer family of Noetherian topological spaces. This includes the cases of finite words, of multisets, of finite trees, notably. Those representations are given as finite unions of ideals, or more generally of irreducible closed subsets. All the representations we explore are computable, in the sense that we exhibit algorithms that decide inclusion, and compute finite unions and finite intersections. The origin of this work lies in the need for computing finite representations of sets of successors of the downward closure of one state, or more generally of a downward-closed set of states, in a well-structured transition system, and this is where we start: we define adequate notions of completions of well-quasi-orders, and more generally, of Noetherian spaces. For verification purposes, we argue that the required completions must be ideal completions, or more generally sobrifications, that is, spaces of irreducible closed subsets.

Keywords: Well-structured transition systems, well-quasi-orderings, downwards-closed subsets, Noetherian topological spaces, finite computable representations, ideal completions.

1. Introduction

Well-structured transition systems (WSTSs) are a paradigmatic class of infinite-state transition systems on which many properties of interest in verification are decidable (Abdulla et al. 1996; Finkel 1987; Finkel and Schnoebelen 2001). They include Petri nets, affine counter systems, lossy channel systems, data nets, and many more.

Briefly put, a *WSTS* is a triple $(X, \to, \leq)$ where $\leq$ is a well-quasi-order on the (possibly infinite) state space X, and $\to$ is a monotonic transition relation on X. (We define well-quasi-orders in Section 3.1.) To simplify things slightly, by *monotonic* we mean strongly monotonic, namely that if $x \to x'$ and $x \leq y$, then there is a state y' such that $x' \leq y'$ and $y \to y'$. The set of one-step predecessors $\mathrm{Pre}(E) = \{x \in X \mid \exists x' \in E, x \to x'\}$ of any upward-closed subset E is then upward-closed again, where E is *upward-closed* if and only if $x \in E$ and $x \leq y$ imply $y \in E$. Similarly, the sets $\mathrm{Pre}^k(E)$ of k-step predecessors, $\mathrm{Pre}^{\leq k}(E)$ of at-most-k-step predecessors, and $\mathrm{Pre}^*(E) = \bigcup_{k\in\mathbb{N}} \mathrm{Pre}^{\leq k}(E)$ of iterated predecessors of the upward-closed set E are upward-closed. The fact that $\leq$ is a well-quasi-order implies (see Section 3.1 again) that every upward-closed subset is the upward closure $\uparrow A$ of a *finite* set of points A (a *basis* of the set). This implies that $\mathrm{Pre}^*(E) = \mathrm{Pre}^{\leq k}(E)$ for some $k \in \mathbb{N}$: write $\mathrm{Pre}^*(E)$ as $\uparrow A$ with $A = \{x_1, \ldots, x_n\}$, realize that for each i, x_i must be in $\mathrm{Pre}^{\leq k}(E)$

†An extended abstract already appeared in Proc. 26th International Symposium on Theoretical Aspects of Computer Science (STACS'09).

for some k, and take the largest of these ks. In particular, there is a simple algorithm that decides *coverability* in WSTS, namely, which decides whether, given $x \in X$ and a basis for an upward-closed subset E, whether one can reach an element of E in finitely many $\rightarrow$ steps starting from x (Abdulla et al. 2000): compute $\mathrm{Pre}^{\leq 0}(E) = E$, iterate using $\mathrm{Pre}^{\leq k+1}(E) = E \cup \mathrm{Pre}(\mathrm{Pre}^{\leq k}(E))$ until $\mathrm{Pre}^{\leq k+1}(E) \subseteq \mathrm{Pre}^{\leq k}(E)$, at which point $\mathrm{Pre}^{\leq k}(E) = \mathrm{Pre}^*(E)$, then test whether $x \in \mathrm{Pre}^*(E)$. For this to work, we need the WSTS to be effective, which means that we can compute $\mathrm{Pre}(E)$ for E upward-closed, and we can test inclusion.

This was generalized to topological WSTS by the second author (Goubault-Larrecq 2010); the algorithm is the same, but X must now be a Noetherian space (see Section 3.3), the sets E are required to be open, and $\rightarrow$ is a lower semi-continuous relation. Topological WSTS include WSTS, but also some other infinite-state systems, among which the class of lossy concurrent polynomial games is probably the most interesting new instance – see Section 6 of Goubault-Larrecq (2010).

The algorithm described above works *backward*, but sometimes we would prefer a forward algorithm that would compute $\downarrow \mathrm{Post}^*(\{x\})$, where $\downarrow$ denotes downward closure, $\mathrm{Post}(E) = \{x' \in X \mid \exists x \in E, x \rightarrow x'\}$ is the set of one-step successors of E and $\mathrm{Post}^*(E) = \bigcup_{k \in \mathbb{N}} \mathrm{Post}^k(E)$. The set $\downarrow \mathrm{Post}^*(\{x\})$ is called the *cover* of x and can be used to decide coverability as well: one can reach $\uparrow E$ from x if and only if $\downarrow \mathrm{Post}^*(\{x\})$ intersects E.

However, although the backward procedure always terminates, it is often slow. Forward procedures, when they exist, may fail to terminate: on lossy channel systems, any terminating forward procedure would enable us to decided boundedness, which is undecidable (Mayr 2003). But they often give results faster in practice. For this reason, only the non-terminating forward procedure is implemented in the tool TREX (Abdulla et al. 1998).

The cover also provides more useful information than the set computed by the backward algorithm. For example, the cover is a good first approximation of the reachability set $\mathrm{Post}^*(\{x\})$, and the original reachability algorithms for Petri nets rely on the computation of covers (Kosaraju 1982; Lambert 1992; Mayr 1981). This can also serve as a first step toward model checking liveness properties, as in Emerson and Namjoshi (1998) and more recently in Blondin et al. (2017a,b).

For Petri nets, the cover can be computed by the so-called coverability tree algorithm of Karp and Miller (1969). Part II of this paper generalizes this to a large class of WSTS (Finkel and Goubault-Larrecq 2012). Part III of this paper defines and studies very-WSTS, a subclass of WSTS, for which the cover is computable and for which linear temporal logic (introduced in Pnueli 1977) model checking is decidable. The present part I deals with an important preparatory step: characterizing downward-closed subsets of well-quasi-ordered sets X. We shall see that such downward-closed subsets can always be written as the downward closure of *finitely* many points in a completion $\widehat{X}$ of X. In fact, we start by defining the possible relevant completions from a verification perspective and realize that the smallest possible one is the ideal completion or equivalently (in the more general, Noetherian case) the sobrification of X. We shall then explore concrete computer representations for elements of $\widehat{X}$, for a large class of Noetherian spaces X (in particular, well-quasi-orders) that includes most of the spaces needed in the verification of WSTS today.

This paper is an extended version of Finkel and Goubault-Larrecq (2009). Before this paper, and except for some partial results (Emerson and Namjoshi 1998; Finkel 1990; Geeraerts et al. 2006), a general theory of downward-closed sets was missing. This may explain the scarcity of forward algorithms for WSTS. Quoting Abdulla et al. (2004b): "Finally, we aim at developing generic methods for building downward-closed languages, in a similar manner to the methods we have developed for building upward-closed languages in Abdulla et al. (2000). This would give a general theory for forward analysis of infinite-state systems, in the same way the work in Abdulla et al. (2000) is for backward analysis." Our contribution is to provide such a theory of downward-closed sets.

2. Related Work

The coverability for general WSTS was shown decidable using a backward algorithm presented in 1996 (Abdulla et al. 1996); this algorithm was an abstraction of the coverability algorithm for lossy channel systems (Abdulla and Jonsson 1993). Coverability for vector addition systems with resets had been shown decidable by Arnold and Latteux (1978, Theorem 5, p. 391). Interestingly, the latter algorithm is an early instance of the backward algorithm presented in Abdulla et al. (1996) and applied to $\mathbb{N}^n$.

While this paper is not about algorithms, it is worth recalling that the inspiration for our line of work, which culminates in part III (Blondin et al. 2017a,b), comes from Karp and Miller's celebrated finite coverability tree algorithm (Karp and Miller 1969) for Petri nets. This arguably computes a finite representation of the cover $\downarrow \text{Post}^*(\{x\})$, and we expand on that in Finkel and Goubault-Larrecq (2012, Section 4.1) and in Blondin et al. (2017a,b). Further related work on this issue can be found in that paper. What matters to us here is that, while the state space of a Petri net is $\mathbb{N}^k$, Karp and Miller's finite representation is given by finitely many points in the *completion* $\mathbb{N}_\omega^k$, where $\mathbb{N}_\omega$ is $\mathbb{N}$ plus a fresh, infinite element ω.

The focus of this paper is on *finite representations* of downward-closed subsets of well-quasi-ordered sets and more generally of closed subsets of Noetherian spaces. In computer speak, we focus on data structures rather than algorithms. Mathematically, we shall need to define the right notion of completion $\widehat{X}$ for well-quasi-ordered sets, resp., Noetherian spaces X – these will be the familiar constructions of ideal completion, resp. sobrification – and to study finite representations of their (downward-)closed subsets.

Data structures are a prerequisite to define algorithms. In our context, one may argue that what we need is the so-called *adequate domain of limits* (*ADLs*), as defined by Ganty, Geeraerts, Raskin, and van Begin (Ganty et al. 2006; Geeraerts et al. 2006). An ADL is an axiomatization of a data structure on which the authors' expand, enlarge, and check procedure, which computes the cover, works. Alternatively, an ADL is an axiomatization for a relevant completion $\widehat{X}$. We shall see that these completions have strong ties with the ideal completion, resp., sobrification, mentioned above.

In the special case of finite words, such finite representations were developed by Abdulla et al. (1998) as specific regular expressions called SREs (simple regular expressions) and word-products. In their case, the alphabet is finite, with equality as well-quasi-order. Similar representations also apply to certain more complex well-quasi-ordered sets of letters, as demonstrated in Abdulla et al. (2004b) for example. More generally, it had been shown by Kabil and Pouzet that this representation is in fact valid for any arbitrary well-quasi-ordered set of letters (Kabil and Pouzet 1992). We improve on this slightly by showing that this even works for all Noetherian sets of letters (Section 7). Interestingly, one of the key notions we use in the proof is that of an irreducible (closed) subset, which comes from topology. This is also a central concept in Kabil and Pouzet's proof, narrowed down to well-quasi-orders.

We define finite representations for (downward-)closed subsets of a large class of Noetherian data types, including tuples of natural numbers or finite words, as mentioned above, but also many more. Some of them are representations of (downward-)closed subsets in well-known well-quasi-orders, as in the case of finite multisets (Section 8), or of finite labeled trees (Section 11, by far the most technical part of this paper. The fact that the right completion for languages of trees can be described as certain regular tree languages was also observed by Wies, Zufferey, and Henzinger, for finite sets of labels (Wies et al. 2010). They did not characterize what kind of regular tree language is required precisely, which we do). Some others are representations of closed subsets of Noetherian spaces that do not arise from well-quasi-orders. For instance, we deal with polynomial ideals in Section 6, with finite words again but with a different topology, the prefix topology, in Section 9.

3. Preliminaries

We shall borrow from theories of well-quasi-orderings (wqos), as used classically in WSTSs (Abdulla et al. 2000; Finkel and Schnoebelen 2001), from domain theory (Abramsky and Jung 1994; Gierz et al. 2003) and from topology (Goubault-Larrecq 2013). We recap most of what we need. The purpose is not to give a crash course on these three fields, rather to fix notations and notions.

3.1 Order

A *quasi-ordering* $\leq$ is a reflexive and transitive relation on a set X. It is a (partial) *ordering* iff it is antisymmetric. A set X equipped with a partial ordering is a *poset*.

We write $\geq$ for the opposite quasi-ordering, $\approx$ for the equivalence relation $\leq \cap \geq$, $<$ for the associated strict ordering ($\leq \smallsetminus \approx$), and $>$ for the converse ($\geq \smallsetminus \approx$) of $<$. The *upward closure* $\uparrow E$ of a set E is $\{y \in X \mid \exists x \in E, x \leq y\}$. The *downward closure* $\downarrow E$ is $\{y \in X \mid \exists x \in E, y \leq x\}$. A subset E of X is *upward-closed* if and only if $E = \uparrow E$, that is, any element greater than or equal to some element in E is again in E, which was the definition we gave in the introduction. The notion of *downward-closed* sets is defined similarly. When the ambient space X is not clear from context, we shall write $\downarrow_X E$, $\uparrow_X E$ instead of $\downarrow E$, $\uparrow E$. We also write $\uparrow x$ instead of $\uparrow\{x\}$ and $\downarrow x$ instead of $\downarrow\{x\}$.

A quasi-ordering is *well-founded* iff it has no infinite strictly descending chain, that is, $x_0 > x_1 > \cdots > x_i > \cdots$. An *antichain* is a set of pairwise incomparable elements. A quasi-ordering is *well* if and only it is well founded and has no infinite antichain.

There are a number of equivalent definitions for *wqos*. One is that, from any infinite sequence $x_0, x_1, \ldots, x_i, \ldots$, one can extract an infinite ascending chain $x_{i_0} \leq x_{i_1} \leq \cdots \leq x_{i_k} \leq \cdots$, with $i_0 < i_1 < \cdots < i_k < \cdots$. Another one is that any upward-closed subset can be written $\uparrow E$, with E *finite*. Such a finite E is called a *finite basis* for the upward-closed set. In a wqo, every upward-closed set has a minimal finite basis, composed of the subset of its pairwise incomparable, minimal elements. We shall see another, topological, characterization of wqos below.

There is a rich supply of wqos. First, for any $k \in \mathbb{N}$, $\mathbb{N}^k$ is a wqo in the product ordering ($(x_1, \ldots, x_k) \leq (y_1, \ldots, y_k)$ iff $x_i \leq y_i$ for every i, $1 \leq i \leq k$): this is *Dickson's Lemma* (Dickson 1913). $\mathbb{N}^k$ is the set of configurations of Petri nets, or more generally, of counter machines.

For every well-quasi-ordered alphabet Σ, Σ^* with the embedding (a.k.a. scattered subword, a.k.a. divisibility) quasi-ordering is wqo: this is *Higman's Lemma* (Higman 1952). This is instrumental in the backward analysis of lossy channel systems (Abdulla and Jonsson 1993). Under the same assumptions, the collection of finite trees labeled with elements from Σ, with the tree embedding quasi-ordering, is wqo: this is *Kruskal's tree theorem* (Kruskal 1960).

A map f from a quasi-ordered set X to a quasi-ordered set Y is *monotonic* if and only if $x \leq x'$ implies $f(x) \leq f(x')$, for all $x, x' \in X$. (We write $\leq$ for the underlying ordering of any poset, unless mentioned otherwise.) We call it a *quasi-order embedding* if and only if $x \leq x'$ is *equivalent* to $f(x) \leq f(x')$. The *order embeddings* are the injective quasi-order embeddings; there is no difference between the two notions when X is a poset. An *order isomorphism* is a surjective (hence bijective) order embedding. Hence, $f : X \to Y$ is an order embedding if and only f is an order isomorphism onto its image.

Given any quasi-ordered set X, the *order quotient* of X is defined as the set of equivalence classes $[x]$ of elements $x \in X$ under $\approx$, quasi-ordered by letting $[x]$ be below $[y]$ iff $x \leq y$. (We then write $[x] \leq [y]$.) This is well defined and a partial order.

We shall say that a set is *well ordered* by $\leq$ iff it is well-quasi-ordered by $\leq$ and $\leq$ is an ordering. The well-ordered posets are exactly the order quotients of wqos.

In a quasi-ordered set X, an *upper bound* of a family $(x_i)_{i \in I}$ of points of X is an element $x \in X$ such that $x_i \leq x$ for every $i \in I$. A *least upper bound* is one that is less than or equal to all other upper bounds of the same family. If X is a poset, then the least upper bound of a family is unique if it exists at all.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

3.2 Domain theory

Domain theory is, prima facie, concerned with certain posets, called directed-complete partial orders (dcpos), where certain least upper bounds exist, and so-called Scott-continuous maps, which are not just monotonic but also preserve these least upper bounds. Over the years, domain theory has revealed itself as having firm grounds in general topology as well. Let us start with the order-theoretic view.

A *directed family* in a poset X is any non-empty family $(x_i)_{i\in I}$ such that, for all $i, j \in I$, there is a $k \in I$ with $x_i, x_j \leq x_k$. A *dcpo* is a poset X in which every directed family $(x_i)_{i\in I}$ of points of X has a least upper bound $\sup_{i\in I} x_i$.

A map f from a poset X to a poset Y is *Scott-continuous* if and only if it is monotonic and preserves least upper bounds of directed families, that is, if $(x_i)_{i\in I}$ is a directed family in X with least upper bound x, then $f(x)$ is the least upper bound of the (necessarily directed) family $(f(x_i))_{i\in I}$.

An element $x \in X$ is *finite* iff, for every directed family $(z_i)_{i\in I}$ that has a least upper bound $z \geq x$, then $z_i \geq x$ for some $i \in I$ already. The poset X is *algebraic* iff the family of finite elements below any given element x is directed and admits x as least upper bound. The finite elements are often much simpler to describe than arbitrary elements and act as approximants to the latter.

Let us give a few examples. The power $\mathbb{P}(X)$ of a set X is a dcpo, in fact a complete lattice, under inclusion $\subseteq$. Its finite elements, in the sense above, are the finite subsets of X, in the usual sense of the word, and $\mathbb{P}(X)$ is algebraic. We write $\mathbb{P}_{\text{fin}}(X)$ for the set of finite subsets of X, ordered by inclusion.

Neither $\mathbb{P}(X)$ nor $\mathbb{P}_{\text{fin}}(X)$ is wqo under inclusion, unless X is finite. $\mathbb{N}$, with its natural ordering, is an algebraic poset, which is also a wqo. $\mathbb{N}$ is not a dcpo, since $\mathbb{N}$ itself is a directed family without a least upper bound. However, $\mathbb{N}_\omega$, obtained by adjoining a new top element ω to $\mathbb{N}$, is a dcpo. Its finite elements are the elements of $\mathbb{N}$, and $\mathbb{N}_\omega$ is algebraic.

Some dcpos fail to be algebraic, for example, the only finite element of $[0, 1]$, with its natural ordering, is 0. However, $[0, 1]$ is continuous, in the following sense.

Define the *way below* relation $\ll$ on a poset X by $x \ll y$ iff, for every directed family $(z_i)_{i\in I}$ that has a least upper bound $z \geq y$, then $z_i \geq x$ for some $i \in I$ already. So, in particular, the finite elements are those that are way below themselves.

Note that $x \ll y$ implies $x \leq y$, and that $x' \leq x \ll y \leq y'$ implies $x' \ll y'$. However, $\ll$ is not reflexive or irreflexive in general. Write $\twoheaduparrow E = \{y \in X \mid \exists x \in E, x \ll y\}$, $\twoheaddownarrow E = \{y \in X \mid \exists x \in E, y \ll x\}$.

The poset X is *continuous* iff, for every $x \in X$, $\twoheaddownarrow x$ is a directed family, and has x as least upper bound. More finely, call a *basis* (not to be confused with the finite bases of upward-closed subsets of wqos) any subset B of X such that any element $x \in X$ is the least upper bound of a directed family of elements way below x *in B*. Then, X is continuous if and only if it has a basis, and in this case X itself is the largest basis. On the other hand, every algebraic poset is continuous and has a least basis, namely its set of finite elements.

An essential property of continuous posets is *interpolation* (Mislove 1998, Lemma 4.16): if $x \ll y$, then $x \ll z \ll y$ for some $z \in X$. We may even choose z to be in any prescribed basis B. For example, in $[0, 1]$, $x \ll y$ iff $x < y$ or $x = 0$, and we may choose B to be, say, the set of rational points in $[0, 1]$. Interpolation fails in general, non-continuous posets, even non-continuous dcpos.

Any finite product of dcpos is a dcpo, where product is taken in the order-theoretic sense, that is, with the product ordering. Then, any finite product of algebraic (resp., continuous) posets is again algebraic (resp., continuous).

Given a poset X, which might fail to be a dcpo, there is a canonical way to obtain a completion, called the *ideal completion* $\mathbf{I}(X)$ of X. An *ideal* I of X is any downward-closed set that is also directed. $\mathbf{I}(X)$ is defined as the poset of all ideals of X, ordered by inclusion. $\mathbf{I}(X)$ is then a dcpo, where directed suprema are computed as unions, and X order-embeds into $\mathbf{I}(X)$ through the function $\eta^{\mathbf{I}} \colon X \to \mathbf{I}(X)$ that maps x to $\downarrow x$. For example, $\mathbf{I}(\mathbb{N})$ consists of all the ideals $\downarrow n$, $n \in \mathbb{N}$, plus

Order	Topology
Upward-closed	Open
Downward-closed	Closed
Monotonic	Continuous
wqo	Noetherian
Ideal	Irreducible closed
Ideal completion $\mathbf{I}(X)$	Sobrification $\mathcal{S}(X)$

Figure 1. Informal, order versus topology glossary.

a fresh element above all others, which we write ω and, as an ideal, is just the whole of $\mathbb{N}$. In this sense, $\mathbf{I}(\mathbb{N})$ is the completion $\mathbb{N}_\omega$ we have already mentioned in the context of the Karp–Miller algorithm.

$\mathbf{I}(X)$ is the *free dcpo* over X, meaning that for every monotonic map f from X to a dcpo Y extends to a unique Scott-continuous map g from $\mathbf{I}(X)$ to Y – namely, $f = g \circ \eta^{\mathbf{I}}$, see Goubault-Larrecq (2013, Exercise 5.5.3, or comment pages 175–176). $\mathbf{I}(X)$ is also an algebraic dcpo (Goubault-Larrecq 2013, Proposition 5.1.46), with the elements of X forming a basis.

3.3 Topology

A *topology* $\mathcal{O}$ on a set X is a collection of subsets (the *opens*) of X that is closed under arbitrary unions and finite intersections. In particular, considering empty unions and empty intersections, both $\emptyset$ and X itself are open. We say that X itself is a topological space, leaving $\mathcal{O}$ implicit. The complements of opens are the *closed sets*. The largest open contained in A is its *interior* and the smallest closed subset $cl(A)$ containing it is its *closure*.

A famous topology in domain theory is the *Scott topology* on a poset X. Its opens, the *Scott opens*, are all upward-closed subsets U such that every directed family $(x_i)_{i\in I}$ that has a least upper bound x in U intersects U, that is, $x_i \in U$ for some $i \in I$. In other words, the closed subsets of the topology, namely the *Scott closed* subsets, are the downward-closed subsets F that are stable under taking least upper bounds of directed families of elements of F. The non-empty Scott closed subsets of $[0, 1]$ are the intervals $[0, t]$, $0 \leq t \leq 1$, and its Scott-open subsets are the half-open intervals $(t, 1]$, $0 \leq t \leq 1$, plus $[0, 1]$ itself.

A topology is *coarser* than another iff it contains less opens. Conversely, a topology if *finer* than another iff it contains more opens.

For example, consider the *Alexandroff topology* of a quasi-order X whose opens are all upward-closed subsets. This is finer than the Scott topology and in general strictly finer: on $[0, 1]$, $[1/2, 1]$ is Alexandroff open but not Scott-open. On $\mathbb{N}$, the Scott and Alexandroff topologies agree, and the non-empty opens are of the form $\uparrow n$, $n \in \mathbb{N}$. The *discrete topology* is the finest possible topology, where all subsets are open. Note that this is also the Alexandroff topology of the equality ordering.

The Alexandroff topology converts a quasi-order into a topological space and suggests a glossary of generalizations of order-theoretic notions as topological notions, see Figure 1. We have just explained the first row: the upward-closed subsets of a quasi-order are the opens in the Alexandroff topology. We shall explain the other rows below.

We shall write X_σ for X with its Scott topology, and X_a for X with its Alexandroff topology. It is easy to see that the downward-closed subsets of X are exactly the closed subsets of X_a, and we shall use this fact several times. This is the second row of Figure 1.

A map f from a topological space X to a topological space Y is *continuous* if and only if $f^{-1}(V)$ is open in X for every open subset V of Y. When both X and Y are posets equipped with the Alexandroff topology, a map $f: X \to Y$ is continuous if and only if it is monotonic. This is the third row of Figure 1. When X and Y are posets equipped with their Scott topology, then f is continuous if and only if it is Scott-continuous.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

A *homeomorphism* is a topological isomorphism, that is, a continuous, bijective map whose inverse is also continuous.

Given any collection $\mathcal{C}$ of subsets of a set X, there is a smallest (*coarsest*) topology containing all elements of $\mathcal{C}$. This is the topology *generated by* $\mathcal{C}$, $\mathcal{C}$ is then called a *subbase* for the topology, and the elements of $\mathcal{C}$ are *subbasic opens*. Their complements are the *subbasic closed subsets*. Any open in the topology is then a (possibly infinite) union of finite intersections of subbasic opens. If any open can be written as a union of elements of $\mathcal{C}$, then one says that $\mathcal{C}$ is a *base* of the topology, and the elements of $\mathcal{C}$ are *basic opens*. This occurs typically when $\mathcal{C}$ contains X and is closed under binary intersections.

In a continuous poset, $\twoheaduparrow x$ is Scott-open for all x, and every Scott-open set U is a union of such sets, viz. $U = \bigcup_{x \in U} \twoheaduparrow x$ (Abramsky and Jung 1994), that is , the subsets $\twoheaduparrow x$ form a base of the Scott topology. Note that the subsets $\uparrow x$ form a base of the Alexandroff topology instead.

Every topology comes with a *specialization quasi-ordering* $\leq$, defined as $x \leq y$ iff every open that contains x also contains y; equivalently, iff $x \in cl\{y\}$. It is easy to see that every open is upward-closed with respect to $\leq$. The converse need not hold. A subset A of X is *saturated* iff A equals the intersection of all opens U containing A, equivalently iff it is upward-closed with respect to $\leq$. The specialization quasi-ordering of both the Scott and Alexandroff topologies of a poset X ordered by $\leq$ is $\leq$ again.

In fact, the Alexandroff topology is the finest having this property. The coarsest is called the *upper topology*; its opens are arbitrary unions of complements of sets of the form $\downarrow E$, E finite. And the Scott topology is somewhere inbetween. The sets $\downarrow E$, with E finite, will play an important role and we call them the *finitary closed* subsets. These are closed in the upper, Scott, and Alexandroff topologies.

Paralleling the notations X_σ, X_a, we write X_u for X with its upper topology.

A topological space X is T_0 iff for any two distinct points $x, y \in X$, there is an open subset containing x but not y, or conversely. X is T_0 if and only if its specialization quasi-ordering $\leq$ is a partial ordering, that is, $x \leq y$ and $y \leq x$ imply $x = y$.

A *subspace* of a topological space X is a subset A of X with the so-called *subspace topology*, whose opens are $A \cap U$, U open in X.

The *product* $\prod_{i \in I} X_i$ of a family $(X_i)_{i \in I}$ of topological spaces is the space of tuples $\vec{x} = (x_i)_{i \in I}$ where each x_i is in X_i, and with the *product topology*. The latter is the coarsest that makes the projection maps $\pi_i \colon \vec{x} \mapsto x_i$ continuous. In other words, the sets $\pi_i^{-1}(U)$, $i \in I$, U open in X_i, form a subbase of the product topology. The binary product of X and Y is written $X \times Y$, and the open subsets of the product topology on the latter are the unions $\bigcup_{i \in I} U_i \times V_i$, where I is an arbitrary index set, U_i is open in X, and V_i is open in Y.

A *topological embedding* f of X into Y is a map from X to Y that is a homeomorphism of X onto the image $f[X] = \{f(x) \mid x \in X\}$ of f, seen as a subspace of Y. Equivalently, f is a topological embedding if and only if it is injective, continuous, and *almost open* in the sense that every open subset U of X is the inverse image $f^{-1}(V)$ of some open subset V of Y. A trivial example is the canonical injection $i \colon A \to X$ of a subspace A of X into X. Up to homeomorphism, these are the only topological embeddings: any topological embedding $f \colon X \to Y$ is by definition the composition of the canonical injection of $f[X]$ into Y with the homeomorphism $f \colon X \to f[X]$.

Given an equivalence relation $\equiv$ on a topological space X, we can form the *quotient space* $X/\equiv$. Its elements are the equivalence classes of elements of X modulo $\equiv$. The map $q \colon X \to X/\equiv$ sending every element to its equivalence class is called the *quotient map*, and $X/\equiv$ is then given the *quotient topology*, defined as the finest topology on $X/\equiv$ that makes q continuous. Explicitly, the opens of the quotient topology are exactly the subsets V of $X/\equiv$ such that $q^{-1}(V)$ is open in X.

A crucial notion in topology is compactness. A subset K of X is *compact* iff every *open cover* $(U_i)_{i \in I}$ (a family of opens U_i whose union contains K) contains a finite subcover. Alternatively, K is compact iff, for every directed family $(U_i)_{i \in I}$ of opens (directed with respect to inclusion) such that $K \subseteq \bigcup_{i \in I} U_i$, then $K \subseteq U_i$ for some $i \in I$ already.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

A topological space X is *Noetherian* iff every open subset of X is compact (Grothendieck 1960, chapitre 0, § 2). A less intimidating definition is that X is Noetherian if and only if its lattice of open subsets has the *ascending chain condition*: every properly ascending chain $U_0 \subsetneq U_1 \subsetneq \cdots \subsetneq U_n \subsetneq$ of opens must be finite (Goubault-Larrecq 2013, Proposition 9.7.6).

There is a strong link between Noetherian spaces and wqos: a poset X is wqo if and only if X is Noetherian in its Alexandroff topology (Goubault-Larrecq 2013, Proposition 9.7.17). So wqos are a special case of Noetherian spaces, yielding the fourth row of Figure 1. But there are more Noetherian spaces. We shall see a few of them in this paper, and we only mention two examples for now.

One of the simplest examples, although somehow artificial, is $\mathbb{N}$ with the *cofinite topology*, whose closed subsets are $\mathbb{N}$ plus all finite subsets of $\mathbb{N}$. This is Noetherian, because every properly descending chain of closed sets must be finite; by taking complements, every properly ascending chain of open sets is finite. If that were a wqo with the Alexandroff topology of some quasi-ordering, that quasi-ordering would have to be the specialization quasi-ordering of the space, which is equality. However, equality on $\mathbb{N}$ is not wqo, since $\mathbb{N}$ itself is an infinite antichain. In fact, the Alexandroff topology of $=$ is the discrete topology, which is much finer than the cofinite topology.

The primary example of a Noetherian space, $\mathbb{C}^k$ with its Zariski topology (Goubault-Larrecq 2013, Exercise 9.7.53), is far from arising from a wqo as well: its specialization quasi-ordering is equality $=$ again, and the whole space is an infinite antichain. This is one of the ingredients used in the study of the lossy concurrent polynomial games mentioned in the introduction.

3.4 Sobriety

For this section, we refer to Abramsky and Jung (1994, Section 7.2.1) or to Chapter 8 of Goubault-Larrecq (2013).

A closed subset C of a topological space is *irreducible* if and only if C is non-empty, and whenever $C \subseteq F_1 \cup F_2$ with F_1, F_2 closed, then $C \subseteq F_1$ or $C \subseteq F_2$. Equivalently, if C is included in a finite union of closed subsets $F_1, \ldots, F_n$ (whatever $n \in \mathbb{N}$), then $C \subseteq F_i$ for some i, $1 \leq i \leq n$.

The finitary closed subset $\downarrow x = cl(\{x\})$ ($x \in X$) is always irreducible. (When we write $\downarrow x$ in a topological space, this is relative to its specialization quasi-ordering.) A space X is *sober* iff every irreducible closed subset C is the closure of a unique point, that is, $C = \downarrow x$ for some unique x. Every sober space is T_0, and every continuous dcpo is sober in its Scott topology, see Abramsky and Jung (1994, Proposition 7.2.27) or Goubault-Larrecq (2013, Proposition 8.2.12 (*b*)).

Much as we could complete a poset X to a dcpo $\mathbf{I}(X)$, we can complete a topological space to its *sobrification* $\mathcal{S}(X)$. The elements of $\mathcal{S}(X)$ are the irreducible closed subsets of X. Its opens are the subsets of the form $\diamond U = \{C \in \mathcal{S}(X) \mid C \cap U \neq \emptyset\}$, U open in X. (This is a topology, not just a subbase.)

As an example, in a poset X with its Alexandroff topology, not only all sets of the form $\downarrow x$ are irreducible closed, but every ideal is irreducible closed, too. We let the reader check this, and also that the converse holds: the ideals of a poset X are exactly the irreducible closed subsets of X_a, leading to the fifth row of Figure 1. This goes much further: by *Hoffmann's theorem* (Hoffmann 1979b), for a poset X, the sobrification $\mathcal{S}(X_a)$ coincides with the ideal completion $\mathbf{I}(X)$ exactly (Goubault-Larrecq 2013, Fact 8.2.49). This means that the points are the same, but also the topologies, that is, the topology of $\mathcal{S}(X_a)$ is the Scott topology of $\mathbf{I}(X)$. This justifies the sixth row of Figure 1.

$\mathcal{S}(X)$ is always sober, and the map $\eta^{\mathcal{S}}_X : x \mapsto \downarrow x$ is a topological embedding of X inside $\mathcal{S}(X)$ as soon as X is T_0, that is , up to isomorphism, any T_0 space can be seen as a subspace of its sobrification $\mathcal{S}(X)$, equating $x \in X$ with $\downarrow x$ in $\mathcal{S}(X)$.

The sobrification $\mathcal{S}(X)$ of X can be thought of as X together with all missing limits from X. Note in particular that a sober space is always a dcpo in its specialization ordering, see Abramsky and Jung (1994, Proposition 7.2.13) or Goubault-Larrecq (2013, Corollary 8.2.23).

A topological space X is Noetherian if and only if $\mathcal{S}(X)$ is Noetherian (Goubault-Larrecq 2013, Lemma 9.7.9). This is clear from the fact that, up to natural order isomorphism, X and $\mathcal{S}(X)$ have the same opens, see Gierz et al. (2003, Proposition V-4.7(i)) or Goubault-Larrecq (2013, Lemma 8.2.26). Actually, $\mathcal{S}(X)$ is the *free sober space* above the topological space X, meaning that every continuous map f from X to a sober space Y extends to a unique continuous map g from $\mathcal{S}(X)$ to Y, in the sense that $f = g \circ \eta_X^{\mathcal{S}}$, see Gierz et al. (2003, Exercise V-4.9) or Goubault-Larrecq (2013, Theorem 8.2.44). This is a form of extension by continuity theorem and is the proper categorical way of saying that $\mathcal{S}(X)$ is X plus all missing limits.

$\mathcal{S}(X)$, as a space of specific closed subsets of X, embeds into the *Hoare powerdomain* $\mathcal{H}_{\mathsf{V}}(X)$, namely the space of all non-empty closed subsets of X. Let also $\mathcal{H}_{\mathsf{V}}(X)_\perp$ be the *lifted Hoare powerdomain* of X, which one can see either as $\mathcal{H}_{\mathsf{V}}(X)$ plus a fresh bottom element $\perp$ added, or as the set of all closed subsets of X, including the empty set. The topology of both $\mathcal{H}_{\mathsf{V}}(X)$ and $\mathcal{H}_{\mathsf{V}}(X)_\perp$ is the so-called *lower Vietoris topology* whose subbasic opens are $\Diamond U = \{F \in \mathcal{H}_{\mathsf{V}}(X) \mid F \cap U \neq \emptyset\}$, U open in X. With this topology, $\mathcal{S}(X)$ can be considered a subspace of $\mathcal{H}_{\mathsf{V}}(X)$ and the latter as a subspace of $\mathcal{H}_{\mathsf{V}}(X)_\perp$. We use a slightly different symbol $\Diamond U$ here, compared to the open subsets $\diamond U$ of $\mathcal{S}(X)$: although they denote very similar sets (and $\diamond U = \Diamond U \cap \mathcal{S}(X)$), the sets $\Diamond U$ only form a subbase of the lower Vietoris topology on $\mathcal{H}_{\mathsf{V}}(X)$ and $\mathcal{H}_{\mathsf{V}}(X)_\perp$, while the sets $\diamond U$ are *exactly* all the open subsets of $\mathcal{S}(X)$.

Remarkably, $\mathcal{H}_{\mathsf{V}}(X)$ and $\mathcal{H}_{\mathsf{V}}(X)_\perp$ are Noetherian for every Noetherian space X (Goubault-Larrecq 2013, Exercise 9.7.14), even though their specialization quasi-ordering, which is inclusion, is in general not wqo.

4. Completions of wqos

We have announced that the proper completion $\widehat{X}$ of a wqo, or more generally of a Noetherian space X, would be its ideal completion, or more generally its sobrification. Before we compute finite representations, it is in order to vindicate this choice.

A rational way to define a completion is to state the properties we need for it first and then derive what it should be. In our case, there are various properties we might want for a completion, depending on the point of view we take. In the conference version of this paper (Finkel and Goubault-Larrecq 2009), we had explored several of these points of view.

Let us concentrate on just one: Geeraerts et al.'s axiomatization of so-called *ADLs* for well-quasi-ordered sets X, used in their expand, enlarge, and check forward procedure (Geeraerts et al. 2006). We stress that this notion is independent of their algorithm, and of any particular algorithm: adequate domains of limits are merely an axiomatization of some basic requirements on the representability of downward-closed subsets. These requirements are also needed in our own approach (Finkel and Goubault-Larrecq 2012).

An *ADL* (Geeraerts et al. 2006) for a well-ordered set X is a triple $(L, \trianglelefteq, \gamma)$ where L is a set disjoint from X (the set of *limits*); (L_1) the map $\gamma : L \cup X \to \mathbb{P}(X)$ is such that $\gamma(z)$ is downward-closed for all $z \in L \cup X$, and $\gamma(x) = \downarrow_X x$ for all non-limit points $x \in X$; (L_2) there is a limit point $\top \in L$ such that $\gamma(\top) = X$; (L_3) for all $z, z' \in L \cup X$, $z \trianglelefteq z'$ if and only if $\gamma(z) \subseteq \gamma(z')$; and (L_4) for any downward-closed subset D of X, there is a finite subset $E \subseteq L \cup X$ such that $\widehat{\gamma}(E) = D$. Here $\widehat{\gamma}(E) = \bigcup_{z \in E} \gamma(z)$.

No explicit construction for such adequate domains of limits is given by Geeraerts et al., and they have to be found by trial and error. Our first result, below, is that there is a unique least (weak) ADL of X, and this is $\mathbf{I}(X) = \mathcal{S}(X_a)$ minus X. This not only gives a concrete construction of such an ADL but also shows that we do not have much freedom in defining one: any other one must contain $\mathcal{S}(X_a)$.

The definition of ADLs above is slightly awkward. Let us simplify it.

Requirement (L_2) in Geeraerts et al. (2006) only serves to ensure that all closed subsets of $L \cup X$ can be represented as $\downarrow_{L \cup X} E$ for some finite subset E: the closed subset $L \cup X$ itself is then exactly

$\downarrow_{L\cup X}\{\top\}$. However, ($L_2$) is unnecessary for this, since $L \cup X$ already equals $\downarrow_{L\cup X} E$ by (L_3), where E is the finite subset of $L \cup X$ such that $\widehat{\gamma}(E) = L \cup X$ as ensured by (L_4). We will not need (L_2) either in our own subsequent work (Finkel and Goubault-Larrecq 2012) and shall call *weak adequate domain of limits* (*WADLs*) any triple $(L, \trianglelefteq, \gamma)$ satisfying (L_1), (L_3), and (L_4).

Even so, this definition remains awkward. First, the real space of interest is not L, but $Z = L \cup X$; L can always be recovered as $Z \smallsetminus X$. Then $\gamma(z)$ should be downward-closed for every $z \in Z$, that is, it should be closed in X_a. The space of downward-closed subsets is the Hoare powerdomain $\mathcal{H}_\mathsf{V}(X_a)_\perp$. As a consequence, ($L_1$) can be expressed more succinctly by requiring that γ be a map from Z to $\mathcal{H}_\mathsf{V}(X_a)_\perp$, and that every subset of the form $\downarrow_X x$, $x \in X$, is obtained as $\gamma(x)$. Requirement (L_3) means that γ is a quasi-order embedding. In other words, the elements z of Z can be thought as *syntax* for particular elements $\gamma(z)$ of $\mathcal{H}_\mathsf{V}(X_a)_\perp$, and we define $\trianglelefteq$ on syntax by $z \trianglelefteq z'$ iff $\gamma(z) \subseteq \gamma(z')$. So we may safely omit $\trianglelefteq$ from the definition and remove requirement (L_3).

The only important requirement is (L_4), which states that every downward-closed subset of X should be describable as a finite union of representable subsets, that is, of elements of the form $\gamma(z)$, $z \in Z$. (L_1) also requires all elements of the form $\downarrow_X x$, $x \in X$, to be representable. However, this is a consequence of (L_4): $\downarrow_X x$ is a finite union of representable subsets $\gamma(z_1), \ldots, \gamma(z_n)$; then $x \in \gamma(z_i)$ for some i, $1 \leq i \leq n$, from which we deduce that $\gamma(z_i) = \downarrow_X x$.

We therefore arrive at the following definition.

Definition 4.1 (ADL, WADL). *Let X be a quasi-ordered set. A* WADL*, on X is a pair (Z, γ) of a set Z and a map $\gamma : Z \to \mathcal{H}_\mathsf{V}(X_a)_\perp$ (the* representation map*) such that every downward-closed subset of X is a finite union of representables. A* representable subset *of X is by definition one of the form $\gamma(z)$ for some $z \in Z$.*

(Z, γ) is an ADL *iff, additionally, the whole set X is representable.*

In any case, the limit points *of Z are those $z \in Z$ such that $\gamma(z)$ is not of the form $\downarrow_X x$, $x \in X$.*

We check the formal relationship with Geeraerts et al.'s conditions. The easy proof is left to the reader.

Lemma 4.2. *Let X be a quasi-ordered set.*

If $(L, \trianglelefteq, \gamma)$ satisfies (L_1), (L_3), and (L_4), then $(L \cup X, \gamma)$ is a WADL.

If $(L, \trianglelefteq, \gamma)$ satisfies (L_1), (L_2), (L_3), and (L_4), then $(L \cup X, \gamma)$ is an ADL.

Conversely, if (Z, γ) is a WADL (resp., ADL) on X, then $(L, \trianglelefteq, \gamma')$ satisfies (L_1), (resp., and (L_2)), (L_3), and (L_4) where L is the set of limit points of Z, $\trianglelefteq$ is defined by $z \trianglelefteq z'$ iff $\gamma'(z) \subseteq \gamma'(z')$, and γ' is defined by $\gamma'(z) = \gamma(z)$ if $z \in L$, $\gamma'(x) = \downarrow_X x$ if $x \in X$.

Definition 4.1 displays a tension between mathematical practice and computer science needs. That every downward-closed subset of X be a finite union of representables $\gamma(z_1), \ldots, \gamma(z_n)$ means that we can represent any downward-closed set by finitely many pieces of information $z_1, \ldots, z_n$. However, from a computer science perspective, we have not (yet) put any computability conditions on WADL. We repair this now.

Definition 4.3 (Effective WADL). *A WADL (Z, γ) on X is an* effective WADL *iff the relation $\trianglelefteq$ on Z, defined by $z \trianglelefteq z'$ iff $\gamma(z) \subseteq \gamma(z')$, is decidable.*

This naturally assumes that Z is a domain of objects representable on a computer, for example, a word, or a natural number.

From a mathematical standpoint, on the other hand, one usually reasons up to order quotients and order isomorphisms. Then γ and Z are useless in Definition 4.1, and the only relevant part of a WADL is the collection of representable subsets, that is, a WADL is, up to these details, a collection

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

of downward-closed subsets, the representable subsets, such that every downward-closed subset is a finite union of representables.

We can then bound precisely all WADLs between two well-known spaces of downward-closed subsets (i.e., closed in X_a).

Lemma 4.4. *Let X be a poset and* (Z, γ) *be a WADL on X. The set* $\gamma[Z]$ *of representable subsets is such that:*

$$\mathcal{S}(X_a) \subseteq \gamma[Z] \subseteq \mathcal{H}_{\mathsf{V}}(X_a)_{\perp}$$

Proof. We must show that $\mathcal{S}(X_a) \subseteq \gamma[Z]$, the other inclusion being by definition. Let $C \in \mathcal{S}(X_a)$, that is, assume C is irreducible closed. C must be a finite union of representables $\bigcup_{i=1}^{n} \gamma(z_i)$ by the definition of WADLs. So $C \subseteq \gamma(z_i)$ for some i, $1 \leq i \leq n$, by irreducibility (and since each $\gamma(z_i)$ is closed in X_a). It follows that $C = \gamma(z_i)$, hence $C \in \gamma[Z]$. □

So, up to order quotients and order isomorphisms, there cannot be any WADL smaller than the sobrification $\mathcal{S}(X_a)$. We shall see later that the latter is effective in a large number of practical cases.

Naturally, the statement of Lemma 4.4 does not require any topology. Purely order-theoretically, Lemma 4.4 states that the collection of representable subsets must lie between the collection of ideals ($\mathbf{I}(X) = \mathcal{S}(X_a)$) and the collection of all downward-closed subsets ($\mathcal{H}_{\mathsf{V}}(X_a)_{\perp}$).

The interest in using topology is in the proof of Lemma 4.4: the point is that the key notion is irreducibility, a topological notion. In turn, these notions and proofs will generalize to the topological, Noetherian case with no effort later.

When X is a wqo, the ideal completion $\mathbf{I}(X) = \mathcal{S}(X_a)$ is not just a lower bound below any WADL, it is itself a WADL. This follows from the following more general topological statement.

Proposition 4.5. *Let X be a Noetherian space. Then* $\mathcal{S}(X)$ *is the least collection* $\mathcal{C}$ *of closed subsets of X such that every closed subset of X can be expressed as a finite union of elements of* $\mathcal{C}$.

Proof. First, if $\mathcal{C}$ is as above, then $\mathcal{S}(X) \subseteq \mathcal{C}$. The proof is as in Lemma 4.4, which is in fact a topological proof: every element C of $\mathcal{S}(X_a)$ must be written as a finite union of elements of $\mathcal{C}$, and by irreducibility it must equal one of them.

Conversely, we need to show that every closed subset of X is a finite union of irreducible closed subsets, provided that X is Noetherian. This is a well-known fundamental result and occurs as part of Goubault-Larrecq (2013, Theorem 9.7.12). We give an elementary proof of it in Lemma 4.6 below, for the sake of completeness. □

Lemma 4.6. *In a Noetherian space, every closed subset is a finite union of irreducibles.*

Proof. By taking complements, in a Noetherian space X every properly descending chain $F_0 \supsetneq F_1 \supsetneq \cdots \supsetneq F_n \supsetneq \cdots$ of closed subsets must be finite, in other words $\mathcal{H}_{\mathsf{V}}(X_a)$, ordered by inclusion, is well founded. Imagine there were a closed subset C that cannot be written as a finite union of irreducibles. By well-foundedness, we can choose C minimal. C is not empty, since the empty set can be written as a finite union of irreducibles, namely none. C cannot be irreducible either, so there are two closed subsets F_1 and F_2 such that $C \subseteq F_1 \cup F_2$, but $C \not\subseteq F_1$ and $C \not\subseteq F_2$. Because of the latter, $C \cap F_1$ and $C \cap F_2$ are strictly smaller than C. By the minimality of C, $C \cap F_1$ and $C \cap F_2$ can be written as finite unions of irreducibles, so $C = (C \cap F_1) \cup (C \cap F_2)$ is also a finite union of irreducibles: contradiction. □

As a special case, we obtain that every downward-closed subset of a wqo X is a finite union of ideals. This can also be deduced from the observation by Erdős and Tarski (1943) that a poset has no infinite antichain if and only if every downward-closed subset is a finite union of ideals.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Remark 4.7. Generalizing the above cited result of Erdős and Tarski to the topological setting, we have the following: a topological space has no infinite discrete subspace if and only if every closed subset is a finite union of irreducibles (Goubault-Larrecq 2019).

Proposition 4.5, once toned down to wqos, translates to the following.

Corollary 4.8 (Least WADL). *Let X be a wqo. The ideal completion* $\mathbf{I}(X) = \mathcal{S}(X_a)$ *is the least WADL, in the sense that:*

(1) for any WADL (Z, γ) *on X, every element of* $\mathcal{S}(X_a)$ *is representable;*
(2) $(\mathcal{S}(X_a), i)$ *is itself a WADL, where i is the canonical injection of* $\mathcal{S}(X_a)$ *into* $\mathcal{H}_\vee(X_a)_\perp$*.*

In other words, up to the coding function γ, there is a unique *minimal* WADL on any given wqo X. We contend that $\mathcal{S}(X_a)$ is, in all practical cases, the sole WADL worth considering and will in particular be effective.

Our treatment so far uses topology for no particular good reason apart from mathematical elegance. Our presentation, however, lends itself to the following natural topological extension of WADLs. We have claimed that the additional generality obtained by shifting focus from wqos to the larger class of Noetherian spaces was useful in Goubault-Larrecq (2010). Notably, the class of polynomial concurrent programs introduced there is naturally seen as a *topological* WSTS, that is, as a pair $(X, \to)$ where X is Noetherian space (instead of a wqo) and $\to$ is a lower semi-continuous binary relation – this is the natural generalization of WSTS to a topological setting. Using Noetherianness, and algorithms and proof arguments that are variants of WSTS arguments, it was shown in that same paper that the reachability of sets of states defined by so-called forbidden patterns is decidable for polynomial concurrent programs. Note that the latter are *not* WSTS. In that context, studying topological WADLs instead of WADLs is the logical next step.

Definition 4.9. *Let X be a topological space. A* topological WADL *on X is a pair* (Z, γ) *of a set Z and a map* $\gamma : Z \to \mathcal{H}_\vee(X)_\perp$ *(the* representation map*) such that every closed subset of X is a finite union of representables. A* representable *subset of X is by definition one of the form* $\gamma(z)$ *for some* $z \in Z$*.*

So WADLs are topological WADLs, in the special case where X comes with the Alexandroff topology of some quasi-ordering. We have just seen (Proposition 4.5) that, when X is Noetherian, $\mathcal{S}(X)$ is the least topological WADL. We state it as follows.

Proposition 4.10 (Least topological WADL). *Let X be a Noetherian space. The sobrification* $\mathcal{S}(X)$ *is the least topological WADL, in the sense that:*

(1) for any topological WADL (Z, γ) *on X, every element of* $\mathcal{S}(X)$ *is representable;*
(2) $(\mathcal{S}(X), i)$ *is itself a topological WADL, where i is the canonical injection of* $\mathcal{S}(X)$ *into* $\mathcal{H}_\vee(X)_\perp$*.*

5. S-representations

We shall devote the rest of this paper to describe completions $\widehat{D} = \mathcal{S}(D)$ for those datatypes of Figure 2. As we shall see, these datatypes include most of the datatypes encountered in the literature on WSTS (e.g., Petri nets and more generally counter machines, lossy channel systems, data nets) and contain several new ones.

All the datatypes in this figure are Noetherian spaces, as can be gathered from Section 9.7 of Goubault-Larrecq (2013). We also state the relevant theorem is in each case. Stars indicate constructs that, while preserving Noetherianness, do not preserve well-quasi-orderedness. What

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

$D ::=$	A	(finite poset; Theorem 5.3)	
$\mid$	$\mathbb{N}$	(natural numbers; Theorem 5.4)	
$\mid$	$D_1 \times D_2 \times \cdots \times D_n$	(product; Theorem 5.5)	
$\mid$	$D_1 + D_2 + \cdots + D_n$	(coproduct; Theorem 5.6)	
$\mid$	$\mathcal{S}(D)$	(sobrification; Theorem 5.7)	$*$
$\mid$	$\mathbb{P}(D)$	(powerset; Theorem 5.11)	$*$
$\mid$	$\mathbb{P}^*(D)$	(non-empty powerset; Theorem 5.11)	$*$
$\mid$	$\mathcal{H}_{\mathsf{V}}(D)$	(Hoare powerdomain; Theorem 5.8)	$*$
$\mid$	$\mathcal{H}_{\mathsf{V}}(D)_\perp$	(lifted Hoare powerdomain; Theorem 5.8)	$*$
$\mid$	$\mathrm{Spec}(R)$	(spectrum of ring R; Proposition 6.1)	$*$
$\mid$	D^*	(finite words; Theorem 7.15)	
$\mid$	$D^{\circledast}$	(finite multisets; Theorem 8.7)	
$\mid$	$\rhd_{n=1}^{+\infty} D_n$	(words, prefix; Theorem 9.10)	$*$
$\mid$	$\mathcal{T}(D)$	(finite trees; Theorem 11.36)	

Figure 2. An algebra of Noetherian datatypes.

the values of these types are and what their topologies are (and associated specialization quasi-orderings) will also be made precise in each corresponding section.

The completion process is modular: the completion $\widehat{D}$ of a type D, built from $D_1, \ldots, D_n$, will be defined as a function of $\widehat{D}_1, \ldots, \widehat{D}_n$. In each case, we shall show that if $\widehat{D}_1, \ldots, \widehat{D}_n$ are effective, then so is $\widehat{D}$.

As a result, all the datatypes defined in Figure 2 will be effective. This is important: Definition 4.3, applied to the WADL $\widehat{X} = \mathcal{S}(X_a)$ (when X is wqo), requires us to decide the ordering (i.e., inclusion) on $\mathcal{S}(X)$. We shall require – and obtain – more: we shall be able to compute finite intersections of closed subsets (i.e., downward-closed subsets in wqos) as well.

We consider topological WADLs – for example, the starred rows in Figure 2 – for added generality, but also because the topological approach, relying on the notion of irreducibility, provides a unifying perspective on the matter. This leads to the following notion of an effective, finite representation of irreducible closed subsets. The closed, not necessarily irreducible, subsets are all finite unions of irreducibles (Lemma 4.6) and can therefore be represented as finite sets of codes. Below, this is how we represent the closed sets X (item D) and $[\![a]\!] \cap [\![b]\!]$ (item E).

Definition 5.1 (S-representation). *Let X be a topological space. An* S-representation *of X is a tuple $(S, [\![\_]\!], \unlhd, \tau, \wedge)$ where:*

(A) S is a recursively enumerable set of so-called codes *(of irreducible closed subsets);*
(B) $[\![\_]\!]$ is a surjective map from S to $\mathcal{S}(X)$;
(C) $\unlhd$ is a decidable relation such that, for all codes $a, b \in S$, $a \unlhd b$ iff $[\![a]\!] \leq [\![b]\!]$;
(D) τ is a finite subset of S, such that $X = \bigcup_{a \in \tau} [\![a]\!]$;
(E) $\wedge$ is a computable map from $S \times S$ to the collection $\mathbb{P}_{fin}(S)$ of finite subsets of S (and we write $a \wedge b$ for $\wedge(a, b)$) such that $[\![a]\!] \cap [\![b]\!] = \bigcup_{c \in a \wedge b} [\![c]\!]$.

We call $\wedge$ the intersection map.

The idea is that codes represent irreducible closed subsets, through the semantic function $[\![\_]\!]$, that $\unlhd$ implements inclusion, τ denotes the whole set X, and $\wedge$ implements intersection.

We justify this now, in a more precise way. We represent closed subsets F through finite sets $\{a_1, \ldots, a_m\}$ of codes. The denotation of such a finite set is the union $\bigcup_{i=1}^m [\![a_i]\!]$. Since $[\![\_]\!]$ is surjective, and using Lemma 4.6, *every* closed subset of a Noetherian space X can be represented

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

this way. This is in particular true for the whole set X (item D) and the intersection $[\![a]\!] \cap [\![b]\!]$ in item E.

While $\trianglelefteq$ allows us to test two (codes of) irreducible closed subsets for inclusion, one can extend the inclusion test to arbitrary closed subsets: this is what we show now.

Lemma 5.2. *Given irreducible closed subsets $C_1, \ldots, C_m, C'_1, \ldots, C'_n$ of a topological space X, the following are equivalent:*

- $C_1 \cup \cdots \cup C_m \subseteq C'_1 \cup \cdots \cup C'_n$;
- $\{C_1, \ldots, C_m\} \subseteq^\flat \{C'_1, \ldots, C'_n\}$, *that is, for every i $(1 \leq i \leq m)$, there is a j $(1 \leq j \leq n)$ with $C_i \subseteq C'_j$.*

Proof. $C_1 \cup \cdots \cup C_m \subseteq C'_1 \cup \cdots \cup C'_n$ if and only if for every i, C_i is included in $C'_1 \cup \cdots \cup C'_n$. Since C_i is irreducible, the latter is equivalent to the existence of j such that $C_i \subseteq C'_j$. □

In general, $\leq^\flat$ is the *Hoare quasi-ordering* on subsets, also called the *domination quasi-ordering*: $A \leq^\flat B$ iff for every $a \in A$, there is a $b \in B$ such that $a \leq b$. We will use $\leq^\flat$ for various quasi-orderings $\leq$ and will accordingly use the notations $\subseteq^\flat$ as above, or $\trianglelefteq^\flat$ later.

Given an S-representation, we can then test two closed sets for inclusion: given two finite sets $\{a_1, \ldots, a_m\}$ and $\{b_1, \ldots, b_n\}$ of codes, $\bigcup_{i=1}^m [\![a_i]\!]$ is included in $\bigcup_{j=1}^n [\![b_j]\!]$ iff for every i, there is a j such that $a_i \trianglelefteq b_j$.

Finite intersections are computable, too, using $\wedge$: the intersection of two closed sets represented by finite sets $\{a_1, \ldots, a_m\}$ and $\{b_1, \ldots, b_n\}$ of codes is $\bigcup_{i=1}^m [\![a_i]\!] \cap \bigcup_{j=1}^n [\![b_j]\!] = \bigcup_{i,j} [\![a_i]\!] \cap [\![b_j]\!]$ and is therefore represented by the finite set $\bigcup_{i,j} a_i \wedge b_j$. Finite unions are, of course, easily computable as well.

