Rapport de Mi-Parcours

Well-quasi-orderings and Database Driven Systems

Aliaume Lopez
LMF, École Normale Supérieure Paris-Saclay, Université Paris-Saclay
IRIF, Université de Paris

29th March 2022

UNDER THE SUPERVISION OF

- Sylvain Schmitz
- Jean Goubault-Larrecq

WITH REVIEWERS

- Sernella Cerrito
- Arnaud Durand

Abstract

This thesis takes places at the crossroad of several fields of computer science: combinatorics (through well-quasi-orderings), topology (through Nœtherian spaces), and logic (through Finite Model Theory). Ideally, one would study database related problems through the lens of logic, and provide combinatorial arguments for their resolution using the theory of well-quasi-orderings. However, well-quasi-orderings are not well-suited to describe logical properties, and Finite Model Theory is more prone to defining topologies. As a glue between these two realms, Nœtherian topologies, the analogue of well-quasi-orderings in the topological setting, allows to amorce a dialogue between these classical order-theoretic results and finite model theory.

A particular instance of this interplay, studied in this thesis, is the wide variety of preservation theorems, relating syntactic fragments of First Order Logic (e.g.unions of conjunctive queries) to semantic fragments of First Order Logic (e.g.queries preserved under extensions). There exists a plethora of such preservation theorems in classical model theory but most of them fail in the finite. As the classical proofs crucially rely on the compactness theorem of first order logic, this is a fitting place to study the interplay between finite model theory and notions of compactness in topological spaces. Our first contribution is to study preservation theorems through the lens of topology via "topological preservation theorems", or, more formally, logically presented pre-spectral spaces, where the topological setting is ideal to describe algebraic properties of those theorems. However, this topological descriptions has its limits and we provided a way to "decouple" topology from logic though the study of localisable classes of structures and a positive variant of the Gaifman Normal Form, a fundamental tool of finite model theory.

A second interplay, with well-quasi-ordering, is to generalise the "minimal bad sequence arguments" at the heart of many proofs in the theory, to a categorical setting. A main consequence is the existence of a canonical Nœtherian topology associated to an initial algebra. Using these results, we re-interpret existing Nœtherian topologies, and develop new topologies, beyond the reach of well-quasi-orderings, such as the recurrent topology over transfinite words.

Finally, bridging finite model theory and well-quasi-orderings, this thesis explored graph classes that are well-quasi-ordered for the induced substructure ordering. This line of research was driven by the hope that one could characterise concrete classes of graphs defined by an interpretation from tree-like structures that are well-quasi-ordered for the induced substructure ordering. The hope was to provide a decision procedure for the existence of simple algorithms based on well-quasi-ordering theory, but this did not go through.

The contributions and their relative position with respect to the three main domains of this thesis are described in Figure 1.

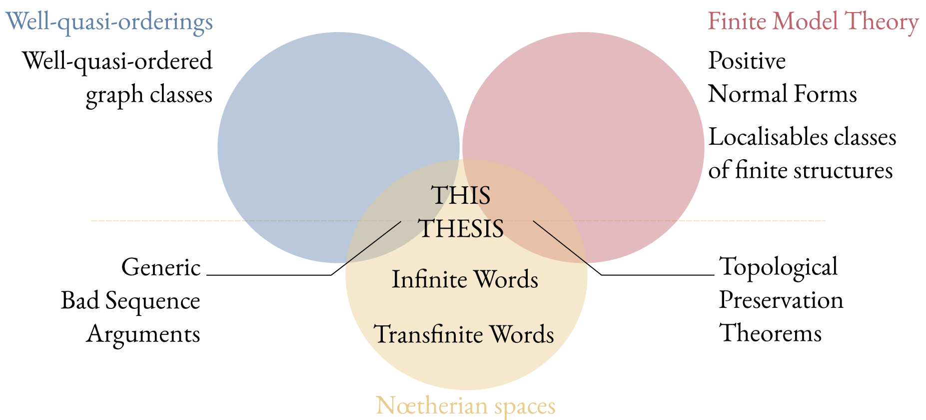


Figure 1: Topology as a bridge between model theory and well-quasi-orderings.

CONTENTS

SCIENTIFIC CONTEXT

I

This section is devoted to the introduction of three core notions of this thesis: 1. well-quasi-orders in Section 1.1 2. Noetherian spaces in Section 1.2 3. finite model theory in Section 1.3. These notions will have numerous and fruitful interplay in Section 2, devoted to the scientific production of this thesis. None of the results and definitions appearing hereafter are new, but may be spelled out in a slightly unorthodox manner to keep consistency between the different fields of research.

1 WELL QUASI ORDERINGS

A *poset* is a set X equipped with a transitive, reflexive and antisymmetric relation $\leq$. A *quasi-order* is a set X equipped with a transitive and reflexive relation $\leq$. As one can always quotient a quasi-order to obtain a poset, we will sometimes mix those denominations to match the convention in the literature. In a poset, two elements might be incomparable for the relation $\leq$, this will be written $x \perp y$.

