Submitted to the Symposium on Theoretical Aspects of Computer Science
www.stacs-conf.org

FORWARD ANALYSIS FOR WSTS, PART I: COMPLETIONS

ALAIN FINKEL [1] AND JEAN GOUBAULT-LARRECQ [1,2]

[1] LSV, ENS Cachan, CNRS; 61 avenue du président Wilson, F-94230 Cachan

[2] INRIA Saclay Ile-de-France
E-mail address: {finkel,goubault}@lsv.ens-cachan.fr

ABSTRACT. Well-structured transition systems provide the right foundation to compute a finite basis of the set of predecessors of the upward closure of a state. The dual problem, to compute a finite representation of the set of successors of the downward closure of a state, is harder: Until now, the theoretical framework for manipulating downward-closed sets was missing. We answer this problem, using insights from domain theory (dcpos and ideal completions), from topology (sobrifications), and shed new light on the notion of adequate domains of limits.

1. Introduction

The theory of well-structured transition systems (WSTS) is 20 years old [9, 10, 2]. The most often used result of this theory [10] is the backward algorithm for computing a finite basis of the set $\uparrow Pre^*(\uparrow s)$ of predecessors of the upward closure $\uparrow s$ of a state s. The starting point of this paper is our desire to compute $\downarrow Post^*(\downarrow s)$ in a similar way. We then need a theory to finitely (and effectively) represent downward-closed sets, much as upward-closed subsets can be represented by their finite sets of minimal elements. This will serve as a basis for constructing forward procedures.

The *cover*, $\downarrow Post^*(\downarrow s)$, contains more information than the set of predecessors $\uparrow Pre^*(\uparrow s)$ because it characterizes a good approximation of the reachability set, while the set of predecessors describes the states from which the system may fail; the cover may also allow the computation of a finite-state abstraction of the system as a symbolic graph. Moreover, the backward algorithm needs a finite basis of the upward closed set of bad states, and its implementation is, in general, less efficient than a forward procedure: e.g., for lossy channel systems, although the backward procedure always terminates, only the non-terminating forward procedure is implemented in the tool TREX [1].

Except for some partial results [9, 7, 12], a general theory of downward-closed sets is missing. This may explain the scarcity of forward algorithms for WSTS. Quoting Abdulla *et al.* [3]: "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 [2]. This would give a general theory for forward analysis of infinite state systems, in the same way the work in [2] is for backward analysis." Our contribution is to provide such a theory of downward-closed sets.

Key words and phrases: WSTS, forward analysis, completion, Karp-Miller procedure, domain theory, sober spaces, Noetherian spaces.

SYMPOSIUM ON THEORETICAL ASPECTS OF COMPUTER SCIENCE

Related Work. Karp and Miller [15] proposed an algorithm that computes a finite representation of the downward closure of the reachability set of a Petri net. Finkel [9] introduced the WSTS framework and generalized the Karp-Miller procedure to a class of WSTS. This is done by constructing the completion of the set of states (by ideals, see Section 3) and in replacing the ω-acceleration of an increasing sequence of states (in Petri nets) by its least upper bound (lub). However, there are no effective finite representations of downward closed sets in [9]. Emerson and Namjoshi [7] considered a variant of WSTS (using cpos, but still without a theory of effective finite representations of downward-closed subsets) for defining a Karp-Miller procedure to broadcast protocols—termination is then not guaranteed [8]. Abdulla *et al.* [1] proposed a forward procedure for lossy channel systems using downward-closed languages, coded as SREs. Ganty, Geeraerts, and others [12, 11] proposed a forward procedure for solving the coverability problem for WSTS equipped with an effective adequate domain of limits. This domain ensures that every downward closed set has a finite representation; but no insight is given how these domains can be found or constructed. They applied this to Petri nets and lossy channel systems. Abdulla *et al.* [3] proposed another symbolic framework for dealing with downward closed sets for timed Petri nets.

We shall see that these constructions are special cases of our completions (Section 3). We shall illustrate this in Section 4, and generalize to a comprehensive hierarchy of data types in Section 5. We briefly touch the question of computing approximations of the cover in Section 6, although we shall postpone most of it to future work. We conclude in Section 7.

2. Preliminaries

We shall borrow from theories of order, both from the theory of well quasi-orderings, as used classically in well-structured transition systems [2, 10], and from domain theory [5, 13]. We should warn the reader that this is one bulky section on preliminaries. We invite her to skip technical points first, returning to them on demand.

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

A quasi-ordering is *well-founded* iff it has no infinite strictly descending chain, i.e., $x_0 > x_1 > \ldots > x_i > \ldots$. 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 well quasi-orderings (wqo). One is that, from any infinite sequence $x_0, x_1, \ldots, x_i, \ldots$, one can extract an infinite ascending chain $x_{i_0} \leq x_{i_1} \leq \ldots \leq x_{i_k} \leq \ldots$, with $i_0 < i_1 < \ldots < i_k < \ldots$. Another one is that any upward closed subset can be written $\uparrow E$, with E *finite*. Yet another, topological definition [14, Proposition 3.1] is to say that X, with its Alexandroff topology, is Noetherian. The *Alexandroff topology* on X is that whose opens are exactly the upward closed subsets. A subset K is compact if it satisfies the Heine-Borel property, i.e., every one may extract a finite subcover from any open cover of K. A topology is *Noetherian* iff every open subset is compact, iff any increasing chain of opens stabilizes [14, Proposition 3.2]. We shall cite results from the latter paper as the need evolves.

We shall be interested in rather particular topological spaces, whose topology arises from order. A *directed family* of 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 $x_i, x_j \leq x_k$. The *Scott topology* on X has as opens all upward closed subsets U such that every directed family $(x_i)_{i \in I}$ that has a least upper bound x in X intersects U, i.e., $x_i \in U$ for some $i \in I$. The Scott topology is coarser than the Alexandroff topology, i.e., every Scott-open is Alexandroff-open (upward closed); the converse fails in general. The Scott topology is particularly interesting on *dcpos*, i.e., posets X in which every directed family $(x_i)_{i \in I}$ has a least upper bound $\sup_{i \in I} x_i$.

The *way below* relation $\ll$ on a poset X is defined 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. 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 $\twoheaduparrow E = \{y \in X \mid \exists x \in E \cdot x \ll y\}$, $\twoheaddownarrow E = \{y \in X \mid \exists x \in E \cdot y \ll x\}$. X is *continuous* iff, for every $x \in X$, $\twoheaddownarrow x$ is a directed family, and has x as least upper bound. One may be more precise: A *basis* is a 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. In a continuous dcpo, $\twoheaduparrow x$ is Scott-open for all x, and every Scott-open set U is a union of such sets, viz. $U = \bigcup_{x \in U} \twoheaduparrow x$ [5].

X is *algebraic* iff every element x is the least upper bound of the set of finite elements below x—an element y is *finite* if and only if $y \ll y$. Every algebraic poset is continuous, and has a least basis, namely its set of finite elements.

$\mathbb{N}$, with its natural ordering, is a wqo and an algebraic poset. All its elements are finite, so $x \ll y$ iff $x \leq y$. $\mathbb{N}$ is not a dcpo, since $\mathbb{N}$ itself is a directed family without a least upper bound. Any finite product of continuous posets (resp., continuous dcpos) is again continuous, and the Scott-topology on the product coincides with the product topology. Any finite product of wqos is a wqo. In particular, $\mathbb{N}^k$, for any integer k, is a wqo and a continuous poset: this is the set of configurations of Petri nets.

It is clear how to complete $\mathbb{N}$ to make it a cpo: let $\mathbb{N}_\omega$ be $\mathbb{N}$ with a new element ω such that $n \leq \omega$ for all $n \in \mathbb{N}$. Then $\mathbb{N}_\omega$ is still a wqo, and a continuous cpo, with $x \ll y$ if and only if $x \in \mathbb{N}$ and $x \leq y$. In general, completing a wqo is necessary to extend coverability tree techniques [9, 12]. Geeraerts *et al.* (op. cit.) axiomatize the kind of completions they need in the form of so-called *adequate domains of limits*. We discuss them in Section 3. For now, let us note that the second author also proposed to use another notion of completion in another context, known as *sobrification* [14]. We need to recap what this is about.

A topological space X is always equipped with a *specialization quasi-ordering*, which we shall write $\leq$ again: $x \leq y$ if and only if any open subset containing x also contains y. X is T_0 if and only if $\leq$ is a partial ordering. Given any quasi-ordering $\leq$ on a set X, both the Alexandroff and the Scott topologies admit $\leq$ as specialization quasi-ordering. In fact, the Alexandroff topology is the finest (the one with the most opens) 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. The latter sets $\downarrow E$, with E finite, will play an important role, and we call them the *finitary closed* subsets. Note that finitary closed subsets are closed in the upper, Scott, and Alexandroff topologies, recalling that a subset is *closed* iff its complement is open. The *closure* $cl(A)$ of a subset A of X is the smallest closed subset containing A. A closed subset F is *irreducible* if and only if F is non-empty, and whenever $F \subseteq F_1 \cup F_2$ with F_1, F_2 closed, then $F \subseteq F_1$ or $F \subseteq F_2$. The finitary closed subset $\downarrow x = cl(\{x\})$ ($x \in X$) is always irreducible. A space X is *sober* iff every irreducible closed subset F is the closure of a unique point, i.e., $F = \downarrow x$ for some unique x. Any sober space is T_0, and any continuous cpo is sober in its Scott topology. Conversely, given a T_0 space X, the space $\mathcal{S}(X)$

of all irreducible closed subsets of X, equipped with upper topology of the inclusion ordering $\subseteq$, is always sober, and the map $\eta_{\mathcal{S}} : x \mapsto \uparrow x$ is a topological embedding of X inside $\mathcal{S}(X)$. $\mathcal{S}(X)$ is the *sobrification* of X, and can be thought as X together with all missing limits from X. Note in particular that a sober space is always a cpo in its specialization ordering [5, Proposition 7.2.13].

It is an enlightening exercise to check that $\mathcal{S}(\mathbb{N})$ is $\mathbb{N}_\omega$. Also, the topology on $\mathcal{S}(\mathbb{N})$ (the upper topology) coincides with that of $\mathbb{N}_\omega$ (the Scott topology). In general, X is Noetherian if and only if $\mathcal{S}(X)$ is Noetherian [14, Proposition 6.2], however the upper and Scott topologies do not always coincide [14, Section 7]. In case of ambiguity, given any poset X, we write X_a the space X with its Alexandroff topology.

Another important construction is the *Hoare powerdomain* $\mathcal{H}(X)$ of X, whose elements are the closed subsets of X, ordered by inclusion. (We do allow the empty set.) We again equip it with the corresponding upper topology.

3. Completions of Wqos

One of the central problems of our study is the definition of a *completion* of a wqo X, with all missing limits added. Typically, the Karp-Miller construction [15] works not with $\mathbb{N}^k$, but with $\mathbb{N}_\omega^k$. We examine several ways to achieve this, and argue that they are the same, up to some details.

ADLs, WADLs. We start with Geeraerts *et al.*'s axiomatization of so-called *adequate domain of limits* for well-quasi-ordered sets X [12]. No explicit constructions for such adequate domains of limits is given, and they have to be found by trial and error. Our main result, below, is that there is a unique least adequate domain of limits: the *sobrification* $\mathcal{S}(X_a)$ of X_a. (Recall that X_a is X with its Alexandroff topology.) This not only gives a concrete construction of such an adequate domain of limits, but also shows that we do not have much freedom in defining one.

An *adequate domain of limits* [12] (ADL) for a well-ordered set X is a triple $(L, \preceq, \gamma)$ 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) $z \preceq z'$ 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 $\widehat{\gamma}(E) = \bigcup_{z \in E} \gamma(z)$.

Requirement (L_2) in [12] 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). Accordingly, we drop requirement (L_2):

Definition 3.1 (WADL)**.** Let X be a poset. A *weak adequate domain of limits* (WADL) on X is any triple $(L, \preceq, \gamma)$ satisfying (L_1), (L_3), and (L_4).

Proposition 3.2. *Let X be a poset. Given a WADL $(L, \preceq, \gamma)$ on X, γ defines an order-isomorphism from $(L \cup X, \preceq)$ to some subset of $\mathcal{H}(X_a)$ containing $\mathcal{S}(X_a)$.*

Conversely, assume X wqo, and let Y be any subset of $\mathcal{H}(X_a)$ containing $\mathcal{S}(X_a)$. Then $(Y \setminus \eta_{\mathcal{S}}(X_a), \preceq, \gamma)$ is a weak adequate domain of limits, where γ maps each $x \in X$ to $\downarrow_X x$ and each $F \in Y \setminus \eta_{\mathcal{S}}(X_a)$ to itself; $\preceq$ is defined by requirement (L_3).

Proof. The Alexandroff-closed subsets of X are just its downward-closed subsets. So $\gamma(z)$ is in $\mathcal{H}(X_a)$ for all z, by (L_1). Let Y be the image of γ. By (L_3), γ defines an order-isomorphism of $L \cup X$ onto Y. It remains to show that Y must contain $\mathcal{S}(X_a)$. Let F be any irreducible closed subset of X_a. By (L$_4$), there is a finite subset $E \subseteq L \cup X$ such that $F = \bigcup_{x \in E} \gamma(x)$. Since F is irreducible, there must be a single $x \in E$ such that $F = \gamma(x)$. So F is in Y.

Conversely, let X be wqo, $L = Y \setminus \eta_{\mathcal{S}}(X_a)$, and γ, $\preceq$ be as in the Lemma. Properties (L$_1$) and (L$_3$) hold by definition. For (L$_4$), note that X_a is a Noetherian space, hence $\mathcal{S}(X_a)$ is, too [14, Proposition 6.2]. However, by [14, Corollary 6.5], every closed subset of a sober Noetherian space is finitary. In particular, take any downward closed subset D of X. This is closed in X_a, hence its image $\eta_{\mathcal{S}}(D)$ by the topological embedding $\eta_{\mathcal{S}}$ is closed in $\eta_{\mathcal{S}}(X_a)$, i.e., is of the form $\eta_{\mathcal{S}}(X_a) \cap F$ for some closed subset F of $\mathcal{S}(X_a)$. Also, $D = \eta_{\mathcal{S}}^{-1}(F)$. Since $\mathcal{S}(X_a)$ is both sober and Noetherian, F is finitary, hence is the downward-closure $\downarrow_{\mathcal{S}(X)} E'$ of some finite subset E' in $\mathcal{S}(X)$. Let E be the set consisting of the (limit) elements in $E' \cap L$, and of the (non-limit) elements $x \in X$ such that $\downarrow_X x \in E'$. We obtain $\widehat{\gamma}(E) = \bigcup_{z \in E'} z$. On the other hand, $D = \eta_{\mathcal{S}}^{-1}(F) = \{x \in X \mid \downarrow x \in \downarrow_{\mathcal{S}(X)} E'\} = \{x \in X \mid \exists z \in E' \cdot \downarrow x \subseteq z\} = \bigcup_{z \in E'} z = \widehat{\gamma}(E)$. So (L$_4$) holds. ∎

I.e., up to the coding function γ, there is a unique *minimal* WADL on any given wqo X: its sobrification $\mathcal{S}(X_a)$. There is also a unique largest one: its Hoare powerdomain $\mathcal{H}(X_a)$. An adequate domain of limits in the sense of Geeraerts *et al.* [12], i.e., one that additionally satisfies (L$_2$) is, up to isomorphism, any subset of $\mathcal{H}(X_a)$ containing $\mathcal{S}(X_a)$ plus the special closed set X itself as top element. We contend that $\mathcal{S}(X_a)$ is, in general, the sole WADL worth considering.

Ideal completions. We have already argued that $\mathcal{S}(X)$, for any Noetherian space X, was in a sense of completion of X, adding missing limits. Another classical construction to add limits to some poset X is its *ideal completion* $Idl(X)$. The elements of the ideal completion of X are its *ideals*, i.e., its downward-closed directed families, ordered by inclusion. $Idl(X)$ can be visualized as a form of Cauchy completion of X: we add all missing limits of directed families $(x_i)_{i \in I}$ from X, by declaring these families to be their limits, equating two families when they have the same downward-closure. In $Idl(X)$, the finite elements are the elements of X; formally, the map $\eta_{Idl} : X \to Idl(X)$ that sends x to $\downarrow x$ is an embedding, and the finite elements of $Idl(X)$ are those of the form $\eta_{Idl}(x)$. It turns out that sobrification and ideal completion coincide, in a strong sense:

Proposition 3.3 ([16]). *For any poset X, $\mathcal{S}(X_a) = Idl(X)$.*

This is not just an isomorphism: the irreducible closed subsets of X_a are *exactly* the ideals. Note also that $Idl(X)$ is always an algebraic dcpo [5, Proposition 2.2.22, Item 4].

When X is wqo, any downward-closed subset of X is a *finite* union of ideals. So $(Idl(X) \setminus X, \subseteq, \mathrm{id})$ is a WADL on X. Proposition 3.2 and Proposition 3.3 entail this, and a bit more:

Theorem 3.4. *For any wqo X, $\mathcal{S}(X_a) = Idl(X)$ is the smallest WADL on X.*

Well-based continuous cpos. There is a natural notion of limit in dcpos: whenever $(x_i)_{i \in I}$ is a directed family, consider $\sup_{i \in I} x_i$. Starting from a wqo X, it is then natural to look at some dcpo Y that would contain X as a basis. In particular, Y would be continuous. This prompts us to define a *well-based continuous dcpo* as one that has a well-ordered basis—namely the original poset X.

This has several advantages. First, in general there are several notions of "sets of limits" of a given subset $A \subseteq Y$, but we shall see that they all coincide in continuous posets. Such sets of limits are important, because these are what we would like Karp-Miller-like procedures to compute, through acceleration techniques. Here are the possible notions. First, define $\mathrm{Lub}_Y(A)$ as the set of all least upper bounds in Y of directed families in A. Second, $\mathrm{Ind}_Y(A)$, the *inductive hull* of A in Y, is the smallest sub-dcpo of Y containing A. Finally, the (Scott-topological) closure $cl(A)$ of A. It is well-known that $cl(A)$ is the smallest *downward closed* sub-dcpo of Y containing A.

(Recall that any open is upward closed, so that any closed set must be downward closed.) In any dcpo Y, one has $A \subseteq \mathrm{Lub}_Y(A) \subseteq \mathrm{Ind}_Y(A) \subseteq cl(A)$, and all inclusions are strict in general. E.g., in $Y = \mathbb{N}_\omega$, take A to be the set of even numbers. Then $\mathrm{Lub}_Y(A) = \mathrm{Ind}_Y(A) = A \cup \{\omega\}$ while $cl(A) = \mathbb{N}_\omega$. While $\mathrm{Lub}_Y(A) = \mathrm{Ind}_Y(A)$ in this case, there are cases where $\mathrm{Lub}_Y(A)$ is itself not closed under least upper bounds of directed families, and one has to iterate the Lub_Y operator to compute $\mathrm{Ind}_Y(A)$. On continuous posets however, all these notions coincide (see Appendix A).

Proposition 3.5. *Let Y be a continuous poset. Then, for every downward-closed subset A of Y,* $\mathrm{Ind}_Y(A) = \mathrm{Lub}_Y(A) = cl(A)$.

We shall use this in Section 6. The key point now is that, again, well-based continuous dcpos coincide with completions of the form $\mathcal{S}(X_a)$ or $Idl(X)$, and are therefore WADLs (see Appendix B). This even holds for continuous dcpos having a well-founded (not well-ordered) basis:

Proposition 3.6. *Any continuous dcpo Y with a well-founded basis is order-isomorphic to $Idl(X)$ for some well-ordered set X. One may take the subset of finite elements of X for Y. If Y is well-based, then X is well-ordered.*

4. Some Concrete WADLs

We now build WADLs for several concrete posets X. Following Proposition 3.2, it suffices to characterize $\mathcal{S}(X_a)$. Although $\mathcal{S}(X_a) = Idl(X)$ (Proposition 3.3), the mathematics of $\mathcal{S}(X_a)$ is easier to deal with than $Idl(X)$.

$\mathbb{N}^k$. We start with $X = \mathbb{N}^k$, with the pointwise ordering. We have already recalled from [14] that $\mathcal{S}(\mathbb{N}^k_a)$ was, up to isomorphism, $(\mathbb{N}_\omega)^k$, ordered with the pointwise ordering, where ω is a new element above any natural number. This is the structure used in the standard Karp-Miller construction for Petri nets [15].

Σ^*. Let Σ be a finite alphabet. The *divisibility ordering* $\mid$ on Σ^*, a.k.a. the subsequence (non-continuous subword) ordering, is defined by $a_1 a_2 \ldots a_n \mid w_0 a_1 w_1 a_2 \ldots a_n w_n$, for any letters $a_1, a_2, \ldots, a_n \in \Sigma$ and words $w_0, w_1, \ldots, w_n \in \Sigma^*$. There is a more general definition, where letters themselves are quasi-well-ordered. Our definition is the special case where the wqo on letters is $=$, and is the one required in verifying lossy channel systems [4]. Higman's Lemma states that $\mid$ is wqo on Σ^*.

Any upward closed subset U of Σ^* is then of the form $\uparrow E$, with E finite. For any element $w = a_1 a_2 \ldots a_n$ of E, $\uparrow w$ is the regular language $\Sigma^* a_1 \Sigma^* a_2 \Sigma^* \ldots \Sigma^* a_n \Sigma^*$. Forward analysis of lossy channel systems is instead based on simple regular expressions (SREs). Recall from [1] that an *atomic expression* is any regular expression of the form $a^?$, with $a \in \Sigma$, or A^*, where A is a non-empty subset of Σ. When $A = \{a_1, \ldots, a_m\}$, we take A^* to denote $(a_1 + \ldots + a_m)^*$; $a^?$ denotes $\{a, \epsilon\}$. A *product* is any regular expression of the form $e_1 e_2 \ldots e_n$ ($n \in \mathbb{N}$), where each e_i is an atomic expression. A *simple regular expression*, or *SRE*, is a sum, either $\emptyset$ or $P_1 + \ldots + P_k$, where $P_1, \ldots, P_k$ are products. Sum is interpreted as union. That SREs and products are relevant here is no accident, as the following proposition shows.

Proposition 4.1. *The elements of $\mathcal{S}(\Sigma^*_a)$ are exactly the denotations of products. The downward closed subsets of Σ^* are exactly the denotations of SREs.*

Proof. The second part is well-known. If $F = P_1 + \ldots + P_k$ is irreducible closed, then by irreducibility k must equal 1, hence F is denoted by a product. Conversely, it is easy to show that any product denotes an ideal, hence an element of $Idl(X) = \mathcal{S}(X_a)$ (Proposition 3.3). ∎

Inclusion between products can then be checked in quadratic time [1]. Inclusion between SREs can be checked in polynomial time, too, because of the remarkable property that $P_1 + \ldots + P_m \subseteq P'_1 + \ldots + P'_n$ if and only if, for every i $(1 \leq i \leq m)$, there is a j $(1 \leq j \leq n)$ with $P_i \subseteq P'_j$ [1, Lemma 1].Similar lemmas are given by Abdulla *et al.* [3, Lemma 3, Lemma 4] for more general notions of SREs on words on infinite alphabets, and for a similar notion for finite multisets of elements from a finite set (both will be special cases of our constructions of Section 5). This is again no accident, and is a general fact about Noetherian spaces:

Proposition 4.2. *Let X be a Noetherian space, e.g., a wqo with its Alexandroff topology. Every closed subset F of X is a finite union of irreducible closed subsets $C_1, \ldots, C_m$. If $C'_1, \ldots, C'_n$ are also irreducible closed, Then $C_1 \cup \ldots \cup C_m \subseteq C'_1 \cup \ldots \cup C'_n$ if and only if for every i ($1 \leq i \leq m$), there is a j ($1 \leq j \leq n$) with $C_i \subseteq C'_j$.*

Proof. For the first part, see Appendix C. The second part is an easy consequence of irreducibility. ∎

Proposition 4.2 suggests to represent closed subsets of X as finite subsets A of $\mathcal{S}(X)$, interpreted as the closed set $\bigcup_{C \in A} C$. When $X = \Sigma_a^*$, A is a finite set of products, i.e., an SRE. When $X = \mathbb{N}_a^k$, A is a finite subset of $\mathbb{N}_\omega^k$, interpreted as $\downarrow A \cap \mathbb{N}^k$.