Our purpose is to show that every space X that occurs as the space of values of some type D in Figure 2 has an S-representation. In each case, we will actually define an S-representation of D as a function of given S-representation of its constituent datatypes, and we shall use a uniform presentation: each result will be given in the form of a proposition of the following shape.

"**Proposition XXX** (S-representation, datatype D) Let X_i be Noetherian spaces, and $(S_i, [\![\_]\!]_i, \trianglelefteq_i, \tau_i, \wedge_i)$ be an S-representation of D_i for each i. Then $(S', [\![\_]\!]', \trianglelefteq', \tau', \wedge')$ is an S-representation of D, where:

(A) S' is …
(B) $[\![\cdots]\!]'$ is defined as …
(C) $\trianglelefteq'$ is defined as …
(D) τ' is defined as …
(E) $\wedge'$ is defined by $a' \wedge' b' = \ldots$"

We start with the easiest cases. The more difficult cases will be dealt with in separate sections. Our first instance is trivial: in a finite quasi-ordered set A, irreducible closed subsets, that is, ideals, are all of the form $\downarrow x$, $x \in A$, so $\widehat{A} = \mathcal{S}(X_a) = \mathbf{I}(X)$ is isomorphic to A.

Theorem 5.3 (S-representation, finite quasi-orders). *Let A be any finite quasi-ordered set. An S-representation of A is $(S, [\![\_]\!], \trianglelefteq, \tau, \wedge)$ where:*

(A) S is A itself,
(B) $[\![\_]\!]$ is the identity map,
(C) $\trianglelefteq$ is the given ordering on A,
(D) τ is the set of maximal elements of A,
(E) $a \wedge b$ is the set of maximal lower bounds of a and b.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

All this is computable, by just maintaining all needed information in tables. As a particular case, one finds the finite *sets*: these are the posets whose ordering is equality. In particular, the above provides an S-representation for finite sets Q of control states of various kinds of machines. In this case, $\tau = Q$, $a \wedge b = \{a\}$ if $a = b$, and $a \wedge b = \emptyset$ otherwise.

The next case is an easy exercise.

Theorem 5.4 (S-representation, $\mathbb{N}$). *An S-representation of $\mathbb{N}$ is $(S, [\![\_]\!], \trianglelefteq, \tau, \wedge)$ where:*

(A) $S = \mathbb{N}_\omega$,
(B) $[\![\_]\!]$ *maps* n *to* $\downarrow n$ *and* ω *to* $\mathbb{N}$,
(C) $\trianglelefteq$ *is the usual ordering on* $\mathbb{N}_\omega$,
(D) $\tau = \{\omega\}$,
(E) $m \wedge n = \{\min(m, n)\}$.

Theorem 5.5 (S-representation, products). *Let* $X_1, \ldots, X_n$ *be* n *Noetherian spaces,* $X = X_1 \times \cdots \times X_n$, *and* $(S_i, [\![\_]\!]_i, \trianglelefteq_i, \tau_i, \wedge_i)$ *an S-representation of* X_i *for each* i, $1 \le i \le n$. *Then* $(S, [\![\_]\!], \trianglelefteq, \tau, \wedge)$ *is an S-representation of* X, *where:*

(A) $S = S_1 \times \cdots \times S_n$;
(B) $[\![(a_1, \ldots, a_n)]\!] = [\![a_1]\!] \times \cdots \times [\![a_n]\!]$;
(C) $(a_1, \ldots, a_n) \trianglelefteq (b_1, \ldots, b_n)$ *iff* $a_1 \trianglelefteq_1 b_1$ *and ... and* $a_n \trianglelefteq_n b_n$;
(D) $\tau = \tau_1 \times \cdots \times \tau_n$;
(E) $(a_1, \ldots, a_n) \wedge (b_1, \ldots, b_n) = (a_1 \wedge_1 b_1) \times \cdots \times (a_n \wedge_n b_n)$.

Proof. The elements of $\mathcal{S}(X)$ are the products $C_1 \times \cdots \times C_n$ of irreducible closed subsets C_1 of X_1, ..., C_n of X_n (see Lemma A.2 in the Appendix), which justifies items A and B: we represent $C_1 \times \cdots \times C_n$ as the n-tuple of codes for $C_1, C_2, \ldots, C_n$. The if direction of item C follows from the fact that product is monotonic with respect to inclusion. Conversely, if $\prod_{i=1}^n [\![a_i]\!] \subseteq \prod_{i=1}^n [\![b_i]\!]$, then $[\![a_i]\!] \subseteq [\![b_i]\!]$ for every i: since $[\![a_j]\!]$ is non-empty, we can pick an element x_j from $[\![a_j]\!]$ for every $j \neq i$; then, for every $x \in [\![a_i]\!]$, the tuple $(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n)$ is in $\prod_{i=1}^n [\![a_i]\!]$, hence in $\prod_{i=1}^n [\![b_i]\!]$, showing that x is in $[\![b_i]\!]$. Items D and E are clear. □

So, for example, an S-representation for $\mathbb{N}^k$, the datatype of configurations of Petri nets, and more generally, of counter machines, is as expected: $S = \mathbb{N}_\omega^k$, $(m_1, \ldots, m_k) \trianglelefteq (n_1, \ldots, n_k)$ iff $m_i \le n_i$ for every i, $1 \le i \le k$, $\tau = \{(\omega, \ldots, \omega)\}$, and $(m_1, \ldots, m_k) \wedge (n_1, \ldots, n_k) = \{\min(m_1, n_1), \ldots, \min(m_k, n_k))\}$.

Theorem 5.6 (S-representation, coproducts). *Let* $X_1, \ldots, X_n$ *be* n *Noetherian spaces, and* $X = X_1 + \cdots + X_n$. *Then* $\mathcal{S}(X)$ *is homeomorphic to* $\mathcal{S}(X_1) + \cdots + \mathcal{S}(X_n)$.

Let $(S_i, [\![\_]\!]_i, \trianglelefteq_i, \tau_i, \wedge_i)$ *be an S-representation of* X_i *for each* i, $1 \le i \le n$. *Then* $(S, [\![\_]\!], \trianglelefteq, \tau, \wedge)$ *is an S-representation of* $X = X_1 + \cdots + X_n$, *where:*

(A) $S = \{(i, a) \mid 1 \le i \le n, a \in S_i\}$;
(B) $[\![(i, a)]\!] = [\![a]\!]_i$ *(up to the homeomorphism between* $\mathcal{S}(X)$ *and* $\mathcal{S}(X_1) + \cdots + \mathcal{S}(X_n)$*)*;
(C) $(i, a) \trianglelefteq (j, b)$ *iff* $i = j$ *and* $a \trianglelefteq_i b$;
(D) $\tau = \bigcup_{i=1}^n \{i\} \times \tau_i$;
(E) $(i, a) \wedge (j, b) = \emptyset$ *if* $i \neq j$, $(i, a) \wedge (i, b) = \{(i, c) \mid c \in a \wedge_i b\}$.

Proof. For the first part, see Goubault-Larrecq (2013, Fact 8.4.3), which states that $\mathcal{S}$ commutes with coproducts (in fact with all colimits, since $\mathcal{S}$ is a left adjoint). The rest is clear. □

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

A trivial case of S-representation is provided by sobrifications themselves, because $\mathcal{S}(\mathcal{S}(X))$ is canonically isomorphic to $\mathcal{S}(X)$. Indeed, $Y = \mathcal{S}(X)$ is sober, and for every sober space Y, $\eta^{\mathcal{S}}$ is an isomorphism between Y and $\mathcal{S}(Y)$ (Goubault-Larrecq 2013, Fact 8.2.5). It follows that any S-representation $(S, [\![\_]\!], \trianglelefteq, \tau, \wedge)$ for X yields an S-representation for $\mathcal{S}(X)$ with the same set S of codes and the same operations $\trianglelefteq$, τ, and $\wedge$, namely $(S, \eta^{\mathcal{S}}_X \circ [\![\_]\!], \trianglelefteq, \tau, \wedge)$:

Theorem 5.7 (S-representation, sobrifications). *Let X be a Noetherian space, and $X' = \mathcal{S}(X)$. Let $(S, [\![\_]\!], \trianglelefteq, \tau, \wedge)$ be an S-representation of X. Then $(S', [\![\_]\!]', \trianglelefteq', \tau', \wedge')$ is an S-representation of X' where:*

- *(A) $S' = S$;*
- *(B) for every $a \in S$, $[\![a]\!]' = \downarrow_{X'} [\![a]\!]$;*
- *(C) $a \trianglelefteq' b$ iff $a \trianglelefteq b$;*
- *(D) $\tau' = \tau$;*
- *(E) $a \wedge' b = a \wedge b$.*

Let us deal with the Hoare powerdomain $\mathcal{H}_{\mathsf{V}}(X)$ of X and its lifted version.

Theorem 5.8 (S-representation, Hoare powerdomains). *Let X be a Noetherian space, and $X' = \mathcal{H}_{\mathsf{V}}(X)_\perp$ (resp., $X' = \mathcal{H}_{\mathsf{V}}(X)$). Let $(S, [\![\_]\!], \trianglelefteq, \tau, \wedge)$ be an S-representation of X. Then $(S', [\![\_]\!]', \trianglelefteq', \tau', \wedge')$ is an S-representation of X' where:*

- *(A) $S' = \mathbb{P}_{fin}(S)$ (resp., $S' = \mathbb{P}^*_{fin}(S)$);*
- *(B) for every $a' \in S'$, $[\![a']\!]' = \downarrow_{X'} \{\bigcup_{a \in a'} [\![a]\!]\}$;*
- *(C) $a' \trianglelefteq' b'$ iff $a' \trianglelefteq^\flat b'$, where $a' \trianglelefteq^\flat b'$ iff for every $a \in a'$, there is a $b \in b'$ such that $a \trianglelefteq b$ (compare Lemma 5.2);*
- *(D) $\tau' = \{\tau\}$;*
- *(E) $a' \wedge' b' = \{\bigcup_{a \in a', b \in b'} (a \wedge b)\}$.*

Proof. By Lemma 4.6, every element F of $\mathcal{H}_{\mathsf{V}}(X)_\perp$ is a finite union of irreducible closed subsets, which are each of the form $[\![a]\!]$ with $a \in S$ by assumption, since $[\![\_]\!]$ is surjective. So $[\![\_]\!]'$ is surjective.

Next, $a' \trianglelefteq' b'$ iff $[\![a']\!]' \subseteq [\![b']\!]'$, iff $\bigcup_{a \in a'} [\![a]\!] \subseteq \bigcup_{b \in b'} [\![b]\!]$, iff $a' \trianglelefteq^\flat b'$, by Lemma 5.2.

We must check that $X' = \bigcup_{a' \in \tau'} [\![a']\!]'$. The right-hand side is $[\![\tau]\!]' = \downarrow_{X'} \{\bigcup_{a \in \tau} [\![a]\!]\} = \downarrow_{X'} \{X\} = X'$.

Finally, let us compute $[\![a']\!]' \cap [\![b']\!]'$. This is $\downarrow_{X'} \{\bigcup_{a \in a'} [\![a]\!]\} \cap \downarrow_{X'} \{\bigcup_{b \in b'} [\![b]\!]\} = \downarrow_{X'} \{\bigcup_{a \in a'} [\![a]\!] \cap \bigcup_{b \in b'} [\![b]\!]\} = \downarrow_{X'} \{\bigcup_{a \in a', b \in b'} ([\![a]\!] \cap [\![b]\!])\} = \downarrow_{X'} \{\bigcup_{a \in a', b \in b'} \bigcup_{c \in a \wedge b} [\![c]\!]\} = \downarrow_{X'} \{\bigcup_{c \in \bigcup_{a \in a', b \in b'} (a \wedge b)} [\![c]\!]\} = [\![\bigcup_{a \in a', b \in b'} (a \wedge b)]\!]'$. ☐

Let $\mathbb{P}^*(X)$ be the set of non-empty subsets of X. We topologize $X' = \mathbb{P}(X)$ (resp., $X' = \mathbb{P}^*(X)$) by the lower Vietoris topology, generated by the subbasic opens $\Diamond U = \{A \in X' \mid A \cap U \neq \emptyset\}$, where U ranges over the open subsets of X. (Although there is a risk of confusion with the lower Vietoris topology on $\mathcal{H}_{\mathsf{V}}(X)_\perp$, we shall see that the two are strongly tied.)

It is worth to point out that A is below B in the specialization quasi-ordering of those spaces if and only if $cl(A) \subseteq cl(B)$. This is well known and appears for example as part of Proposition 7.3 of Goubault-Larrecq (2007). We include the proof for completeness.

Lemma 5.9. *The specialization quasi-ordering of $\mathbb{P}(X)$, resp. $\mathbb{P}^*(X)$, is inclusion of closures.*

Proof. Let us temporarily write $\preceq$ for that specialization quasi-ordering.

If $A \preceq B$ in $\mathbb{P}(X)$ (resp., $\mathbb{P}^*(X)$), then consider the (open) complement U of $cl(B)$. Since $A \preceq B$, if $A \in \Diamond U$ then $B \in \Diamond U$. However, B does not intersect U, since U is the complement of $cl(B)$, so

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

B is not in $\Diamond U$. It follows that A is not in $\Diamond U$ either. This means that A does not intersect U, and therefore that it is included in its complement, $cl(B)$. Since $cl(B)$ is closed, contains A, and $cl(A)$ is by definition the smallest closed subset of X containing A, $cl(A)$ is included in $cl(B)$.

In the converse direction, we use the standard fact that, for an open subset U of X, $cl(A)$ intersects U if and only if A intersects U. Let us assume that $cl(A) \subseteq cl(B)$. Let $\mathcal{U} = \bigcup_{i \in I} \bigcap_{j \in J_i} \Diamond U_{ij}$ (where each J_i is finite) be any open subset of $\mathbb{P}(X)$ (resp., $\mathbb{P}^*(X)$) containing A. Then, for some $i \in I$, A intersects U_{ij} for every $j \in J_i$. The larger set $cl(B)$ must then intersect each U_{ij} as well. Hence, (see standard fact) B also intersects each U_{ij}. It follows that B is in $\mathcal{U}$. Since $\mathcal{U}$ is arbitrary, $A \preceq B$. □

When X is equipped with the Alexandroff topology of a quasi-ordering $\leq$, $cl(A) = \downarrow A$ and $cl(B) = \downarrow B$, and therefore the specialization quasi-ordering of $\mathbb{P}(X)$ and $\mathbb{P}^*(X)$ is the familiar Hoare quasi-ordering $\leq^\flat$ in that case.

Even when $\leq$ is wqo, $\leq^\flat$ fails to be wqo in general. However, $\mathbb{P}(X)$ and $\mathbb{P}^*(X)$ are Noetherian for X Noetherian, as first remarked in Goubault-Larrecq (2007). This stems from the following result, and the fact that sobrifications of Noetherian spaces are Noetherian.

Lemma 5.10. *For a topological space X, $\mathcal{H}_\vee(X)_\perp$ is the sobrification of $\mathbb{P}(X)$, and $\mathcal{H}_\vee(X)$ is the sobrification of $\mathbb{P}^*(X)$, up to homeomorphism.*

Proof. We deal with the first claim, by exhibiting a homeomorphism $\Box$ between $\mathcal{H}_\vee(X)_\perp$ and $\mathcal{S}(\mathbb{P}(X))$. To distinguish closure in X and closure in $\mathbb{P}(X)$, let us write cl_X for the former and $cl_{\mathbb{P}(X)}$ for the latter.

For every closed subset C of X, $cl_{\mathbb{P}(X)}(\{C\})$ is the closure of the point C in $\mathbb{P}(X)$, and that is equal to the downward closure of $\{C\}$ with respect to the specialization quasi-ordering of $\mathbb{P}(X)$. By Lemma 5.9, this is the set of subsets A of X such that $cl_X(A) \subseteq cl_X(C)$. Since $cl_X(C) = C$, and since $cl_X(A) \subseteq C$ is equivalent to $A \subseteq C$ (since C is closed in X), $cl_{\mathbb{P}(X)}(C)$ is therefore just the set $\Box C$ of subsets of C.

The notation $\Box C$ is justified by the fact that it is the complement of $\Diamond U$ where U is the complement of C. (Admittedly, we could also have written it as $\downarrow C$.) Since $\Box C$ is equal to $cl_{\mathbb{P}(X)}(\{C\})$, it is in particular irreducible closed. This defines a map $\Box : \mathcal{H}_\vee(X)_\perp \to \mathcal{S}(\mathbb{P}(X))$.

Conversely, let $\mathcal{I}$ be an irreducible closed subset of $\mathbb{P}(X)$. As a closed set, we can write it as $\bigcap_{i \in I} \bigcup_{j \in J_i} \Box C_{ij}$, where each J_i is finite and each C_{ij} is closed. For each $i \in I$, $\mathcal{I} \subseteq \bigcup_{j \in J_i} \Box C_{ij}$, and since $\mathcal{I}$ is irreducible, there is a $j_i \in J_i$ such that $\mathcal{I} \subseteq \Box C_{ij_i}$. Therefore $\mathcal{I} \subseteq \bigcap_{i \in I} \Box C_{ij_i}$, and as the right-hand side is clearly included in $\mathcal{I}$, this inequality is in fact an equality. It is easy to see that $\Box$ commutes with arbitrary intersections. As a consequence, $\mathcal{I}$ is of the form $\Box C$, where $C = \bigcap_{i \in I} C_{ij_i}$. It follows that $\Box$ is surjective.

Note furthermore that C is unique: if $\mathcal{I} = \Box C$, then C is necessarily the largest closed set that is an element of $\mathcal{I}$. Hence, the map $\Box : \mathcal{H}_\vee(X)_\perp \to \mathcal{S}(\mathbb{P}(X))$ is bijective.

To show that a map is continuous, it is enough to show that the inverse image of a subbasic open is open. A subbase of opens of $\mathcal{S}(\mathbb{P}(X))$ consists of the sets of the form $\diamond\Diamond U$, U open in X, because the outer $\diamond$ commutes with all unions and finite intersections. (Both $\diamond$ and $\Diamond$ commute with unions; we let the reader check that the outer $\diamond$ also commutes with finite intersections, as a consequence of irreducibility.)

To show that $\Box$ is continuous, it therefore suffices to show that $\Box^{-1}(\diamond\Diamond U)$ is open for every open subset U of X. For every C in $\mathcal{H}_\vee(X)_\perp$, $C \in \Box^{-1}(\diamond\Diamond U)$ if and only if $\Box C$ intersects $\Diamond U$, if and only if some closed subset C' of C intersects U, if and only if C itself intersects U. So $\Box^{-1}(\diamond\Diamond U) = \Diamond U$, showing that $\Box$ is continuous.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Conversely, the inverse image of $\Diamond U$ by the inverse map $\Box^{-1}$ is $\diamond\Diamond U$, showing that $\Box^{-1}$, too, is continuous. Therefore $\Box$ is a homeomorphism. $\Box$

This not only shows that $\mathbb{P}(X)$ and $\mathbb{P}^*(X)$ are Noetherian for X Noetherian, but also that they have the same sobrifications as $\mathcal{H}_\mathsf{V}(X)_\perp$ and $\mathcal{H}_\mathsf{V}(X)$, respectively, hence can be given the same S-representations:

Theorem 5.11 (S-representation, powersets). *Let X be a Noetherian space, and $X' = \mathbb{P}(X)$ (resp., $X' = \mathbb{P}^*(X)$). Let $(S, [\![\_]\!], \trianglelefteq, \tau, \wedge)$ be an S-representation of X. Then, $(S', [\![\_]\!]', \trianglelefteq', \tau', \wedge')$ is an S-representation of X' where:*

(A) $S' = \mathbb{P}_{fin}(S)$ (resp., $S' = \mathbb{P}^*_{fin}(S)$);
(B) for every $a' \in S'$, $[\![a']\!]' = \downarrow_{X'}\{\bigcup_{a \in a'} [\![a]\!]\}$;
(C) $a' \trianglelefteq' b'$ iff $a' \trianglelefteq^\flat b'$;
(D) $\tau' = \{\tau\}$;
(E) $a' \wedge' b' = \{\bigcup_{a \in a', b \in b'} (a \wedge b)\}$.

6. Completing Ring Ideals

The primary example of Noetherian spaces, historically, are the spectra of Noetherian rings. Mentioning them is therefore mandatory. The non-algebraically inclined reader is invited to proceed to finite words (Section 7).

Let R be a commutative ring (with unit). Recall that an *ideal* I is any additive subgroup of R such that for any $r \in I$, $r' \in R$, the product rr' is in I. A *prime ideal* p is an ideal that does not contain the multiplicative unit 1 of R (equivalently, which is different from the whole of R), and such that whenever $rr' \in p$, then r or r' is in p. The *spectrum* $\mathrm{Spec}(R)$ of R is the set of all prime ideals of R. It is equipped with the *Zariski topology*, whose closed subsets are $F_I = \{p \in \mathrm{Spec}(R) \mid I \subseteq p\}$, where I ranges over the ideals of R.

Union and intersection is computed on such sets by $F_I \cap F_{I'} = F_{I+I'}$, where $I + I' = \{r + r' \mid r \in I, r' \in I'\}$, and $F_I \cup F_{I'} = F_{I \cap I'}$.

A ring R is *Noetherian* iff every $\subseteq$-increasing sequence of ideals $I_0 \subseteq I_1 \subseteq \cdots \subseteq I_n \subseteq \cdots$ in R is stationary: for some $n \in \mathbb{N}$, all the ideals I_n, I_{n+1}, …, are equal. For example, the ring $K[X_1, \ldots, X_k]$ of all polynomials over the variables $X_1, \ldots, X_k$ with coefficients in K is Noetherian for any field K, in fact even for any Noetherian ring K. For any Noetherian ring R, $\mathrm{Spec}(R)$ is a Noetherian topological space (Grothendieck 1960, corollaire 1.1.6, p. 81). The specialization ordering of $\mathrm{Spec}(R)$ is reverse inclusion $\supseteq$ (Grothendieck 1960, corollaire 1.1.7, p.81). By Grothendieck (1960, proposition 1.1.10, (i), p. 82), the sets $\mathrm{Spec}(R) \setminus \downarrow (r)$ form a base of the Zariski topology, where (r) is the (prime) ideal generated by $r \in R$, so that $\downarrow (r) = \{p \mid p \supseteq (r)\} = \{p \mid r \in p\}$. In particular, the Zariski topology coincides with the upper topology of $\supseteq$ (even when R is not Noetherian).

There are in general several ideals I that yield the same closed set F_I. In fact, two ideals yield the same closed set if and only if they have the same radical; the *radical* $\sqrt{I}$ is defined as $\{r \in R \mid \exists k \geq 1, r^k \in I\}$.

Whatever the ring R, $\mathrm{Spec}(R)$ is always sober (Grothendieck 1960, corollaire 1.1.14, (ii), p. 82). It follows that its irreducible closed subsets are exactly its subsets of the form F_p, p a prime ideal, which are exactly the downward closure (with respect to $\supseteq$) of p. When $\mathrm{Spec}(R)$ is also Noetherian, it follows from Lemma 4.6 that every closed subset F_I of $\mathrm{Spec}(R)$ is a finite union of irreducible closed subsets $F_{p_1} \cup \cdots \cup F_{p_n}$. Since the latter is equal to $F_{p_1 \cap \cdots \cap p_n}$, $\sqrt{I} = \sqrt{p_1 \cap \cdots \cap p_n}$, and the latter equals $p_1 \cap \cdots \cap p_n$ since radical commutes with intersections and since $\sqrt{p} = p$ for every prime ideal p. So, every radical ideal I in a Noetherian ring is the intersection of finitely many prime ideals: this is *Kaplansky's theorem* (Faith 1999, Theorem 14.34). Applying this to the closed subset $F_p \cap F_{p'}$, where p and p' are prime ideals, we obtain that $F_p \cap F_{p'} = F_{p+p'}$ is a finite union of irreducible closed subsets $F_{p_1} \cup \cdots \cup F_{p_n}$ by Lemma 4.6. So $\sqrt{p+p'} = p_1 \cap \cdots \cap p_n$. Applying this to the whole space $\mathrm{Spec}(R) = F_{\{0\}}$, we obtain that $\sqrt{\{0\}} = \{0\} = p_1 \cap \cdots \cap p_n$ for finitely many prime ideals $p_1, \ldots, p_n$.

We therefore obtain an S-representation for $\mathrm{Spec}(R)$, with enough computability assumptions on the ring R. The following proposition is almost vacuous and only reflects our needs for S-representations at the level of rings.

Proposition 6.1 ($\mathrm{Spec}(R)$). *Let R be a Noetherian ring and assume that the set $\mathrm{Spec}(R)$ of prime ideals of R is recursively enumerable that the relation $\trianglelefteq$ defined by $p \trianglelefteq p'$ iff $F_p \subseteq F_{p'}$ iff $\sqrt{p} \supseteq \sqrt{p'}$ is decidable, and that given $p, p' \in \mathrm{Spec}(R)$ one can compute a finite set $p \wedge p'$ of elements $p_1, \ldots, p_n \in \mathrm{Spec}(R)$ such that $\sqrt{p+p'} = p_1 \cap \cdots \cap p_n$. Let also τ be a finite set of prime ideals whose intersection is $\{0\}$.*

Then $(\mathrm{Spec}(R), \mathrm{id}_{\mathrm{Spec}(R)}, \trianglelefteq, \tau, \wedge)$ is an S-representation of $\mathrm{Spec}(R)$.

An important special case is given by taking the polynomial ring $K[X_1, \ldots, X_k]$ for R, where K is a Noetherian ring. For the purpose of computability, we shall even concentrate on $\mathbb{Q}[X_1, \ldots, X_k]$. The latter is an interesting space as far as verification of so-called polynomial programs is concerned (Müller-Olm and Seidl 2002): such programs have k rational-valued variables, and the only allowed operations are $+$, $-$, $\times$, assigning an arbitrary value to a variable non-deterministically, and testing for non-equality. The natural state space for such programs is $\mathbb{Q}^k$. However, $\mathbb{Q}^k$ embeds into $\mathrm{Spec}(\mathbb{Q}[X_1, \ldots, X_k])$, by mapping every tuple $(v_1, \ldots, v_k)$ of values to the prime ideal generated by the polynomials $X_1 - v_1, \ldots, X_k - v_k$. While Müller-Olm and Seidl computed with polynomial ideals directly (Müller-Olm and Seidl 2002), one can alternatively notice that polynomial programs form a topological WSTS, where the state space $\mathbb{Q}^k$ has the subspace topology from $\mathrm{Spec}(\mathbb{Q}[X_1, \ldots, X_k])$ (Goubault-Larrecq 2010).[1]

To satisfy the requirements of Proposition 6.1 for $\mathrm{Spec}(\mathbb{Q}[X_1, \ldots, X_k])$, we represent polynomial ideals using *Gröbner bases* (Buchberger and Loos 1983, Section 11), which are certain finite sets of polynomials $u = \{P_1, \ldots, P_n\}$ representing the ideal $(u) = \{Q_1P_1 + \cdots + Q_nP_n \mid Q_1, \ldots, Q_n \in \mathbb{Q}[X_1, \ldots, X_k]\}$. Given a Gröbner basis u, one can decide whether (u) is a prime ideal: see Adams and Loustaunau (1994, Algorithm 4.4.1, p. 244) or Grieco and Zucchetti (1989, Section 5, end). So the set S of all Gröbner bases u such that (u) is prime is recursively enumerable.

We can now define $[\![u]\!]$ as $F_{(u)}$.

Given two Gröbner bases u and v, it is easy to check whether $(u) \supseteq (v)$. It suffices to check whether $P \in (u)$ for every $P \in v$, and this proceeds using the polynomials of u as rewriting rules and checking whether P rewrites to 0 (Buchberger and Loos 1983, Section 11). However, one needs to decide whether $\sqrt{(u)} \supseteq \sqrt{(v)}$, equivalently, $\sqrt{(u)} \supseteq (v)$, that is, to decide whether $P \in \sqrt{(u)}$ for every $P \in v$. The easiest way to decide this is to use the *Rabinowitch trick* (Rabinowitch 1929): $P \in \sqrt{(u)}$ iff $1 \in (u \cup \{1 - YP\})$, where Y is a fresh variable.

It is clear that one can take $\tau = \{\{0\}\}$, the ideal generated by 0, or equivalently by the empty family of polynomials, since $\{0\}$ is a prime ideal in $\mathbb{Q}[X_1, \ldots, X_k]$, and in fact the minimal prime ideal, so $F_{\{0\}}$ is the unique largest element of $\mathrm{Spec}(\mathbb{Q}[X_1, \ldots, X_k])$.

The really tricky part is in defining the intersection map $\wedge$, that is, to give an effective version of Kaplansky's theorem. The algorithms that allow us to do this are too complicated to even give a glimpse of here. One may consult Laplagne (2006).

Theorem 6.2 (S-representation, spectrum of a polynomial ring). *An S-representation $(S, [\![\_]\!], \trianglelefteq, \tau, \wedge)$ of $\mathrm{Spec}(\mathbb{Q}[X_1, \ldots, X_k])$ in its Zariski topology is given by:*

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

(A) *S is the collection of Gröbner bases u on $\mathbb{Q}[X_1, \ldots, X_k]$ such that (u) is a prime ideal.*
(B) $[\![u]\!] = F_{(u)}$.
(C) *$u \trianglelefteq v$ iff $1 \in (u \cup \{1 - YP\})$, where Y is a fresh variable, for every $P \in v$.*
(D) $\tau = \{\{0\}\}$.
(E) *$u \wedge v$ is a finite collection of Gröbner bases $u_1, \ldots, u_n$ such that $(u_1), \ldots, (u_n)$ are prime ideals and $\sqrt{u+v} = (u_1) \cap \cdots \cap (u_n)$, computed by Laplagne's algorithm (Laplagne 2006).*

An alternative S-representation of $\mathrm{Spec}(R)$ is given using for S the set of all those finite sets u of polynomials such that (u) is a *primary ideal*, instead of a prime ideal. A primary ideal p is such that whenever rr' is in p, then $r \in p$ or some power of r' is in p. Every prime ideal is primary, but the converse fails. The radical $\sqrt{p}$ of a primary ideal is always prime. Given a set u of polynomials, one can decide whether (u) is primary (Grieco and Zucchetti 1989, Theorem 3.2), and in fact one can compute a Gröbner basis for $\sqrt{(u)}$ in this case. So S is again, in particular, recursively enumerable. We define again $[\![u]\!]$ as $F_{(u)}$. Since $F_{(u)} = F_{\sqrt{(u)}}$ and $\sqrt{(u)}$ is prime, $F_{(u)}$ is certainly an irreducible closed subset. Next, $\trianglelefteq$ and τ are defined as above, while $\wedge$ is now based on a computable variant of the *Lasker–Noether theorem*, instead of Kaplansky's theorem. This states that every ideal I in a Noetherian ring R can be written as the intersection of finitely many primary ideals. When $R = \mathbb{Q}[X_1, \ldots, X_k]$, then one can even compute a finite collection of Gröbner bases $w_1, \ldots, w_m$ such that $(u+v) = (w_1) \cap \cdots \cap (w_m)$ and $(w_1), \ldots, (w_m)$ are primary ideals, see Sturmfels (2002, Chapter 5). Now given $u, v \in S$, $[\![u]\!] \cap [\![v]\!] = F_{(u)} \cap F_{(v)} = F_{(u+v)} = F_{\sqrt{(u+v)}}$. One can compute a finite collection of Gröbner bases $w_1, \ldots, w_k$ such that $\sqrt{(u+v)} = \sqrt{(w_1)} \cap \cdots \cap \sqrt{(w_k)}$ and $(w_1), \ldots, (w_k)$ are primary ideals. Then, $[\![u]\!] \cap [\![v]\!] = F_{\sqrt{(w_1)}} \cup \cdots \cup F_{\sqrt{(w_k)}} = F_{(w_1)} \cup \cdots \cup F_{(w_k)} = \bigcup_{i=1}^{k} [\![w_i]\!]$: define $u \wedge v$ as $\{w_1, \ldots, w_k\}$.

To sum up

Theorem 6.3 (S-representation, spectrum of a polynomial ring, alternate). *An S-representation $(S, [\![\_]\!], \trianglelefteq, \tau, \wedge)$ of $\mathrm{Spec}(\mathbb{Q}[X_1, \ldots, X_k])$ in its Zariski topology is given by:*

(A) *S is the collection of Gröbner bases u on $\mathbb{Q}[X_1, \ldots, X_k]$ such that (u) is a primary ideal.*
(B) $[\![u]\!] = F_{(u)}$.
(C) *$u \trianglelefteq v$ iff $1 \in (u \cup \{1 - YP\})$, where Y is a fresh variable, for every $P \in v$.*
(D) $\tau = \{1\}$.
(E) *$u \wedge v$ is a finite collection of Gröbner bases $u_1, \ldots, u_n$ such that $(u_1), \ldots, (u_n)$ are primary ideals and $\sqrt{u+v} = (u_1) \cap \cdots \cap (u_n)$, computed as in Sturmfels (2002, Chapter 5).*

We finish this section by mentioning an issue with our polynomial program example. We really think of the state space as $\mathbb{Q}^k$, not the larger space $\mathrm{Spec}(\mathbb{Q}[X_1, \ldots, X_k])$. To make things formal, this means equipping $\mathbb{Q}^k$ with the subspace topology, whose closed subsets are exactly those sets of the form $\mathcal{Z}(u) = \{\vec{x} \in \mathbb{Q}^k \mid \forall P \in u, P(\vec{x}) = 0\}$, for u an ideal in $\mathbb{Q}[X_1, \ldots, X_k]$. That topology is usually called the *Zariski topology* on $\mathbb{Q}^k$ and makes polynomial programs topological WSTS. Whether we use $\mathbb{Q}^k$ or $\mathrm{Spec}(\mathbb{Q}[X_1, \ldots, X_k])$ is of little consequence if we use the backward algorithm mentioned in the introduction, because the only thing it cares about is open subsets, which can be encoded as complements of sets $\mathcal{Z}(u)$, namely as ideals u.

The situation is different with S-representations, since S-representations do not encode closed sets, but *irreducible* closed subsets, and $\mathrm{Spec}(\mathbb{Q}[X_1, \ldots, X_k])$ contains many more irreducible closed subsets than $\mathbb{Q}^k$. This boils down to the fact that we do not know an S-representation for $\mathbb{Q}^k$ with its Zariski topology: the situation for the apparently more complex space $\mathrm{Spec}(\mathbb{Q}[X_1, \ldots, X_k])$ is simpler.

The situation is the following: we have two spaces $X = \mathbb{Q}^k$ and $Y = \mathrm{Spec}(\mathbb{Q}[X_1, \ldots, X_k])$, and X is a subspace of Y; we know of an S-representation for Y, can we infer one for X?

Proposition 6.4. *Let X be a subspace of a topological space Y. Then, $\mathcal{S}(X)$ embeds into $\mathcal{S}(Y)$, that is, every irreducible closed subset of X can be equated, in a canonical way, with some irreducible closed subset of Y.*

Proof. Let $m: X \to Y$ be the inclusion map. Then, $\mathcal{S}(m)$ is a topological embedding (Goubault-Larrecq 2013, Lemma 8.4.11). In other words, every irreducible closed subset C of X can be equated with $\mathcal{S}(m)(C)$, namely the closure of C in Y, and that is irreducible closed in Y. □

In our case, this means that an S-representation for $\mathbb{Q}^k$ consists in a subset of either set of codes considered in Proposition 6.2 or in Proposition 6.3. Characterizing those codes remains to be elucidated.

7. Completing Words

If X is a wqo, then X^* is a wqo again under the embedding quasi-ordering by Higman's Lemma. This is often used when X is a finite alphabet Σ, with equality as quasi-ordering, but more general wqos are sometimes needed. For instance, Abdulla et al. (2004b) need to use X^* where $X = \Sigma^{\circledast}$, the set of finite multisets on a finite alphabet Σ. (We will deal with multisets in Section 8.) In that case, X itself is infinite. That paper is also one where a suitable theory of downward-closed subsets was first developed, on $(\Sigma^{\circledast})^*$, and our constructions will generalize theirs. Data nets (Lazič et al. 2008) are transition systems on a state space of the form X^* with $X = \mathbb{N}^k$, for some $k \in \mathbb{N}$, and again X is infinite in this case. More recently, Leroux and Schmitz have analyzed the question of reachability in Petri nets (Leroux and Schmitz 2015) and required to work on ideals in the space of runs of Petri nets, which is a subspace of $(\mathbb{N}^k)^*$.

We work at the more general level of Noetherian spaces. In that context, the analog of Higman's Lemma reads: for every Noetherian space X, the set X^* of finite words over X taken as alphabet is Noetherian again, with the so-called *word topology* (Goubault-Larrecq 2013, Theorem 9.7.33). (The converse also holds.) The latter topology is generated by basic open subsets $X^* U_1 X^* U_2 X^* \cdots X^* U_n X^*$, where $n \in \mathbb{N}$ and $U_1, \ldots, U_n$ are open subsets of X. We write AB for the sets of concatenations ww' of words $w \in A$ and $w' \in B$ and equate subsets of X such as U_i with the set of one-letter words whose letter is in U_i. So $X^* U_1 X^* U_2 X^* \cdots X^* U_n X^*$ is the (open) subset of words containing a not necessarily contiguous word $a_1 a_2 \cdots a_n$ with $a_1 \in U_1$, $a_2 \in U_2$, $\ldots$, $a_n \in U_n$. We stress that such subsets form a *base*, not just a subbase:

Lemma 7.1. *Let X be a topological space. Call* elementary open *of X^* any subset of the form $X^* U_1 X^* U_2 X^* \cdots X^* U_n X^*$, with all U_i open in X. Every finite intersection of elementary opens can be expressed as a finite union of elementary opens. In particular, the elementary opens form a base of the word topology.*

Proof. This is Exercise 9.7.28 of Goubault-Larrecq (2013). An empty intersection is just X^*, and the intersection of $X^* U_1 X^* U_2 X^* \ldots X^* U_m X^*$ and $X^* V_1 X^* V_2 X^* \ldots X^* V_n X^*$ is computed by induction on $m + n$ using the auxiliary formulae $X^* \cap \mathcal{V} = \mathcal{V}$, $\mathcal{U} \cap X^* = \mathcal{U}$, and $X^* U_1 \mathcal{U} \cap X^* V_1 \mathcal{V} = X^* U_1 (\mathcal{U} \cap X^* V_1 \mathcal{V}) \cup X^* V_1 (X^* U_1 \mathcal{U} \cap \mathcal{V}) \cup X^* (U_1 \cap V_1)(\mathcal{U} \cap \mathcal{V})$. □

If $\leq$ is the specialization quasi-ordering of X, then the specialization quasi-ordering of X^* is the standard *embedding quasi-ordering* $\leq^*$, a.k.a. Higman's divisibility quasi-ordering (Higman 1952): $w \leq^* w'$ iff, writing w as the sequence of m letters $a_1 a_2 \cdots a_m$, one can write w' as $w_0 a'_1 w_1 a'_2 w_2 \cdots w_{m-1} a'_m w'_m$ with $a_1 \leq a'_1, a_2 \leq a'_2, \ldots, a_m \leq a'_m$. Higman's Lemma states that if X is well-quasi-ordered by $\leq$, then X^* is well-quasi-ordered by $\leq^*$ (Higman 1952). The fact that X^* is Noetherian if and only if X is Noetherian is a natural generalization of Higman's Lemma: the latter can be obtained as a special case by considering Alexandroff topologies (Goubault-Larrecq 2013, Exercise 9.7.34).

The completion $\mathcal{S}(X^*)$ is well known in case X is wqo. As mentioned in the introduction, this is due to Kabil and Pouzet (1992). Kabil and Pouzet also look at the (ideal) completion of spaces of finite words over more general ordered sets X. We explore another direction, that where X is Noetherian. This will include the result by Kabil and Pouzet in the wqo case as a by-product. Additionally, we give a simple, dynamic programming algorithm for deciding inclusion between irreducible closed subsets, and computing intersections, retrieving formulae that were known in the case where X is finite (Abdulla et al. 1998).

To study the completion $\mathcal{S}(X^*)$, we start by examining the shape of closed subsets of X^*. For any subset A of X, let A^* denote the set of all words $a_1a_2\cdots a_n$ with $a_1, a_2, \ldots, a_n \in A$, $n \in \mathbb{N}$ (n is possibly equal to 0). Let $A^?$ be $A \cup \{\epsilon\}$. We delegate the proof of the following Lemma to Appendix B, and similarly for a certain number of other results of this section. Our aim is to avoid disrupting the flow of arguments and to proceed as fast as we can to the final result.

Lemma 7.2. *Let X be a topological space. The complement of $X^*U_1X^*U_2X^* \cdots X^*U_nX^*$ ($n \in \mathbb{N}$, $U_1, U_2, \ldots, U_n$ open in X) in X^* is $\emptyset$ when $n = 0$, and $F_1^*X^?F_2^*X^? \cdots X^?F_{n-1}^*X^?F_n^*$ otherwise, where $F_1 = X \smallsetminus U_1, \ldots, F_n = X \smallsetminus U_n$.*

If X is Noetherian, then this complement can be expressed as a finite union of sets of the form $F_1^*C_1^?F_2^*C_2^? \cdots C_{n-1}^?F_n^*$, where $C_1, C_2, \ldots, C_{n-1}$ range over irreducible closed subsets of X.

Definition 7.3 (Word-product, word-SRE). *Let X be a topological space. Call a* word-product *P on X any expression of the form $e_1e_2 \cdots e_n$, where $n \geq 0$, and each e_i is an* atomic expression, *that is, either F_i^* with F_i closed in X, or $F_i^?$ with F_i irreducible closed in X. The* components *of P are the closed sets $F_1, \ldots, F_n$. Word-products are interpreted as the obvious subsets of X^*. When $n = 0$, this notation is abbreviated as ϵ and denotes the one-element set $\{\epsilon\}$.*

Call word-SRE *any finite sum of word-products, where sum is interpreted as union.*

There is no harm in requiring F_i non-empty in addition, in atomic expressions F_i^*: indeed $\emptyset^* = \{\epsilon\}$, so such atomic expressions can simply be erased.

This definition is inspired from the products and SREs of Abdulla et al. (2004a). Indeed, we get back the latter from Definition 7.3 in the case where X is a finite alphabet Σ, with the discrete topology (hence its specialization quasi-ordering is $=$). Then each closed subset F_i is just a finite subset, and each irreducible closed subset C_i is just a singleton.

Lemma 7.4. *Let X be a topological space. For every closed subset F of X, for every closed subset $\mathcal{F}$ of X^*, $F^?\mathcal{F}$ is closed in X^*.*

Lemma 7.5. *Let X be a topological space. For every closed subset F of X, for every closed subset $\mathcal{F}$ of X^*, $F^*\mathcal{F}$ is closed in X^*.*

Corollary 7.6. *Let X be a topological space. For every word-product, every word-SRE is closed in X^*.*

We can in fact say more:

Lemma 7.7. *Let X be a topological space. Every word-product is irreducible closed in X^*.*

It is instructive to see how X^* embeds in its completion $\mathcal{S}(X)$. Recall that the topological closure $\eta_X^{\mathcal{S}}(x)$ of a point $x \in X$ is also its downward closure $\downarrow x$, for the specialization quasi-ordering of X.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Lemma 7.8 (Embedding). *Let X be a topological space. The closure $\eta^{\mathcal{S}}_{X^*}(x_1x_2\cdots x_n)$ of the word $x_1x_2\cdots x_n$ in X^* is the word-product $\eta^{\mathcal{S}}_X(x_1)^?\eta^{\mathcal{S}}_X(x_2)^?\cdots\eta^{\mathcal{S}}_X(x_n)^?$.*

Proof. The latter is easily seen to be the downward closure of $x_1x_2\cdots x_n$ with respect to $\leq^*$, which is the specialization quasi-ordering of X^*. □

We shall see that the converse of Lemma 7.7 holds the irreducible closed subsets of X^*, that is, the elements of $\mathcal{S}(X^*)$ are exactly the word-products when X is Noetherian. The following lemmas will serve to show this, as well as to give some ways of computing on word-products. We do not make an explicit distinction between syntax and semantics, on purpose, so as to avoid excessively formal notation.

Lemma 7.9. *Let X be a topological space. Inclusion between word-products can be checked in polynomial time (precisely in time proportional to the product of the lengths of the two word-products), modulo an oracle testing inclusion of closed subsets of X.*

Explicitly, we have $\epsilon\subseteq P$ for any word-product P, $P\not\subseteq\epsilon$ unless all the atomic expressions in P are syntactically equal to $\emptyset^*$, and for all $C, C'\in\mathcal{S}(X)$, for all $F, F'\in\mathcal{H}_V(X)$, and for all word-products P, P':

- *$C^?P\subseteq C'^?P'$ if and only if $C\subseteq C'$ and $P\subseteq P'$, or $C\not\subseteq C'$ and $C^?P\subseteq P'$.*
- *$C^?P\subseteq F'^*P'$ if and only if $C\subseteq F'$ and $P\subseteq F'^*P'$, or $C\not\subseteq F'$ and $C^?P\subseteq P'$.*
- *$F^*P\subseteq C'^?P'$ if and only if F is empty and $P\subseteq C'^?P'$, or F is non-empty and $F^*P\subseteq P'$.*
- *$F^*P\subseteq F'^*P'$ if and only if $F\subseteq F'$ and $P\subseteq F'^*P'$, or $F\not\subseteq F'$ and $F^*P\subseteq P'$.*

The above formulae lend themselves immediately to a dynamic programming algorithm, modulo an oracle O testing inclusion of closed subsets of X. Assume that we wish to test whether $P\subseteq P'$, where $P=e_1e_2\cdots e_m$ and $P'=e_1e_2\cdots e_n$. We create an $(m+1)\times(n+1)$ array $A=(a_{ij})_{0\leq i\leq m, 0\leq j\leq n}$. At the end of the algorithm, a_{ij} will be true if and only if $e_{i+1}\cdots e_m\subseteq e_{j+1}\cdots e_n$. We initialize A by letting a_{mj} be true for every j, $0\leq j\leq n$. For every i, $0\leq i<m$, we set a_{in} to false, unless $e_{i+1},\ldots,e_m$ are all equal to $\emptyset^*$, in which case a_{in} is set to true; explicitly, we initialize a flag b to true, and enumerating i from $m-1$ to 0, we do the following: if e_{i+1} is not of the form F^*, or is of the form F^* with $F\not\subseteq\emptyset$ (which we can decide using the oracle O), then set b to false, otherwise leave b unchanged, then set a_{in} to b. This completes the initialization phase. Then, using two nested loops on i and j, one enumerating i from $m-1$ to 0, the other one enumerating j from $n-1$ to 0 (for each value of i), we set a_{ij} to true if and only if:

- e_{i+1} is of the form $C^?$, e_{j+1} is of the form $C'^?$, and either $C\subseteq C'$ (which we decide using the oracle O) and $a_{(i+1)(j+1)}$ is true, or $C\not\subseteq C'$ and $a_{i(j+1)}$ is true;
- or e_{i+1} is of the form $C^?$, e_{j+1} is of the form F'^*, and either $C\subseteq F'$ and $a_{(i+1)j}$ is true or $C\not\subseteq F'$ and $a_{i(j+1)}$ is true;
- or e_{i+1} is of the form F^*, e_{j+1} is of the form $C'^?$, and either F is empty (which we decide using O on F and $\emptyset$, as in the second part of the initialization phase) and $a_{(i+1)j}$ is true, or F is non-empty and $a_{i(j+1)}$ is true;
- or e_{i+1} is of the form F^*, e_{j+1} is of the form F'^*, and either $F\subseteq F'$ and $a_{(i+1)j}$ is true, or $F\not\subseteq F'$ and $a_{i(j+1)}$ is true.

Otherwise, we set a_{ij} to false. At the end of the nested loops, we return a_{00}, which is true if and only if $P\subseteq P'$. Alternatively to dynamic programming, we may use a directed recursive implementation, with memoization (Michie 1968).

We can rephrase the equations of Lemma 7.9 in the slightly more synthetic, following form. This happens to be the inclusion of products as specified in Abdulla et al. (2004a), in the case

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

where X is a finite set. The fourth case is not needed if we first remove all atomic expressions $\emptyset^*$. We will refer to this specific formulation in the proof of Lemma 11.22, and in Definition 11.31.

Lemma 7.10. *Let X be a topological space. Given two atomic expressions e_1 and e'_1, and two word-products P_1 and P'_1, letting $P = e_1 P_1$ and $P' = e'_1 P'_1$, then $P \sqsubseteq P'$ if and only if:*

(1) $e_1 \not\sqsubseteq e'_1$ and $P \subseteq P'_1$,
(2) or $e_1 = C^?$, $e'_1 = C'^?$, $C \subseteq C'$ and $P_1 \subseteq P'_1$,
(3) or $e'_1 = F'^*$, $e_1 \sqsubseteq e'_1$ and $P_1 \subseteq P'$,
(4) or $e_1 = \emptyset^*$ and $P_1 \subseteq P'$.

The relation $\sqsubseteq$ on atomic expressions is defined by: $C^? \sqsubseteq C'^?$ if and only if $C \subseteq C'$; $F^* \sqsubseteq F'^*$ if and only if $F \subseteq F'$; $C^? \sqsubseteq F'^*$ if and only if $C \subseteq F'$; and $F^* \sqsubseteq C'^?$ if and only if F is empty.

Corollary 7.11. *Let X be a topological space. Inclusion between word-SREs can be checked in polynomial time, modulo an oracle testing inclusion of closed subsets of X.*

Proof. By Lemma 5.2, and since word-products are irreducible closed (Lemma 7.7), inclusion of word-SREs $P_1 \cup \cdots \cup P_m$ and $P'_1 \cup \cdots \cup P'_n$ reduces to mn inclusion tests $P_i \subseteq P'_j$ between word-products, which we decide using the dynamic programming algorithm mentioned after Lemma 7.9. □

We can also compute intersections of word-products.

Lemma 7.12. *Let X be a topological space. Any finite intersection of word-products is expressible as a finite union of word-products. Specifically, the intersection of two word-products is given by: $\epsilon \cap P = \epsilon$ for every word-product P, and by the recursive formulae:*

- $C^? P \cap C'^? P' = (C^? P \cap P') \cup (P \cap C'^? P') \cup (C \cap C')^? (P \cap P')$;
- $C^? P \cap F'^* P' = (C \cap F')^? (P \cap F'^* P') \cup (C^? P \cap P')$;
- $F^* P \cap F'^* P' = (F \cap F')^* (P \cap F'^* P') \cup (F \cap F')^* (F^* P \cap P')$.

Recall that the components of a word-product $P = e_1 e_2 \cdots e_n$ are the components of each e_i, where the component of $C^?$ is C, and the component of F^* is F. Lemma 7.12 yields the following, more computation-oriented description of the intersection algorithm for word-products. Our particular way of presenting it will be helpful in Theorem 7.15, in Lemma 11.26, and in the proof of Lemma 11.34.

Lemma 7.13. *Let X be a Noetherian space. Define the finite set $Meet^{\mathcal{E}}(P, P')$ of word-products as follows, where P and P' are word-products, and the oracle $\mathcal{E}$ maps pairs (F, F') of a component F of P and a component F' of P' to a finite set of irreducible closed subsets of X.*

First, let $Meet^{\mathcal{E}}(\epsilon, P') = \{\epsilon\}$, $Meet^{\mathcal{E}}(P, \epsilon) = \{\epsilon\}$. Then, let

$$Meet^{\mathcal{E}}(C^? P, C'^? P') = \{C''^? P'' \mid C'' \in \mathcal{E}(C, C'), P'' \in Meet^{\mathcal{E}}(P, P')\} \cup Meet^{\mathcal{E}}(C^? P, P') \cup Meet^{\mathcal{E}}(P, C'^? P')$$

$$Meet^{\mathcal{E}}(C^? P, F'^* P') = \begin{cases} \{C''^? P'' \mid C'' \in \mathcal{E}(C, F'), \\ \quad P'' \in Meet^{\mathcal{E}}(P, F'^* P')\} \cup Meet^{\mathcal{E}}(C^? P, P') & \text{if } \mathcal{E}(C, F') \neq \emptyset, \\ Meet^{\mathcal{E}}(P, F'^* P') \cup Meet^{\mathcal{E}}(C^? P, P') & \text{otherwise} \end{cases}$$

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

$$Meet^{\mathcal{E}}(F^*P, C'^?P') = \begin{cases} \{C''^?P'' \mid C'' \in \mathcal{E}(F, C'), \\ \qquad P'' \in Meet^{\mathcal{E}}(F^*P, P')\} \cup Meet^{\mathcal{E}}(P, C'^?P') & \textit{if } \mathcal{E}(F, C') \neq \emptyset, \\ Meet^{\mathcal{E}}(F^*P, P') \cup Meet^{\mathcal{E}}(P, C'^?P') & \textit{otherwise} \end{cases}$$

$$Meet^{\mathcal{E}}(F^*P, F'^*P') = \{(\bigcup_{C'' \in \mathcal{E}(F,F')} C'')^*P'' \mid P'' \in Meet^{\mathcal{E}}(F^*P, P') \cup Meet^{\mathcal{E}}(P, F'^*P')\}.$$

If $\mathcal{E}$ computes intersections of closed subsets of X, that is, is such that for any component F of P and any component F' of P', $\mathcal{E}(F, F')$ is a finite family of irreducible closed subsets of X whose union is $F \cap F'$, then $Meet^{\mathcal{E}}(P, P')$ is a finite family of word-products whose union is $P \cap P'$.

Note that the map $(F, F') \mapsto \mathcal{E}(F, F')$ is well defined, by Lemma 4.6. We will later require to be able to compute it.

Proof. The lemma is a simple consequence of Lemma 7.12.

In the first case, if $C \cap C'$ is non-empty, we conclude since $(C \cap C')^?(P \cap P') = \bigcup_{C'' \in \mathcal{E}(C,C')} C''^?(P \cap P')$, $P \cap P'$ is the union of the word-products in $Meet^{\mathcal{E}}(P, P')$, and unions distribute over concatenation. There is a subtle issue when $C \cap C'$ is empty. In that subcase, $(C \cap C')^?(P \cap P')$ is equal to $P \cap P'$, and that is different from $\bigcup_{C'' \in \mathcal{E}(C,C')} C''^?(P \cap P')$, which is empty; however, $(C \cap C')^?(P \cap P')$ is equal to $(P \cap P') \cup (C^?P \cap P') \cup (P \cap C'^?P')$, hence also to $(C^?P \cap P') \cup (P \cap C'^?P')$, because $P \cap P'$ is included in $C^?P \cap P'$ (or in $P \cap C'^?P'$), and that justifies the indicated formula again. In the second case (and symmetrically, the third case), we rely on $(C \cap F')^?(P \cap F'^*P') = \bigcup_{C'' \in \mathcal{E}(C,F')} C''^?(P \cap F'^*P')$, which is valid if $\mathcal{E}(C, F')$ is non-empty. If $\mathcal{E}(C, F')$ is empty, then $C \cap F'$ is empty, and then $(C \cap F')^?$ is not equal to $\bigcup_{C'' \in \mathcal{E}(C,F')} C''^?$, rather to $\{\epsilon\}$; so $(C \cap F')^?(P \cap F'^*P') = P \cap F'^*P'$ in that (sub)case. In the final case, we use the fact that $F \cap F' = \bigcup_{C'' \in \mathcal{E}(F,F')} C''$.

Finally, the definition of $Meet^{\mathcal{E}}(P, P')$ is well founded, by induction on the number of atomic expressions in P and P'. □

Proposition 7.14. *Let X be a Noetherian space. The closed subsets of X^* are the (languages of) word-SREs, and the irreducible closed subsets of X^* are the (languages of) word-products.*

Proof. Lemma 7.7 states that every word-product is irreducible closed.

Conversely, we observe that, in a Noetherian space Y with a base $\mathcal{B}$ of opens, every open is a *finite* union of elements of $\mathcal{B}$. This is an easy consequence of the fact that every open, which is a union of elements of $\mathcal{B}$, is also compact.

Consider $Y = X^*$, $\mathcal{B}$ consisting of the subsets of the form $X^*U_1X^*U_2X^* \cdots X^*U_nX^*$, where each U_i is open in X (Lemma 7.1). Taking the complements of finite unions of such basic opens, and using Lemma 7.2, one obtains that every closed subset of X^* is a finite intersection of finite unions of word-products. Distributing unions over intersections, and using Lemma 7.12, we conclude that every closed subset F is expressible as a word-SRE, that is, as a finite union of word-products.

If F is also irreducible, it follows immediately that F is one of these word-products. □

We now state the final S-representation we obtain, in a way that we hope will be readable. The pedantic, formal statement is given in the Appendix (Proposition B.2).

Theorem 7.15 (S-representation, words). *Let X be a Noetherian space, $X' = X^*$, and $(S, [\![\_]\!], \unlhd, \tau, \wedge)$ be an S-representation of X. Then, $(S', [\![\_]\!]', \unlhd', \tau', \wedge')$ is an S-representation of X', where:*

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

(A) *S' is the collection of all word-products over the alphabet S, and $[\![\_]\!]'$ is defined in the obvious way.*
(B) *$\trianglelefteq'$ is defined using the procedure of Lemma 7.9, where inclusion of finite sets of elements of S is tested by: u is included in u' iff for every $a \in u$, there is an $a' \in u'$ such that $a \trianglelefteq a'$ (Lemma 5.2).*
(C) *τ' is $\{\tau^*\}$.*
(D) *$\wedge'$ is implemented by the procedure $Meet^{\mathcal{E}}$ of Lemma 7.13, where the oracle $\mathcal{E}$ is defined by $\mathcal{E}(u, u') = \bigcup_{a \in u, a' \in u'} (a \wedge a')$.*

Note that $[\![\_]\!]'$ is surjective, as required: the irreducible closed subsets of X' are the word-products by Proposition 7.14.