Sequences in a Poset

To describe termination of programs, one relies on well-foundedness of some quasi-ordered set $(X, \leq)$. Concretely, one wants to avoid *disastrous sequences*, that are sequences $(x_n)_{n\in\mathbb{N}}$ of $(X, \leq)$ that are strictly decreasing, i.e. such that for all $i < j$, $x_i > x_j$. A quasi-ordered set $(X, \leq)$ is *well-founded* whenever no sequence is disastrous. The prototypical example of well-founded total ordering is $(\mathbb{N}, \leq)$, the natural numbers with the usual ordering. Given an alphabet Σ, one can order the finite words over Σ using the *suffix ordering* $\sqsubseteq$, building the well-founded poset $(\Sigma^*, \sqsubseteq)$.

More often than not, one cannot prove that a given sequence is disastrous, but rather that it cannot contain an increasing pair, which happens to be two different properties when the ordering is not total. For that purpose, many authors introduced the notion of good sequence. A sequence $(x_n)_{n\in\mathbb{N}}$ is *good* whenever there exists $i < j$ such that $x_i \leq x_j$, as shown in Figure 2. A *well-quasi-order* is a quasi-order where every infinite sequence is good. A *bad sequence* is a sequence that is not good. Hence, a wqo is a poset having no infinite bad sequence.

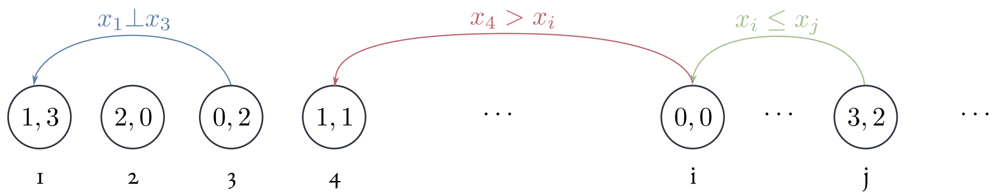


Figure 2: A good sequence in the poset $\mathbb{N} \times \mathbb{N}$ ordered pointwise.

Example 1.1 (Examples of well-quasi-orderings)**.** The following quasi-orders are wqos:

- $(\mathbb{N}, \leq)$, the natural numbers equipped with the usual ordering,
- $(\mathbb{F}, =)$ a finite set equipped with equality,
- $(\mathcal{G}, \leq_{\text{minor}})$ the class $\mathcal{G}$ of finite graphs equipped with the minor ordering $\leq_{\text{minor}}$.

Example 1.2 (Non-examples of well-quasi-orderings)**.** The following quasi-orders are not wqos:

- Words with the suffix ordering $\sqsubseteq$.
- An infinite set equipped with equality.
- $(\mathcal{G}, \subseteq_i)$ the class of finite graphs equipped with the induced substructure ordering.

To better understand the notions of good sequence and bad sequence we provide in Table 1 some examples of such sequences in various posets. In this table, we will refer to the finite simple cycle of size i as C_i.

The *downwards closure* of a set $E \subseteq X$ is written $\downarrow E$ and is the set of all elements of X below some element of E, i.e. $\downarrow E \triangleq \{x \in X \mid \exists y \in E, x \leq y\}$. Similarly, one defines the *upwards closure* via $\uparrow E \triangleq \{x \in X \mid \exists y \in E, y \leq x\}$. A subset E of X is *upwards closed* whenever $\uparrow E = E$.

An *antichain* is a sequence $(x_i)_{i\in I}$ of elements such that for all $i \neq j$, $x_i \perp x_j$. By definition, an antichain is a bad sequence, and the following proposition states that it is enough to forbid antichains and infinite decreasing sequences to forbid arbitrary bad sequences.

Quasi-Order	Sequence	Good
$(\mathbb{N}, =)$	$i \mapsto i$	✗
$(\mathbb{N}, \leq)$	$i \mapsto i$	✓
$(\{a,b\}, \sqsubseteq)$	$i \mapsto a^i$	✓
$(\{a,b\}, \sqsubseteq)$	$i \mapsto ba^i$	✗
$(\{a,b\}, \leq_h)$	$i \mapsto ba^i$	✓
$(\mathcal{G}, \subseteq_i)$	$i \mapsto C_i$	✗
$(\mathcal{G}, \leq_{\text{minor}})$	$i \mapsto C_i$	✓

Table 1: Examples and non examples good sequences $(x_i)_{i \geq 1}$.

Proposition 1.3 (Equivalent definitions of wqos)**.** *The following definitions are equivalent for a poset* $(X, \leq)$*.*

(i) X *is a wqo.*
(ii) X *has no infinite decreasing sequence and no infinite antichains.*
(iii) X *has no strictly decreasing sequence of downwards closed sets for inclusion.*
(iv) X *has no strictly increasing sequence of upwards closed sets for inclusion.*

A Grammar for well-quasi-orders

We now turn our attention to the relations between wqos and their algebraic properties. These properties mentioned are crucial as we will compare other properties by their ability to enjoy them.

Interpretations. A map $f \colon X \to Y$ is *monotone* whenever $x \leq_X y \implies f(x) \leq_Y f(y)$ for all $(x, y) \in X$. It is an easy check that if X is a well-quasi-order and f is a monotone surjective function, then Y is also a well-quasi-order. In particular, this allows to quotient wqos.

Conversely, if there exists a map $f \colon X \to Y$ such that $f(x) \leq_Y f(y) \implies x \leq_X y$ for all $(x, y) \in X$, then Y wqo implies X wqo, such a map is called an *order reflection*. This implies that subsets of a wqo are wqos.