Finite Trees. All the examples given above are well-known. Here is one that is new, and also more involved than the previous ones. Let $\mathcal{F}$ be a finite signature of function symbols with their arities. We let $\mathcal{F}_k$ the set of function symbols of arity k; $\mathcal{F}_0$ is the set of *constants*, and is assumed to be non-empty. The set $\mathcal{T}(\mathcal{F})$ is the set of ground terms built from $\mathcal{F}$. Kruskal's Tree Theorem states that this is well-quasi-ordered by the *homeomorphic embedding* ordering $\trianglelefteq$, defined as the smallest relation such that, whenever $u = f(u_1, \ldots, u_m)$ and $v = g(v_1, \ldots, v_n)$, $u \trianglelefteq v$ if and only if $u \trianglelefteq v_j$ for some j, $1 \leq j \leq n$, or $f = g$, $m = n$, and $u_1 \trianglelefteq v_1$, $u_2 \trianglelefteq v_2$, …, $u_m \trianglelefteq v_m$. (As for Σ^*, we take a special case, where each function has fixed arity.)

The structure of $\mathcal{S}(\mathcal{T}(\mathcal{F})_a)$ is described using an extension of SREs to the tree case. This uses regular tree expressions as defined in [6, Section 2.2]. Let $\mathcal{K}$ be a countably infinite set of additional constants, called *holes* $\Box$. Most tree regular expressions are self-explanatory, except Kleene star $L^{*,\Box}$ and concatenation $L.{}_\Box L'$. The latter denotes the set of all terms obtained from a term t in L by replacing all occurrences of $\Box$ by (possibly different) terms from L'. The language of a hole $\Box$ is just $\{\Box\}$. $L^{*,\Box}$ is the infinite union of the languages of $\Box$, L, $L.{}_\Box L$, $L.{}_\Box L.{}_\Box L$, etc.

Definition 4.3 (STRE)**.** *Tree products* and *product iterators* are defined inductively by:

- Every hole $\Box$ is a tree product.
- $f^?(P_1, \ldots, P_k)$ is a tree product, for any $f \in \Sigma_k$ and any tree products $P_1, \ldots, P_k$. We take $f^?(P_1, \ldots, P_k)$ as an abbreviation for $f(P_1, \ldots, P_k) + P_1 + \ldots + P_k$.
- $(\sum_{i=1}^n C_i)^{*,\Box}.{}_\Box P$ is a tree product, for any tree product P, any $n \geq 1$, and any product iterators C_i over $\Box$, $1 \leq i \leq n$. We write $\sum_{i=1}^n C_i$ for $C_1 + C_2 + \ldots + C_n$.
- $f(P_1, \ldots, P_k)$ is a product iterator over $\Box$ for any $f \in \Sigma_k$, where: 1. each P_i, $1 \leq i \leq k$ is either $\Box$ itself or a tree product such that $\Box$ is not in the language of P_i; and 2. $P_i = \Box$ for some i, $1 \leq i \leq k$.

A *simple tree regular expression* (STRE) is a finite sum of tree products.

A tree regular expression is *closed* iff it has no free hole, where a hole is free in $f(L_1, \ldots, L_k)$, $L_1 + \ldots + L_k$, or in $f^?(L_1, \ldots, L_k)$ iff it is free in some L_i, $1 \leq i \leq k$; the only free hole in $\Box$ is $\Box$ itself; the free holes of $L^{*,\Box}$ are those of L, plus $\Box$; the free holes of $L.{}_\Box L'$ are those of L', plus those of L except $\Box$. E.g., $f^?(a^?, b^?)$ and $(f(\Box, g^?(a^?)) + f(g^?(b^?), \Box))^{*,\Box}.{}_\Box f^?(a^?, b^?)$ are closed tree products. We prove the following in Appendix D.

Theorem 4.4. *The elements of $\mathcal{S}(\mathcal{T}(\mathcal{F})_a)$ are exactly the denotations of closed tree products. The downward closed subsets of $\mathcal{T}(\mathcal{F})$ are exactly the denotations of closed STREs. Inclusion is decidable in polynomial time for tree products and for STREs.*

5. A Hierarchy of Data Types

The sobrification WADL can be computed in a compositional way, as we now show. Consider the following grammar of data types of interest in verification:

D	$::=$	$\mathbb{N}$	natural numbers
	$\mid$	$A_{\leq}$	finite set A, quasi-ordered by $\leq$
	$\mid$	$D_1 \times \ldots \times D_k$	finite product
	$\mid$	$D_1 + \ldots + D_k$	finite, disjoint sum
	$\mid$	D^*	finite words
	$\mid$	$D^{\circledast}$	finite multisets

By *compositional*, we mean that the sobrification of any data type D is computed in terms of the sobrifications of its arguments. E.g., $\mathcal{S}(D_a^*)$ will be expressed as some extended form of products over $\mathcal{S}(D_a)$. The semantics of data types is the intuitive one. Finite products are quasi-ordered by the pointwise quasi-ordering, finite disjoint sums by comparing elements in each summand—elements from different summands are incomparable. For any poset X (even infinite), X^* is the set of finite words over X ordered by the *embedding* quasi-ordering $\leq^*$: $w \leq^* w'$ iff, writing w as the sequence of m letters $a_1 a_2 \ldots a_m$, one can write w' as $w_0 a'_1 w_1 a'_2 w_2 \ldots w_{m-1} a'_m w'_m$ with $a_1 \leq a'_1$, $a_2 \leq a'_2$, ..., $a_m \leq a'_m$. $X^{\circledast}$ is the set of finite multisets $\{|x_1, \ldots, x_n|\}$ of elements of X, and is quasi-ordered by $\leq^{\circledast}$, defined as: $\{|x_1, x_2, \ldots, x_m|\} \leq^{\circledast} \{|y_1, y_2, \ldots, y_n|\}$ iff there is an injective map $r : \{1, \ldots, m\} \to \{1, \ldots, n\}$ such that $x_i \leq y_{r(i)}$ for all i, $1 \leq i \leq m$. When $\leq$ is just equality, $m \leq^{\circledast} m'$ iff every element of m occurs at least as many times in m' as in m: this is the $\leq^m$ quasi-ordering considered, on finite sets X, by Abdulla *et al.* [3, Section 2].

The analogue of products and SREs for D^* is given by the following definition, which generalizes the Σ^* case of Section 4. Note that D is in general an *infinite* alphabet, as in [3]. The following definition should be compared with [1]. The only meaningful difference is the replacement of $(a + \epsilon)$, where a is a letter, with $C^?$, where $C \in \mathcal{S}(X_a)$. It should also be compared with the *word language generators* of [3, Section 6]. Indeed, the latter are exactly our products on $A^{\circledast}$, where A is a finite alphabet (in our notation, $A_{\leq}$, with $\leq$ given as equality).

Definition 5.1 (Product, SRE)**.** Let X be a topological space. Let X^* be the set of finite words on X. For any $A, B \subseteq X^*$, let AB be $\{ww' \mid w \in A, w' \in B\}$, A^* be the set of words on A, $A^? = A \cup \{\epsilon\}$.

Atomic expressions are either of the form $C^?$, with $C \in \mathcal{S}(X)$, or A^*, with A a non-empty finite subset of $\mathcal{S}(X)$. *Products* are finite sequences $e_1 e_2 \ldots e_k$, $k \in \mathbb{N}$, and *SREs* are finite sums of products. The denotation of atomic expressions is given by $[\![C^?]\!] = C^?$, $[\![A^*]\!] = (\bigcup_{C \in A} [\![C]\!])^*$; of products by $[\![e_1 e_2 \ldots e_k]\!] = [\![e_1]\!]\, [\![e_2]\!] \ldots [\![e_k]\!]$; of SREs by $[\![P_1 + \ldots + P_k]\!] = \bigcup_{i=1}^k [\![P_i]\!]$.

Atomic expressions are ordered by $C^? \sqsubseteq C'^?$ iff $C \subseteq C'$; $C^? \sqsubseteq A'^*$ iff $C \subseteq C'$ for some $C' \in A'$; $A^* \not\sqsubseteq C'^?$; $A^* \sqsubseteq A'^*$ iff for every $C \in A$, there is a $C' \in A'$ with $C \subseteq C'$. Products are quasi-ordered by $eP \sqsubseteq e'P'$ iff (1) $e \not\sqsubseteq e'$ and $eP \sqsubseteq P'$, or (2) $e = C^?$, $e' = C'^?$, $C \subseteq C'$ and $P \sqsubseteq P'$, or (3) $e' = A'^*$, $e \sqsubseteq A'^*$ and $P \sqsubseteq e'P'$. We let $\equiv$ be $\sqsubseteq \cap \sqsupseteq$.

Definition 5.2 ($\circledast$-Product, $\circledast$-SRE)**.** Let X be a topological space. For any $A, B \subseteq X$, let $A \odot B = \{m \uplus m' \mid m \in A, m' \in B\}$, $A^{\circledast}$ be the set of multisets comprised of elements from A, $A^{?} = \{\{|x|\} \mid x \in A\} \cup \{\emptyset\}$, where $\emptyset$ is the empty multiset.

The $\circledast$-*products* P are the expressions of the form $A^{\circledast} \odot C_1^{?} \odot \ldots \odot C_n^{?}$, where A is a finite subset of $\mathcal{S}(X)$, $n \in \mathbb{N}$, and $C_1, \ldots, C_n \in \mathcal{S}(X)$. Their denotation $[\![P]\!]$ is $(\bigcup_{C \in A} C)^{\circledast} \odot [\![C_1]\!]^{?} \odot \ldots \odot [\![C_n]\!]^{?}$. They are quasi-ordered by $P \sqsubseteq P'$, where $P = A^{\circledast} \odot C_1^{?} \odot C_2^{?} \odot \ldots \odot C_m^{?}$ and $P' = A'^{\circledast} \odot C_1'^{?} \odot C_2'^{?} \odot \ldots \odot C_n'^{?}$, iff: (1) for every $C \in A$, there is a $C' \in A'$ with $C \subseteq C'$, and (2) letting I be the subset of those indices i, $1 \leq i \leq m$, such that $C_i \subseteq C'$ for no $C' \in A'$, there is an injective map $r : I \to \{1, \ldots, n\}$ such that $C_i \subseteq C'_{r(i)}$ for all $i \in I$. Let $\equiv$ be $\sqsubseteq \cap \sqsupseteq$.

Theorem 5.3. *For every data type D, $\mathcal{S}(D_a)$ is Noetherian, and is computed by: $\mathcal{S}(\mathbb{N}_a) = \mathbb{N}_\omega$; $\mathcal{S}(A_{\leq_a}) = A_{\leq}$; $\mathcal{S}((D_1 \times \ldots \times D_k)_a) = \mathcal{S}(D_{1a}) \times \ldots \times \mathcal{S}(D_{ka})$; $\mathcal{S}((D_1 + \ldots + D_k)_a) = \mathcal{S}(D_{1a}) + \ldots + \mathcal{S}(D_{ka})$; $\mathcal{S}(D^*)$ is the set of products on D modulo $\equiv$, ordered by $\sqsubseteq$ (Definition 5.1); $\mathcal{S}(D^{\circledast})$ is the set of $\circledast$-products on D modulo $\equiv$, ordered by $\sqsubseteq$ (Definition 5.2).*

For any data type D, equality and ordering (inclusion) in $\mathcal{S}(D_a)$ is decidable in the polynomial hierarchy.

Proof. We show that $\mathcal{S}(D_a)$ is Noetherian and is computed as given above, by induction on the construction of D. We in fact prove the following two facts separately: (1) $\mathcal{S}(D)$ is Noetherian (D, not D_a), where D is topologized in a suitable way, and (2) $D = D_a$.

To show (1), we topologize $\mathbb{N}$ and $A_{\leq}$ with their Alexandroff topologies, sums and products with the sum and product topologies respectively; X^* with the *subword topology*, viz. the smallest containing the open subsets $X^* U_1 X^* U_2 X^* \ldots X^* U_n X^*$, $n \in \mathbb{N}$, $U_1, U_2, \ldots, U_n$ open in X; and $X^{\circledast}$ with the *sub-multiset topology*, namely the smallest containing the subsets $X^{\circledast} \odot U_1 \odot U_2 \odot \ldots \odot U_n$, $n \in \mathbb{N}$, where $U_1, U_2, \ldots, U_n$ are open subsets of X. The case of $\mathbb{N}$ has already been discussed above. When $A_{\leq}$ is finite, it is both Noetherian and sober. The case of finite products is by [14, Section 6], that of finite sums by [14, Section 4]. The case of X^* is dealt with in Appendix E, while the case of $X^{\circledast}$ is dealt with in Appendix F. We also need to show that the quasi-orderings $\sqsubseteq$ on products in X^*, resp. $\circledast$-products in $X^{\circledast}$, denote inclusion in $\mathcal{S}(X^*)$, resp. $\mathcal{S}(X^{\circledast})$. This is also done in the appendices.

To show (2), we appeal to a series of coincidence lemmas, showing that $(X^*)_a = X_a^*$ (Lemma E.4) and that $(X^{\circledast})_a = X_a^{\circledast}$ (Lemma F.10) notably. The other cases are obvious.

Finally, we show that inclusion and equality are decidable in the polynomial hierarchy. For this, we show in the appendices that inclusion on $\mathcal{S}(D^*)$ is $\sqsubseteq$ on products, and is decidable by a polynomial time algorithm modulo calls to an oracle deciding inclusion in $\mathcal{S}(D)$. This is by dynamic programming. Inclusion in $\mathcal{S}(D^{\circledast})$ is $\sqsubseteq$ on $\circledast$-products, and is decidable by a non-deterministic polynomial time algorithm modulo a similar oracle. We conclude since the orderings on $\mathbb{N}_\omega$ and on $A_{\leq}$ are polynomial-time decidable, while inclusion in $\mathcal{S}(D_1 \times \ldots \times D_k) \cong \mathcal{S}(D_1) \times \ldots \times \mathcal{S}(D_k)$ and in $\mathcal{S}(D_1 + \ldots + D_k) \cong \mathcal{S}(D_1) + \ldots + \mathcal{S}(D_k)$ are polynomial time modulo oracles deciding inclusion in $\mathcal{S}(D_i)$, $1 \leq i \leq k$. ∎

Look at some special cases of this construction. First, $\mathbb{N}^k$ is the data type $\mathbb{N} \times \ldots \times \mathbb{N}$, and we retrieve that $\mathcal{S}(\mathbb{N}^k) = \mathbb{N}_\omega^k$. Second, when A is a finite alphabet, A^* is given by products, as given in the Σ^* paragraph of Section 4; i.e., we retrieve the products (and SREs) of Abdulla *et al.* [1]. The more complicated case $(A^{\circledast})^*$ was dealt with by Abdulla *et al.* [3]. 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. Additionally, we can deal with more complex data structures such as, e.g., $(((\mathbb{N} \times A_{\leq})^* \times \mathbb{N})^{\circledast})^{\circledast}$.

Finally, note that (1) and (2) are two separate concerns in the proof of Theorem 5.3. If we are ready to relinquish orderings for the more general topological route, as advocated in [14], we could also enrich our grammar of data types with infinite constructions such as $\mathbb{P}(D)$, where $\mathbb{P}(D)$ is interpreted as the powerset of D with the so-called lower Vietoris topology. See Appendix G, where we show that $\mathcal{S}(\mathbb{P}(X)) \cong \mathcal{H}(X)$ is Noetherian whenever X is, and that its elements can be represented as *finite* subsets A of $\mathcal{S}(X)$, interpreted as $\bigcup_{C \in A} C$. In a sense, while $\mathcal{S}(X_a) = Idl(X)$ for all ordered spaces X, the sobrification construction is more robust than the ideal completion.

6. Completing WSTS, or: Towards Forward Procedures Computing the Cover

We show how one may use our completions on wqos to deal with forward analysis of well-structured systems. We shall describe this in more detail in another paper. First note that any data type D of Section 5 is suited to applying the expand, enlarge and check algorithm [12] out of the box to this end, since then $\mathcal{S}(D_a)$ is (the least) WADL for D. We instead explore extensions of the Karp-Miller procedure [15], in the spirit of Finkel [9] or Emerson and Namjoshi [7]. While the latter assumes an already built completion, we construct it. Also, we make explicit how this kind of acceleration-based procedure really computes the cover, i.e., $\downarrow Post^*(\downarrow x)$, in Proposition 6.1.

Recall that a *well-structured transition system* (WSTS) is a triple $S = (X, \leq, (\delta_i)_{i=1}^n)$, where X is well-quasi-ordered by $\leq$, and each $\delta_i : X \to X$ is a partial monotonic transition function. (By "partial monotonic" we mean that the domain of δ_i is upward closed, and δ_i is monotonic on its domain.) Letting $Pre(A) = \bigcup_{i=1}^n \delta_i^{-1}(A)$, $Pre^0(A) = A$, and $Pre^*(A) = \bigcup_{k \in \mathbb{N}} Pre^k(A)$, it is well-known that any upward closed subset of X is of the form $\uparrow E$ for some finite $E \subseteq X$, and that $Pre^*(\uparrow E)$ is an upward-closed subset $\uparrow E'$, E' finite, that arises as $\bigcup_{k=0}^m Pre^k(\uparrow E)$ for some $m \in \mathbb{N}$. Hence, provided $\leq$ is decidable and $\delta_i^{-1}(\uparrow E)$ is computable for each finite E, it is decidable whether $x \in Pre^*(\uparrow E)$, i.e., whether one may reach $\uparrow E$ from x in finitely many steps. It is equivalent to check whether $y \in \downarrow Post^*(\downarrow x)$ for some $y \in E$, where $Post(A) = \bigcup_{i=1}^n \delta_i(A)$, $Post^0(A) = A$, and $Post^*(A) = \bigcup_{k \in \mathbb{N}} Post^k(A)$.

All the existing symbolic procedures that attempt to compute $\downarrow Post^*(\downarrow x)$, even with a finite number of accelerations (e.g., Fast, Trex, Lash), can only compute subsets of the larger set $\mathrm{Lub}(\downarrow Post^*(\downarrow x))$. In general, $\mathrm{Lub}(\downarrow Post^*(\downarrow x))$ does not admit a finite representation. On the other hand, we know that the Scott-closure $cl(Post^*(\downarrow x))$, as a closed subset of $Idl(X)$ (intersected with X itself), is always finitary. Indeed, it is also a closed subset of $\mathcal{S}(X_a)$ (Proposition 3.3), which is represented as the downward closure of finitely many elements of $\mathcal{S}(X_a)$. Since $Y = Idl(X)$ is continuous, Proposition 3.5 allows us to conclude that $\mathrm{Lub}_Y(\downarrow Post^*(\downarrow x)) = cl(Post^*(\downarrow x))$ is finitary—hence representable provided X is one of the data types of Section 5.

This leads to the following construction. Any partial monotonic map $f : X \to Y$ between quasi-ordered sets lifts to a *continuous* partial map $\mathcal{S}f : \mathcal{S}(X_a) \to \mathcal{S}(Y_a)$: for each irreducible closed subset (a.k.a., ideal) C of $\mathcal{S}(X_a)$, either $C \cap \mathrm{dom}\, f \neq \emptyset$ and $\mathcal{S}f(C) = \downarrow f(C) = \{y \in Y \mid \exists x \in C \cap \mathrm{dom}\, f \cdot y \leq f(x)\}$, or $C \cap \mathrm{dom}\, f = \emptyset$ and $\mathcal{S}f(C)$ is undefined. The *completion* of a WSTS $S = (X, \leq, (\delta_i)_{i=1}^n)$ is then the transition system $\widehat{S} = (\mathcal{S}(X_a), \subseteq, (\mathcal{S}\delta_i)_{i=1}^n)$.

For example, when $X = \mathbb{N}^k$, and S is a Petri net with transitions δ_i defined as $\delta_i(\vec{x}) = \vec{x} + \vec{d_i}$ (where $\vec{d_i} \in \mathbb{Z}^k$; this is defined whenever $\vec{x} + \vec{d} \in \mathbb{N}^k$), then $\widehat{S}$ is the transition system whose set of states is $\mathcal{S}(X) = \mathbb{N}_\omega^k$, and whose transition functions are: $\mathcal{S}\delta_i(\vec{x}) = \vec{x} + \vec{d_i}$, whenever this has only non-negative coordinates, taking the convention that $\omega + d = \omega$ for any $d \in \mathbb{Z}$.

We may emulate lossy channel systems through the following *functional-lossy* channel systems (FLCS). For simplicity, we assume just one channel and no local state; the general case would only make the presentation more obscure. An FLCS differs from an LCS in that it loses only the least amount of messages needed to enable transitions. Take $X = \Sigma^*$ for some finite alphabet Σ of messages; the transitions are either of the form $\delta_i(w) = wa_i$ for some fixed letter a_i (sending a_i onto the channel), or of the form $\delta_i(w) = w_2$ whenever w is of the form $w_1 a_i w_2$, with w_1 not containing a_i (expecting to receive a_i). Any LCS is cover-equivalent to the FLCS with the same sends and receives, where two systems are *cover-equivalent* if and only if they have the same sets $\downarrow Post^*(F)$ for any downward-closed F. Equating $\mathcal{S}(\Sigma_a^*)$ with the set of products, as advocated in Section 4, we find that transition functions of the first kind lift to $\mathcal{S}\delta_i(P) = Pa_i^?$, while transition functions of the second kind lift to: $\mathcal{S}\delta_i(\epsilon)$ is undefined, $\mathcal{S}\delta_i(a^?P) = \mathcal{S}\delta_i(P)$ if $a_i \neq a$, $\mathcal{S}\delta_i(a_i^?P) = P$, $\mathcal{S}\delta_i(A^*P) = \mathcal{S}\delta_i(P)$ if $a_i \notin A$, $\mathcal{S}\delta_i(A^*P) = A^*P$ otherwise. This is exactly how Trex computes successors [1, Lemma 6].

In general, the results of Section 5 allow us to use any domain of datatypes D for the state space X of S. The construction $\widehat{S}$ then generalizes all previous constructions, which used to be defined specifically for each datatype.

The Karp-Miller algorithm in Petri nets, or the Trex procedure for lossy channel systems, gives information about the cover $\downarrow Post^*(\downarrow x)$. This is true of *any* completion $\widehat{S}$ as constructed above:

Proposition 6.1. *Let S be a WSTS. Let $\widehat{Post}$ be the $Post$ map of the completion $\widehat{S}$. For any closed subset F of $\mathcal{S}(X_a)$, $\widehat{Post}(F) = cl(Post(F \cap X))$, and $\widehat{Post}^*(F) = cl(Post^*(F \cap X))$. Hence, for any downward closed subset F of X, $\downarrow Post(F) = X \cap \widehat{Post}(F)$, $\downarrow Post^*(F) = X \cap \widehat{Post}^*(F)$.*

Proof. Let F be closed in $\mathcal{S}(X_a)$. $\widehat{Post}(F) = \bigcup_{i=1}^n cl(\delta_i(F)) = cl(\bigcup_{i=1}^n \delta_i(F)) = cl(Post(F))$, since closure commutes with (arbitrary) unions. We then claim that $\widehat{Post}^k(F) = cl(Post^k(F))$ for each $k \in \mathbb{N}$. This is by induction on k. The cases $k = 0, 1$ are obvious. When $k \geq 2$, we use the fact that, for any continuous partial map f: $(*)$ $cl(f(cl(A))) = cl(f(A))$. Then $\widehat{Post}^k(F) = \bigcup_{i=1}^n cl(\delta_i(\widehat{Post}^{k-1}(F))) = \bigcup_{i=1}^n cl(\delta_i(cl(Post^{k-1}(F)))) = \bigcup_{i=1}^n cl(\delta_i(Post^{k-1}(F)))$ (by $(*)$) $= cl(Post^k(F))$. Finally, $\widehat{Post}^*(F) = \bigcup_{k\in\mathbb{N}} \widehat{Post}^k(F) = \bigcup_{k\in\mathbb{N}} cl(Post^k(F)) = cl(Post^*(F))$. We conclude, since for any $A \subseteq X$, $\downarrow A$ is the closure of A in X_a; the topology of X_a is the subspace topology of that of $\mathcal{S}(X_a)$; so, writing cl for closure in $\mathcal{S}(X_a)$, $\downarrow A = X \cap cl(A)$. ∎

Writing F as the finite union $C_1 \cup \ldots \cup C_k$, where $C_1, \ldots, C_k \in \mathcal{S}(X_a)$, $\widehat{Post}(F)$ is computable as $\bigcup_{1\leq i_1,\ldots,i_n\leq k} \mathcal{S}\delta_1(C_{i_1}) \cup \ldots \cup \mathcal{S}\delta_n(C_{i_n})$, assuming $\mathcal{S}\delta_i$ computable for each i. (We take $\mathcal{S}\delta_j(C_i)$ to mean $\emptyset$ if undefined, for notational convenience.) Although $\mathcal{S}\delta_i$ may be uncomputable even when δ_i is, it is computable on most WSTS in use. This holds, for example, for Petri nets and lossy channel systems, as exemplified above.