8. Completing Multisets

If X is a wqo, then the space of finite multisets $X^{\circledast}$ of elements of X, with a quasi-ordering $\leq^{\circledast}$ to be defined below, is a wqo again. This is again typically used when X is a finite alphabet Σ: the multiset language generators of Abdulla et al. (2004b) are the ideals of such a wqo $\Sigma^{\circledast}$.

Beyond finite alphabets, branching vector addition systems with states (BVASS) are a generalization of Petri nets with a form of branching, with applications in security (Verma and Goubault-Larrecq 2005), in linear logic (de Groote et al. 2004), in structured databases (Bojańczyk et al. 2009; Jacquemard et al. 2016), and are a rediscovery of Rambow's multiset-valued linear indexed grammars (Rambow 1994) in computational linguistics, see Schmitz (2010). BVASS, and some of their extensions, can be conveniently represented as transition systems on the space $(\mathbb{N}^k)^{\circledast}$ of finite multisets of k-tuples of natural numbers (Jacobé de Naurois 2014). Note that the alphabet $(\mathbb{N}^k)$ is infinite in this case.

As a final example, the synchronous polyadic π-calculus processes investigated in Acciai and Boreale (2012) are encoded as trees, which can be seen as nested multisets, with bounded nesting depth. Encoding processes by trees of this form was pioneered by Meyer (2008). Precisely, for a finite set Σ, which consists of channel names and so-called unit processes in that case, let $\mathcal{T}_0^{\circledast}(\Sigma)$ be defined recursively as the set of finite trees $f(m)$ where $f \in \Sigma$ and m is a finite multiset of elements of $\mathcal{T}_0^{\circledast}(\Sigma)$, quasi-ordered by the universal relation ($s \leq_0 t$ is always true). Define $\mathcal{T}_{k+1}^{\circledast}(\Sigma)$, for every $k \in \mathbb{N}$, as the set of finite trees $f(m)$ where $f \in \Sigma$ and m is a finite multiset of elements of $\mathcal{T}_k^{\circledast}(\Sigma)$, quasi-ordered by $\leq_{k+1}$, defined by $f(m) \leq_{k+1} f'(m')$ if and only if $f = f'$ and $m(\leq_k)^{\circledast} m'$. All processes are encoded as elements of $\mathcal{T}_k^{\circledast}(\Sigma)$ for some $k \in \mathbb{N}$. Equivalently, $\mathcal{T}_k^{\circledast}(\Sigma)$ is $\Sigma \times (\Sigma \times (\Sigma \times \cdots \times (\Sigma \times Y^{\circledast})^{\circledast} \cdots)^{\circledast}$, where there are k nested uses of $\_^{\circledast}$, and $Y = \mathcal{T}_0^{\circledast}(\Sigma)$.

We again turn to a more general topological setting. Given any topological space, let $X^{\circledast}$ be the set of all finite multisets on X. We shall write $\{\!|x_1, \ldots, x_n|\!\}$ for the multiset containing exactly the elements $x_1, \ldots, x_n$. We write $\emptyset$ for the empty multiset, and $m \uplus m'$ for the multiset union of m and m'.

On the order-theoretic side, we quasi-order $X^{\circledast}$, not with the multiset extension $\leq^{mul}$ of the specialization quasi-ordering $\leq$ of X, rather with the following quasi-ordering.

Definition 8.1 (Sub-multiset). *The* sub-multiset quasi-ordering $\leq^{\circledast}$ *is defined by:* $\{\!|x_1, x_2, \ldots, x_m|\!\} \leq^{\circledast} \{\!|y_1, y_2, \ldots, y_n|\!\}$ *if and only if there is an injective map* $r\colon \{1, 2, \ldots, m\} \to \{1, 2, \ldots, n\}$ *such that* $x_i \leq y_{r(i)}$ *for every* i, $1 \leq i \leq m$.

When $\leq$ is just equality, this quasi-ordering makes $m \leq^{\circledast} m'$ if and only if every element of m occurs at least as many times in m' as it occurs in m: this is the $\leq^m$ quasi-ordering considered, on finite sets X, by Abdulla et al. (2004b, Section 2).

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

The multiset extension $\leq^{mul}$ of $\leq$ is more mainstream than $\leq^{\circledast}$. One usually defines $m \leq^{mul} m'$ if and only if one can obtain m from m' in finitely many steps, repeatedly replacing one element by finitely many strictly smaller ones.

Clearly, $m \leq^{\circledast} m'$ implies $m \leq^{mul} m'$. It turns out that $\leq^{\circledast}$ is wqo for every wqo $\leq$. This implies that $\leq^{mul}$ is wqo, too: any infinite sequence of multisets m_n, $n \in \mathbb{N}$, is such that there are indices $i < j$ such that $m_i \leq^{\circledast} m_j$, and therefore $m_i \leq^{mul} m_j$.

On the topological side, we simply observe that multisets are equivalence classes of finite words up to permutation. Accordingly, we topologize $X^{\circledast}$ with the quotient topology (Goubault-Larrecq 2013, Exercise 9.7.35). The quotient map $\Psi\colon X^* \to X^{\circledast}$ sends every word $x_1 x_2 \cdots x_n$ to the multiset $\{\!|x_1, x_2, \ldots, x_n|\!\}$ and is sometimes called the *Parikh mapping* (Parikh 1966). We have the following results.

Proposition 8.2. *For every Noetherian space X, $X^{\circledast}$ is Noetherian.*

A base of the topology on $X^{\circledast}$ is given by the sets $\langle U_1, U_2, \ldots, U_n\rangle$ with U_1, U_2, …, U_n open in X. The set $\langle U_1, U_2, \ldots, U_n\rangle$ is defined as containing all multisets that contain one element from U_1, another one from U_2, …, another one from U_n, or more precisely all multisets of the form $\{\!|x_1, x_2, \ldots, x_n|\!\} \uplus m$ with $x_1 \in U_1$, $x_2 \in U_2$, …, $x_n \in U_n$.

The specialization quasi-ordering of $X^{\circledast}$ is $\leq^{\circledast}$, where $\leq$ is the specialization quasi-ordering of X.

If X has the Alexandroff topology of $\leq$, then $X^{\circledast}$ has the Alexandroff topology of $\leq^{\circledast}$.

If $\leq$ is wqo, then $\leq^{\circledast}$ is wqo.

Proof. If X is Noetherian, then $X^{\circledast}$ is, too, since every quotient of a Noetherian space is Noetherian (Goubault-Larrecq 2013, Proposition 9.7.18 (v)).

The inverse image $\Psi^{-1}(\langle U_1, U_2, \ldots, U_n\rangle)$ is equal to the union over all permutations π of $\{1, 2, \ldots, n\}$ of the opens $X^* U_{\pi(1)} X^* \cdots X^* U_{\pi(n)} X^*$, hence is open in X^* (see Section 7). For every subset V of $X^{\circledast}$, V is open (in the quotient topology) if and only if $\Psi^{-1}(V)$ is open, so $\langle U_1, U_2, \ldots, U_n\rangle$ is open in $X^{\circledast}$. To show that these sets form a base, take a multiset m and an open neighborhood V of m in $X^{\circledast}$. Write m as $\Psi(w)$ for some word $w \in X^*$. Since $w \in \Psi^{-1}(V)$, and we know of a base of the topology of X^* (Lemma 7.1), we can find open subsets U_1, U_2, …, U_n of X such that $w \in X^* U_1 X^* U_2 X^* \cdots X^* U_n X^* \subseteq \Psi^{-1}(V)$. Then $m = \Psi(w)$ is in $\Psi[X^* U_1 X^* U_2 X^* \cdots X^* U_n X^*] = \langle U_1, U_2, \ldots, U_n\rangle$, which is included in V.

If $m \leq^{\circledast} m'$, then every basic open subset $\langle U_1, U_2, \ldots, U_n\rangle$ that contains m also contains m', so m is below m' in the specialization quasi-ordering. Conversely, we shall show that the downward closure $\downarrow^{\circledast} m'$ of m' with respect to $\leq^{\circledast}$ is closed: if m is below m' in the specialization quasi-ordering, then m will be in the closure of m', hence in $\downarrow^{\circledast} m'$, and this will imply that $m \leq^{\circledast} m'$. To show that $\downarrow^{\circledast} m'$ is closed, it is enough to show that $\Psi^{-1}(\downarrow^{\circledast} m')$ is closed in X^*, since Ψ is quotient. Write m' as $\Psi(w')$, where w' is the word $x_1 x_2 \cdots x_n$. Then $\Psi^{-1}(\downarrow^{\circledast} m')$ is the union over all permutations π of $\{1, 2, \ldots, n\}$ of the sets $(\downarrow x_{\pi(1)})^? (\downarrow x_{\pi(2)})^? \cdots (\downarrow x_{\pi(n)})^?$, which are closed by Lemma 7.7.

Assume now that X has the Alexandroff topology of $\leq$. The upward closure of a multiset $m = \{\!|x_1, x_2, \ldots, x_n|\!\}$ in $X^{\circledast}$ is equal to the open subset $\langle \uparrow x_1, \uparrow x_2, \ldots \uparrow x_n\rangle$. Every upward-closed subset is the union of the upward closures of its points, hence is open, too, so $X^{\circledast}$ has the Alexandroff topology of $\leq^{\circledast}$.

In particular, if $\leq$ is wqo, then X with the Alexandroff topology of $\leq$ is Noetherian, so $X^{\circledast}$ is Noetherian, too. Since $X^{\circledast}$ has the Alexandroff topology of $\leq^{\circledast}$, the latter is wqo. □

Our candidates for (irreducible) closed subsets of $X^{\circledast}$ are the Parikh images of word-products and word-SREs. Write $F \mid C_1, C_2, \ldots, C_n$ for the family of all multisets that one can obtain by picking at most one element from C_1 (possibly zero), at most one from C_2, …, at most one from C_n, and as many as we wish from F. We think of the enumeration $C_1, C_2, \ldots, C_n$ as a multiset itself, hence invariant under permutation. Formally, $m \in F \mid C_1, C_2, \ldots, C_n$ if and only if one can

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

write m as $m_0 \uplus \{\!|x_i \mid i \in I|\!\}$, where all the elements of m_0 are in F, I is a subset of $\{1, 2, \ldots, n\}$ and for each $i \in I$, x_i is in C_i (implicitly, up to permutation of the x_i, or equivalently of the C_i).

Definition 8.3 (m-product, m-SRE). *Let X be a topological space. Call an* m-product *on X any subset of the form $F \mid C_1, C_2, \ldots, C_n$, where $n \in \mathbb{N}$, F is a closed subset of X, and $C_1, C_2, \ldots, C_n$ range over irreducible closed subsets of X.*

When F is empty, we shall also write this as simply $\mid C_1, C_2, \ldots, C_n$. When $n = 0$, we just write $F \mid$, and when $n = 0$ and $F = \emptyset$, we write this $\mid$.

An m-SRE *is any finite union of m-products.*

The proof of the following Proposition, as well as for most other results of this Section, are to be found in Appendix C.

Proposition 8.4. *Let X be a topological space. Then, the m-SREs are closed in $X^{\circledast}$, and the m-products are irreducible closed.*

If X is Noetherian, then every irreducible closed subset of $X^{\circledast}$ is an m-product, and every closed subset of $X^{\circledast}$ is an m-SRE.

Again, it is instructive to see how $X^{\circledast}$ embeds in its completion $\mathcal{S}(X^{\circledast})$.

Lemma 8.5 (Embedding). *Let X be a topological space. The closure $\eta^{\mathcal{S}}_{X^{\circledast}} \{\!|x_1, x_2, \ldots, x_n|\!\}$ of the multiset $\{\!|x_1, x_2, \ldots, x_n|\!\}$ in $X^{\circledast}$ is the m-product $\mid \eta^{\mathcal{S}}_X(x_1), \eta^{\mathcal{S}}_X(x_2), \ldots, \eta^{\mathcal{S}}_X(x_n)$.*

Proof. By Proposition 8.4, $\mid \eta^{\mathcal{S}}_X(x_1), \eta^{\mathcal{S}}_X(x_2), \ldots, \eta^{\mathcal{S}}_X(x_n)$ is (irreducible) closed and is clearly the downward closure of $\{\!|x_1, x_2, \ldots, x_n|\!\}$ with respect to $\leq^{\circledast}$. □

One can decide inclusion between m-products using Ψ again. This leads to the following algorithm.

Lemma 8.6. *Let X be a topological space. Inclusion between m-products can be checked in polynomial time, modulo an oracle testing inclusion of closed subsets of X.*

Explicitly, let $P = F \mid C_1, C_2, \ldots, C_m$ and $P' = F' \mid C'_1, C'_2, \ldots, C'_n$ be two m-products. Let $I = \{i_1, i_2, \ldots, i_k\}$ be the subset of those indices i, $1 \leq i \leq m$, such that $C_i \not\subseteq F'$.

Then, $P \subseteq P'$ if and only if $F \subseteq F'$ and there is an injective map $r \colon I \to \{1, 2, \ldots, n\}$ such that $C_i \subseteq C'_{r(i)}$ for every $i \in I$ – in other words, $\{\!|C_{i_1}, C_{i_2}, \ldots, C_{i_k}|\!\} \subseteq^{\circledast} \{\!|C'_1, C'_2, \ldots, C'_n|\!\}$.

It may not be immediately obvious why this leads to a polynomial time algorithm. The reason is the following observation, due to Simon Halfon (Halfon 2018, Corollary 7.14). Let G be the bipartite graph whose vertex set is the disjoint union of I and of $\{1, 2, \ldots, n\}$, and such that there is an edge from $i \in I$ to $j \in \{1, 2, \ldots, n\}$ if and only if $C_i \subseteq C'_j$. Finding r means finding a matching of G that covers all the vertices in I. Let $N(G)$ be the number of edges in any maximum matching of G. $N(G)$ can be computed in polynomial time, say by the Ford–Fulkerson algorithm (Cormen et al. 2001, Section 26.2), and r exists if and only if $N(G) \geq k$, where k is the cardinality of I.

We now turn to S-representations.

Theorem 8.7 (S-representation, multisets). *Let X be a Noetherian space, $X' = X^{\circledast}$, and $(S, [\![\_]\!], \trianglelefteq, \tau, \wedge)$ be an S-representation of X. Then, $(S', [\![\_]\!]', \trianglelefteq', \tau', \wedge')$ is an S-representation of X', where:*

(A) *S' is the collection of all* m-product notations*, that is, of all expressions of the form $A \mid u$, where A is a finite subset of S, and u is a multiset of elements of S. When $u = \{\!|b_1, \ldots, b_n|\!\}$, we also write $A \mid b_1, \ldots, b_n$ for $A \mid u$.*

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

(B) $[\![A \mid b_1, \ldots, b_n]\!]' = (\bigcup_{a \in A} [\![a]\!]) \mid [\![b_1]\!], \ldots, [\![b_n]\!]$.
(C) $A \mid u \trianglelefteq' A' \mid u'$ *if and only if* $A \trianglelefteq^\flat A'$ *and* $u_1 \trianglelefteq^\circledast u'$ *where* u_1 *is the subset of those elements* $a \in u$ *such that* $a \trianglelefteq a'$ *for no* $a' \in A'$.
(D) τ' *is* $\{\tau \mid \emptyset\}$.
(E) $\wedge'$ *is defined as follows. A* matching $f\colon \{1, \ldots, m\} \to \{1, \ldots, n\}$ *is any bijection from some subset of* $\{1, \ldots, m\}$ *(the* domain $\mathrm{dom}\, f$*) to some subset of* $\{1, \ldots, n\}$ *(the* codomain $\mathrm{cod}\, f$*). Then,* $(A \mid a_1, \ldots, a_m) \wedge' (A' \mid a'_1, \ldots, a'_n)$ *is the collection of all m-product notations of the form* $A'' \mid m_{1f} \uplus m_{2f} \uplus m_{3f}$*, where:*
- $A'' = \bigcup_{a \in A, a' \in A'} (a \wedge a')$;
- f *ranges over all matchings from* $\{1, \ldots, m\}$ *to* $\{1, \ldots, n\}$;
- m_{1f} *ranges over all multisets of the form* $\{\!| c_i \mid i \in \mathrm{dom}\, f |\!\}$ *where* $c_i \in a_i \wedge a'_{f(i)}$ *for every* $i \in \mathrm{dom}\, f$;
- m_{2f} *ranges over all multisets of the form* $\{\!| c_i \mid 1 \le i \le m, i \notin \mathrm{dom}\, f |\!\}$*, where* $c_i \in \bigcup_{a' \in A'} (a_i \wedge a')$ *for each* i*,* $1 \le i \le m$*,* $i \notin \mathrm{dom}\, f$;
- m_{3f} *ranges over all multisets of the form* $\{\!| c'_j \mid 1 \le j \le n, j \notin \mathrm{cod}\, f |\!\}$*, where* $c'_j \in \bigcup_{a \in A} (a \wedge a'_j)$ *for each* j*,* $1 \le j \le n$*,* $j \notin \mathrm{cod}\, f$.

As a final note to this section, Abdulla et al. (2004b) required a completion of $(A^\circledast)^*$, for some finite set A. We note that the elements of $\mathcal{S}((A^\circledast)^*_a)$ are exactly their *word language generators*, which we retrieve here in a principled way.

9. Completing Words, Prefix Topology

The word topology is not the only interesting topology on X^* that makes it Noetherian, assuming X Noetherian. The prefix topology is another (Goubault-Larrecq 2013, Exercise 9.7.36), and its specialization quasi-ordering is a form of the prefix ordering. We mention that topology because its specialization quasi-ordering is *never* a wqo, unless X is trivial. Also, this is the topology needed to decide reachability of sets defined by forbidden patterns in the so-called oblivious k-stack system model of Goubault-Larrecq (2010, Section 5).

The prefix topology is defined not just on X^*, but more generally on sets of *heterogeneous words*, that is, words whose letters are taken from possibly distinct spaces, depending on their position. This of course includes the case of words in X^*, but heterogeneous words are a natural generalization to consider, and incur no additional difficulty.

Let $X_1, X_2, \ldots, X_n, \ldots$ be countably many topological spaces. A heterogeneous word over these spaces is any tuple $(x_1, x_2, \ldots, x_m)$ in $X_1 \times X_2 \times \cdots \times X_m$, $m \in \mathbb{N}$. We write it as $x_1 x_2 \cdots x_m$ and call $m = |w|$ the *length* of the form $w = x_1 x_2 \cdots x_m$.

A *telescope* on $(X_n)_{n \ge 1}$ is a sequence $\mathcal{U} = U_0, U_1, \ldots, U_n, \ldots$ of opens, where U_n is open in $\prod_{i=1}^n X_i$ for each $n \in \mathbb{N}$, and such that $U_n X_{n+1} \subseteq U_{n+1}$ for every $n \in \mathbb{N}$. (We write $U_n X_{n+1}$ instead of $U_n \times X_{n+1}$. When $n = 0$, $\prod_{i=1}^n X_i$ just contains the empty word ϵ, so that $\prod_{i=1}^n X_i$ is just $\{\epsilon\}$ in that case, with the only possible topology. U_0 must be open in that, and that means that U_0 must be empty or equal to $\{\epsilon\}$ itself.)

A telescope is a *wide telescope* iff $U_n = \prod_{i=1}^n X_i$ for some $n \in \mathbb{N}$ – equivalently, for all sufficiently large $n \in \mathbb{N}$. Given any telescope $\mathcal{U} = U_0, U_1, \ldots, U_n, \ldots$ on $(X_n)_{n \ge 1}$, let $\lfloor \mathcal{U} \rangle$ be the set of heterogeneous words w over $X_1, X_2, \ldots, X_k, \ldots$, such that $w \in U_{|w|}$.

We write $\rhd_{n=1}^{+\infty} X_n$ for the space of all *heterogeneous words* over $(X_n)_{n \ge 1}$, that is, the disjoint union of all spaces $\prod_{i=1}^n X_i$, $n \in \mathbb{N}$, with the *prefix topology*, which is given by the trivial open $\emptyset$, plus all subsets of the form $\lfloor \mathcal{U} \rangle, \mathcal{U}$ a wide telescope on $(X_n)_{n \ge 1}$. One checks easily that those form a topology, and we shall say so explicitly in Proposition 9.1 below.

The point of the definition of the prefix topology is that its specialization quasi-ordering is the prefix quasi-ordering, defined by $a_1 a_2 \cdots a_m \le^\rhd b_1 b_2 \cdots b_n$, where $a_i, b_i \in X_i$ for all i, iff $m \le n$, $a_1 \le b_1$, $a_2 \le b_2$, …, and $a_m \le b_m$. (Here, $a_i \le b_i$ means that a_i is less than or equal to b_i in

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

the specialization quasi-ordering of X_i.) This is part of the following result, which appears as Goubault-Larrecq (2013, Exercise 9.7.36).

Proposition 9.1. *Let $X_1, X_2, \ldots, X_n, \ldots$ be countably many topological spaces. The prefix topology on $\triangleright_{n=1}^{+\infty} X_n$ is indeed a topology. Its specialization quasi-ordering is the prefix quasi-ordering $\leq^{\triangleright}$.*

If $X_1, X_2, \ldots, X_n, \ldots$ are all Noetherian, then $\triangleright_{n=1}^{+\infty} X_n$ is Noetherian.

Proof. Let us abbreviate $X_1X_2\cdots X_n$ as All_n, for each $n \in \mathbb{N}$. When $n = 0$, $All_0 = \{\varepsilon\}$.

It is easy to see that given wide telescopes $\mathcal{U}_i = U_{i0}, U_{i1}, \ldots, U_{in}, \ldots, i \in I$, the sequence $\bigcup_{i\in I} \mathcal{U}_i$ defined as $\bigcup_{i\in I} U_{i0}, \bigcup_{i\in I} U_{i1}, \ldots, \bigcup_{i\in I} U_{in}, \ldots$ is an infinite sequence of empty sets (if I is empty), or a wide telescope (if $I \neq \emptyset$) and $\lfloor \bigcup_{i\in I} \mathcal{U}_i \rangle = \bigcup_{i\in I} \lfloor \mathcal{U}_i \rangle$; moreover, if I is finite, then the sequence $\bigcap_{i\in I} \mathcal{U}_i$ defined as $\bigcap_{i\in I} U_{i0}, \bigcap_{i\in I} U_{i1}, \ldots, \bigcap_{i\in I} U_{in}, \ldots$ is also a wide telescope (when $I = \emptyset$, this is the wide telescope $All_0, All_1, \ldots, All_n, \ldots$), and $\lfloor \bigcap_{i\in I} \mathcal{U}_i \rangle = \bigcap_{i\in I} \lfloor \mathcal{U}_i \rangle$. Therefore, the prefix topology, as we have defined it, is indeed a topology.

Write $\sqsubseteq$ temporarily for the specialization quasi-ordering of $\triangleright_{n=1}^{+\infty} X_n$. For every telescope $\mathcal{U} = U_0, U_1, \ldots, U_n, \ldots$, if $a_1a_2\cdots a_m \in \lfloor \mathcal{U} \rangle$, and $a_1a_2\cdots a_m \leq^{\triangleright} b_1b_2\cdots b_n$, then $b_1b_2\cdots b_m$ is in U_m, since U_m is open hence upward-closed in $\leq_m$. Since $U_mX \subseteq U_{m+1}$ by the definition of telescopes, $b_1b_2\cdots b_mb_{m+1}$ is in U_{m+1}, and by an easy induction, $b_1b_2\cdots b_mb_{m+1}\cdots b_n$ is in U_n. So $\mathcal{U}$ is upward-closed in $\leq^{\triangleright}$: $w \leq^{\triangleright} w'$ implies $w \sqsubseteq w'$.

Conversely, assume $w \sqsubseteq w'$, where $w = a_1a_2\cdots a_m$ and $w' = b_1b_2\cdots b_n$. We shall examine various wide telescopes $\mathcal{U}$ such that $w \in \lfloor \mathcal{U} \rangle$ and draw consequences from the fact that $w' \in \lfloor \mathcal{U} \rangle$. Considering the telescope $\emptyset, \ldots, \emptyset, All_m, All_{m+1}, \ldots$, one sees that $m \leq n$. Considering the telescopes $\emptyset, \ldots, \emptyset, U_m, All_{m+1}, All_{m+2}, \ldots$, where U_m is an arbitrary open set of X^m of which w is a member, one sees that $b_1b_2\cdots b_m$ is in U_m, so $a_1a_2\cdots a_m$ is less than or equal to $b_1b_2\cdots b_m$ in the specialization quasi-ordering $\leq_1 \times \leq_2 \times \cdots \times \leq_m$ of $X_1X_2\cdots X_m$. As a consequence, $w \leq^{\triangleright} w'$.

Let us show that $X = \triangleright_{n=1}^{+\infty} X_n$ is Noetherian, assuming that $X_1, X_2, \ldots, X_n, \ldots$, all are. For every non-empty wide telescope $\mathcal{U} = U_0, U_1, \ldots, U_n, \ldots$, there is a least number m such that U_m is non-empty, and a least number n such that $U_n = All_n$. Moreover, $m \leq n$. Call m the *small end* $m(\mathcal{U})$ of $\mathcal{U}$, n its *big end* $n(\mathcal{U})$. If $\lfloor \mathcal{U} \rangle \subseteq \lfloor \mathcal{V} \rangle$, then $m(\mathcal{U}) \geq m(\mathcal{V})$ (consider any word of length $m(\mathcal{U})$ in $\lfloor \mathcal{U} \rangle$), and $n(\mathcal{U}) \geq n(\mathcal{V})$ (otherwise, consider any word of length $n(\mathcal{U})$ that is not in $\lfloor \mathcal{V} \rangle$). It follows that, in any infinite ascending chain $\lfloor \mathcal{U}_1 \rangle \subseteq \lfloor \mathcal{U}_2 \rangle \subseteq \cdots \subseteq \lfloor \mathcal{U}_k \rangle \subseteq \cdots$, all small ends coincide, and all big ends coincide, for all k large enough, say $k \geq p$. Let m be this common small end, n be the common big end. Then, for each $k \geq p$, $\mathcal{U}_k$ is a telescope of the form $\emptyset, \ldots, \emptyset, U_{km}, U_{k(m+1)}, \ldots, U_{k(n-1)}, All_n, All_{n+1}, \ldots$ In addition, for each j with $m \leq j < n$, $U_{pj} \subseteq U_{(p+1)j} \subseteq \cdots \subseteq \cdots U_{kj} \subseteq \cdots$ is an infinite ascending chain of opens in $X_1X_2\cdots X_j$. The latter is Noetherian, so the chain stabilizes, say at $k_j \geq p$. Therefore, $\lfloor \mathcal{U}_1 \rangle \subseteq \lfloor \mathcal{U}_2 \rangle \subseteq \cdots \subseteq \lfloor \mathcal{U}_k \rangle \subseteq \cdots$ stabilizes at $\max(p, k_m, k_{m+1}, \ldots, k_{n-1})$. This holds for ascending chains of opens that exclude the empty open subset; the general case is easy. It follows that $\triangleright_{n=1}^{+\infty} X_n$ is Noetherian. $\square$

When $X_1 = X_2 = \cdots = X_n = \cdots$ are the same space X, we write $X^{\triangleright}$ for the space $\triangleright_{n=1}^{+\infty} X_n$. Although it has the same elements as X^*, it has a definitely distinct topology, for example, while $\leq^*$ is wqo when the specialization ordering $\leq$ of X is, $\leq^{\triangleright}$ is well founded but not well, as soon as X contains two incomparable elements a and b. Indeed, in this case $a, ba, bba, bbba, \ldots$ is an infinite antichain.

In order to characterize the completion $\mathcal{S}(\triangleright_{n=1}^{+\infty} X_n)$, we define the subset $\lceil F_1F_2\cdots F_k \rangle$ of $\triangleright_{n=1}^{+\infty} X_n$, where $k \in \mathbb{N}$ and each F_i is closed in X_i, as the set of all heterogeneous words $a_1a_2\cdots a_m$ of $\prod_{i=1}^{m} X_i$ such that $m \leq k$, $a_1 \in F_1, a_2 \in F_2, \ldots, a_m \in F_m$. (When $k = 0$, $\lceil F_1F_2\cdots F_k \rangle$ is just $\{\epsilon\}$.) The following, as well as other results of this Section, are proved in Appendix D.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Proposition 9.2. *Let $X_1, X_2, \ldots, X_n, \ldots$ be countably many topological spaces. The sets of the form $\lceil F_1 F_2 \cdots F_n\rangle$, where each F_i is closed in X_i, form a subbase of closed sets for $\rhd_{n=1}^{+\infty} X_n$: these sets are closed, and every closed subset is an intersection of finite unions of such sets.*

Lemma 9.3. *Let $X_1, X_2, \ldots, X_n, \ldots,$ be countably many topological spaces. The subsets of the form $\lceil C_1 C_2 \cdots C_n\rangle$, where C_i is irreducible closed in X_i for each i, $1 \le i \le n$, are irreducible closed in $\rhd_{n=1}^{+\infty} X_n$.*

This is enough to state how $\rhd_{n=1}^{+\infty} X_n$ embeds in its completion.

Lemma 9.4 (Embedding). *Let $X_1, X_2, \ldots, X_n, \ldots,$ be countably many topological spaces, and $Y = \rhd_{n=1}^{+\infty} X_n$. The closure $\eta_Y^{\mathcal{S}}(x_1 x_2 \cdots x_n)$ of the word $x_1 x_2 \cdots x_n$ in $X = \rhd_{n=1}^{+\infty} X_n$ is $\lceil \eta_X^{\mathcal{S}}(x_1)\eta_X^{\mathcal{S}}(x_2)\cdots\eta_X^{\mathcal{S}}(x_n)\rangle$.*

Proof. By Lemma 9.3, $\lceil \eta_X^{\mathcal{S}}(x_1)\eta_X^{\mathcal{S}}(x_2)\cdots\eta_X^{\mathcal{S}}(x_n)\rangle$ is irreducible closed and contains $x_1 x_2 \cdots x_n$, so must contain $\eta_Y^{\mathcal{S}}(x_1 x_2 \cdots x_n)$. Conversely, every element of $\lceil \eta_X^{\mathcal{S}}(x_1)\eta_X^{\mathcal{S}}(x_2)\cdots\eta_X^{\mathcal{S}}(x_n)\rangle$ is a ($\le^{\rhd}$-) prefix of $x_1 x_2 \cdots x_n$, and must therefore be in $\eta_Y^{\mathcal{S}}(x_1 x_2 \cdots x_n)$. □

There is just one extra irreducible closed subset in $\rhd_{n=1}^{+\infty} X_n$, unless some X_n is empty:

Lemma 9.5. *Let $X_1, X_2, \ldots, X_n, \ldots,$ be countably many non-empty topological spaces. The whole space $\rhd_{n=1}^{+\infty} X_n$ is irreducible closed in itself.*

So we obtain the following description of all irreducible closed subsets.

Lemma 9.6. *Let $X_1, X_2, \ldots, X_n, \ldots,$ be countably many non-empty topological spaces. The only irreducible closed subsets of $\rhd_{n=1}^{+\infty} X_n$ are $\rhd_{n=1}^{+\infty} X_n$ itself, and the subsets of the form $\lceil C_1 C_2 \cdots C_n\rangle$, where C_i is irreducible closed in X_i for each i, $1 \le i \le n$.*

This suggests that $\mathcal{S}(\rhd_{n=1}^{+\infty} X_n)$ coincides with $\rhd_{n=1}^{+\infty} \mathcal{S}(X_n)$, with a new top element $\top$ added, at least up to isomorphism. For any space Y, let $Y^\top$ be the space obtained by adding a fresh element $\top$ to Y, and whose closed subsets are those of Y, plus $Y^\top$ itself. The specialization quasi-ordering of $Y^\top$ is given by: $y \preceq^\top y'$ iff $y, y' \in Y$ and $y \preceq y'$, or $y' = \top$, where $\preceq$ is the specialization quasi-ordering of Y.

Proposition 9.7. *Let $X_1, X_2, \ldots, X_n, \ldots,$ be countably many non-empty topological spaces. The map $i\colon (\rhd_{n=1}^{+\infty} \mathcal{S}(X_n))^\top \to \mathcal{S}(\rhd_{n=1}^{+\infty} X_n)$ that sends $\top$ to $\rhd_{n=1}^{+\infty} X_n$ and the word $C_1 C_2 \cdots C_n$ (where $C_i \in \mathcal{S}(X_i)$ for each i) to $\lceil C_1 C_2 \cdots C_n\rangle$ is an order isomorphism and a homeomorphism.*

To complete the picture, we examine the case where some of the spaces X_n are empty. Taking n to be the largest index such that $X_1, \ldots, X_n$ are non-empty (and 0 if every X_i is empty), we then write $X_1 \rhd X_2 \rhd \cdots \rhd X_n$, or $\rhd_{k=1}^{n} X_k$, instead of $\rhd_{k=1}^{+\infty} X_k$. Since there cannot be any $(n+1)$st letter, this is a space of words of length at most n. Clearly as well $\rhd_{k=1}^{n} X_k$ then does not depend on the actual spaces $X_{n+1}, X_{n+2}, \ldots$, provided X_{n+1} is empty, which justifies the notation.

Lemma 9.8. *Let $X_1, X_2, \ldots, X_n$ be non-empty topological spaces. The only irreducible closed subsets of $\rhd_{k=1}^{n} X_k$ are the subsets of the form $\lceil C_1 C_2 \cdots C_m\rangle$, where C_i is irreducible closed in X_i for each i, $1 \le i \le m$, and $m \le n$.*

We then obtain an isomorphism as in Proposition 9.7, without the need to add a top element $\top$.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Proposition 9.9. *Let $X_1, X_2, \ldots, X_n$ be non-empty topological spaces. The map $i\colon \rhd_{k=1}^{n} \mathcal{S}(X_k) \to \mathcal{S}(\rhd_{k=1}^{n} X_k)$ that sends the word $C_1C_2 \cdots C_k$ (where $k \leq n$ and $C_i \in \mathcal{S}(X_i)$ for each i) to $\lceil C_1C_2 \cdots C_k\rangle$ is an order isomorphism and a homeomorphism.*

We therefore obtain

Theorem 9.10 (S-representation, prefix). *Let $X_1, X_2, \ldots, X_n, \ldots$ be countably many Noetherian spaces, $X' = \rhd_{n=1}^{+\infty} X_n$, and $(S_i, [\![\_]\!]_i, \trianglelefteq_i, \tau_i, \wedge_i)$ be an S-representation of X_i for each $i \geq 1$. Assume that the disjoint sum $\coprod_{i=1}^{+\infty} S_i$ is recursively enumerable that $a \trianglelefteq_i b$ is decidable in a, b, i, that τ_i is computable in i, and that $a \wedge_i b$ is computable in a, b, i.*

Then $(S', [\![\_]\!]', \trianglelefteq', \tau', \wedge')$ is an S-representation of X', where:

(A) S' is the set of all heterogeneous words over $S_1, S_2, \ldots, S_n, \ldots$, plus a fresh element ω in case no S_n is empty (i.e., no X_n is empty).
(B) $[\![a_1a_2 \cdots a_n]\!]' = \lceil [\![a_1]\!]_1 [\![a_2]\!]_2 \cdots [\![a_n]\!]_n\rangle$ where $a_1 \in S_1$, $a_2 \in S_2$, $\ldots$, $a_n \in S_n$, and $[\![\omega]\!] = \rhd_{n=1}^{+\infty} X_n$ (if no X_n is empty).
(C) $\trianglelefteq'$ is defined by: $u \trianglelefteq' \omega$ for all $u \in S'$ and $\omega \trianglelefteq' u$ for no word $u \neq \omega$ in S' (in case no X_n is empty), and $a_1a_2 \cdots a_m \trianglelefteq' a'_1a'_2 \cdots a'_n$ iff $m \leq n$, $a_1 \trianglelefteq_1 a'_1$, $a_2 \trianglelefteq_2 a'_2, \ldots,$ and $a_m \trianglelefteq_m a'_m$.
(D) τ' is $\{\omega\}$ if no X_n is empty, $\{a_1a_2 \cdots a_n \mid a_1 \in \tau_1, a_2 \in \tau_2, \ldots, a_n \in \tau_n\}$ otherwise, where n is the largest index such that X_n is non-empty.
(E) $\wedge'$ is defined by: $\omega \wedge' u' = \{u'\}$, $u \wedge' \omega = \{u\}$ (if no X_n is empty), and $a_1a_2 \cdots a_m \wedge' a'_1a'_2 \cdots a'_n = \{c_1c_2 \cdots c_{\min(m,n)} \mid c_1 \in a_1 \wedge_1 a'_1, c_2 \in a_2 \wedge_2 a'_2, \ldots, c_{\min(m,n)} \in a_{\min(m,n)} \wedge_{\min(m,n)} a'_{\min(m,n)}\}$.

Proof. Follows easily from Proposition 9.7 in case no X_n is empty, or from Proposition 9.9 otherwise. □

When $X_1 = X_2 = \ldots = X_n = \ldots$ are all the same space, one can drop the subscripts to S, $\trianglelefteq$, τ, $\wedge$. Then, $u \trianglelefteq' u'$ is decidable in polynomial time modulo an oracle for $\trianglelefteq$.

Remark 9.11. We have already seen an example of spaces that are Noetherian, but not wqo in their specialization preordering, for example, $\mathbb{P}(X)$ where X is Noetherian. The construction $X^{\rhd}$ is another example: while $X^{\rhd}$ is Noetherian for X Noetherian, the prefix ordering $\leq^{\rhd}$ is not wqo, even if $\leq$ is wqo.

10. Completing Finite Trees in a Simple Case

The case of finite trees is by far the most complex one. We start with a simple case: that of *ranked* trees, whose vertices are decorated by function symbols from a *finite* set, with equality as ordering. The ordering on trees is the so-called homeomorphic embedding $\preceq$, a wqo by Kruskal's tree theorem (Kruskal 1960). In that case, the topology will be the Alexandroff topology, and irreducible closed subsets will be ideals. We shall also take advantage of the assumptions to give short, automata-theoretic proofs of some of our results.

The completions of sets of trees over a finite set of function symbols were already considered in Wies et al. (2010), where it was used to decide coverability for depth-bounded processes without requiring one to know the depth bound in advance, and in Goubault-Larrecq and Schmitz (2016), where it was used to decide piecewise testable separability of regular tree languages.

The more general case of unranked trees with function symbols taken from a Noetherian space will be dealt with in Section 11. This will definitely be more complicated but will share many similarities with the present case – apart from the fact that we will not be able to use automata-theoretic methods.

Let Σ be a finite *signature*, namely a finite set of so-called *function symbols* $f, g, \ldots$, each coming with a natural number called its *arity*. Let Σ_r be the subset of those elements of Σ that have arity exactly r.

The set $\mathcal{T}(\Sigma)$ of *terms* over Σ (Baader and Nipkow 1998) is the smallest such that, for every $f \in \Sigma_r$, for all $t_1, \ldots, t_r \in \mathcal{T}(\Sigma)$, $f(t_1, \ldots, t_r)$ is in $\mathcal{T}(\Sigma)$. By $f(t_1, \ldots, t_n)$, we mean the tree whose root is labeled f and has a list of r subtrees $t_1, \ldots, t_r$, from left to right.

Our terms are ground: there is no variable involved here, although one may code variables as specific constants, that is, as specific function symbols of arity 0.

The embedding ordering $\preceq$ is defined by $s \preceq t$ if and only if:

— either $t = g(t_1, \ldots, t_n)$ and $s \preceq t_j$ for some j, $1 \leq j \leq n$;
— or $s = f(s_1, \ldots, s_m)$, $t = g(t_1, \ldots, t_n)$, $f = g$ (hence, $m = n$) and $s_1 \preceq t_1, \ldots, s_n \preceq t_n$.

Since the canonical S-representation for finite words under the word topology (or the $\leq^*$ quasi-ordering) consists in certain regular expressions, we shall similarly define an S-representation for $\mathcal{T}(\Sigma)$ as certain tree regular expressions (Comon et al. 2004, Section 2.2).

We define *simple tree regular expressions* (over Σ), a.k.a. *STRE*s, by the following abstract syntax:

$$S ::= 0 \mid P \mid S + S \qquad P ::= f^{\bar{?}}(S, \ldots, S) \mid \mathfrak{C}^{\bar{*}}.S$$
$$\mathfrak{C} ::= 0 \mid A \mid \mathfrak{C} + \mathfrak{C} \qquad A ::= f(S_\Box, \ldots, S_\Box) \qquad S_\Box ::= S \mid \Box$$

where $f \in \Sigma_r$ in $f^{\bar{?}}(P_1, \ldots, P_r)$ and in $f(S_{\Box 1}, \ldots, S_{\Box r})$, the sum operation $+$ is associative and commutative (we shall sometimes write $\sum_{i=1}^m P_i$ for $P_1 + \cdots + P_m$) with 0 denoting the empty sum, and $\Box \notin \Sigma$ is a placeholder called the *hole*. Note that $\Box$ is not meant to denote a family of placeholders, rather a single one. The extended trees over the signature $\Sigma \cup \{\Box\}$, where $\Box$ has arity 0, are called *contexts*.

The standard notations for $\bar{?}$ and $\bar{*}$ are ? and $*$. We want to distinguish $\bar{?}$ visually from ?, since we will need both in Section 11, and similarly for $\bar{*}$ and $*$.

The STREs of the form P are called *tree pre-products*. Among them, the *tree-products* will be our notations for ideals. They will satisfy additional constraints, which we shall define later (Definition 10.5).

To define the *semantics* of STREs, we write $t \in S$ as an abbreviation for "t is in the language of S" and define this language by structural induction on S.

Accordingly, $t \in f^{\bar{?}}(S_1, \ldots, S_n)$ if and only if either t is of the form $f(t_1, \ldots, t_n)$ with $t_i \in S_i$ for every i, $1 \leq i \leq n$, *or* if $t \in \bigcup_{i=1}^n S_i$. The latter is necessary for S to denote a downward-closed language.

As the notation suggests, for $S = P_1 + \cdots + P_m$, $t \in S$ iff $t \in P_j$ for some j, $1 \leq j \leq m$, and the language of 0 is empty.

The productions of $\mathfrak{C}$, A, and $S_\Box$ serve to form *iterators* $\mathfrak{C}^{\bar{*}}.S$. The language of $A = f(S_{\Box 1}, \ldots, S_{\Box n})$ consists of those contexts in $\mathcal{T}(\Sigma \cup \{\Box\})$ of the form $f(c_1, \ldots, c_n)$ where $c_i \in S_{\Box i}$ for every i. In turn, $c \in S_\Box$ if and only if either $S_\Box = \Box$ and c is the trivial context $\Box$, or $S_\Box$ is an STRE S, c is a tree t in $\mathcal{T}(\Sigma)$, and $t \in S$. The language of $\mathfrak{C} = A_1 + \cdots + A_m$ is the union of the languages of A_j, $1 \leq j \leq m$.

Intuitively, the language of $\mathfrak{C}^{\bar{*}}.S$ should consist of all trees obtained by applying contexts in $\mathfrak{C}$, repeatedly, until one reaches a tree in S. For example, $(f(\Box))^{\bar{*}}.a^{\bar{?}}$ will recognize all trees of the form $f^n(a)$, $n \in \mathbb{N}$. There are however two catches.

(1) The first one has to do with patterns A where $\Box$ occurs more than once: as usual with tree regular expressions, in replacing $\Box$ by a tree from S, several occurrences of $\Box$ can be replaced by *different* trees from S. Hence, $(f(\Box, \Box))^{\bar{*}}.(a^{\bar{?}} + b^{\bar{?}})$ consists of all binary-branching trees with inner nodes labeled f and leaves labeled a or b, including $f(f(a, a), a)$

and $f(f(b, b), b)$ but also $f(f(a, b), a))$ or $f(f(b, b), a)$ among others. (We assume f binary, and a and b of arity 0.) For a context c, and a set of trees S, accordingly, we shall write $c[S]$ for the set of trees obtained from c by replacing each occurrence of $\Box$ by a (possibly different) tree in S.

(2) The second catch has to do with downward closure. It is tempting to define the trees of $\overline{\mathfrak{C}^*}.S$ as those in $c_1[\cdots[c_k[S]]\cdots]$, for some $k \in \mathbb{N}$ and some $c_1, \ldots, c_k \in \mathfrak{C}$. However, there are cases where that language would fail to be downward-closed, for example, $\overline{(f(a^{\overline{?}}, \Box))^*}.b^{\overline{?}}$ would contain $f(a, b)$ but not a, according to that semantics.

We repair that as follows. For $A = f(S_{\Box 1}, \ldots, S_{\Box n})$, define args A, the *argument support* of A, as the set of trees $t \in \mathcal{T}(\Sigma)$ such that some context $f(\cdots, t, \cdots)$ (i.e., with one of its arguments equal to t) is in the language of A. Alternatively, if those $S_{\Box i}$, $1 \leq i \leq n$, that are different from $\Box$ define non-empty languages, then args A is the union of those languages; if some $S_{\Box i} \neq \Box$ has an empty language, then args $A = \emptyset$. Hence, for example, args $f(\Box, \Box) = \emptyset$, args $f(a^{\overline{?}}, \Box) = a^{\overline{?}} = \{a\}$, and args $f'(a^{\overline{?}}, \Box, 0) = \emptyset$. For $\mathfrak{C} = A_1 + \cdots + A_m$, let args $\mathfrak{C} = \bigcup_{j=1}^{m}$ args A_j.

For every $c \in \mathfrak{C}$, let us write $c^{\overline{?}}[S]$ for $c[S] \cup S$. We are now ready to define the language of $\overline{\mathfrak{C}^*}.S$, as the language of trees in $c_1^{\overline{?}}[\cdots[c_k^{\overline{?}}[S \cup \text{args}\ \mathfrak{C}]]\cdots]$, for some $k \in \mathbb{N}$ and some $c_1, \ldots, c_k \in \mathfrak{C}$ – in writing $S \cup \text{args}\ \mathfrak{C}$, we equate the STRE S with the language it defines.

Proposition 10.1. *Every STRE defines a downward-closed language of $\mathcal{T}(\Sigma)$ with respect to $\preceq$. Conversely, every downward-closed language of $\mathcal{T}(\Sigma)$ with respect to $\preceq$ is the language of some STRE.*

Proof. For the first part, we use induction on the size of the STRE, and the main point consists in checking that if $t \in \overline{\mathfrak{C}^*}.S$ then any smaller tree s (w.r.t. $\preceq$) is also in $\overline{\mathfrak{C}^*}.S$. By induction hypothesis, S is downward-closed. We use a secondary induction on the k used in the definition of the language of $\overline{\mathfrak{C}^*}.S$ as $c_1^{\overline{?}}[\cdots[c_k^{\overline{?}}[S \cup \text{args}\ \mathfrak{C}]]\cdots]$. If $k = 0$, then $t \in S \cup \text{args}\ \mathfrak{C}$, hence $s \in S \cup \text{args}\ \mathfrak{C}$ as well since both S and args $\mathfrak{C}$ are downward-closed – args $\mathfrak{C}$ is downward-closed by the outer induction hypothesis. Otherwise, $k \geq 1$. If t is in $c_2^{\overline{?}}[\cdots[c_k^{\overline{?}}[S \cup \text{args}\ \mathfrak{C}]]\cdots]$, then we conclude by the inner induction hypothesis directly. Otherwise, one of the summands in c_1 is of the form $f(S_{\Box 1}, \ldots, S_{\Box n})$, and $t = f(t_1, \ldots, t_n)$. If s is smaller than some t_i, then either $S_{\Box i} = \Box$ and $t_i \in c_2^{\overline{?}}[\cdots[c_k^{\overline{?}}[S \cup \text{args}\ \mathfrak{C}]]\cdots]$, which allows us to conclude by the inner induction hypothesis; or $S_{\Box i}$ is of the form P and is downward-closed by the outer induction hypothesis. If instead $s = f(s_1, \ldots, s_n)$ where s_i is smaller than t_i for each i, we conclude similarly that $s_i \in \overline{\mathfrak{C}^*}.S$ for each position i such that $S_{\Box i} = \Box$. For all other positions i, $S_{\Box i}$ is an STRE, which is downward-closed by the outer induction hypothesis, so $s_i \in S_{\Box i}$. Hence, s is in $c_1[\overline{\mathfrak{C}^*}.S]$, and therefore in $\overline{\mathfrak{C}^*}.S$.

For the converse direction, let L be a downward-closed language. The complement $\mathcal{T}(\Sigma) \smallsetminus L$ of L is upward-closed, and since $\preceq$ is a wqo, $\mathcal{T}(\Sigma) \smallsetminus L$ can be written as the upward closure $\uparrow\{t_1, t_2, \ldots, t_n\}$ of finitely many trees. For each i, $\uparrow t_i$ is easily seen to be recognizable by a finite (bottom-up) tree automaton. Since finite unions and complements of recognizable languages are recognizable, L is recognized by some finite tree automaton $\mathcal{A}$.

We now convert $\mathcal{A}$ to an STRE. In general, we describe a procedure that converts any (ε-free) finite tree automaton $\mathcal{A}$ to an STRE whose language is the downward closure $\downarrow L(\mathcal{A})$ of the language recognized by $\mathcal{A}$. This is best explained on an example: see Figure 3(a), where there is one transition a of arity 0 (from no state) to state q_1, one binary transition f from the pair of states q_1, q_2 to q_3, and so on. In textual form, we write these transitions as rewrite rules (Comon et al. 2004): $a \to q_1$, $f(q_1, q_2) \to q_3$, $h(q_3, q_4) \to q_2$, $d \to q_3$, $g(q_4) \to q_4$, $b \to q_4$. A tree t is *recognized at a state s* if and only if $t \to^* s$, using the rewrite rules of $\mathcal{A}$. There is a set of final states, marked

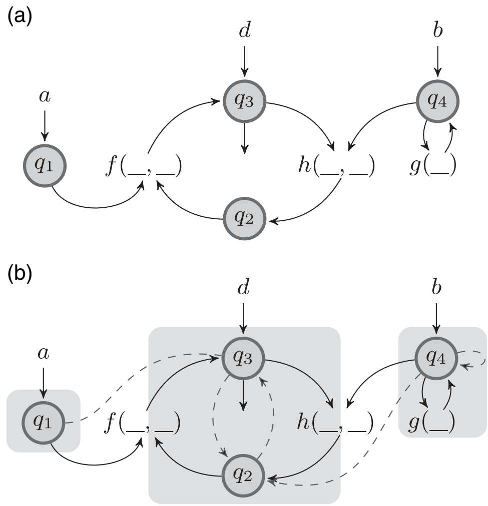

Figure 3. Converting tree automata to STREs. (a) Initial tree automaton. (b) ε-strongly connected components.

with an outgoing arrow with dangling end: in our example, just q_3. The language $L(\mathcal{A})$ of $\mathcal{A}$ is the set of trees recognized at some final state.

We first extend our automaton with ε-transitions. An ε-transition from s to s' will be drawn as a dashed arrow, see Figure 3(b), and is just a rewrite rule of the new form $s \to s'$. This implies that every tree recognized at s is also recognized at s'. For each transition, say $f(s_1, s_2, \ldots, s_n) \to s$, of $\mathcal{A}$, we add n ε-transitions $s_1 \to s$, $s_2 \to s, \ldots, s_n \to s$. (To make things clear, we are assuming that $\mathcal{A}$ does not originally contain any ε-transition.) Call the resulting automaton $\downarrow \mathcal{A}$. It is an easy exercise to show that $L(\downarrow \mathcal{A}) = \downarrow L(\mathcal{A})$.

There is a graph underlying $\downarrow \mathcal{A}$, whose vertices are the states of $\downarrow \mathcal{A}$, and whose edges are the ε-transitions. Build its strongly connected components: on Figure 3(b), they are shown against a gray background. By construction, any two states in the same strongly connected component C recognize exactly the same trees, so it makes sense to talk of the language $L_C(\downarrow \mathcal{A})$ of those trees recognized at any state of C. Let $C \to C'$ if and only if $s \to s'$ for some $s \in C$, $s' \in C'$, $C \neq C'$. The strict ordering $<$ defined as the transitive closure $\to^+$ is well founded, and we shall build an STRE S_C whose language is $L_C(\downarrow \mathcal{A})$, by induction along $<$.

If C is a trivial strongly connected component (one state s, no self-edge), then enumerate its incoming non-ε transitions $f_i(s_{i1}, s_{i2}, \ldots, s_{in_i}) \to s$, $1 \leq i \leq m$. Let S_{ij} be an STRE whose language is the set of trees recognized at s_{ij}, which is given by induction hypothesis. Then $S_C = \sum_{i=1}^m f_i^{?}(S_{i1}, S_{i2}, \ldots, S_{in_i})$ is the desired STRE. For instance, the set of trees recognized at the leftmost state q_1 is the language of $a^{?}$.

If C is a non-trivial strongly connected component, then enumerate the non-ε transitions $f_i(s_{i1}, s_{i2}, \ldots, s_{in_i}) \to s_i$, $1 \leq i \leq m$ whose end state s_i is in C. For each pair i, j, if s_{ij} is in C, then let $S_{\Box ij} = \Box$; otherwise, let $S_{\Box ij}$ be an STRE whose language is the set of trees recognized at s_{ij}, which we obtain by induction hypothesis. It is not too hard to see that $S_C = (\sum_{i=1}^m f_i(S_{\Box i1}, S_{\Box i2}, \ldots, S_{\Box in_i}))^{\overline{*}}.0$ is an STRE that suits our needs. For example, the rightmost strongly connected component $\{q_4\}$ yields the STRE $S_4 = (b + g(\Box))^{\overline{*}}.0$. One might have expected an STRE of the more intuitive form $(g(\Box))^{\overline{*}}.b^{\overline{?}}$; however, note that they define exactly the same language.

Finally, $\downarrow L(A)$ is the union of the languages of the strongly connected components containing a final state; in our example, the strongly connected component in the middle yields the final STRE $(d + f(a^{\overline{?}}, \Box) + h(\Box, S_4))^{\overline{*}}.0$. $\Box$

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

$$
\begin{array}{llcl}
(R1) & P + P' & \to_1 & P' \quad \text{if } P \subseteq P' \\
(R2) & A + A' & \to_1 & A' \quad \text{if } A \subseteq A' \\
(R3) & 0 + P & \to_1 & P \\
(R4) & 0 + A & \to_1 & A \\
(R5) & 0^{\overline{*}}.S & \to_1 & S \\
(R6) & (\mathfrak{C} + f(S_1, \cdots, S_n))^{\overline{*}}.S & \to_1 & \mathfrak{C}^{\overline{*}}.(S + f^{\overline{?}}(S_1, \cdots, S_n)) \\
(R7) & f^{\overline{?}}(\vec{S}_1, 0, \vec{S}_2) & \to_1 & 0 \\
(R8) & f^{\overline{?}}(\vec{S}_1, S + S', \vec{S}_2) & \to_1 & f^{\overline{?}}(\vec{S}_1, S, \vec{S}_2) + f^{\overline{?}}(\vec{S}_1, S', \vec{S}_2) \\
(R9) & f(\vec{S}_{\Box 1}, 0, \vec{S}_{\Box 2}) & \to_1 & 0 \\
(R10) & f(\vec{S}_{\Box 1}, S + S', \vec{S}_{\Box 2}) & \to_1 & f(\vec{S}_{\Box 1}, S, \vec{S}_{\Box 2}) + f(\vec{S}_{\Box 1}, S', \vec{S}_{\Box 2}) \\
(R11) & \mathfrak{C}^{\overline{*}}.0 & \to_1 & 0 \quad \text{if } \mathfrak{C} = \sum_{i=1}^{m} f_i(\Box, \cdots, \Box) \text{ and no } f_i \text{ has arity } 0 \\
(R12) & \mathfrak{C}^{\overline{*}}.(S + S') & \to_1 & \mathfrak{C}^{\overline{*}}.S + \mathfrak{C}^{\overline{*}}.S' \quad \text{if } \mathfrak{C} \text{ is } \Box\text{-linear}
\end{array}
$$

Figure 4. The rewrite relation $\to_1$.

We characterize the STREs that define ideals of $\mathcal{T}(\Sigma)$ with respect to $\preceq$. Let us define a rewrite relation $\to_1$ on STREs that moves all $+$ signs to the outside: for a $\to_1$-normal STRE $S = P_1 + \cdots + P_m$, each P_i will be irreducible, hence S will be an ideal, that is, an irreducible closed subset, if and only if $m = 1$. (Recall that $\mathcal{S}(P_a) = \mathbf{I}(P)$ for every poset P.)

The rewrite relation $\to_1$ is defined in Figure 4. Recall that $+$ is understood modulo associativity and commutativity. Letters matter, too: S, S', S_1, …, S_n are STREs, while P, P' are those special STREs of the form $f^{\overline{?}}(S_1, \ldots, S_n)$ or $\mathfrak{C}^{\overline{*}}.S$, etc. In particular, rule ($R6$) applies provided the pattern $f(S_1, \ldots, S_n)$ does not contain $\Box$ at all. Similarly, in rule ($R10$), S and S' cannot contain $\Box$.

For rule ($R12$), we need some auxiliary definitions.

Definition 10.2. *A pattern $A = f(S_{\Box 1}, \ldots, S_{\Box n})$ is* $\Box$-linear *if and only if at most one $S_{\Box i}$ is the hole $\Box$.*

Writing $\mathfrak{C}$ as $A_1 + \cdots + A_m$, we say that $\mathfrak{C}$ is $\Box$-linear if and only if every non-empty A_i is $\Box$-linear.

The $\Box$-linearity restriction imposed on the last rule is needed for the following to hold.