Concretely, one can use an order reflection to prove that a subset of a wqo is a wqo, and monotone surjective maps to prove that the quotient of a wqo is wqo. These stabilities under subsets and quotients, while natural on posets, will be lacking when studying the otherwise analogous notion of "preservation theorem" in Section 1.3.

Sums and Products. If A and B are two wqos, then $A + B$ with the disjoint orderings is wqo, and $A \times B$ with the usual product ordering is wqo. This immediately allows considering finite sums and products of wqos to build new wqos.

Example 1.4. The set of unordered pairs $\{a, b\}$ equipped with $\{a, b\} \leq \{c, d\}$ when $a \leq c \wedge b \leq d$ or $a \leq d \wedge b \leq c$ is a wqo.

Proof. The set $(\mathbb{N}, \leq)$ is a wqo, hence $\mathbb{N} \times \mathbb{N}$ ordered pointwise is a wqo, therefore its quotient under the action of the permutation group S_2 is a wqo. □

Noticeably, infinite sums and infinite products fail to be wqo in general. This fails because infinite sums of non-empty wqos contain an infinite antichain. For infinite products $\prod_{i \in I} X_i$, taking $X_i = \{a, b\}$ with $a < b$ allows to build a strictly decreasing infinite sequence, although X_i was wqo for all $i \in I$. In categorical terms, this means that wqos have equalisers (quotients) but only finite products and co-products, hence that one cannot, in general, consider the limit/co-limit of a diagram of wqos. This is a problem because most of the data structures in computer science are defined inductively, and therefore naturally appear as such co-limits.

Finite words. Surprisingly, one can consider words over a wqo and endow it with an ordering making it wqo. Note that the subfactor ordering, suffix ordering, and prefix ordering are not well-quasi-orders over finite words over a binary alphabet. One could hope that by seeing words Σ^* as the construction $\sum_{i \geq 0} \Sigma^i$, the associated *product ordering* on words defined as $u \leq v$ if $|u| = |v|$ and $u_i \leq v_i$ for $1 \leq i \leq |u|$ would be wqo. As expected, the product ordering is not wqo on words as the sequence a^n is an infinite antichain. To tackle this issue, one defines the *Higman word embedding*, where $u \leq_h v$ if there exists a strictly increasing function f such that $u_i \leq v_{f(i)}$ for $1 \leq i \leq |u|$. In practise, bigger words are obtained by "increasing the letters" and "inserting arbitrary factors", as depicted in Figure 3.

Higman's word embedding can be seen "algebraically" as the transitive closure of the usual product ordering and the suffix ordering, but is also directly related to the notion of structure embedding in finite model theory. Indeed, a word w is a finite structure equipped with a total order and unary predicates, and Higman's word embedding is the actual structure embedding of words. Over a unary alphabet $\Sigma \triangleq \{\star\}$, the poset $(\Sigma^*, \leq_h)$ is isomorphic to $(\mathbb{N}, \leq)$.

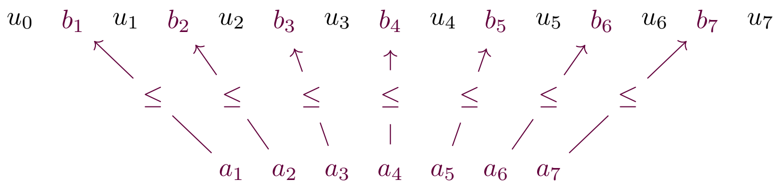


Figure 3: Higman ordering

Extended constructors. One can quotient finite words by permutation of letters to obtain a representation of finite multisets with the *multiset embedding*[1], and forget about the number of occurrences to obtain finite sets with the *Hoare embedding*. Concretely, $E \leq^\flat E'$ whenever $\downarrow E \subseteq \downarrow E'$. This provides new datatype constructors for wqos.

Following the intuition of structure embeddings, we can define *Kruskal's tree embedding* over finite trees by considering them as structures equipped with the tree ancestor relation. This provides one more constructor of wqo, and completes the "basic grammar of well-quasi-orderings" described in Figure 4.

$D ::= (\mathbb{F}, =)$	finite set
$\mid (\mathbb{N}, \leq)$	natural numbers
$\mid \Sigma_{i=1}^{n} D_i$	finite disjoint sums
$\mid \Pi_{i=1}^{n} D_i$	finite products
$\mid D^*$	finite words, subword embedding
$\mid D^{\circledast}$	finite multisets, multiset embedding
$\mid \wp_f(D)$	finite sets, Hoare embedding
$\mid T(D)$	finite trees, Kruskal embedding

Figure 4: Grammar for building well-quasi-orderings.

Extensions of the Grammar

Building up on the ideas from the Higman word embedding and the Kruskal tree embedding one can extend the grammar of wqos pursuing either the idea of "quotients of inductive datatypes", the one of "structure ordering", or the one of "word combinatorics".

Structure orderings. Given a relational signature σ, one can consider $\mathrm{Struct}(\sigma)$ the class of structures over this relational signature. As relational structures, elements of $\mathrm{Struct}(\sigma)$ can be compared by the way of *homomorphisms*, i.e. functions f from the domain of one structure to the domain of another one, respecting the interpretation of relations. This comparison gives rise to the notion of *homomorphism pre-order*, $A \leq_{\mathrm{hom}} B$ whenever there exists a homomorphism $h\colon A \to B$. This ordering has no particular reason to be wqo in general, and might not even be well-founded.