So it is easy to compute $\downarrow Post(\downarrow x)$, as (the intersection of X with) $\widehat{Post}(\downarrow x)$. Computing $\downarrow Post^*(\downarrow x)$ (our goal) is also easily computed as $\widehat{Post}^*(\downarrow x)$ (intersected with X again), using acceleration techniques for loops. This is what the Karp-Miller construction does for Petri nets, what Trex does for lossy channel systems [1]. (We examine termination issues below.) Our framework generalizes all these procedures, using a weak acceleration assumption, whereby we assume that we can compute the least upper bound of the values of loops iterated k times, $k \in \mathbb{N}$. For any continuous partial map $g : Y \to Y$ (with open domain) on a dcpo Y, let the *iteration* $\overline{g}$ be the map of domain $\operatorname{dom} g$ such that $\overline{g}(y)$ is the least upper bound of $(g^k(y))_{k\in\mathbb{N}}$ if $y < g(y)$, and $g(y)$ otherwise. Let $\Delta = \{\mathcal{S}\delta_1, \ldots, \mathcal{S}\delta_n\}$, Δ^* be the set of all composites of finitely many maps from Δ. Our *acceleration assumption* is that one can compute $\overline{g}(y)$ for any $g \in \Delta^*$, $y \in \mathcal{S}(X_a)$. The following procedure then computes $\downarrow Post^*(\downarrow x)$, as (the intersection of X with) $\widehat{Post}^*(\downarrow x)$, itself represented as a finite union of elements of $\mathcal{S}(X_a)$: initially, let A be $\{x\}$; then, while $\widehat{Post}(A) \not\subseteq \downarrow A$, choose fairly $(g, a) \in \Delta^* \times A$ such that $a \in \operatorname{dom} g$ and add $\overline{g}(a)$ to A. If this terminates, A is a finite set whose downward closure is exactly $\downarrow Post^*(\downarrow x)$. Despite its simplicity, this is the essence of the Karp-Miller procedure, generalized to a large class of spaces X.

Termination is ensured for flat systems, i.e., systems whose control graph has no nested loop, as one only has to compute the effect of a finite number of loops. In general, the procedure terminates on *cover-flattable* systems, that is systems that are cover-equivalent to some flat system. Petri nets are cover-flattable, while, e.g., not all LCS are: recall that, in an LCS, $\downarrow Post^*(\downarrow x)$ is *always* representable as an SRE, however not effectively so.

7. Conclusion and Perspectives

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. 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 $\mathcal{S}(X_a)$ was in a sense inevitable (Section 3), and described how this applied to compute downward closures of reachable sets of configurations in WSTS (Section 6). We plan to describe such new forward analysis algorithms, in more detail, in papers to come.

References

[1] P. A. Abdulla, A. Bouajjani, and B. Jonsson. On-the-fly analysis of systems with unbounded, lossy fifo channels. In *CAV'98*, Vancouver, Canada, 1998. Springer Verlag LNCS 1427.

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

[3] P. A. Abdulla, J. Deneux, P. Mahata, and A. Nylén. Forward reachability analysis of timed Petri nets. In Y. Lakhnech and S. Yovine, editors, *FORMATS/FTRTFT*, pages 343–362. Springer Verlag LNCS 3253, 2004.

[4] P. A. Abdulla and B. Jonsson. Verifying programs with unreliable channels. In *LICS'93*, pages 160–170, 1993.

[5] S. Abramsky and A. Jung. Domain theory. In S. Abramsky, D. M. Gabbay, and T. S. E. Maibaum, editors, *Handbook of Logic in Comp. Sci.*, volume 3, pages 1–168. OUP, 1994.

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

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

[8] J. Esparza, A. Finkel, and R. Mayr. On the verification of broadcast protocols. In *LICS'99*, pages 352–359, 1999.

[9] A. Finkel. Reduction and covering of infinite reachability trees. *Inf. Comput.*, 89(2):144–179, 1990.

[10] A. Finkel and P. Schnoebelen. Well-structured transition systems everywhere! *Theor. Comp. Sci.*, 256(1–2):63–92, 2001.

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

[12] G. Geeraerts, J.-F. Raskin, and L. van Begin. Expand, enlarge and check: New algorithms for the coverability problem of WSTS. *J. Comp. Sys. Sci.*, 72(1):180–203, 2006.

[13] G. Gierz, K. H. Hofmann, K. Keimel, J. D. Lawson, M. Mislove, and D. S. Scott. Continuous lattices and domains. In *Encyc. Math. and its Applications*, volume 93. CUP, 2003.

[14] J. Goubault-Larrecq. On Noetherian spaces. In *LICS'07*, pages 453–462, 2007.

[15] R. M. Karp and R. E. Miller. Parallel program schemata. *J. Comp. Sys. Sci.*, 3(2):147–195, 1969.

[16] M. Mislove. Algebraic posets, algebraic cpo's and models of concurrency. In *Topology and Category Theory in Computer Science*, pages 75–109. Clarendon Press, 1981.

Appendix A. All Sets of Limits Coincide on Continuous Posets

Continuous posets are nice spaces, in that one can compute the inductive closure in just one Lub_Y step, provided we start from a downward-closed set. This is Proposition 3.5, which also states that we get the Scott-topological closure this way.

Lemma A.1. *Let Y be a poset, and A a downward closed subset of Y. Then $\mathrm{Lub}_Y(A) \subseteq \downarrow \mathrm{Lub}_Y(A)$, and equality holds whenever Y is a continuous poset.*

Proof. Inclusion is obvious. Let us show equality, assuming Y is continuous. Let $x \in \downarrow \mathrm{Lub}_Y(A)$: for some $z \geq x$, z is the least upper bound of some directed family $(z_i)_{i \in I}$ in A. Since Y is continuous, $x = \sup_{y \ll x} y$, so by definition of $\ll$, every $y \ll x$ is less than or equal to some z_i, $i \in I$. In particular, every such y is in $\downarrow A$. So $x \in \mathrm{Lub}_Y(\downarrow A)$. ■

Equality may fail without the continuity assumption. E.g., let Y be $\mathbb{N}_\omega$ union a fresh element $*$ with $0 < * < \omega$, but incomparable with all other elements. Then $\mathbb{N}$ is a downward closed subset of Y, however $\mathrm{Lub}_Y(A) = \mathbb{N}_\omega$, and $\downarrow \mathrm{Lub}_Y(A) = Y = \mathbb{N}_\omega \cup \{*\}$.

We use the following technical lemma. This is folklore.

Lemma A.2. *Let Y be a continuous poset, $(x_i)_{i \in I}$ a directed family of elements of Y, with least upper bound x, and assume that each x_i is the least upper bound of a directed family $(x_{ij})_{j \in J_i}$, $i \in I$. Then $(x_{ij})_{\substack{i \in I \\ j \in J_i}}$ is a directed family, and has x as least upper bound.*

Proof. Given x_{ij} and $x_{i'j'}$, one can find $i'' \in I$ such that $x_i, x_{i'} \leq x_{i''}$; since $x_{ij}, x_{i'j'} \ll x_{i''} = \sup_{j'' \in J_{i''}} x_{i''j''}$, there are $k, k' \in J_{i''}$ such that $x_{ij} \leq z_{i''k}$ and $x_{i'j'} \leq x_{i''k'}$; by directedness again, there is an $j'' \in J_{i''}$ such that $x_{i''k}, x_{i''k'} \leq x_{i''j''}$, whence $x_{ij}, x_{i'j'} \leq x_{i''j''}$. So $(x_{ij})_{\substack{i \in I \\ j \in J_i}}$ is directed. It is clear that $\sup_{\substack{i \in I \\ j \in J_i}} x_{ij} = \sup_{i \in I} \sup_{j \in J_i} x_{ij} = \sup_{i \in I} x_i = x$. ■

We recall the statement of Proposition 3.5: Let Y be a continuous poset, then, for every downward-closed subset A of Y, $\mathrm{Ind}_Y(A) = \mathrm{Lub}_Y(A) = cl(A)$.

Proof. Clearly, $\mathrm{Lub}_Y(A) \subseteq \mathrm{Ind}_Y(A) \subseteq cl(A)$. It remains to show that $cl(A) \subseteq \mathrm{Lub}_Y(A)$, i.e., that $\mathrm{Lub}_Y(A)$ is downward-closed and closed under directed least upper bounds. This is downward-closed by Lemma A.1.

We note that: $(*)$ every element x of $\mathrm{Lub}_Y(A)$ is the least upper bound of some directed family of elements of A way-below x. Indeed, we just take $\twoheaddownarrow x$, using the fact that Y is continuous, and check that it is contained in A. Because $x \in \mathrm{Lub}_Y(A)$, x is the least upper bound of some directed family $(x_i)_{i \in I}$ in A. For any $y \in \twoheaddownarrow x$, we obtain $y \ll x = \sup_{i \in I} x_i$, so $y \leq x_i$ for some $i \in I$. Since $x_i \in A$ and A is downward closed, $y \in A$. Since y is arbitrary, $\twoheaddownarrow x \subseteq A$.

Now let $z \in \mathrm{Lub}_Y(\mathrm{Lub}_Y(A))$. There is a directed family $(z_j)_{j \in J}$ of elements of $\mathrm{Lub}_Y(A)$ that has z as least upper bound. Using $(*)$, write z_j is the least upper bound of a family $(z_{ji})_{i \in I_j}$ of elements of A such that $z_{ji} \ll z_j$ for all j, i. Then the family $(z_{ji})_{j \in J, i \in I_j}$ is again directed, and has z as least upper bound by Lemma A.2. So $z \in \mathrm{Lub}_Y(A)$. It follows that $\mathrm{Lub}_Y(\mathrm{Lub}_Y(A)) \subseteq \mathrm{Lub}_Y(A)$, i.e., that $\mathrm{Lub}_Y(A)$ is closed under directed least upper bounds. ■

Appendix B. Well-Based Continuous Dcpos and Ideal Completions

Lemma B.1. *Any continuous poset with a well-founded basis is algebraic, with a well-ordered set of finite elements.*

Proof. Assume Y is a continuous poset, has a well-founded basis B, but is not algebraic. There is an element $x \in Y$ that is not the least upper bound of a directed family of finite elements below x. We first claim that we can assume $x \in B$.

Since Y is continuous with basis B, x is the least upper bound of a directed family of elements $(x_i)_{i \in I}$ in B that are way-below x. If every x_i were the least upper bound of a directed family $(x_{ij})_{j \in J_i}$ of finite elements, Lemma A.2 would entail that x would be the least upper bound of the directed family $(x_{ij})_{\substack{i \in I \\ j \in J_i}}$, consisting of finite elements, contradiction.

So there is an $x \in B$ that is not the least upper bound of a directed family of finite elements below x. Since B is well-founded, we may choose x minimal. Since B is a basis of Y, write x as the least upper bound of some directed family $(x_i)_{i \in I}$ of elements of B way below x. Since x was chosen minimal, every x_i were the least upper bound of a directed family $(x_{ij})_{j \in J_i}$ of finite elements. As above, Lemma A.2 entails that x is the least upper bound of the directed family $(x_{ij})_{\substack{i \in I \\ j \in J_i}}$, consisting of finite elements, contradiction.

So Y is algebraic. Since every finite element is in every basis, the set of finite elements is contained in B, and is therefore well-ordered, since B is. ■

Recall the statement of Proposition 3.6:

> Any continuous dcpo Y with a well-founded basis is order-isomorphic to $Idl(X)$ for some well-ordered set X. One may take the subset of finite elements of Y for X. If Y is well-based, then X is well-ordered.

Proof. Let X be the set of finite elements of Y. By Lemma B.1, X is well-ordered, and Y is algebraic. Now Y is order-isomorphic to $Idl(X)$, using the well-known fact that any two continuous dcpos with isomorphic bases are isomorphic. Concretely, here, the map $\eta : Y \to Idl(X)$ that sends each $y \in Y$ to $\{x \in X \mid x \leq y\}$ is monotonic and continuous: for each directed family $(y_i)_{i \in I}$ in Y with least upper bound y, $\eta(y) = \{x \in X \mid x \leq \sup_{i \in I} y_i\} = \{x \in X \mid \exists i \in I \cdot x \leq y_i\}$ (because x is finite, i.e., $x \ll x) = \bigcup_{i \in I} \eta(y_i)$. Conversely, the map $\epsilon : Idl(X) \to Y$ that sends each ideal F to $\sup_{x \in F} x$ is also continuous: for every directed family $(F_i)_{i \in I}$ of ideals of X, $\epsilon(\bigcup_{i \in I} F_i) = \sup_{x \in \bigcup_{i \in I} F_i} x = \sup_{\exists i \in I \cdot y \in F_i} x = \sup_{i \in I} \sup_{x \in F_i} x = \sup_{i \in I} \epsilon(F_i)$. It is easy to check that η and ϵ are inverse of each other: $\epsilon(\eta(y)) = \sup_{x \in \{x \in X \mid x \leq y\}} x = y$, $\eta(\epsilon(F)) = \{x \in X \mid x \leq \sup_{x' \in F} x'\} = \{x \in X \mid \exists x' \in F \cdot x \leq x'\}$ (since each $x \in X$ is finite) $= \{x \in X \mid x \in F\}$ (since F is downward closed) $= F$. ■

In other words, well-based continuous posets are *special cases* of the notion of weak adequate domains of limits. These are the minimal cases where one takes a wqo X, and adds all limits in $Idl(X) = \mathcal{S}(X_a)$.

Appendix C. Proof of Proposition 4.2

Let X be a Noetherian space. We show that every closed subset of X is a finite union of irreducible closed subsets.

By [14, Proposition 6.2], $\mathcal{S}(X)$ is Noetherian, too. The key to this result is the fact that $\mathcal{S}(X)$ has exactly the same opens as X, in the sense described in op.cit.: the map that sends each open

U of X to the open $\Diamond U = \{F \in \mathcal{S}(X) \mid F \cap U \neq \emptyset\}$ is an isomorphism. This extends to an isomorphism between the lattices of closed subsets, mapping each closed subset F' of X to the closed subset $\Box F' = \mathcal{S}(X) \setminus \Diamond(X \setminus F') = \{F \in \mathcal{S}(X) \mid F \subseteq F'\}$.

Now, by [14, Corollary 6.5], since $\mathcal{S}(X)$ is both sober and Noetherian, every closed subset of $\mathcal{S}(X)$ is finitary, i.e., of the form $\downarrow E$ for some finite subset E of $\mathcal{S}(X)$. In particular, every closed subset of $\mathcal{S}(X)$ is a finite union of irreducible closed subsets, namely $\downarrow x$, $x \in E$. Using the isomorphism $F' \mapsto \Box F'$, every closed subset of X must also be a finite union of irreducible closed subsets. Concretely, for any closed subset F' of X, $\Box F'$ is a finite union of irreducible closed subsets $\downarrow F_i = \{F'' \in \mathcal{S}(X) \mid F'' \subseteq F_i\} = \Box F_i$, where F_i, $1 \leq i \leq n$, ranges over some finite set of irreducible closed subsets. Now note that $\Box F' = \Box F_1 \cup \ldots \cup \Box F_n$ equals $\Box(F_1 \cup \ldots \cup F_n)$. Indeed, for every $F'' \in \mathcal{S}(X)$ that is contained in some F_i, F'' is contained in $F_1 \cup \ldots \cup F_n$; conversely, if $F'' \in \mathcal{S}(X)$ is contained in $F_1 \cup \ldots \cup F_n$, then it must be contained in some F_i, $1 \leq i \leq n$, since F'' is irreducible. From $\Box F' = \Box(F_1 \cup \ldots \cup F_n)$, we conclude that $F' = F_1 \cup \ldots \cup F_n$. ■

Appendix D. Finite Trees, with Homeomorphic Embedding

The situation for finite trees is very similar to finite words. Let $\mathcal{F}$ be a finite signature of function symbols with their arities. We let $\mathcal{F}_p$ the set of function symbols of arity k; $\mathcal{F}_0$ is the set of *constants*, and is assumed to be non-empty. The set $\mathcal{T}(\mathcal{F})$ is the set of ground terms built from $\mathcal{F}$.

We rest on a version of Kleene's Theorem for trees [6, Section 2.2]. Let $\mathcal{K}$ be a countably infinite set of constants, disjoint from $\mathcal{F}$. The set of regular tree expressions on $\mathcal{F}$ and $\mathcal{K}$ is defined by the grammar:

$$\begin{array}{rcl} L & ::= & f(L_1, \ldots, L_p) \mid \emptyset \\ & \mid & \Box \mid L + L \mid L ._{\Box} L \mid L^{*,\Box} \end{array}$$

where $f \in \mathcal{F}_p$, $p \in \mathbb{N}$, $\Box \in \mathcal{K}$. The new thing, compared to word regular expressions, is the notion of *hole* $\Box \in \mathcal{K}$. This is used to give meaning to concatenation $L_1 ._{\Box} L_2$ and to Kleene star $L^{*,\Box}$.

Each tree regular expression defines a language of terms in $\mathcal{T}(\mathcal{F} \cup \mathcal{K})$ by: the language of $f(L_1, \ldots, L_p)$ is the set of terms $f(t_1, \ldots, t_p)$ with t_1 is in the language of L_1, …; t_p is in the language of L_p; the language of $\emptyset$ is the empty set; the language of $\Box \in \mathcal{K}$ is $\{\Box\}$; the language of $L_1 + L_2$ is the union of those of L_1 and L_2; the language of $L^{*,\Box}$ is the union of the languages of $\Box$, $L = L._{\Box}\Box$, $L._{\Box}L._{\Box}\Box$, …, $L._{\Box}L._{\Box} \ldots ._{\Box}L._{\Box}\Box$ (n times), …The subtle point is the definition of the language of $L_1 ._{\Box} L_2$. For any term $t \in \mathcal{F}(\mathcal{T} \cup \mathcal{K})$ and any language L, define $t._{\Box}L$ as the language defined by induction on t as follows: $\Box._{\Box}L = L$, $\Box'._{\Box}L = \{\Box'\}$ if $\Box' \neq \Box$, $f(t_1, \ldots, t_p)._{\Box}L = \{f(u_1, \ldots, u_p) \mid u_1$ in the language of $t_1._{\Box}L, \ldots, u_p$ in the language of $t_p._{\Box}L\}$. Then the language of $L_1._{\Box}L_2$ is the union of the languages $t_1._{\Box}L_2$ over all terms t_1 in the language of L_1. The subtlety is that this is *not* the set of terms $t_1[\Box := t_2]$ with t_1 in the language of L_1 and t_2 in the language of L_2 (where substitution of terms for holes is defined in the obvious way). The difference arises when $\Box$ occurs several times in L_1. For example, if $L_1 = f(\Box, \Box)$ and $L_2 = a + b$, for two constants a, b, the set of terms $t_1[\Box := t_2]$ would be $\{f(a, a), f(b, b)\}$. However, the language of $L_1._{\Box}L_2$ is $\{f(a, a), f(a, b), f(b, a), f(b, b)\}$. In other words, we may replace different occurrences of the same hole $\Box$ by different terms from L_2.

Kleene's Theorem for trees [6, Theorem 19, Section 2.2] states that a tree language is regular if and only if it is the language of some tree expression. There is subtlety here, related to the set of function symbols we allow ourselves: we wish to define languages of terms in $\mathcal{T}(\mathcal{F})$, while tree regular expressions give languages of terms in $\mathcal{T}(\mathcal{F} \cup \mathcal{K})$. So we need to restrict tree regular expressions so that they recognize terms on $\mathcal{T}(\mathcal{F})$.

Definition D.1. The set of *free holes* $\mathrm{fh}(L)$ of a tree regular expression L is defined by:

$$\mathrm{fh}(f(L_1,\ldots,L_n)) = \bigcup_{i=1}^n \mathrm{fh}(L_i) \quad \mathrm{fh}(\emptyset) = \emptyset \quad \mathrm{fh}(\Box) = \{\Box\} \quad \mathrm{fh}(L^{*,\Box}) = \mathrm{fh}(L) \cup \{\Box\}$$
$$\mathrm{fh}(L_1 + L_2) = \mathrm{fh}(L_1) \cup \mathrm{fh}(L_2) \quad \mathrm{fh}(L_1 ._\Box L_2) = (\mathrm{fh}(L_1) \setminus \{\Box\}) \cup \mathrm{fh}(L_2)$$

A regular tree expression L is *closed* if and only if it has no free hole.

It is easy to see that any term in the language of L is in $\mathcal{T}(\mathcal{F} \cup \mathrm{fh}(L))$, i.e., contains only holes that are free in L.

Proposition D.2. *Let $\mathcal{F}$ contain at least one constant. A language of terms in $\mathcal{T}(\mathcal{F})$ is regular if and only if it is the language of some closed regular tree expression.*

Proof. By the above remark, the language of any closed regular tree expression is not only regular, but also in $\mathcal{T}(\mathcal{F})$. Conversely, any regular language of terms in $\mathcal{T}(\mathcal{F})$ is definable as the language of some regular tree expression L. Let a be a constant in $\mathcal{F}$. Letting $\Box_1$, …, $\Box_k$ be the free holes in L, let L' be $L ._{\Box_1} a ._{\Box_2} a \ldots ._{\Box_k} a$. Since the language of L only contains terms without holes, the language of L' is the same as that of L. Moreover, it is easy to check that L' is closed. ∎

The *homeomorphic embedding* ordering $\trianglelefteq$ on $\mathcal{T}(\mathcal{F})$ (or on $\mathcal{T}(\mathcal{F} \cup \mathcal{K})$) is defined as the smallest relation such that, whenever $u = f(u_1,\ldots,u_m)$ and $v = g(v_1,\ldots,v_n)$, $u \trianglelefteq v$ if and only if $u \trianglelefteq v_j$ for some j, $1 \leq j \leq n$, or $f = g$, $m = n$, and $u_1 \trianglelefteq v_1$, $u_2 \trianglelefteq v_2$, …, $u_m \trianglelefteq v_m$. (In general, the homeomorphic embedding ordering is defined relative to a well-quasi-ordering $\preceq$ on function symbols, and, instead of $f = g$ in the second case, we would require that $f \preceq g$ and there is an increasing subsequence $1 \leq j_1 < j_2 < \ldots < j_m \leq n$ such that $u_1 \trianglelefteq v_{j_1}$, $u_2 \trianglelefteq v_{j_2}$, …, $u_m \trianglelefteq v_{j_m}$.) Kruskal's Theorem states that $\trianglelefteq$ is a well ordering on $\mathcal{T}(\mathcal{F})$.

We elucidate the structure of the adequate domain of limits $\mathcal{S}(\mathcal{T}(\mathcal{F})_a)$.

Lemma D.3. *Any downward closed subset of $\mathcal{T}(\mathcal{F})$ is the language of a closed tree regular expression of the form:*

$$\begin{array}{rcl} L & ::= & f^?(L_1,\ldots,L_p) \mid \emptyset \\ & \mid & \Box \mid L + L \mid L ._\Box L \mid L^{*,\Box} \end{array}$$

where the language of $f^?(L_1,\ldots,L_p)$ is by convention the one of $f(L_1,\ldots,L_p) + L_1 + \ldots + L_p$.

Accordingly, we extend Definition D.1 so that $\mathrm{fh}(f^?(L_1,\ldots,L_p)) = \bigcup_{i=1}^p \mathrm{fh}(L_i)$.

Recall from Definition 4.3 that *tree products* and *product iterators* are defined inductively by:

- Every hole $\Box$ is a tree product.
- $f^?(P_1,\ldots,P_k)$ is a tree product, for any $f \in \Sigma_k$ and any tree products $P_1,\ldots,P_k$.
- $(\sum_{i=1}^n C_i)^{*,\Box} ._\Box P$ is a tree product, for any tree product P, any integer $n \geq 1$, and any product iterators C_i over $\Box$, $1 \leq i \leq n$. We write $\sum_{i=1}^n C_i$ for $C_1 + C_2 + \ldots + C_n$.
- $f(P_1,\ldots,P_k)$ is a product iterator over $\Box$ for any $f \in \Sigma_k$, where: 1. each P_i, $1 \leq i \leq k$ is either $\Box$ itself or a tree product such that $\Box$ is not in the language of P_i; and 2. $P_i = \Box$ for some i, $1 \leq i \leq k$.

A *simple tree regular expression* (STRE) is a finite sum of tree products (possibly empty, in which case $\emptyset$ is meant).

In the case of $(\sum_{i=1}^n C_i)^{*,\Box} ._\Box P$, note that we may rename $\Box$ to any other hole, in a way reminiscent to α-renaming in the λ-calculus. Formally, we may define $C[\Box := \Box']$, for any product iterator $C = f(P_1,\ldots,P_k)$ over $\Box$, as $f(P'_1,\ldots,P'_k)$, where for each i, $1 \leq i \leq k$, either $P_i = \Box$ and then $P'_i = \Box'$, or $P'_i = P_i$. When $\Box'$ is not free in C (or $\Box' = \Box$), $C[\Box := \Box']$ is a product iterator over $\Box'$. When $\Box'$ is not free in any C_i, moreover, $(\sum_{i=1}^n C_i[\Box := \Box'])^{*,\Box'}.{}_{\Box'}P$ defines the same language as $(\sum_{i=1}^n C_i)^{*,\Box}.{}_{\Box}P$, and has the same free holes.