Lemma 10.3. *If $S \to_1^* S'$ then S and S' define the same language.* $\Box$

Lemma 10.4. *Every STRE S has a normal form with respect to $\to_1$.*

Proof. Using Bachmair and Plaisted's associative path ordering $>_{apo}$ (Bachmair and Plaisted 1985) on a precedence where $+$ is minimal, $f > f^{\overline{?}}$ for each symbol f, and the $(\_)^{\overline{*}}.\_$ operator has lexicographic status, we see that $\to_1$ is even a terminating relation: every sequence of rewrite steps terminates. (Bachmair and Plaisted's ordering has been improved upon many times but is sufficient in the case of just one associative commutative symbol $+$.) $\Box$

Definition 10.5. *A* tree-product *is any $\to_1$-normal tree pre-product P.*

Lemma 10.6. *Every ideal, that is, every irreducible closed subset of $\mathcal{T}(\Sigma)$ is the language of some tree-product.*

Proof. By Proposition 10.1, an ideal I is the language of some STRE S. S has a $\to_1$-normal form by Lemma 10.4, write it $P_1 + \cdots + P_m$. Since I is non-empty, $m \geq 1$, and since I is irreducible closed, it is included in, hence equal to, the language of some P_i. $\Box$

Conversely, we check that the language of every tree-product is irreducible closed. In the special case that we are in, it is easier to show that they are directed sets.

Let us introduce some additional notation.

Definition 10.7. *A pattern* $A = f(S_{\Box 1}, \ldots, S_{\Box n})$ *is* $\Box$-generated *if and only if at least one* $S_{\Box i}$ *is the hole* $\Box$; *it is* empty *if and only if some* $S_{\Box i}$ *is different from* $\Box$ *and has an empty language.*

Writing C *as* $A_1 + \cdots + A_m$, *we say that* C *is* $\Box$*-generated if and only if every non-empty* A_i *is* $\Box$*-generated and is empty if and only if every* A_i *is empty.*

By inspection of the rules defining $\rightarrow_1$, we see

Lemma 10.8. *The tree-products are exactly the STREs of the form:*

- $f^{\overline{?}}(P_1, \ldots, P_n)$ *where* $P_1, \ldots, P_n$ *are tree-products;*
- *or* $\mathfrak{C}^{\overline{*}}.(P_1 + \cdots + P_n)$ *where* $n \in \mathbb{N}$, $P_1, \ldots, P_n$ *are pairwise incomparable tree-products,* $\mathfrak{C} = \sum_{i=1}^{m} f_i(P_{\Box i1}, \ldots, P_{\Box in_i})$, $m \geq 1$, *each summand* $f_i(P_{\Box i1}, \ldots, P_{\Box in_i})$ *is incomparable with any other, each pattern* $P_{\Box ij}$ *is either a tree-product or the hole* $\Box$, $\mathfrak{C}$ *is* $\Box$*-generated, and one of the following conditions is satisfied:* (*a*) $\mathfrak{C}$ *is not* $\Box$*-linear and* $n \geq 1$, *or* (*b*) $\mathfrak{C}$ *is not* $\Box$*-linear,* $n = 0$, *and* $P_{\Box ij} \neq \Box$ *for some* i, j, *or* (*c*) $\mathfrak{C}$ *is* $\Box$*-linear and* $n \leq 1$.

Proof. A tree pre-product of the form $f^{\overline{?}}(S_1, \ldots, S_n)$ is $\rightarrow_1$-normal if and only if $S_1, \ldots, S_n$ are $\rightarrow_1$-normal and neither ($R7$) nor ($R8$) applies. The latter means that each S_i is a tree-product P_i.

Next, consider a tree pre-product of the form $\mathfrak{C}^{\overline{*}}.S$. Write $\mathfrak{C}$ as $\sum_{i=1}^{m} A_i$ where $A_i = f_i(S_{\Box i1}, \ldots, S_{\Box in_i})$, and S as a sum of tree pre-products $P_1 + \cdots + P_n$.

If $\mathfrak{C}^{\overline{*}}.S$ is $\rightarrow_1$-normal, then ($R5$) does not apply, so $m \geq 1$. Since ($R1$) and ($R3$) do not apply (and $+$ is understood modulo associativity and commutativity), $P_1, \ldots, P_n$ are pairwise incomparable (and different from 0, but that is implied). Similarly, since ($R2$) and ($R4$) do not apply, each summand $f_i(P_{\Box i1}, \ldots, P_{\Box in_i})$ is incomparable with any other. Additionally, ($R9$) and ($R10$) do not apply, so each $S_{\Box ij}$ is either equal to $\Box$ or equal to some tree-product. Henceforth, write $S_{\Box ij}$ as $P_{\Box ij}$. If $\mathfrak{C}$ were not $\Box$-generated, then for some non-empty A_i, $P_{\Box ij}$ would be different from $\Box$ for every j, then rule ($R6$) would apply. We now prove that (a), (b), or (c) holds depending on the shape of $\mathfrak{C}$. If $\mathfrak{C}$ is not $\Box$-linear but (a) does not hold, then $n = 0$. Since ($R11$) does not apply, some $P_{\Box ij}$ is different from $\Box$, or some f_i has arity 0. However, if some f_i has arity 0, then A_i would just be $f_i()$, and ($R6$) would apply, so (b) holds. It remains to consider the case where $\mathfrak{C}$ is $\Box$-linear. Since ($R12$) does not apply, it must be that $n \leq 1$, so (c) holds.

Conversely, it is easy to check that if the conditions listed in the statement of the lemma for $\mathfrak{C}^{\overline{*}}.(P_1 + \cdots + P_n)$ are satisfied, then $\mathfrak{C}^{\overline{*}}.(P_1 + \cdots + P_n)$ is $\rightarrow_1$-normal. $\Box$

Lemma 10.9. *If* $S_1, \ldots, S_n$ *are directed STREs, then so is* $f^{\overline{?}}(S_1, \ldots, S_n)$.

Proof. Non-emptiness is clear, since $S_1, \ldots, S_n$ are non-empty. Let t, t' be any two trees in $f^{\overline{?}}(S_1, \ldots, S_n)$. If $t = f(t_1, \ldots, t_n)$ and $t' = f(t'_1, \ldots, t'_n)$ with $t_i, t'_i \in S_i$ for every i, then we can find $t''_i \geq t_i, t'_i$ in S_i, and then $f(t''_1, \ldots, t''_n) \geq t, t'$ is in $f^{\overline{?}}(S_1, \ldots, S_n)$.

If t is in some S_j and $t' = f(t'_1, \ldots, t'_n)$ with $t'_i \in S_i$ for every i, then build the tree $s = f(s_1, \ldots, s_{j-1}, t, s_{j+1}, \ldots, s_n)$, where s_i is an arbitrary tree from the non-empty set S_i, $i \neq j$. Clearly, s is in $f^{\overline{?}}(S_1, \ldots, S_n)$, so, by reduction to the previous case, there is a tree $t'' \in f^{\overline{?}}(S_1, \ldots, S_n)$ such that $s, t' \preceq t''$. Since $t \preceq s$, we obtain that $t, t' \preceq t''$.

Similarly if t' is in some S_k, we build a new tree $s' = f(s'_1, \ldots, s'_{k-1}, t', s'_{k+1}, \ldots, s'_n)$ and conclude by a similar argument that there is a tree $t'' \in f^{\overline{?}}(S_1, \ldots, S_n)$ such that $t' \preceq s' \preceq t''$ and $t \preceq t''$. $\Box$

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

The case of STREs of the form $\mathfrak{C}^{\overline{*}}.S$ is more complex. The three cases (*a*), (*b*), and (*c*) in the lemma below match those of Lemma 10.8, second item.

Lemma 10.10. *Let $\mathfrak{C} = A_1 + \cdots + A_m$ be a sum of patterns $A_i = f_i(S_{\Box i1}, \ldots, S_{\Box in_i})$, where each $S_{\Box ij}$ that is different from $\Box$ has a non-empty language. If any of the following conditions is satisfied, then $\mathfrak{C}^{\overline{*}}.S$ is directed:*

- (*a*) *$\mathfrak{C}$ is not $\Box$-linear, and S has a non-empty language, or*
- (*b*) *$\mathfrak{C}$ is not $\Box$-linear, and some $S_{\Box ij}$ is different from $\Box$, or*
- (*c*) *$\mathfrak{C}$ is $\Box$-linear and $\Box$-generated, and S is irreducible.*

Proof. The fact that $\mathfrak{C}^{\overline{*}}.S$ is non-empty is an easy exercise: in case (*a*), $\mathfrak{C}^{\overline{*}}.S$ contains S; in case (*b*), it contains $S_{\Box ij}$; and in case (*c*), it contains S, which, as an irreducible subset, is necessarily non-empty. Let t and t' be any two trees in $\mathfrak{C}^{\overline{*}}.S$.

In case (*a*), some non-empty A_i is of the form $f(S_{\Box 1}, \ldots, S_{\Box n})$, where $\Box$ occurs at least twice, say at positions j and j', $j \neq j'$. For every k, $1 \leq k \leq n$, define a tree t_k as follows: if $S_{\Box k} = \Box$ and $k \neq j'$ (including the case $k = j$), let $t_k = t$, if $k = j'$ then let $t_k = t'$, and if $S_{\Box k} \neq \Box$ then pick any tree for t_k from the language of $S_{\Box k}$, which is non-empty by assumption. We check that $f(t_1, \ldots, t_n)$ is in $\mathfrak{C}^{\overline{*}}.S$, and $t = t_j$, $t' = t_{j'}$ both embed into $f(t_1, \ldots, t_n)$.

Case (*b*) reduces to (*a*), since if $S_{\Box ij} \neq \Box$, then $\mathfrak{C}^{\overline{*}}.S$ defines the same language as $\mathfrak{C}^{\overline{*}}.(S + S_{\Box ij})$.

In case (*c*), every non-empty A_i is of the form $f(\cdots, \Box, \cdots)$ with a unique occurrence of $\Box$. In that case, the language of $\mathfrak{C}^{\overline{*}}.S$ can be described more simply: it consists of those trees of the form $c_1[c_2[\cdots c_k[s] \cdots]]$, where $k \in \mathbb{N}$, each c_k is a context in the language of $\mathfrak{C}$, with just one occurrence of $\Box$ each, and s is a tree in $S \cup \operatorname{args} \mathfrak{C}$. For short, say that a context c is *in* $\mathfrak{C}^{\overline{*}}$ if and only if it is of the form $c_1[c_2[\cdots c_k[\Box] \cdots]]$, where $k \in \mathbb{N}$ and each c_k is in $\mathfrak{C}$. Such contexts have exactly one occurrence of $\Box$. Hence, the language of $\mathfrak{C}^{\overline{*}}.S$ consists of those trees $c[s]$ where c is in $\mathfrak{C}^{\overline{*}}$ and $s \in S \cup \operatorname{args} \mathfrak{C}$. Given any two such trees $t = c[s]$ and $t' = c'[s']$, we find a tree $t'' \in \mathfrak{C}^{\overline{*}}.S$ in which both t and t' embed, as follows.

If both s and s' are in S, then by directedness there is an $s'' \in S$ such that $s, s' \preceq s''$: we define t'' as $c[c'[s'']]$.

If $s \in S$ and $s' \in \operatorname{args} \mathfrak{C}$, then there is an $A_i = f_i(S_{\Box i1}, \ldots, S_{\Box in_i})$ and a position j', $1 \leq j' \leq n_i$, such that $S_{\Box ij'} \neq \Box$ and the language of $S_{\Box ij'}$ contains s'. The unique position j at which $S_{\Box ij} = \Box$ is different from j'. Let $u = f_i(u_1, \ldots, u_n)$ where: $u_j = s$, $u_{j'} = s'$, for every position $k \neq j, j'$ u_k is an arbitrary tree from $S_{\Box ik}$. By construction, $u \in \mathfrak{C}^{\overline{*}}.S$, hence so is $t'' = c[c'[u]]$. Additionally, since s and s' both embed in u, t and t' both embed in t''.

The same argument applies when $s \in \operatorname{args} \mathfrak{C}$ and $s' \in S$. Finally, we consider the case where s and s' are both in $\operatorname{args} \mathfrak{C}$. Then, $s \in \operatorname{args} A_i$ for some i, say $A_i = f_i(S_{\Box i1}, \ldots, S_{\Box in_i})$, $S_{\Box ij'} \neq \Box$ and s in in the language of $S_{\Box ij'}$. Since S is irreducible, it is non-empty, hence we can pick a tree s_0 from it. Let $u_{j'} = s$, $u_j = s_0$ where j is the unique position where $S_{\Box ij} = \Box$ and pick u_k from the non-empty language $S_{\Box ik}$ for every $k \neq j, j'$; define $u = f_i(u_1, \ldots, u_{n_i})$, a tree from $\mathfrak{C}^{\overline{*}}.S$ in which s embeds. Similarly, since $s' \in \operatorname{args} \mathfrak{C}$, s' is in the support of some $A_{i'}$, say $A_{i'} = f_{i'}(S_{\Box i'1}, \ldots, S_{\Box i'n_{i'}})$, $S_{\Box i'j'''} \neq \Box$ and s' is in the language of $S_{\Box i'j'''}$. Let $v_{j'''} = s'$ (instead of s in our previous step), $v_{j''} = u$ (instead of s_0) where j'' is the unique position where $S_{\Box i'j''} = \Box$ and pick v_k from the non-empty language $S_{\Box i'k}$ for every $k \neq j'', j'''$. The tree $v = f_{i'}(v_1, \ldots, v_{n_{i'}})$ is again in $\mathfrak{C}^{\overline{*}}.S$, and now both s and s' embed into it. Finally, we define t'' as $c[c'[v]]$. $\Box$

Theorem 10.11. *The ideals, that is, the irreducible closed subsets of $\mathcal{T}(\Sigma)$ are exactly the languages of tree-products.*

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Proof. One direction is Lemma 10.6. In the other direction, we show that the language of every tree-product P is directed by structural induction on P, using Lemma 10.9 or Lemma 10.10, depending on its shape, as given by Lemma 10.8. In doing the proof, one needs to observe that any $\rightarrow_1$-normal STRE $S = P_1 + \cdots + P_m$ (where the language of each P_i is an ideal by induction hypothesis) has an empty language if and only if $m = 0$ – because ideals are never empty. $\square$

Exceptionally, let us dispense with the traditional shape of our S-representation theorem, and let us just state the following.

Theorem 10.12 (S-representation, finite ranked trees). *Let Σ be a finite signature. There is an S-representation $(S, [\![\_]\!], \trianglelefteq, \tau, \wedge)$ of $\mathcal{T}(\Sigma)$, where S is the collection of all tree-products over Σ.*

Proof. Tree-products are regular tree languages. Hence, inclusion ($\trianglelefteq$) is decidable (Thatcher and Wright 1968, Section 2), see also Comon et al. (2004). Since $\mathcal{T}(\Sigma)$ is equal to the language of $(\sum_{f \in \Sigma} f(\square, \ldots, \square))^{\overline{*}}.0$ (where each $f \in \Sigma_r$ is applied to a list of r boxes), we obtain τ by normalizing with respect to $\rightarrow_1$. To compute finite intersections ($\wedge$) of two tree-products, we first convert those tree-products to tree automata, compute their intersection, convert the result back to an STRE by the construction of Proposition 10.1, and normalize it by $\rightarrow_1$ to obtain its expression as a sum of tree-products. $\square$

The above procedures are not optimal, and notably the inclusion procedure takes exponential time. As in the word case (see Lemma 7.9), there is a polynomial time inclusion test, but it is complex, and its correctness proof is difficult. We shall study it – by necessity – in the general case of unranked trees over a Noetherian signature. (See Corollary 11.33.) We let the interested reader do the required modifications to adapt it to the finite, ranked case.

Anticipating on that general case, the ranked trees in $\mathcal{T}(\Sigma)$ embed into the set of all (unranked) trees over the finite set Σ (seen as a Noetherian space, with the discrete topology). Proposition 6.4 tells us that the sobrification of the former can be be seen, up to isomorphism, as a subspace of the sobrification of the latter. We leave it as an exercise to check that this embedding $S \mapsto S^\circ$ is defined by:

— if $S = P_1 + \cdots + P_m$, then $S^\circ = P_1^\circ + \cdots + P_m^\circ$;
— $(f^{\overline{?}}(P_1, \ldots, P_n))^\circ = f^{\overline{?}}(P_1^{\circ ?} \cdots P_n^{\circ ?})$;
— $((\sum_{i=1}^m f_i(P_{\square i1}, \ldots, P_{\square in_i}))^{\overline{*}}.S)^\circ = (\sum_{i=1}^m f_i(P_{\square i1}^{\circ ?} \cdots P_{\square in_i}^{\circ ?}))^{\overline{*}}.S^\circ$
— $\square^\circ = \square$.

Combining this with the polynomial time inclusion test we shall see in Definition 11.31, this provides us with a polynomial time inclusion test for the tree-products and STREs over Σ.

11. Completing Finite Trees: the General Case

For every set X, let $\mathcal{T}(X)$ denote the set of all (ground, first-order) terms built using function symbols from X. Function symbols are now unranked and may be applied to arbitrarily long lists of arguments. Since lists can be seen as finite words, $\mathcal{T}(X) = X \times \mathcal{T}(X)^*$. The leaves of such terms are the constants $f()$, formed from a function symbol $f \in X$ and an empty list of arguments; we shall also simply write f instead of $f()$. In general, a term t will be written as $f(t_1, t_2, \ldots, t_n)$. When we wish to stress that the list $t_1, t_2, \ldots, t_n$ is a word, we shall also write $f(t_1 t_2 \cdots t_n)$ or $f(\vec{t})$ for a word $\vec{t} \in \mathcal{T}(X)^*$.

Again, our terms are ground. We shall later use the notation $\mathcal{T}(X, \{\square\})$ to refer to the set of terms with exactly one variable $\square$. These are the terms of $\mathcal{T}(X \cup \{\square\})$ where $\square$ is always applied to the empty list of arguments.

The *subterms* of a term t are defined inductively as usual: writing t as $f(t_1, \ldots, t_n)$, they are t itself plus all subterms of t_i, $1 \leq i \leq n$.

Given any quasi-ordering $\leq$ on X, the *embedding quasi-ordering* $\preceq_\leq$ on $\mathcal{T}(X, \mathcal{V})$ is defined by induction on the sum of the sizes of the terms s, t, by $s \preceq_\leq t$ iff:

— either $t = g(t_1, \ldots, t_n)$ and $s \preceq_\leq t_j$ for some j, $1 \leq j \leq n$;
— or $s = f(\vec{s})$, $t = g(\vec{t})$, $f \leq g$ and $\vec{s} \preceq^*_\leq \vec{t}$;

where we recall that $\preceq^*_\leq$ is the embedding quasi-ordering on words, understanding that the letters (which are terms themselves here) are quasi-ordered by $\preceq_\leq$ (this is a recursive definition). As with several other notions here, we reuse freely some notations and some notions that we had introduced in the special case considered in Section 10.

Kruskal's tree theorem – in a more complete form than stated earlier – states that $\preceq_\leq$ is wqo on $\mathcal{T}(X)$ iff $\leq$ is wqo on X.

The latter extends to a topological setting: for a topological space X, $\mathcal{T}(X)$ is Noetherian iff X is Noetherian (Goubault-Larrecq 2013, Theorem 9.7.46). For this to make sense, we need to put a topology on $\mathcal{T}(X)$, and this is the *tree topology* (Goubault-Larrecq 2013, Definition 9.7.39), defined as follows. For short, let $Y = \mathcal{T}(X)$. The *simple tree expressions* on X are given by the grammar $\pi ::= \Diamond U(\pi_1 \mid \cdots \mid \pi_n)$ where U is open in X, and $n \in \mathbb{N}$. (The base case is obtained when $n = 0$.) Such a simple tree expression denotes the set of terms t that have a subterm of the form $f(\vec{t})$ with $f \in U$ and $\vec{t} \in Y^*\pi_1 Y^* \cdots Y^*\pi_n Y^*$. The simple tree expressions generate a topology, called the tree topology.

Here are a few basic facts about the tree topology. These can be found in Goubault-Larrecq (2013, Exercises 9.7.40, 9.7.43). The proof of the following proposition, and of most subsequent results, can be found in Appendix E.

Proposition 11.1. *Let X be a topological space. Every finite intersection of simple tree expressions can be rewritten as a finite union of simple tree expressions. In particular, the simple tree expressions form a base of the tree topology.*

Letting $\leq$ be the specialization quasi-ordering of X, the specialization quasi-ordering of $\mathcal{T}(X)$ is the embedding quasi-ordering $\preceq_\leq$.

The reader might be under the impression that the tree topology is far removed from the embedding quasi-ordering $\preceq_\leq$. Not so: the situation is exactly as for words (Proposition 8.2), and when X is a poset, then the tree topology is the Alexandroff topology of $\preceq_\leq$:

Proposition 11.2. *Let X be a set quasi-ordered by $\leq$. The tree topology on $\mathcal{T}(X_a)$ is exactly the Alexandroff topology of $\preceq_\leq$ on $\mathcal{T}(X)$.*

Proof. This is the first part of Exercise 9.7.48 of Goubault-Larrecq (2013). We give a proof just before Section E.1 in Appendix E. □

This means, as for most other cases studied in this paper (except the case of rings, of words under the prefix ordering, and of powersets), that in the familiar case where X is a poset, the completion $\widehat{\mathcal{T}(X)}$ is both the sobrification of $\mathcal{T}(X)$ with the tree topology, and the ideal completion of $\mathcal{T}(X)$ with the $\preceq_\leq$ quasi-ordering. (Recall Hoffmann's theorem that for a poset Y, $\mathcal{S}(Y_a) = \mathbf{I}(Y)$ Hoffmann 1979b.) The present section is therefore merely an extension of Section 10 to a more general format of trees, and more general spaces of function symbols.

11.1 Tree steps

To characterize $\mathcal{S}(X)$, we rely again on specific forms of regular tree expressions, this time for terms. We start with regular expressions based on the $\overline{?}$ operator, which we call *tree steps*.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Definition 11.3 (Tree step). *Let X be a topological space.*

For every word-product $\vec{P}$ on $\mathcal{T}(X)$, the support *$supp\,\vec{P}$ of $\vec{P}$ is defined as the set of terms t such that the one-element sequence t is in the language of $\vec{P}$. Equivalently, when $\vec{P} = e_1 e_2 \cdots e_n$, $supp\,\vec{P} = \bigcup_{i=1}^{n} supp\,e_i$, where $supp\,S^* = S$ and $supp\,P^? = P$.*

For every closed subset F of X, let $F^{\overline{?}}(\vec{P})$ denote the union of $supp\,\vec{P}$ with the set of all terms of the form $f(\vec{t})$, $f \in F$, $\vec{t} \in \vec{P}$.

The tree steps *are the expressions of the form $C^{\overline{?}}(\vec{P})$ where C is irreducible closed in X and $\vec{P}$ is a word-product on $\mathcal{T}(X)$.*

For example, when $C = \{f\}$, $\vec{P} = \{a\}^?\{b\}^?$, then $C^{\overline{?}}(\vec{P})$ is the set of all terms $f(a, b), f(a), f(b), f()$, but also the terms a and b from supp $\vec{P}$.

Lemma 11.4. *Let X be a topological space. For every closed subset F of X, and every word-product $\vec{P}$ on $\mathcal{T}(X)$, $supp\,\vec{P}$ and $F^{\overline{?}}(\vec{P})$ are closed in $\mathcal{T}(X)$. If moreover $F = C$ is irreducible, then so is the tree step $C^{\overline{?}}(\vec{P})$.*

Proof. See Appendix E, as for most other results of this section. □

11.2 Tree iterators

We turn to the needed generalization of tree regular expressions of the form $\mathfrak{C}^{\overline{*}}.S$. Recall that $\mathcal{T}(X, \{\Box\})$ is the set of terms of $\mathcal{T}(X \cup \{\Box\})$ where $\Box$ is always applied to the empty list of arguments – $\Box$ acts as a hole, meant to be replaced by terms.

Definition 11.5 (Context). *Let X be a topological space, and $\Box$ be an element called the* hole, *and assumed not to be in X. A* context *is a term of $\mathcal{T}(X, \{\Box\})$.*

Given any context $\mathcal{C}$, and any subset S of $\mathcal{T}(X)$, $\mathcal{C}[S]$ denotes the set of all terms in $\mathcal{T}(X)$ obtained by replacing each occurrence of $\Box$ in $\mathcal{C}$ by (possibly different) terms from S.

The hole $\Box$ can be replaced by different terms, for example, when $S = \{a, b\}$, $f(\Box, \Box)[S]$ denotes $\{f(a, a), f(a, b), f(b, a), f(b, b)\}$. Notice that terms without the hole are also considered as contexts: take $\mathcal{C} = f(c)$ for example, then $\mathcal{C}[S]$ contains just $f(c)$, for any subset S.

By extension, where t is a single term, $\mathcal{C}[t]$ denotes the unique term obtained by replacing every hole $\Box$ in $\mathcal{C}$ by t. One should beware that $\mathcal{C}[S]$ is *not* in general equal to the set of all terms $\mathcal{C}[t]$, $t \in S$, as the example $S = \{a, b\}$, $\mathcal{C} = f(\Box, \Box)$ demonstrates.

Given two contexts $\mathcal{C}$ and $\mathcal{C}'$, the notation $\mathcal{C}[\mathcal{C}'[S]]$ makes sense: this is $\mathcal{C}[S']$, where $S' = \mathcal{C}'[S]$. One can also read this as $\mathcal{C}[\mathcal{C}'][S]$, where $\mathcal{C}[\mathcal{C}']$ is the context obtained by replacing each occurrence of $\Box$ in $\mathcal{C}$ by the (same) context $\mathcal{C}'$.

Given a single hole $\Box$, we equip $\{\Box\}$ with the only possible topology and write $\mathcal{T}(X) + \{\Box\}$ for the topological coproduct. Its open subsets are the sets of the form U or $U \cup \{\Box\}$, where U is open in $\mathcal{T}(X)$, and similarly for closed subsets. Its *irreducible* closed subsets are the irreducible closed subsets of $\mathcal{T}(X)$, and $\{\Box\}$.

Definition 11.6 (Tree iterators). *Let X be a topological space, and $\Box$ be a distinguished element outside X, called the hole.*

The tree iterators *are formal expressions of the form $\mathfrak{C}^{\overline{*}}.S$, where $\mathfrak{C}$ is a closed subset of $X \times (\mathcal{T}(X) + \{\Box\})^*$, and S is a closed subset of $\mathcal{T}(X)$. We equate pairs $(f, \vec{u})$ in $\mathfrak{C}$ with the contexts $c = f(\vec{u})$ – such contexts are* elementary contexts, *in that $\Box$ can only occur directly under f.*

Call argument support *$args\,\mathfrak{C}$ of $\mathfrak{C}$ the set of all terms $t \in \mathcal{T}(X)$ such that $f(t) \in \mathfrak{C}$ for some $f \in X$. (Equivalently, such that some term of the form $f(\ldots, t, \ldots)$ belongs to $\mathfrak{C}$.) Then, $\mathfrak{C}^{\overline{*}}.S$ denotes the smallest set of terms in $\mathcal{T}(X)$ such that:*

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

(1) *every term in* S *is in* $\overline{\mathfrak{C}^*}.S$;
(2) *every term in args* $\mathfrak{C}$ *is in* $\overline{\mathfrak{C}^*}.S$;
(3) *for every elementary context* $c \in \mathfrak{C}$, $c[\overline{\mathfrak{C}^*}.S]$ *is included in* $\overline{\mathfrak{C}^*}.S$.

For example, let $\mathfrak{C} = \{f\} \times \{\Box\}^*$. Its argument support is the set of all terms $t \in \mathcal{T}(X)$ such that $t \in \{\Box\}^*$, that is, $t = \Box$: this is impossible since $\Box$ is not a term of $\mathcal{T}(X)$. So args $\mathfrak{C}$ is empty. Let $S = \{a, b\}$. $\overline{\mathfrak{C}^*}.S$ is the following set of terms. First, there are the terms from S, namely a and b. Then, there are the terms obtained from the latter by applying the elementary contexts $c \in \mathfrak{C}$ to the above terms. These contexts are $f(), f(\Box), f(\Box\Box)$, etc. So we obtain the terms $f, f(a), f(b), f(a, a), f(a, b), f(b, a), f(b, b)$, etc. In a third step, we find the terms obtained by applying the contexts $f()$, $f(\Box), f(\Box\Box)$, etc., to the latter terms: we obtain $f(f), f(f(a)), f(f(b)), f(f(a, a)), \ldots, f(a, f(a))$, etc. Continuing this way, we realize that $\overline{\mathfrak{C}^*}.S$ is the set of terms whose non-leaf nodes are labeled with f, and whose leaves are labeled by a or $b \ldots$ or f.

For a more complex example, take $\mathfrak{C} = \{f\} \times \{c\}^? \{\Box\}^*$. Now args $\mathfrak{C} = \{c\}$ – c is the only part of $\{c\}^? \{\Box\}^*$ that does not contain the hole $\Box$. Take $S = \{a, b\}$ again. Then, $\overline{\mathfrak{C}^*}.S$ is the following set of terms. First, we find the terms from S, namely a and b. Second, we find the terms from args $\mathfrak{C}$, that is, c. Third, we find the results of applying contexts of the form $f(), f(c), f(\Box), f(c\Box), f(\Box\Box)$, $f(c\Box\Box), f(\Box\Box\Box)$, etc., to the above terms: these are the terms of the form $f(t_1, t_2, \ldots, t_n)$ where each t_i is in $\{a, b, c\}$. Continuing this way, we obtain that $\overline{\mathfrak{C}^*}.S$ is the set of all terms built using f, a, b, c, where a, b, c are applied to no argument. In particular, $\overline{\mathfrak{C}^*}.S$ is the same set as $\overline{\mathfrak{C}'^*}.S'$, where $\mathfrak{C}' = \{f\} \times \{\Box\}^*$, and $S' = \{a, b, c\}$.

Remark 11.7. One should beware that $\overline{\mathfrak{C}^*}.S$ can be non-empty even when S is empty, and even when args $\mathfrak{C}$ is empty. For example, for $\mathfrak{C} = \{f\} \times \{\Box\}^*$, $\overline{\mathfrak{C}^*}.\emptyset$ is the set of terms whose only function symbol is f, at all positions: $f(), f(f()), f(f(), f()), \ldots$ This unexpected kind of non-empty STRE will be used in case 3 of Lemma 11.11, and in the definition of τ' in Theorem 11.36, notably.

The argument support args $\mathfrak{C}$ mimics the eponymous notion of Section 10. While we do not commit yet to a specific syntax for closed subsets $\mathfrak{C}$ of $X \times (\mathcal{T}(X) + \{\Box\})^*$, the case that will be of most interest to us – and which will be the general case when X is Noetherian – is when $\mathfrak{C}$ is of the form $\bigcup_{i=1}^m C_i \times Q_i$, where C_i is irreducible closed in X and each Q_i is a word-product over $\mathcal{T}(X) + \{\Box\}$ for each i, $1 \leq i \leq m$. In that case, args $\mathfrak{C}$ simplifies:

Lemma 11.8. *Let* $\mathfrak{C} = \bigcup_{i=1}^m C_i \times Q_i$, *where* C_i *is a closed non-empty subset of* X *and each* Q_i *is a word-product over* $\mathcal{T}(X) + \{\Box\}$ *for each* i, $1 \leq i \leq m$. *Then, args* $\mathfrak{C} = \bigcup_{i=1}^m \text{supp}\, Q_i \cap \mathcal{T}(X)$.

Each Q_i *is closed in* $\mathcal{T}(X) + \{\Box\}$. *In particular, args* $\mathfrak{C}$ *is closed in* $\mathcal{T}(X)$.

Proof. Let $t \in \text{args}\, \mathfrak{C}$. For some i, $1 \leq i \leq m$, there is a term $f(t)$ with $f \in C_i$ and such that the one-element word t is in Q_i. The latter means that $t \in \text{supp}\, Q_i$. Also, $t \in \mathcal{T}(X)$ by definition. Conversely, any element of supp $Q_i \cap \mathcal{T}(X)$ is a term (not the hole) t in supp Q_i, that is, such that the one-element word t is in Q_i. Since C_i is non-empty, pick f from C_i. Then $f(t)$ is in $\mathfrak{C}$, so t is in args $\mathfrak{C}$.

For any space Y, the function $i\colon Y \to Y^*$ that maps any $y \in Y$ to the one-letter word y is continuous (Lemma B.1 in the Appendix). Since supp $Q_i = i^{-1}(Q_i)$, it is closed in $Y = \mathcal{T}(X) + \{\Box\}$. It follows that supp $Q_i \cap \mathcal{T}(X)$ is closed in $\mathcal{T}(X)$, hence also args $\mathfrak{C}$. ☐

Lemma 11.9. *Let* X *be a topological space, and* $\Box$ *be a hole outside* $\mathcal{T}(X)$. *Every tree iterator* $\overline{\mathfrak{C}^*}.S$ *such that args* $\mathfrak{C}$ *is closed in* $\mathcal{T}(X)$ *denotes a closed subset of* $\mathcal{T}(X)$.

The cases where tree iterators are irreducible require an analysis of the number of holes that can occur in each context, which parallels the analysis we ran in Lemma 10.8, leading to three cases (a), (b), and (c).

Definition 11.10. *Let X be a topological space, and $\Box$ be a hole outside $\mathcal{T}(X)$. Let $\mathfrak{C}$ be a subset of $X \times (\mathcal{T}(X) + \{\Box\})^*$. $\mathfrak{C}$ is* $\Box$-linear *iff its elements have at most one occurrence of $\Box$ each, that is, for every elementary context $f(\vec{t}) \in \mathfrak{C}$, at most one element of $\vec{t}$ equals $\Box$. $\mathfrak{C}$ is* $\Box$-generated *iff for every $f(\vec{u}) \in \mathfrak{C}$ such that $\Box$ does not occur in $\vec{u}$, one can split $\vec{u}$ as a concatenation $\vec{u}_1\vec{u}_2$ so that $f(\vec{u}_1\Box\vec{u}_2)$ is in $\mathfrak{C}$.*

The $\Box$-generated closed sets $\mathfrak{C}$ are those such that every element in $\mathfrak{C}$ contains the hole $\Box$ or has a larger element in $\mathfrak{C}$ that contains the hole. So, for example, $\mathfrak{C} = \{f\} \times \{\Box\}^*$ is $\Box$-generated. Indeed, its elements are the pairs $(f, \Box^n)$, $n \in \mathbb{N}$; when $n \geq 1$, this contains $\Box$, and when $n = 0$, (f, ϵ) is below $(f, \Box)$, which is inside $\mathfrak{C}$ and contains $\Box$, that is, we apply the above definition, picking $\vec{u}_1 = \vec{u}_2 = \epsilon$. $\mathfrak{C}$ is not $\Box$-linear, since, say, $(f, \Box\Box)$ is in it, and contains two holes.

On the other hand, $\{f\} \times \{\Box\}^?$ is both $\Box$-linear and $\Box$-generated.

The tree iterators $\mathfrak{C}^{\overline{*}}.S$ are easier to understand when $\mathfrak{C}$ is $\Box$-linear, just as in Section 10. Let $\mathfrak{C}^{\Box*}$ be the set of all contexts of the form $c_1[c_2[\cdots[c_k]\cdots]]$, where $k \in \mathbb{N}$, and $c_1, c_2, \ldots, c_k \in \mathfrak{C}$. (When $k = 0$, this denotes $\Box$.) Whenever $\mathfrak{C}$ is $\Box$-linear, all these contexts have at most occurrence of $\Box$. Then $\mathfrak{C}^{\overline{*}}.S$ is the set of all terms obtained from a context c in $\mathfrak{C}^{\Box*}$ by replacing the unique occurrence of the hole $\Box$ (if any) by a term from $S \cup \operatorname{args} \mathfrak{C}$.

Lemma 11.11. *Let X be a topological space, and $\Box$ be a hole outside $\mathcal{T}(X)$, $\mathfrak{C}$ be a closed subset of $X \times (\mathcal{T}(X) + \{\Box\})^*$, S be a closed subset of $\mathcal{T}(X)$, and assume that $\operatorname{args} \mathfrak{C}$ is closed. Then, the tree iterator $\mathfrak{C}^{\overline{*}}.S$ is irreducible in the following cases:*

(1) if $\mathfrak{C}$ is non-$\Box$-linear, and S is non-empty;
(2) or if $\mathfrak{C}$ is $\Box$-generated and $\Box$-linear and S is irreducible;
(3) or if $\mathfrak{C}$ is non-empty, $\Box$-generated, and S is empty.

In case 3, $\mathfrak{C}^{\overline{*}}.\emptyset$ is in particular non-empty. That should not be a surprise, considering Remark 11.7.

Using these two forms of closed subsets – tree steps and tree iterators – one can express the complements of all simple tree expressions.

Lemma 11.12. *Let X be a topological space. The complement $\complement\pi$ of the open subset denoted by the simple tree expression $\pi = \Diamond U(\pi_1 \mid \pi_2 \mid \cdots \mid \pi_n)$ is given by structural induction on π by:*

— $\complement\pi = ((F \times \{\Box\}^*) \cup (X \times (\complement\pi_1^*\{\Box\}^?\complement\pi_2^*\{\Box\}^? \cdots \{\Box\}^?\complement\pi_n^*)))^{\overline{*}}.\emptyset$ *if $n \geq 1$, where F is the complement of U in X;*
— *if $n = 0$, then $\complement\pi = (F \times \{\Box\}^*)^{\overline{*}}.\emptyset$.*

Lemma 11.12 can be made explicit, especially when $U = X$ and $n = 1$:

Lemma 11.13. *Let X be a topological space. The complement $\complement\pi$ of the open subset denoted by the simple tree expression $\pi = \Diamond U(\pi_1 \mid \pi_2 \mid \cdots \mid \pi_n)$ is given by structural induction on π by:*

— *if $U = X$ and $n = 0$, then $\complement\pi$ is empty;*
— *if $U = X$ and $n = 1$, then $\complement\pi$ is $X^{\overline{?}}(\complement\pi_1^*)$;*
— *if $U = X$ and $n \geq 2$, then $\complement\pi$ is $(X \times (\complement\pi_1^*\{\Box\}^?\complement\pi_2^*\{\Box\}^? \cdots \{\Box\}^?\complement\pi_n^*))^{\overline{*}}.\emptyset$;*
— *if $U \neq X$ and $n = 0$, then $\complement\pi$ is $(F \times \{\Box\}^*)^{\overline{*}}.\emptyset$, where F is the complement of U;*

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

— *if* $U \neq X$ *and* $n = 1$, *then* $\complement\pi$ *is* $((F \times \{\Box\}^*))^{\overline{*}}.X^?(\complement\pi_1^*)$, *where* F *is the complement of* U;
— *if* $U \neq X$ *and* $n \geq 2$, *then* $\complement\pi$ *is* $((F \times \{\Box\}^*) \cup (X \times (\complement\pi_1^*\{\Box\}^?\complement\pi_2^*\{\Box\}^? \cdots \{\Box\}^?\complement\pi_n^*)))^{\overline{*}}.\emptyset$, *where* F *is the complement of* U.

Proof. The case $U = X$, $n = 0$ is clear. When $U = X$ and $n = 1$, Lemma 11.12 tells us that $\complement\pi = (X \times \complement\pi_1^*)^{\overline{*}}.\emptyset$. This is by definition the smallest set A of terms containing $\complement\pi_1 = \text{supp}\,(X \times \complement\pi_1^*)$ and such that for every elementary context $f(\vec{u}) \in X \times \complement\pi_1^*$, $f(\vec{u})$ (which is a term, i.e., $\Box$ does not occur in it) is in A. Thus, $A = X^{\overline{?}}(\complement\pi_1^*)$. When $U \neq X$ and $n = 1$, by Lemma 11.12 $\complement\pi = ((F \times \{\Box\}^*) \cup (X \times (\complement\pi_1^*)))^{\overline{*}}.\emptyset$. The argument support of $\mathfrak{C} = (F \times \{\Box\}^*) \cup (X \times (\complement\pi_1^*))$ is $\complement\pi_1$, so the elements of $\complement\pi$ are those of $\complement\pi_1$, those of the form $f(\vec{t})$ where f is arbitrary and $\vec{t} \in \complement\pi_1^*$ (hence, all the terms of $X^?(\complement\pi_1^*)$), and those obtained from the latter by applying any number of function symbols from F. The other cases follow directly from Lemma 11.12. ☐

11.3 Checking inclusion

We start with tree step inclusion.

Lemma 11.14. *Let* X *be a topological space,* C *and* C' *be two irreducible closed subsets of* X, $\vec{P}$ *and* $\vec{P}'$ *be two word-products over* $\mathcal{T}(X)$. *Then* $C^{\overline{?}}(\vec{P}) \subseteq C'^{\overline{?}}(\vec{P}')$ *iff* $C \subseteq C'$ *and* $\vec{P} \subseteq \vec{P}'$, *or* $C^{\overline{?}}(\vec{P}) \subseteq \text{supp}\,\vec{P}'$.

We turn to the cases where one of the closed set to compare is of the form $\mathfrak{C}^{\overline{*}}.S$. For every closed subset $\mathfrak{C}$ of $X \times (\mathcal{T}(X) + \{\Box\})^*$ and every closed subset S of $\mathcal{T}(X)$, we write $\mathfrak{C}[S]$ for the set of pairs $(f, \vec{t})$, where $(f, \vec{u})$ ranges over $\mathfrak{C}$ and where $\vec{t}$ is obtained from $\vec{u}$ by replacing each occurrence of $\Box$ by possible different terms from S.

Lemma 11.15. *Let* X *be a topological space,* C *be an irreducible closed subset of* X, $\vec{P}$ *be a word-product over* $\mathcal{T}(X)$, $\mathfrak{C}$ *be a closed subset of* $X \times (\mathcal{T}(X) + \{\Box\})^*$ *such that* $\text{args}\,\mathfrak{C}$ *is closed in* $\mathcal{T}(X)$, *where* $\Box$ *is a hole outside* $\mathcal{T}(X)$, *and* S *be a closed subset of* $\mathcal{T}(X)$.

Then, $C^{\overline{?}}(\vec{P}) \subseteq \mathfrak{C}^{\overline{*}}.S$ *iff* $C \times \vec{P} \subseteq \mathfrak{C}[\mathfrak{C}^{\overline{*}}.S]$ *and* $\text{supp}\,\vec{P} \subseteq \mathfrak{C}^{\overline{*}}.S$, *or* $C^{\overline{?}}(\vec{P}) \subseteq \text{args}\,\mathfrak{C} \cup S$.

The expression $\mathfrak{C}[\mathfrak{C}^{\overline{*}}.S]$ in Lemma 11.15 is arguably not syntactically smaller than $\mathfrak{C}^{\overline{*}}.S$, and this would cause some problems in designing an algorithm for inclusion testing.

We shall show below that we can replace $\mathfrak{C}[\mathfrak{C}^{\overline{*}}.S]$ by $\mathfrak{C}[\mathcal{C}]$ for any irreducible closed set $\mathcal{C}$ containing $\mathfrak{C}^{\overline{*}}.S$. The one that will suit us best is a set $\mathcal{T}^{\Box}(X)$, which we now define. While $\mathcal{T}^{\Box}(X)$ is semantically very large – larger than the set of all terms! –, one can think of it denoted by some specific piece of syntax that one would naturally call the *wildcard*.

Definition 11.16 (Wildcard $\mathcal{T}^{\Box}(X)$). *Let* $\mathcal{T}^{\Box}(X)$ *be the disjoint union of* $\mathcal{T}(X)$ *with a fresh element* $\Box$. *Its topology is described by letting the closed subsets of* $\mathcal{T}^{\Box}(X)$ *be those of* $\mathcal{T}(X)$, *plus* $\mathcal{T}^{\Box}(X)$ *itself.*

That is, $\mathcal{T}^{\Box}(X)$ is obtained from $\mathcal{T}(X)$ by adding a new top element to it. Notice that the whole of $\mathcal{T}^{\Box}(X)$ is then irreducible closed in $\mathcal{T}^{\Box}(X)$.

$\mathcal{T}^{\Box}(X)$ is different from $\mathcal{T}(X) + \{\Box\}$. While both spaces have the same elements, the topology of $\mathcal{T}^{\Box}(X)$ is strictly coarser than that of $\mathcal{T}(X) + \{\Box\}$: the only closed subset of $\mathcal{T}^{\Box}(X)$ that contains $\Box$ is $\mathcal{T}^{\Box}(X)$ itself, while any set $F \cup \{\Box\}$, F closed in $\mathcal{T}(X)$, is closed in $\mathcal{T}(X) + \{\Box\}$. In terms of opens, every non-empty open subset of $\mathcal{T}^{\Box}(X)$ contains $\Box$, while any set $U \cup \{\Box\}$, with U open in $\mathcal{T}(X)$, is also open in $\mathcal{T}(X) + \{\Box\}$.

In the following, $\mathfrak{C}[\mathcal{T}^{\Box}(X)]$ is just what it looks like it should be the set of contexts obtained by instantiating a context from $\mathfrak{C}$ by replacing some (but not necessarily all) of its holes by terms.

Lemma 11.17. *Let X be a topological space, C be an irreducible closed subset of X, $\vec{P}$ be a word-product over $\mathcal{T}(X)$, $\mathfrak{C}$ be a closed subset of $X \times (\mathcal{T}(X) + \{\Box\})^*$ such that $args\,\mathfrak{C}$ is closed in $\mathcal{T}(X)$, where $\Box$ is a hole outside $\mathcal{T}(X)$, and S be a closed subset of $\mathcal{T}(X)$.*

Then $C^{\vec{?}}(\vec{P}) \subseteq \mathfrak{C}^{\vec{*}}.S$ iff $C \times \vec{P} \subseteq \mathfrak{C}[\mathcal{T}^{\Box}(X)]$ and $supp\,\vec{P} \subseteq \mathfrak{C}^{\vec{*}}.S$, or $C^{\vec{?}}(\vec{P}) \subseteq args\,\mathfrak{C} \cup S$.

Proof. Using Lemma 11.15, it is enough to show that if $C \times \vec{P} \subseteq \mathfrak{C}[\mathcal{T}^{\Box}(X)]$ and supp $\vec{P} \subseteq \mathfrak{C}^{\vec{*}}.S$, then $C \times \vec{P} \subseteq \mathfrak{C}[\mathfrak{C}^{\vec{*}}.S]$. Indeed, let $(f, \vec{t})$ be any element of $C \times \vec{P}$. By assumption, $f(\vec{t})$ is in $\mathfrak{C}[\mathcal{T}^{\Box}(X)]$. So there is an elementary context $f(\vec{u})$ in $\mathfrak{C}$ such that $\vec{t}$ is obtained from $\vec{u}$ by replacing those u_j that are equal to $\Box$ by some terms t_j, $1 \le j \le m$. Since $t_j \in$ supp $\vec{P}$, $t_j \in \mathfrak{C}^{\vec{*}}.S$. So $(f, \vec{t})$ is in $\mathfrak{C}[\mathfrak{C}^{\vec{*}}.S]$. □

One may wonder above why we used $\mathcal{T}^{\Box}(X)$ instead of, for example, $\mathcal{T}(X)$. This would work, too. But using $\mathcal{T}^{\Box}(X)$ will be what we shall need, in particular in the case where we need to compare two tree iterators.

Let us refine our understanding of the construction $\mathfrak{C}[S]$.

Lemma 11.18. *Let X be a Noetherian space, and S be a closed subset of $\mathcal{T}^{\Box}(X)$. Let $\mathfrak{C}$ be a closed subset of $X \times (\mathcal{T}(X) + \{\Box\})^*$ of the form $\bigcup_{i=1}^m C_i \times Q_i$, where each C_i is irreducible closed in X and each Q_i is a word-product over $\mathcal{T}(X) + \{\Box\}$.*

Then, $\mathfrak{C}[S]$ equals $\bigcup_{i=1}^m C_i \times Q_i[S]$, where for each word-product $Q = e_1 e_2 \cdots e_n$, $Q[S] = e_1[S]e_2[S] \cdots e_n[S]$, and for each atomic expression e, $e[S]$ equals $S^?$ if $e = \{\Box\}^?$, $I^?$ if $e = I^?$ for some irreducible closed subset I of $\mathcal{T}(X)$, $((\mathcal{F} \smallsetminus \{\Box\}) \cup S)^*$ if $e = \mathcal{F}^*$ and $\Box \in \mathcal{F}$, $\mathcal{F}^*$ if $e = \mathcal{F}^*$ and $\Box \notin \mathcal{F}$. □

That is, $\mathfrak{C}[S]$ is obtained from $\mathfrak{C}$ by literally replacing $\Box$ by S throughout. In particular, when $S = \mathcal{T}^{\Box}(X)$: $\{\Box\}^?[\mathcal{T}^{\Box}(X)] = \mathcal{T}^{\Box}(X)^?$, $I^?[\mathcal{T}^{\Box}(X)] = I^?$ when I is irreducible closed in $\mathcal{T}(X)$, $\mathcal{F}^*[\mathcal{T}^{\Box}(X)] = \mathcal{F}^*$ when $\mathcal{F}$ is closed in $\mathcal{T}(X)$, and $\mathcal{F}^*[\mathcal{T}^{\Box}(X)] = \mathcal{T}^{\Box}(X)^*$ when $\Box \in \mathcal{F}$.

This allows us to give our final characterization of inclusion between tree steps and tree iterators, in a way that will lend itself more directly to a recursive algorithm.

Lemma 11.19. *Let X be a topological space, C be an irreducible closed subset of X, $\vec{P}$ be a word-product over $\mathcal{T}(X)$, $\mathfrak{C}$ be a closed subset of $X \times (\mathcal{T}(X) + \{\Box\})^*$ of the form $\bigcup_{i=1}^m C_i \times Q_i$, where each C_i is irreducible closed in X and each Q_i is a word-product over $\mathcal{T}(X) + \{\Box\}$, where $\Box$ is a hole outside $\mathcal{T}(X)$, and S be a closed subset of $\mathcal{T}(X)$.*

Then $C^{\vec{?}}(\vec{P}) \subseteq \mathfrak{C}^{\vec{*}}.S$ iff either: supp $\vec{P} \subseteq \mathfrak{C}^{\vec{*}}.S$, $C \subseteq C_i$ and $\vec{P} \subseteq Q_i[\mathcal{T}^{\Box}(X)]$ for some i, $1 \le i \le m$; or $C^{\vec{?}}(\vec{P}) \subseteq args\,\mathfrak{C} \cup S$.

Proof. Lemma 11.17 applies since args $\mathfrak{C} = \bigcup_{i=1}^n$ supp $Q_i \cap \mathcal{T}(X)$ is closed. Then $\vec{P}$ is irreducible closed (in $\mathcal{T}^{\Box}(X)^*$, not just in $\mathcal{T}(X)^*$) by Lemma 7.7. So $C \times \vec{P}$ is irreducible closed in $X \times \mathcal{T}^{\Box}(X)^*$, whence $C \times \vec{P} \subseteq \mathfrak{C}[\mathcal{T}^{\Box}(X)] = \bigcup_{i=1}^n C_i \times Q_i[\mathcal{T}^{\Box}(X)]$ iff $C \times \vec{P} \subseteq C_i \times Q_i[\mathcal{T}^{\Box}(X)]$ for some i, $1 \le i \le n$, iff $C \subseteq C_i$ and $\vec{P} \subseteq Q_i[\mathcal{T}^{\Box}(X)]$ for some i, $1 \le i \le n$. □

Let us now turn to the third case, where we compare tree iterators and tree steps in the other order.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Lemma 11.20. *Let X be a topological space, C be an irreducible closed subset of X, $\vec{P}$ be a word-product over $\mathcal{T}(X)$, $\mathfrak{C}$ be a closed subset of $X \times (\mathcal{T}(X) + \{\Box\})^*$ such that args $\mathfrak{C}$ is closed in $\mathcal{T}(X)$, where $\Box$ is a hole outside $\mathcal{T}(X)$, and S be a closed subset of $\mathcal{T}(X)$. Assume also that there is an elementary context $f(\vec{u})$ in $\mathfrak{C}$ such that $\Box$ occurs as some element of $\vec{u}$.*

Then $\mathfrak{C}^{\overline{*}}.S \subseteq C^{\overline{?}}(\vec{P})$ iff $\mathfrak{C}^{\overline{*}}.S \subseteq \text{supp}\,\vec{P}$.

Proof. Assume $\mathfrak{C}^{\overline{*}}.S \subseteq C^{\overline{?}}(\vec{P})$. By assumption, there is an elementary context $f(\vec{u})$ in $\mathfrak{C}$ such that $\Box$ occurs as some element of $\vec{u}$. To make things simpler, observe that this implies that $f(\Box)$ is in $\mathfrak{C}$. For every $t \in \mathfrak{C}^{\overline{*}}.S$, $f(t)$ is then again in $\mathfrak{C}^{\overline{*}}.S$. So $f(t)$ is in $C^{\overline{?}}(\vec{P})$. It follows that either $f \in C$ and $t \in \text{supp}\,\vec{P}$, or $f(t) \in \text{supp}\,\vec{P}$. Since supp $\vec{P}$ is closed, it is downward-closed in $\preceq_{\leq}$, so if $f(t) \in \text{supp}\,\vec{P}$, then $t \in \text{supp}\,\vec{P}$. In any case, $t \in \text{supp}\,\vec{P}$. Since t is arbitrary, $\mathfrak{C}^{\overline{*}}.S \subseteq \text{supp}\,\vec{P}$. The converse inclusion is obvious. $\Box$

Note that the condition that there is an elementary context $f(\vec{u})$ in $\mathfrak{C}$ such that $\Box$ occurs as some element of $\vec{u}$ is satisfied in all of the cases 1–3 where we proved $\mathfrak{C}^{\overline{*}}.S$ to be irreducible (Lemma 11.11), since $\mathfrak{C}$ is not $\Box$-linear in the first case and $\Box$-generated in the remaining cases.

Finally, we deal with the case where we try to compare two tree iterators. We do this in several steps, and start, as above, with a lemma with no obvious algorithmic content, but which gives the basic characterization we need. The proof of this heavily depends on irreducibility, as usual. Let @ be the application map from $X \times \mathcal{T}(X)^*$ to $\mathcal{T}(X)$. This sends $(f, \vec{t})$ to $f(\vec{t})$ and is continuous (see Exercise 9.7.47 of Goubault-Larrecq 2013, or Lemma E.6 in Appendix E).

Lemma 11.21. *Let X be a topological space, $\mathfrak{C}$ and $\mathfrak{C}'$ be two closed subsets of $X \times (\mathcal{T}(X) + \{\Box\})^*$ such that args $\mathfrak{C}$ and args $\mathfrak{C}'$ are closed in $\mathcal{T}(X)$, where $\Box$ is a hole outside $\mathcal{T}(X)$, and let S, S' be two closed subsets of $\mathcal{T}(X)$.*

Then $\mathfrak{C}^{\overline{*}}.S \subseteq \mathfrak{C}'^{\overline{*}}.S'$ iff $\mathfrak{C}[\mathfrak{C}^{\overline{*}}.S] \subseteq @^{-1}(\text{args}\,\mathfrak{C}' \cup S') \cup \mathfrak{C}'[\mathcal{T}^{\Box}(X)]$ and args $\mathfrak{C} \cup S \subseteq \mathfrak{C}'^{\overline{*}}.S'$.

Again, we need to make the above lemma clearer, in the case that is of primary interest to us.

Lemma 11.22. *Let X be a topological space, $\mathfrak{C}$ and $\mathfrak{C}'$ be two closed subsets of $X \times (\mathcal{T}(X) + \{\Box\})^*$, where $\Box$ is a hole outside $\mathcal{T}(X)$, and let S, S' be two closed subsets of $\mathcal{T}(X)$. Assume also that $\mathfrak{C}$ is of the form $\bigcup_{i=1}^m C_i \times Q_i$, and that $\mathfrak{C}'$ is of the form $\bigcup_{j=1}^n C'_j \times Q'_j$, where each C_i and each C'_j is irreducible closed in X, and Q_i and Q'_j are word-products over $\mathcal{T}(X) + \{\Box\}$ for each i, $1 \leq i \leq m$, and each j, $1 \leq j \leq n$. Assume finally that $\mathfrak{C}^{\overline{*}}.S$ is irreducible, and that $\Box \in Q_i$ for every i, $1 \leq i \leq m$.*

Then $\mathfrak{C}^{\overline{*}}.S \subseteq \mathfrak{C}'^{\overline{*}}.S'$ iff:

- *either $\mathfrak{C}^{\overline{*}}.S \subseteq \text{args}\,\mathfrak{C}' \cup S'$,*
- *or args $\mathfrak{C} \cup S \subseteq \mathfrak{C}'^{\overline{*}}.S'$, and for every i, $1 \leq i \leq m$, there is a j, $1 \leq j \leq n$, such that $C_i \subseteq C'_j$ and $Q_i[\mathcal{T}^{\Box}(X)] \subseteq Q'_j[\mathcal{T}^{\Box}(X)]$.*

Note that the assumptions that $\mathfrak{C}^{\overline{*}}.S$ is irreducible, and that $\Box \in Q_i$ for every i, $1 \leq i \leq m$, are satisfied as soon as any of the cases 1–3 of Lemma 11.11 hold.

11.4 Intersections of tree steps and tree iterators

We now compute intersections, and we start with tree steps. All missing proofs are in Appendix E.6.