By restricting the range of accepted homomorphism, one obtains finer quasi-orders. The one that is studied extensively in this thesis is *structure embedding* or *induced substructure* that puts the restriction on homomorphisms $h\colon A \to B$ to be both *injective* and *strong*, i.e. a and b are in relation if A if and only if $h(a)$ and $h(b)$ are in

[1] That may differ from the usual multiset extension of a total order.

relation in B. When there exists a structure embedding between A and B, we write $A \subseteq_i B$, and can also write that B is an extension of A. The quasi-order $\subseteq_i$ is called the *extension quasi-order*. This ordering may not be wqo but is well-founded over the class of finite structures $\mathrm{Fin}(\sigma)$. As we have seen before, both Higman's word embedding and Kruskal's tree embedding are directly connected to this notion of structure embeddings, as they relate to the extension quasi-order of the underlying structures. However, there has not been characterisations of classes of finite structures for which the extension quasi-order is wqo, apart from the very specific case of words, trees and similar inductive data-structures.

Divisibility Quasi-Orderings on Inductive Datatypes. There are similarities between the word embedding and tree embedding pre-orders beyond their model theoretic definition through induced substructures. As both the constructions are inductive datatypes, they enjoy a notion of *subterm*, e.g. subtrees or suffixes. We can therefore devise a notion of *subterm ordering*, defined by $A \sqsubseteq_{\text{term}} B$ whenever A is a subterm of B.

It turns out that the tree embedding is also the transitive closure of the subterm ordering with the "pointwise ordering" of trees. In the case of binary trees, the pointwise ordering is defined recursively as in Figure 5. Moreover, the actual proofs that the tree embedding and the word embedding are wqos use, at their core, an identical combinatorial argument: Nash-Willam's *minimal bad sequence argument*.

$$\frac{a \leq b}{a \cdot () \leq_p b \cdot ()} \qquad \frac{a \leq b \qquad t_1 \leq_p t_1' \qquad t_2 \leq_p t_2'}{a \cdot (t_1, t_2) \leq_p b \cdot (t_1', t_2')}$$

Figure 5: Recursive definition of the pointwise ordering on binary trees.

This ressemblance has been noted and used by Hasegawa to define, in full generality, the *divisibility ordering* of an inductive construction [29, Theorem 2.10]. For different reasons, this idea of divisibility ordering appears in recent work from Freund [18]. This allows to add a "fixpoint combinator" to the grammar in Figure 4. As a sanity check, one can see that Σ^* is the least fixed point of the constructor $F_\Sigma(X) \triangleq 1 + \Sigma \times X$, and the divisibility ordering over $\Sigma^* = \mu X.F_\Sigma(X)$ is exactly Higman's word embedding.

These inductive constructions can be used to recover a well-known "complex" wqo, namely the gap embedding. The concurrent approaches from Hasegawa and Freund are both the result of a categorical study of wqos, and we will not provide related definitions in this mid-term report.

Induced Substructures on Graph Classes. Although the graph minor ordering is a wqo, the more natural model theoretic notion is the one of structure embedding, which corresponds over graphs to the *induced subgraph ordering*, defined by $G \subseteq_i G'$ whenever there exists an injective map $f : G \to G'$ such that $(f(u), f(v)) \in G'$ if and only if $(u, v) \in G$ (see Figure 6 for examples).

Unfortunately, the induced subgraph ordering is not wqo, as the family of cycles forms an infinite antichain of finite graphs. This started a quest for large classes of finite graphs that are well-quasi-ordered with respect to $\subseteq_i$, the latest positive result [11] provides a characterisation of classes of bounded clique-width that are wqo. Note that a characterisation of classes of finite graphs that are wqo for the "subgraph ordering" (not induced) has been provided by Ding [15], but does not extend to the notion of induced subgraph.

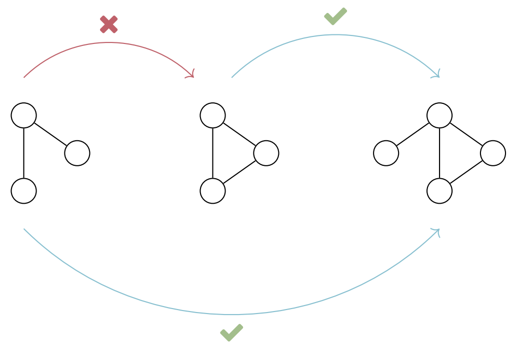
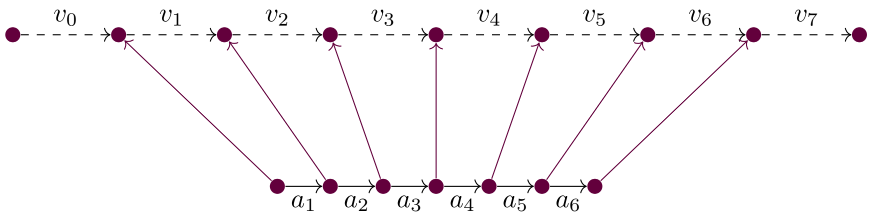


Figure 6: Examples and non examples of induced subgraphs.