For example, $f^?(a^?, b^?)$ and $(f(\Box, g^?(a^?)) + f(g^?(b^?), \Box))^{*,\Box}.{}_{\Box}f^?(a^?, b^?)$ are tree products. They are also closed tree products.

Lemma D.4. *Let P, P' be two tree products, and $\Box$ be a hole.*

(1) *If $\Box \notin \mathrm{fh}(P)$, then $P.{}_{\Box}P'$ defines the same language as P.*
(2) *If $P = \Box$, then $P.{}_{\Box}P'$ defines the same language as P'.*
(3) *If $P = f^?(P_1, \ldots, P_k)$, where $P_1, \ldots, P_k$ are tree products, then $P.{}_{\Box}P'$ defines the same language as $f^?(P_1.{}_{\Box}P', \ldots, P_k.{}_{\Box}P')$.*
(4) *If $P = (\sum_{i=1}^n C_i)^{*,\Box'}.{}_{\Box'}P_0$, where P_0 is a tree product, and $\Box' \neq \Box$, $\Box' \notin \mathrm{fh}(P')$, then $P.{}_{\Box}P'$ defines the same language as $(\sum_{i=1}^n C_i.{}_{\Box}P')^{*,\Box'}.{}_{\Box'}(P_0.{}_{\Box}P')$.*

Finally, if C is a product iterator $f(P_1, \ldots, P_k)$ over $\Box'$, P' is a tree product, and $\Box' \neq \Box$, $\Box' \notin \mathrm{fh}(P')$, then $f(P_1.{}_{\Box}P', \ldots, P_k.{}_{\Box}P')$ defines the same language as C.

Lemma D.5. *Let P, P' be two tree products, and $\Box$ be a hole. Then there is a tree product P'' that defines the same language as $P.{}_{\Box}P'$. Moreover, $\mathrm{fh}(P'') \subseteq \mathrm{fh}(P.{}_{\Box}P')$.*

Proof. By induction on P. If $\Box \notin \mathrm{fh}(P)$, then take $P'' = P$. By Lemma D.4, item 1, this defines the same language as $P.{}_{\Box}P'$. Moreover, $\mathrm{fh}(P.{}_{\Box}P') = \mathrm{fh}(P) \cup \mathrm{fh}(P')$, since $\Box \notin \mathrm{fh}(P)$, hence contains $\mathrm{fh}(P'') = \mathrm{fh}(P)$.

If $\Box \in \mathrm{fh}(P)$, then we consider three cases, corresponding to the different possible forms for P. If P is a box, then $P = \Box$, and $P.{}_{\Box}P'$ defines the same language as P' by Lemma D.4, item 2. So take $P'' = P'$. We check that $\mathrm{fh}(P.{}_{\Box}P') \supseteq \mathrm{fh}(P') = \mathrm{fh}(P'')$.

If $P = f^?(P_1, \ldots, P_k)$, where $P_1, \ldots, P_k$ are tree products, then by induction hypothesis there are tree products P_1'', …, P_k'' that define the same languages as $P_1.{}_{\Box}P'$, …, $P_k.{}_{\Box}P'$ respectively. Moreover $\mathrm{fh}(P_i'') \subseteq \mathrm{fh}(P_i.{}_{\Box}P')$ for each i, $1 \leq i \leq k$. Using Lemma D.4, item 3, $P.{}_{\Box}P'$ defines the same language as $f^?(P_1'', \ldots, P_k'')$. Moreover, $\mathrm{fh}(f^?(P_1'', \ldots, P_k'')) = \bigcup_{i=1}^k \mathrm{fh}(P_i'') \subseteq \bigcup_{i=1}^k \mathrm{fh}(P_i.{}_{\Box}P') = \mathrm{fh}(P.{}_{\Box}P')$.

Finally, if $P = (\sum_{i=1}^n C_i)^{*,\Box'}.{}_{\Box'}P_0$, where P_0 is a tree product, first, we may assume that $\Box'$ is fresh, i.e., that $\Box' \neq \Box$ and $\Box' \notin \mathrm{fh}(P')$, by α-renaming. We use the fact that: $(*)$ if C is a product iterator over $\Box'$, then there is a product iterator C'' over $\Box'$ such that $\mathrm{fh}(C'') \subseteq \mathrm{fh}(C)$, and which defines the same language as $C.{}_{\Box}P'$. We defer the proof of this for a moment. Knowing $(*)$, we can conclude that there is a product iterator C_i'' over $\Box'$ for each i, $1 \leq i \leq n$, such that $\mathrm{fh}(C_i'') \subseteq \mathrm{fh}(C_i)$, and defining the same language as $C_i.'_P$. Also, by induction hypothesis there is a product P_0'' that defines the same language as $P_0.{}_{\Box}P'$, and with $\mathrm{fh}(P_0'') \subseteq (P_0.{}_{\Box}P')$. By Lemma D.4, item 4, $(\sum_{i=1}^n C_i'')^{*,\Box'}.{}_{\Box'}P_0''$ is then a tree product that defines the same language as $P.{}_{\Box}P'$. Moreover, it is easy to check that its set of free holes is contained in $\mathrm{fh}(P.{}_{\Box}P')$.

We now come to prove $(*)$. Let $C = f(P_1, \ldots, P_k)$ be a product iterator over $\Box'$, $\Box' \neq \Box$, $\Box' \notin \mathrm{fh}(P')$. We construct tree products P_1'', …, P_k'' defining the same languages as $P_1.{}_{\Box}P'$, …, $P_k.{}_{\Box}P'$ respectively, and with $\mathrm{fh}(P_i'') \subseteq \mathrm{fh}(P_i.{}_{\Box}P')$ for every i, $1 \leq i \leq k$. For each i, if $P_i = \Box'$, then we may take $P_i'' = \Box'$ again, using Lemma D.4, item 1, since $\Box' \neq \Box$; otherwise, we use the induction hypothesis. Observe that in the first case $P_i'' = \Box'$, while in the second case $\Box'$ is not in $\mathrm{fh}(P_i'') \subseteq \mathrm{fh}(P_i.{}_{\Box}P') = (\mathrm{fh}(P_i) \setminus \{\Box\}) \subseteq \mathrm{fh}(P')$. Indeed, $\Box' \notin \mathrm{fh}(P_i)$ because $P_i \neq \Box'$, using property 1 of product iterators, and $\Box' \notin \mathrm{fh}(P')$ by assumption. So $C'' = f(P_1'', \ldots, P_k'')$ satisfies property 1 of product iterators. It also satisfies property 2: there is an i, $1 \leq i \leq k$, such that $P_i = \Box'$, whence $P_i'' = \Box'$ again. ■

Lemma D.6. *Let L, L' be two STREs. Then there is an STRE L'' defining the same language as $L._\Box L'$. Moreover,* $\mathrm{fh}(L'') \subseteq \mathrm{fh}(L._\Box L')$.

Proof. Because $._\Box$ distributes over sums, using Lemma D.5. ∎

Lemma D.7. *Let L be a tree regular expression, $\Box$ be a hole, and P be regular tree expression.*

(1) *If $\Box$ is not in the language of P, then $(P+L)^{*,\Box}$ defines the same language as $L^{*,\Box}._\Box P + L^{*,\Box}$.*
(2) *$(\Box + L)^{*,\Box}$ defines the same language as $L^{*,\Box}$.*
(3) *If $P = f^?(P_1, \ldots, P_k)$ where $P_1, \ldots, P_k$ are tree products, then let $1 \leq i_1 < \ldots < i_\ell \leq k$ be the sequence of indices i such that $\Box$ is in the language of P_i. Then $(P+L)^{*,\Box}$ defines the same language as $(P' + P_1 + \ldots + P_k + L)^{*,\Box}$, where P' is obtained from P by replacing $f^?$ by f and each P_{i_j} by $\Box$. Formally, $P' = f(P_1, \ldots, P_{i_1-1}, \Box, P_{i_1+1}, \ldots, P_{i_2-1}, \Box, P_{i_2+1}, \ldots, P_{i_\ell-1}, \Box, P_{i_\ell+1}, \ldots, P_k)$.*
(4) *If $P = (\sum_{i=1}^n C_i)^{*,\Box'}._{\Box'} P_0$, with $\Box' \neq \Box$, and $\Box$ is in the language of P_0, then $(P+L)^{*,\Box}$ defines the same language as $(\sum_{i=1}^n C_i[\Box' := \Box] + P_0 + L)^{*,\Box}$.*
(5) *If $P = f(P_1, \ldots, P_k)$, where $P_1, \ldots, P_k$ are tree products, then let $1 \leq i_1 < \ldots < i_\ell \leq k$ be the sequence of indices i such that $\Box$ is in the language of P_i. Then $(P+L)^{*,\Box}$ defines the same language as $(P' + P_{i_1} + \ldots + P_{i_\ell} + L)^{*,\Box}$, where P' is obtained from P by replacing each P_{i_j} by $\Box$. Formally, $P' = f(P_1, \ldots, P_{i_1-1}, \Box, P_{i_1+1}, \ldots, P_{i_2-1}, \Box, P_{i_2+1}, \ldots, P_{i_\ell-1}, \Box, P_{i_\ell+1}, \ldots, P_k)$.*

Proof. The only technical point is item 4. Recall that we defined $C_i[\Box' := \Box]$ after Definition 4.3; this was used to define α-renaming. Note in particular that we do not assume that $\Box \notin \mathrm{fh}(C_i)$, so this is not a case of α-renaming here. For example, it might be that $C_i = f(\Box, \Box')$, then $C_i[\Box' := \Box] = f(\Box, \Box)$.

Assume t is a term in $L_0 = (P+L)^{*,\Box}$. So t is in $(P+L)._\Box (P+L)._\Box \ldots ._\Box (P+L)._\Box \Box$, with n times $P+L$, for some $p \in \mathbb{N}$. We show that t is in $L_1 = (\sum_{i=1}^n C_i[\Box' := \Box] + P_0 + L)^{*,\Box}$ by induction on p. If $p = 0$, then $t = \Box$, and the claim is clear. Otherwise, there is a term t_0 in the language of $P + L$, with, say, k occurrences of the hole $\Box$, and terms $t_1, \ldots, t_k$ in the language of $(P+L)._\Box (P+L)._\Box \ldots ._\Box (P+L)$ ($p-1$ times) such that t is obtained from t_0 by replacing the jth occurrence of $\Box$ by t_j, $j \leq j \leq k$. We shall use the convention to write this as $t = t_0[\Box := t_1, \ldots, t_k]$. By induction hypothesis, each t_j is in the language of L_1. If t_0 is in the language of L, then clearly t is again in the language of L_1. The interesting case is when t_0 is in the language of $P = (\sum_{i=1}^n C_i)^{*,\Box'}._{\Box'} P_0$. Then t_0 is in $(\sum_{i=1}^n C_i)._{\Box'} (\sum_{i=1}^n C_i)._{\Box'} \ldots ._{\Box'} (\sum_{i=1}^n C_i)._{\Box'} P_0$ (q times, for some $q \in \mathbb{N}$). We show that $t_0[\Box := t_1, \ldots, t_k]$ is in L_1 by a second induction on q. If $q = 0$, then t_0 is in P_0, so that $t_0[\Box := t_1, \ldots, t_k]$ is in $P_0._{\Box'} L_1$, hence in L_1. Otherwise, we can write t_0 as $u_0[\Box' := u_1, \ldots, u_\ell]$ for some term u_0 in the language of some product iterator C_i, $1 \leq i \leq n$, with ℓ occurrences of $\Box'$, and where $u_1, \ldots, u_\ell$ are in $(\sum_{i=1}^n C_i)._{\Box'} (\sum_{i=1}^n C_i)._{\Box'} \ldots ._{\Box'} (\sum_{i=1}^n C_i)._{\Box'} P_0$ ($q-1$ times). By induction hypothesis, each u_j is in L_1. Assume u_0 has ℓ' (other) occurrences of $\Box$. Let u_0' be u_0 where each occurrence of $\Box'$ is replaced by $\Box$. Then u_0' is in the language of $C_i[Box' := \Box]$. Moreover, $u_0[\Box' := u_1, \ldots, u_\ell]$ is obtained from u_0' by replacing ℓ occurrences of $\Box$ by u_1, …, u_ℓ, and not replacing the others, i.e., replacing them by $\Box$ itself. However, it is easy to see that $\Box$ is in the language of L_1. Since each u_j is also in L_1, $u_0[\Box' := u_1, \ldots, u_\ell]$ is in $C_i[\Box' := \Box]._\Box L_1$, hence in L_1.

Now assume t is in $L_1 = (\sum_{i=1}^n C_i[\Box' := \Box] + P_0 + L)^{*,\Box}$. So t is in $(\sum_{i=1}^n C_i[\Box' := \Box] + P_0 + L)._\Box \ldots ._\Box (\sum_{i=1}^n C_i[\Box' := \Box] + P_0 + L)._\Box \Box$ (p times, for some $p \in \mathbb{N}$). We show that t is in $L_0 = (P+L)^{*,\Box}$ by induction on p. If $p = 0$, then $t = \Box$, so $t \in L_0$. Otherwise,

$t = t_0[\Box := t_1, \ldots, t_m]$ for some term t_0 in $\sum_{i=1}^n C_i[\Box' := \Box] + P_0 + L$ with m occurrences of $\Box$, and terms $t_1, \ldots, t_m$ in L_1. By induction hypothesis, $t_1, \ldots, t_m$ are in L_0. If t_0 is in L, then t is in $L.{}_\Box L_0$, hence in L_0. If t_0 is in P_0, then it is in $P = (\sum_{i=1}^n C_i)^{*,\Box'}.{}_{\Box'} P_0$, so t is in $P.{}_\Box L_0$, hence in L_0. The interesting case is when t_0 is in some $C_i[\Box' := \Box]$ for some i, $1 \leq i \leq n$. One checks that t_0 is then of the form $u_0[\Box' := \Box]$ for some term u_0 in the language of C_i. (Write C_i as $f(P_1, \ldots, P_k)$. Up to inessential permutation of the arguments, we may assume that $P_1 = \ldots = P_{\ell-1} = \Box$, $P_\ell = \ldots = P_{\ell+\ell'-1} = \Box'$, and $\Box'$ is not free in $P_{\ell+\ell'}, \ldots, P_k$. Then t_0 is of the form $f(\underbrace{\Box, \ldots, \Box}_{\ell+\ell'}, u_{\ell+\ell'+1}, \ldots, u_k)$, and we let $u_0 = f(\underbrace{\Box, \ldots, \Box}_{\ell}, \underbrace{\Box', \ldots, \Box'}_{\ell'}, u_{\ell+\ell'+1}, \ldots, u_k)$.) Now $t_0 = u_0[\Box' := \Box]$ is in $C_i.{}_{\Box'} P_0$, because by assumption $\Box$ is in P_0. So t_0 is in the language of P. Then $t = t_0[\Box := t_1, \ldots, t_m]$ is in $P.{}_\Box L_0$, hence in L_0. ■

Lemma D.8. *Let L be an STRE, and $\Box$ be a hole. Then there is an STRE L'' that defines the same language as $L^{*,\Box}$. Moreover,* $\mathrm{fh}(L'') \subseteq \mathrm{fh}(L^{*,\Box})$.

Proof. Call a *quasi-iterator* C^0 over $\Box$ any tree regular expression of the form $f(P_1, \ldots, P_k)$, $f \in \Sigma_k$, that satisfies property 2 of product iterators, but not necessarily property 1. The *defect* of C^0 is the sum of the sizes of those products P_k that are different from $\Box$ yet contain $\Box$ in their language. The *size* of a tree product is defined as follows. the size of $\Box$ is 1, that of $f^?(P_1, \ldots, P_k)$ and of $f(P_1, \ldots, P_k)$ is one plus the sum of the sizes of $P_1, \ldots, P_k$, and the size of $(\sum_{j=1}^m C_j^0)^{*,\Box'}.{}_{\Box'} P_0$ is one plus the sum of the sizes of $C_1^0, \ldots, C_m^0, P_0$. The size of an STRE is the sum of the sizes of its tree products.

We prove the more general statement that $(\sum_{j=1}^m C_j^0 + L)^{*,\Box}$ is definable by an STRE, for any STRE L, and for any quasi-iterators $C_1^0, \ldots, C_m^0$ over $\Box$. We prove this by induction on the sum of the defects of $C^0, \ldots, C^m$ and of the size of L.

If this sum is zero, then $(\sum_{j=1}^m C_j^0 + L)^{*,\Box}$ is definable by the tree product $(\sum_{j=1}^m C_j^0 + L)^{*,\Box}.{}_\Box \Box$, since L is the empty sum.

If some C_j^0 is not a product iterator, say $j = 1$, then $C_1^0 = f(P_1, \ldots, P_k)$, and we apply Lemma D.7, item 5. Using the notations used there, $(\sum_{j=1}^m C_j^0 + L)^{*,\Box}$ defines the same language as $(\sum_{j=2}^m C_j^0 + P' + P_{i_1} + \ldots + P_{i_\ell} + L)^{*,\Box}$. Let $C_{m+1}^0 = P'$, and note that this is a product iterator over $\Box$. So $(\sum_{j=1}^m C_j^0 + L)^{*,\Box}$ defines the same language as $(\sum_{j=2}^{m+1} C_j^0 + P_{i_1} + \ldots + P_{i_\ell} + L)^{*,\Box}$. Since $P_i = \Box$ for some i, $1 \leq i \leq k$, $\ell < k$, so the measure of the latter expression is less than the measure of $(\sum_{j=1}^m C_j^0 + L)^{*,\Box}$ by at least the size of P_i. We can therefore apply the induction hypothesis.

Otherwise, every C_j^0 is a product iterator over $\Box$, and L is not the empty sum. L can be written as a tree product P, or as a sum $P + L'$. Without loss of generality, we may assume L is $P + L'$, since P defines the same language and has the same size as $P + \emptyset$.

If $\Box$ is not in the language of P, then by Lemma D.7, item 1, $(\sum_{j=1}^m C_j^0 + L)^{*,\Box}$ defines the same language as $(\sum_{j=1}^m C_j^0 + L')^{*,\Box}.{}_\Box P + (\sum_{j=1}^m C_j^0 + L')^{*,\Box}$. By induction hypothesis, there is an SLRE L'' defining the same language as $(\sum_{j=1}^m C_j^0 + L')^{*,\Box}$. By Lemma D.6, there is an SLRE L''' defining the same language as $L''.{}_\Box P$. Then $L''' + L''$ fits the bill.

Otherwise, $\Box$ is in the language of P, and we distinguish three cases.

If $P = \Box$, then $(\sum_{j=1}^m C_j^0 + L)^{*,\Box}$ defines the same language as $(\sum_{j=1}^m C_j^0 + L')^{*,\Box}$ by Lemma D.7, item 2, and we conclude by the induction hypothesis.

If P is of the form $f^?(P_1, \ldots, P_k)$, then $(\sum_{j=1}^m C_j^0 + L)^{*,\Box}$ defines the same language as $(\sum_{j=1}^m C_j^0 + P' + P_1 + \ldots + P_k + L')^{*,\Box}$ by Lemma D.7, item 3, using the notations introduced there.

Note that P' is a product iterator over $\Box$: property 1 is by construction, and property 2 is because $\Box$ is in the language of P, so it must be in the language of some P_i, whence $\ell \neq 0$. Writing C^0_{m+1} for P', it follows that $(\sum_{j=1}^m C^0_j + L)^{*,\Box}$ defines the same language as $(\sum_{j=1}^{m+1} C^0_j + P_1 + \ldots + P_k + L')^{*,\Box}$. Since the size of $P_1 + \ldots + P_k + L'$ is smaller than that of L, we conclude using the induction hypothesis.

Finally, if P is of the form $(\sum_{i=1}^m C_i)^{*,\Box'} \cdot_{\Box'} P_0$, we may first assume that $\Box' \neq \Box$ by α-renaming. Also, since $\Box$ is in the language of P, $\Box$ must also be in the language of P_0. So we may apply Lemma D.7, item 4: $(\sum_{j=1}^m C^0_j + L)^{*,\Box}$ defines the same language as $(\sum_{j=1}^m C^0_j + \sum_{i=1}^n C_i[\Box' := \Box] + P_0 + L')^{*,\Box}$. Note that $C_i[\Box' := \Box]$ is *not* in general a product iterator over $\Box$: for example, if $C_i = f(\Box', g(\Box))$, then $C_i[\Box' := \Box] = f(\Box, g(\Box))$. This is the reason why we are using quasi-iterators. Let $C^0_{m+1} = C_1[\Box' := \Box]$, …, $C^0_{m+n} = C_n[\Box' := \Box]$. One may check that the measure of $(\sum_{j=1}^m C^0_j + \sum_{i=1}^n C_i[\Box' := \Box] + P_0 + L')^{*,\Box}$, i.e., of $(\sum_{j=1}^{m+n} C^0_j + P_0 + L')^{*,\Box}$ is stricly less than that of $(\sum_{j=1}^m C^0_j + P + L')^{*,\Box}$. We therefore conclude by the induction hypothesis. ∎

Proposition D.9. *Any STRE (resp., closed STRE) defines a downward closed subset of $\mathcal{T}(\mathcal{F} \cup \mathcal{K})$ (resp., of $\mathcal{T}(\mathcal{F})$). Conversely, any downward closed subset of $\mathcal{T}(\mathcal{F} \cup \mathcal{K})$ (resp., of $\mathcal{T}(\mathcal{F})$) is the language of some (closed) STRE.*

Proof. The first claim is clear. Conversely, any downward closed set of $\mathcal{T}(\mathcal{F} \cup \mathcal{K})$ (resp., $\mathcal{T}(\mathcal{F})$) is the language of some (closed) regular expression L as given in Lemma D.3. We now induct on L to show that this is also the language of some STRE L'' with $\mathrm{fh}(L'') \subseteq \mathrm{fh}(L)$. First, we use the following trick, to simplify the presentation. We can always assume that any subexpression of L'' of the form $f^?(L_1, \ldots, L_p)$ is in fact such that L_1, …, L_p are holes. Indeed, we can always replace $f^?(L_1, \ldots, L_p)$ by the tree regular expression $f^?(\Box_1, \ldots, \Box_p) \cdot_{\Box_1} L_1 \ldots \cdot_{\Box_p} L_p$, where $\Box_1$, …, $\Box_p$ are fresh disjoint holes. This defines the same language.

So let us induct on L, under this simplification. When L is of the form $f^?(\Box_1, \ldots, \Box_p)$ or $\Box$, L is already a tree product. When L is $\emptyset$, L is an STRE. When L is a sum $L_1 + L_2$, we appeal to the induction hypothesis. When $L = L_1 \cdot_\Box L_2$, we conclude by Lemma D.6 and the induction hypothesis. When $L = L^{*,\Box}$, we conclude by Lemma D.8 instead. ∎

In other words, $\mathcal{H}(\mathcal{T}(\mathcal{F})_a)$ is the space of languages defined by STREs, ordered by inclusion.

Theorem D.10. *$\mathcal{S}(\mathcal{T}(\mathcal{F})_a)$ is the set of languages defined by closed tree products, ordered by inclusion.*

Proof. Take any irreducible closed subset F of $\mathcal{T}(\mathcal{F})$. By Proposition D.9, F can be expressed as an STRE $P_1 + \ldots + P_k$. Since F is irreducible, $k \leq 1$; since every irreducible closed set is non-empty by definition, $k \neq 0$. So F is definable by a tree product. Note that the language of a tree product is never empty.

Conversely, we must show that the language of any tree product P is irreducible closed. It is clearly downward closed, i.e., closed. We shall show that it is in fact a directed set. Because $\mathcal{S}(\Sigma^*_a) = Idl(\Sigma^*)$, and since the language of P is clearly (downward) closed, directedness of P is equivalent to it being irreducible closed. However, the fact that any downward closed directed subset is irreducible closed is elementary, so we prove it here. Let F be downward closed, directed. Assume $F \subseteq F_1 \cup F_2$, where F_1, F_2 are closed. we must show that $F \subseteq F_1$ or $F \subseteq F_2$. If on the contrary there were $x_1 \in F \setminus F_1$ and $x_2 \in F \setminus F_2$, there would be $x \in F$ with $x_1, x_2 \leq x$ by directedness. Now x is either in F_1 or in F_2. Say $x \in F_1$. Since F_1 is downward closed, x_1 is in F_1, too, contradiction. Similarly if $x \in F_2$. So F is irreducible.

So we show that P is directed, by induction on P.