In the following lemma, recall that, by Lemma 7.13, the intersection of any two word-products $\vec{P}$ and $\vec{P}'$ (here, on $\mathcal{T}(X)$) can be expressed as a finite union of word-products $\vec{P}''_j$.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Lemma 11.23. *Let X be a topological space. The intersection of two tree steps* $P = C^{\vec{?}}(\vec{P})$ *and* $P' = C'^{\vec{?}}(\vec{P}')$ *is equal to* $\bigcup_{j=1}^n (C \cap C')^{\vec{?}}(\vec{P}''_j) \cup (supp \vec{P} \cap P') \cup (P \cap supp\, \vec{P}')$*, where* $\vec{P} \cap \vec{P}'$ *is expressed as a finite union* $\bigcup_{j=1}^n \vec{P}''_j$ *of word-products on* $\mathcal{T}(X)$*. If* $C \cap C'$ *can be written as the union of finitely many irreducible closed subsets* C_i*,* $1 \leq i \leq m$*, then* $P \cap P'$ *is also equal to the union of the tree steps* $C_i^{\vec{?}}(\vec{P}''_j)$ $(1 \leq i \leq n, 1 \leq j \leq m)$*, of* $supp\, \vec{P} \cap P'$*, and of* $P \cap supp\, \vec{P}'$*.*

Lemma 11.24. *Let X be a Noetherian space, and S be a closed subset of* $\mathcal{T}(X)$*. Let* $C^{\vec{?}}(\vec{P})$ *be a tree step,* $\mathfrak{C}$ *be a closed subset of* $X \times (\mathcal{T}(X) + \{\Box\})^*$ *of the form* $\bigcup_{j=1}^n C_j \times Q_j$*, where each* C_j *is irreducible closed in X and each* Q_j *is a word-product over* $\mathcal{T}(X) + \{\Box\}$*.*

The intersection of the tree step $P = C^{\vec{?}}(\vec{P})$ *and of the tree iterator* $P' = \mathfrak{C}^{\vec{*}}.S$ *is the union of* $supp\, \vec{P} \cap P'$*, of* $P \cap (S \cup args\, \mathfrak{C})$*, and of* $(C \cap C_j)^{\vec{?}}(\vec{P} \cap Q_j[P'])$*,* $1 \leq j \leq n$*.*

If, for each j, one can write $C \cap C_j$ *as the union of finitely many irreducible subsets* C_{ij}*,* $1 \leq i \leq m_j$*, and if* $\vec{P} \cap Q_j[P']$ *can be expressed as the union of finitely many word-products* $\vec{P}_{\ell j}$*,* $1 \leq \ell \leq q_j$*, then* $P \cap P'$ *is also equal to the union of* $supp\, \vec{P} \cap P'$*, of* $P \cap (S \cup args\, \mathfrak{C})$*, and of* $C_{ij}^{\vec{?}}(\vec{P}_{\ell j})$*,* $1 \leq j \leq n$*,* $1 \leq i \leq m_j$*,* $1 \leq \ell \leq q_j$*.*

Lemma 11.26 below, which deals with intersections of tree iterators, is only valid if the word-products Q_i and Q'_j on $\mathcal{T}(X) + \{\Box\}$ are normalized. A *normalized word-product* on $\mathcal{T}(X) + \{\Box\}$ is of the form $e_1 e_2 \cdots e_n$, where each atomic expression e_i is of the form $P^?$ (with P irreducible closed in $\mathcal{T}(X)$), $\{\Box\}^?$, F^* (with F closed in $\mathcal{T}(X)$), or $\{\Box\}^*$. In other words, we forbid atomic expressions of the form $(F \cup \{\Box\})^*$ where F is a non-empty closed subset of $\mathcal{T}(X)$. Note that the components of a normalized product are either closed subsets of $\mathcal{T}(X)$ (not containing $\Box$) or just $\{\Box\}$.

Lemma 11.25. *Let X be a Noetherian space, Q and* Q' *be two normalized word-products on* $\mathcal{T}(X) + \{\Box\}$*, and P and* P' *be two closed subsets of* $\mathcal{T}(X)$*. The intersection* $Q[P] \cap Q'[P']$ *can be written as a finite union* $\bigcup_{i=1}^n Q_i[P \cap P']$*, where each* Q_i *is a normalized word-product over* $\mathcal{T}(X) + \{\Box\}$*. Explicitly, the set* $\{Q_i \mid 1 \leq i \leq n\}$ *is obtained as* $Meet^{\mathcal{E}}(Q, Q')$*, where, for all components F of Q and* F' *of* Q'*,* $\mathcal{E}(F, F')$ *is a finite set of irreducible closed subsets whose union is:*

- $F \cap F'$ *if* $F, F' \neq \{\Box\}$*;*
- $\{\Box\}$ *if* $F = F' = \{\Box\}$*;*
- $P \cap F'$ *if* $F = \{\Box\}$ *and* $F' \neq \{\Box\}$*;*
- $F \cap P'$ *if* $F \neq \{\Box\}$ *and* $F' = \{\Box\}$*.*

Moreover, for every i, $supp\, Q_i$ *is included in* $(supp Q \cap supp\, Q') \cup (supp Q \cap P') \cup (P \cap supp\, Q') \cup \{\Box\}$*.*

Proof. Since X is Noetherian, $\mathcal{T}(X)$ is, too. By Lemma 7.13, $Q[P] \cap Q'[P']$ is equal to the union of the finitely many elements of $Meet^{\cap}(Q[P], Q'[P'])$.

For every set F that is either closed in $\mathcal{T}(X)$ or equal to $\{\Box\}$, and every closed subset S of $\mathcal{T}(X)$, we write $F[S]$ for F if $F \neq \{\Box\}$, for S otherwise. We observe that for all components F of Q and F' of Q', $F[P] \cap F'[P']$ is the union of the elements $C[P \cap P']$, where C ranges over the elements of $\mathcal{E}(F, F')$: indeed both are equal to $F \cap F'$ in the first case defining $\mathcal{E}$, to $P \cap P'$ in the second case, to $P \cap F'$ in the third case and to $F \cap P'$ in the fourth case. It follows, by induction on the definition of *Meet*, that $Meet^{\cap}(Q[P], Q'[P'])$ is equal to (the union of the finitely many elements of) $Meet^{\mathcal{E}}(Q, Q')[P \cap P']$. It is also easy to check that $Meet^{\mathcal{E}}(Q, Q')$ consists of normalized word-products only, because $\mathcal{E}$ only returns (sets of irreducible) closed subsets of $\mathcal{T}(X)$, or $\{\Box\}$.

For the final part of the Lemma, supp Q_i consists of unions of closed sets as returned by $\mathcal{E}(F, F')$ on components F of Q and F' of Q', as inspection of the $Meet^{\mathcal{E}}$ procedure reveals. If $F, F' \neq \{\Box\}$,

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

$$
\begin{array}{lll}
S ::= 0 \mid P \mid S+S & P ::= C^{\overline{?}}(\vec{P}) \mid \mathfrak{C}^{\overline{*}}.S & \\
 & \vec{P} ::= e_1 e_2 \cdots e_n & e ::= P^{?} \mid S^{*} \\
\mathfrak{C} ::= 0 \mid A \mid \mathfrak{C}+\mathfrak{C} & A ::= C(Q) & \\
 & Q ::= \Box^{*} \mid \vec{P}_1 \Box^{?} \vec{P}_2 \Box^{?} \cdots \Box^{?} \vec{P}_m &
\end{array}
$$

Figure 5. STREs, tree-products ($C \in \mathcal{S}(X)$, $n \geq 0$, $m \geq 1$).

then $\mathcal{E}(F, F') = F \cap F'$ is included in supp $Q \cap$ supp Q'. If $F = F' = \{\Box\}$, then $\mathcal{E}(F, F') = \{\Box\}$. If $F = \{\Box\}$ and $F' \neq \{\Box\}$, then $\mathcal{E}(F, F') = P \cap F'$ is included in $P \cap$ supp Q'. Finally, if $F \neq \{\Box\}$ and $F' = \{\Box\}$, then $\mathcal{E}(F, F') = F \cap P'$ is included in supp $Q \cap P'$. $\Box$

Lemma 11.26. *Let X be a Noetherian space and S and S' be closed subsets of $\mathcal{T}(X)$. Let also $\mathfrak{C}$ (resp., $\mathfrak{C}'$) be a closed subset of $X \times (\mathcal{T}(X) + \{\Box\})^*$ of the form $\bigcup_{i=1}^{m} C_i \times Q_i$ (resp., $\bigcup_{j=1}^{n} C'_j \times Q'_j$), where each C_i and each C'_j is irreducible closed in X and each Q_i and each Q'_j is a normalized word-product over $\mathcal{T}(X) + \{\Box\}$. For all i, j, write $C_i \cap C_j$ as $\bigcup_{k=1}^{p_{ij}} C''_{ijk}$ where each C''_{ijk} is irreducible closed in X, and let $Q''_{ij\ell}$, $1 \leq \ell \leq q_{ij}$ enumerate the elements of $Meet^{\mathcal{E}}(Q_i, Q'_j)$, where the oracle $\mathcal{E}$ is defined in Lemma 11.25.*

Then the intersection of the tree iterators $P = \mathfrak{C}^{\overline{*}}.S$ and $P' = \mathfrak{C}'^{\overline{*}}.S'$ is the tree iterator $\mathfrak{C}''^{\overline{*}}.S''$, where $\mathfrak{C}'' = \bigcup_{i,j,k,\ell} C''_{ijk} \times Q''_{ij\ell}$ and where S'' is the union of $P \cap (\text{args}\, \mathfrak{C}' \cup S')$ and of $(\text{args}\, \mathfrak{C} \cup S) \cap P'$.

11.5 STREs, tree-products

Assume X Noetherian. We now claim that the closed subsets of $\mathcal{T}(X)$ are exactly those denoted by simple tree regular expressions (STREs), and those that are irreducible are exactly those denoted by tree-products, which we now define.

The STREs S and the tree-products P are defined in Figure 5, with additional constraints that we describe in Requirement 11.28 below. We understand those up to associativity and commutativity for $+$, and the fact that 0 is neutral for $+$. Hence, for example, every STRE S can be written as a sum of finitely many tree-products $P_1, \ldots, P_n$, and those are unique up to permutation; in particular $S = 0$ if and only if $n = 0$.

Those expressions have the obvious semantics, once we understand 0 as the empty set, $+$ as union, and in productions such as Q, $\Box$ as the set $\{\Box\}$. The semantics of $A = C(Q)$ is the product of the semantics of C and of Q. We have:

Lemma 11.27. *The languages of S and P are closed subsets of $\mathcal{T}(X)$, the languages of $\vec{P}$ are closed subsets of $\mathcal{T}(X)^*$, the languages of $\mathfrak{C}$ and A are closed subsets of $X \times (\mathcal{T}(X) + \{\Box\})^*$, and the languages of Q are closed subsets of $(\mathcal{T}(X) + \{\Box\})^*$.*

Proof. By induction on syntax. If $P = C^{\overline{?}}(\vec{P})$, then P is closed by Lemma 11.4. If $P = \mathfrak{C}^{\overline{*}}.S$, then P is closed by Lemma 11.9, which applies since args $\mathfrak{C}$ is closed, due to Lemma 11.8. The case of S follows from the fact that finite unions of closed sets are closed. The case of $\vec{P}$ follows from Corollary 7.6, and similarly for Q, noticing that $\{\Box\}$ is closed in $\mathcal{T}(X) + \{\Box\}$. The case of A is because products of closed sets are closed, and the case of $\mathfrak{C}$ follows, again, from the fact that finite unions of closed sets are closed. $\Box$

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Requirement 11.28. We say that $A = C(Q)$ is *syntactically* $\Box$*-linear* if and only if Q is of the form $\vec{P}_1\Box^?\vec{P}_2\Box^? \cdots \Box^?\vec{P}_m$ with $m \leq 2$, and is *syntactically* $\Box$*-generated* if and only if it is of this form with $m \geq 2$, or $Q = \Box^*$.

$\mathfrak{C} = \sum_{i=1}^m C_i(Q_i)$ is syntactically $\Box$-linear (resp., generated) if and only if every $C_i(Q_i)$ is syntactically $\Box$-linear (resp., generated).

We require every subexpression $\mathfrak{C}^{\overline{*}}.S$ to be such that $\mathfrak{C}$ is syntactically $\Box$-generated, $\mathfrak{C} \neq 0$, and one of the following conditions hold:

(1) $\mathfrak{C}$ is not syntactically $\Box$-linear, and $S \neq 0$, or
(2) $\mathfrak{C}$ is syntactically $\Box$-linear and S is a tree-product P;
(3) or $S = 0$.

Lemma 11.29. *The language of every tree-product is an irreducible closed subset of $\mathcal{T}(X)$.*

Proof. We first show that the language of every tree-product P is non-empty. If P is of the form $C^{\overline{?}}(\vec{P})$, then its language contains $f()$, for any f in C. If P is of the form $\mathfrak{C}^{\overline{*}}.S$, where $\mathfrak{C} = C_1(Q_1) + \cdots + C_n(Q_n)$, then since $\mathfrak{C} \neq 0$, n is non-zero. (Remember that we reason up to the fact that $+$ is associative and commutative, and 0 is neutral for $+$.) In that case, pick f from C_1, and we note that $f()$ is in the language of $\mathfrak{C}^{\overline{*}}.S$.

It follows that $(*)$ for every STRE S such that $S \neq 0$, the language of S is non-empty.

We now prove the claim by induction on expressions. We need to show both that the language of every tree-product is irreducible closed, and that every expression of the form Q is an irreducible closed subset of $(\mathcal{T}(X) + \{\Box\})^*$. The latter follows from Lemma 7.7 (every word-product is irreducible closed). The fact that every expression of the form $\mathfrak{C}^{\overline{*}}.S$ is irreducible closed follows from Lemma 11.11, once we observe that every non-syntactically $\Box$-linear context $\mathfrak{C}$ has a non-$\Box$-linear language, and using $(*)$ in case 1; and for the remaining cases, that every syntactically $\Box$-linear context $\mathfrak{C}$ has a $\Box$-linear language, and similarly for $\Box$-generatedness. Every expression of the form $C^{\overline{?}}(\vec{P})$ is irreducible closed by Lemma 11.4. $\Box$

We shall see below that the converse holds.

An example of tree-products is given in the following result, which also serves to state how $\mathcal{T}(X)$ embeds into its completion $\widehat{\mathcal{T}(X)} = \mathcal{S}(\mathcal{T}(X))$.

Lemma 11.30 (Embedding). *Let X be a topological space. The closure $\eta^{\mathcal{S}}_{\mathcal{T}(X)}(t)$ of the term t in $\mathcal{T}(X)$ is defined by structural induction on t by: if $t = f(t_1, t_2, \ldots, t_m)$, then $\eta^{\mathcal{S}}_{\mathcal{T}(X)}(t) = (\eta^{\mathcal{S}}_X f)^{\overline{?}}(\eta^{\mathcal{S}}_{\mathcal{T}(X)}(t_1)^? \eta^{\mathcal{S}}_{\mathcal{T}(X)}(t_2)^? \cdots \eta^{\mathcal{S}}_{\mathcal{T}(X)}(t_n)^?)$.*

Proof. By Lemma 11.4, $(\eta^{\mathcal{S}}_X f)^{\overline{?}}(\eta^{\mathcal{S}}_{\mathcal{T}(X)}(t_1)^? \eta^{\mathcal{S}}_{\mathcal{T}(X)}(t_2)^? \cdots \eta^{\mathcal{S}}_{\mathcal{T}(X)}(t_n)^?)$ is closed, since $\eta^{\mathcal{S}}_{\mathcal{T}(X)}(t_1)$, $\eta^{\mathcal{S}}_{\mathcal{T}(X)}(t_2)$, ..., $\eta^{\mathcal{S}}_{\mathcal{T}(X)}(t_n)$ are closed by induction hypothesis. It also contains t, so it contains the closure of t. Conversely, it is easy to see that whenever $s \in (\eta^{\mathcal{S}}_X(f))^{\overline{?}}(\eta^{\mathcal{S}}_{\mathcal{T}(X)}(t_1)^? \eta^{\mathcal{S}}_{\mathcal{T}(X)}(t_2)^? \cdots \eta^{\mathcal{S}}_{\mathcal{T}(X)}(t_n)^?)$, then $s \preceq_{\leq} t$, so s is in $\eta^{\mathcal{S}}_{\mathcal{T}(X)}(t)$. $\Box$

We can define a syntactic inclusion test $\leq$ as follows, following Lemmas 11.14, 11.19, 11.20, and 11.22. We write semantic inclusion as $\subseteq$. The following notion of syntactic support is meant to mimic the semantical notion of support of Definition 11.3 in the syntax, while the notion of syntactic argument support mimics the argument support of Definition 11.6; accordingly, we use the same notations $\operatorname{supp} \vec{P}$ and $\operatorname{args} \mathfrak{C}$.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Definition 11.31. *Let the* syntactic support *of* $\vec{P} = e_1 e_2 \cdots e_n$ *be* $supp\, \vec{P} = \sum_{i=1}^n supp\, e_i$, *where* $supp\, P^? = P$, $supp\, S^* = S$.

Let $supp'\, Q$ *be* 0 *if* $Q = \Box^*$, $\sum_{i=1}^m supp\, \vec{P}_i$ *if* $Q = \vec{P}_1 \Box^? \vec{P}_2 \Box^? \cdots \Box^? \vec{P}_m$. *(This denotes the intersection of the support of* Q *with* $\mathcal{T}(X)$.*) The* syntactic argument support $args\, \mathfrak{C}$ *of the context* $\mathfrak{C} = \sum_{i=1}^n C_i(Q_i)$ *is* $\sum_{i=1}^n supp'\, Q_i$.

We define $S \leq S'$, $P \leq P'$, *etc., by induction on the sum of the sizes of the two expressions involved, by:*

(1) *for* $S = \sum_{i=1}^m P_i$ *and* $S' = \sum_{j=1}^n P'_j$, $S \leq S'$ *if and only if for every i, there is a j such that* $P_i \leq P'_j$;
(2) *for* $P = C^{\bar{?}}(\vec{P})$ *and* $P' = C'^{\bar{?}}(\vec{P}')$, $P \leq P'$ *if and only if* $C \subseteq C'$ *and* $\vec{P} \leq \vec{P}'$, *or* $P \leq supp\, \vec{P}'$;
(3) *for* $P = C^{\bar{?}}(\vec{P})$ *and* $P' = \mathfrak{C}^{\bar{*}}.S$ *with* $\mathfrak{C} = \sum_{i=1}^m C_i(Q_i)$, $P \leq P'$ *if and only if* $supp\, \vec{P} \leq P'$ *and* $C \subseteq C_i$, $\vec{P} \leq Q_i$ *for some i*, $1 \leq i \leq m$, *or* $P \leq args\, \mathfrak{C} + S$;
(4) *for* $P = \mathfrak{C}^{\bar{*}}.S$ *and* $P' = C^{\bar{?}}(\vec{P})$, $P \leq P'$ *if and only if* $P \leq supp\, \vec{P}$;
(5) *for* $P = \mathfrak{C}^{\bar{*}}.S$ *and* $P' = \mathfrak{C}'^{\bar{*}}.S'$ *with* $\mathfrak{C} = \sum_{i=1}^m C_i(Q_i)$ *and* $\mathfrak{C}' = \sum_{j=1}^n C'_j(Q'_j)$, $P \leq P'$ *if and only if* $P \leq args\, \mathfrak{C}' + S'$, *or* $args\, \mathfrak{C} + S \leq P'$ *and for every i there is a j such that* $C_i \subseteq C'_j$ *and* $Q_i \leq Q_j$;
(6) $Q \leq Q'$ *if and only if Q is less than* Q' *as a word-product (see Lemma 7.10), comparing letters by* $\leq$, *recursively, and considering* $\Box$ *as a letter above all others. (This case subsumes tests of the form* $\vec{P} \leq Q$ *as well, for instance, since any* $\vec{P}$ *is a Q.)*

One easily sees that $S \leq S'$ if and only if (the language of) S is included in S', $P \leq P'$ if and only if P is included in P', etc., by induction. One should pay attention to the fact that $Q \leq Q'$ is equivalent to the inclusion of $Q[\mathcal{T}^{\Box}(X)]$ into $Q'[\mathcal{T}^{\Box}(X)]$, as needed in cases 3 and 5 above. This is why the comparison of expressions of the form Q assumes that $\Box$ is a letter outside $\mathcal{T}(X)$, and above all elements of $\mathcal{T}(X)$ (see Item 6): indeed, $\Box$ denotes the whole set $\mathcal{T}^{\Box}(X)$ there.

In case Item 6 of the definition is not clear enough, here is a complete explanation. This will also help understand how the conditions $\vec{P} \leq Q_i$ of Item 3 and $Q_i \leq Q_j$ of Item 5 are checked. We let $\epsilon \leq Q'$ for every Q', $Q \leq \epsilon$ iff $Q = \epsilon$, and, if $Q = e_1 Q_1$ and $Q' = e'_1 Q'_1$, then $Q \leq Q'$ if and only if:

(1) $e_1 \not\sqsubseteq e'_1$ and $Q \leq Q'_1$,
(2) or e_1 is of the form $E^?$, e'_1 is of the form $E'^?$, $e_1 \sqsubseteq e'_1$ and $Q_1 \leq Q'_1$,
(3) or e'_1 is of the form E'^*, $e_1 \sqsubseteq e'_1$ and $Q_1 \leq Q'$,
(4) or $e_1 = \emptyset^*$ and $Q_1 \leq Q'$.

Additionally, $\sqsubseteq$ is defined by: $P^? \sqsubseteq P'^?$ iff $P \leq P'$, $P^? \sqsubseteq \Box^?$ and $\Box^? \sqsubseteq \Box^?$ are always true, $\Box^? \sqsubseteq P'^?$ is always false; $S^* \sqsubseteq S'^*$ iff $S \leq S'$, $S^* \sqsubseteq \Box^*$ and $\Box^* \sqsubseteq \Box^*$ are always true, $\Box^* \sqsubseteq S'^*$ is always false; $P^? \sqsubseteq S^*$ iff $P \leq S$, $P^? \sqsubseteq \Box^*$ and $\Box^? \sqsubseteq \Box^*$ are always true, $\Box^? \subseteq S^*$ is always false; $S^* \sqsubseteq P^?$ if and only if $S = \emptyset$; similarly $S^* \sqsubseteq \Box^?$ if and only if $S = \emptyset$; and $\Box^* \sqsubseteq P^?$, $S^* \sqsubseteq \Box^?$ (with S non-empty), $\Box^* \sqsubseteq \Box^?$ are always false. One should be conscious that, for instance, $P^? \sqsubseteq \Box^?$ is always true (and the converse inequality is false unless $P = \Box$), even when P contains further subexpressions with occurrences of $\Box$, for example, $((\{f\} \times \{\Box\}^*)^{\bar{*}}.S)^? \sqsubseteq \Box^?$.

Lemma 11.32. *Let X be a Noetherian space. Then the language of P is included in that of* P' *if and only if* $P \leq P'$. *Similarly for* $S \leq S'$. ☐

Corollary 11.33. *Let X be a Noetherian space. Inclusion of word-products, resp. of STREs, on X, can be checked in polynomial time modulo an oracle testing inclusion of closed subsets of X.*

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Proof. Here memoization (Michie 1968) is easier to implement and analyze than dynamic programming. We write a recursive program `incl` that takes two word-products P and P' and returns true if $P \leq P'$, false otherwise. This program maintains a table A, initially empty, that maps pairs of word-products to Booleans. The program `incl`, applied to P and P', first checks whether (P, P') is in A, and if so returns immediately the Boolean associated with it in A; otherwise, it tests whether $P \leq P'$, depending on the shape of P and P', using the corresponding case (1 through 6), as given in Definition 11.31; once this is done, it adds (P, P') to A, associated with the Boolean value true if P was found to be below P', false otherwise. Each case, 1 through 6, involves further inclusion tests between word-products, which are evaluated by calling `incl` recursively.

Instead of describing all cases, we will focus on case 3 below. It will be practical to consider another program `Incl`, defined in mutual recursion with `incl` by: `Incl`(S, S') is true, where $S = P_1 + \cdots + P_m$ and $S' = P'_1 + \cdots + P'_n$, if and only if for every $i \in \{1, \ldots, m\}$, there is a $j \in \{1, \ldots, n\}$ such that `incl`(P_i, P'_j) returns true; `Incl` returns true on every pair whose second component is $\Box$, and false on every pair $(\Box, S')$ where S' is an STRE.

In case 3, then P is of the form $C^?(\vec{P})$ and P' is of the form $\mathfrak{C}^{\bar{*}}.S$ with $\mathfrak{C} = \sum_{i=1}^m C_i(Q_i)$. Following the definition of case 3, `incl` does the following.

(1) First, it tests whether $\text{supp}\, \vec{P} \leq P'$. We refer to Definition 11.3 for the definition of $\text{supp}\, \vec{P}$. Accordingly, writing $\vec{P}$ as $e_1 e_2 \cdots e_n$, testing whether $\text{supp}\, \vec{P} \leq P'$ means testing whether, for every $i \in \{1, \ldots, n\}$:
  — $P'' \leq P'$, in case e_i is of the form $P''^?$;
  — or $P''_j \leq P'$ for every $j \in \{1, \ldots, m\}$, in case e_i is of the form S^*, where $S = P''_1 + \cdots + P''_m$.
  All the tests $P'' \leq P'$ (in the first case), and $P''_j \leq P'$ (in the second case) are done by calling `incl` recursively.

(2) If item 1 succeeded, `incl` then tests whether $C \subseteq C_i$ and $\vec{P} \leq Q_i$ for some i, $1 \leq i \leq m$. This is done by enumerating the indices $i \in \{1, \ldots, m\}$, testing whether $C \subseteq C_i$ for each one (using the given oracle that tests inclusion of closed subsets of X) and whether $\vec{P} \leq Q_i$. For the latter, we use Item 6 of Definition 11.3: we need to compare $\vec{P}$ and Q_i as word-products; this allows us to use the dynamic programming procedure given after Lemma 7.9, fed with `Incl` as an oracle. (You may wish to return to the explanation before Lemma 11.32 in order to better understand where recursive calls to `Incl` are involved; namely in the $\leq$ tests involved in the definition of $\sqsubseteq$.)

(3) If item 1 failed, `incl` instead tests whether $P \leq \text{args}\, \mathfrak{C} + S$. In other words, using the fact that P is irreducible, it tests whether $P \leq S$ (using `Incl`), or whether $P \leq \text{args}\, \mathfrak{C}$. For the latter, we use Lemma 11.8, and therefore we test whether $P \leq \text{supp}\, Q_i \cap \mathcal{T}(X)$ for some i, $1 \leq i \leq m$. Writing Q_i as $e_1 e_2 \cdots e_n$, this means testing whether $P \leq \text{supp}\, e_j$ for some j, $1 \leq j \leq n$, such that $e_j \neq \Box^?, \Box^*$. Finally, if e_j is of the form $P'^?$, we test whether $P \leq \text{supp}\, e_j = P'$ by calling `incl` recursively on (P, P'), and if it is of the form S'^*, we test whether $P \leq \text{supp}\, e_j = S'$ by calling `Incl` on (P, S').

All other cases are dealt with similarly. Up to a multiplicative constant, the time complexity a memoized procedure such as `incl` on input (P, P') is bounded by the product of two values: the first one is the number of pairs on which `incl` can be called recursively, and this is bounded by the product of the sizes of P and P'; the second one is the time complexity of each case (such as case 3, explained above), including table lookups and table updates on A, but counting the complexity of each recursive call to `incl` (including those obtained through an intermediate call to `Incl`) as one. Analyzing each case shows that the second value is also polynomial in the sizes of P and P', whence the conclusion. $\Box$

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Lemma 11.34. *Let X be a Noetherian space. Given two tree-products P, P', one can express their intersection as a finite sum of tree-products. If $(S, [\![\_]\!], \trianglelefteq, \tau, \wedge)$ is an S-representation of X, then this intersection can be implemented as a map $\wedge'$ modulo an oracle that implements $\wedge$.*

Proof. Let $|E|$ denote the number of subexpressions of E that are tree-products, for every kind of expression in Figure 5. Explicitly: $|0| = 0$, $|S_1 + S_2| = |S_1| + |S_2|$, $|C^{\overline{?}}(\vec{P})| = 1 + |\vec{P}|$, $|\mathfrak{C}^{\overline{*}}.S| = 1 + |\mathfrak{C}| + |S|$, $|e_1 e_2 \cdots e_n| = \sum_{i=1}^n |e_i|$, $|P^?| = |P|$, $|S^*| = |S|$, $|\mathfrak{C}_1 + \mathfrak{C}_2| = |\mathfrak{C}_1| + |\mathfrak{C}_2|$, $|C(Q)| = |Q|$, $|\Box^*| = 0$, $|\vec{P}_1 \Box^? \vec{P}_2 \Box^? \cdots \Box^? \vec{P}_m| = \sum_{j=1}^m |\vec{P}_j|$.

The fact that $P \cap P'$ can be expressed as a finite sum of tree-products is by induction on $|P| + |P'|$. When both are tree steps, say $P = C^{\overline{?}}(\vec{P})$ and $P' = C'^{\overline{?}}(\vec{P}')$, we use Lemma 11.23, using the subprocedure $Meet^{\mathcal{E}}$ of Lemma 7.13 to compute the intersection of the word-products $\vec{P}$ and $\vec{P}'$ as a finite union $\bigcup_{j=1}^n \vec{P}''_j$, where $\mathcal{E}$ merely computes the intersection of a component of $\vec{P}$ and of a component of $\vec{P}'$, by induction.

When P is a tree step $C^{\overline{?}}(\vec{P})$ and P' is a tree iterator $P' = \mathfrak{C}^{\overline{*}}.S$ (or conversely), we use Lemma 11.24 instead. Again, we use $Meet^{\mathcal{E}}$ in order to compute $\vec{P} \cap Q_j[P']$, where $\mathcal{E}$ is intersection again. Note that the components of $Q_j[P']$ are either components of Q_j, which are finite sums of tree-products that are strictly smaller than P' or equal to P' (at the positions where Q_j holds a $\Box^?$ or a $\Box^*$). In any case, $\mathcal{E}$ is only ever applied to pairs of word-products (P_0, P'_0) such that $|P_0| < |P|$ and $|P'_0| \leq |P'|$, so the induction hypothesis applies and $\mathcal{E}(P_0, P'_0)$ therefore computes $P_0 \cap P'_0$, as desired. The fact that the induction hypothesis really applies is probably clearer with an example. Imagine that $\vec{P} = P_1^? S_1^*$ and $Q_j = P_2^? \Box^? S_2^* \Box^* P_3^?$. Following Lemma 7.13, we compute $Meet^{\mathcal{E}}(\vec{P}, Q_j[P'])$ as follows. We use the first clause that defines it, and we are led (among other things) to compute $\mathcal{E}(P_1, P_2)$ – since $|P_1| < |P|$ and $|P_2| < |P'|$, the induction hypothesis applies – and $Meet^{\mathcal{E}}(S_1^*, (\Box^? S_2^* \Box^* P_3^?)[P'])$. For the latter, we use the third clause, and we are led to compute $\mathcal{E}(S_1, P')$ – here $|S_1| < |P|$ and the size of the second argument is equal to $|P'|$, so the induction hypothesis applies again – and then $Meet^{\mathcal{E}}(S_1^*, (S_2^* \Box^* P_3^?)[P'])$ and $Meet^{\mathcal{E}}(\epsilon, (\Box^? S_2^* \Box^* P_3^?)[P'])$. The latter is trivial. For the former, we are led to compute $\mathcal{E}(S_1, S_2)$ (by induction hypothesis, since $|S_1| < |P|$ and $|S_2| < |P'|$) and then $Meet^{\mathcal{E}}(S_1^*, (\Box^* P_3^?)[P'])$ and $Meet^{\mathcal{E}}(\epsilon, (S_2^* \Box^* P_3^?)[P'])$. The latter is again trivial, so we look at the former. This requires us to compute $\mathcal{E}(S_1, P')$ – and the induction hypothesis applies again since $|S_1| < |P|$ and the size of the second argument is equal to $|P'|$ – and then $Meet^{\mathcal{E}}(S_1^*, (P_3^?)[P'])$ and $Meet^{\mathcal{E}}(\epsilon, (\Box^* P_3^?)[P'])$. The last call to $\mathcal{E}$ that we need has arguments S_1 and P_3, with $|S_1| < |P|$ and $|P_3| < |P'|$.

Finally, when P is a tree iterator $\mathfrak{C}^{\overline{*}}.S$ and P' is a tree iterator $\mathfrak{C}'^{\overline{*}}.S'$, we use Lemma 11.26. (Note that the syntax of Figure 5 forces all word-products Q to be normalized, so that Lemma 11.26 indeed applies.) This involves calling $Meet^{\mathcal{E}}$ again, where $\mathcal{E}$ is now the oracle of Lemma 11.25. The latter is only ever applied to pairs of word-products (P_0, P'_0) such that $|P_0| < |P|$ and $|P'_0| \leq |P'|$, or such that $|P_0| \leq |P|$ and $|P'_0| < |P'|$, so that the induction hypothesis applies again.

However, the result $\mathfrak{C}''^{\overline{*}}.S''$ given by Lemma 11.26 may fail to satisfy Requirement 11.28. We repair this using a rewriting process akin to the relation $\to_1$ of Figure 4. (This is very similar, except that (R11) does not have an equivalent here, since there is no such thing as a function of arity 0.) First, we split $\mathfrak{C}''$ into a sum $\mathfrak{C}''_0 + \mathfrak{C}''_1$, where $\Box$ does not occur at all in $\mathfrak{C}''_0$ (hence, $\mathfrak{C}''_0$ is of the form $\sum_a C''_a(\vec{P}_a)$) and $\Box$ occurs in each summand of $\mathfrak{C}''_1$ (so $\mathfrak{C}''_1$ is syntactically $\Box$-generated). We can then rewrite $\mathfrak{C}''^{\overline{*}}.S''$ as $\mathfrak{C}_1''^{\overline{*}}.(S'' + \sum_a C_a''^{\overline{?}}(\vec{P}_a))$ (mimicking rule (R6)). Modulo this rewrite, we can assume that we have expressed $P \cap P'$ as $\mathfrak{C}''^{\overline{*}}.S''$, where $\mathfrak{C}''$ is syntactically $\Box$-generated. If $\mathfrak{C}'' = 0$, then we simplify that to S''. Otherwise, if $\mathfrak{C}''$ is syntactically $\Box$-linear, then $\mathfrak{C}''^{\overline{*}}.\_$ distributes

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

over + (as with rule ($R12$)): writing S'' as $\sum_b P''_b$, we return $\sum_b \mathfrak{C}''^{\overline{*}}.P''_b$. In all other cases, we return $\mathfrak{C}''^{\overline{*}}.S''$. Whatever the situation, the expression we return satisfies Requirement 11.28. □

Proposition 11.35. *Let X be a Noetherian space. The closed subsets of $\mathcal{T}(X)$ are exactly the languages of STREs, and the irreducible closed subsets of $\mathcal{T}(X)$ are exactly the languages of tree-products.*

Proof. One direction follows from Lemma 11.29. Conversely, we show that every closed subset of terms is expressible as some STRE.

For every simple tree expression π, the complement of π can be expressed as an STRE, by induction on (the syntax of) π and following Lemma 11.13. We deal with the second and sixth items of that lemma; the other cases are similar. In the second case, the complement of π is $X^{\overline{?}}(\complement\pi_1^*)$. Since X is Noetherian, we can express X itself as a finite union of irreducible closed subsets C_i, $1 \leq i \leq m$. By induction hypothesis, $\complement\pi_1$ is expressible as some STRE S. Then π is expressible as $\sum_{i=1}^m C_i^{\overline{?}}(S^*)$. In the sixth case, the complement of π is equal to $((F \times \{\Box\}^*) \cup (X \times (\complement\pi_1^*\{\Box\}^?\complement\pi_2^*\{\Box\}^? \cdots \{\Box\}^?\complement\pi_n^*)))^{\overline{*}}.\emptyset$, where F is closed in X and $n \geq 2$. Write F as a finite union of irreducible closed subsets C'_k of X, $1 \leq k \leq p$, and each $\complement\pi_j$, $1 \leq j \leq n$, as some STRE S_j. Recall that $X = \bigcup_{i=1}^m C_i$. Then π is expressible as the expression $\mathfrak{C}''^{\overline{*}}.S''$, where $\mathfrak{C}'' = \sum_{k=1}^p C'_k(\Box^*) + \sum_{i=1}^n C_i(S_1^*\Box^?S_2^*\Box^? \cdots \Box^?S_n^*)$ and $S'' = 0$. Observe that this satisfies Requirement 11.28.

Since X is Noetherian, and simple tree expressions form a base of the tree topology (Proposition 11.1), every open subset of $\mathcal{T}(X)$ is a finite union of tree expressions. Therefore every closed subset is a finite intersection of STREs, hence itself an STRE by Lemma 11.34.

Finally, given an irreducible closed subset of $\mathcal{T}(X)$, expressed as an STRE $S = \sum_{i=1}^n P_i$, that irreducible closed subset must be the language of some P_i – by irreducibility. □

We can now conclude.

Theorem 11.36 (S-representation, trees). *Let X be a Noetherian space, $X' = \mathcal{T}(X)$, and $(S, [\![\_]\!], \trianglelefteq, \tau, \wedge)$ be an S-representation of X. Then $(S', [\![\_]\!]', \trianglelefteq', \tau', \wedge')$ is an S-representation of X', where $[\![\_]\!]'$ is defined in terms of auxiliary maps that we write $[\![\_]\!]^\circ$:*

(A) S' is the collection of all tree-products over the signature S;
(B) $\trianglelefteq'$ is implemented using the procedure outlined in Definition 11.31, where the inclusions between closed subsets of X are decided using $\trianglelefteq$ (and Lemma 5.2);
(C) τ' is $(\sum_{i=1}^n C_i(\Box^*))^{\overline{*}}.\emptyset$, where $\tau = \{C_1, \ldots, C_n\}$;
(D) $\wedge'$ is defined by the procedure of Lemma 11.34.

One should not be surprised of the shape of τ', which is non-empty despite the mention of the empty set: see Remark 11.7.

12. Conclusion

We have developed the first comprehensive theory of (downward-)closed subsets, as required for a general understanding of forward analysis techniques of WSTS. This generalizes previous domain proposals on tuples of natural numbers, on words, on multisets, allowing for nested datatypes, and infinite alphabets.

We have also done this on new domains such as trees, words under prefix, infinite powersets, and Noetherian rings or $\mathbb{Q}^k$.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Each of these domains is effective, in the sense that each has finite presentations with a decidable ordering.

We have also shown how the notion of sobrification $\mathcal{S}(X)$ was in a sense inevitable (Section 4). In the special case of wqos, the latter coincides with the ideal completion of X, and it is important to stress that we have, in particular, characterized the shapes of ideals in several wqos of interest in verification ($\mathbb{N}^k$, X^*, $X^{\circledast}$, $\mathcal{T}(X)$). Ideals in wqos are growingly appearing to be a central concept in verification.

The natural generalization from wqos to Noetherian spaces was also naturally in order, not only because it allows us to deal with extra constructions (words under prefix, rings, infinite powersets), but also because the proofs, as well as the computations (witness Lemma 5.2) intimately rely on irreducibility, a concept that occurs naturally from the study of sobrification.

Acknowledgements. We thank Philippe Schnoebelen, Simon Halfon, Sylvain Schmitz, Ranko Lazic, and all those who expressed interest in this work since 2009 and encouraged us into submitting this paper again – not just out of friendship but also because they need the results. Simon Halfon found the polynomial time decision procedure mentioned in Lemma 8.6 and is gratefully acknowledged for this. The referees have done a remarkably thorough job in rereading this paper, and we thank them for that. Any remaining mistake is, of course, ours.

Note

1 The state space was erroneously claimed to be $\mathbb{C}^k$ in that paper.

References

Abdulla, P. A., Bouajjani, A. and Jonsson, B. (1998). On-the-fly analysis of systems with unbounded, lossy FIFO channels. In: *CAV'98*, LNCS, vol. 1427, Vancouver, Canada, Springer Verlag, 305–318.

Abdulla, P. A., Čerans, K., Jonsson, B. and Tsay, Y.-K. (1996). General decidability theorems for infinite-state systems. In: *LICS*, 313–321.

Abdulla, P. A., Čerāns, K., Jonsson, B. and Tsay, Y.-K. (2000). Algorithmic analysis of programs with well quasi-ordered domains. *Information and Computation* **160** (1–2) 109–127.

Abdulla, P. A., Collomb-Annichini, A., Bouajjani, A. and Jonsson, B. (2004a). Using forward reachability analysis for verification of lossy channel systems. *Formal Methods in System Design* **25** (1) 39–65.

Abdulla, P. A., Deneux, J., Mahata, P. and Nylén, A. (2004b). Forward reachability analysis of timed Petri nets. In: Lakhnech, Y. and Yovine, S. (eds.) *FORMATS/FTRTFT*, LNCS, vol. 3253, Springer Verlag, 343–362.

Abdulla, P. A. and Jonsson, B. (1993). Verifying programs with unreliable channels. In: *LICS'93*, 160–170.

Abramsky, S. and Jung, A. (1994). Domain theory. In: Abramsky, S., Gabbay, D. M. and Maibaum, T. S. E. (eds.) *Handbook of Logic in Computer Science*, vol. 3, OUP, 1–168.

Acciai, L. and Boreale, M. (2012). Deciding safety properties in infinite-state pi-calculus via behavioural types. *Information and Computation* **212** 92–117.

Adams, W. W. and Loustaunau, P. (1994). *An Introduction to Gröbner Bases*, Graduate Studies in Mathematics, vol. 3, American Mathematical Society, 289.

Arnold, A. and Latteux, M. (1978). Récursivité et cônes rationnels fermés par intersection. *CALCOLO* **15** (4) 381–394. doi: https://doi.org/10.1007/BF02576519.

Baader, F. and Nipkow, T. (eds.) (1998). *Term Rewriting and All That*, Cambridge University Press, 301.

Bachmair, L. and Plaisted, D. A. (1985). Termination orderings for associative-commutative rewriting systems. *Journal of Logic and Computation* **1** (4) 329–349.

Blondin, M., Finkel, A. and Goubault-Larrecq, J. (2017a). Forward analysis for WSTS, part III: Karp-Miller trees. In: Lokam, S. and Ramanujam, R. (eds.) *Proceedings of the 37th Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS'17)*, Leibniz International Proceedings in Informatics, vol. 93, Kanpur, India, Leibniz-Zentrum für Informatik, 16:1–16:15.

Blondin, M., Finkel, A. and Goubault-Larrecq, J. (2017b). Forward analysis for WSTS, Part III: Karp-Miller trees. Logical Methods in Computer Science 16 (2), 2020. doi: 10.23638/LMCS-16(2:13)2020. Long and improved version of Blondin et al. (2017a).

Bojańczyk, M., Muscholl, A., Schwentick, T. and Ségoufin, L. (2009). Two-variable logic on data trees and XML reasoning. *Journal of the ACM* **56** (3) 13:1–13:48.

Buchberger, B. and Loos, R. (1982–1983). Algebraic simplification. In: Buchberger, B., Collins, G. E., Loos, R. and Albrecht, R. (eds.) *Computer Algebra, Symbolic and Algebraic Computation*, Springer Verlag.

Comon, H., Dauchet, M., Gilleron, R., Jacquemard, F., Lugiez, D., Tison, S. and Tommasi, M. (2004). Tree automata techniques and applications. www.grappa.univ-lille3.fr/tata.

Cormen, T. H., Leiserson, C. E., Rivest, R. L. and Stein, C. (2001). *Introduction to Algorithms*, 2nd edition, MIT Press and McGraw-Hill.

de Groote, P., Guillaume, B. and Salvati, S. (2004). Vector addition tree automata. In: *LICS'04*, IEEE Computer Society, 64–73.

Dickson, L. E. (1913). Finiteness of the odd perfect and primitive abundant numbers with *n* distinct prime factors. *American Journal of Mathematics* **35** (4) 413–422.

Emerson, E. A. and Namjoshi, K. S. (1998). On model checking for non-deterministic infinite-state systems. In: *LICS'98*, 70–80.

Erdős, P. and Tarski, A. (1943). On families of mutually exclusive sets. *Annals of Mathematics* **44** 315–329.

Faith, C. C. (1999). *Rings and Things and a Fine Array of Twentieth Century Associative Algebra,* American Mathematical Society.

Finkel, A. (1987). A generalization of the procedure of Karp and Miller to well structured transition system. In: Ottmann, T. (ed.) *Proceedings of the 14th International Colloquium on Automata, Languages and Programming (ICALP'87)*, Lecture Notes in Computer Science, vol. 267, Karlsruhe, Germany, Springer-Verlag, 499–508.

Finkel, A. (1990). Reduction and covering of infinite reachability trees. *Information and Computation* **89** (2) 144–179.

Finkel, A. and Goubault-Larrecq, J. (2009). Forward analysis for WSTS, part I: Completions. In: Albers, S. and Marion, J.-Y. (eds.) *Proceedings of the 26th Annual Symposium on Theoretical Aspects of Computer Science (STACS'09)*, Leibniz International Proceedings in Informatics, vol. 3, Freiburg, Germany, Leibniz-Zentrum für Informatik, 433–444.

Finkel, A. and Goubault-Larrecq, J. (2012). Forward analysis for WSTS, part II: Complete WSTS. *Logical Methods in Computer Science* **8** (3:28).

Finkel, A. and Schnoebelen, P. (2001). Well-structured transition systems everywhere! *Theoretical Computer Science* **256** (1–2) 63–92.

Ganty, P., Raskin, J.-F. and van Begin, L. (2006). A complete abstract interpretation framework for coverability properties of WSTS. In: *VMCAI'06*, LNCS, vol. 3855, Springer Verlag, 49–64.

Geeraerts, G., Raskin, J.-F. and van Begin, L. (2006). Expand, enlarge and check: New algorithms for the coverability problem of WSTS. *Journal of Computer and System Sciences* **72** (1) 180–203.

Gierz, G., Hofmann, K. H., Keimel, K., Lawson, J. D., Mislove, M. and Scott, D. S. (2003). Continuous lattices and domains. In: *Encyclopedia of Mathematics and its Applications*, vol. 93, CUP.

Goubault-Larrecq, J. (2007). On Noetherian spaces. In: *LICS'07*, 453–462.

Goubault-Larrecq, J. (2010). Noetherian spaces in verification. In: Abramsky, S., Meyer auf der Heide, F. and Spirakis, P. (eds.) *Proceedings of the 37th International Colloquium on Automata, Languages and Programming (ICALP'10) – Part II*, Lecture Notes in Computer Science, vol. 6199, Bordeaux, France, Springer, 2–21.

Goubault-Larrecq, J. (2013). *Non-Hausdorff Topology and Domain Theory, Selected Topics in Point-Set Topology*, New Mathematical Monographs, vol. 22, Cambridge University Press.

Goubault-Larrecq, J. (2019). Spaces with no infinite discrete subspace. *Topology Proceedings* **53** 27–36.

Goubault-Larrecq, J. and Schmitz, S. (2016). Deciding piecewise testable separability for regular tree languages. In: Calamoneri, T., Gorla, D., Rabani, Y., Sangiorgi, D. and Mitzenmacher, M. (eds.) *43rd International Colloquium on Automata, Languages, and Programming*, Leibniz-Zentrum für Informatik, Leibniz International Proceedings in Informatics, vol. 95, 97:1–97:15.

Grieco, M. and Zucchetti, B. (1989). How to decide whether a polynomial ideal is primary or not. In: *Proceedings of the 5th International Conference on Applied Algebra, Algebraic Algorithms and Error-Correcting Codes (AAECC-5), 1987*, LNCS, vol. 356, Menorca, Spain, Springer-Verlag, 258–268.

Grothendieck, A. (1960). *Éléments de géométrie algébrique (rédigés avec la collaboration de Jean Dieudonné): I. Le langage des schémas*, vol. 4, Publications mathématiques de l'I.H.É.S, 5–228.

Halfon, S. (2018). *On Effective Representations of Well Quasi-Orderings*. Phd thesis, ENS Paris-Saclay, Université Paris-Saclay.

Higman, G. (1952). Ordering by divisibility in abstract algebras. *Proceedings of the London Mathematical Society* **2** (7) 326–336.

Hoffmann, R.-E. (1979a). On the sobrification remainder $^{s}X - X$. *Pacific Journal of Mathematics* **83** (1), 145–156.

Hoffmann, R.-E. (1979b). Sobrification of partially ordered sets. *Semigroup Forum* **17** 123–138.

Jacobé de Naurois, P. (2014). Coverability in a nonfunctional extension of BVASS. hal-00947136. https://hal.archives-ouvertes.fr/hal-00947136.

Jacquemard, F., Ségoufin, L. and Dimino, J. (2016). $\mathrm{FO}^2(<, +1, \sim)$ on data trees, data tree automata and branching vector addition systems. *LMCS* **12** (2) 1–28.

Kabil, M. and Pouzet, M. (1992). Une extension d'un théorème de P. Jullien sur les âges de mots. *Informatique théorique et applications* **26** (5) 449–482.

Karp, R. M. and Miller, R. E. (1969). Parallel program schemata. *Journal of Computer and System Sciences* **3** (2) 147–195.

Kosaraju, S. R. (1982). Decidability of reachability in vector addition systems (preliminary version). In: *Proceedings of the 14th Annual ACM Symposium on the Theory of Computing (STOC'82)*, San Francisco, California, USA, ACM Press, 267–281.
Kruskal, J. B. (1960). Well-quasi-ordering, the tree theorem, and Vazsonyi's conjecture. *Transactions of the American Mathematical Society* **95** (2) 210–225.
Lambert, J.-L. (1992). A structure to decide reachability in Petri nets. *Theoretical Computer Science* **99** (1) 79–104.
Laplagne, S. (2006). Computation of the minimal associated primes. In: Decker, W., Dewar, M., Kaltofen, E. and Watt, S. (eds.) *Challenges in Symbolic Computation Software*, Dagstuhl Seminar Proceedings, vol. 06271, Internationales Begegnungs- und Forschungszentrum fuer Informatik (IBFI), Schloss Dagstuhl, Germany. http://drops.dagstuhl.de/opus/volltexte/2006/774 (date of citation: 2006-01-01).
Lazič, R., Newcomb, T., Ouaknine, J., Roscoe, A. W. and Worrell, J. (2008). Nets with tokens which carry data. *Fundamenta Informaticae* **88** (3) 251–274.
Leroux, J. and Schmitz, S. (2015). Demystifying reachability in vector addition systems. In: *LICS'15*, 56–67.
Mayr, E. W. (1981). An algorithm for the general Petri net reachability problem. In: *Proceedings of the 13th Annual ACM Symposium on the Theory of Computing (STOC'81)*, Milwaukee, Wisconsin, USA, ACM Press, 238–246.
Mayr, R. (2003). Undecidable problems in unreliable computations. *Theoretical Computer Science* **297** (1–3) 337–354.
Meyer, R. (2008). On boundedness in depth in the pi-calculus. *IFIP TCS* **273** 477–489.
Michie, D. (1968). Memo functions and machine learning. *Nature* **218** 19–22.
Mislove, M. (1998). Topology, domain theory and theoretical computer science. *Topology and Its Applications* **89** 3–59.
Müller-Olm, M. and Seidl, H. (2002). Polynomial constants are decidable. In: Hermenegildo, M. V. and Puebla, G. (eds.) *Proceedings of the 9th International Symposium on Static Analysis (SAS'02)*, LNCS, vol. 2477, Springer-Verlag, 4–19.
Parikh, R. (1966). On context-free languages. *Journal of the ACM* **13** (4) 570–581.
Pnueli, A. (1977). The temporal logic of programs. In: *18th Annual Symposium on Foundations of Computer Science*, Providence, RI, USA, IEEE Computer Society, 46–57.
Rabinowitch, S. (1929). Zum Hilbertschen Nullstellensatz. *Mathematische Annalen* **102** 520.
Rambow, O. (1994). Multiset-valued linear index grammars: Imposing dominance constraints on derivations. In: *ACL'94*, ACL Press, 263–270.
Schmitz, S. (2010). On the computational complexity of dominance links in grammatical formalisms. In: *ACL'10*, Association for Computer Linguistics, 514–524.
Sturmfels, B. (2002). *Solving Systems of Polynomial Equations*, CBMS Regional Conferences Series, vol. 97, American Mathematical Society.
Thatcher, J. W. and Wright, J. B. (1968). Generalized finite automata theory with an application to a decision problem of second-order logic. *Mathematical Systems Theory* **2** (1) 57–81.
Verma, K. N. and Goubault-Larrecq, J. (2005). Karp-Miller trees for a branching extension of VASS. *Discrete Mathematics & Theoretical Computer Science* **7** (1) 217–230.
Wies, T., Zufferey, D. and Henzinger, T. A. (2010). Forward analysis of depth-bounded processes. In: Ong, L. (ed.) *Proceedings of the 13th International Conference Foundations of Software Science and Computational Structures (FoSSaCS'10)*, LNCS, vol. 6014, Springer Verlag, 94–108.

Appendix A. Auxiliary Proofs on Irreducible Closed Sets

The results of this section are well known. We include their proofs for convenience.

Lemma A.1. *For every continuous map $f : Y \to Z$ between topological spaces Y and Z, for every irreducible closed subset C of Y, $cl(f[C])$ is irreducible closed in Z.*

Proof. This is a direct consequence of the fact that $\mathcal{S}$ is a functor, in particular that $\mathcal{S}(f)$ maps every $C \in \mathcal{S}(Y)$ to an element of $\mathcal{S}(Z)$, and of the fact that $\mathcal{S}(f)(C) = cl(f[C])$. Here is a direct argument. Assume that $cl(f([C]))$ is included in the union of two closed subsets F_1 and F_2 of Z. Then $f[C] \subseteq F_1 \cup F_2$, so $C \subseteq f^{-1}(F_1 \cup F_2) = f^{-1}(F_1) \cup f^{-1}(F_2)$. Since C is irreducible, C is included in $f^{-1}(F_1)$ or in $f^{-1}(F_2)$. Assume without loss of generality that $C \subseteq f^{-1}(F_1)$. Then $f[C]$ is included in F_1, and since $cl(f[C])$ is the smallest closed set containing $f[C]$, and F_1 is a closed set containing $f[C]$, $cl(f[C])$ is included in F_1. □

Lemma A.2. *Let $(X_i)_{i \in I}$ be any family of topological spaces. The irreducible closed subsets of $\prod_{i \in I} X_i$ are exactly the products $\prod_{i \in I} C_i$ where each C_i is irreducible closed in X_i.*

Proof. This is Proposition 8.4.7 of Goubault-Larrecq (2013) and leads to Hoffmann's theorem (Hoffmann 1979a, Theorem 1.4) that the sobrification functor preserves products. We give a direct proof. Let $X = \prod_{i \in I} X_i$. This has a base of open sets of the form $\bigcap_{i \in J} \pi_i^{-1}(U_i)$, where J is a finite subset of I and each U_i is open in X_i.

We first note that a closed set C is irreducible if and only if it is non-empty, and whenever it intersects two opens U_1 and U_2, then it also intersects their intersection $U_1 \cap U_2$. (Just take the contrapositive of the definition of irreducibility, with the complement of U_1 for F_1 and the complement of U_2 for F_2.)

If each C_i is closed in X_i, then $\prod_{i \in I} C_i$ is closed, being the intersection $\bigcap_{i \in I} \pi_i^{-1}(C_i)$. If additionally each C_i is irreducible, we claim that $C = \prod_{i \in I} C_i$ is also irreducible. It is certainly non-empty.

Assume that C intersects two open subsets W and W' of X. Write W as a union of basic open subsets $\bigcap_{j \in J_k} \pi_j^{-1}(U_{jk})$, $k \in K$, where each J_k is a finite subset of I, and each U_{jk} is open in X_j. Then C intersects $\bigcap_{j \in J_k} \pi_j^{-1}(U_{jk})$ for some $k \in K$, say at $(x_i)_{i \in I}$. Similarly, write W' as a union of basic open subsets $\bigcap_{j \in J'_{k'}} \pi_j^{-1}(U_{jk'})$, $k' \in K'$, where each $J'_{k'}$ is a finite subset of I, and each $U_{jk'}$ is open in X_j. Then, C intersects $\bigcap_{j \in J'_{k'}} \pi_j^{-1}(U_{jk'})$ for some $k' \in K'$, say at $(x'_i)_{i \in I}$. Without loss of generality, we may assume that $J_k = J'_{k'}$: otherwise replace J_k, resp. $J'_{k'}$, by $J_k \cup J'_{k'}$ and define the missing sets U_{jk}, resp. $U_{jk'}$, as X_j. For every $j \in J_k$, C_j intersects U_{jk} (at x_j) and also $U_{jk'}$ (at x'_j). Since C_j is irreducible, it therefore also intersects their intersection $U_{jk} \cap U_{jk'}$, say at y_j. For every $j \in I \setminus J_k$, define y_j as some arbitrary point from C_j. Then $(y_j)_{j \in I}$ is in C, and in $\bigcap_{j \in J_k} \pi_j^{-1}(U_{jk} \cap U_{jk'})$, hence in $W \cap W'$.