Monoid Induced Well-Quasi-Orders. Another way to augment the range of wqos beyond divisibility orderings is to take "context" into account, that is, disallow to compare arbitrary substructures regardless of their respective context. This quite informal statement can be made rigorous on words over an alphabet Σ using a monoid M and a morphism $\mu\colon \Sigma \to M$. We say that a word $u = a_1 \dots a_n$ is below a word v whenever $v = v_1 \dots v_n$ and $\mu(a_i) = \mu(v_i)$ for $1 \leq i \leq n$. This is the *composition preorder* described in Figure 7, that already has been quite extensively studied (in different formats) by Ehrenfeucht [17], Bucher, Ehrenfeucht, and Haussler [6], Kirsten [31], Kunc [34], Tzameret [49], Dershowitz and Tzameret [13].

This composition preorder can be then used to build new classes of graphs for which the extension preorder is wqo via interpretations. Given a word w, a monoid M, a part $P \subseteq M$ and a morphism $\mu\colon \Sigma \to M$, one can build a graph G_w whose set of vertices is the positions in w and with an edge (i,j) whenever $\mu(w_{i,j}) \in P$. This recovers the notion of graphs of bounded linear clique width. If the composition quasi-order is wqo, then the class of graphs G_w is also wqo for induced substructures as $w \leq_\mu^M w'$ implies $G_w \subseteq_i G_{w'}$. This remark lead to the characterisation of classes graphs of bounded clique width that are wqo by Daligault, Rao, and Thomassé [11].



$$\forall i \in \{1, \dots, 6\}, \mu(v_i) = \mu(a_i)$$

Figure 7: Composition ordering

2 NOETHERIAN SPACES

While being better behaved than well-founded pre-orderings with respect to most of the usual set constructions (products, words, trees, etc.), some natural ones fail to preserve well-quasi-orderings, such as the (infinite) powerset construction [42].

There are two main possibilities to tackle this issue. The first one is to strengthen the definition of well-quasi-ordering to ensure that Rado's structure cannot be built. This is the path leading to the theory of better-quasi-orderings [41, 38]. A different approach is taken by Goubault-Larrecq [22], who proceeds to weaken the definition of a well-quasi-order to a notion of Noetherian topological space. This shift from quasi-orders to topologies resolves these stability properties through careful adaptation of the topologies [23]. However, this approach suffers from the fact that many topologies can correspond to a single quasi-ordering and there usually does not exist a finest one that is Noetherian. Notice that this lack of canonicity already arises from the definitions of the well-quasi-orderings on words or trees, although this is partially answered by the divisibility preorder [29, 18].

Topological Spaces Cheatsheet

A set X can be equipped with a *topology*, that is, a set $\tau \subseteq \wp(X)$ stable under finite intersections and arbitrary unions. In particular, $\emptyset \in \tau$ as the empty union and $X \in \tau$ as the empty intersection. Given a topological space (X, τ) we say that a given subset $E \subseteq X$ is an *open set* if $E \in \tau$ and a *closed set* if $X \setminus E \in \tau$. We usually denote open sets with letters U, V and closed sets F, C. The sets $\{\emptyset, X\}$ and $\wp(X)$ are always topologies over X, the first one is called the *trivial topology* and the latter the *discrete topology*. Note that the topologies over a given set X form a complete lattice for inclusion.

Be careful that "usual" topological spaces come from a notion of distance between points: those are metric spaces, but many the topologies considered in computer science will not be metric at all. An example of a usually non-metric topology is the *cofinite topology*, where open sets are complement of finite sets. To build the right intuition on the spaces that are going to be considered, we provide a translation between posets and topological spaces. Given a quasi-ordered set $(X, \leq)$, one can build the *Alexandroff topology* $\tau_{\leq}$ over X by collecting all subsets of X that are upwards-closed. We now refer to Table 2 for a conversion between topological properties and order properties.

Pre-order $\leq$	**Topology** $\tau_{\leq}$
U is upwards-closed f is monotone	U is open f is continuous
E has finitely many minimal elements	E is compact
wqo	Noetherian

Table 2: Cheatsheet relating concepts between a quasi-order and its Alexandroff topology.

In Table 2, we use the following definition of compactness: E is *compact* in (X, τ) whenever for every family $(U_i)_{i \in I}$ of opens such that $E \subseteq \bigcup_{i \in I} U_i$, there exists a finite subset $J \subseteq I$ such that $E \subseteq \bigcup_{j \in J} U_j$. The family $(U_i)_{i \in I}$ is called an *open cover* of E and $(U_j)_{j \in I}$ is a *subcover* of $(U_i)_{i \in I}$. Recall from Proposition 1.3 that $(X, \leq)$ is a wqo if and only if strictly increasing sequences of upwards closed subsets of X are finite. Using our cheatsheet, we can define *Noetherian spaces* as those where increasing sequences of open sets stabilise in finite time. In particular, $(X, \leq)$ is wqo if and only if $(X, \tau_{\leq})$ is Noetherian.

We provide in Table 3 examples of compact and non-compact spaces, using simple topologies that we define hereafter. Given a poset $(X, \leq)$, we already defined the Alexandroff topology, but it is possible to build a coarser topology from this order: the *upper topology* $\mathrm{Upper}(\leq)$ is the smallest topology containing the downwards closures of points in X as closed sets. Every open in the upper topology is open in the Alexandroff topology but the converse is not true in general.