Clearly the language of $\Box$ is directed. If $P = f^?(P_1, \ldots, P_k)$, and t, t' are two terms in the language of P, we must consider several cases. If t and t' are in the language of the same P_i, then by induction there is another term t'' such that $t, t' \trianglelefteq t''$, in the language of P_i, hence in that of P. If t is in the language of P_i and t' is in the language of P_j with $j \neq i$, then $f(t_1, \ldots, t_{i-1}, t, t_{i+1}, \ldots, t_{j-1}, t', t_{j+1}, \ldots, t_k)$ is in the language of P, where the only occurrence of t is at position i, the only occurrence of t' is at position j, and the terms t_ℓ ($\ell \neq i, j$) are taken from the languages of P_ℓ respectively. (Recall these languages are non-empty.) Clearly this is a term t'' such that $t, t' \trianglelefteq t''$. If t is in the language of P_i and t' is in the language of $f(P_1, \ldots, P_k)$, i.e., $t' = f(t_1, \ldots, t_k)$ where each t_j is in the language of P_j, $1 \leq j \leq k$, then since P_i is directed, there is a term t'_i with $t, t_i \trianglelefteq t'_i$ in the language of P_i. Then $t'' = f(t_1, \ldots, t_{i-1}, t'_i, t_{i+1}, \ldots, t_k)$ is in $f(P_1, \ldots, P_k)$, hence in P, and $t, t' \trianglelefteq t''$. The case where t is in the language of $f(P_1, \ldots, P_k)$ and t' is in some P_i is symmetrical. Finally, if t and t' both are in $f(P_1, \ldots, P_k)$, then we can write $t = f(t_1, \ldots, t_k)$, $t' = f(t'_1, \ldots, t'_k)$, with t_i, t'_i in P_i for each i. Since each P_i is directed, there are terms t''_i in P_i such that $t_i, t'_i \trianglelefteq t''_i$. Then take $t'' = f(t''_1, \ldots, t''_k)$: this is in P, and $t, t' \trianglelefteq t''$.

Finally, if $P = (\sum_{i=1}^n C_i)^{*,\Box}.{}_\Box P_0$, where P is a tree product and $C_1, \ldots, C_n$ are product iterators over $\Box$, and t is in the language of P, then there is a term u in the language of $\sum_{i=1}^n C_i$, with m occurrences of the hole $\Box$, and terms $u_1, \ldots, u_m$ in the language of P_0 such that $t = u[\Box := u_1, \ldots, u_m]$. (We reuse a notation introduced in the proof of Lemma D.7.) For any other term t' in the language of P, we construct similarly u' and $u'_1, \ldots, u'_{m'}$ so that $t' = u'[\Box := u'_1, \ldots, u'_{m'}]$. Note that m and m' are both non-zero. This rests on property 2 of product iterators. Let now u'' equal $u[\Box := u', \ldots, u']$. This is a term with mm' occurrences of $\Box$. Each can be described as the kth occurrence of $\Box$ in the jth occurrence of u', $1 \leq j \leq m$, $1 \leq j \leq m'$. For each j and k, there is a term u''_{jk} in the language of P_0 such that $u_j, u'_k \trianglelefteq u''_{jk}$. Then define t'' as obtained from u'' by replacing the j, k occurrence of $\Box$ by u''_{jk}. Since $m \neq 0$ and $m' \neq 0$, we obtain $t, t' \trianglelefteq t''$. Also, by construction t'' is in the language of $P = (\sum_{i=1}^n C_i)^{*,\Box}.{}_\Box P_0$. ■

Testing the inclusion of closed tree products is more complex than testing the inclusion of products over words. This is computed by way of the following lemmas. Write $P \subseteq P'$, by abuse of language, for "the language of P is contained in that of P'".

Lemma D.11. *The language of the tree product $f^?(P_1, \ldots, P_m)$ is included in that of the tree product $g^?(P'_1, \ldots, P'_n)$ if and only if:*

- *either $f \neq g$, and $f^?(P_1, \ldots, P_m) \subseteq P'_j$ for some j, $1 \leq j \neq n$;*
- *or $f = g$, $m = n$, and then either $f^?(P_1, \ldots, P_m) \subseteq P'_j$ for some j, $1 \leq j \neq n$, or $P_i \subseteq P'_i$ for all i, $1 \leq i \leq m$.*

Proof. The if direction is clear. For example, if $f^?(P_1, \ldots, P_m) \subseteq P'_j$, then also $f^?(P_1, \ldots, P_m) \subseteq g^?(P'_1, \ldots, P'_n)$, since $g^?(P'_1, \ldots, P'_n) = g(P'_1, \ldots, P'_n) + P'_1 + \ldots + P'_n$.

Conversely, assume $f^?(P_1, \ldots, P_m) \subseteq g^?(P'_1, \ldots, P'_n)$, i.e., that $f(P_1, \ldots, P_m) \subseteq g^?(P'_1, \ldots, P'_n)$, since the latter language is downward closed. Assume also that $f^?(P_1, \ldots, P_m)$ is not contained in any P'_j, for any j. Again, this is equivalent to assuming that $f(P_1, \ldots, P_m)$ is not contained in any P'_j, for any j. So there is a term t_j in $f(P_1, \ldots, P_m)$ which is not in P'_j, for all j, $1 \leq j \leq n$. Note that $f(P_1, \ldots, P_m)$ is directed, so there is a term t in $f(P_1, \ldots, P_m)$ such that $t_1, \ldots, t_n \trianglelefteq t$. (If $n = 0$, we take *any* term in $f(P_1, \ldots, P_m)$.) For each j, since $t_j \trianglelefteq t$, t_j is not in P'_j, and P'_j is downward closed, t cannot be in P'_j either.

Now if $f \neq g$, since t is in $f(P_1, \ldots, P_m) \subseteq g^?(P'_1, \ldots, P'_n)$, it must be the case that t is in some P'_j, contradiction. This proves the first case.

If $f = g$, write t as $f(u_1, \ldots, u_m)$, with u_1 in P_1, …, u_m in P_m. Assume by contradiction that for some i, P_i is not contained in P'_i. Then there is a term v in P_i that is not in P'_i. Since P_i is directed, there is a term w_i in P_i with $u_i, v \trianglelefteq w_i$. For all $j \neq i$, define w_j as u_j, and consider $w = f(w_1, \ldots, w_m)$. This is a term in $f(P_1, \ldots, P_m)$, hence in $f^?(P'_1, \ldots, P'_m)$. If w were in some P'_j, then t, which satisfies $t \trianglelefteq w$, would be in P'_j, too, but t was constructed not to be in any P'_j. So w is in $f(P'_1, \ldots, P'_m)$. That is, each w_j is in P'_j, $1 \leq j \leq m$. However, for $j = i$, this entails that w_i is in P'_i, hence that v is in P'_i since the latter is downward closed: contradiction. ∎

Lemma D.12. *The language of the tree product $f^?(P_1, \ldots, P_m)$ is included in that of the tree product $(\sum_{j=1}^n C_j)^{*,\Box}.{}_\Box P'$ if and only if:*

- *either $f^?(P_1, \ldots, P_m) \subseteq P'$;*
- *or for some j, $1 \leq j \leq n$, C_j can be written $f(P'_1, \ldots, P'_m)$, so that for all i, $1 \leq i \leq m$:*
  - *$P'_i = \Box$ and $P_i \subseteq (\sum_{j=1}^n C_j)^{*,\Box}.{}_\Box P'$;*
  - *or $P'_i \neq \Box$ and $P_i \subseteq P'_i$.*

Proof. As in the previous lemma, the if direction is easy. This is left as an exercise, and uses property 1 of product iterators. Conversely, assume by contradiction that $f^?(P_1, \ldots, P_m)$ is not included in P', equivalently that $f(P_1, \ldots, P_m)$ is not included in P'. Let t be a term in $f(P_1, \ldots, P_m)$ that is not in P'.

Again for the sake of contradiction, assume that for all j, $1 \leq j \leq n$, such that C_j has head symbol f, say $C_j = f(P'_{j1}, \ldots, P'_{jm})$, then there is a subscript $i = i_j$, $1 \leq i_j \leq m$ such that either $P'_i = \Box$ and P_i is not contained in $(\sum_{j=1}^n C_j)^{*,\Box}.{}_\Box P'$, or $P'_i \neq \Box$ and P_i is not contained in P'_i.

We observe that the first case (when $P'_i = \Box$) cannot happen: otherwise there would be a term w_i in P_i not in $(\sum_{j=1}^n C_j)^{*,\Box}.{}_\Box P'$; letting, for each $i' \neq i$, $w_{i'}$ be an arbitrary term in $P_{i'}$, the term $f(w_1, \ldots, w_m)$ would be in $f(P_1, \ldots, P_m)$, hence in $(\sum_{j=1}^n C_j)^{*,\Box}.{}_\Box P'$; since the latter is downward closed and $w_i \trianglelefteq f(w_1, \ldots, w_m)$, w_i would also be in $(\sum_{j=1}^n C_j)^{*,\Box}.{}_\Box P'$, contradiction.

So, for each j, $1 \leq j \leq n$, such that C_j has head symbol f, $P'_{i_j} \neq \Box$, and P_{i_j} is not contained in P'_{i_j}. Let therefore w_j be a term in P_{i_j} but not in P'_{i_j}. Now, for each i, $1 \leq i \leq m$, since P_i is directed, we may build a term u_i in P_i such that $w_j \trianglelefteq u_i$ whenever j is such that $i_j = i$. (In case no such w_j exists, we take an arbitrary u_i from P_i.) Let $u = f(u_1, \ldots, u_n)$, a term in $f(P_1, \ldots, P_m)$. Since t is also in $f(P_1, \ldots, P_m)$ and $f(P_1, \ldots, P_m)$ is directed, we may find a term v in $f(P_1, \ldots, P_m)$ with $t, u \trianglelefteq v$. Since t is not P' and P' is downward closed, v is not in P' either. But since v is in $f(P_1, \ldots, P_m)$, v is in $(\sum_{j=1}^n C_j)^{*,\Box}.{}_\Box P'$. So there is a j, $1 \leq j \leq n$, where C_j has head symbol f, such that v is in $C_j.{}_\Box(\sum_{j=1}^n C_j)^{*,\Box}.{}_\Box P'$. Since $u \trianglelefteq v$, u is also in the latter language. However $u = f(u_1, \ldots, u_n)$, so u_{i_j} must be in P'_{i_j} (remember i_j is a position i such that $P'_i \neq \Box$, hence u_{i_j} is in $P'_{i_j}.{}_\Box(\sum_{j=1}^n C_j)^{*,\Box}.{}_\Box P'$, which defines the same language as P'_{i_j} by property 1 of product iterators). By construction, $w_j \trianglelefteq u_{i_j}$, so w_j is also in P'_{i_j}, since P'_{i_j} is downward-closed: contradiction. ∎

Lemma D.13. *The language of the tree product $(\sum_{i=1}^m C_i)^{*,\Box}.{}_\Box P$ is included in that of the tree product $g^?(P'_1, \ldots, P'_n)$ if and only if it is included in the language of P'_j for some j, $1 \leq j \leq n$.*

Proof. Otherwise, let t_j be a term in $(\sum_{i=1}^m C_i)^{*,\Box}.{}_\Box P$ that is not in P'_j, for each j, $1 \leq j \leq n$. Since $(\sum_{i=1}^m C_i)^{*,\Box}.{}_\Box P$ is directed, there is a term t in $(\sum_{i=1}^m C_i)^{*,\Box}.{}_\Box P$ such that $t_1, \ldots, t_n \trianglelefteq t$. (If $n = 0$, take t arbitrary in $(\sum_{i=1}^m C_i)^{*,\Box}.{}_\Box P$.) Since each P'_j is downward closed, t is not in P'_j either. Recall that $m \geq 1$, so C_1 exists. By property 2 of product iterators, we may write C_1 as $f(P_1, \ldots, P_k)$ where $P_{i_0} = \Box$ for some i_0, $1 \leq i_0 \leq k$. Build the term $f(u_1, \ldots, u_k)$, where: if $P_i = \Box$ (in particular for $i = i_0$), $u_i = t$; otherwise, let u_i be an arbitrary term of P_i. Using property 1 of product iterators, $u = f(u_1, \ldots, u_k)$ is in $C_1.{}_\Box(\sum_{i=1}^m C_i)^{*,\Box}.{}_\Box P)$, hence in $\sum_{i=1}^m C_i)^{*,\Box}.{}_\Box P$. Also, u cannot be in any P'_j, $1 \leq j \leq n$, otherwise $t = u_{i_0}$ would also be in P'_j. By assumption, u must be in $g^?(P'_1, \ldots, P'_n)$, and since it is in no P'_j, u must be of the form $g(u'_1, \ldots, u'_n)$ with u'_j in P'_j for each j. This not only forces $g = f$, but also that $u'_j = u_j$ for all j; however u'_{i_0} is in P'_{i_0}, is equal to $u_{i_0} = t$, which is in no P'_j, contradiction. ∎

Lemma D.14. *The language of the tree product $(\sum_{i=1}^m C_i)^{*,\Box}.{}_\Box P$ is included in that of the tree product $(\sum_{j=1}^n C'_j)^{*,\Box'}.{}_{\Box'} P'$ if and only if, for every i, $1 \leq i \leq m$, writing C_i as $f(P_1, \ldots, P_k)$:*

- *either $f^?(P''_1, \ldots, P''_k) \subseteq P'$, where for each p, $1 \leq p \leq k$, $P''_p = \sum_{i=1}^m C_i)^{*,\Box}.{}_\Box P$ if $P_p = \Box$, and $P''_p = P_p$ otherwise;*
- *or for every there is a j, $1 \leq j \leq n$, such that C'_j is of the form $f(P'_1, \ldots, P'_k)$ and for every p, $1 \leq p \leq k$:*
  - *if $P'_p = \Box'$, then $P''_p \subseteq (\sum_{j=1}^n C'_j)^{*,\Box'}.{}_{\Box'} P'$;*
  - *if $P'_p \neq \Box'$, then $P''_p \subseteq P'_p$.*

Proof. Note that $f(P''_1, \ldots, P''_k)$ defines the same language as $C_i.{}_\Box(\sum_{j=1}^n C'_j)^{*,\Box'}.{}_{\Box'} P'$. This uses properties 1 and 2 of product iterators. The only difficult direction is the only if direction.

Assume by contradiction that there is an i, $1 \leq i \leq m$, with C_i written as $f(P_1, \ldots, P_k)$, such that, first, $f^?(P''_1, \ldots, P''_k)$ is not contained in P'; in particular, $f(P''_1, \ldots, P''_k)$ is not contained in P', so there is a term $t = f(t_1, \ldots, t_k)$ in $f(P''_1, \ldots, P''_k)$ but not in P'. Second, we assume that for every j, $1 \leq j \leq n$, with C'_j of the form $f(P'_1, \ldots, P'_k)$, there is an index $p = p_j$, $1 \leq p \leq k$, with:

- either $P'_p \neq \Box$, and there is a term t'_{p_j} in P''_p but not in $(\sum_{j=1}^n C'_j)^{*,\Box'}.{}_{\Box'} P'$;
- or $P'_p = \Box'$, and there is a term t'_{p_j} in P''_p but not in P'_p.

For each p, $1 \leq p \leq k$, t_p and every t'_{p_j} with $p_j = p$ is in P''_p. Let u_p be a term in P''_p such that $t_p \trianglelefteq u_p$, and $t'_{p_j} \trianglelefteq u_p$ for every j such that $p_j = p$. This is possible since P''_p is directed. Then let $u = f(u_1, \ldots, u_k)$, so that u is in $f(P''_1, \ldots, P''_k)$; in particular, u is in $C_i.{}_\Box(\sum_{i=1}^m C_i)^{*,\Box}.{}_\Box P$, hence in $(\sum_{i=1}^m C_i)^{*,\Box}.{}_\Box P$. By assumption u is also in $(\sum_{j=1}^n C'_j)^{*,\Box'}.{}_{\Box'} P'$. Now $t \trianglelefteq u$ since $t_p \trianglelefteq u_p$ for each p; since t is not in P', and P' is downward closed, u is not in P' either. So u is in $C'_j.{}_{\Box'}(\sum_{j=1}^n C'_j)^{*,\Box'}.{}_{\Box'} P'$ for some j, $1 \leq j \leq n$.

Consider $p = p_j$. If $P'_p = \Box'$, then u_p is in $(\sum_{j=1}^n C'_j)^{*,\Box'}.{}_{\Box'} P'$. Since $t'_{p_j} \trianglelefteq u_p$, t'_{p_j} is also in $(\sum_{j=1}^n C'_j)^{*,\Box'}.{}_{\Box'} P'$: contradiction. If $P'_p \neq \Box'$, then u_p is in P'_p. Since $t'_{p_j} \trianglelefteq u_p$, t'_{p_j} is also in P'_p: contradiction again. ∎

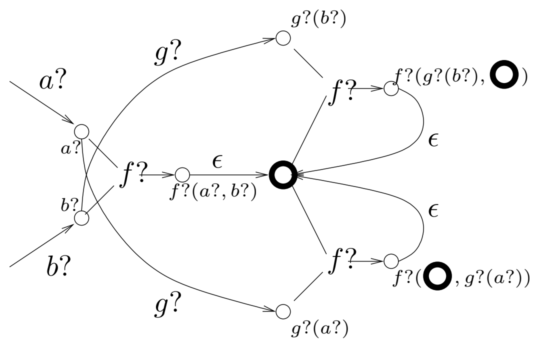
These four lemmas allow us to decide the inclusion of tree products. We represent tree products in a tree-automata-like notation, where transitions between vertices are labeled by symbols $f^?$, where f is a function symbol, or by boxes $\Box$. If f has arity k, then the corresponding transition takes k vertices as input, and has one vertex as output. Boxes are thought of as having arity 0. We also allow for ϵ-transitions from vertices to vertices. We equate vertices with specific tree products. The set of tree products used as vertices of the hypergraph for the tree product P is not the set of subexpressions of P, rather it is a larger set, reminiscent of the notion of Fisher-Ladner closure from modal logic. Explicitly,

Definition D.15. For every tree product P, we build a hypergraph G_P as follows:

- If $P = \Box$, then G_P has one vertex, $\Box$, and one 0-ary transition labeled $\Box$ reaching it;

- If $P = f^?(P_1, \dots, P_k)$, then G_P is obtained from G_{P_1}, …, G_{P_k} by adding a transition labeled $f^?$ from the tuple of vertices P_1, …, P_k to the state P;
- If $P = (\sum_{i=1}^n C_i)^{*,\Box}.{}_\Box P_0$, where $C_i = f_i(P_{i1}, \dots, P_{ik_i})$ for each i, $1 \leq i \leq n$, then G_P is obtained from G_{P_0}, as well as from the hypergraphs $G_{P_{ij}}$ for all i, j with $P_{ij} \neq \Box$, as follows. First, add an ϵ-transition from P_0 to P. Then, for each pair i, j let P'_{ij} be the vertex P_{ij} if $P_{ij} \neq \Box$, otherwise $P'_{ij} = P$. Then, for each i, $1 \leq i \leq n$, add transitions $f_i?$ from the k_i-tuple of vertices P'_{i1}, …, P'_{ik_i} to the vertex $f_i?(P'_{i1}, \dots, P'_{ik_i})$, and an ϵ-transition from the latter to P.

For example, the hypergraph of $(f(\Box, g^?(a^?)) + f(g^?(b^?), \Box))^{*,\Box}.{}_\Box f^?(a^?, b^?)$ would be:



whose root is shown as the big circle ○, and is the vertex $(f(\Box, g^?(a^?)) + f(g^?(b^?), \Box))^{*,\Box}.{}_\Box f^?(a^?, b^?)$ itself.

Proposition D.16. *For any tree products P and P', we can check whether the language of P is contained in that of P' in time $O(m^2 m'^2)$, where m is the size of G_P, m' is the size of $G_{P'}$.*

Proof. By dynamic programming. Let n be the number of vertices in G_P, n' the number of vertices in $G_{P'}$. We allocate a Boolean array of nn' entries, where $a[\ell, \ell']$ will denote whether the tree product at vertex ℓ of G_P is contained in that at vertex ℓ' of $G_{P'}$. We order these vertices by a topological ordering, i.e., such that for any vertex ℓ of G_P, any subformula of this vertex occurs at indices at most ℓ, and similarly for $G_{P'}$. Then we fill in the a array in increasing order of ℓ', and for fixed ℓ', in increasing order of ℓ, using Lemma D.11, Lemma D.12, Lemma D.13, and Lemma D.14. We deal with just one case that requires the latter lemma, to show one subtlety. Assume vertex ℓ of G_P is of the form $(\sum_{i=1}^m C_i)^{*,\Box}.{}_\Box P_0$, and vertex ℓ' of $G_{P'}$ is of the form $(\sum_{j=1}^n C'_j)^{*,\Box'}.{}_{\Box'} P'_0$. Then, for each i, $1 \leq j \leq n$, we need to test the inclusions of $f^?(P''_1, \dots, P''_k)$ inside P'_0 first, where for each p, $1 \leq p \leq k$, $P''_p = \sum_{i=1}^m C_i)^{*,\Box}.{}_\Box P$ if $P_p = \Box$, and $P''_p = P_p$ otherwise. Note that $f^?(P''_1, \dots, P''_k)$ occurs as a vertex, say ℓ_1, in the graph G_P, however ℓ_1 is not necessarily smaller than or equal to ℓ. But P'_0 does occur with an index ℓ'_1, strictly less than ℓ', so the entry $a[\ell_1, \ell'_1]$ has already been filled in. We must then test whether there is a j such that certain conditions hold (see second item of Lemma D.14), all involving entries $a[\ell_1, \ell'_1]$ with ℓ'_1 strictly less than ℓ'.

The a array contains $nn' \leq mm'$ entries, and each can be filled using at most mm' operations, whence the $O(m^2 m'^2)$ complexity. ∎

Appendix E. Words, and a Topological Variant of Higman's Lemma

We show that, when X is Noetherian, then the set X^* of finite words over X, with a suitable topology, is Noetherian again, and that its sobrification consists of natural analogues of the notion of products used in SREs, built on an alphabet of points in $\mathcal{S}(X)$. Note this is only interesting when the alphabet X is infinite, and suitably topologized. For example, we may take X to be the set of all vectors in $\mathbb{N}^k$.

For any topological space X, let X^* be the set of all finite words over X. We write ϵ for the empty word, ww' the concatenation of the words w and w'; we also use ambiguously a for a letter (in X) and for the corresponding one-letter word. Whether we mean a letter or a word will be disambiguated by context, and by the convention that a, b, …, denote letters, while w, w', …, denote words. Whenever A and B are subsets of X, we also write AB the set $\{ww' \mid w \in A, w' \in B\}$ of all concatenations of a word in A with a word in B.

The right topology on X^* is defined as follows. We call it the *subword topology*:

Definition E.1 (Subword Topology)**.** The *subword topology* on X^* is the least one containing the subsets $X^*U_1X^*U_2X^* \dots X^*U_nX^*$ as opens, where $n \in \mathbb{N}$, and U_1, U_2, …, U_n are open subsets of X.

We shall see later that, if $\leq$ is the specialization quasi-ordering of X, then the embedding quasi-ordering $\leq^*$ is the specialization quasi-ordering of X^* with the subword topology. Remember that $\leq^*$ is defined by: $w \leq^* w'$ iff, writing w as the sequence of m letters $a_1a_2 \dots a_m$, one can write w' as $w_0a'_1w_1a'_2w_2 \dots w_{m-1}a'_mw'_m$ with $a_1 \leq a'_1$, $a_2 \leq a'_2$, …, $a_m \leq a'_m$. Higman's Lemma states that if X is well-quasi-ordered by $\leq$, then X^* is well-quasi-ordered by $\leq^*$.

Any open is, by definition a union of finite intersections of opens of the form $X^*U_1X^*U_2X^* \dots X^*U_nX^*$. One may simplify this statement:

Lemma E.2. *The subsets $X^*U_1X^*U_2X^* \dots X^*U_nX^*$ as defined in Definition E.1 form a* basis *of the subword topology: any open is a union of such opens. We call them the* basic opens.