Conversely, we claim that every irreducible closed subset C of X must be a product $\prod_{i \in I} C_i$ of irreducible closed subsets C_i of X_i. We define C_i as $cl(\pi_i[C])$: by Lemma A.1, C_i is irreducible closed in X_i. Clearly, C is included in $\prod_{i \in I} C_i$. Assume for the sake of contradiction that the inclusion were strict: there is a point $\vec{x} = (x_i)_{i \in I}$ in $\prod_{i \in I} C_i$ and not in C. Then $\vec{x}$ is in the open complement W of C. By definition of the product topology, W contains a basic open subset $\bigcap_{j \in J} \pi_j^{-1}(U_j)$ containing $\vec{x}$, where J is a finite subset of I, and each U_j is open in X_j. Since it contains $\vec{x}$, x_j is in U_j for every $j \in J$. Since it is included in W, $\bigcap_{j \in J} \pi_j^{-1}(U_j)$ is disjoint from C. Since C is irreducible, if C intersected $\pi_j^{-1}(U_j)$ for every $j \in J$, it would also intersect $\bigcap_{j \in J} \pi_j^{-1}(U_j)$, which is impossible, as we have just seen. Hence, C is disjoint from $\pi_j^{-1}(U_j)$ for some $j \in J$. This implies that $\pi_j[C]$ is disjoint from U_j, hence that $C_j = cl(\pi_j[C])$ is also disjoint from U_j. (A set intersects an open set if and only if its closure does.) However, since $\vec{x} \in \prod_{i \in I} C_i$, x_j is in C_j, and x_j is also in U_j, contradiction. □

Appendix B. Proofs of Results on Words (Section 7)

Lemma 7.2 (recap). *Let X be a topological space. The complement of $X^* U_1 X^* U_2 X^* \cdots X^* U_n X^*$ ($n \in \mathbb{N}$, $U_1, U_2, \ldots, U_n$ open in X) in X^* is $\emptyset$ when $n = 0$, and $F_1^* X^? F_2^* X^? \cdots X^? F_{n-1}^* X^? F_n^*$ otherwise, where $F_1 = X \setminus U_1, \ldots, F_n = X \setminus U_n$.*

If X is Noetherian, then this complement can be expressed as a finite union of sets of the form $F_1^* C_1^? F_2^* C_2^? \cdots C_{n-1}^? F_n^*$, where $C_1, C_2, \ldots, C_{n-1}$ range over irreducible closed subsets of X.

Proof. When $n = 0$, this is clear: the complement of $X^* U_1 X^* U_2 X^* \cdots X^* U_n X^*$ is the empty set. So let $n \geq 1$.

We first claim that the complement of $X^* U_1 X^* U_2 X^* \cdots X^* U_n X^*$ is $F_1^* X^? F_2^* X^? \cdots X^? F_{n-1}^* X^? F_n^*$. We show this by induction on n. If $n = 1$, then the complement of $X^* U_1 X^*$ is the set of words that contain no letter from U_1, that is, F_1^*. If $n \geq 1$, let w be an arbitrary element of the complement of

$X^*U_1X^*U_2X^* \cdots X^*U_nX^*$. Let w_1 be the longest prefix of w comprised of letters not in U_1. Note that w_1 is in F_1^*. If $w_1 = w$, then certainly w is in $F_1^* \subseteq F_1^*X^?F_2^*X^? \cdots X^?F_{n-1}^*X^?F_n^*$. Otherwise, w is of the form w_1xw', where $x \in U_1$ and w' is not in $X^*U_2X^* \cdots X^*U_nX^*$. By induction hypothesis w' is in $F_2^*X^? \cdots X^?F_{n-1}^*X^?F_n^*$, hence again w is in $F_1^*X^?F_2^*X^? \cdots X^?F_{n-1}^*X^?F_n^*$.

Conversely, let w be any word in $F_1^*X^?F_2^*X^? \cdots X^?F_{n-1}^*X^?F_n^*$. Let w_1 be the longest prefix of w that lies in F_1^*. Then either $w = w_1$, then $w \in F_1^*$ cannot be in $X^*U_1X^*U_2X^* \cdots X^*U_nX^*$, since all the words in the latter set must contain at least one letter in U_1; or $w = w_1xw'$ for some $x \notin F_1$, that is, $x \in U_1$, and $w' \in F_2^*X^? \cdots X^?F_{n-1}^*X^?F_n^*$. By induction hypothesis, w' cannot be in $X^*U_2X^* \cdots X^*U_nX^*$. By construction, x would be the first occurrence of an element of U_1 in w. If $w = w_1xw'$ were in $X^*U_1X^*U_2X^* \cdots X^*U_nX^*$, then, some suffix w'' of w' would be in $X^*U_2X^* \cdots X^*U_nX^*$. Then $w'' \leq^* w'$, hence w' would be in $X^*U_2X^* \cdots X^*U_nX^*$, which is open hence upward-closed: contradiction.

By Lemma 4.6, X, as a closed subset of itself, is a finite union of irreducible closed subsets, that is, there is a finite subset E of $\mathcal{S}(X)$ such that $X = \bigcup_{C \in E} C$. Distributing across the $\_^?$ operator and concatenation in the expression $F_1^*X^?F_2^*X^? \cdots X^?F_{n-1}^*X^?F_n^*$ yields that the complement of $X^*U_1X^*U_2X^* \cdots X^*U_nX^*$ equals:

$$\bigcup_{C_1,\ldots,C_{n-1} \in E} F_1^*C_1^?F_2^*C_2^? \cdots C_{n-2}^?F_{n-1}^*C_{n-1}^?$$

from which the desired conclusion follows. □

Lemma 7.4 (recap). *Let X be a topological space. For every closed subset F of X, for every closed subset $\mathcal{F}$ of X^*, $F^?\mathcal{F}$ is closed in X^*.*

This is part of Exercise 9.7.29 of Goubault-Larrecq (2013), but the argument is non-trivial. Here is a complete proof. This requires the following construction, which we shall also require in the next Lemma. For any open $\mathcal{U}$ of X^*, and any open U of X, define $\mathcal{U}/U$ as follows. If $\mathcal{U} = X^*$, then $\mathcal{U}/U = \emptyset$; otherwise, $\mathcal{U}$ is a union of basic opens of the form $X^*U_{i1}X^*U_{i2}X^* \cdots X^*U_{in_i}X^*$, $i \in I$, where $n_i \geq 1$ for every $i \in I$, then we let $\mathcal{U}/U$ be the union of all basic opens $X^*(U_{i1} \cap U)X^*U_{i2}X^* \cdots X^*U_{in_i}X^*$. In all formality, $\mathcal{U}/U$ depends on a presentation of $\mathcal{U}$ as a union of basic opens, not on $\mathcal{U}$ itself, but this will cause no problem in the sequel. Note that this definition is made possible by the fact that $X^*U_1X^*U_2X^* \cdots X^*U_nX^*$ form a base, not just a subbase, of the topology (Lemma 7.1).

Proof. If $\mathcal{F}$ is empty, then $F^?\mathcal{F}$ is empty hence closed. Henceforth, we assume that $\mathcal{F}$ is non-empty, so that the complement $\mathcal{U}$ is different from X^*. Then, $\mathcal{U}/U$ is open. Also, $X^*U\mathcal{U}$ is open. We claim that the complement of $F^?\mathcal{F}$ in X^* is $X^*X\mathcal{U} \cup \mathcal{U}/U$, which will show the claim. We first make the following remark. Let L_1 and L_2 be two subsets of X^* that are downward-closed with respect to $\leq^*$. For any word w not in L_1L_2, we can write w as $w_1w'w_2$, where w_1 is the longest prefix of w in L_1, w_2 is the longest suffix of w in L_2, and w' is not empty. Indeed, any prefix of a word in L_1 is again in L_1, and any suffix of a word in L_2 is in L_2, since both are downward-closed with respect to $\leq^*$. That remark applies, notably, to $L_1 = F^?$ and $L_2 = \mathcal{F}$ fit, in the second case because $\leq^*$ is the specialization quasi-ordering of X^* and every closed subset is downward-closed.

If $\mathcal{F} = X^*$, then the complement of $F^?\mathcal{F} = X^*$ is empty, so the claim is proved. Otherwise, write $\mathcal{U}$ as the union of basic opens $X^*U_{i1}X^*U_{i2}X^* \cdots X^*U_{in_i}X^*$, $i \in I$, $n_i \geq 1$.

Assume w is in the complement of $F^?\mathcal{F}$ and write w as $w_1w'w_2$, as above. Since w' is not empty, it starts with some letter $x \in X$. Then by the maximality property of w_1, w_1 is in $F^?$, but w_1x is not. Again, by the maximality property of w_2, $w'w_2$ is not in $\mathcal{F}$, hence in $\mathcal{U}$. If $w_1 \neq \varepsilon$, then w_1 is a single letter, so w is in $X\mathcal{U} \subseteq X^*X\mathcal{U}$. If $w_1 = \varepsilon$, then x is not in F (otherwise w_1x would be in $F^?$), hence is in U. So $w'w_2$ starts with a letter in U, since $w'w_2$ is in $\mathcal{U}$, $w'w_2$ is in $X^*U_{i1}X^*U_{i2}X^* \cdots X^*U_{in_i}X^*$ for some $i \in I$. If $w'w_2$ is in $U_{i1}X^*U_{i2}X^* \cdots X^*U_{in_i}X^*$, then $w'w_2$ is in $(U \cap U_{i1})X^*U_{i2}X^* \cdots X^*U_{in_i}X^*$, so $w = w_1w'w_2$ is in $X^*(U \cap U_{i1})X^*U_{i2}X^* \cdots X^*U_{in_i}X^* \subseteq \mathcal{U}/U$;

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

otherwise, $w'w_2$ is in $UX^*U_{i1}X^*U_{i2}X^* \cdots X^*U_{in_i}X^*$, so $w = w_1w'w_2$ is in $X^*UX^*U_{i1}X^*U_{i2}X^* \cdots X^*U_{in_i}X^* \subseteq X^*U\mathcal{U} \subseteq X^*X\mathcal{U}$.

Conversely, assume w is in $X^*X\mathcal{U} \cup \mathcal{U}/U$. If $w \in X^*X\mathcal{U}$, then w contains a subsequence of the form $a_0a_1a_2 \cdots a_{n_i}$, for some $i \in I$, where a_0 is arbitrary, $a_1 \in U_{i1}$, $a_2 \in U_{i2}$, …, $a_{n_i} \in U_{in_i}$. Note that $a_1a_2 \cdots a_{n_i}$ is in $U_{i1}U_{i_2} \cdots U_{in_i} \subseteq X^*U_{i1}X^*U_{i2}X^* \cdots X^*U_{in_i}X^* \subseteq \mathcal{U}$. If w were in $F^?\mathcal{F}$, then since $F^?\mathcal{F}$ is downward-closed with respect to $\leq^*$, $a_0a_1a_2 \cdots a_{n_i}$ would be in $F^?\mathcal{F}$; hence, $a_1a_2 \cdots a_{n_i}$ would be in $\mathcal{F}$: contradiction. (More slowly, from $a_0a_1a_2 \cdots a_{n_i} \in F^?\mathcal{F}$, we deduce that $a_0 \in F$ and $a_1a_2 \cdots a_{n_i} \in \mathcal{F}$, or $a_0a_1a_2 \cdots a_{n_i} \in \mathcal{F}$, but the latter also implies $a_1a_2 \cdots a_{n_i} \in \mathcal{F}$, since $\mathcal{F}$ is downward-closed.) So w is in the complement of $F^?\mathcal{F}$. If, on the other hand, $w \in \mathcal{U}/U$, then w contains a subsequence of the form $a_1a_2 \cdots a_{n_i}$, for some $i \in I$, where $a_1 \in U \cap U_{i1}$, $a_2 \in U_{i2}, \ldots, a_{n_i} \in U_{in_i}$. In particular, $a_1a_2 \cdots a_{n_i}$ is in $U_{i1}U_{i_2} \cdots U_{in_i} \subseteq X^*U_{i1}X^*U_{i2}X^* \cdots X^*U_{in_i}X^* \subseteq \mathcal{U}$. If w were in $F^?\mathcal{F}$, then since $F^?\mathcal{F}$ is downward-closed with respect to $\leq^*$, $a_1a_2 \cdots a_{n_i}$ would be in $F^?\mathcal{F}$. However, a_1 is in U, so is not in F, and this implies that $a_1a_2 \cdots a_{n_i}$ would be in $\mathcal{F}$: contradiction. So, again, w is in the complement of $F^?\mathcal{F}$.□ □

Lemma 7.5 (recap). *Let X be a topological space. For every closed subset F of X, for every closed subset $\mathcal{F}$ of X^*, $F^*\mathcal{F}$ is closed in X^*.*

More specifically, in the case where $\mathcal{F}$ is non-empty, let U be the open complement of F in X and $\mathcal{U}$ be the open complement of $\mathcal{F}$ in X^*. Then, the complement of $F^*\mathcal{F}$ is $X^*U\mathcal{U} \cup \mathcal{U}/U$ and is therefore open.

Proof. Write $\mathcal{U}$ is a union of basic opens of the form $X^*U_{i1}X^*U_{i2}X^* \cdots X^*U_{in_i}X^*$, $i \in I$, as above.

Assume w is in the complement of $F^*\mathcal{F}$ and write w as $w_1w'w_2$, where w_1 is the longest prefix of w in F^*, w_2 is the longest suffix of w in $\mathcal{F}$, and w' is not empty. Let x be the first letter of w' and note that $w_1 \in F^*$ but w_1x is not in F^*: so x is in U; and that $w'w_2$ is in $\mathcal{U}$, so $w'w_2$ is in some basic open set $X^*U_{i1}X^*U_{i2}X^* \cdots X^*U_{in_i}X^*$, $i \in I$. Depending on whether the first letter x of $w'w_2$ is in U_{i1} or not, $w'w_2$ is in $(U \cap U_{i1})X^*U_{i2}X^* \cdots X^*U_{in_i}X^*$ or in $UX^*U_{i1}X^*U_{i2}X^* \cdots X^*U_{in_i}X^*$, so that w is in $X^*U\mathcal{U}$ or in $\mathcal{U}/U$.

Conversely, if $w \in X^*U\mathcal{U}$, then w contains a subword $a_0a_1a_2 \cdots a_{n_i}$ for some $i \in I$, $a_0 \in U$, $a_1 \in U_{i1}$, $a_2 \in U_{i2}$, …, $a_{n_i} \in U_{in_i}$. If w were in $F^*\mathcal{F}$, then $a_0a_1a_2 \cdots a_{n_i}$ would be, too: it is indeed clear that $F^*\mathcal{F}$ is downward-closed, since F^* is and $\mathcal{F}$ is closed. Since $a_0 \in U$, a_0 is not in F, so $a_0a_1a_2 \cdots a_{n_i}$ is in $\mathcal{F}$. Again by downward closure, $a_1a_2 \cdots a_{n_i}$ is in $\mathcal{F}$: contradiction. So w is in the complement of $F^*\mathcal{F}$. And if $w \in \mathcal{U}/U$, then w contains a subword of the form $a_1a_2 \cdots a_{n_i}$, for some $i \in I$, where $a_1 \in U \cap U_{i1}$, $a_2 \in U_{i2}, \ldots, a_{n_i} \in U_{in_i}$. In particular, $a_1a_2 \cdots a_{n_i}$ is in $U_{i1}U_{i_2} \cdots U_{in_i} \subseteq X^*U_{i1}X^*U_{i2}X^* \cdots X^*U_{in_i}X^* \subseteq \mathcal{U}$. If w were in $F^*\mathcal{F}$, then so would be this subword, and as $a_1 \in U$ is not in F, $a_1a_2 \cdots a_{n_i}$ would be in $\mathcal{F}$: contradiction. So w is in the complement of $F^*\mathcal{F}$. □

Lemma B.1. *The concatenation function cat: $X^* \times X^* \to X^*$ is continuous. The function i: $X \to X^*$ that maps the letter x to x as a word is also continuous.*

Proof. This is Exercise 9.7.27 of Goubault-Larrecq (2013). To show that i is continuous, observe that $i^{-1}(X^*U_1X^*U_2X^* \cdots X^*U_nX^*) = U_1$ if $n = 1$, X if $n = 0$, and $\emptyset$ otherwise. As far as *cat* is concerned, $cat^{-1}(X^*U_1X^*U_2X^* \cdots X^*U_nX^*) = \bigcup_{i=0}^{n} (X^*U_1X^*U_2X^* \cdots X^*U_iX^*) \times (X^*U_{i+1}X^*U_2X^* \cdots X^*U_nX^*)$. □

Lemma 7.7 (recap). *Let X be a topological space. Every word-product is irreducible closed in X^*.*

Proof. By induction on syntax, starting with the fact that the base case ϵ denotes a one-element downward-closed set, hence, is trivially irreducible closed. It suffices to show that sets of the form F^* and $C^?$ are irreducible closed, where F is closed and C is irreducible closed, and that if $\mathcal{C}_1$ and $\mathcal{C}_2$ are irreducible closed and $\mathcal{C}_1\mathcal{C}_2$ is closed, then $\mathcal{C}_1\mathcal{C}_2$ is irreducible closed.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

1. Let us show that F^* is irreducible closed in X^*, for any closed subset F of X. Assume $F^* \subseteq \mathcal{F}_1 \cup \mathcal{F}_2$, where $\mathcal{F}_1$ and $\mathcal{F}_2$ are closed in X^*. If F^* was not contained in $\mathcal{F}_1$ or in $\mathcal{F}_2$, then there would be a word $w_1 \in F^* \smallsetminus \mathcal{F}_1$ and a word $w_2 \in F^* \smallsetminus \mathcal{F}_2$. Then, $w_1 w_2$ would again be in F^*, hence either in $\mathcal{F}_1$ or in $\mathcal{F}_2$. Assume by symmetry that $w_1 w_2$ is in $\mathcal{F}_1$. Since $w_1 \leq^* w_1 w_2$, and closed sets such as $\mathcal{F}_1$ are downward-closed, we would have $w_1 \in \mathcal{F}_1$: contradiction. So F^* is irreducible.
2. We claim that $C^?$ is irreducible closed in X^* whenever C is irreducible closed in X. It is enough to observe that $C^? = cl(i[C])$ – recall that i is continuous by Lemma B.1 – and to use Lemma A.1. The inclusion $cl(i[C]) \subseteq C^?$ stems for the fact that $C^?$ is closed, and $i[C] \subseteq C^?$, which is clear. The converse inclusion is obvious.
3. Finally, we show that whenever $\mathcal{C}_1$ and $\mathcal{C}_2$ are irreducible closed in X^*, and $\mathcal{C}_1\mathcal{C}_2$ is closed, it is irreducible. Indeed, $\mathcal{C}_1 \times \mathcal{C}_2$ is irreducible closed by Lemma A.2, and since cat is continuous by Lemma B.1, $cl(cat[\mathcal{C}_1 \times \mathcal{C}_2])$ is irreducible by Lemma A.1. Then, $cat[\mathcal{C}_1 \times \mathcal{C}_2] = \mathcal{C}_1\mathcal{C}_2$, and since the latter is closed by assumption, it is equal to $cl(cat[\mathcal{C}_1 \times \mathcal{C}_2])$, hence irreducible closed. □

Lemma 7.9 (recap). *Let X be a topological space. Inclusion between word-products can be checked in polynomial time (precisely in time proportional to the product of the lengths of the two word-products), modulo an oracle testing inclusion of closed subsets of X.*

Explicitly, we have $\epsilon \subseteq P$ for any word-product P, $P \not\subseteq \epsilon$ unless all the atomic expressions in P are syntactically equal to $\emptyset^*$, and for all $C, C' \in \mathcal{S}(X)$, for all $F, F' \in \mathcal{H}_\mathsf{V}(X)$, and for all word-products P, P':

- *$C^? P \subseteq C'^? P'$ if and only if $C \subseteq C'$ and $P \subseteq P'$, or $C \not\subseteq C'$ and $C^? P \subseteq P'$.*
- *$C^? P \subseteq F'^* P'$ if and only if $C \subseteq F'$ and $P \subseteq F'^* P'$, or $C \not\subseteq F'$ and $C^? P \subseteq P'$.*
- *$F^* P \subseteq C'^? P'$ if and only if F is empty and $P \subseteq C'^? P'$, or F is non-empty and $F^* P \subseteq P'$.*
- *$F^* P \subseteq F'^* P'$ if and only if $F \subseteq F'$ and $P \subseteq F'^* P'$, or $F \not\subseteq F'$ and $F^* P \subseteq P'$.*

Proof. The cases $\epsilon \subseteq P$ and $P \not\subseteq \epsilon$ are obvious.

We first examine when $C^? P \subseteq C'^? P'$ holds. The if direction is easy. Conversely, assume $C^? P \subseteq C'^? P'$. For every $x \in C$, either $x \in C'$ or xw is in P' for every $w \in P$: indeed, when $w \in P$, then $xw \in C^? P \subseteq C'^? P'$, and if $x \notin C'$ this can only happen if $xw \in P'$. This means that C is contained in $C' \cup F$, where $F = \{x \in X \mid \forall w \in P \cdot xw \in P'\} = \bigcap_{w \in P} f_w^{-1}(P')$, and f_w is the map $x \mapsto xw$. Note that $f_w(x) = cat(i(x), w)$, hence f_w is continuous. Using the fact that P' is closed (Corollary 7.6), F is closed. Since $C \subseteq C' \cup F$ and C is irreducible, we have proved $(*)$ either $C \subseteq C'$ or $C \subseteq F$. We also note that $(**)$ $P \subseteq P'$, in any case, fixing some element $x \in C$ (recall that C is non-empty), for every $w \in P$, xw is in $C^? P$, hence in $C'^? P'$, so w or xw is in P'; since P' is closed and $w \leq^* xw$, w must be in P' in any case. Using $(*)$ and $(**)$, we now have two cases: either $C \subseteq C'$ and $P \subseteq P'$, or $C \not\subseteq C'$, $C \subseteq F$, and $P \subseteq P'$. In the latter case, $C \subseteq F$ entails $CP \subseteq P'$ by definition of F, so $C^? P = P \cup CP \subseteq P'$.

We now examine when $C^? P \subseteq F'^* P'$ holds. This is similar. The if direction is obvious. Conversely, if $C^? P \subseteq F'^* P'$, then for every $x \in C$, either $x \in F'$ or $xw \in P'$ for every $w \in P$. So $C \subseteq F' \cup F$, where F is the closed set $\{x \in X \mid \forall w \in P \cdot xw \in P'\}$. Since C is irreducible, $(*)$ either $C \subseteq F'$ or $C \subseteq F$. Also, $(**)$ $P \subseteq F'^* P'$, because $P \subseteq C^? P \subseteq F'^* P'$. If $C \not\subseteq F'$, then $C \subseteq F$, hence $CP \subseteq P'$ by the definition of F; since P' is downward-closed, P is also included in P', so $C^? P = P \cup CP \subseteq P'$.

Let us proceed to the case $F^* P \subseteq C'^? P'$. When F is empty, since $F^* P = P$, the equivalence between $F^* P \subseteq C'^? P'$ and $P \subseteq C'^? P'$ is obvious. Otherwise, since F is non-empty, let x be some element in F. For any $w \in F^* P$, xw is also in $F^* P$, so is in $C'^? P'$. This implies that xw or w is in P'. But, as P' is downward-closed, $w \in P'$ in any case. So $F^* P \subseteq P'$. The converse is again easy.

Finally, assume $F^* P \subseteq F'^* P'$. If $F \subseteq F'$, then $P \subseteq F'^* P'$, since $P \subseteq F^* P$. Otherwise, let x be in F but not in F'. For any word $w \in F^* P$, xw is again in $F^* P$, hence in $F'^* P'$. Since $x \notin F'$, xw must be in P', hence also $w \in P'$. So $F^* P \subseteq P'$.

We obtain the desired algorithm (up to an oracle) by dynamic programming. □

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Lemma 7.12 (recap). *Let X be a topological space. Any finite intersection of word-products is expressible as a finite union of word-products. Specifically, the intersection of two word-products is given by:* $\epsilon \cap P = \epsilon$ *for every word-product P, and by the recursive formulae:*

— $C^? P \cap C'^? P' = (C^? P \cap P') \cup (P \cap C'^? P') \cup (C \cap C')^? (P \cap P')$;
— $C^? P \cap F'^* P' = (C \cap F')^? (P \cap F'^* P') \cup (C^? P \cap P')$;
— $F^* P \cap F'^* P' = (F \cap F')^* (P \cap F'^* P') \cup (F \cap F')^* (F^* P \cap P')$.

Proof. Let us deal with the first case. Any word w in $C^? P \cap C'^? P'$ is either in $P \cap P'$, or is in CP and in P', or in P and in $C'P'$, or is of the form xw', with $x \in C \cap C'$ and $w' \in P \cap P'$. So $C^? P \cap C'^? P' \subseteq (P \cap P') \cup (C^? P \cap P') \cup (P \cap C'^? P') \cup (C \cap C')^? (P \cap P') = (C^? P \cap P') \cup (P \cap C'^? P') \cup (C \cap C')^? (P \cap P')$. It is easy to see that conversely, $(C^? P \cap P') \cup (P \cap C'^? P') \cup (C \cap C')^? (P \cap P')$ is included in $C^? P \cap C'^? P'$.

Next, any word w in $C^? P \cap F'^* P'$ is either in $P \cap F'^* P'$, or is of the form xw' with $x \in C$, $w' \in P$, and $xw' \in F'^* P'$. In the latter case, either $x \in C \cap F'$ and $w' \in P \cap F'^* P'$, so $w \in (C \cap F')^? (P \cap F'^* P')$ or $x \in C$, x is not in F' so $w = xw'$ is in P', hence w is in $C^? P \cap P'$. In any case, $C^? P \cap F'^* P' \subseteq (P \cap F'^* P') \cup (C \cap F')(P \cap F'^* P') \cup (C^? P \cap P') = (C \cap F')^? (P \cap F'^* P') \cup (C^? P \cap P')$. The converse inclusion is clear.

Finally, for every word w in $F^* P \cap F'^* P'$, write w as $w_1 w_2$ where w_1 is the longest prefix of w in F^*, and $w_2 \in P$; also, as $w'_1 w'_2$ where w'_1 is the longest prefix of w in F'^*, and $w'_2 \in P'$. If w_1 is shorter than w'_1, then w_2 is also in $F'^* P'$, so $w \in (F \cap F')^* (P \cap F'^* P')$, otherwise $w \in (F \cap F')^* (F^* P \cap P')$. So $F^* P \cap F'^* P' \subseteq (F \cap F')^* (P \cap F'^* P') \cup (F \cap F')^* (F^* P \cap P')$. The converse inclusion is obvious.

These formulae define the intersection of two word-products, by induction on the sum of the number of atomic formulae in each of them. So the intersection of two word-products is a finite union of word-products. The empty intersection, the space X^* itself, is clearly a word-product. By induction on the number of word-products, any finite intersection of word-products can then be rewritten as a finite union of word-products. □

Proposition B.2. *Let X be a Noetherian space,* $X' = X^*$, *and* $(S, [\![\_]\!], \trianglelefteq, \tau, \wedge)$ *be an S-representation of X. Then* $(S', [\![\_]\!]', \trianglelefteq', \tau', \wedge')$ *is an S-representation of* X', *where:*

— S' *is the collection of all* word-product notations, *that is, of all expressions of the form* $e_1 e_2 \cdots e_n$ *where each* e_i *is either of the form* u^* *where u is a finite subset of S, or of the form* $a^?$ *where* $a \in S$. *We write* ϵ *for the empty word-product notation.*
— $[\![e_1 e_2 \cdots e_n]\!]' = [\![e_1]\!]' [\![e_2]\!]' \cdots [\![e_n]\!]' = \{x_1 x_2 \cdots x_n \mid x_1 \in [\![e_1]\!]', x_2 \in [\![e_2]\!]', \ldots, x_n \in [\![e_n]\!]'\}$, *where* $[\![u^*]\!]'$ *is the set of all finite words whose letters are in* $\bigcup_{a \in u} [\![a]\!]$, *and* $[\![a^?]\!]'$ *is the set of words containing at most one letter, and this letter is in* $[\![a]\!]$;
— $\trianglelefteq'$ *is the relation* star$[\trianglelefteq]$, *defined recursively from* $\trianglelefteq$ *by:* $\epsilon \trianglelefteq' w$ *for every word-product notation w,* $w \trianglelefteq' \epsilon$ *iff* $w = \epsilon$, *and for all* $a, a' \in S$, *for all non-empty finite subsets u, u' of S:*
  - $a^? w \trianglelefteq' a'^? w'$ *iff* $a \trianglelefteq a'$ *and* $w \trianglelefteq' w'$, *or* $a \ntrianglelefteq a'$ *and* $a^? w \trianglelefteq' w'$;
  - $a^? w \trianglelefteq' u'^* w'$ *iff* $a \trianglelefteq a'$ *for some* $a' \in u'$ *and* $w \trianglelefteq' u'^* w'$, *or* $a \trianglelefteq a'$ *for no* $a' \in u'$ *and* $a^? w \trianglelefteq' w'$;
  - $u^* w \trianglelefteq' a'^? w'$ *iff u is empty and* $w \trianglelefteq' a'^? w'$, *or u is non-empty and* $u^* w \trianglelefteq' w'$;
  - $u^* w \trianglelefteq' u'^* w'$ *iff either for every* $a \in u$, *there is an* $a' \in u'$ *such that* $a \trianglelefteq a'$ *and* $w \trianglelefteq' u'^* w'$, *or there is an* $a \in u$ *such that* $a \trianglelefteq a'$ *for no* $a' \in u'$, *and* $u^* w \trianglelefteq' w'$.
— τ' *is* $\{\tau^*\}$.
— $\wedge'$ *is the map* meet$[\wedge]$, *parametrized by* $\trianglelefteq$, *and defined recursively by:* $\epsilon \wedge' w' = \{\epsilon\}$, $w \wedge' \epsilon = \{\epsilon\}$, *and for all* $a, a' \in S$, *for all non-empty finite subsets u, u' of S:*
  - $a^? w \wedge' a'^? w' = \{a''^? w'' \mid a'' \in a \wedge a', w'' \in w \wedge' w'\} \cup (a^? w \wedge' w') \cup (w \wedge' a'^? w')$;

– $a^? w \wedge' u'^* w' = \{a''^? w'' \mid a' \in u', a'' \in a \wedge a', w'' \in w \wedge' u'^* w'\} \cup (a^? w \wedge' w')$ *if there is an* $a' \in u'$ *such that* $a \wedge a' \neq \emptyset$, $a^? w \wedge' u'^* w' = (w \wedge' u'^* w') \cup (a^? w \wedge' w')$ *otherwise; and similarly for* $u^* w \wedge' a'^? w'$;
– $u^* w \wedge' u'^* w' = \left\{ \left(\bigcup_{\substack{a \in u \\ a' \in u'}} a \wedge a' \right)^* w'' \mid w'' \in (u^* w \wedge' w') \cup (w \wedge' u'^* w') \right\}$.

Proof. $[\![\_]\!]'$ is surjective: the irreducible closed subsets of X' are the word-products by Proposition 7.14, that is, of the form $[\![w]\!]$ for some word-product notation $w \in X'$.

The formulae defining $\mathsf{star}[\trianglelefteq]$ are obtained from Lemma 7.9. The formula defining τ' is justified by the fact that $\bigcup_{w \in \tau'} [\![w]\!]' = [\![\tau^*]\!]'$ is the set of all words whose letters are all in $\bigcup_{a \in \tau} [\![a]\!] = X$, this is, the whole set X^*. The formulae defining $\wedge'$ are obtained from Lemma 7.13. □

Appendix C. Proofs of Results on Multisets (Section 8)

Proposition 8.4 (recap). *Let X be a topological space. Then, the m-SREs are closed in* $X^{\circledast}$, *and the m-products are irreducible closed.*

If X is Noetherian, then every irreducible closed subset of $X^{\circledast}$ *is an m-product, and every closed subset of* $X^{\circledast}$ *is an m-SRE.*

Proof. Consider any m-product $P = F \mid C_1, C_2, \ldots, C_n$. We observe that $\Psi^{-1}(P)$ is the union over all permutations π of $\{1, 2, \ldots, n\}$ of the word-products $F^* C^?_{\pi(1)} F^* C^?_{\pi(2)} F^* \cdots F^* C^?_{\pi(n)} F^*$. This means that the words whose multiset of letters can be split as at most one letter from each of C_1, $C_2, \ldots, C_n$, plus remaining letters from F, are just the words that are comprised of letters from F, except for zero or one letter from C_i, $i \in \{1, 2, \ldots, n\}$, sprinkled here and there in some order. So $\Psi^{-1}(P)$ is closed in X^*. Because Ψ is quotient, P is closed in $X^{\circledast}$.

It also follows that any m-SRE is closed in $X^{\circledast}$.

Next we show that the m-products $F \mid C_1, C_2, \ldots, C_n$ are indeed irreducible. By Lemma 7.7, $F^* C^?_1 C^?_2 \cdots C^?_n$ is irreducible closed in X^*. So the closure of its image by Ψ is irreducible closed by Lemma A.1. However, the image of $F^* C^?_1 C^?_2 \cdots C^?_n$ by Ψ is $F \mid C_1, C_2, \ldots, C_n$. We have seen that it is closed, hence equal to its closure. Therefore, it is irreducible closed.

Now assume X Noetherian. Let $\mathcal{F}$ be any closed subset of $X^{\circledast}$. Since Ψ is continuous, $\Psi^{-1}(\mathcal{F})$ is closed in X^*, hence a finite union of word-products, by Proposition 7.14. Since Ψ is surjective, $\mathcal{F}$ is equal to the image $\Psi[\Psi^{-1}(\mathcal{F})]$ of $\Psi^{-1}(\mathcal{F})$ by Ψ. So $\mathcal{F}$ is a finite union of subsets $\Psi[P_i]$, $i \in I$, where each P_i is a word-product. We calculate $\Psi[P_i]$ as follows. Write P_i as $e_1 e_2 \cdots e_n$, and since it will not change its image by Ψ, reorder the atomic expressions e_i in $e_1 e_2 \cdots e_n$ so that the starred ones come first. Doing so allows us to write our word-product as $F^*_1 F^*_2 \cdots F^*_m C^?_1 C^?_2 \cdots C^?_p$, up to permutation of factors. Its image by Ψ is the set of multisets obtained by picking at most one element from C_1, at most one from C_2, ..., at most one C_p, then arbitrarily many from F_1, arbitrarily many from $F_2, \ldots,$ arbitrarily many from F_m. It follows that $\Psi[e_1 e_2 \cdots e_n] = (F_1 \cup F_2 \cup \cdots \cup F_m) \mid C_1, C_2, \ldots, C_p$. In particular, $\Psi[P_i]$ is an m-product. Therefore, $\mathcal{F}$ is a finite union of m-products, hence an m-SRE.

If $\mathcal{F}$ is also irreducible, then this finite union must be the union of a single m-product, hence is an m-product. □

Lemma 8.6 (recap). *Let X be a topological space. Inclusion between m-products can be checked in polynomial time, modulo an oracle testing inclusion of closed subsets of X.*

Explicitly, let $P = F \mid C_1, C_2, \ldots, C_m$ *and* $P' = F' \mid C'_1, C'_2, \ldots, C'_n$ *be two m-products. Let* $I = \{i_1, i_2, \ldots, i_k\}$ *be the subset of those indices i,* $1 \leq i \leq m$, *such that* $C_i \not\subseteq F'$.

Then, $P \subseteq P'$ *if and only if* $F \subseteq F'$ *and there is an injective map* $r \colon I \to \{1, 2, \ldots, n\}$ *such that* $C_i \subseteq C'_{r(i)}$ *for every* $i \in I$ *– in other words,* $\{\!|C_{i_1}, C_{i_2}, \ldots, C_{i_k}|\!\} \subseteq^{\circledast} \{\!|C'_1, C'_2, \ldots, C'_n|\!\}$.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Proof. Assume $P \subseteq P'$. If $F \not\subseteq F'$, then pick $x \in F \setminus F'$: the multiset consisting of $n+1$ copies of x is in P but not in P': contradiction. So $F \subseteq F'$.

Let now $I = \{i_1, i_2, \ldots, i_k\}$ be as above. Let $D_1 = C_{i_1}$, $D_2 = C_{i_2}$, ..., $D_k = C_{i_k}$. Let also $E_1, E_2, \ldots, E_{m-k}$ be an enumeration of those C_i, $1 \leq i \leq n$, with $i \notin I$. Consider the word-product P_1 defined as $E_1^? E_2^? \cdots E_{m-k}^? F^* D_1^? D_2^? \cdots D_k^?$. Note that $P_1 \subseteq \Psi^{-1}(P)$, so $P_1 \subseteq \Psi^{-1}(P')$. On the other hand, $\Psi^{-1}(P')$ is the union over all permutations π of $\{1, 2, \ldots, n\}$ of $F'^* {C'_{\pi(1)}}^? F'^* {C'_{\pi(2)}}^? F'^* \cdots F'^* {C'_{\pi(n)}}^? F'^*$. Since P_1 is irreducible (Lemma 7.7), there a permutation π of $\{1, 2, \ldots, n\}$ such that $P_1 \subseteq F'^* {C'_{\pi(1)}}^? F'^* {C'_{\pi(2)}}^? F'^* \cdots F'^* {C'_{\pi(n)}}^? F'^*$. Using Lemma 7.9, and the fact that $E_1, E_2, \ldots, E_{m-k}$ are contained in F', and $F \subseteq F'$, and recalling the definition of P_1, we obtain that $D_1^? D_2^? \cdots D_k^?$ is included in $F'^* {C'_{\pi(1)}}^? F'^* {C'_{\pi(2)}}^? F'^* \cdots F'^* {C'_{\pi(n)}}^? F'^*$.

We show that there is an injective map $r\colon I \to \{\pi(1), \pi(2), \ldots, \pi(n)\}$ such that $C_i \subseteq C'_{r(i)}$ for every $i \in I$, by induction on $k+n$. If $k=0$, the empty map fits. Otherwise, since $D_1 \not\subseteq F'$, using Lemma 7.9, we must have $D_1^? D_2^? \cdots D_k^? \subseteq {C'_{\pi(1)}}^? F'^* {C'_{\pi(2)}}^? F'^* \cdots F'^* {C'_{\pi(n)}}^? F'^*$. Now we have two cases, again following Lemma 7.9. In the first case $D_1 = C_{i_1} \subseteq C'_{\pi(1)}$ and $D_2^? \cdots D_k^? \subseteq F'^* {C'_{\pi(2)}}^? F'^* \cdots F'^* {C'_{\pi(n)}}^? F'^*$, so there is an injective map $r'\colon \{i_2, \ldots, i_k\} \to \{\pi(2), \ldots, \pi(n)\}$ such that $C_i \subseteq C'_{r'(i)}$ for every $i \in \{i_2, \ldots, i_k\}$, by induction hypothesis. Then taking $r(i_1) = \pi(1)$ and $r(i) = r'(i)$ for every $i \in \{i_2, \ldots, i_k\}$ fits. In the second case, $D_1^? D_2^? \cdots D_k^? \subseteq F'^* {C'_{\pi(2)}}^? F'^* \cdots F'^* {C'_{\pi(n)}}^? F'^*$, and we conclude directly by the induction hypothesis.

Conversely, if there is an injective map $r\colon I \to \{1, 2, \ldots, n\}$ such that $C_i \subseteq C'_{r(i)}$ for every $i \in I$, it is clear that $P \subseteq P'$. □

Theorem 8.7 (recap). *Let X be a Noetherian space, $X' = X^{\circledast}$, and $(S, [\![\_]\!], \trianglelefteq, \tau, \wedge)$ be an S-representation of X. Then, $(S', [\![\_]\!]', \trianglelefteq', \tau', \wedge')$ is an S-representation of X', where:*

- *(A) S' is the collection of all* m-product notations, *that is, of all expressions of the form $A \mid u$, where A is a finite subset of S, and u is a multiset of elements of S. When $u = \{\!|b_1, \ldots, b_n|\!\}$, we also write $A \mid b_1, \ldots, b_n$ for $A \mid u$.*
- *(B) $[\![A \mid b_1, \ldots, b_n]\!]' = (\bigcup_{a \in A} [\![a]\!]) \mid [\![b_1]\!], \ldots, [\![b_n]\!]$.*
- *(C) $A \mid u \trianglelefteq' A' \mid u'$ if and only if $A \trianglelefteq^{\flat} A'$ and $u_1 \trianglelefteq^{\circledast} u'$ where u_1 is the subset of those elements $a \in u$ such that $a \trianglelefteq a'$ for no $a' \in A'$.*
- *(D) τ' is $\{\tau \mid \emptyset\}$.*
- *(E) $\wedge'$ is defined as follows. A* matching *$f\colon \{1, \ldots, m\} \to \{1, \ldots, n\}$ is any bijection from some subset of $\{1, \ldots, m\}$ (the* domain $\operatorname{dom} f$*) to some subset of $\{1, \ldots, n\}$ (the* codomain $\operatorname{cod} f$*). Then, $(A \mid a_1, \ldots, a_m) \wedge' (A' \mid a'_1, \ldots, a'_n)$ is the collection of all m-product notations of the form $A'' \mid m_{1f} \uplus m_{2f} \uplus m_{3f}$, where:*
  - *— $A'' = \bigcup_{\substack{a \in A \\ a' \in A'}} (a \wedge a')$;*
  - *— f ranges over all matchings from $\{1, \ldots, m\}$ to $\{1, \ldots, n\}$;*
  - *— m_{1f} ranges over all multisets of the form $\{\!|c_i \mid i \in \operatorname{dom} f|\!\}$ where $c_i \in a_i \wedge a'_{f(i)}$ for every $i \in \operatorname{dom} f$;*
  - *— m_{2f} ranges over all multisets of the form $\{\!|c_i \mid 1 \leq i \leq m, i \notin \operatorname{dom} f|\!\}$, where $c_i \in \bigcup_{a' \in A'} (a_i \wedge a')$ for each i, $1 \leq i \leq m$, $i \notin \operatorname{dom} f$;*
  - *— m_{3f} ranges over all multisets of the form $\{\!|c'_j \mid 1 \leq j \leq n, j \notin \operatorname{cod} f|\!\}$, where $c'_j \in \bigcup_{a \in A} (a \wedge a'_j)$ for each j, $1 \leq j \leq n$, $j \notin \operatorname{cod} f$.*

Proof. First, $[\![\_]\!]'$ is surjective by Proposition 8.4. The formula for $\trianglelefteq'$ is justified by Lemma 8.6. The fact that $X^{\circledast} = \bigcup_{A \mid u \in \tau'} [\![A \mid u]\!]'$ is clear: the union on the right-hand side is $[\![\tau \mid \emptyset]\!]'$, which is by definition the set of multisets whose elements are all in $\bigcup_{a \in \tau} [\![a]\!] = X$.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

To justify the formula for $\wedge'$, finally, compute the intersection of $[\![A \mid a_1, \ldots, a_m]\!]' = F \mid [\![a_1]\!], \ldots, [\![a_m]\!]$ (where $F = \bigcup_{a \in A} [\![a]\!]$) and of $[\![A' \mid a'_1, \ldots, a'_n]\!]' = F' \mid [\![a'_1]\!], \ldots, [\![a'_n]\!]$ (where $F' = \bigcup_{a' \in A'} [\![a']\!]$). Any multiset m in the intersection can be split into four parts: first, the multiset m_0 of those elements that are in $F \cap F'$; then, among the remaining elements, the multiset m_1 of those elements that are both in some $[\![a_i]\!]$ and in some $[\![a'_j]\!]$: reasoning on indices, there must be a matching $f : \{1, \ldots, m\} \to \{1, \ldots, n\}$ such that m_1 is a multiset of elements x_i, $i \in \operatorname{dom} f$, where x_i is in $[\![a_i]\!] \cap [\![a'_{f(i)}]\!]$; then, the remaining elements are obtained by taking at most one element from each $[\![a_i]\!]$, $i \notin \operatorname{dom} f$, provided they are in F', and at most one element from each $[\![a'_j]\!]$, $j \notin \operatorname{cod} f$, provided they are in F. Let E_f be the set $(F \cap F') \mid \underbrace{[\![a_i]\!] \cap [\![a'_{f(i)}]\!]}_{i \in \operatorname{dom} f}, \underbrace{[\![a_i]\!] \cap F'}_{\substack{1 \le i \le m \\ i \notin \operatorname{dom} f}}, \underbrace{F \cap [\![a'_j]\!]}_{\substack{1 \le j \le n \\ j \notin \operatorname{cod} f}}$.

We have just shown that $[\![A \mid a_1, \ldots, a_m]\!]' \cap [\![A' \mid a'_1, \ldots, a'_n]\!]'$ was contained in the union of all E_f, when f ranges over the matchings from $\{1, \ldots, m\}$ to $\{1, \ldots, n\}$. The converse inclusion is obvious.

We then observe that:

— $F \cap F' = \bigcup_{\substack{a \in A \\ a' \in A'}} ([\![a]\!] \cap [\![a']\!]) = \bigcup_{\substack{a \in A \\ a' \in A' \\ c \in a \wedge a'}} [\![c]\!] = \bigcup_{c \in A''} [\![c]\!]$;

— for each $i \in \operatorname{dom} f$, $[\![a_i]\!] \cap [\![a'_{f(i)}]\!] = \bigcup_{c_i \in a_i \wedge a'_{f(i)}} [\![c_i]\!]$;

— for each i, $1 \le i \le m$, $i \notin \operatorname{dom} f$, $[\![a_i]\!] \cap F' = [\![a_i]\!] \cap \bigcup_{a' \in A'} [\![a']\!] = \bigcup_{a' \in A'} ([\![a_i]\!] \cap [\![a']\!]) = \bigcup_{\substack{a' \in A' \\ c_i \in a_i \wedge a'}} [\![c_i]\!]$;

— and similarly, for each j, $1 \le j \le n$, $j \notin \operatorname{cod} f$, $F \cap [\![a'_j]\!] = \bigcup_{\substack{a \in A \\ c'_j \in a \wedge a'_j}} [\![c'_j]\!]$.

Finally, we notice that unions distribute over the $\mid$ construction, meaning that $F \mid A \cup B, C_2, \ldots, C_n$ is equal to the union of $F \mid A, C_2, \ldots, C_n$ and $F \mid B, C_2, \ldots, C_n$. By distributing all unions across the $\mid$ construction, we obtain the indicated formula for $\wedge'$. □

Appendix D. Proofs of Results on Words with the Prefix Topology (Section 9)

Proposition 9.2 (recap). *Let $X_1, X_2, \ldots, X_n, \ldots$ be countably many topological spaces. The sets of the form $\lceil F_1 F_2 \cdots F_n \rangle$, where each F_i is closed in X_i, form a subbase of closed sets for $\rhd_{n=1}^{+\infty} X_n$: these sets are closed, and every closed subset is an intersection of finite unions of such sets.*

Proof. We first observe that the complement of $\lceil F_1 F_2 \cdots F_n \rangle$ is the open set $\lfloor \emptyset, X_1 \setminus F_1, X_1 X_2 \setminus F_1 F_2, \ldots, X_1 X_2 \cdots X_m \setminus F_1 F_2 \cdots F_m, \cdots, X_1 X_2 \cdots X_n \setminus F_1 F_2 \cdots F_n, X_1 X_2 \cdots X_n X_{n+1}, \cdots, X_1 X_2 \cdots X_k, \cdots \rangle = \lfloor \emptyset, U_1, U_1 X_2 \cup X_1 U_2, \cdots, \bigcup_{i=1}^m X_1 \cdots X_{i-1} U_i X_{i+1} \cdots X_m, \cdots, \bigcup_{i=1}^n X_1 \cdots X_{i-1} U_i X_{i+1} \cdots X_n, X_1 X_2 \cdots X_n X_{n+1}, \cdots, X_1 X_2 \cdots X_k, \cdots \rangle$, where U_i is the complement of F_i, $1 \le i \le n$.

Conversely, we claim that every open in the prefix topology is a union of subsets of the form ${}_m\lfloor V_1 V_2 \cdots V_m \rangle_n$ ($0 \le m \le n$, V_i open in X_i for every i), where the latter denotes the set of all words w of length at least m such that either $|w| \ge n$ or the m-letter prefix of w is in $V_1 V_2 \cdots V_m$. Indeed, consider any wide telescope $\mathcal{U} = U_0, U_1, \ldots, U_n, \ldots$, where $U_k = \prod_{i=1}^k X_i$ for all $k \ge n$. $\mathcal{U}$ certainly contains ${}_m\lfloor V_1 V_2 \cdots V_m \rangle_n$ for any open rectangle $V_1 V_2 \cdots V_m$ contained in U_m: for any $w \in {}_m\lfloor V_1 V_2 \cdots V_m \rangle_n$, either $|w| \ge n$, then $w \in \prod_{i=1}^{|w|} X_i = U_{|w|}$, or $m \le |w| < n$, in which case w is in $V_1 V_2 \cdots V_m X_{m+1} \cdots X_{|w|}$, hence in $U_m X_{m+1} \cdots X_{|w|}$, and therefore in $U_{|w|}$ by the definition of telescopes. We claim that $\mathcal{U}$ is equal to the union of all ${}_m\lfloor V_1 V_2 \cdots V_m\rangle_n$, where $0 \leq m \leq n$, and $V_1 V_2 \cdots V_m$ ranges over the open rectangles contained in U_m. Indeed, given any word $w \in \lfloor\mathcal{U}\rangle$, either $|w| \geq n$ and we can take $m = n$, $V_1 = X_1, \ldots, V_m = X_m$; or $|w| = m < n$, then w is in some open rectangle $V_1 V_2 \cdots V_m$ contained in U_m, by definition of the product topology on $\prod_{i=1}^m U_i$, whence $w \in {}_m\lfloor V_1 V_2 \cdots V_m\rangle_n$.

Finally, we observe that the complement of ${}_m\lfloor V_1 V_2 \cdots V_m\rangle_n$ is the set of words w such that either $|w| < m$, or $m \leq |w| < n$ and for some i, $1 \leq i \leq m$, the ith letter of w is not in V_i. When $m = 0$, this is empty. When $m = n$, the condition $m \leq |w| < n$ is always false, so the complement of ${}_m\lfloor V_1 V_2 \cdots V_m\rangle_n$ is equal to $\lceil X^{m-1}\rangle$. Finally, when $m \neq 0$ and $m < n$, write F_i for the complement of V_i, then the complement of ${}_m\lfloor V_1 V_2 \cdots V_m\rangle_n$ is equal to $\lceil X^{m-1}\rangle \cup \bigcup_{i=1}^m \lceil X^{i-1} F_i X^{n-i-1}\rangle$ if $m \geq 1$.

So the complement of any open of the prefix topology is an intersection of finite unions of the claimed subbasic closed sets, and we conclude. □

We need the following, which also appears in Goubault-Larrecq (2013, Exercise 9.7.36).

Lemma D.1. *Let $X_1, X_2, \ldots, X_n, \ldots$ be countably many topological spaces. The map $i_n \colon X_1 \times \cdots \times X_n \to \rhd_{n=1}^{+\infty} X_n$ that sends each n-tuple $(a_1, a_2, \ldots, a_n)$ to the word $a_1 a_2 \cdots a_n$, and the map $cons \colon X_1 \times \rhd_{n=2}^{+\infty} X_n \to \rhd_{n=1}^{+\infty} X_n$ that sends $a_1, a_2 a_3 \cdots a_n$ to $a_1 a_2 a_3 \cdots a_n$ are both continuous.*

Proof. To show that i_n is continuous, we note that the inverse image of the open subset $\lfloor\mathcal{U}\rangle$, where $\mathcal{U}$ is any telescope $U_0, U_1, \ldots, U_n, \ldots$, is the open subset U_n. To show that *cons* is continuous, it is easier to show that the inverse image of a subbasic closed set $\lceil F_1 F_2 \cdots F_n\rangle$ (see Proposition 9.2) is closed. Indeed, this inverse image is empty if $n = 0$ and equal to $F_1 \times \lceil F_2 \cdots F_n\rangle$ otherwise. □

Lemma 9.3 (recap). *Let $X_1, X_2, \ldots, X_n, \ldots$, be countably many topological spaces. The subsets of the form $\lceil C_1 C_2 \cdots C_n\rangle$, where C_i is irreducible closed in X_i for each i, $1 \leq i \leq n$, are irreducible closed in $\rhd_{n=1}^{+\infty} X_n$.*

Proof. $C_1 \times C_2 \times \cdots \times C_n$ is irreducible closed in $X_1 \times \cdots \times X_n$ by Lemma A.2. The map i_n is continuous, so $cl(i_n[C_1 \times C_2 \times \cdots \times C_n])$ is irreducible closed by Lemma A.1. We claim that the latter is exactly $\lceil C_1 C_2 \cdots C_n\rangle$, and this will prove the lemma. Indeed, $i_n[C_1 \times C_2 \times \cdots \times C_n]$ is contained in $\lceil C_1 C_2 \cdots C_n\rangle$, hence so does its closure. Conversely, any word in $\lceil C_1 C_2 \cdots C_n\rangle$ is also in the downward closure of $i_n[C_1 \times C_2 \times \cdots \times C_n]$ (with respect to $\leq^{\rhd}$), hence in the set $cl(i_n[C_1 \times C_2 \times \cdots \times C_n])$, since the latter is closed hence downward-closed. □

Lemma 9.5 (recap). *Let $X_1, X_2, \ldots, X_n, \ldots$, be countably many non-empty topological spaces. The whole space $\rhd_{n=1}^{+\infty} X_n$ is irreducible closed in itself.*

Proof. Passing to complements, it is equivalent to show that the intersection of two non-empty opens is again non-empty. Any two non-empty open subsets of $\rhd_{n=1}^{+\infty} X_n$ are of the form $\lfloor\mathcal{U}\rangle$ and $\lfloor\mathcal{U}'\rangle$ for two wide telescopes $\mathcal{U} = (U_n)_{n\in\mathbb{N}}$ and $\mathcal{U}' = (U'_n)_{n\in\mathbb{N}}$. For n large enough, $U_n = U'_n = \prod_{i=1}^n X_i$, so any length n heterogeneous word is in $\lfloor\mathcal{U}\rangle \cap \lfloor\mathcal{U}'\rangle$. □

To show that there is no other irreducible closed subset, we rest on the following general-purpose lemma.

Lemma D.2. *Let Y be a topological space and $\mathcal{B}$ be a subbase of closed sets of Y. Assume that any (possibly infinite) non-empty intersection of elements of $\mathcal{B}$ can be written as a finite union of elements of $\mathcal{B}$. Then every irreducible closed subset of Y is a member of $\mathcal{B}$ or equals Y itself.*

Proof. Let C be an arbitrary irreducible closed subset of Y. As a closed set, C can be written as $\bigcap_{i\in I} \bigcup_{j\in J_i} F_{ij}$, where I is some index set, J_i is finite for every $i \in I$, and $F_{ij} \in \mathcal{B}$ for all i, j. If I is

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

empty, then $C = Y$. Otherwise, for each $i \in I$, C is contained in $\bigcup_{j \in J_i} F_{ij}$, so C is contained in F_{ij_i} for some $j_i \in J_i$, since C is irreducible. Since F_{ij_i} is clearly contained in $\bigcup_{j \in J_i} F_{ij}$, $C = \bigcap_{i \in I} F_{ij_i}$. By assumption, C can be written as a finite union $\bigcup_{k=1}^{n} B_k$ of elements of $\mathcal{B}$. Since C is irreducible, again, C must equal some B_k, $1 \leq k \leq n$, whence $C \in \mathcal{B}$. □

Lemma 9.6 (recap). *Let $X_1, X_2, \ldots, X_n, \ldots,$ be countably many non-empty topological spaces. The only irreducible closed subsets of $\rhd_{n=1}^{+\infty} X_n$ are $\rhd_{n=1}^{+\infty} X_n$ itself, and the subsets of the form $\lceil C_1 C_2 \cdots C_n \rangle$, where C_i is irreducible closed in X_i for each i, $1 \leq i \leq n$.*

Proof. Let $\mathcal{B}$ be the subbase of closed sets of the form $\lceil F_1 F_2 \cdots F_n \rangle$. Any non-empty intersection of such sets is again of this form. In fact, whenever I is non-empty, the intersection $\bigcap_{i \in I} \lceil F_{i1} F_{i2} \cdots F_{in_i} \rangle$ equals $\lceil \bigcap_{i \in I} F_{i1} \bigcap_{i \in I} F_{i2} \cdots \bigcap_{i \in I} F_{i \min_{j \in I} n_j} \rangle$. So Lemma D.2 applies the only irreducible closed subsets of $\rhd_{n=1}^{+\infty} X_n$ other than $\rhd_{n=1}^{+\infty} X_n$ itself are in $\mathcal{B}$.

Now assume $\lceil F_1 F_2 \cdots F_n \rangle$ is irreducible. Without loss of generality, no F_i is empty: otherwise, letting k be the smallest index such that F_k is empty, one can rewrite $\lceil F_1 F_2 \cdots F_n \rangle$ as $\lceil F_1 F_2 \cdots F_{k-1} \rangle$.

For each i, $1 \leq i \leq n$, since F_i is non-empty, fix an element x_i of F_i. We claim that F_i must be irreducible for each i, $1 \leq i \leq n$. Otherwise, there would be two closed subsets F' and F'' such that $F_i \subseteq F' \cup F''$, but F_i is contained neither in F' nor in F''. In this case, let x' be an element of F_i outside F', and x'' an element of F_i outside F''. Then $x_1 x_2 \cdots x_{i-1} x'$ is in $\lceil F_1 F_2 \cdots F_i \cdots F_n \rangle$ but not in $\lceil F_1 F_2 \cdots F' \cdots F_n \rangle$ (where F' replaces F_i at position i), and $x_1 x_2 \cdots x_{i-1} x''$ is in $\lceil F_1 F_2 \cdots F_i \cdots F_n \rangle$ but not in $\lceil F_1 F_2 \cdots F'' \cdots F_n \rangle$ (where F'' replaces F_i at position i), although $\lceil F_1 F_2 \cdots F_i \cdots F_n \rangle$ is contained in the union $\lceil F_1 F_2 \cdots F' \cdots F_n \rangle \cup \lceil F_1 F_2 \cdots F'' \cdots F_n \rangle$. This would contradict the fact that $\lceil F_1 F_2 \cdots F_i \cdots F_n \rangle$ is irreducible. So each F_i is irreducible. □

Proposition 9.7 (recap). *Let $X_1, X_2, \ldots, X_n, \ldots,$ be countably many non-empty topological spaces. The map $i \colon (\rhd_{n=1}^{+\infty} \mathcal{S}(X_n))^{\top} \to \mathcal{S}(\rhd_{n=1}^{+\infty} X_n)$ that sends $\top$ to $\rhd_{n=1}^{+\infty} X_n$ and the word $C_1 C_2 \cdots C_n$ (where $C_i \in \mathcal{S}(X_i)$ for each i) to $\lceil C_1 C_2 \cdots C_n \rangle$ is an order isomorphism and a homeomorphism.*

Proof. First, i is well defined, by Lemmas 9.5 and 9.3. It is surjective by Lemma 9.6.