Proposition 1.5 (Equivalent definitions of Noetherian spaces)**.** *The following propositions are equivalent for a space (X, τ)*

(i) (X, τ) is Noetherian.

(ii) Every subset of X is compact.

(iii) Strictly decreasing sequences of closed sets are finite.

Space	Topology	Compact	Nœtherian
$\mathbb{N}$	$\tau_{\leq}$	✔	✔
$\mathbb{N}$	cofinite	✔	✔
$\mathbb{N}$	discrete	✘	✘
Σ^*	$\mathrm{Upper}(\sqsubseteq)$	✔	✔
Σ^*	$\tau_{\sqsubseteq}$	✔	✘
Σ^*	$\tau_{\leq_h}$	✔	✔
$\mathbb{R}$	metric	✘	✘
$[0,1]$	metric	✔	✘
$\mathbb{C}$	Zariski	✔	✔
Σ^*	regular subword	✔	✔

Table 3: Examples and non examples of Nœtherian spaces.

A Grammar for Noetherian Spaces

True Noetherian Spaces. One may wonder whether some Noetherian spaces can be found without building them from an already existing well-quasi-ordering. This natural question has two very different answers. One may argue that Noetherian spaces are closed under some operations for which well-quasi-orderings are not, such as the powerset [22], ω-words [24] or even transfinite ordinal words [25]. In that sense, even if some Noetherian spaces come from well-quasi-orderings, they quickly leave the realm of Alexandroff topologies.

A second argument is to consider spaces that have either a pre-existing topological nature, or better, an algebraic one. A classical example following this pattern is the complex plane $\mathbb{C}$ endowed with the *Zariski topology* where the closed sets are the *algebraic sets* $V(S) \triangleq \{z \in \mathbb{C} \mid \forall f \in S, f(z) = 0\}$ where S is a subset of $\mathbb{C}[X]$.

A Noetherian spaces program. While there exists canonical topologies for product spaces and sum spaces, there can be a vast range of topologies to place on finite words or finite trees. A first try would be to mimic the construction on wqos via Alexandroff topologies, but we also have to treat the case where the input space is not endowed with an Alexandroff topology, in which case we have no idea where to start.

A sanity check can be given through a conversion *from* topologies *to* quasi-orders: the *specialisation pre-order* $\mathrm{spec}(\tau)$ of a topology τ is defined by $x \ \mathrm{spec}(\tau)\ y$ if $\forall U \in \tau, x \in U \implies y \in U$. The specialisation pre-order of the Alexandroff topology of a quasi-order $\leq$ is the preorder itself, but several topologies can share the same specialisation pre-order, and this is the key observation allowing to escape the classical roadblocks in the theory of well-quasi-orderings.

In his "Noetheran spaces program", Goubault-Larrecq developed topological analogues of constructions on wqos having the following properties: (a) they preserve Noetherian spaces (b) if F_T is the operation on topologies and F_O the operation on quasi-orderings, the following equation holds $\mathrm{spec}(F_T(\tau)) = F_O(\mathrm{spec}(\tau))$. We provide the grammar in Figure 8 with fewer details than in the case of wqos Figure 4.

$D ::= (X, \tau_{\leq})$	$(X, \leq)$ wqo
$\mid \Sigma_{i=1}^n D_i$	finite disjoint sums
$\mid \Pi_{i=1}^n D_i$	finite products
$\mid D^*$	finite words, regular subword topology
$\mid D^{\circledast}$	finite multisets, multiset topology
$\mid \wp(D)$	arbitrary subsets, lower Vietoris topology
$\mid T(D)$	finite trees, regular subtree topology
$\mid \mathcal{S}(D)$	sobrification
$\mid X^\omega$	infinite words, regular subword topology

Figure 8: Grammar for building Noetherian spaces.

3 FINITE AND ALGORITHMIC MODEL THEORY

We are interested in first-order logic over a finite relational signature σ, written FO[σ], and several syntactic variants of FO, such as EFO[σ] the set of *existential sentences*, that is, sentences of the form $\exists \vec{x}.\psi$ where ψ is quantifier free. One can also be interested in PFO, the set of sentences having no negations but an explicit "inequality" predicate, EPFO, the set of sentences having nor negations, nor universal quantifiers.

Preservation Theorems. In classical model theory, preservation theorems characterise first-order definable sets enjoying some semantic property as those definable in a suitable syntactic fragment [e.g., 7, Section 5.2]. A well-known instance of a preservation theorem is the Łoś-Tarski Theorem [48, 36]: a first-order sentence φ is *preserved under extensions* over all structures—i.e., $A \models \varphi$ and A is an induced substructure of B imply $B \models \varphi$—if and only if it is equivalent to an existential sentence. Over a class that $\mathcal{C} \subseteq$ Struct(σ), the statement that every first-order sentence preserved under extensions is equivalent to an existential sentence is called the *preservation under extensions* property. Similarly, the Lyndon Positivity Theorem applied to a unary predicate X [37] connects surjective homomorphisms that are *strong* in every predicate except X to sentences that are positive in X.