Proof. Let w a word in the intersection of $X^*U_1X^*U_2X^* \dots X^*U_mX^*$ and $X^*V_1X^*V_2X^* \dots X^*V_nX^*$. That is, w contains a subword $a_1a_2 \dots a_m \leq^* w$, where $a_1 \in U_1$, $a_2 \in U_2$, …, $a_m \in U_m$. Let $I = \{\iota_1, \iota_2, \dots, \iota_m\}$ be the set of positions where the letters a_i can be found; i.e., a_1 is the letter at position ι_1 in w, a_2 is the letter at position $\iota_2 > \iota_1$ in w, and so on. Also, w contains a subword $b_1b_2 \dots b_n \leq^* w$ where $b_1 \in V_1$, $b_2 \in V_2$, …, $b_n \in V_n$. Let $J = \{\eta_1, \eta_2, \dots, \eta_n\}$ be the set of positions where the letters b_j can be found. Now let $\kappa_1 < \kappa_2 < \dots < \kappa_p$ be the increasing sequence of positions in $I \cup J$, and consider the open subset $X^*W_1X^*W_2X^* \dots X^*W_pX^*$, where for each k, W_k equals $U_i \cap V_j$ if $k \in I \cap J$ (where i, j are defined by $\kappa_k = \iota_i = \eta_j$), U_i if $k \in I \setminus J$ (where $\kappa_k = \iota_i$), and V_j if $k \in J \setminus I$ (where $\kappa_k = \eta_j$). Clearly w is in $X^*W_1X^*W_2X^* \dots X^*W_pX^*$, and the latter is contained in both $X^*U_1X^*U_2X^* \dots X^*U_mX^*$ and $X^*V_1X^*V_2X^* \dots X^*V_nX^*$.

It follows that the intersection of $X^*U_1X^*U_2X^* \dots X^*U_mX^*$ and $X^*V_1X^*V_2X^* \dots X^*V_nX^*$ if the union of the thus obtained sets $X^*W_1X^*W_2X^* \dots X^*W_pX^*$, when w varies over the intersection, and is therefore a union of basic open sets.

By induction on n, the same holds for the intersection of n basic open sets. The intersection of 0 basic open set is just the basic open X^*, the claim is clear for $n = 1$, and follows in the other cases from the binary case, treated above. ∎

We can show half of the statement that $\leq^*$ is the specialization quasi-ordering of the subword topology.

Lemma E.3. *Let X be a topological space, with specialization quasi-ordering $\leq$. Any open subset of X^* is upward closed with respect to $\leq^*$. Any closed subset of X^* is downward closed with respect to $\leq^*$.*

Proof. We first show that sets of the form $X^*U_1X^*U_2X^* \dots X^*U_nX^*$, with $U_1, U_2, \dots, U_n$ open in X, are upward closed with respect to $\leq^*$. The Lemma will follow, since every open of X^* is a union of finite intersections of such sets.

Let therefore w be any word from $X^*U_1X^*U_2X^* \dots X^*U_nX^*$. One may write w as $w_0x_1w_1x_2 w_2 \dots w_{n-1}x_nw_n$, with $x_1 \in U_1$, $x_2 \in U_2$, …, $x_n \in U_n$. For any w' with $w \leq^* w'$, one may write w' as $w'_0x'_1w'_1x'_2w'_2 \dots w'_{n-1}x'_nw'_n$, with $w_0 \leq^* w'_0$, $x_1 \leq x'_1$, $w_1 \leq^* w'_1$, $x_2 \leq x'_2$, $w_2 \leq^* w'_2$, …, $w_{n-1} \leq^* w'_{n-1}$, $x_n \leq x'_n$, $w_n \leq^* w'_n$. Since every open is upward closed, $x'_1 \in U_1$, $x'_2 \in U_2$, …, $x'_n \in U_n$. So w' is in $X^*U_1X^*U_2X^* \dots X^*U_nX^*$.

The statement on closed sets follows by complementation. ∎

In fact, when X is just a quasi-ordered set, seen as a topological space through the Alexandroff topology of its ordering, X^* is just the space of finite words quasi-ordered by $\leq^*$, again equipped with its Alexandroff topology.

Lemma E.4 (Coincidence Lemma)**.** *Let X be a set equipped with a quasi-ordering $\leq$. We see X as equipped with the Alexandroff topology of $\leq$. Then the subword topology on X^* is the Alexandroff topology of $\leq^*$.*

Proof. Any upward-closed subset A of X^* is a union of sets of the form $X^*(\uparrow x_1)X^*(\uparrow x_2)X^* \dots X^*(\uparrow x_n)X^*$, namely all those obtained by taking the upward closures of words $x_1x_2 \dots x_n$ in A; indeed $X^*(\uparrow x_1)X^*(\uparrow x_2)X^* \dots X^*(\uparrow x_n)X^*$ is just the upward closure of $x_1x_2 \dots x_n$ in $\leq^*$. Since these are basic opens of the subword topology, the subword topology on X^* is contained in the Alexandroff topology of $\leq^*$. The converse direction is by Lemma E.3. ∎

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 $a_1a_2 \dots a_n$ with $a_1, a_2, \dots, a_n \in A$. Let $A^?$ be $A \cup \{\epsilon\}$.

Lemma E.5. *Let X be a topological space. The complement of $X^*U_1X^*U_2X^* \dots X^*U_nX^*$ ($n \in \mathbb{N}$, $U_1, U_2, \dots, U_n$ open in X) in X^* is $\emptyset$ when $n = 0$, and $F_1^*X^?F_2^*X^? \dots X^?F_{n-1}^*X^?F_n^*$ otherwise, where $F_1 = X \setminus U_1$, …, $F_n = X \setminus U_n$.*

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

Proof. When $n = 0$, this is clear: the complement of $X^*U_1X^*U_2X^* \dots X^*U_nX^*$ is the empty set. So let $n \geq 1$.

We first claim that the complement of $X^*U_1X^*U_2X^* \dots X^*U_nX^*$ is $F_1^*X^?F_2^*X^? \dots X^?F_{n-1}^* X^?F_n^*$. We show this by induction on n. If $n = 1$, then the complement of $X^*U_1X^*$ is the set of words that contain no letter from U_1, i.e., F_1^*. If $n \geq 1$, let w be an arbitrary element of the complement of $X^*U_1X^*U_2X^* \dots X^*U_nX^*$. 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 $F_1^* \subseteq F_1^*X^?F_2^*X^? \dots X^?F_{n-1}^*X^?F_n^*$. Otherwise, w is of the form w_1xw', where $x \in U_1$ and w' is not in $X^*U_2X^* \dots X^*U_nX^*$. By induction hypothesis w' is in $F_2^*X^? \dots X^?F_{n-1}^*X^?F_n^*$, hence again w is in $F_1^*X^?F_2^*X^? \dots X^?F_{n-1}^*X^?F_n^*$. Conversely, let w be any word in $F_1^*X^?F_2^*X^? \dots X^?F_{n-1}^*X^?F_n^*$. 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 $X^*U_1X^*U_2X^* \dots X^*U_nX^*$, since all the words in the latter set must contain at least one letter in U_1; or $w = w_1xw'$ for some $x \notin F_1$, i.e. $x \in U_1$, and $w' \in F_2^*X^? \dots X^?F_{n-1}^*X^?F_n^*$. By induction hypothesis, w' cannot be in $X^*U_2X^* \dots X^*U_nX^*$. By construction, x would be the first occurrence of an element of U_1 in w. If w were in $X^*U_1X^*U_2X^* \dots X^*U_nX^*$, then, some suffix w'' of w' would be in $X^*U_2X^* \dots X^*U_nX^*$. Then $w'' \leq^* w'$, hence by Lemma E.3, w' would be in $X^*U_2X^* \dots X^*U_nX^*$: contradiction.

By Proposition 4.2, X, as a closed subset of itself, is a finite union of irreducible closed subsets. I.e., 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 $F_1^* X^? F_2^* X^? \dots X^? F_{n-1}^* X^? F_n^*$ yields that the complement of $X^* U_1 X^* U_2 X^* \dots X^* U_n X^*$ equals:

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

from which the desired result obtains. ■

This prompts for the following natural generalization of products and SREs to the topological case. Note that, when X is a finite alphabet Σ, with the discrete topology (hence its specialization quasi-ordering is $=$), each closed subset F_i is just a finite subset, and each irreducible closed subset C_i is just a singleton. So the following definition specializes to ordinary products and SREs in this simple case.

Definition E.6 (Top-Product, Top-SRE)**.** Let X be a topological space. Call a *top-product* on X any expression of the form $F_1^* C_1^? F_2^* C_2^? \dots C_{n-1}^? F_n^*$, where $n \in \mathbb{N}$, $F_1, \dots, F_n$ are non-empty closed subsets, and $C_1, C_2, \dots, C_{n-1}$ range over irreducible closed subsets of X. Top-products are interpreted as the obvious subsets of X^*. When $n = 0$, this notation is abbreviated as ϵ, and denotes $\{\epsilon\}$.

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

Lemma E.5 shows that any complement of a basic open $X^* U_1 X^* U_2 X^* \dots X^* U_n X^*$ is (the denotation of) a top-SRE, of a special form. We shall show that any complement of a basic open is (the denotation of) a top-SRE, i.e., top-SREs denote exactly the closed sets. But first, let us check that indeed top-products and top-SREs define closed sets.

Lemma E.7. *Let X be a topological space. 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 $X^* U_{i1} X^* U_{i2} X^* \dots X^* U_{in_i} X^*$, $i \in I$, where $n_i \geq 1$ for all $i \in I$, then we let $\mathcal{U}/U$ be the union of all basic opens $X^*(U_{i1} \cap U) X^* U_{i2} X^* \dots X^* U_{in_i} X^*$.*

Then $\mathcal{U}/U$ is open. The subset $X^* U \mathcal{U}$ is also open for any open subset U of X. For any closed subset F of X, for any closed subset L of X^*, let $U = X \setminus C$, $\mathcal{U} = X^* \setminus L$, then:

- *the complement of $F^? L$ is $X^* X \mathcal{U} \cup \mathcal{U}/U$;*
- *the complement of $F^* L$ is $X^* U \mathcal{U} \cup \mathcal{U}/U$.*

In particular, $F^? L$ and $F^* L$ are closed in X^*.

Proof. We must first check that, if $\mathcal{U} \neq X^*$, then $\mathcal{U}$ is a union of basic opens of the form $X^* U_{i1} X^* U_{i2} X^* \dots X^* U_{in_i} X^*$, $i \in I$, where $n_i \geq 1$ for all $i \in I$. $\mathcal{U}$ is a union of basic opens by Lemma E.2. If n_1 were not at least 1 for all $i \in I$, then the basic open number i would be X^* for some i, so $\mathcal{U}$ would be X^*, contradiction.

$\mathcal{U}/U$ is open, as a union of basic opens. The subset $X^* U \mathcal{U}$ is also open, as the union of all basic opens $X^* U X^* U_{i1} X^* U_{i2} X^* \dots X^* U_{in_i} X^*$, $i \in I$.

To compute complements of $F^? L$ and of $F^* L$, we first make the following remark. Let L_1 and L_2 be two subsets of X^* that are downward closed with respect to $\leq^*$. Note that $L_1 = F^?$ or $L_1 = F^*$, and $L_2 = L$ fit, by Lemma E.3. For any word w not in $L_1 L_2$, we can write w as $w_1 w' w_2$, 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 $\leq^*$.

Note also that both $F^?L$ and F^*L are downward closed with respect to $\leq^*$.

Let F be closed in X, L be closed in X^*, $U = X \setminus C$, $\mathcal{U} = X^* \setminus L$. If $L = X^*$, then the complements of $F^?L = X^*$ and of $F^*L = X^*$ are empty, $\mathcal{U}$ is empty, so $X^*U\mathcal{U}$, $X^*X\mathcal{U}$ and $\mathcal{U}/U$ are empty, too, so the claim is proved. Otherwise, write L as the union of $X^*U_{i1}X^*U_{i2}X^* \dots X^*U_{in_i}X^*$, $i \in I$.

Let us compute the complement of $F^?L$. Assume w is in the complement of $F^?L$, 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 L, hence in $\mathcal{U}$. If $w_1 \neq \epsilon$, then w is in $X\mathcal{U} \subseteq X^*X\mathcal{U}$. If $w_1 = \epsilon$, 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 $X^*U_{i1}X^*U_{i2}X^* \dots X^*U_{in_i}X^*$ for some $i \in I$. If the first letter in $w'w_2$ is in U_{i1}, then $w'w_2$ is in $(U \cap U_{i1})X^*U_{i2}X^* \dots X^*U_{in_i}X^*$, so $w = w_1w'w_2$ is in $X^*(U \cap U_{i1})X^*U_{i2}X^* \dots X^*U_{in_i}X^* \subseteq \mathcal{U}/U$; otherwise, $w'w_2$ is in $UX^*U_{i1}X^*U_{i2}X^* \dots X^*U_{in_i}X^*$, so $w = w_1w'w_2$ is in $X^*UX^*U_{i1}X^*U_{i2}X^* \dots X^*U_{in_i}X^* \subseteq X^*U\mathcal{U}$.

Conversely, assume w is in $X^*X\mathcal{U} \cup \mathcal{U}/U$. If $w \in X^*X\mathcal{U}$, then w contains a subword of the form $a_0a_1a_2 \dots a_{n_i}$, for some $i \in I$, where a_0 is arbitrary, $a_1 \in U_{i1}$, $a_2 \in U_{i2}$, $\dots$, $a_{n_i} \in U_{in_i}$. Note that $a_1a_2 \dots a_{n_i}$ is in $U_{i1}U_{i_2} \dots U_{in_i} \subseteq X^*U_{i1}X^*U_{i2}X^* \dots X^*U_{in_i}X^* \subseteq \mathcal{U}$. If w were in $F^?L$, then since $F^?L$ is downward closed with respect to $\leq^*$, $a_0a_1a_2 \dots a_{n_i}$ would be in $F^?L$, hence $a_1a_2 \dots a_{n_i}$ would be in L: contradiction. So w is in the complement of $F^?L$. If, on the other hand, $w \in \mathcal{U}/U$, then w contains a subword of the form $a_1a_2 \dots a_{n_i}$, for some $i \in I$, where $a_1 \in U \cap U_{i1}$, $a_2 \in U_{i2}$, $\dots$, $a_{n_i} \in U_{in_i}$. In particular, $a_1a_2 \dots a_{n_i}$ is in $U_{i1}U_{i_2} \dots U_{in_i} \subseteq X^*U_{i1}X^*U_{i2}X^* \dots X^*U_{in_i}X^* \subseteq \mathcal{U}$. If w were in F^*L, then since F^*L is downward closed with respect to $\leq^*$, $a_1a_2 \dots a_{n_i}$ would be in F^*L. However, a_1 is in U, so is not in F, and this implies that $a_1a_2 \dots a_{n_i}$ would be in L: contradiction. So, again, w is in w is in the complement of $F^?L$.

The computation of the complement of F^*L follows similar lines. Assume w is in the complement of F^*L, and write w as $w_1w'w_2$ where w' starts with some letter $x \in X$, w_1 is in F^* but w_1x is not, and $w'w_2$ is in $\mathcal{U}$. In particular, x is in U, and $w'w_2$ is in some $X^*U_{i1}X^*U_{i2}X^* \dots X^*U_{in_i}X^*$. Depending on whether the first letter (x) of $w'w_2$ is in U_{i1} or not, $w'w_2$ is in $(U \cap U_{i1})X^*U_{i2}X^* \dots X^*U_{in_i}X^*$ or in $UX^*U_{i1}X^*U_{i2}X^* \dots X^*U_{in_i}X^*$, 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 $a_0a_1a_2 \dots a_{n_i}$ for some $i \in I$, $a_0 \in U$, $a_1 \in U_{i1}$, $a_2 \in U_{i2}$, $\dots$, $a_{n_i} \in U_{in_i}$. If w were in F^*L, then $a_0a_1a_2 \dots a_{n_i}$, too, by down-closure. Since $a_0 \in U$, a_0 is not in F, so $a_0a_1a_2 \dots a_{n_i}$ would be in L. Again by down-closure, $a_1a_2 \dots a_{n_i}$ would be in L: contradiction. So w is in the complement of F^*L. And if $w \in \mathcal{U}/U$, then w contains a subword of the form $a_1a_2 \dots a_{n_i}$, for some $i \in I$, where $a_1 \in U \cap U_{i1}$, $a_2 \in U_{i2}$, $\dots$, $a_{n_i} \in U_{in_i}$. In particular, $a_1a_2 \dots a_{n_i}$ would be in $U_{i1}U_{i_2} \dots U_{in_i} \subseteq X^*U_{i1}X^*U_{i2}X^* \dots X^*U_{in_i}X^* \subseteq \mathcal{U}$. If w were in F^*L, then so would be this subword, and as $a_1 \in U$ is not in F, $a_1a_2 \dots a_{n_i}$ would be in L: contradiction. So w is in the complement of F^*L. ■

Corollary E.8. *Let X be a topological space. Every top-product, every top-SRE is closed in X^*.*

Proof. Let P be any top-product. We show that P is closed by induction on the length n of P. If $n = 0$, i.e., $P = \epsilon$, then we must show that $\{\epsilon\}$ is closed: its complement is indeed the basic open X^*XX^*. If $n = 1$, then $P = F^*$, whose complement is X^*UX^*, where $U = X \setminus F$. Whenever $n \geq 2$, this follows from the induction hypothesis and Lemma E.7. Any top-SRE then denotes a finite union of closed sets, and is therefore closed. ■

We can in fact say more: the top-products are irreducible. For this, we need to recall the following lemma. We give a proof, as we have been unable to find one in standard references.

Lemma E.9. *Let X, Y be two topological spaces, F a closed subset of X, F' a closed subset of Y. If F and F' are irreducible, then $F \times F'$ is an irreducible closed subset of $X \times Y$.*

Proof. It is well known, and also easy to check, that a closed subset F of X is irreducible if and only if $\Diamond F = \{U \text{ open in } X \mid U \cap F \neq \emptyset\}$ is a completely irreducible filter of opens. A *filter* (of opens) is an upward closed family of opens such that any intersection of two elements of the filter is again in the filter. It is *completely prime* if and only if, whenever a union of opens (possibly infinite) lies in the filter, then one of the opens is already in it [5, Section 7.1].

Now consider $\Diamond(F \times F')$. This is clearly upward closed.

If W_1 and W_2 are two elements of $\Diamond(F \times F')$, then both W_1 and W_2 intersect $F \times F'$. Now a basis for the product topology is given by the *open rectangles*, i.e., the product of two opens. So W_1 can be written as a union $\bigcup_{i \in I} U_{1i} \times V_{1i}$ Since W_1 intersects $F \times F'$, for some $i \in I$, $U_{1i} \times V_{1i}$ must intersect $F \times F'$. In particular, U_{1i} intersects F, and V_{1i} intersects F'. Similary, W_2 contains an open rectangle $U_{2j} \times V_{2j}$ where U_{2j} intersects F, and V_{2j} intersects F'. In other words, U_{1i} and U_{2j} are in $\Diamond F$, and V_{1i} and V_{2j} are in $\Diamond F'$. Since F and F' are irreducible, $\Diamond F$ and $\Diamond F'$ are filters, so $U_{1i} \cap U_{2j}$ intersects F, and $V_{1i} \cap V_{2j}$ intersects F'. It follows that $W_1 \cap W_2$, which contains $(U_{1i} \cap U_{2j}) \times (V_{1i} \cap V_{2j})$, intersects $F \times F'$, i.e., is in $\Diamond(F \times F')$. So $\Diamond(F \times F')$ is a filter.

To show that $\Diamond(F \times F')$ is completely prime, consider any union $\bigcup_{i \in I} W_i$ of opens of $F \times F'$ that intersects $F \times F'$. Then some W_i must intersect $F \times F'$, and we are done. ∎

Lemma E.10. *The concatenation function $cat : X^* \times X^* \to X^*$ is continuous. The embedding function $i : X \to X^*$ that maps the letter x to x as a word, is continuous. Every top-product is irreducible closed in X^*. Every top-SRE is closed in X^*.*

Proof. The inverse image of $X^*U_1X^*U_2X^* \dots X^*U_nX^*$ by cat is clearly the union of all rectangles $(X^*U_1X^* \dots X^*U_{j-1}X^*) \times (X^*U_jX^* \dots X^*U_nX^*)$, $1 \leq j \leq n+1$. Since the latter are open in $X^* \times X^*$, we easily check that the inverse image of any open of X^* by cat is open in $X^* \times X^*$. Indeed any open of X^* is a union of finite intersections of such opens. So cat is continuous.

Similarly, the inverse image of $X^*U_1X^*U_2X^* \dots X^*U_nX^*$ by i is $\emptyset$ if $n \geq 2$, U_1 if $n = 1$, and X itself if $n = 0$. In any case, it is open, so i is continuous.

We now claim 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 $w_1 \in F^* \setminus \mathcal{F}_1$ and a word $w_2 \in F^* \setminus \mathcal{F}_2$. Then w_1w_2 would again be in F^*, hence either in $\mathcal{F}_1$ or in $\mathcal{F}_2$. Assume by symmetry that w_1w_2 is in $\mathcal{F}_1$. Since $w_1 \leq^* w_1w_2$, and closed sets such as $\mathcal{F}_1$ are downward closed (w.r.t. the specialization quasi-ordering of X^*, hence also w.r.t. $\leq^*$ by Lemma E.3), we would have $w_1 \in \mathcal{F}_1$: contradiction. So F^* is irreducible.

Second, we claim that $C^?$ is irreducible closed in X^* whenever C is irreducible closed in X. Assume that $C^? \subseteq \mathcal{F}_1 \cup \mathcal{F}_2$, where $\mathcal{F}_1$ and $\mathcal{F}_2$ are closed in X^*. In particular, $i(C) \subseteq \mathcal{F}_1 \cup \mathcal{F}_2$, that is, $C \subseteq i^{-1}(\mathcal{F}_1 \cup \mathcal{F}_2) = i^{-1}(\mathcal{F}_1) \cup i^{-1}(\mathcal{F}_2)$. Since i is continuous, $i^{-1}(\mathcal{F}_1)$ and $i^{-1}(\mathcal{F}_2)$ are closed. Since C is irreducible, $C \subseteq i^{-1}(\mathcal{F}_1)$ or $C \subseteq i^{-1}(\mathcal{F}_2)$. Assume $C \subseteq i^{-1}(\mathcal{F}_1)$, by symmetry. Then $i(C) \subseteq \mathcal{F}_1$. Since C is non-empty, $\mathcal{F}_1$ is non-empty; $\mathcal{F}_1$ is downward closed with respect to $\leq^*$, by Lemma E.3, so ϵ is in $\mathcal{F}_1$. It follows that $C^? = i(C) \cup \{\epsilon\}$ is contained in $\mathcal{F}_1$. Hence $C^?$ is irreducible.

We now observe 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. Assume that $\mathcal{C}_1\mathcal{C}_2 \subseteq \mathcal{F}_1 \cup \mathcal{F}_2$, where $\mathcal{F}_1$ and $\mathcal{F}_2$ are closed in X^*. That is, the image of $\mathcal{C}_1 \times \mathcal{C}_2$ by cat is contained in $\mathcal{F}_1 \cup \mathcal{F}_2$, i.e., $\mathcal{C}_1 \times \mathcal{C}_2 \subseteq cat^{-1}(\mathcal{F}_1) \cup cat^{-1}(\mathcal{F}_2)$. Then the claim follows from the fact that $cat^{-1}(\mathcal{F}_1)$ and $cat^{-1}(\mathcal{F}_2)$ are closed, since cat is continuous, and from the fact that $\mathcal{C}_1 \times \mathcal{C}_2$ is irreducible closed by Lemma E.9. Indeed, we obtain that $\mathcal{C}_1 \times \mathcal{C}_2$ is contained in $cat^{-1}(\mathcal{F}_1)$ or in $cat^{-1}(\mathcal{F}_2)$, i.e., that $\mathcal{C}_1\mathcal{C}_2 \subseteq \mathcal{F}_1$ or $\mathcal{C}_1\mathcal{C}_2 \subseteq \mathcal{F}_2$.

By induction on syntax, it follows that every top-product is irreducible closed. The base case is ϵ, which, since it denotes a one-element set, is clearly irreducible. Then, any top-SRE is a finite union of top-products, and hence closed. ∎

Recall that the topological closure of a point $x \in X$ is also its downward closure $\downarrow x$, for the specialization quasi-ordering of X.

Lemma E.11. *Let X be a topological space. The closure of the word $x_1x_2 \ldots x_n$ in X^* is the top-product $(\downarrow x_1)^?(\downarrow x_2)^? \ldots (\downarrow x_n)^?$.*

Proof. By Lemma E.10, this top-product is closed. The closure of $x_1x_2 \ldots x_n$ must contain this top-product, because any word $w \leq^* x_1x_2 \ldots x_n$ must be in this closure, by Lemma E.3. Whence the equality. ∎

Proposition E.12. *Let X be a topological space. The specialization quasi-ordering of X^* is the embedding quasi-ordering $\leq^*$, where $\leq$ is the specialization quasi-ordering of X.*

Proof. Let $\preceq$ denote the specialization quasi-ordering of X^* for the time being. If $w \leq^* w'$ then $w \preceq w'$: indeed, any open $\mathcal{U}$ containing w is upward closed with respect to $\leq^*$, so contains w' as well, by Lemma E.3. Conversely, if $w \preceq w'$, then w is in the topological closure of w'. This is an alternative definition of the specialization quasi-ordering, which is easily seen to be equivalent. However, Lemma E.11 states precisely that w must then be such that $w \leq^* w'$. ∎

We can compare top-products for inclusion, algorithmically. This is analogous to the case of products [1]. For short, write C, C' for irreducible closed subsets of X; F, F' for non-empty closed subsets of X; P, P' for top-products.