The specialization quasi-ordering on $(\rhd_{n=1}^{+\infty} \mathcal{S}(X_n))^{\top}$ is $\subseteq^{\rhd\top}$.

Notice that $C_1 \cdots C_m \subseteq^{\rhd\top} C'_1 \cdots C'_n$ iff $C_1 \cdots C_m \subseteq^{\rhd} C'_1 \cdots C'_n$ iff $m \leq n$ and $C_1 \subseteq C'_1, \ldots, C_m \subseteq C'_m$. We claim that this is equivalent to $\lceil C_1 \cdots C_m \rangle \subseteq \lceil C'_1 \cdots C'_n \rangle$. The only if direction is clear. In the if direction, pick $x_1 \in C_1, \ldots, x_m \in C_m$: the word $x_1 \cdots x_m$ is in $\lceil C_1 \cdots C_m \rangle \subseteq \lceil C'_1 \cdots C'_n \rangle$, so $x_1 \in C'_1, \ldots, x_m \in C'_m$. In particular, $m \leq n$; also, letting $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m$ remain fixed, but varying x_i in C_i, we obtain that $C_i \subseteq C'_i$.

Notice also that $\top \subseteq^{\rhd\top} C'_1 \cdots C'_n$ never holds, and that $\rhd_{n=1}^{+\infty} X_n \subseteq \lceil C'_1 \cdots C'_n \rangle$ never holds either, since $\rhd_{n=1}^{+\infty} X_n$ contains words of arbitrary lengths.

Notice finally that $C_1 \cdots C_m \subseteq^{\rhd\top} \top$ always holds, and correspondingly $\lceil C_1 \cdots C_m \rangle \subseteq \rhd_{n=1}^{+\infty} X_n$ always holds.

It follows that for every $w, w' \in (\rhd_{n=1}^{+\infty} \mathcal{S}(X_n))^{\top}$, $w \subseteq^{\rhd\top} w'$ iff $i(w) \subseteq i(w')$. In particular, $i(w) = i(w')$ entails $w = w'$, so that i is injective. Since i is surjective, it is bijective. This also shows that i and its inverse are monotonic, so that i is an order isomorphism.

For each closed subset F of a space X, write $\Box F$ for the family of all irreducible closed subsets C of X such that $C \subseteq F$. Alternatively, $\Box F$ is the complement of the open subset $\diamond U$ of $\mathcal{S}(X)$, for U the complement of F in X. In particular, $\Box F$ is closed in $\mathcal{S}(X)$, and all closed subsets of $\mathcal{S}(X)$ are of this form. Since $\diamond$ commutes with arbitrary unions and finite intersections, $\Box$ commutes with arbitrary intersections and finite unions, so given a subbase $\mathcal{B}$ of closed subsets of X, the sets $\Box F$ with $F \in \mathcal{B}$ form a subbase of the closed subsets of $\mathcal{S}(X)$. In particular, the sets $\Box \lceil F_1, F_2, \ldots, F_n \rangle$ with each F_i closed in X_i form a subbase of closed sets of $\mathcal{S}(\rhd_{n=1}^{+\infty} X_n)$. Their

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

inverse image by i is the set of words $C_1C_2\cdots C_m$ in $\rhd_{n=1}^{+\infty}\mathcal{S}(X_n)$ such that $\lceil C_1, C_2, \ldots, C_m\rangle \subseteq \lceil F_1, F_2, \ldots, F_n\rangle$. This is equivalent to $C_1\cdots C_m \subseteq^{\rhd\top} F_1F_2\cdots F_n$, by an argument similar to one we have already seen at the beginning of the present proof. (The difference is that $F_1, F_2, \ldots, F_n$ are no longer irreducible, contrarily to $C'_1, \ldots, C'_n$ – but we never used irreducibility there.) Therefore, $i^{-1}(\Box\lceil F_1, F_2, \ldots, F_n\rangle) = \lceil\Box F_1, \Box F_2, \ldots, \Box F_n\rangle$. This shows that i is continuous.

Since i is bijective, let j be its inverse. We have just shown that $j^{-1}(\lceil\Box F_1, \Box F_2, \ldots, \Box F_n\rangle) = \Box\lceil F_1, F_2, \ldots, F_n\rangle$. The sets $\Box F_i$, for F_i closed in X_i, span all the closed subsets of $\mathcal{S}(X_i)$, since their complements $\diamond U$ for U open span all the open subsets of $\mathcal{S}(X_i)$. Using Proposition 9.2, the sets $\lceil\Box F_1, \Box F_2, \ldots, \Box F_n\rangle$ with all F_i closed form a subbase of closed subsets of $\rhd_{n=1}^{+\infty}\mathcal{S}(X_n)$. Together with the whole set $(\rhd_{n=1}^{+\infty}\mathcal{S}(X_n))^\top$, they form a subbase of closed subsets of $(\rhd_{n=1}^{+\infty}\mathcal{S}(X_n))^\top$, whose inverse images by j are closed: either the closed set $\Box\lceil F_1, F_2, \ldots, F_n\rangle$ or the whole space $\mathcal{S}(\rhd_{n=1}^{+\infty} X_n)$. Therefore, $j = i^{-1}$ is also continuous, hence i is a homeomorphism. □

Lemma 9.8 (recap). *Let $X_1, X_2, \ldots, X_n$ be non-empty topological spaces. The only irreducible closed subsets of $\rhd_{k=1}^{n} X_k$ are the subsets of the form $\lceil C_1C_2\cdots C_m\rangle$, where C_i is irreducible closed in X_i for each i, $1 \le i \le m$, and $m \le n$.*

Proof. The proof of Lemma 9.6, with minor changes shows that all irreducible closed subsets are of this form, except possibly for the whole space $\rhd_{k=1}^{n} X_k$. We show that the latter cannot be irreducible, unless it is itself of one of the above form.

Assume that $\rhd_{k=1}^{n} X_k$ is irreducible, that is, any two non-empty opens have a non-empty intersection. In particular, given any two non-empty open subsets U_k and V_k of X_k, the open subset $\lfloor\emptyset, \ldots, \emptyset, X_1\cdots X_{k-1}U_k, X_1\cdots X_{k-1}U_kX_{k+1}, \ldots, X_1\cdots X_{k-1}U_kX_{k+1}\cdots X_n, \cdots\rangle$ and the open subset $\lfloor\emptyset, \ldots, \emptyset, X_1\cdots X_{k-1}V_k, X_1\cdots X_{k-1}V_kX_{k+1}, \ldots, X_1\cdots X_{k-1}V_kX_{k+1}\cdots X_n, \cdots\rangle$ are both non-empty and must have non-empty intersection. Any word in this intersection must be of length at least k, and its kth letter must be both in U_k and in V_k. So any two non-empty open subsets U_k and V_k of X_k must have non-empty intersection: X_k is irreducible. Then, $\rhd_{k=1}^{n} X_k = \lceil X_1X_2\cdots X_n\rangle$ is irreducible closed by Lemma 9.3, and we are done. □

Proposition 9.9 (recap). *Let $X_1, X_2, \ldots, X_n$ be non-empty topological spaces. The map $i\colon \rhd_{k=1}^{n}\mathcal{S}(X_k) \to \mathcal{S}(\rhd_{k=1}^{n} X_k)$ that sends the word $C_1C_2\cdots C_k$ (where $k \le n$ and $C_i \in \mathcal{S}(X_i)$ for each i) to $\lceil C_1C_2\cdots C_k\rangle$ is an order isomorphism and a homeomorphism.*

Proof. The proof is as for Proposition 9.7, using now Lemma 9.8 instead of Lemma 9.6. □

Appendix E. Proofs of Results on Trees (Section 11)

It is sometimes convenient to be able to talk about subterms and positions p, together with the *subterm $t_{|p}$ of t at position p*. A *position* is a finite word over $\mathbb{N}$. The empty word ϵ is always a position in any term t, and $t_{|\epsilon} = t$. Whenever $t_{|p}$ is defined, and, $t_{|p}$ is of the form $f(t_1, \ldots, t_n)$, then $p\,i$ is a position in t for every i, $1 \le i \le n$, and $t_{|pi} = t_i$. The *size* of a term is the number of its positions. We write $t[s]_p$ for the term t, except that the subterm at position p has been replaced by s.

The following generalizes the notion of simple tree expression: for U open in X and $\mathcal{U}$ open in $\mathcal{T}(X)^*$, let $\Diamond U\cdot\mathcal{U}$ be the set of all terms that have a subterm of the form $f(\vec{t})$ with $f \in U$ and $\vec{t} \in \mathcal{U}$. We use them in proving the first part of Proposition 11.1:

Lemma E.1. *Let X be a topological space. Every finite intersection of simple tree expressions can be rewritten as a finite union of simple tree expressions. In particular, the simple tree expressions form a base of the tree topology.*

Proof. Let $Y = \mathcal{T}(X)$. The empty intersection is $\Diamond X()$, and it remains to compute binary intersections. We do this in two steps. First, for all opens U and U' of X, for all opens $\mathcal{U}$ and $\mathcal{U}'$ of

Y^*, we show that the intersection of $\Diamond U \cdot \mathcal{U}$ and $\Diamond U' \cdot \mathcal{U}'$ can be expressed as the union of simpler expressions of the form $\Diamond U'' \cdot \mathcal{U}''$. Next, in the special case where $\Diamond U \cdot \mathcal{U}$ is a simple tree expression $\Diamond U(\pi_1 \mid \cdots \mid \pi_n)$, namely when $\mathcal{U}$ is of the form $Y^*\pi_1 Y^* \cdots Y^*\pi_n Y^*$, and similarly for $\Diamond U' \cdot \mathcal{U}'$, we show that each of the simpler expressions $\Diamond U'' \cdot \mathcal{U}''$ obtained in Step 1 can themselves be expressed as finite unions of simple tree expressions.

Step 1. The intersection of $\Diamond U \cdot \mathcal{U}$ and $\Diamond U' \cdot \mathcal{U}'$ is the union of:

1. $\Diamond U \cdot (\mathcal{U} \cap (Y^*(\Diamond U' \cdot \mathcal{U}')Y^*)$,
2. $\Diamond U' \cdot (\mathcal{U}' \cap (Y^*(\Diamond U \cdot \mathcal{U})Y^*)$,
3. $\Diamond(U \cap U') \cdot (\mathcal{U} \cap \mathcal{U}')$,
4. $\Diamond X \cdot (\Diamond U \cdot \mathcal{U} \mid \Diamond U' \cdot \mathcal{U}')$
5. $\Diamond X \cdot (\Diamond U' \cdot \mathcal{U}' \mid \Diamond U \cdot \mathcal{U})$.

Indeed, the terms t of $\Diamond U \cdot \mathcal{U} \cap \Diamond U' \cdot \mathcal{U}'$ are those that have a subterm $t_{|p} = f(\vec{t})$ with $f \in U$ and $\vec{t} \in \mathcal{U}$ and that have a subterm $t_{|p'} = f'(\vec{t}')$ with $f' \in U'$ and $\vec{t}' \in \mathcal{U}'$, for some positions p and p'. If p is a proper prefix of p' (we say that $f(\vec{t})$ is *above* $f'(\vec{t}')$), then t is in $\Diamond U \cdot (\mathcal{U} \cap (Y^*(\Diamond U' \cdot \mathcal{U}')Y^*)$ (case 1 above); if p' instead is a proper prefix of p, then t is in case 2; if $p = p'$, then t is in case 3; if p and p' are incomparable, then t is in case 4 if $f(\vec{t})$ is to the left of $f'(\vec{t}')$ (i.e., the first element that differs in p and p' is less in p than in p'), and in case 5 if $f(\vec{t})$ is to the right of $f'(\vec{t}')$. Conversely, each of the opens 1–5 are clearly contained both in $\Diamond U \cdot \mathcal{U}$ and in $\Diamond U' \cdot \mathcal{U}'$.

Note. The operator $\Diamond V\cdot$ commutes with finite unions, that is, $\Diamond V \cdot \bigcup_{i=1}^m \mathcal{V}_i = \bigcup_{i=1}^m \Diamond V \cdot \mathcal{V}_i$. This is easy: both sides are the set of terms such that there is a subterm $f(\vec{t})$ with $f \in V$ and there is an i, $1 \leq i \leq m$, such that $\vec{t} \in \mathcal{V}_i$. We will use that freely below.

Step 2. Call an *elementary open* any open subset of Y^* of the form $Y^*\pi_1 Y^* \cdots Y^*\pi_n Y^*$, where $\pi_1, \ldots, \pi_n$ are simple tree expressions. We claim that, for all open subsets U and U' of X, for all elementary opens $\mathcal{U}$ and $\mathcal{U}'$, $\Diamond U \cdot \mathcal{U} \cap \Diamond U' \cdot \mathcal{U}'$ is a finite union of (denotations of) simple tree expressions. We show this by induction over the size of the expressions $\Diamond U \cdot \mathcal{U}$ and $\Diamond U' \cdot \mathcal{U}'$. Write $\mathcal{U}$ as $Y^*\pi_1 Y^* \cdots Y^*\pi_m Y^*$, and $\mathcal{U}'$ as $Y^*\pi'_1 Y^* \cdots Y^*\pi'_n Y^*$. Then $\mathcal{U} \cap (Y^*(\Diamond U' \cdot \mathcal{U}')Y^*)$ (case 1) is the union of all elementary opens of the form $Y^*\pi_1 Y^* \cdots Y^*\pi_{i-1} Y^*(\Diamond U' \cdot \mathcal{U}')Y^*\pi_i Y^* \cdots Y^*\pi_m Y^*$, $1 \leq i \leq m+1$, plus all opens of the form $Y^*\pi_1 Y^* \cdots Y^*(\pi_i \cap \Diamond U' \cdot \mathcal{U}')Y^* \cdots Y^*\pi_m Y^*$, $1 \leq i \leq m$. For each i, the latter is the (finite) union of the elementary opens $Y^*\pi_1 Y^* \cdots Y^*\pi'' Y^* \cdots Y^*\pi_m Y^*$, where π'' ranges over the (finitely many) simple tree expressions given by the induction hypothesis, and whose union equals $\pi_i \cap \Diamond U' \cdot \mathcal{U}'$. So $\mathcal{U} \cap (Y^*(\Diamond U' \cdot \mathcal{U}')Y^*)$ is a (finite) union of elementary opens, say $\mathcal{U}_1, \ldots, \mathcal{U}_k$: then the open $\Diamond U \cdot (\mathcal{U} \cap (Y^*(\Diamond U' \cdot \mathcal{U}')Y^*)$ of case 1 is the (finite) union $\bigcup_{i=1}^k \Diamond U \cdot \mathcal{U}_i$. Case 2 is symmetric.

Cases 4 and 5 are already in the form of simple tree expressions.

For case 3, we show that $\mathcal{U} \cap \mathcal{U}'$ is a finite union of elementary opens, by induction on $m+n$, using the formulae that we have already used in the proof of Lemma 7.1. When $m = 0$ or $n = 0$, this is clear. Otherwise, write $\mathcal{U} = Y^*\pi_1\mathcal{V}$, $\mathcal{U}' = Y^*\pi'_1\mathcal{V}'$, so $\mathcal{U} \cap \mathcal{U}'$ is the union of $Y^*\pi_1(\mathcal{V} \cap Y^*\pi'_1\mathcal{V}')$, of $Y^*\pi'_1(Y^*\pi_1\mathcal{V} \cap \mathcal{V}')$, and of $Y^*(\pi_1 \cap \pi'_1)(\mathcal{V} \cap \mathcal{V}')$. By induction hypothesis, $\mathcal{V} \cap Y^*\pi'_1\mathcal{V}'$ is a finite union of elementary opens, so $Y^*\pi_1(\mathcal{V} \cap Y^*\pi'_1\mathcal{V}')$ is, too, since unions distribute over concatenations. Similarly for $Y^*\pi'_1(Y^*\pi_1\mathcal{V} \cap \mathcal{V}')$. For $Y^*(\pi_1 \cap \pi'_1)(\mathcal{V} \cap \mathcal{V}')$, $\pi_1 \cap \pi'_1$ is a finite union of simple tree expressions by induction hypothesis (the first one, on π and π'), and $\mathcal{V} \cap \mathcal{V}'$ is a finite union of elementary opens by induction hypothesis (the second one, on $\mathcal{U}$ and $\mathcal{U}'$). We then distribute unions over concatenations again to conclude. □

Lemma E.2. *For every open subset U of X, for every open subset $\mathcal{U}$ of $\mathcal{T}(X)^*$, $\Diamond U \cdot \mathcal{U}$ is open in $\mathcal{T}(X)$.*

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Proof. Let $Y = \mathcal{T}(X)$. Notice that $\mathcal{U}$ can be written as a union of opens of the form $Y^* U_1 Y^* \cdots Y^* U_n Y^*$, where each U_i is open in Y, because those form a base of the word topology on Y^* (Lemma 7.1). By Lemma E.1, $U_1, \ldots, U_n$ can all be written as unions of simple tree expressions. Distributing unions over concatenations, $\mathcal{U}$ is then a union of elementary opens $\mathcal{U}_i$ of Y^*, $i \in I$ (in the sense already used in Lemma E.1). Then, $\Diamond U \cdot \mathcal{U} = \bigcup_{i \in I} \Diamond U \cdot \mathcal{U}_i$ is a union of simple tree expressions, hence is open. □

Lemma E.3. *Letting $\leq$ be the specialization quasi-ordering of X, every open subset of $\mathcal{T}(X)$ is upward-closed with respect to $\preceq_\leq$.*

Proof. This is Exercise 9.7.43 of Goubault-Larrecq (2013). For short, let Y denote $\mathcal{T}(X)$. We show that whenever $s \preceq_\leq t$ and $s \in \pi$, then $t \in \pi$. This is by induction on the structure of π. Write π as $\Diamond U(\pi_1 \mid \ldots \mid \pi_n)$. There must be a subterm, say at position p, of s, of the form $f'(\vec{s}')$, with $f' \in U$ and $\vec{s}' \in Y^* \pi_1 Y^* \ldots Y^* \pi_n Y^*$. Let $(*)$ be our induction hypothesis: whenever $u \preceq_\leq v$ and $u \in \pi_i$ for some i, $1 \leq i \leq n$, then $v \in \pi_i$.

We show that $t \in \pi$ for any term t such that $s \preceq_\leq t$, whenever s contains a subterm at some position p of the form $f'(\vec{s}')$, with $f' \in U$ and $\vec{s}' \in Y^* \pi_1 Y^* \ldots Y^* \pi_n Y^*$, by a secondary induction on the size of t. If $s \preceq_\leq t$ by the first case of the definition, that is, if $t = g(t_1, \ldots, t_p)$ and $s \preceq_\leq t_j$ for some j, $1 \leq j \leq p$, then by induction hypothesis $t_j \in \pi$, from which $t \in \pi$ follows immediately, by definition of (the denotation of) π.

So assume that $s \preceq_\leq t$ by the second case of the definition, that is, $s = f(\vec{s})$, $t = g(\vec{t})$, $f \leq g$, and $\vec{s} \preceq_\leq^* \vec{t}$. Write $\vec{s}$ as $s_1 s_2 \ldots s_m$, $\vec{t}$ as $t_1 t_2 \ldots t_n$, so that there is an (injective) increasing map $h\colon \{1, 2, \ldots, m\} \to \{1, 2, \ldots, n\}$ with $s_1 \preceq_\leq t_{h(1)}, s_2 \preceq_\leq t_{h(2)}, \ldots, s_m \preceq_\leq t_{h(m)}$.

If $p = \varepsilon$, then $f' = f$ and $\vec{s}' = \vec{s}$, so $\vec{s} \in Y^* \pi_1 Y^* \ldots Y^* \pi_n Y^*$. Let $s_{i_1} \ldots s_{i_n}$ be a subword of $\vec{s}$ satisfying $s_{i_1} \in \pi_1, \ldots, s_{i_n} \in \pi_n$. Then $t_{h(i_1)} \in \pi_1, \ldots, t_{h(i_n)} \in \pi_n$ by $(*)$, and $t_{h(i_1)} \ldots t_{h(i_n)}$ forms a subword of $\vec{t}$. So $\vec{t} \in Y^* \pi_1 Y^* \ldots Y^* \pi_n Y^*$; since U is upward-closed, $g \in U$; so $t \in \pi$.

Finally, if $p \neq \varepsilon$, then $f'(\vec{s}')$ must be a subterm of some s_i, $1 \leq i \leq m$. Since $s_i \preceq_\leq t_{h(i)}$, by induction hypothesis, $t_{h(i)}$ is in π, from which $t \in \pi$ follows immediately. □

Lemma E.4. *For every open subset U of X, every open subset $\mathcal{U}$ of $\mathcal{T}(X)^*$, and every open subset V of $\mathcal{T}(X)$, let $\Diamond U \cdot \mathcal{U} \mathbin{/\!/} V$ be the set of all terms containing a subterm $f(\vec{t}) \in V$ with $f \in U$ and $\vec{t} \in \mathcal{U}$. Then $\Diamond U \cdot \mathcal{U} \mathbin{/\!/} V$ is open in $\mathcal{T}(X)$.*

Proof. This is the first part of Exercise 9.7.44 of Goubault-Larrecq (2013). Let $Y = \mathcal{T}(X)$. We first note that $\Diamond U \cdot \mathcal{U} \mathbin{/\!/} V$ is open if V is (the denotation of) a simple tree expression $\pi = \Diamond U'(\pi'_1 \mid \ldots \mid \pi'_n)$: in that case $\Diamond U \cdot \mathcal{U} \mathbin{/\!/} \pi = \Diamond(U \cap U') \cdot (\mathcal{U} \cap Y^* \pi'_1 Y^* \ldots Y^* \pi'_n Y^*) \cup \Diamond U \cdot (\mathcal{U} \cap Y^* \pi Y^*)$. This is open by Lemma E.2. In the general case, by Lemma E.1, we can write V as a union of simple tree expressions π_i, $i \in I$. We conclude that $\Diamond U \cdot \mathcal{U} \mathbin{/\!/} V = \bigcup_{i \in I} \Diamond U \cdot \mathcal{U} \mathbin{/\!/} \pi_i$ is indeed open. □

Lemma E.5. *For every closed subset F of X, and all closed subsets $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n$ of $Y = \mathcal{T}(X)$, let $F^{\overline{?}}(\mathcal{F}_1^? \mathcal{F}_2^? \cdots \mathcal{F}_n^?)$ denote the union of $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n$ with the set of those terms $f(\vec{t})$ such that $f \in F$ and $\vec{t} \in \mathcal{F}_1^? \mathcal{F}_2^? \cdots \mathcal{F}_n^?$. (Recall the word-products of Section 7.) Then $F^{\overline{?}}(\mathcal{F}_1^? \mathcal{F}_2^? \cdots \mathcal{F}_n^?)$ is closed in Y: let $U = X \setminus F$, V be the complement of $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \cdots \cup \mathcal{F}_n$ in Y, $\mathcal{U}$ be the complement of $\mathcal{F}_1^? \mathcal{F}_2^? \ldots \mathcal{F}_n^?$ in Y^*, then $F^{\overline{?}}(\mathcal{F}_1^? \mathcal{F}_2^? \cdots \mathcal{F}_n^?)$ is the complement of $(\Diamond X \cdot \mathcal{U} \mathbin{/\!/} V) \cup (\Diamond U \cdot Y^* \mathbin{/\!/} V)$.*

Proof. This is the second part of Exercise 9.7.44 of Goubault-Larrecq (2013). Let us characterize the complement of $F^{\overline{?}}(\mathcal{F}_1^? \mathcal{F}_2^? \ldots \mathcal{F}_n^?)$. Note that U and V are open by definition, while $\mathcal{U}$ is open by Corollary 7.6. Then $\Diamond X \cdot \mathcal{U} \mathbin{/\!/} V$ and $\Diamond U \cdot Y^* \mathbin{/\!/} V$ are open, as we have seen in Lemma E.4.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Let t be a term outside $F^{\overline{?}}(\mathcal{F}_1^?\mathcal{F}_2^? \ldots \mathcal{F}_n^?)$. Since t is not in $F^{\overline{?}}(\mathcal{F}_1^?\mathcal{F}_2^? \ldots \mathcal{F}_n^?)$, we first observe that t is in V, otherwise it would be in $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \ldots \cup \mathcal{F}_n$. Write t as $f(\vec{t})$. If $f \in F$, then $\vec{t}$ cannot be in $\mathcal{F}_1^?\mathcal{F}_2^? \ldots \mathcal{F}_n^?$. Recall that $\mathcal{U}$ is the complement of $\mathcal{F}_1^?\mathcal{F}_2^? \ldots \mathcal{F}_n^?$. So, if $f \in F$, then t is in $\Diamond X \cdot \mathcal{U} \;/\!/\; V$. If on the other hand $f \notin F$, then t is in $\Diamond U \cdot Y^* \;/\!/\; V$.

Conversely, consider any element t of $(\Diamond X \cdot \mathcal{U} \;/\!/\; V) \cup (\Diamond U \cdot Y^* \;/\!/\; V)$. We claim that t cannot be in $F^{\overline{?}}(\mathcal{F}_1^?\mathcal{F}_2^? \ldots \mathcal{F}_n^?)$. Notice first that any subterm of a term in $F^{\overline{?}}(\mathcal{F}_1^?\mathcal{F}_2^? \ldots \mathcal{F}_n^?)$ is again in $F^{\overline{?}}(\mathcal{F}_1^?\mathcal{F}_2^? \ldots \mathcal{F}_n^?)$: this follows easily from the definition, and the fact that $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \ldots \cup \mathcal{F}_n$, being closed, is downward-closed with respect to $\preceq_\leq$ (as a consequence of Lemma E.3), hence is closed under taking subterms. Now let t be both in $(\Diamond X \cdot \mathcal{U} \;/\!/\; V) \cup (\Diamond U \cdot Y^* \;/\!/\; V)$ and in $F^{\overline{?}}(\mathcal{F}_1^?\mathcal{F}_2^? \ldots \mathcal{F}_n^?)$. If t is in $\Diamond X \cdot \mathcal{U} \;/\!/\; V$, then t has a subterm $f(\vec{t}) \in V$ with $\vec{t} \in \mathcal{U}$. Since t is in $F^{\overline{?}}(\mathcal{F}_1^?\mathcal{F}_2^? \ldots \mathcal{F}_n^?)$, its subterm $f(\vec{t})$ is in $F^{\overline{?}}(\mathcal{F}_1^?\mathcal{F}_2^? \ldots \mathcal{F}_n^?)$, too. But since $\vec{t} \in \mathcal{U}$, $\vec{t} \notin \mathcal{F}_1^?\mathcal{F}_2^? \ldots \mathcal{F}_n^?$, so $f(\vec{t})$ must be in $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \ldots \cup \mathcal{F}_n$; but this would contradict the fact that $f(\vec{t}) \in V$. If t is instead in $\Diamond U \cdot Y^* \;/\!/\; V$, then t has a subterm $f(\vec{t}) \in V$ with $f \in U$, that is, $f \notin F$. Again $f(\vec{t})$ is in $F^{\overline{?}}(\mathcal{F}_1^?\mathcal{F}_2^? \ldots \mathcal{F}_n^?)$ and $f \notin F$ entails that $f(\vec{t})$ must be in $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \ldots \cup \mathcal{F}_n$, again contradicting $f(\vec{t}) \in V$.

So $F^{\overline{?}}(\mathcal{F}_1^?\mathcal{F}_2^? \ldots \mathcal{F}_n^?)$ is the complement of $(\Diamond X \cdot \mathcal{U} \;/\!/\; V) \cup (\Diamond U \cdot Y^* \;/\!/\; V)$. Since the latter is open, the former is closed. □

Proposition 11.1 (recap). *Let X be a topological space. Every finite intersection of simple tree expressions can be rewritten as a finite union of simple tree expressions. In particular, the simple tree expressions form a base of the tree topology.*

Letting $\leq$ be the specialization quasi-ordering of X, the specialization quasi-ordering of $\mathcal{T}(X)$ is the embedding quasi-ordering $\preceq_\leq$.

Proof. The first part is Lemma E.1. For the second part, let $\preceq$ denote temporarily the specialization quasi-ordering of $\mathcal{T}(X)$. Using Lemma E.3, $t \preceq_\leq t'$ implies $t \preceq t'$. Conversely, we show by structural induction on t' that its downward closure $\downarrow_{\preceq_\leq} t'$ in $\preceq_\leq$ is closed: write t' as $f(t'_1, t'_2, \ldots, t'_n)$, then $\downarrow_{\preceq_\leq} t'$ is equal to $(\downarrow_X f)^{\overline{?}}((\downarrow_{\preceq_\leq} t'_1)^?(\downarrow_{\preceq_\leq} t'_2)^? \ldots (\downarrow_{\preceq_\leq} t'_n)^?)$, which is closed by Lemma E.5 and the induction hypothesis. So, if $t \preceq t'$ then t is in the closure of $\{t'\}$, hence in the closed set $\downarrow_{\preceq_\leq} t'$, whence $t \preceq_\leq t'$. □

Proposition 11.2 (recap). *Let X be a set quasi-ordered by $\leq$. The tree topology on $\mathcal{T}(X_a)$ is exactly the Alexandroff topology of $\preceq_\leq$ on $\mathcal{T}(X)$.*

Proof. Any upward-closed subset A of $Y = \mathcal{T}(X)$ is the union of $\uparrow_Y s$, $s \in A$, where upward closure is taken relative to $\preceq_\leq$. We claim that $\uparrow_Y s$ is obtained recursively by $\uparrow_Y s = \Diamond(\uparrow_X f)(\uparrow_Y s_1 \mid \ldots \mid \uparrow_Y s_m)$, where $s = f(s_1, \ldots, s_m)$. Indeed, let $t = g(t_1, \ldots, t_n)$ be any term. We first show that if $t \in \uparrow_Y s$, that is, if $s \preceq_\leq t$, then $t \in \Diamond(\uparrow_X f)(\uparrow_Y s_1 \mid \ldots \mid \uparrow_Y s_m)$: s clearly belongs to $\Diamond(\uparrow_X f)(\uparrow_Y s_1 \mid \ldots \mid \uparrow_Y s_m)$, and since open sets are upward-closed in the specialization quasi-ordering, which is $\preceq_\leq$ by Proposition 11.1, t is also in $\Diamond(\uparrow_X f)(\uparrow_Y s_1 \mid \ldots \mid \uparrow_Y s_m)$. Conversely, if $t \in \Diamond(\uparrow_X f)(\uparrow_Y s_1 \mid \ldots \mid \uparrow_Y s_m)$, then either $t = g(\vec{t})$ is itself the subterm such that $g \in \uparrow_X f$ and $\vec{t} \in Y^*(\uparrow_Y s_1)Y^* \ldots Y^*(\uparrow_Y s_m)Y^*$, so $f \leq g$ and $\vec{s} \preceq^*_\leq \vec{t}$, in particular $s \preceq_\leq t$; or t_j is in $\Diamond(\uparrow_X f)(\uparrow_Y s_1 \mid \ldots \mid \uparrow_Y s_m)$ for some j, $1 \leq j \leq n$, so $s \preceq_\leq t_j$ by induction hypothesis, therefore $s \preceq_\leq t$.

It follows that $\uparrow_Y s$ is open in the tree topology for every $s \in A$, so $\uparrow_Y A$ is also open in the tree topology. Conversely, every open subset if upward-closed in $\preceq_\leq$ by Proposition 11.1. So the tree topology is the Alexandroff topology of $\preceq_\leq$. □

E.1 Tree steps

Lemma E.6. *The application map $@ \colon X \times \mathcal{T}(X)^* \to \mathcal{T}(X)$, which sends $(f, \vec{t})$ to $f(\vec{t})$, is continuous.*

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Proof. This is the first part of Exercise 9.7.47 of Goubault-Larrecq (2013). Let $Y = \mathcal{T}(X)$, and $\pi = \Diamond U(\pi_1, \ldots, \pi_n)$ be a simple tree expression. We show that $@^{-1}(\pi)$ is open: $@^{-1}(\pi)$ is the union of the open $U \times (Y^* \pi_1 Y^* \ldots Y^* \pi_n Y^*)$ with $X \times (Y^* \pi Y^*)$; indeed, $f(\vec{t}) \in \pi$ iff either $f \in U$ and $\vec{t} \in Y^* \pi_1 Y^* \ldots Y^* \pi_n Y^*$ (case where the needed subterm of $f(\vec{t})$ that witnesses the fact that $f(\vec{t})$ is in π is $f(\vec{t})$ itself), or some element of the sequence $\vec{t}$ is in π. So $@^{-1}(\pi)$ is open, and therefore $@$ is continuous. □

Lemma 11.4 (recap). *Let X be a topological space. For every closed subset F of X, and every word-product $\vec{P}$ on $\mathcal{T}(X)$, supp $\vec{P}$ and $F^{\vec{?}}(\vec{P})$ are closed in $\mathcal{T}(X)$. If moreover $F = C$ is irreducible, then so is the tree step $C^{\vec{?}}(\vec{P})$.*

Proof. Let $Y = \mathcal{T}(X)$. The first part of the Lemma is a slight extension of Lemma E.5.

First, supp $\vec{P}$ is just $i^{-1}(\vec{P})$, where $i \colon Y \to Y^*$ is the continuous map such that $i(t)$ is the word with just one letter, t. By Corollary 7.6, $\vec{P}$ is closed, so supp $\vec{P}$ is closed too.

Let t be a term outside $F^{\vec{?}}(\vec{P})$. Let V be the complement of supp $\vec{P}$. We have just seen that V is open. Moreover, since t is not in $F^{\vec{?}}(\vec{P})$, t is not in supp $\vec{P}$, so t is in V. Write t as $f(\vec{t})$. If $f \in F$, then $\vec{t}$ cannot be in $\vec{P}$. Let $\mathcal{U}$ be the complement of $\vec{P}$ in $\mathcal{T}(X)^*$. This is open. So, if $f \in F$, then t is in $\Diamond X \cdot \mathcal{U} \mathbin{//} V$, which is open by Lemma E.4. If on the other hand $f \notin F$, then t is in $\Diamond U \cdot Y^* \mathbin{//} V$, where U is the complement of F in X; therefore, $t \in (\Diamond X \cdot \mathcal{U} \mathbin{//} V) \cup (\Diamond U \cdot Y^* \mathbin{//} V)$.

Conversely, consider any element t of $(\Diamond X \cdot \mathcal{U} \mathbin{//} V) \cup (\Diamond U \cdot Y^* \mathbin{//} V)$. We claim that t cannot be in $F^{\vec{?}}(\vec{P})$. Notice that any subterm of a term in $F^{\vec{?}}(\vec{P})$ is again in $F^{\vec{?}}(\vec{P})$: this follows easily from the definition, and the fact that supp $\vec{P}$, being closed, is downward-closed with respect to $\preceq_{\leq}$, hence is closed under taking subterms. So let t be both in $(\Diamond X \cdot \mathcal{U} \mathbin{//} V) \cup (\Diamond U \cdot Y^* \mathbin{//} V)$ and in $F^{\vec{?}}(\vec{P})$. If t is in $\Diamond X \cdot \mathcal{U} \mathbin{//} V$, then t has a subterm $f(\vec{t}) \in V$ with $\vec{t} \in \mathcal{U}$. Since t is in $F^{\vec{?}}(\vec{P})$, its subterm $f(\vec{t})$ is in $F^{\vec{?}}(\vec{P})$, too. But since $\vec{t} \in \mathcal{U}$, $\vec{t} \notin \vec{P}$, so $f(\vec{t})$ must be in supp $\vec{P}$; this contradicts the fact that $f(\vec{t}) \in V$. If t is instead in $\Diamond U \cdot Y^* \mathbin{//} V$, then t has a subterm $f(\vec{t}) \in V$ with $f \in U$, that is, $f \notin F$. Again $f(\vec{t})$ is in $F^{\vec{?}}(\vec{P})$, and $f \notin F$ entails that $f(\vec{t})$ must be in supp $\vec{P}$, again contradicting $f(\vec{t}) \in V$.

So $F^{\vec{?}}(\vec{P})$ is the complement of $(\Diamond X \cdot \mathcal{U} \mathbin{//} V) \cup (\Diamond U \cdot Y^* \mathbin{//} V)$. Since the latter is open, the former is closed.

Let us now assume that $F = C$ is irreducible.

By Lemma 7.7, the word-product $\vec{P}$ over $\mathcal{T}(X)$ is irreducible closed in $\mathcal{T}(X)^*$. So $C \times \vec{P}$ is irreducible closed in $X \times \mathcal{T}(X)^*$, since the product of two irreducible closed subsets is irreducible closed (Lemma A.2). Since $@$ is continuous (Lemma E.6), one concludes that $cl(@[C \times \vec{P}])$ is irreducible closed in $\mathcal{T}(X)$ (Lemma A.1). Since $C^{\vec{?}}(\vec{P})$ is closed and clearly contains $@[C \times \vec{P}]$, it contains $cl(@[C \times \vec{P}])$. Conversely, the latter contains $\{f(\vec{t}) \mid f \in C, \vec{t} \in \vec{P}\}$ and is downward-closed in $\preceq_{\leq}$, so is closed under taking subterms, whence $cl(@[C \times \vec{P}])$ contains $C^{\vec{?}}(\vec{P})$. Therefore, the latter is the closure $cl(@[C \times \vec{P}])$ and must then be irreducible. □

E.2 Tree iterators

We need the following lemma to show that tree iterators define closed sets. A relation R from a space X to a space Y is a subset of $X \times Y$. It is *lower semi-continuous* iff $\mathrm{Pre}^{\exists} R(V) = \{x \in X \mid \exists y \in V \cdot x \mathrel{R} y\}$ is open for every open subset V of Y. It is *upper semi-continuous* iff $\mathrm{Pre}^{\forall} R(V) = \{x \in X \mid \forall y \cdot x \mathrel{R} y \Rightarrow y \in V\}$ is open for every open subset V of Y. It is *continuous* if and only if it is both lower semi-continuous and upper semi-continuous.

Lemma E.7. *Let Z be a topological space, $\Box$ be a hole outside Z, and* inst-of *be the relation from Z^* to $(Z+\{\Box\})^*$ defined by w* inst-of *w' iff w is obtained from w' by replacing each occurrence of $\Box$ by (possibly distinct) elements from Z. Formally, iff w and w' have the same length and for every index i, the ith letter of w' is either $\Box$ or equal to the ith letter of w.*

Then inst-of *is continuous.*

Proof. For short, let Y be $Z+\{\Box\}$.

Lower semi-continuity. Consider any basic open $V=Y^*V_1Y^*\cdots Y^*V_nY^*$ of Y. Let $U_i=V_i\subseteq Z$ if $\Box\notin V_i$, $U_i=Z$ otherwise. Then, $\mathrm{Pre}^{\exists}\,\mathsf{inst\text{-}of}\,(V)=Z^*U_1Z^*\cdots Z^*U_nZ^*$ is open. Since every open subset of Y is a union of basic open sets, and since $\mathrm{Pre}^{\exists}\,\mathsf{inst\text{-}of}$ commutes with unions, $\mathrm{Pre}^{\exists}\,\mathsf{inst\text{-}of}\,(V)$ is open for every open subset V of Y^*.

Upper semi-continuity. We first observe that, for every word-product $P=e_1e_2\cdots e_n$ on Y, $\mathrm{Pre}^{\exists}\,\mathsf{inst\text{-}of}\,(P)$ is a word-product on Z. Indeed, $\mathrm{Pre}^{\exists}\,\mathsf{inst\text{-}of}\,(P)$ is equal to $\mathrm{Pre}^{\exists}\,\mathsf{inst\text{-}of}\,(e_1)$ $\mathrm{Pre}^{\exists}\,\mathsf{inst\text{-}of}\,(e_2)\cdots\mathrm{Pre}^{\exists}\,\mathsf{inst\text{-}of}\,(e_n)$, while $\mathrm{Pre}^{\exists}\,\mathsf{inst\text{-}of}\,(F^?)$ equals $Z^?$ if $\Box\in F$ and $F^?$ otherwise, and $\mathrm{Pre}^{\exists}\,\mathsf{inst\text{-}of}\,(F^*)$ equals Z^* if $\Box\in F$ and F^* otherwise. Call a *monotone Boolean combination* of word-products any finite union of finite intersections of word-products. Lemma 7.12 shows that any finite intersection of word-products can be rewritten as a finite union of word-products. So the monotone Boolean combinations of word-products are the finite unions of word-products $\bigcup_{i=1}^m P_i$. Now $\mathrm{Pre}^{\exists}\,\mathsf{inst\text{-}of}$ commutes with unions, so $\mathrm{Pre}^{\exists}\,\mathsf{inst\text{-}of}\,(F)$ is a finite union of word-products on Z (hence closed in Z^* by Corollary 7.6) for every monotone Boolean combination F of word-products on Y.

Using Lemma 7.2, the complement of any monotone Boolean combination U of basic opens of Y^* is a monotone Boolean combination of word-products. So $\mathrm{Pre}^{\forall}\,\mathsf{inst\text{-}of}\,(U)$, which is the complement of $\mathrm{Pre}^{\exists}\,\mathsf{inst\text{-}of}\,(F)$, assuming that F is the complement of U, is open in Z^*.

Consider now any open subset U of Y^*. U is a union of basic opens, hence a directed union $\bigcup_{i\in I}U_i$, where each U_i is a finite union (in particular, a monotone Boolean combination) of basic opens. We observe that $\mathrm{Pre}^{\forall}\,\mathsf{inst\text{-}of}$ commutes with directed unions. This is because each word w (say of length m) in Z^* only has finitely many images $w_1, w_2, \ldots, w_{2^m}$, namely the 2^m words obtained from w by replacing each letter by $\Box$, or not: if $w\in\mathrm{Pre}^{\forall}\,\mathsf{inst\text{-}of}\,(\bigcup_{i\in I}U_i)$, then for every j, $1\le j\le 2^m$, there is an $i\in I$ such that $w_j\in U_i$; we may take the same i for every j, by directedness, whence $w\in\mathrm{Pre}^{\forall}\,\mathsf{inst\text{-}of}\,(U_i)$; the converse direction is obvious. So $\mathrm{Pre}^{\forall}\,\mathsf{inst\text{-}of}\,(U)$ is the directed union $\bigcup_{i\in I}\mathrm{Pre}^{\forall}\,\mathsf{inst\text{-}of}\,(U_i)$ and is therefore open: inst-of is upper semi-continuous. $\Box$

Lemma E.8. *Let X an Z be topological spaces. The relation $\mathrm{id}_X\times$* inst-of *that relates $(f,\vec{t})\in X\times Z^*$ with $(f,\vec{u})\in X\times(Z+\{\Box\})^*$ if and only if $\vec{t}$* inst-of *$\vec{u}$ is continuous.*

Proof. Let $Y=Z+\{\Box\}$, and fix an arbitrary open subset $\mathcal{V}$ of $X\times(Z+\{\Box\})^*$. Write $\mathcal{V}$ as a union of open rectangles $\bigcup_{i\in I}U_i\times W_i$ (where every U_i and every W_i is open; this is the definition of the product topology).

Lower semi-continuity. $\mathrm{Pre}^{\exists}(\mathrm{id}_X\times\mathsf{inst\text{-}of}\,)(\mathcal{V})$ is equal to $\bigcup_{i\in I}U_i\times\mathrm{Pre}^{\exists}\,\mathsf{inst\text{-}of}\,(W_i)$, hence is open.

Upper semi-continuity. We must show that $\mathcal{U}=\mathrm{Pre}^{\forall}(\mathrm{id}_X\times\mathsf{inst\text{-}of}\,)(\mathcal{V})$ is open. For that, fix $(f,\vec{t})$: it is enough to find an open rectangle containing $(f,\vec{t})$ and included in $\mathcal{U}$. As in Lemma E.8, note that there are only finitely many elements $\vec{u}$ such that $\vec{t}$ inst-of $\vec{u}$. List them as $\vec{u}_1,\ldots,\vec{u}_m$. For each one, $(f,\vec{u}_j)$ is in some $U_i\times W_i$: pick one such i and call it i_j. Our desired open rectangle is $U\times V$ where $U=\bigcap_{j=1}^m U_{i_j}$ and $V=\mathrm{Pre}^{\forall}\,\mathsf{inst\text{-}of}\,(\bigcup_{j=1}^m W_{i_j})$. By construction, $(f,\vec{t})$ is in $U\times V$. For every element $(g,\vec{s})$ of $U\times V$, by definition every element $(g,\vec{u})$ that is related to $(g,\vec{s})$ by $\mathrm{id}_X\times$ inst-of is such that $g\in U$ and $\vec{u}\in\bigcup_{j=1}^m W_{i_j}$. Let j be such $\vec{u}\in W_{i_j}$. Since $g\in U$, g is in U_{i_j}, so $(g,\vec{u})$ is in $U_{i_j}\times W_{i_j}\subseteq\mathcal{U}$. $\Box$

Lemma E.9. *Let $\mathfrak{C}^{\overline{*}}.S$ be a tree iterator such that $\text{args}\,\mathfrak{C}$ is closed. Any subterm s of a term t in $\mathfrak{C}^{\overline{*}}.S$ is again in $\mathfrak{C}^{\overline{*}}.S$.*

Proof. This is proved by structural induction on t. If t is in S, then any subterm s of t is such that $s \preceq_{\leq} t$, hence $s \in S$. If $t \in \text{args}\,\mathfrak{C}$, we argue similarly, since $\mathfrak{C}$ is closed (this is the first place where we need this assumption). Otherwise, either $s = t$ and the claim is obvious, or s is a proper subterm of t. In the latter case, there is an elementary context $c = f(u_1 u_2 \cdots u_m) \in \mathfrak{C}$ such that t is obtained from c by replacing those u_j, $1 \leq j \leq m$, that equal $\Box$ by terms from $\mathfrak{C}^{\overline{*}}.S$, that is, t can be written $f(t_1, t_2, \ldots, t_m)$, where $t_j = u_j$ if $u_j \neq \Box$, $t_j \in \mathfrak{C}^{\overline{*}}.S$ otherwise. For some j, s is a subterm of t_j. If $u_j = \Box$, then $t_j \in \mathfrak{C}^{\overline{*}}.S$, so that $s \in \mathfrak{C}^{\overline{*}}.S$ by induction hypothesis. Otherwise, we claim that t_j is in $\text{args}\,\mathfrak{C}$. Indeed, $c = f(u_1 u_2 \cdots u_m) \in \mathfrak{C}$, so the smaller $f(u_j)$ is in $\mathfrak{C}$, too, since $\mathfrak{C}$ is closed in $X \times (\mathcal{T}(X) + \{\Box\})^*$, hence downward-closed. Since $u_j = t_j$, $f(t_j)$ is in $\mathfrak{C}$, so $t_j \in \text{args}\,\mathfrak{C}$. It follows that $t_j \in \mathfrak{C}^{\overline{*}}.S$. Since s is a subterm of t_j, by induction hypothesis $s \in \mathfrak{C}^{\overline{*}}.S$. $\Box$

Lemma 11.9 (recap). *Let X be a topological space, and $\Box$ be a hole outside $\mathcal{T}(X)$. Every tree iterator $\mathfrak{C}^{\overline{*}}.S$ such that $\text{args}\,\mathfrak{C}$ is closed in $\mathcal{T}(X)$ denotes a closed subset of $\mathcal{T}(X)$.*

Proof. Let V be the complement of $S \cup \text{args}\,\mathfrak{C}$ in $\mathcal{T}(X)$; this is open, since both S and $\text{args}\,\mathfrak{C}$ are closed. Let also $\mathcal{V}$ be the open complement of $\mathfrak{C} \cap (X \times (\text{args}\,\mathfrak{C} + \{\Box\})^*)$ in $X \times Y^*$, where we let Y abbreviate $\mathcal{T}(X) + \{\Box\}$. Using Lemma E.8, the binary relation $\text{id}_X \times \textsf{inst-of}$ between $X \times \mathcal{T}(X)^*$ and $X \times Y^*$ is continuous, hence upper semi-continuous. Therefore, $\mathcal{U} = \text{Pre}^{\forall}(\text{id}_X \times \textsf{inst-of})(\mathcal{V}) = \{(f, \vec{t}) \mid \forall \vec{u}. \vec{t}\ \textsf{inst-of}\ \vec{u} \Rightarrow (f, \vec{u}) \in \mathcal{V}\}$ is open.

Write $\mathcal{U}$ as a union of open rectangles $\bigcup_{i \in I} U_i \times W_i$. We claim that the complement of $\mathfrak{C}^{\overline{*}}.S$ is $\bigcup_{i \in I} \Diamond U_i \cdot W_i \,/\!/\, V$. It will follow that $\mathfrak{C}^{\overline{*}}.S$ is closed.

Let t be any term not in $\mathfrak{C}^{\overline{*}}.S$. Consider a minimal subterm $f(\vec{t})$ of t that is not in $\mathfrak{C}^{\overline{*}}.S$. By minimal, we mean that all its proper subterms are in $\mathfrak{C}^{\overline{*}}.S$. Since $f(\vec{t})$ is not in $\mathfrak{C}^{\overline{*}}.S$, it is in particular not in $S \cup \text{args}\,\mathfrak{C}$, hence it is in V. For any tuple $\vec{u} \in Y^*$ of which $\vec{t}$ is an instance, that is, such that $\vec{t}\ \textsf{inst-of}\ \vec{u}$, and such that the components of $\vec{u}$ that are different from $\Box$ are in $\text{args}\,\mathfrak{C}$ (i.e., $\vec{u} \in (\text{args}\,\mathfrak{C} + \{\Box\})^*$), $(f, \vec{u})$ cannot be in $\mathfrak{C}$: otherwise $f(\vec{t})$ would be obtained from $f(\vec{u})$ by replacing each occurrence of $\Box$ by some components of the tuple $\vec{t}$, which are all in $\mathfrak{C}^{\overline{*}}.S$, so $f(\vec{t})$ would again be in $\mathfrak{C}^{\overline{*}}.S$, which is impossible. Another way of stating this is that whenever $(f, \vec{t})\ (\text{id}_X \times \textsf{inst-of})\ (f, \vec{u})$, then either $\vec{u}$ is not in $(\text{args}\,\mathfrak{C} + \{\Box\})^*$ or $(f, \vec{u})$ is not in $\mathfrak{C}$. That is, $(f, \vec{t})$ is in $\mathcal{U}$. It follows that, for some $i \in I$, $f \in U_i$, and $\vec{t} \in W_i$. Recall that $f(\vec{t}) \in V$. So t is in $\Diamond U_i \cdot W_i \,/\!/\, V$.

Conversely, assume that $t \in \Diamond U_i \cdot W_i \,/\!/\, V$, for some $i \in I$. That is, t has a subterm $f(\vec{t})$ in V, with $f \in U_i$ and $\vec{t} \in W_i$. Assume, for the sake of contradiction, that t is in $\mathfrak{C}^{\overline{*}}.S$. By Lemma E.9, $f(\vec{t})$ is also in $\mathfrak{C}^{\overline{*}}.S$. Since $f(\vec{t}) \in V$, $f(\vec{t})$ is neither in S nor in $\text{args}\,\mathfrak{C}$, so there is an elementary context $f(\vec{u})$ in $\mathfrak{C}$ such that $\vec{t}$ is obtained from $\vec{u}$ by replacing the $\Box$ elements in $\vec{u}$ by some terms in $\mathfrak{C}^{\overline{*}}.S$. Since $f(\vec{u})$ is in $\mathfrak{C}$, the components of $\vec{u}$ that are different from $\Box$ are in $\text{args}\,\mathfrak{C}$, by definition of the argument support. So $\vec{u}$ is in $(\text{args}\,\mathfrak{C} + \{\Box\})^*$, hence $(f, \vec{u}) \in \mathfrak{C} \cap (X \times (\text{args}\,\mathfrak{C} + \{\Box\})^*)$, the complement of $\mathcal{V}$. However, $(f, \vec{t})\ (\text{id}_X \times \textsf{inst-of})\ (f, \vec{u})$, so $(f, \vec{t})$ cannot be in $\mathcal{U} = \text{Pre}^{\forall}(\text{id}_X \times \textsf{inst-of})(\mathcal{V})$. This contradicts the fact that $(f, \vec{t})$ is in $U_i \times W_i$.

This concludes our proof that the complement of $\mathfrak{C}^{\overline{*}}.S$ is $\bigcup_{i \in I'} \Diamond U_i \cdot W_i \,/\!/\, V$, so that $\mathfrak{C}^{\overline{*}}.S$ is closed. $\Box$

Lemma 11.11 (recap). *Let X be a topological space, and $\Box$ be a hole outside $\mathcal{T}(X)$, $\mathfrak{C}$ be a closed subset of $X \times (\mathcal{T}(X) + \{\Box\})^*$, S be a closed subset of $\mathcal{T}(X)$, and assume that $\text{args}\,\mathfrak{C}$ is closed. Then, the tree iterator $\mathfrak{C}^{\overline{*}}.S$ is irreducible in the following cases:*

(1) if $\mathfrak{C}$ is non-$\Box$-linear, and S is non-empty;
(2) or if $\mathfrak{C}$ is $\Box$-generated and $\Box$-linear and S is irreducible;
(3) or if $\mathfrak{C}$ is non-empty, $\Box$-generated, and S is empty.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Proof. (1) $\mathfrak{C}$ is non-$\Box$-linear, and S is non-empty. Since $S \subseteq \overline{\mathfrak{C}^*.S}$, $\overline{\mathfrak{C}^*.S}$ is non-empty. Since $\mathfrak{C}$ is non-$\Box$-linear, in particular there is an elementary context $f(\vec{u}) \in \mathfrak{C}$ such that $\vec{u}$ has at least two occurrences of $\Box$. More precisely, there is an element of the form $f(\vec{u}_1\Box\vec{u}_2\Box\vec{u}_3)$ in $\mathfrak{C}$, that is, one where $\Box$ occurs at least twice. If $\overline{\mathfrak{C}^*.S}$ is included in the union of two closed subsets S' and S'', but not in S' or S'', then pick t' in $\overline{\mathfrak{C}^*.S}$ outside S', and t'' in $\overline{\mathfrak{C}^*.S}$ outside S''. Pick some term t in $\overline{\mathfrak{C}^*.S}$ (e.g., t' or t''), and let $\vec{t}_1, \vec{t}_2, \vec{t}_3$ be obtained from $\vec{u}_1, \vec{u}_2, \vec{u}_3$, respectively, by replacing all occurrences of $\Box$ by t. Clearly $f(\vec{t}_1 t' \vec{t}_2 t'' \vec{t}_3)$ is in $\overline{\mathfrak{C}^*.S}$, hence in S' or in S''. Assume without loss of generality that it is in S'. Then its subterm t' is in S', contradiction. So $\overline{\mathfrak{C}^*.S}$ is irreducible.

(2) $\mathfrak{C}$ is $\Box$-generated and $\Box$-linear, and S is irreducible. Assume that $\overline{\mathfrak{C}^*.S}$ is included in the union of two closed subsets S' and S'', but not in S' or in S''. We claim that there is a context $c' \in \mathfrak{C}^{\Box *}$, with exactly one occurrence of $\Box$, and a term $t' \in S \cup \text{args}\,\mathfrak{C}$, such that $c'[t']$ is not in S'. Indeed, since $\mathfrak{C}$ is $\Box$-linear, there is a context $c = c_1[c_2[\cdots[c_k]\cdots]]$, $k \in \mathbb{N}$, where each c_i is in $\mathfrak{C}$, and such that one obtains a term outside S' by replacing the unique occurrence of $\Box$ (if any) in c by a term from $S \cup \text{args}\,\mathfrak{C}$. If $\Box$ actually occurs (once) in c, let $c' = c$, and the term outside S' obtained above can be written $c'[t']$ for some $t' \in S \cup \text{args}\,\mathfrak{C}$. Otherwise, some c_i does not contain an occurrence of $\Box$. Pick i minimal: so $\Box$ occurs (once) in $c_1, c_2, \ldots, c_{i-1}$, but not in c_i; moreover, $c = c_1[c_2[\cdots[c_{i-1}[c_i]]\cdots]]$ is a term (i.e., where $\Box$ does not occur) outside S'. Write c_i as $f(\vec{u})$. Since $\mathfrak{C}$ is $\Box$-generated, one can split $\vec{u}$ as $\vec{u}_1\vec{u}_2$ so that $f(\vec{u}_1\Box\vec{u}_2) \in \mathfrak{C}$. Pick any term t' from S: this is easy since irreducible sets are non-empty. Let $c' = c_1[c_2[\cdots[c_{i-1}[f(\vec{u}_1\Box\vec{u}_2)]]\cdots]]$. Then $c_i = f(\vec{u}_1\vec{u}_2) \preceq_{\leq} f(\vec{u}_1\Box\vec{u}_2)[t'] = f(\vec{u}_1 t' \vec{u}_2)$, so $c \preceq_{\leq} c'[t']$. Since c is not in S', $c'[t']$ is not in S' either.

In any case, there is a context $c' \in \mathfrak{C}^{\Box *}$, with exactly one occurrence of $\Box$, and a term $t' \in S \cup \text{args}\,\mathfrak{C}$, such that $c'[t']$ is not in S'. Similarly, there is a context $c'' \in \mathfrak{C}^{\Box *}$, with exactly one occurrence of $\Box$, and a term $t'' \in S \cup \text{args}\,\mathfrak{C}$, such that $c''[t'']$ is not in S''. Note that both $c'[t']$ and $c''[t'']$ are in $\overline{\mathfrak{C}^*.S}$.