As most of classical model theory, preservation theorems typically rely on compactness. Recall that the *compactness theorem of first-order logic* states that if a set of sentences is inconsistent, then a finite subset of it is already inconsistent. This theorem is known to fail in the finite case. Consequently, preservation theorems generally do not relativise to classes of structures, and in particular to the class Fin(σ) of all finite structures [see the discussions in 43, Section 2 and 33, Section 3.4]. For example, Lyndon's Positivity Theorem fails even over the class of finite words [35]. Nevertheless, there are a few known instances of classes of finite structures where some preservation theorems hold [3, 44, 45, 28], and this type of question is still actively investigated [e.g. 8, 35, 12]. We refer to Table 4 for a brief recap of the most well-known preservation theorems and their behaviour over Fin(σ).

	Struct(σ)	Order	Fragment	Fin(σ)
Łós-Tarski	✓	$\subseteq_i$	EFO	✗ [47]
Tarski-Lyndon	✓	$\subseteq$	PFO	✗ [1]
H.P.T.	✓	$\leq_{\text{hom}}$	EPFO	✓ [44]

Table 4: Example of preservation theorems and their relativisation in the finite.

Given a class $C \subseteq$ Fin(σ) and a sentence $\varphi \in$ FO[σ], one can define a set $[\![\varphi]\!]_C \triangleq \{A \in C \mid A \models \varphi\}$ and a function $[\![\varphi]\!] : X \to \mathbb{S}$, where $\mathbb{S}$ is the Sierpiński space $\{\top, \bot\}$ with $\bot \leq \top$. Let us now fix a preorder on C, for instance extensions $\subseteq_i$; the following properties are equivalent for a sentence φ: (a) φ is preserved under $\subseteq_i$, (b) $[\![\varphi]\!]$ is monotone, (c) $[\![\varphi]\!]_C$ is upwards closed. As a consequence, if the class C is wqo for $\leq$, every sentence preserved under $\leq$ is the upwards closure of finitely many structures. It is an easy check that for every order appearing in Table 4, the upwards closure of finitely many points is definable in the expected fragment of first-order logic. Therefore, whenever C is wqo for $\leq\ \in \{\subseteq, \subseteq_i, \leq_{\text{hom}}\}$, the corresponding preservation theorem holds.

However, most of the classes considered in the literature are not well-quasi-ordered, and some enjoy preservation theorems. This is because preservation theorems arise from the interplay between the ordering and the expressiveness of first-order logic. As an example, the class of finite cycles enjoys preservation under extensions, although the set is an infinite antichain for $\subseteq_i$, because first-order logic cannot distinguish between large cycles (see Figure 9).

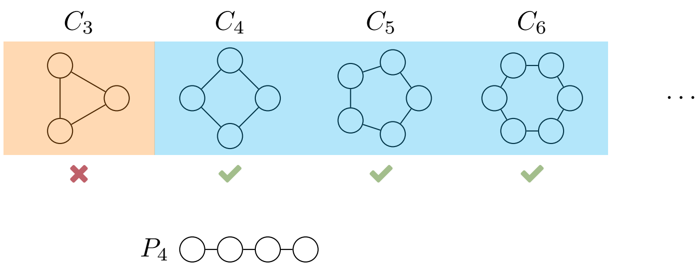


Figure 9: Evaluating $\forall x.\exists y.\neg(xEy) \land x \neq y$ over finite cycles.

Notice that the class of finite cycles is not *hereditary*, that is, not downwards closed with respect to induced substructures, but that its "minimal models" can be defined through equivalent existential sentences, and in this case can be taken to be finite paths.

Locality. As a way to tackle the lack of compactness theorem over $\mathrm{Fin}(\sigma)$, a more combinatorial description of the expressiveness of first-order logic is introduced. One of the main tools developed is the notion of "locality", with the broad meaning that first-order sentences cannot define "global" properties of structures, hence can be studied by their behaviour on local parts of said structures.

Given a structure A over a finite relational signature σ, its *Gaifman graph* has the elements of A as vertices and an edge (a, b) whenever both a and b are in relation in A. The distance $\mathrm{d}_A(a, b)$ between two elements of A is their distance in the Gaifman graph of A For a tuple $\vec{a} \in A$, and a radius $r \in \mathbb{N}$ one can consider the *r-neighbourhood* around $\vec{a}$ in A defined as $\mathcal{N}_A(\vec{a}, r) \triangleq \{a' \in A \mid \exists a \in \vec{a}, \mathrm{d}_A(a, a') \leq r\} = \bigcup_{a \in \vec{a}} \mathcal{N}_A(a, r)$; we emphasise that this union is not required to be disjoint. Slightly abusing notations, we identify the set $\mathcal{N}_A(\vec{a}, r) \subseteq_i A$ with the corresponding induced substructure of A. We call the structure $\mathcal{N}_A(\vec{a}, r)$ the *induced neighbourhood* of $\vec{a}$ in A.

A first-order formula $\varphi(\vec{x})$ is said to be *r-local* if its evaluation over a structure A and a tuple $\vec{a}$ from A only depends on the r-neighbourhood of $\vec{a}$ in A, i.e. $A, \vec{a} \models \varphi$ if and only if $\mathcal{N}_A(\vec{a}, r), \vec{a} \models \varphi$. In the particular case of $r = 0$, local formulas are equivalent to quantifier free formulas. Since σ is a finite relational signature, every formula φ can be *relativised* to a r-local formula, which coincides with φ over neighbourhoods of size r, although it might have a higher quantifier rank.