Lemma E.13. *Let X be a topological space. Inclusion between top-products can be checked in quadratic time, modulo an oracle testing inclusion of closed subsets of X. We have: $\epsilon \subseteq P$ for any top-product P, $P \not\subseteq \epsilon$ unless P is syntactically the top-product ϵ, and:*

- *$C^?P \subseteq C'^?P'$ 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 $F^*P \subseteq P'$.*
- *$F^*P \subseteq F'^*P'$ 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.

- Assume $C^?P \subseteq C'^?P'$. If $C \subseteq C'$, then let x be an arbitrary element of C. This is possible, as C is irreducible, hence non-empty. For every $w \in P$, xw is in $C^?P$, hence in $C'^?P'$. So xw or w is in P', and since P' is downward closed under $\leq^*$ by Lemma E.3, in any case $w \in P'$. So $P \subseteq P'$. If on the other hand C is not contained in C', then there is an element x of C which is not in C'. Since $C^?P \subseteq C'^?P'$, every word of the form xw with $x \in C$, $w \in P$, is in $C'^?P'$. However since x is not in C', xw must be in P'. So $CP \subseteq P'$. Since P' is downward closed, $C^?P \subseteq P'$. The converse direction is easy.
- Assume $C^?P \subseteq F'^*P'$. If $C \subseteq F'$, then $P \subseteq F'^*P'$, since $P \subseteq C^?P$. If $C \not\subseteq F'$, then let x be in C but not in F'. For every $xw \in CP$, xw is in F'^*P', hence in P' since $x \notin F'$. So $CP \subseteq P'$. Since P' is downward closed, $C^?P \subseteq P'$. The converse direction is easy.
- Assume $F^*P \subseteq C'^*P'$. 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.

- 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 by dynamic programming. ∎

Now, testing inclusion between closed subsets of X is as easy as testing inclusion between elements of $\mathcal{S}(X)$. This is a general fact about Noetherian spaces

We may also compute intersections of top-products.

Lemma E.14. *Let X be a Noetherian space. One may compute the intersection of two top-products, modulo an oracle that computes intersections of closed subsets of X, i.e., such that given two closed subsets F, F' of X, computes a finite set $\mathcal{E}(F, F')$ of irreducible closed subsets of X whose union is $F \cap F'$.*

We have: $\epsilon \cap P = \epsilon$ for every product P, and:

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

Proof. Note that the map $\mathcal{E}(F, F')$ is well-defined, by Proposition 4.2. We require to be able to compute it.

One may also note that the purpose of the Lemma is to show how to define an oracle computing this for irreducible closed subsets of X^*, knowning one for closed subsets of X. This much depends on the fact that irreducible closed subsets of X^* are exactly the denotations of top-products, which we shall prove later. Then, provided closed sets of X^* are represented as finite unions of irreducible closed sets, i.e., as top-SREs, and distributing intersections over unions, we obtain a similar oracle for X^*, knowing one for X.

- 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 $C^? P \cap C'^? P'$, so equality obtains. We conclude since $(C \cap C')^? (P \cap P') = \bigcup_{C'' \in \mathcal{E}(C,C')} C''^? (P \cap P')$.
- 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 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. We conclude since $(C \cap F')^? (P \cap F'^* P') = \bigcup_{C'' \in \mathcal{E}(C,F')} C''^? (P \cap F'^* P')$.
- 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. We conclude because $F \cap F' = \bigcup_{C'' \in \mathcal{E}(F,F')} C''$.

Rewriting left hand sides to right hand sides clearly defines a terminating procedure to compute intersections of top-products. ∎

Lemma E.15. *Let X be Noetherian. In X^*, the intersection of any two top-products is a finite union of top-products.*

Proof. The algorithm of Lemma E.14 rewrites any such intersection as a finite union of top-products, recursively. ∎

As in the case of SREs, any top-product P can be written as $e_1 e_2 \dots e_n$, where each e_i is an *atomic expression* of the form $C^?$ or F^*, and additionally $e_i e_{i+1}$ is contained neither in e_i nor in e_{i+1} for all i, $1 \leq i < n$. Indeed, if $e_i e_{i+1}$ is contained in e_i, then its denotation in fact equals that of e_i, and similarly for e_{i+1}. Call such a sequence a *reduced* top-product. Clearly, every top-product denotes the same set of some reduced top-product.

Lemma E.16. *Let X be a topological space. For every top-product $P = e_1 e_2 \dots e_n$, let $\mu(P)$ be the multiset consisting of $e_1, \dots, e_n$. Define $\sqsubseteq$ on atomic expressions 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 $F^* \not\sqsubseteq {C'}^?$. Let $\sqsubseteq_{mul}$ be the multiset extension of $\sqsubseteq$.*

For every top-products P, P', if P is reduced and $P \subseteq P'$ then $\mu(P) \sqsubseteq_{mul} \mu(P')$.

Proof. We show this by induction on $|P| + |P'|$, where $|P|$ is the number of atomic expressions in P. If $|P'| = 0$, i.e., $P' = \epsilon$, then $P = \epsilon$, and the claim is clear. In general, if $|P| = 0$, i.e., $P = \epsilon$, then $\mu(P)$ is the empty multiset, so $\mu(P) \sqsubseteq_{mul} \mu(P')$. Otherwise, there are four cases, using Lemma E.13. Observe that Lemma E.13 can be stated equivalently as follows: $P = e_1 P_1 \subseteq e'_1 P'_1 = P'$ if and only if: (1) $e_1 \not\sqsubseteq e'_1$ and $P \subseteq P'_1$, or (2) $e_1 = C^?$, $e'_1 = {C'}^?$, $C \subseteq C'$ and $P_1 \subseteq P'_1$, or (3) $e'_1 = {F'}^*$, $e_1 \sqsubseteq {F'}^*$ and $P_1 \subseteq P'$. Write $\sqsubset$ the strict part of $\sqsubseteq$, i.e., $e \sqsubset e'$ iff $e \sqsubseteq e'$ and $e' \not\sqsubseteq e$.

- In case (1), we have $\mu(P) \sqsubseteq \mu(P'_1)$ by induction hypothesis. Since $\mu(P'_1) \sqsubset \mu(P')$, $\mu(P) \sqsubset \mu(P')$.
- In case (2), $e_1 \sqsubseteq e'_1$ and $\mu(P_1) \sqsubseteq_{mul} \mu(P'_1)$, so $\mu(P) = \mu(e_1 P_1) \sqsubseteq_{mul} \mu(e'_1 P'_1) = \mu(P')$.
- Case (3) is the trickiest. We have $e_1 \sqsubseteq {F'}^*$, and $P_1 \subseteq P'$. Write P as $e_1 e_2 \dots e_k P_0$, where k is the largest integer such that $e_1, e_2, \dots, e_k \sqsubseteq {F'}^*$. For each i, since $e_i \sqsubseteq {F'}^*$, in particular $e_i \subseteq {F'}^*$.

  We first deal with the case where some e_i has the same denotation as ${F'}^*$. If we had $k \geq 2$, then either $i \geq 2$ and then $e_{i-1} \subseteq e_i$, so $e_{i-1} e_i = e_i$, or $i < k$ and then $e_{i+1} \subseteq e_i$, so $e_i e_{i+1} = e_i$; in any case, this would contradict the fact that P is reduced. So $k = 1$, $P_0 = P_1$, and P is of the form ${F'}^* P_1$. If $|P_0| = 0$, i.e, $P_0 = \epsilon$, then $\mu(P) \sqsubseteq_{mul} \mu({F'}^* P'_1) = \mu(P')$. Otherwise, $P_0 = P_1$ is of the form eP_+, where $e \not\sqsubseteq {F'}^*$ by maximality of k. Since $P_1 \subseteq P'$, i.e., $eP_+ \subseteq {F'}^* P'_1$, but $e \not\sqsubseteq {F'}^*$, we must be in case (1), so $P_1 = eP_+ \subseteq P'_1$. Since $P = {F'}^* P_1$, $P' = {F'}^* P'_1$, and $P_1 \subseteq P'_1$, it follows, using the induction hypothesis, that $\mu(P) \sqsubseteq_{mul} \mu(P')$.

  Otherwise, $P = e_1 e_2 \dots e_k P_0$, where no e_i has the same denotation as ${F'}^*$. It is easy to check that, then, $e_i \sqsubset {F'}^*$ for all i, $1 \leq i \leq k$. It is then enough to show that: $(*)$ $\mu(P_0) \sqsubseteq_{mul} \mu(P'_1)$. From $(*)$ it will follow that $\mu(P)$ is obtained from $\mu(P')$ by replacing one copy of ${F'}^*$ by finitely many (k) copies of atomic expressions $e_1, e_2, \dots, e_k$ that are strictly smaller than ${F'}^*$ in $\sqsubset$; so $\mu(P)$ will be (strictly) smaller than $\mu(P')$ in $\sqsubseteq_{mul}$.

  To show $(*)$, we observe that, by induction on $k - i$, $e_i e_{i+1} \dots e_k P_0$ must be contained in ${F'}^* P'_1$, for all i, $1 \leq i \leq k+1$: just use case (3) repetitively. So $P_0 \subseteq {F'}^* P'_1$. If $|P_0| = 0$, then $\mu(P_0) \sqsubseteq_{mul} \mu(P'_1)$, as claimed. Otherwise, write P_0 as eP_+. By the maximality of k, $e \not\sqsubseteq {F'}^*$, so case (1) applies, and therefore $P_0 \subseteq P'_1$, whence $\mu(P_0) \sqsubseteq_{mul} \mu(P'_1)$ by induction hypothesis. ∎

Proposition E.17. *Let X be Noetherian. The inclusion ordering on the set of (denotations of) top-products is well-founded.*

Proof. We observe that, since X is Noetherian, $\subseteq$ is well-founded on the set of closed sets. Indeed, by Proposition 3.2 of [14], X is Noetherian if and only if no ascending chain of open sets is infinite: so there is no infinite descending chain of closed sets. It follows that $\sqsubseteq$, and therefore also $\sqsubseteq_{mul}$, is well-founded. The claim then follows from Lemma E.16. ∎

Corollary E.18. *Let X be Noetherian. The inclusion ordering on the set of (denotations of) SREs is well-founded.*

Proof. Let $\mathcal{F} = P_1 \cup \ldots \cup P_m$ and $\mathcal{F}' = P'_1 \cup \ldots \cup P'_n$ be two SREs. Without loss of generality, assume the P_is are pairwise incomparable, and similarly that the P'_js are pairwise incomparable. Since every top-product P_i, $1 \leq i \leq m$, is irreducible closed by Lemma E.10, $\mathcal{F} \subseteq \mathcal{F}'$ if and only if for every i, $1 \leq i \leq m$, there is a j, $1 \leq j \leq n$ with $P_i \subseteq P'_j$. Since the P_is are pairwise incomparable, it follows that the multiset consisting of $P_1, \ldots, P_m$ is smaller that the one consisting of $P'_1, \ldots, P'_n$ in the multiset extension of the inclusion ordering $\subseteq$ on (denotations of) top-products. Since the latter is well-founded by Proposition E.17, so is inclusion between SREs. ∎

Proposition E.19. *Let X be a Noetherian space. The irreducible closed subsets of X^* are the (denotations of) top-products. The closed subsets of X^* are the (denotations of) top-SREs.*

Proof. Lemma E.10 states that Every top-product is irreducible closed, and every top-SRE is closed. Conversely, let $\mathcal{F}$ be any closed subset of X^*. $\mathcal{F}$ is the complement of a union of basic subsets, of the form $X^*U_1X^*U_2X^* \ldots X^*U_nX^*$, by Lemma E.2. So $\mathcal{F}$ is an intersection of finite unions of top-products, by Lemma E.5. It is well-known that any (possibly infinite) intersection $\bigcap_{i \in I} \mathcal{F}_i$ can also be written as the filtered intersection $\bigcap_{J \text{ finite} \subseteq I} \mathcal{F}_J$, where $\mathcal{F}_J$ is the finite intersection $\bigcap_{i \in J} \mathcal{F}_i$. Filtered means that whatever J and J', there is a J'' such that $\mathcal{F}_{J''}$ is contained in both $\mathcal{F}_J$ and $\mathcal{F}_{J'}$: namely $J'' = J \cup J'$.

Now each $\mathcal{F}_J$ is a finite intersection of finite unions of top-products. By distributing unions over intersections, one may therefore write $\mathcal{F}_J$ as a top-SRE. Using Corollary E.18, it follows that $\mathcal{F}$ equals some $\mathcal{F}_J$. Indeed, otherwise, we could build an infinite descending chain of SREs $\mathcal{F}'_0 \supset \mathcal{F}'_1 \supset \ldots$, all containing $\mathcal{F}$, as follows: pick $\mathcal{F}'_0 = \mathcal{F}_J$ for some arbitrary J; if the chain has been built up to index k, since $\mathcal{F}'_k$ (of the form $\mathcal{F}_J$ for some J) does not coincide with $\mathcal{F}$, there must be a finite subset J' of I such that $\mathcal{F}_J \cap \mathcal{F}_{J'}$ is strictly contained in $\mathcal{F}_J$: then choose $\mathcal{F}'_{k+1} = \mathcal{F}_{J \cup J'}$.

So $\mathcal{F}$ is (the denotation of) a top-SRE, namely $\mathcal{F}_J$.

Let $\mathcal{F}$ be written as the sum of the top-products $P_1, \ldots, P_k$. If additionally $\mathcal{F}$ is irreducible, then $k = 0$ otherwise $\mathcal{F}$ would be empty, and $k \geq 2$ is impossible since $\mathcal{F}$ is irreducible. So $k = 1$, hence $\mathcal{F}$ is (the denotation of) a top-product. ∎

Theorem E.20 (Topological Higman Lemma)**.** *Let X be a topological space. Then X is Noetherian if and only if X^* is.*

Proof. Theorem 6.11 of [14] states that a sober space Y is Noetherian if and only if its topology is the upper topology of a well-founded partial ordering $\preceq$ that obeys:

- property T: there is a finite subset E such that $Y = \downarrow E$ ($\downarrow$ denotes downward closure with respect to $\preceq$ here);
- and property W: for all $y_1, y_2 \in Y$, there is a finite subset E such that $\downarrow y_1 \cap \downarrow y_2 = \downarrow E$.

Any sobrification is equipped with the upper topology of the inclusion ordering $\subseteq$. Assume X Noetherian. Since (denotations of) top-products and irreducible closed subsets of X coincide by

Proposition E.19, Proposition E.17 states exactly that $\subseteq$ is well-founded on $\mathcal{S}(X^*)$; Lemma E.15 states property W for $\mathcal{S}(X^*)$; while property T for $\mathcal{S}(X^*)$ is obvious, since $\mathcal{S}(X^*)$ is the downward closure of the top-product X^*. So $\mathcal{S}(X^*)$ is Noetherian. By [14, Proposition 6.2], a space is Noetherian if and only if its sobrification is. So X^* is Noetherian.

Conversely, recall that a space is Noetherian if and only if it has no infinite ascending chain of opens [14, Proposition 3.2]. If X^* is Noetherian, then any infinite ascending chain $U_0 \subset U_1 \subset \ldots \subset U_k \subset \ldots$ of opens of X induces an infinite ascending chain $X^*U_0X^* \subset X^*U_1X^* \subset \ldots \subset X^*U_kX^* \subset \ldots$ of opens in X^*: contradiction. So X is Noetherian. ■

Theorem E.20 generalizes Higman's Lemma, in the following sense. When X is a set equipped with a quasi-ordering $\leq$, we may see X as a topological space, equipped with the Alexandroff topology of $\leq$. If X is well, then by Proposition 3.1 of [14], X is Noetherian (and conversely). By Lemma E.4, the topology of X^* is the Alexandroff topology of $\leq^*$. Theorem E.20 then states that X^* is Noetherian, hence $\leq^*$ is well, by Proposition 3.1 of [14] again. Such an argument would probably be the most complicated proof of Higman's Lemma in existence. We only aim to clarify that Theorem E.20 indeed generalizes Higman's Lemma to the topological case.

Corollary E.21. *Let X be Noetherian. Any open subset of X^* is a finite union of basic opens, of the form $X^*U_1X^*U_2X^* \ldots X^*U_nX^*$.*

Proof. Any open U is a union of basic opens U_i, $i \in I$, by Lemma E.2. Note that $(U_i)_{i\in I}$ is a cover of U. Since X^* is Noetherian by Theorem E.20, U is compact. So we may extract a finite subcover of $(U_i)_{i\in I}$. ■

We have announced that $\mathcal{S}(X^*)$ would consist of natural analogues of the notion of products used in SREs, built on an alphabet of points in $\mathcal{S}(X)$. These analogues are the top-products, as one can expect. The following theorem is a syntactic rewriting of most of the results obtained above.

Theorem E.22. *If X is Noetherian, then up to homeomorphism, the elements of $\mathcal{S}(X)$ are (denotations of)* products*, which are defined as finite sequences $e_1e_2\ldots e_k$ of atomic expressions, modulo $\equiv$, where:*

- *an atomic expression is either of the form $C^?$ with $C \in \mathcal{S}(X)$, or A^* with A a non-empty finite subset of $\mathcal{S}(X)$;*
- *the denotation of products is $[\![e_1e_2\ldots e_k]\!] = [\![e_1]\!]\,[\![e_2]\!]\ldots[\![e_k]\!]$, where $[\![C^?]\!] = [\![C]\!]^?$ and $[\![A*]\!] = (\bigcup_{C\in A}[\![C]\!])^*$;*
- *$P \equiv P'$ if and only if $[\![P]\!] = [\![P']\!]$.*

This is equipped with the upper topology of the ordering $\sqsubseteq$ (and $\equiv$ is $\sqsubseteq \cap \sqsupseteq$), where:

- *$C^?P \sqsubseteq C'^?P'$ if and only if $C \subseteq C'$ and $P \sqsubseteq P'$, or $C^?P \sqsubseteq P'$.*
- *$C^?P \subseteq A'^*P'$ if and only if $C \subseteq C'$ for some $C' \in A'$ and $P \sqsubseteq A'^*P'$, or $C \subseteq C'$ for no $C' \in A'$ and $C^?P \sqsubseteq P'$.*
- *$A^*P \sqsubseteq C'^*P'$ if and only if $A^*P \sqsubseteq P'$.*
- *$A^*P \subseteq A'^*P'$ if and only if every $C \in A$ is contained in some $C' \in A'$ and $P \sqsubseteq A'^*P'$, or some $C \in A$ is contained in no $C' \in A'$ and $A^*P \sqsubseteq P'$.*

The latter definition can then be simplified, to: $eP \sqsubseteq e'P'$ if and only if (1) $e \not\sqsubseteq e'$ and $eP \sqsubseteq P'$, or (2) $e = C^?$, $e' = C'^?$, $C \subseteq C'$ and $P \sqsubseteq P'$, or (3) $e' = A'^*$, $e \sqsubseteq A'^*$ and $P \sqsubseteq e'P'$. This requires defining $\sqsubseteq$ on atomic expressions, by: $C^? \sqsubseteq C'^?$ if and only if $C \subseteq C'$; $C^? \sqsubseteq A'^*$ if and only if $C \subseteq C'$ for some $C' \in A'$; $A^* \not\sqsubseteq C'^?$; $A^* \sqsubseteq A'^*$ if and only if for every $C \in A$, there is a $C' \in A'$ with $C \subseteq C'$.

Appendix F. Finite Multisets and the $\leq^{\circledast}$ Quasi-Ordering

Given any topological space, let $X^{\circledast}$ be the set of all finite multisets on X. We shall write $\{|x_1, \ldots, x_n|\}$ the multiset containing exacly the elements $x_1, \ldots, x_n$, $\emptyset$ the empty multiset, and $m \uplus m'$ the multiset union of m and m'. For any $A \subseteq X$, let $A^{\circledast}$ be the of those multisets consisting of elements of A only. Let $A^{?}$ be the set consisting of $\emptyset$ and all multisets $\{|x|\}$, $x \in A$. Given two subsets $\mathcal{A}$ and $\mathcal{B}$ of $X^{\circledast}$, $\mathcal{A} \odot \mathcal{B}$ denotes $\{m \uplus m' \mid m \in \mathcal{A}, m' \in \mathcal{B}\}$.

We quasi-order $X^{\circledast}$, not with the multiset extension of the specialization quasi-ordering $\leq$ of X, rather with the *submultiset* quasi-ordering $\leq^{\circledast}$ 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 $r : \{1, 2, \ldots, m\} \to \{1, 2, \ldots, n\}$ such that $x_i \leq y_{r(i)}$ for all 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.* [3, Section 2]. The corresponding topology is:

Definition F.1 (Sub-Multiset Topology)**.** The *sub-multiset topology* on $X^{\circledast}$ is the least one containing the subsets $X^{\circledast} \odot U_1 \odot U_2 \odot \ldots \odot U_n$, $n \in \mathbb{N}$, where $U_1, U_2, \ldots, U_n$ are open subsets of X.

We shall topologize $X^{\circledast}$ with the sub-multiset topology. An important tool to study $X^{\circledast}$ is the *Parikh mapping*, extended here to the topological case, i.e., the case of an infinite alphabet X with a topology.

Definition F.2 (Parikh)**.** The *Parikh mapping* $\Psi : X^* \to X^{\circledast}$ maps every finite word $x_1 x_2 \ldots x_n$ on X to $\{|x_1, x_2, \ldots, x_n|\}$.

We shall see that Ψ is not only continuous, it is *quotient*. A quotient map $f : A \to B$ is by definition a surjective map such that, for every $V \subseteq B$, V is open in B if and only if $f^{-1}(B)$ is open in A. A continuous map satisfies that V open in B implies $f^{-1}(V)$ open in A, but $f^{-1}(V)$ open does not necessarily entail that V is open. Additionally, a quotient map must be surjective. Whenever $\equiv$ is an equivalence relation on a space A, the map sending each $a \in A$ to its equivalence class is a quotient map; conversely, if $f : A \to B$ is quotient, then B is homeomorphic to the quotient of A by the relation $a \equiv a'$ defined as $f(a) = f(a')$, and, up to this homeomorphism, f maps $a \in A$ to its equivalence class. The fact that Ψ is quotient therefore means that $X^{\circledast}$ appears as the quotient of X^* with respect to all reorderings of letters in words.

To show this, we make two comments. First, for any subset B of $X^{\circledast}$, $\Psi(\Psi^{-1}(B)) = B$. This is because Ψ is surjective, which is clear. Second, define $\equiv$ on X^* by $w \equiv w'$ if and only if $\Psi(w) = \Psi(w')$, i.e., w and w' contain the same letters, with the same multiplicities. A subset A of X^* is $\equiv$-saturated if and only if it is a union of equivalence classes. Equivalently, A is $\equiv$-saturated if and only if $\Psi^{-1}(\Psi(A)) = A$. One notes indeed that the $\equiv$-saturation of any subset A of X^*, i.e., the smallest $\equiv$-saturated subset of X^* containing A, is $\Psi^{-1}(\Psi(A))$.

Proposition F.3. *The Parikh mapping Ψ is quotient.*

Proof. We have already noted that Ψ was quotient. I.e., any multiset $\{|x_1, x_2, \ldots, x_n|\}$ appears as $\Psi(x_1 x_2 \ldots x_n)$.

The inverse image of the basic open $X^{\circledast} \odot U_1 \odot U_2 \odot \ldots \odot U_n$ by Ψ is the union over all permutations π of $\{1, 2, \ldots, n\}$ of the basic opens $X^* U_{\pi(1)} X^* U_{\pi(2)} X^* \ldots X^* U_{\pi(n)} X^*$, and is therefore open. This just means that the finite words whose multiset of letters contain one letter from U_1, one from $U_2, \ldots$, one from U_n, are just the finite words containing a subword contain one letter from each in some order. It follows that the inverse image of any open of $X^{\circledast}$ is open in X^*, so Ψ is continuous.

Finally, let V be any subset of $X^{\circledast}$ such that $\Psi^{-1}(V)$ is open in X^*. Then $\Psi^{-1}(V)$ is a union of basic opens of the form $X^*U_{i1}X^* \ldots X^*U_{ik_i}X^*$, $i \in I$. Observe that $V = \Psi(\Psi^{-1}(V))$ is then the union of all subsets of the form $X^{\circledast} \odot U_{i1} \odot \ldots \odot U_{ik_i}$, $i \in I$, and is therefore open. So Ψ is quotient. ∎

Theorem F.4. *Let X be a topological space. Then X is Noetherian if and only if $X^{\circledast}$ is.*

Proof. If X is Noetherian, then X^* is, by Theorem E.20. Since Ψ is surjective and continuous by Proposition F.3, $X^{\circledast}$ is the continuous image of X^*. But the continuous image of any Noetherian space is again Noetherian [14, Lemma 4.4].