Examine the case where t' or t'' is in $\text{args}\,\mathfrak{C}$, say t' by symmetry. So $f(t') \in \mathfrak{C}$ and $\Box$ does not occur in t', for some $f \in X$. Since $\mathfrak{C}$ is $\Box$-generated, $f(t'\Box)$ or $f(\Box t')$ is in $\mathfrak{C}$, too, say $f(t'\Box)$. The term $c'[f(t'\Box)[c''[t'']]]$ is then in $\overline{\mathfrak{C}^*.S}$, hence in S' or in S'', say S'. However, the terms $c'[t']$ and $c''[t'']$ are below the latter term in the $\preceq_{\leq}$ ordering, since $\Box$ occurs in c'. So they are both in S', since S' is closed hence downward-closed. But precisely, $c'[t']$ is not in S', contradiction.

Then examine the case where t' and t'' are both in S. We recall from Lemma E.6 that @ is continuous, and from Lemma B.1 that $i\colon Y \to Y^*$ and $cat\colon Y^* \times Y^* \to Y^*$ are continuous. By a simple induction on c', the function that maps each term t to $c'[t]$ is then continuous. Similarly, the function that maps t to $c''[t]$ is continuous. Consider the map f that sends each term $t \in \mathcal{T}(X)$ to $c'[c''[t]] = c'[c''][t]$: f is continuous. Since $c'[c'']$ is in $\mathfrak{C}^{\Box *}$, every $t \in S$ is such that $c'[c''][t]$ is in $\overline{\mathfrak{C}^*.S}$, hence in $S' \cup S''$. So S is included in $f^{-1}(S' \cup S'') = f^{-1}(S') \cup f^{-1}(S'')$. The latter is a union of two closed sets, since f is continuous. Since S is irreducible, S is included in $f^{-1}(S')$ or in $f^{-1}(S'')$. If $S \subseteq f^{-1}(S')$, then in particular $t' \in f^{-1}(S')$, that is, $c'[c''[t']] \in S'$. However, $c'[t'] \preceq_{\leq} c'[c''[t']]$, since $\Box$ occurs in c''. So $c'[t']$ is in S', a contradiction. Similarly, $S \subseteq f^{-1}(S'')$ also leads to a contradiction.

So $\overline{\mathfrak{C}^*.S}$ is in fact included in S' or in S''. We conclude that $\overline{\mathfrak{C}^*.S}$ is irreducible.

(3) $\mathfrak{C}$ is non-empty, $\Box$-generated, and S is empty. Since $\mathfrak{C}$ is non-empty, it contains an elementary context $f(\vec{u})$. Since $\mathfrak{C}$ is closed, hence downward-closed in $\leq \times \preceq_{\leq}$, $f()$ is also in $\mathfrak{C}$, so f is in $\overline{\mathfrak{C}^*.S}$. Let S' be the closure of f in $\mathcal{T}(X)$. Since $\overline{\mathfrak{C}^*.S}$ is closed, it contains S'. So $\overline{\mathfrak{C}^*.S} \supseteq \overline{\mathfrak{C}^*.S'}$. The converse inclusion follows since $S = \emptyset$, whence $\overline{\mathfrak{C}^*.S} = \overline{\mathfrak{C}^*.S'}$. By construction, S' is irreducible closed. If $\mathfrak{C}$ is $\Box$-linear, then $\overline{\mathfrak{C}^*.S} = \overline{\mathfrak{C}^*.S'}$ is irreducible by case (2). If $\mathfrak{C}$ is not $\Box$-linear, then $\overline{\mathfrak{C}^*.S} = \overline{\mathfrak{C}^*.S'}$ is irreducible by case (1). $\Box$

Lemma 11.12 (recap). *Let X be a topological space. The complement $\complement\pi$ of the open subset denoted by the simple tree expression $\pi = \Diamond U(\pi_1 \mid \pi_2 \mid \cdots \mid \pi_n)$ is given by structural induction on π by:*

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

— $\complement\pi = ((F \times \{\Box\}^*) \cup (X \times (\complement\pi_1^*\{\Box\}^?\complement\pi_2^*\{\Box\}^? \cdots \{\Box\}^?\complement\pi_n^*)))^{\overline{*}}.\emptyset$ *if* $n \geq 1$, *where F is the complement of U in X;*
— *if* $n = 0$, *then* $\complement\pi = (F \times \{\Box\}^*)^{\overline{*}}.\emptyset$.

Proof. We first deal with the case $n = 0$. The terms t that are not in $\Diamond U()$ are those such that no function symbol occurring in t is in U. So they are the terms whose function symbols are all in F, that is, the terms in $(F \times \{\Box\}^*)^{\overline{*}}.\emptyset$.

Next, we deal with the case $n \geq 1$. Let $\pi = \Diamond U(\pi_1 \mid \pi_2 \mid \cdots \mid \pi_n)$. Let us explain the notation first. Notice that $\{\Box\}$ is irreducible closed in $\mathcal{T}(X) + \{\Box\}$. So $\{\Box\}^*$ and $\complement\pi_1^*\{\Box\}^?\complement\pi_2^*\{\Box\}^? \cdots \{\Box\}^?\complement\pi_n^*$ are word-products on $\mathcal{T}(X) + \{\Box\}$, hence are closed in $(\mathcal{T}(X) + \{\Box\})^*$, by Corollary 7.6. Write $\mathfrak{C} = (F \times \{\Box\}^*) \cup (X \times (\complement\pi_1^*\{\Box\}^?\complement\pi_2^*\{\Box\}^? \cdots \{\Box\}^?\complement\pi_n^*))$. We must show that $\complement\pi = \mathfrak{C}^{\overline{*}}.\emptyset$. Notice that args $\mathfrak{C} = \complement\pi_1 \cup \complement\pi_2 \cup \cdots \cup \complement\pi_n$, so args $\mathfrak{C}$ is closed.

For every term $t \in \mathcal{T}(X)$, we show, in one direction, that if t is not in π, then t is in $\mathfrak{C}^{\overline{*}}.\emptyset$, by structural induction on t. Write t as $f(\vec{t})$, where $\vec{t} = t_1 t_2 \cdots t_m$. Necessarily, $t_1, t_2, \ldots, t_m$ are outside π as well. So $t_1, t_2, \ldots, t_m$ are in $\mathfrak{C}^{\overline{*}}.\emptyset$. Moreover, $f \notin U$ or $\vec{t} \notin Y^*\pi_1 Y^* \cdots Y^*\pi_n Y^*$, where $Y = \mathcal{T}(X)$. If $f \notin U$, then $f \in F$, so $t = f(\vec{t})$ is obtained from the context $f(\Box^m)$, in $F \times \{\Box\}^*$, by replacing the holes by terms from $\mathfrak{C}^{\overline{*}}.\emptyset$. Therefore, t is itself in $\mathfrak{C}^{\overline{*}}.\emptyset$. Otherwise, $\vec{t}$ is in $\complement\pi_1^* Y^? \complement\pi_2^* Y^? \cdots Y^? \complement\pi_n^*$, by Lemma 7.2. (Recall that $n \geq 1$.) So one can write $\vec{t}$ as a sequence $\vec{t}_1 \in \complement\pi_1^*$, followed by zero or one term s_1, followed by a sequence $\vec{t}_2 \in \complement\pi_2^*$, followed by zero or one term $s_2, \ldots$, followed by zero or one term s_{n-1}, followed by a sequence $\vec{t}_n \in \complement\pi_n^*$. When there is indeed a term s_i between $\vec{t}_i$ and $\vec{t}_{i+1}$, say that s_i *exists*. Note that those terms among $s_1, s_2, \ldots, s_{n-1}$ that do exist are in $\mathfrak{C}^{\overline{*}}.\emptyset$, since they form a subsequence of $\vec{t}$ (use the induction hypothesis). Let $\vec{u}$ be the sequence obtained by concatenating $\vec{t}_1$, $\Box$ if s_1 exists (and nothing otherwise), $\vec{t}_2$, $\Box$ if s_2 exists, $\vec{t}_3, \ldots, \Box$ if s_{n-1} exists, and $\vec{t}_n$. One obtains t by replacing the occurrences of $\Box$ in $f(\vec{u})$ by terms in $\mathfrak{C}^{\overline{*}}.\emptyset$, and $f(\vec{u})$ is in $X \times (\complement\pi_1^*\{\Box\}^?\complement\pi_2^*\{\Box\}^? \cdots \{\Box\}^?\complement\pi_n^*)$ by construction, so $t \in \mathfrak{C}^{\overline{*}}.\emptyset$ again.

Conversely, we claim that no term in $\mathfrak{C}^{\overline{*}}.\emptyset$ can be in (the language of) π. We start by proving the following claim (a): for every j, $1 \leq j \leq n$, no term in $\complement\pi_j$ can be in π. Indeed, if $t \in \complement\pi_j$ is in $\pi = \Diamond U(\pi_1 \mid \pi_2 \mid \cdots \mid \pi_n)$, then t has a subterm $s \in \pi_j$. Then $s \preceq_{\leq} t$, to $t \in \pi_j$ since opens are upward-closed: contradiction.

We then show that whenever $t \in \mathfrak{C}^{\overline{*}}.\emptyset$, then t cannot be in $\pi = \Diamond U(\pi_1 \mid \pi_2 \mid \cdots \mid \pi_n)$, by structural induction on t, following the definition of $\mathfrak{C}^{\overline{*}}.\emptyset$. Assume that $t \in \pi$: there is a subterm $s = g(\vec{s})$ of t such that $g \in U$ and $\vec{s}$ is in $Y^*\pi_1 Y^* \cdots Y^*\pi_n Y^*$. (Again, $Y = \mathcal{T}(X)$.) Note that s itself is in π. By Claim (a), t cannot be in $\complement\pi_1 \cup \complement\pi_2 \cup \cdots \cup \complement\pi_n =$ args $\mathfrak{C}$. It follows that $t = f(\vec{t})$ must be obtained from some elementary context $f(\vec{u})$ in $\mathfrak{C}$ by replacing the occurrences of the hole $\Box$ by terms, themselves from $\mathfrak{C}^{\overline{*}}.\emptyset$. Write $\vec{t}$ as $t_1 t_2 \cdots t_m$, $\vec{u}$ as $u_1 u_2 \cdots u_m$. There are two cases, corresponding to the definition of $\mathfrak{C}$. If $(f, \vec{u}) \in F \times \{\Box\}^*$, then $f \notin U$, so $s = g(\vec{s})$ must be different from t (since $g \in U$), hence s must be a subterm of some t_j, $1 \leq j \leq m$. Moreover, u_j is a hole, so t_j is in $\mathfrak{C}^{\overline{*}}.\emptyset$. By induction hypothesis t_j cannot be in π, hence its subterm $s \preceq_{\leq} t_j$ is not in π either. This is impossible since s is in π. The other case is when $(f, \vec{u})$ is in $X \times (\complement\pi_1^*\{\Box\}^?\complement\pi_2^*\{\Box\}^? \cdots \{\Box\}^?\complement\pi_n^*)$. Then $\vec{t}$ is in $\complement\pi_1^* Y^? \complement\pi_2^* Y^? \cdots Y^? \complement\pi_n^*$, that is, not in $\mathcal{T}(X)^*\pi_1\mathcal{T}(X)^* \cdots \mathcal{T}(X)^*\pi_n\mathcal{T}(X)^*$, by Lemma 7.2 (recall that $n \geq 1$). So again s must be different from t, hence be a subterm of some t_j, $1 \leq j \leq m$. Either t_j is in some $\complement\pi_i$, $1 \leq i \leq n$ (when $u_j \neq \Box$), or $u_j = \Box$ and t_j is in $\mathfrak{C}^{\overline{*}}.\emptyset$. However, $\complement\pi_i$ is included in args $\mathfrak{C}$, hence in $\mathfrak{C}^{\overline{*}}.\emptyset$. So in any case t_j is in $\mathfrak{C}^{\overline{*}}.\emptyset$. Since t_j contains $s = g(\vec{s})$ as a subterm, $s \preceq_{\leq} t_j$ and therefore t_j is also in $\pi = \Diamond U(\pi_1 \mid \pi_2 \mid \cdots \mid \pi_n)$. This is impossible by induction hypothesis. Having reached a contradiction in each case, we conclude. $\Box$

E.3 Checking inclusion between tree steps

Lemma 11.14 (recap). *Let X be a topological space, C and C' be two irreducible closed subsets of X, $\vec{P}$ and $\vec{P}'$ be two word-products over $\mathcal{T}(X)$. Then $C^{\vec{?}}(\vec{P}) \subseteq C'^{\vec{?}}(\vec{P}')$ iff $C \subseteq C'$ and $\vec{P} \subseteq \vec{P}'$, or $C^{\vec{?}}(\vec{P}) \subseteq \text{supp}\, \vec{P}'$.*

Proof. The if direction is obvious, noting that supp $\vec{P}' \subseteq C'^{\vec{?}}(\vec{P}')$. Conversely, assume $C^{\vec{?}}(\vec{P}) \subseteq C'^{\vec{?}}(\vec{P}')$.

For every pair $(f, \vec{t}) \in C \times \vec{P}$, since $f(\vec{t}) \in C^{\vec{?}}(\vec{P}) \subseteq C'^{\vec{?}}(\vec{P}')$, either $(f, \vec{t}) \in C' \times \vec{P}'$, or $(f, \vec{t})$ is in $S = \{(f, \vec{t}) \in X \times \mathcal{T}(X)^* \mid f(\vec{t}) \in \text{supp}\, \vec{P}'\}$. So $C \times \vec{P}$ is included in $(C' \times \vec{P}') \cup S$. Since @ is continuous (Lemma E.6) and supp $\vec{P}'$ is closed (Lemma 11.4), S is closed. By Lemma 7.7, $\vec{P}$ is irreducible, so $C \times \vec{P}$ is irreducible. Also, $C' \times \vec{P}'$ is closed, since $\vec{P}'$ is closed by Corollary 7.6. So $C \times \vec{P}$ is included in $C' \times \vec{P}'$ or in S.

If $C \times \vec{P} \subseteq S$, then $C^{\vec{?}}(\vec{P}) \subseteq \text{supp}\, \vec{P}'$. Indeed, all the terms $f(\vec{t})$ with $f \in C$ and $\vec{t} \in \vec{P}$ are in supp $\vec{P}'$, by the definition of S. And for every $t \in \text{supp}\, \vec{P}$, fix an arbitrary $f \in C$ (since C, being irreducible, is non-empty) to obtain that $f(t) \in C^{\vec{?}}(\vec{P})$ hence $f(t) \in \text{supp}\, \vec{P}'$; since supp $\vec{P}'$ is closed, hence downward-closed in $\preceq_{\leq}$, $t \in \text{supp}\, \vec{P}'$.

If $C \times \vec{P} \subseteq C' \times \vec{P}'$ on the other hand, then clearly $C \subseteq C'$ and $\vec{P} \subseteq \vec{P}'$, since neither C nor $\vec{P}$ is empty. $\square$

E.4 Checking inclusion between tree steps and tree iterators

Lemma E.10. *Let Z be a topological space, $\Box$ a hole outside Z, and F a closed subset of Z. Let* $\mathsf{inst\text{-}of}_F$ *be the relation from Z^* to $(Z + \{\Box\})^*$ defined by w* $\mathsf{inst\text{-}of}_F$ *w' iff w is obtained from w' by replacing each occurrence of $\Box$ by (possibly distinct) elements from F.*

Then $\mathsf{inst\text{-}of}_F$ *is upper semi-continuous.*

Proof. For short, let Y be $Z + \{\Box\}$. For every word-product $P = e_1 e_2 \cdots e_n$ on Y, $\text{Pre}^{\exists}\mathsf{inst\text{-}of}_F(P)$ is equal to $\text{Pre}^{\exists}\mathsf{inst\text{-}of}_F(e_1)\text{Pre}^{\exists}\mathsf{inst\text{-}of}_F(e_2) \cdots \text{Pre}^{\exists}\mathsf{inst\text{-}of}_F(e_n)$, and for each atomic expression e_j, $\text{Pre}^{\exists}\mathsf{inst\text{-}of}_F(e_j)$ is computed as follows: $\text{Pre}^{\exists}\mathsf{inst\text{-}of}_F(F'^{?})$ is $((F' \smallsetminus \{\Box\}) \cup F)^{?}$ if $\Box \in F'$ and $F'^{?}$ otherwise, $\text{Pre}^{\exists}\mathsf{inst\text{-}of}_F(F'^{*})$ is $((F' \smallsetminus \{\Box\}) \cup F)^{*}$ if $\Box \in F'$ and F'^{*} otherwise. (Notice that $F' \smallsetminus \{\Box\} = F' \cap Z$ is closed in Z.) The rest of the proof is as in Lemma E.7. $\square$

Lemma E.11. *Let X and Z be topological spaces, F be a closed subset of Z. The relation $\text{id}_X \times$* $\mathsf{inst\text{-}of}_F$ *that relates $(f, \vec{t}) \in X \times Z^*$ with $(f, \vec{u}) \in X \times (Z + \{\Box\})^*$ if and only if $\vec{t}$* $\mathsf{inst\text{-}of}_F$ *$\vec{u}$ is upper semi-continuous.*

Proof. As for Lemma E.8, using Lemma E.10 instead of Lemma E.7. $\square$

Lemma E.12. *X be a topological space, $\Box$ a hole outside $\mathcal{T}(X)$, $\mathfrak{C}$ be a closed subset of $X \times (\mathcal{T}(X) + \{\Box\})^*$, and S be a closed subset of $\mathcal{T}(X)$. The set $\mathfrak{C}[S]$, defined as the set of all pairs $(f, \vec{t})$ where $\vec{t}$ is obtained from $\vec{u}$ by replacing each occurrence of $\Box$ by possibly different terms from S, for some $\vec{u}$ such that $(f, \vec{u}) \in \mathfrak{C}$, is closed in $X \times \mathcal{T}(X)^*$.*

Proof. $\mathfrak{C}[S]$ is just $\text{Pre}^{\exists}(\text{id}_X \times \mathsf{inst\text{-}of}_S)(\mathfrak{C})$, then use Lemma E.11. $\square$

Lemma 11.15 (recap). *Let X be a topological space, C be an irreducible closed subset of X, $\vec{P}$ be a word-product over $\mathcal{T}(X)$, $\mathfrak{C}$ be a closed subset of $X \times (\mathcal{T}(X) + \{\Box\})^*$ such that args $\mathfrak{C}$ is closed in $\mathcal{T}(X)$, where $\Box$ is a hole outside $\mathcal{T}(X)$, and S be a closed subset of $\mathcal{T}(X)$.*

Then, $C^{\vec{?}}(\vec{P}) \subseteq \mathfrak{C}^{\vec{*}}.S$ iff $C \times \vec{P} \subseteq \mathfrak{C}[\mathfrak{C}^{\vec{*}}.S]$ and supp $\vec{P} \subseteq \mathfrak{C}^{\vec{*}}.S$, or $C^{\vec{?}}(\vec{P}) \subseteq \text{args}\, \mathfrak{C} \cup S$.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

Proof. If $C \times \vec{P} \subseteq \mathfrak{C}[\mathfrak{C}^{\overline{*}}.S]$ and supp $\vec{P} \subseteq \mathfrak{C}^{\overline{*}}.S$, then we claim that every term $t = f(\vec{t})$ in $C^{\overline{?}}(\vec{P})$ is in $\mathfrak{C}^{\overline{*}}.S$. Indeed, either $f \in C$ and $\vec{t} \in \vec{P}$, or $t \in$ supp $\vec{P}$. In the first case, $(f, \vec{t})$ is obtained from some $(f, \vec{u}) \in \mathfrak{C}$ by replacing each occurrence of $\Box$ in $\vec{u}$ by terms from $\mathfrak{C}^{\overline{*}}.S$, so $f(\vec{t})$ is again in $\mathfrak{C}^{\overline{*}}.S$. In the second case, where $t \in$ supp $\vec{P}$, then $t \in \mathfrak{C}^{\overline{*}}.S$ by assumption.

If $C^{\overline{?}}(\vec{P}) \subseteq \text{args}\, \mathfrak{C} \cup S$, then $C^{\overline{?}}(\vec{P})$ is trivially included in $\mathfrak{C}^{\overline{*}}.S$.

Conversely, assume $C^{\overline{?}}(\vec{P}) \subseteq \mathfrak{C}^{\overline{*}}.S$. For every $f \in C$ and $\vec{t} \in \vec{P}$, $f(\vec{t})$ is in $C^{\overline{?}}(\vec{P})$, hence in $\mathfrak{C}^{\overline{*}}.S$. So either $f(\vec{t}) \in \text{args}\, \mathfrak{C} \cup S$, or there is an elementary context $f(\vec{u})$ in $\mathfrak{C}$ such that $\vec{t}$ is obtained from $\vec{u}$ by replacing the occurrences of $\Box$ by terms from $\mathfrak{C}^{\overline{*}}.S$. That is, $f(\vec{t})$ is in $\text{args}\, \mathfrak{C} \cup S$ or in $\mathfrak{C}[\mathfrak{C}^{\overline{*}}.S]$. So $C \times \vec{P}$ is contained in the union of the set $@^{-1}(\text{args}\, \mathfrak{C} \cup S)$, which is closed since @ is continuous, and the set $\mathfrak{C}[\mathfrak{C}^{\overline{*}}.S]$, which is closed by Lemmas E.12 and 11.9. On the other hand, by Lemma 7.7, $\vec{P}$ is irreducible, so $C \times \vec{P}$ is irreducible. So $C \times \vec{P}$ is included in $@^{-1}(\text{args}\, \mathfrak{C} \cup S)$ or in $\mathfrak{C}[\mathfrak{C}^{\overline{*}}.S]$. If $C \times \vec{P} \subseteq @^{-1}(\text{args}\, \mathfrak{C} \cup S)$, then every term $f(\vec{t})$ with $f \in C$ and $\vec{t} \in \vec{P}$ is in args $\cup S$; since $\text{args}\, \mathfrak{C} \cup S$ is closed hence downward-closed, supp $\vec{P}$ is also included in $\text{args}\, \mathfrak{C} \cup S$, and therefore $C^{\overline{?}}(\vec{P})$ as well. Otherwise, $C \times \vec{P} \subseteq \mathfrak{C}[\mathfrak{C}^{\overline{*}}.S]$. Moreover, supp $\vec{P} \subseteq C^{\overline{?}}(\vec{P}) \subseteq \mathfrak{C}^{\overline{*}}.S$. $\Box$

E.5 Checking inclusion between tree iterators

Lemma 11.21 (recap). *Let X be a topological space, $\mathfrak{C}$ and $\mathfrak{C}'$ be two closed subsets of $X \times (\mathcal{T}(X) + \{\Box\})^*$ such that args $\mathfrak{C}$ and args $\mathfrak{C}'$ are closed in $\mathcal{T}(X)$, where $\Box$ is a hole outside $\mathcal{T}(X)$, and let S, S' be two closed subsets of $\mathcal{T}(X)$.*

Then $\mathfrak{C}^{\overline{*}}.S \subseteq \mathfrak{C}'^{\overline{*}}.S'$ iff $\mathfrak{C}[\mathfrak{C}^{\overline{*}}.S] \subseteq @^{-1}(\text{args}\, \mathfrak{C}' \cup S') \cup \mathfrak{C}'[\mathcal{T}^{\Box}(X)]$ and $\text{args}\, \mathfrak{C} \cup S \subseteq \mathfrak{C}'^{\overline{*}}.S'$.

Proof. A warning, first. Although we have used the notation $f(\vec{u})$ for elementary contexts, we must recall that this is an abbreviation for a pair $(f, \vec{u})$. One obtains the term (or context) $f(\vec{u})$ from $(f, \vec{u})$ by applying @, hence the use of @ in the statement of the lemma.

If $\mathfrak{C}^{\overline{*}}.S \subseteq \mathfrak{C}'^{\overline{*}}.S'$, then in particular $\text{args}\, \mathfrak{C} \cup S \subseteq \mathfrak{C}^{\overline{*}}.S \subseteq \mathfrak{C}'^{\overline{*}}.S'$. Moreover, for every $(f, \vec{t}) \in \mathfrak{C}[\mathfrak{C}^{\overline{*}}.S]$, $f(\vec{t})$ is in $\mathfrak{C}^{\overline{*}}.S$. If $f(\vec{t})$ is not in $\text{args}\, \mathfrak{C}' \cup S'$, then $(f, \vec{t})$ is obtained from some elementary context $f(\vec{u}) = (f, \vec{u})$ in $\mathfrak{C}'$ by replacing all occurrences of the hole $\Box$ by terms (in $\mathfrak{C}^{\overline{*}}.S$, but this is irrelevant). In any case $(f, \vec{t}) \in @^{-1}(\text{args}\, \mathfrak{C}' \cup S') \cup \mathfrak{C}'[\mathcal{T}(X)]$, hence in $@^{-1}(\text{args}\, \mathfrak{C}' \cup S') \cup \mathfrak{C}'[\mathcal{T}^{\Box}(X)]$. It follows that $\mathfrak{C}[\mathfrak{C}^{\overline{*}}.S] \subseteq @^{-1}(\text{args}\, \mathfrak{C}' \cup S') \cup \mathfrak{C}'[\mathcal{T}^{\Box}(X)]$.

Conversely, assume that $\mathfrak{C}[\mathfrak{C}^{\overline{*}}.S] \subseteq @^{-1}(\text{args}\, \mathfrak{C}' \cup S') \cup \mathfrak{C}'[\mathcal{T}^{\Box}(X)]$ and $\text{args}\, \mathfrak{C} \cup S \subseteq \mathfrak{C}'^{\overline{*}}.S'$. Consider any term $t = f(\vec{t})$ in $\mathfrak{C}^{\overline{*}}.S$. We show by induction on the definition of $\mathfrak{C}^{\overline{*}}.S$ that t is in $\mathfrak{C}'^{\overline{*}}.S'$. If $t \in \text{args}\, \mathfrak{C} \cup S$ (base case), then $t \in \mathfrak{C}'^{\overline{*}}.S'$ by assumption. Otherwise, $(f, \vec{t})$ is obtained from some $(f, \vec{u}) \in \mathfrak{C}$ by replacing each occurrence of $\Box$ in $\vec{u}$ by elements of $\mathfrak{C}^{\overline{*}}.S$. Let us make this clear. Write $\vec{t}$ as $t_1 t_2 \cdots t_n$, $\vec{u}$ as $u_1 u_2 \cdots u_n$. For each j, $1 \leq j \leq n$, either $t_j = u_j$ or $u_j = \Box$ and $t_j \in \mathfrak{C}^{\overline{*}}.S$. When $u_j \neq \Box$, observe that $t_j = u_j$ is in $\text{args}\, \mathfrak{C} \subseteq \mathfrak{C}^{\overline{*}}.S$. Therefore, in any case, $t_j \in \mathfrak{C}^{\overline{*}}.S$ for every j, $1 \leq j \leq n$. By induction hypothesis, $t_j \in \mathfrak{C}'^{\overline{*}}.S'$ for every j, $1 \leq j \leq n$. On the other hand, the existence of $(f, \vec{u})$, as specified, means that $(f, \vec{t})$ is in $\mathfrak{C}[\mathfrak{C}^{\overline{*}}.S]$. By assumption, $(f, \vec{t})$ is then in $@^{-1}(\text{args}\, \mathfrak{C}' \cup S')$, or in $\mathfrak{C}'[\mathcal{T}^{\Box}(X)]$. In the first case, $f(\vec{t}) \in \text{args}\, \mathfrak{C}' \cup S' \subseteq \mathfrak{C}'^{\overline{*}}.S'$. In the second case, there is a pair $(f, \vec{v}) \in \mathfrak{C}'$ such that $\vec{t}$ is obtained from $\vec{v} = v_1 v_2 \cdots v_n$ by replacing each v_j that equals $\Box$ by the element t_j from $\vec{t}$. We have noticed that all such elements were in $\mathfrak{C}'^{\overline{*}}.S'$. So $f(\vec{t})$ is in $\mathfrak{C}'^{\overline{*}}.S'$. $\Box$

Lemma 11.22 (recap). *Let X be a topological space, $\mathfrak{C}$ and $\mathfrak{C}'$ be two closed subsets of $X \times (\mathcal{T}(X) + \{\Box\})^*$, where $\Box$ is a hole outside $\mathcal{T}(X)$, and let S, S' be two closed subsets of $\mathcal{T}(X)$. Assume also that $\mathfrak{C}$ is of the form $\bigcup_{i=1}^{m} C_i \times Q_i$, and that $\mathfrak{C}'$ is of the form $\bigcup_{j=1}^{n} C'_j \times Q'_j$, where each C_i and each C'_j is irreducible closed in X, and Q_i and Q'_j are word-products over $\mathcal{T}(X) + \{\Box\}$ for each i,*

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

$1 \leq i \leq m$, and each j, $1 \leq j \leq n$. Assume finally that $\mathfrak{C}^{\overline{*}}.S$ is irreducible, and that $\Box \in Q_i$ for every i, $1 \leq i \leq m$.

Then $\mathfrak{C}^{\overline{*}}.S \subseteq \mathfrak{C}'^{\overline{*}}.S'$ iff:

- *either $\mathfrak{C}^{\overline{*}}.S \subseteq \text{args}\,\mathfrak{C}' \cup S'$,*
- *or $\text{args}\,\mathfrak{C} \cup S \subseteq \mathfrak{C}'^{\overline{*}}.S'$, and for every i, $1 \leq i \leq m$, there is a j, $1 \leq j \leq n$, such that $C_i \subseteq C'_j$ and $Q_i[\mathcal{T}^{\Box}(X)] \subseteq Q'_j[\mathcal{T}^{\Box}(X)]$.*

Proof. Consider the following statements:

(*i*) $\mathfrak{C}^{\overline{*}}.S \subseteq \mathfrak{C}'^{\overline{*}}.S'$;

(*ii*) $\text{args}\,\mathfrak{C} \cup S \subseteq \mathfrak{C}'^{\overline{*}}.S'$, and for every i, $1 \leq i \leq m$, either $C_i^{\overline{?}}(Q_i[\mathfrak{C}^{\overline{*}}.S]) \subseteq \text{args}\,\mathfrak{C}' \cup S'$ or for some j, $1 \leq j \leq n$, $C_i \subseteq C'_j$ and $Q_i[\mathfrak{C}^{\overline{*}}.S] \subseteq Q'_j[\mathcal{T}^{\Box}(X)]$.

(*iii*) either $\mathfrak{C}^{\overline{*}}.S \subseteq \text{args}\,\mathfrak{C}' \cup S'$, or $\text{args}\,\mathfrak{C} \cup S \subseteq \mathfrak{C}'^{\overline{*}}.S'$ and for every i, $1 \leq i \leq m$, there is a j, $1 \leq j \leq n$, such that $C_i \subseteq C'_j$ and $Q_i[\mathcal{T}^{\Box}(X)] \subseteq Q'_j[\mathcal{T}^{\Box}(X)]$.

The Lemma claims that (*i*) is equivalent to (*iii*). We shall show this by proving that (*i*) implies (*ii*) implies (*iii*) implies (*i*).

The differences between (*ii*) and (*iii*) are: first, there is an additional disjunct $\mathfrak{C}^{\overline{*}}.S \subseteq \text{args}\,\mathfrak{C}' \cup S'$ in (*iii*); second, (*iii*) dispenses with the disjunct $C_i^{\overline{?}}(Q_i[\mathfrak{C}^{\overline{*}}.S]) \subseteq \text{args}\,\mathfrak{C}' \cup S'$ that occurs in (*ii*); finally, we use $Q_i[\mathfrak{C}^{\overline{*}}.S]$ versus $Q_i[\mathcal{T}^{\Box}(X)]$ in the last inclusion.

Before we start, note that, using Lemma 11.8, $\text{args}\,\mathfrak{C} = \bigcup_{i=1}^{m} \text{supp}\, Q_i \cap \mathcal{T}(X)$ is closed, and similarly, $\text{args}\,\mathfrak{C}' = \bigcup_{j=1}^{n} \text{supp}\, Q'_j \cap \mathcal{T}(X)$ is closed: so Lemma 11.21 applies.

(*i*) $\Rightarrow$ (*ii*). By Lemma 11.21, $\text{args}\,\mathfrak{C} \cup S \subseteq \mathfrak{C}'^{\overline{*}}.S'$ and $\mathfrak{C}[\mathfrak{C}^{\overline{*}}.S] = \bigcup_{i=1}^{m} C_i \times Q_i[\mathfrak{C}^{\overline{*}}.S] \subseteq @^{-1}(\text{args}\,\mathfrak{C}' \cup S') \cup \mathfrak{C}'[\mathcal{T}^{\Box}(X)] = @^{-1}(\text{args}\,\mathfrak{C}' \cup S') \cup \bigcup_{j=1}^{n} (C'_j \times Q'_j[\mathcal{T}^{\Box}(X)])$. Therefore, for every i, $1 \leq i \leq m$, $C_i \times Q_i[\mathfrak{C}^{\overline{*}}.S] \subseteq @^{-1}(\text{args}\,\mathfrak{C}' \cup S') \cup \bigcup_{j=1}^{n} (C'_j \times Q'_j[\mathcal{T}^{\Box}(X)])$. Since $\mathfrak{C}^{\overline{*}}.S$ is irreducible, and because $Q_i[\mathfrak{C}^{\overline{*}}.S]$ is obtained by syntactically replacing occurrences of $\Box$ by $\mathfrak{C}^{\overline{*}}.S$ (Lemma 11.18), $Q_i[\mathfrak{C}^{\overline{*}}.S]$ is a word-product. So $Q_i[\mathfrak{C}^{\overline{*}}.S]$ is irreducible by Lemma 7.7. It follows that $C_i \times Q_i[\mathfrak{C}^{\overline{*}}.S]$ is irreducible. So $C_i \times Q_i[\mathfrak{C}^{\overline{*}}.S] \subseteq @^{-1}(\text{args}\,\mathfrak{C}' \cup S')$ or, for some $1 \leq j \leq n$, $C_i \times Q_i[\mathfrak{C}^{\overline{*}}.S] \subseteq C'_j \times Q'_j[\mathcal{T}^{\Box}(X)]$. In the first case, $C_i^{\overline{?}}(Q_i[\mathfrak{C}^{\overline{*}}.S]) \subseteq \text{args}\,\mathfrak{C}' \cup S'$, and we conclude.

(*iii*) $\Rightarrow$ (*i*). If $\mathfrak{C}^{\overline{*}}.S \subseteq \text{args}\,\mathfrak{C}' \cup S'$, then (*i*) holds trivially. Otherwise, since $Q_i[\mathfrak{C}^{\overline{*}}.S] \subseteq Q_i[\mathcal{T}^{\Box}(X)]$, we obtain that for every i, there is a j such that $C_i \times Q_i[\mathfrak{C}^{\overline{*}}.S] \subseteq C'_j \times Q'_j[\mathcal{T}^{\Box}(X)]$. Therefore, $\mathfrak{C}[\mathfrak{C}^{\overline{*}}.S] = \bigcup_{i=1}^{m} C_i \times Q_i[\mathfrak{C}^{\overline{*}}.S] \subseteq \bigcup_{j=1}^{n} (C'_j \times Q'_j[\mathcal{T}^{\Box}(X)]) = \mathfrak{C}'[\mathcal{T}^{\Box}(X)]$. Also, $\text{args}\,\mathfrak{C} \cup S \subseteq \mathfrak{C}'^{\overline{*}}.S'$ by assumption, so by Lemma 11.21 $\mathfrak{C}^{\overline{*}}.S \subseteq \mathfrak{C}'^{\overline{*}}.S'$.

(*ii*) $\Rightarrow$ (*iii*). If $\mathfrak{C}^{\overline{*}}.S \subseteq \text{args}\,\mathfrak{C}' \cup S'$, then (*iii*) is clear. So let us also assume $\mathfrak{C}^{\overline{*}}.S \not\subseteq \text{args}\,\mathfrak{C}' \cup S'$.

To show (*iii*) under these assumptions, and in the view of what (*ii*) states, it is enough to show that $C_i^{\overline{?}}(Q_i[\mathfrak{C}^{\overline{*}}.S]) \subseteq \text{args}\,\mathfrak{C}' \cup S'$ is impossible, and that for every i and j, if $Q_i[\mathfrak{C}^{\overline{*}}.S] \subseteq Q'_j[\mathcal{T}^{\Box}(X)]$, then the stronger inclusion $Q_i[\mathcal{T}^{\Box}(X)] \subseteq Q'_j[\mathcal{T}^{\Box}(X)]$ holds.

We start with the former. Since $\Box \in Q_i$, $\mathfrak{C}^{\overline{*}}.S$ is included in $C_i^{\overline{?}}(Q_i[\mathfrak{C}^{\overline{*}}.S])$: indeed, for every $t \in \mathfrak{C}^{\overline{*}}.S$, $f(t) \in C_i^{\overline{?}}(Q_i[\mathfrak{C}^{\overline{*}}.S])$ for some (arbitrary) $f \in C_i$, so $t \preceq_{\leq} f(t)$ is also in $C_i^{\overline{?}}(Q_i[\mathfrak{C}^{\overline{*}}.S])$, which is closed hence downward-closed. Therefore, $C_i^{\overline{?}}(Q_i[\mathfrak{C}^{\overline{*}}.S])$ cannot be included in $\text{args}\,\mathfrak{C}' \cup S'$, since we assumed that $\mathfrak{C}^{\overline{*}}.S \not\subseteq \text{args}\,\mathfrak{C}' \cup S'$.

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

We proceed with the other claim. Write Q_i as the product of atomic expressions $e_1 e_2 \cdots e_m$ over $\mathcal{T}(X) + \{\Box\}$, and similarly Q'_j as $e'_1 e'_2 \cdots e'_n$. Let e_i be written as F_i^* or $F_i^?$, $1 \leq i \leq m$, where F_i is closed in $\mathcal{T}(X) + \{\Box\}$, and also irreducible in case e_i is written $F_i^?$. Similarly, write e'_j as $F_j'^*$ or $F_j'^?$.

The core of the argument is that: $(*)$ when $\Box \in e_i$, that is, when e_i is of the form $\{\Box\}^?$ or F_i^* with $\Box \in F_i$, then $e_i[\overline{\mathfrak{C}^*}.S]$ is not included in any $e'_j[\mathcal{T}^{\Box}(X)]$, $1 \leq j \leq n$, unless $\Box \in e'_j$ as well. Indeed, assume that $\Box$ is not in e'_j. So $e'_j[\mathcal{T}^{\Box}(X)] = e'_j$ is included in args $\mathfrak{C}'$. If $e_i[\overline{\mathfrak{C}^*}.S]$ were included in $e'_j[\mathcal{T}^{\Box}(X)]$, then $\overline{\mathfrak{C}^*}.S$, which is included in $e_i[\overline{\mathfrak{C}^*}.S]$ since $\Box \in e_i$, would be included in e'_j, hence in args $\mathfrak{C}'$. This would contradict the fact that $\overline{\mathfrak{C}^*}.S \not\subseteq \text{args } \mathfrak{C}' \cup S'$.

It follows that $(**)$ $e_i[\overline{\mathfrak{C}^*}.S] \subseteq e'_j[\mathcal{T}^{\Box}(X)]$ iff $e_i[\mathcal{T}^{\Box}(X)] \subseteq e'_j[\mathcal{T}^{\Box}(X)]$. The if direction is obvious. In the only if direction, we distinguish four cases. If $\Box \notin e_i$, then $e_i[\overline{\mathfrak{C}^*}.S] = e_i = e_i[\mathcal{T}^{\Box}(X)]$, and the claim is clear. If $e'_j = \{\Box\}^?$, then the assumption that $e_i[\overline{\mathfrak{C}^*}.S]$ is included in $e'_j[\mathcal{T}^{\Box}(X)]$ means that $e_i[\overline{\mathfrak{C}^*}.S]$ is a collection of sequences of terms of length at most 1; this is then certainly also the case for e_i, hence of $e_i[\mathcal{T}^{\Box}(X)]$, so $e_i[\mathcal{T}^{\Box}(X)] \subseteq e'_j[\mathcal{T}^{\Box}(X)]$. If $e'_j = F_j'^*$ where $\Box \in F'_j$, then $e'_j[\mathcal{T}^{\Box}(X)]$ is just $\mathcal{T}^{\Box}(X)^*$ (see Lemma 11.18), and the claim is obvious. Otherwise, $\Box$ is in e_i, and not in e'_j, so the claim follows from $(*)$.

The algorithmic characterization of inclusion of word-products given in Lemma 7.10 now allows us to conclude that $Q_i[\overline{\mathfrak{C}^*}.S] \subseteq Q'_j[\mathcal{T}^{\Box}(X)]$ if and only if $Q_i[\mathcal{T}^{\Box}(X)] \subseteq Q'_j[\mathcal{T}^{\Box}(X)]$. Concretely, this algorithmic characterization only depends on the answers to queries of the form $e_i[\overline{\mathfrak{C}^*}.S] \subseteq e'_j[\mathcal{T}^{\Box}(X)]$ in the first case, and on answers to the corresponding queries $e_i[\mathcal{T}^{\Box}(X)] \subseteq e'_j[\mathcal{T}^{\Box}(X)]$ in the second case. By $(**)$ these answers must be the same. $\Box$

E.6 Intersections of tree-products

Lemma 11.23 (recap). *Let X be a topological space. The intersection of two tree steps $P = C^?(\vec{P})$ and $P' = C'^?(\vec{P}')$ is equal to $\bigcup_{j=1}^n (C \cap C')^?(\vec{P}''_j) \cup (\text{supp}\vec{P} \cap P') \cup (P \cap \text{supp}\,\vec{P}')$, where $\vec{P} \cap \vec{P}'$ is expressed as a finite union $\bigcup_{j=1}^n \vec{P}''_j$ of word-products on $\mathcal{T}(X)$. If $C \cap C'$ can be written as the union of finitely many irreducible closed subsets C_i, $1 \leq i \leq m$, then $P \cap P'$ is also equal to the union of the tree steps $C_i^?(\vec{P}''_j)$ $(1 \leq i \leq n, 1 \leq j \leq m)$, of $\text{supp}\,\vec{P} \cap P'$, and of $P \cap \text{supp}\,\vec{P}'$.*

Proof. Let $t = f(\vec{t})$ be any term in $P \cap P'$. Since $t \in P$, t is in supp $\vec{P}$, or $f \in C$ and $\vec{t} \in \vec{P}$. In the first case, t is in supp $\vec{P} \cap P'$. Similarly, the claim that t' is in P' splits into two cases. The first one gives $t' \in P \cap \text{supp}\,\vec{P}'$. There remains the case where $f \in C$, $\vec{t} \in \vec{P}$, and $f \in C'$, $\vec{t} \in \vec{P}'$. Then f is in $C \cap C'$ (resp., in some C_i, if $C \cap C'$ can be written as a finite union of irreducible closed subsets C_i), and $\vec{t}$ is in some $\vec{P}''_j$, so t is in $(C \cap C')^?(\vec{P}''_j)$ (resp., in $C_i^?(\vec{P}''_j)$).

Conversely, supp $\vec{P} \cap P'$ and $P \cap \text{supp}\,\vec{P}'$ are included in $P \cap P'$. It remains to show that $(C \cap C')^?(\vec{P}''_j)$ (resp., $C_i^?(\vec{P}''_j)$) is included in $P \cap P'$, namely in both P and P'. We only deal with the first case. For every term t in $(C \cap C')^?(\vec{P}''_j)$ (resp., $C_i^?(\vec{P}''_j)$), either t is in supp $\vec{P}''_j$ or $t = f(\vec{t})$ with $f \in C \cap C'$ (resp., $f \in C_i$) and $\vec{t} \in \vec{P}''_j$. In the first case, the one-element word t is in $\vec{P}''_j$, hence in $\vec{P}$, so t is in supp $\vec{P}$ and therefore in $P = C^?(\vec{P})$. In the second case, f is in C, $\vec{P}$ is in $\vec{P}$, so $t = f(\vec{t})$ is in $P = C^?(\vec{P})$. $\Box$

Let us write supp $Q[S]$ for supp $(Q[S])$.

Lemma E.13. *Let X be a Noetherian space, Q be a word-product over $\mathcal{T}(X) + \{\Box\}$, and S be a closed subset of $\mathcal{T}(X)$. Then $\text{supp}\, Q[S] \subseteq \text{supp}\, Q \cup S$. (See Lemma 11.18 for $Q[S]$.)*

Proof. We first claim that for every atomic expression e, $\text{supp}\, e[S] \subseteq \text{supp}\, e \cup S$. If $e = \{\Box\}^?$, then $\text{supp}\, e[S] = \text{supp}\, S^? = S$. If $e = I^?$ where I is irreducible closed in $\mathcal{T}(X)$, then $\text{supp}\, e[S] = \text{supp}\, I^? = I = \text{supp}\, e$. If $e = \mathcal{F}^*$ and $\Box \in \mathcal{F}$, then $\text{supp}\, e[S] = \text{supp}\, ((\mathcal{F} \smallsetminus \{\Box\}) \cup S)^* = (\mathcal{F} \smallsetminus \{\Box\}) \cup S \subseteq \mathcal{F} \cup S = \text{supp}\, e \cup S$. If $e = \mathcal{F}^*$ and $\Box \notin \mathcal{F}$, then $\text{supp}\, e[S] = \text{supp}\, \mathcal{F}^* = \mathcal{F} = \text{supp}\, e$.

Write Q as $e_1 e_2 \cdots e_n$, where each e_i is an atomic expression. Then, $\text{supp}\, Q[S] = \bigcup_{i=1}^n \text{supp}\, e_i[S] \subseteq \bigcup_{i=1}^n (\text{supp} e_i \cup S) \subseteq \bigcup_{i=1}^n \text{supp}\, e_i \cup S = \text{supp}\, Q \cup S$. ☐

Lemma 11.24 (recap). *Let X be a Noetherian space, and S be a closed subset of $\mathcal{T}(X)$. Let $C^{\bar{?}}(\vec{P})$ be a tree step, $\mathfrak{C}$ be a closed subset of $X \times (\mathcal{T}(X) + \{\Box\})^*$ of the form $\bigcup_{j=1}^n C_j \times Q_j$, where each C_j is irreducible closed in X and each Q_j is a word-product over $\mathcal{T}(X) + \{\Box\}$.*

The intersection of the tree step $P = C^{\bar{?}}(\vec{P})$ and of the tree iterator $P' = \mathfrak{C}^{\bar{*}}.S$ is the union of $\text{supp}\, \vec{P} \cap P'$, of $P \cap (S \cup \text{args}\, \mathfrak{C})$, and of $(C \cap C_j)^{\bar{?}}(\vec{P} \cap Q_j[P'])$, $1 \le j \le n$.

If, for each j, one can write $C \cap C_j$ as the union of finitely many irreducible subsets C_{ij}, $1 \le i \le m_j$, and if $\vec{P} \cap Q_j[P']$ can be expressed as the union of finitely many word-products $\vec{P}_{\ell j}$, $1 \le \ell \le q_j$, then $P \cap P'$ is also equal to the union of $\text{supp}\, \vec{P} \cap P'$, of $P \cap (S \cup \text{args}\, \mathfrak{C})$, and of $C_{ij}^{\bar{?}}(\vec{P}_{\ell j})$, $1 \le j \le n$, $1 \le i \le m_j$, $1 \le \ell \le q_j$.

Proof. Let $t = f(\vec{t})$ be any term in $P \cap P'$. Since $t \in P$, t is in $\text{supp}\, \vec{P}$, or $f \in C$ and $\vec{t} \in \vec{P}$. In the first case, t is in $\text{supp}\, \vec{P} \cap P'$. Similarly, since $t \in P'$, t is in $S \cup \text{args}\, \mathfrak{C}$, or there is an elementary context $c \in \mathfrak{C}$ such that t is in $c[P']$. In the first of these cases, t is in $P \cap (S \cup \text{args}\, \mathfrak{C})$. There remains the case where $f \in C$, $\vec{t} \in \vec{P}$, and t is in $c[P']$ for some elementary context c – necessarily of the form $f(\vec{u})$ – in $\mathfrak{C}$. In that case, $\vec{t}$ is obtained from $\vec{u}$ by replacing each occurrence of $\Box$ by possibly different terms from P', in other words, $(f, \vec{t})$ is in $\mathfrak{C}[P']$. By Lemma 11.18, f is in C_j and $\vec{t}$ is in $Q_j[P']$ for some j. It follows that f is in $C \cap C_j$ and $\vec{t}$ is in $\vec{P} \cap Q_j[P']$, so $t = f(\vec{t})$ is in $(C \cap C_j)^{\bar{?}}(\vec{P} \cap Q_j[P'])$. (Additionally, if $C \cap C_j$ can be written as $\bigcup_{i=1}^{m_j} C_{ij}$ and $\vec{P} \cap Q_j[P']$ can be written as $\bigcup_{\ell=1}^{q_j} \vec{P}_{\ell j}$, then t is in $C_{ij}^{\bar{?}}(\vec{P}_{\ell j})$ for some j, i, and ℓ, too.)

For the converse inclusions, we check that:

— $\text{supp}\, \vec{P} \cap P' \subseteq P \cap P'$: every term in $\text{supp}\, \vec{P} \cap P'$ is in $\text{supp}\, \vec{P}$ hence in $P = C^{\bar{?}}(\vec{P})$ and is also in P';

— $P \cap (S \cup \text{args}\, \mathfrak{C}) \subseteq P \cap P'$: every term in $P \cap (S \cup \text{args}\, \mathfrak{C})$ is in $S \cup \text{args}\, \mathfrak{C}$, hence in $P' = \mathfrak{C}^{\bar{*}}.S$, and also in P; 1

— tem $(C \cap C_j)^{\bar{?}}(\vec{P} \cap Q_j[P'])$ (resp., $C_{ij}^{\bar{?}}(\vec{P}_{\ell j})$) is included in $P \cap P'$.

Every term t in $\text{supp}\, (\vec{P} \cap Q_j[P'])$ (resp., $\text{supp}\, \vec{P}_{\ell j}$) is such that the one-element word t is in $\vec{P} \cap Q_j[P']$ (resp., $\vec{P}_{\ell j}$, hence also in $\vec{P} \cap Q_j[P']$). Since that one-element word is in $\vec{P}$, t is in $\text{supp}\, \vec{P}$ hence in $P = C^{\bar{?}}(\vec{P})$, and since it is also in $Q_j[P']$, t is in $\text{supp}\, Q_j[P']$, hence in $\text{supp}\, Q_j \cup P'$ by Lemma E.13, hence in $\text{args}\, \mathfrak{C}' \cup P' \subseteq P'$.

Next, let t be any term of the form $f(\vec{t})$ with $f \in C \cap C_j$ and $\vec{t} \in \vec{P} \cap Q_j[P']$. (If instead $f \in C_{ij}$ and $\vec{t} \in \vec{P}_{\ell j}$, then f is in $C \cap C_j$ and $\vec{t}$ is in $P \cap Q_j[P']$.) Since f is in C and $\vec{t}$ is in $\vec{P}$, $t = f(\vec{t})$ is in $P = C^{\bar{?}}(\vec{P})$. Since f is in C_j and $\vec{t}$ is in $Q_j[P']$, by Lemma 11.18 $(f, \vec{t})$ is in $\mathfrak{C}[P']$, so $t = f(\vec{t})$ is in $P' = \mathfrak{C}^{\bar{*}}.S$. ☐

Lemma 11.26 (recap). *Let X be a Noetherian space and S and S' be closed subsets of $\mathcal{T}(X)$. Let also $\mathfrak{C}$ (resp., $\mathfrak{C}'$) be a closed subset of $X \times (\mathcal{T}(X) + \{\Box\})^*$ of the form $\bigcup_{i=1}^m C_i \times Q_i$ (resp., $\bigcup_{j=1}^n C'_j \times Q'_j$),*

https://doi.org/10.1017/S0960129520000195 Published online by Cambridge University Press

where each C_i and each C'_j is irreducible closed in X and each Q_i and each Q'_j is a normalized word-product over $\mathcal{T}(X)+\{\Box\}$. For all i, j, write $C_i \cap C_j$ as $\bigcup_{k=1}^{p_{ij}} C''_{ijk}$ where each C''_{ijk} is irreducible closed in X, and let $Q''_{ij\ell}$, $1 \leq \ell \leq q_{ij}$ enumerate the elements of $Meet^{\mathcal{E}}(Q_i, Q'_j)$, where the oracle $\mathcal{E}$ is defined in Lemma 11.25.

Then the intersection of the tree iterators $P = \overline{\mathfrak{C}^*}.S$ and $P' = \overline{\mathfrak{C}'^*}.S'$ is the tree iterator $\overline{\mathfrak{C}''^*}.S''$, where $\mathfrak{C}'' = \bigcup_{i,j,k,\ell} C''_{ijk} \times Q''_{ij\ell}$ and where S'' is the union of $P \cap (args\, \mathfrak{C}' \cup S')$ and of $(args\, \mathfrak{C} \cup S) \cap P'$.

Proof. Let $t = f(\vec{t})$ be in $P \cap P'$, where $\vec{t} = t_1 \cdots t_N$. We show that t is in $\overline{\mathfrak{C}''^*}.S''$ by induction on the size of t. If t is in args $\mathfrak{C} \cup S$, then t is in (args $\mathfrak{C} \cup S) \cap P'$, hence in S'', hence in $\overline{\mathfrak{C}''^*}.S''$. Similarly if t is in args $\mathfrak{C}' \cup S'$. In the remaining case, there is an elementary context $(f, \vec{u})$ in $\mathfrak{C}$ such that $\vec{t}$ is obtained by replacing the occurrences of $\Box$ in $\vec{u}$ by possibly different terms from P. Hence, there is an index i, $1 \leq i \leq m$, such that f is in C_i and $\vec{u}$ is in Q_i, so $\vec{t}$ is in $Q_i[P]$. Similarly, there is an index j, $1 \leq j \leq n$, such that $f \in C'_j$ and $\vec{t}$ is in $Q'_j[P']$. Since f is in $C_i \cap C'_j$, f is in C''_{ijk} for some k, $1 \leq k \leq p_{ij}$. By Lemma 11.25, $\vec{t}$ is in some element of $Meet^{\mathcal{E}}(Q_i, Q'_j)[P \cap P']$, hence in $Q''_{ij\ell}[P \cap P']$ for some ℓ, $1 \leq \ell \leq q_{ij}$. Since every term t_k in the list $\vec{t}$ is strictly smaller than t, every t_k that is in $P \cap P'$ is also in $\overline{\mathfrak{C}''^*}.S''$, by induction hypothesis. It follows that $\vec{t}$ is in $Q''_{ij\ell}[\overline{\mathfrak{C}''^*}.S'']$. Therefore, $(f, \vec{t}) \in C''_{ijk} \times Q''_{ij\ell}[\overline{\mathfrak{C}''^*}.S''] \subseteq \mathfrak{C}''[\overline{\mathfrak{C}''^*}.S''] \subseteq \overline{\mathfrak{C}''^*}.S''$.

This shows that $P \cap P' \subseteq \overline{\mathfrak{C}''^*}.S''$. Conversely, let $t = f(\vec{t})$ be any term in $\overline{\mathfrak{C}''^*}.S''$. We show that t is in $P \cap P'$ by induction on the size of t.

If t is in $P \cap$ (args $\mathfrak{C}' \cup S'$), then t is in P, and in args $\mathfrak{C}' \cup S' \subseteq \overline{\mathfrak{C}'^*}.S' = P'$. If t is in (args $\mathfrak{C} \cup S) \cap P'$, then t is in $P \cap P'$ by a symmetric argument. This shows that if t is in S'', then t is in $P \cap P'$.

If t is in args $\mathfrak{C}''$, then by Lemma 11.8, args $\mathfrak{C}'' = \bigcup_{i,j,k,\ell}$ supp $Q''_{ij\ell} \cap \mathcal{T}(X)$, so t is in supp $Q''_{ij\ell}$ for some i, j and ℓ. By the last part of Lemma 11.25, t is in supp $Q_i \cap$ supp Q'_j, or in supp $Q_i \cap P'$, or in $P \cap$ supp Q'_j (or in $\{\Box\}$, but that is impossible since $t \in \mathcal{T}(X)$). Hence, using Lemma 11.8 again, t is in args $\mathfrak{C} \cap$ args $\mathfrak{C}'$, or in args $\mathfrak{C} \cap P'$, or in $P \cap$ args $\mathfrak{C}'$. In any case, t is in $P \cap P'$.

We have proved that if t is in args $\mathfrak{C}'' \cup S''$, then t is in $P \cap P'$. The other possibility for t to be in $\overline{\mathfrak{C}''^*}.S''$ is for f to be in C''_{ijk} and $\vec{t}$ to be obtained from some elementary context $\vec{u}$ in $Q''_{ij\ell}$ (for some i, j, k, ℓ) by replacing the occurrences of $\Box$ by possibly different terms from $\overline{\mathfrak{C}''^*}.S''$ – hence from $P \cap P'$, by induction hypothesis. Then, f is in both C_i and C'_j, while $\vec{t}$ is in $Q''_{ij\ell}[P \cap P']$, hence in $Q_i[P] \cap Q'_j[P']$ by Lemma 11.25. This shows that $t = f(\vec{t})$ is both in $\mathfrak{C}[P] = \mathfrak{C}[\overline{\mathfrak{C}^*}.S] \subseteq \overline{\mathfrak{C}^*}.S = P$ and in $\mathfrak{C}'[P'] \subseteq P'$. $\Box$

Cite this article: Finkel A and Goubault-Larrecq J (2020). Forward analysis for WSTS, part I: completions. *Mathematical Structures in Computer Science* **30**, 752–832. https://doi.org/10.1017/S0960129520000195