Conversely, recall that a space is Noetherian if and only if it has no infinite ascending chain of opens [14, Proposition 3.2]. If $X^{\circledast}$ is Noetherian, then any infinite ascending chain $U_0 \subset U_1 \subset \ldots \subset U_k \subset \ldots$ of opens of X induces an infinite ascending chain $X^{\circledast} \odot U_0 \subset X^{\circledast} \odot U_1 \subset \ldots \subset X^{\circledast} \odot U_k \subset \ldots$ of opens in $X^{\circledast}$: contradiction. So X is Noetherian. ∎

Definition F.5 (M-Product, M-SRE)**.** Let X be a topological space. Call an *m-product* on X any expression of the form $F^{\circledast} \odot C_1^{\textcircled{?}} \odot C_2^{\textcircled{?}} \odot \ldots \odot C_n^{\textcircled{?}}$, where $n \in \mathbb{N}$, F is a closed subset of X, and $C_1, C_2, \ldots, C_{n-1}$ range over irreducible closed subsets of X. This is interpreted as the obvious subset of X^*. When F is empty, we shall also write this as simply $C_1^{\textcircled{?}} \odot C_2^{\textcircled{?}} \odot \ldots \odot C_n^{\textcircled{?}}$. When $n = 0$, we just write $F^{\circledast}$, and when $n = 0$ and $F = \emptyset$, we write this ϵ. (Note that the denotation of ϵ is then $\{\emptyset\}$.)

An *m-SRE* is any finite sum of m-products, where sum is interpreted as union.

Proposition F.6. *Let X be a topological space. Then the denotations of m-SREs are closed in $X^{\circledast}$, and those of m-products are irreducible closed.*

If X is Noetherian, then the irreducible closed subsets of $X^{\circledast}$ are the denotations of m-products, and the closed subsets of $X^{\circledast}$ are the denotations of m-SREs.

Proof. consider any m-product $P = F^{\circledast} \odot C_1^{\textcircled{?}} \odot C_2^{\textcircled{?}} \odot \ldots \odot C_n^{\textcircled{?}}$. We observe that $\Psi^{-1}(P)$ is the union over all permutations π of $\{1, 2, \ldots, n\}$ of the top-products $F^*C^?_{\pi(1)}F^*C^?_{\pi(2)}F^* \ldots F^*C^?_{\pi(n)}F^*$. This just 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 (Proposition F.3), a subset $\mathcal{F}$ of $X^{\circledast}$ is closed if and only if $\Psi^{-1}(\mathcal{F})$ is closed in X^*. Therefore (the denotation of) P is closed in $X^{\circledast}$.

It also follows that any m-SRE denotes some closed subset of $X^{\circledast}$.

It remains to show that the denotation of m-products are indeed irreducible. Note that $F^{\circledast} \odot C_1^{\textcircled{?}} \odot C_2^{\textcircled{?}} \odot \ldots \odot C_n^{\textcircled{?}}$ equals $\Psi(F^*C_1^?C_2^? \ldots C_n^?)$, hence for any two closed subsets $\mathcal{F}_1$ and $\mathcal{F}_2$ of $X^{\circledast}$, $F^{\circledast} \odot C_1^{\textcircled{?}} \odot C_2^{\textcircled{?}} \odot \ldots \odot C_n^{\textcircled{?}} \subseteq \mathcal{F}_1 \cup \mathcal{F}_2$ if and only if $F^*C_1^?C_2^? \ldots C_n^? \subseteq \Psi^{-1}(\mathcal{F}_1 \cup \mathcal{F}_2) = \Psi^{-1}(\mathcal{F}_1) \cup \Psi^{-1}(\mathcal{F}_2)$. Since $F^*C_1^?C_2^? \ldots C_n^?$ is irreducible (Lemma E.10), and $\Psi^{-1}(\mathcal{F}_1)$ and $\Psi^{-1}(\mathcal{F}_2)$ are closed (Ψ being continuous), $F^*C_1^?C_2^? \ldots C_n^?$ must be contained in $\Psi^{-1}(\mathcal{F}_1)$ or in $\Psi^{-1}(\mathcal{F}_2)$. So $F^{\circledast} \odot C_1^{\textcircled{?}} \odot C_2^{\textcircled{?}} \odot \ldots \odot C_n^{\textcircled{?}}$ is contained in $\mathcal{F}_1$ or in $\mathcal{F}_2$.

Conversely, assume X Noetherian. Let $\mathcal{F}$ be any closed subset of $X^{\circledast}$. Since Ψ is continuous (Proposition F.3), $\Psi^{-1}(\mathcal{F})$ is closed in X^*, hence a finite union of (denotations of) top-products, by Proposition E.19. Since Ψ is surjective, $\mathcal{F} = \Psi(\Psi^{-1}(\mathcal{F}))$ is therefore a finite union of subset $\Psi(P_i)$, $i \in I$, where P_i are (denotations of) top-products. However, for any top-product $e_1e_2 \ldots e_n$, $\Psi(e_1e_2 \ldots e_n) = \{m_1 \uplus m_2 \uplus \ldots \uplus m_n \mid m_1 \in \Psi(e_1), m_2 \in \Psi(e_2), \ldots, m_n \in \Psi(e_n)\}$, where $\Psi(e_j)$ is computed by: $\Psi(F^*) = F^{\circledast}$, $\Psi(C^?) = C^{\textcircled{?}}$; so $\Psi(e_1e_2 \ldots e_n) = \Psi(e_1) \odot \Psi(e_2) \odot \ldots \odot \Psi(e_n)$ can be written (using the fact that $\odot$ is associative and commutative) as $F_1^{\circledast} \odot F_2^{\circledast} \odot \ldots \odot F_k^{\circledast} \odot C_1^{\textcircled{?}} \odot \ldots \odot C_\ell^{\textcircled{?}}$, where the F_is are non-empty closed and the C_js are irreducible closed. Noting the $\emptyset^{\circledast} \odot A = A$, and that $F_i^{\circledast} \odot F_{i'}^{\circledast} = (F_i \cup F_{i'})^{\circledast}$, we conclude that the image of any top-product by Ψ is (the denotation of) an m-product. Hence each $\Psi(P_i)$ is the denotation of an m-product. Therefore $\mathcal{F}$ is a finite union of m-products, hence (the denotation of) an m-SRE.

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

We won't need the following lemma. We mention it because it answers a natural question.

Lemma F.7. *The mappings $j : X \to X^{\circledast}$ sending x to $\{|x|\}$ and $union : X^{\circledast} \times X^{\circledast} \to X^{\circledast}$ sending m, m' to $m \uplus m'$, are continuous.*

Proof. First, $j = \Psi \circ i$, where i is given in Lemma E.10. As a composition of continuous functions, it is continuous. Second, $union(\Psi(w), \Psi(w')) = \Psi(cat(w, w'))$. Since cat is continuous by Lemma E.10, $union$ is, too, by general arguments on quotient maps. Ψ is indeed quotient by Proposition F.3. ∎

Proposition F.8. *Let X be a topological space. The closure of the multiset $\{|x_1, x_2, \ldots, x_n|\}$ in $X^{\circledast}$ is the denotation of the m-product $(\downarrow x_1)^{?} \odot (\downarrow x_2)^{?} \odot \ldots \odot (\downarrow x_n)^{?}$. The specialization quasi-ordering of $X^{\circledast}$ is the sub-multiset extension $\leq^{\circledast}$ of the specialization quasi-ordering $\leq$ of X.*

Proof. We first note that any open subset of $X^{\circledast}$ is upward closed with respect to $\leq^{\circledast}$. This is an easy consequence of the easy fact that $X^{\circledast} \odot U_1 \odot U_2 \odot \ldots \odot U_n$ is upward closed with respect to $\leq_{mul}$ for any opens U_1, U_2, …, U_n, which are upward closed with respect to $\leq$ by definition. It also follows that any closed subset of $X^{\circledast}$ is downward closed with respect to $\leq^{\circledast}$.

By Proposition F.6, $(\downarrow x_1)^{?} \odot (\downarrow x_2)^{?} \odot \ldots \odot (\downarrow x_n)^{?}$ is (irreducible) closed. This is also the downward closure of $m = \{|x_1, x_2, \ldots, x_n|\}$ with respect to $\leq^{\circledast}$, so this must be the smallest closed set containing, i.e., the closure of $\{m\}$. It follows that, if m is smaller than m' in the specialization quasi-ordering of $X^{\circledast}$, then $m \leq^{\circledast} m'$. So $\leq^{\circledast}$ is the specialization quasi-ordering of $X^{\circledast}$.

Lemma F.9. *Let X be a topological space. Inclusion between m-products can be checked in non-deterministic polynomial time, modulo an oracle testing inclusion of closed subsets of X. Explicitly, let $P = F^{\circledast} \odot C_1^{?} \odot C_2^{?} \odot \ldots \odot C_m^{?}$ and $P' = F'^{\circledast} \odot C_1'^{?} \odot C_2'^{?} \odot \ldots \odot C_n'^{?}$ be two m-products. Then $P \subseteq P'$ if and only if $F \subseteq F'$ and, letting $I = \{i_1, i_2, \ldots, i_k\}$ be the subset of those indices i, $1 \leq i \leq m$, such that $C_i \not\subseteq F'$, there is an injective map $r : I \to \{1, 2, \ldots, n\}$ such that $C_i \subseteq C'_{r(i)}$ for all $i \in I$—in other words, $\{|C_{i_1}, C_{i_2}, \ldots, C_{i_j}|\} \subseteq_{\circledast} \{|C'_1, C'_2, \ldots, C'_n|\}$.*

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'. So $F \subseteq F'$.

Let now $I = \{i_1, i_2, \ldots, i_k\}$ be the set of indices i, $1 \leq i \leq m$, such that $C_i \not\subseteq F'$. 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 top-product P_1 defined as $E_1^? E_2^? \ldots E_{m-k}^? F^* D_1^? D_2^? \ldots D_k^?$ (if $F \neq \emptyset$), or $E_1^? E_2^? \ldots E_{m-k}^? D_1^? D_2^? \ldots D_k^?$ (if $F = \emptyset$). 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 $F'^* C'_{\pi(1)}{}^? F'^* C'_{\pi(2)}{}^? F'^* \ldots F'^* C'_{\pi(n)}{}^? F'^*$ (if $F' \neq \emptyset$), or of $C'_{\pi(1)}{}^? C'_{\pi(2)}{}^? \ldots C'_{\pi(n)}{}^?$ (if $F' = \emptyset$). Since P_1 is irreducible (Lemma E.10), there a permutation π of $\{1, 2, \ldots, n\}$ such that $P_1 \subseteq F'^* C'_{\pi(1)}{}^? F'^* C'_{\pi(2)}{}^? F'^* \ldots F'^* C'_{\pi(n)}{}^? F'^*$ (if $F' \neq \emptyset$), or $P_1 \subseteq C'_{\pi(1)}{}^? C'_{\pi(2)}{}^? \ldots C'_{\pi(n)}{}^?$ (if $F' = \emptyset$). Using Lemma E.13, and the fact that E_1, E_2, …, E_{m-k} are contained in F', and $F \subseteq F'$, and recalling the definition of F_1, we obtain that $D_1^? D_2^? \ldots D_k^?$ is included in $F'^* C'_{\pi(1)}{}^? F'^* C'_{\pi(2)}{}^? F'^* \ldots F'^* C'_{\pi(n)}{}^? F'^*$ (if $F' \neq \emptyset$), or in $C'_{\pi(1)}{}^? C'_{\pi(2)}{}^? \ldots C'_{\pi(n)}{}^?$ (if $F' = \emptyset$).

Let us deal with the case $F' \neq \emptyset$, as the case $F' = \emptyset$ is simpler. We show that there is an injective map $r : I \to \{\pi(1), \pi(2), \ldots, \pi(n)\}$ such that $C_i \subseteq C'_{r(i)}$ for all $i \in I$, by induction on k. If $k = 0$, the empty map fits. Otherwise, since $D_1 \not\subseteq F'$, using Lemma E.13, we must have $D_1^? D_2^? \ldots D_k^? \subseteq {C'_{\pi(1)}}^? F'^* {C'_{\pi(2)}}^? F'^* \ldots F'^* {C'_{\pi(n)}}^? F'^*$. Now we have two cases, again following Lemma E.13. In the first case $D_1 = C_{i_1} \subseteq C'_{\pi(1)}$ and $D_2^? \ldots D_k^? \subseteq F'^* {C'_{\pi(2)}}^? F'^* \ldots F'^* {C'_{\pi(n)}}^? F'^*$, so there is an injective map $r' : \{i_2, \ldots, i_k\} \to \{\pi(2), \ldots, \pi(n)\}$ such that $C_i \subseteq C'_{r'(i)}$ for all $i \in \{i_2, \ldots, i_k\}$. Then taking $r(i_1) = \pi(1)$ and $r(i) = r'(i)$ for all $i \in \{i_2, \ldots, i_k\}$ fits. In the second case, $D_1^? D_2^? \ldots D_k^? \subseteq F'^* {C'_{\pi(2)}}^? F'^* \ldots F'^* {C'_{\pi(n)}}^? F'^*$, and we conclude directly by the induction hypothesis.

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

The complexity of the above algorithm can be improved when X is finite, and its quasi-ordering is equality. This is the case considered for the so-called multiset language generators of [3, Section 5]: then irreducible closed subsets C are reduced to single letters, and $C_i \subseteq {C'}_{r(i)}$ is then equivalent to $C_i = {C'}_{r(i)}$. It therefore suffices to check that $\{\!|C_{i_1}, C_{i_2}, \ldots, C_{i_k}|\!\}$ is a sub-multiset of $\{\!|C'_1, C'_2, \ldots, C'_n|\!\}$, which can be done in quadratic time.

Finally, here is an analogue of the Coincidence Lemma E.4.

Lemma F.10 (Coincidence Lemma). *Let X be a set equipped with a quasi-ordering $\leq$. We see X as equipped with the Alexandroff topology of $\leq$. Then the subword topology on $X^\circledast$ is the Alexandroff topology of $\leq^\circledast$.*

Proof. Any upward-closed subset A of $X^\circledast$ is a union of sets of the form $X^\circledast \odot (\uparrow x_1) \odot (\uparrow x_2) \odot \ldots \odot (\uparrow x_n)$, namely all those obtained by taking the upward closures of elements $\{\!|x_1, x_2, \ldots, x_n|\!\}$ in A; indeed $X^\circledast \odot (\uparrow x_1) \odot (\uparrow x_2) \odot \ldots \odot (\uparrow x_n)$ is just the upward closure of $\{\!|x_1, x_2, \ldots, x_n|\!\}$ in $\leq^\circledast$. Since these are basic opens of the sub-multiset topology, the sub-multiset topology on $X^\circledast$ is contained with the Alexandroff topology of $\leq^\circledast$. The converse is by Proposition F.8. ∎

Appendix G. Powersets

For any topological space X, let $\mathbb{P}_\diamond(X)$ denote the powerset $\mathbb{P}(X)$, with the *lower Vietoris* topology, which is the least one containing all opens of the form $\diamond U = \{A \in \mathbb{P}(X) \mid A \cap U \neq \emptyset\}$, where U ranges over the open subsets of X.

Lemma G.1. *Let X be a topological space, with specialization quasi-ordering $\leq$. The specialization quasi-ordering of $\mathbb{P}_\diamond(X)$ is the (topological)* Hoare quasi-ordering $\leq^\flat$*, defined by: $A \leq^\flat B$ if and only if $A \subseteq cl(B)$, if and only if $cl(A) \subseteq cl(B)$, where $cl : \mathbb{P}(X) \to \mathcal{H}(X)$ is the closure operator.*

The closure of $\{A\}$ in $\mathbb{P}_\diamond(X)$ is $\Box cl(A)$, where $\Box F$ is defined as $\{B \in \mathbb{P}(X) \mid B \subseteq F\}$.

Proof. This is well-known. Let $\leq^\flat$ be the specialization quasi-ordering of $\mathbb{P}_\diamond(X)$. We show that $A \leq^\flat B$ if and only if $cl(A) \subseteq cl(B)$.

If $A \leq^\flat B$, then in particular, for every open subset U of X, if $A \in \diamond U$ then $B \in \diamond U$. In particular, take U the complement of $cl(B)$. Clearly B is not in $\diamond U$. So A is not in $\diamond U$ either, i.e., $A \subseteq cl(B)$. So $cl(A) \subseteq cl(B)$.

Conversely, if $cl(A) \subseteq cl(B)$, let $\mathcal{U}$ be any open of $\mathbb{P}_\diamond(X)$ containing A. Note that $A \subseteq cl(B)$. Write $\mathcal{U} = \bigcup_{i \in I} \bigcap_{j=1}^{n_i} \diamond U_{ij}$, where U_{ij} is open in X. Since $A \in \mathcal{U}$, there is an $i \in I$ such that A

intersects each U_{ij}, $1 \leq j \leq n_i$. Since $A \subseteq cl(B)$, $cl(B)$ intersects each U_{ij}, $1 \leq j \leq n_i$. Since each U_{ij} is open, B itself intersects each U_{ij}, $1 \leq j \leq n_i$, so $B \in \mathcal{U}$. That is, $A \leq^\flat B$.

Now the closure of $\{A\}$ is the set of all B with $B \leq^\flat A$, i.e., with $B \subseteq cl(A)$. So this is $\Box cl(A)$. ■

Note that $\mathbb{P}_\diamond(X)$ is far from being T_0, as there are many elements of this space which are equivalent with respect to the equivalence relation generated by $\leq^\flat$. Namely, $A \leq^\flat B$ and $B \leq^\flat A$ if and only if A and B have the same closure.

Corollary G.2. *The topology of $\mathbb{P}_\diamond(X)$ is the upper topology of $\leq^\flat$.*

Proof. Any downward closure of an element A of $\mathbb{P}_\diamond(X)$ is by definition of the form $\{B \in \mathbb{P}(X) \mid B \subseteq cl(A)\} = \Box cl(A)$, and is therefore closed in $\mathbb{P}_\diamond(X)$ by Lemma G.1. So the topology of $\mathbb{P}_\diamond(X)$ is finer than the upper topology. Conversely, the complement of $\diamond U$, U open in X, is $\Box F$, where F is the complement of U in X. But $\Box F$ is the downward closure of F in $\mathbb{P}_\diamond(X)$, and is therefore closed in the upper topology. Hence every $\diamond U$ is open in the upper topology, so the upper topology is finer than the topology of $\mathbb{P}_\diamond(X)$. ■

Theorem G.3. *Let X be a topological space. Then X is Noetherian if and only if $\mathbb{P}_\diamond(X)$ is.*

Proof. Proposition 7.3 of [14] states that, if X is Noetherian, then $\mathbb{P}(X)$ is Noetherian, when equipped with the upper topology of $\leq^\flat$. By Corollary G.2, $\mathbb{P}_\diamond(X)$ is then Noetherian. Conversely, any infinite increasing chain of opens $U_1 \subset U_2 \subset \ldots \subset U_k \subset \ldots$ in X induces an infinte increasing chain of opens $\diamond U_1 \subset \diamond U_2 \subset \ldots \subset \diamond U_k \subset \ldots$ in $\mathbb{P}_\diamond(X)$, so if $\mathbb{P}_\diamond(X)$ is Noetherian, then so is X. ■

Proposition G.4. *Let X be a Noetherian space. Then the sobrification of $\mathbb{P}_\diamond(X)$ is the Hoare powerdomain $\mathcal{H}(X)$ of X, up to homeomorphism. Precisely, the map $F \mapsto \Box F$ is a homeomorphism of $\mathcal{H}(X)$ onto $\mathcal{S}(\mathbb{P}_\diamond(X))$.*

Proof. The closed subsets of $\mathbb{P}_\diamond(X)$ are the intersections of finite unions of closures of single elements. The closure of $A \in \mathbb{P}_\diamond(X)$ is $\Box cl(A)$ by Lemma G.1. Also, $\mathbb{P}_\diamond(X)$ is Noetherian by Theorem G.3, so any intersection of closed sets is a finite intersection. Hence every closed subsets $\mathcal{F}$ of $\mathbb{P}_\diamond(X)$ can be written as a finite intersection of finite unions of sets of the form $\Box F$, F closed in X. Distributing unions over intersections, $\mathcal{F}$ is a finite union of finite intersections of sets of the form $\Box F$, F closed in X. Now it is easy to show that $\bigcap_{i=1}^n \Box F_i = \Box \bigcap_{i=1}^n F_i$ (this denoting $\Box X$ if $n = 0$), so $\mathcal{F}$ is a finite union of closed sets of the form $\Box F$, F closed in X. If $\mathcal{F}$ is irreducible, then it must be of the form $\Box F$.

Conversely, if $\Box F$ (F closed in X) is contained in the union of two closed subsets of $\mathbb{P}_\diamond(X)$, then these closed subsets can be written as $\mathcal{F}_1 = \bigcup_{i=1}^m \Box F_i$ and $\mathcal{F}' = \bigcup_{j=1}^n F'_j$ respectively, where the F_is and the F'_js are closed in X. In particular, F, which is in $\Box F$, is contained in some F_i or in some F'_j. If $F \subseteq F_i$, then $\Box F \subseteq \Box F_i$, hence $\Box F \subseteq \mathcal{F}$. If $F \subseteq F'_j$, then similarly $\Box F \subseteq \mathcal{F}'$. So $\Box F$ is irreducible.

The map $F \mapsto \Box F$ therefore maps any $F \in \mathcal{H}(X)$ to an element of $\mathcal{S}(\mathbb{P}_\diamond(X))$.

This map is clearly surjective: we have shown above that any irreducible closed set of $\mathbb{P}_\diamond(X)$ was of the form $\Box F$ for some closed subset F of X. It is injective. Indeed, $\Box F \subseteq \Box F'$ implies $F \in \Box F'$, hence $F \subseteq F'$. In particular if $\Box F = \Box F'$ then $F \subseteq F'$ and $F' \subseteq F$, hence $F = F'$.

The map $F \mapsto \Box F$ is continuous: it suffices to show that the inverse image of the open subset $\diamond \mathcal{U}$ is open in $\mathcal{H}(X)$ for any open subset $\mathcal{U}$ of $\mathbb{P}_\diamond(X)$. Equivalently, to show that the inverse image of the closed subset $\Box \mathcal{F}$ is closed in $\mathcal{H}(X)$ for any closed subset $\mathcal{F}$ of $\mathbb{P}_\diamond(X)$. Now $\mathcal{F}$ can be written as a finite union $\bigcup_{i=1}^{m} \Box F_i$, where each F_i is closed in X. The inverse image of $\Box \bigcup_{i=1}^{n} \Box F_i$ is the set of closed subsets F of X such that $\Box F \in \Box \bigcup_{i=1}^{n} \Box F_i$, i.e., such that $\Box F \subseteq \bigcup_{i=1}^{n} \Box F_i$, i.e., such that $F \subseteq \Box F_i$ for some i. In other words, this inverse image is $\bigcup_{i=1}^{n} \Box F_i$, which is indeed closed.

Finally, we must show that the inverse of this map is continuous, i.e., that the direct image of a closed set is closed. It is enough to show that the direct image of $\Box F$ is closed in $\mathcal{S}(\mathbb{P}_\diamond(X))$. This direct image is the set of all $\Box F'$, where F' ranges over the closet sets of X with $F' \subseteq F$; equivalently, with $\Box F' \subseteq \Box F$, i.e., with $\Box F' \in \Box\Box F$. So the direct image of $\Box F$ is $\Box\Box F$: the set of irreducible closed subsets $\Box F'$ that are contained in $\Box F$. ∎

Corollary G.5. *For any Noetherian space X, $\mathcal{H}(X)$ is sober.*

Proof. As an homeomorph of $\mathcal{S}(\mathbb{P}_\diamond(X))$, which is sober by construction. ∎

Note that we also know that $\mathcal{H}(X)$ is then Noetherian [14, Theorem 7.2].

There is in general no coincidence Lemma as for words (Lemma E.4) or multisets (Lemma F.10), otherwise powersets of wqos would be wqo, too.

But elements of $\mathcal{S}(\mathbb{P}_\diamond(X))$, i.e., of $\mathcal{H}(X)$ up to homeomorphism, can all be represented finitely, as finite sets A of elements of $\mathcal{S}(X)$. (Assuming X Noetherian.) These are interpreted as $\bigcup_{C \in A} C$. This follows from the fact that any closed subset of X is a finite union of irreducible closed subsets (Proposition 4.2.)