It is more delicate to craft a notion of locality for first order sentences since they have no free variables. The classical approach is to consider *basic local sentences* of the form $\exists \vec{x}. \bigwedge_{i \neq j} \mathrm{d}(x_i, x_j) > 2r \wedge \bigwedge_i \psi_{\leq r}(x_i)$ where $\psi_{\leq r}$ is a r-local formula with a single free variable. Simply put, the evaluation of a local r-basic sentence is determined solely by the evaluation of $\psi_{\leq r}$ over *disjoint* neighbourhoods of radius r. Note that the predicates $\mathrm{d}(x, y) \leq r$ and $y \in \mathcal{N}(\vec{x}, r)$ are definable for each fixed $r \in \mathbb{N}$.

The Gaifman Locality Theorem [19] states that every first-order sentence is equivalent to a Boolean combination of basic local sentences, called its *Gaifman Normal Form*. This can be thought of as $\mathrm{FO}[\sigma]$ being limited to describing the *local behaviour* of structures. In the study of preservation theorems such as preservation under extensions over finite structures, a first step is often to use Gaifman's normal form and rely on the structural properties of such as the "sparsity of the models" as a substitute for compactness [2, 3, 28, 12].

Name	Syntactic Form	Interpretation
Existential sentence	$\exists \vec{x}. \psi_{\mathrm{qf}}(\vec{x})$	Induced substructure
Existential local sentence	$\exists \vec{x}. \psi_{\mathrm{loc}}(\vec{x})$	Induced neighbourhood
Basic local sentence	$\exists \vec{x}. \bigwedge_{i \neq j} \mathrm{d}(x_i, x_j) > 2r \wedge \bigwedge_{i=1}^{n} \psi_{\mathrm{loc}}(x_i)$	n disjoint copies of a one point induced neighbourhood

Table 5: Comparison of the different syntactic local forms. The notation ψ_{qf} denotes a quantifer-free formula, and ψ_{loc} denotes a local formula.

SCIENTIFIC PRODUCTION

II

The goal of this thesis is to bridge these domains through the core notion of "compactness", be it as a logical property, a topological notion, or a combinatorial tool. One of the main problem used as a practical application is the study of so-called "preservation theorems" in first-order logic, introduced in Section 1.3. Those arise from classical model theory, but are highly non-trivial (if not false) in finite model theory, and can be elegantly rephrased as generalisations of well-quasi-orders or Noetherian spaces.

However, we do not restrict our attention to preservation theorems, and also investigate both well-quasi-orders and Noetherian spaces in their own right.

1 COMPOSITIONS ORDERINGS

We tried to study the result by Daligault, Rao, and Thomassé [11] on well-quasi-ordered classes of bounded clique width using automata theory. We rephrased the characterisation from [11] as a requirement on the monoid M: M generates classes of graphs well-quasi-ordered with respect to induced substructure if and only if M is a chain of simple semigroups. Moreover, we encoded the composition preorder of chains of simple semigroups into the gap embedding and vice-versa, proving that they are "equally hard".

As a follow-up question, we tried to measure the loss of expressiveness induced by the restriction to chains of simple semigroups. For that, we compared classes generated by chains of simple semigroups to classes that are m-partite. Classes of m-partite graphs are classes of bounded clique width that are using a trivial monoid $M = \{id\}$, hence are wqo. We exhibited a family $H_{k,n}$ coming from a very simple monoid that leads to a well-quasi-ordered class that is not m-partite for any finite m, see Figure 10. This result was strengthened by proving that $H_{k,n}$ can be built from a word rather than a tree, hence is of linear bounded clique width.

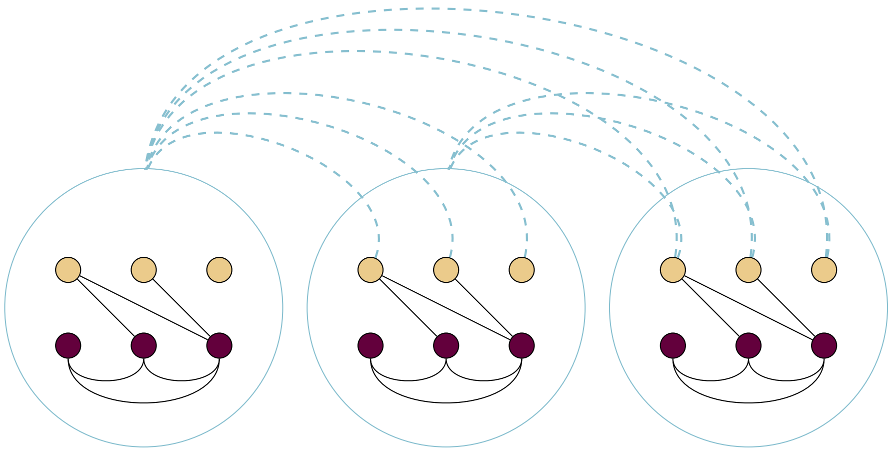


Figure 10: The graph $H_{3,3}$.

One key property in the characterisation of Daligault, Rao, and Thomassé [11] is that one asks for the monoid M to generate only well-quasi-ordered classes, regardless of the way the elements of the monoid are used to build the edges afterwards. We tried to extend the characterisations to pairs (M, P), and classify those pairs such that the "language embedding" is wqo on M^*, that is a word $u = a_1 \dots a_n$ is below a word v if and only if $v = v_1 \dots v_n$ and $\mu(a_i \dots a_j) \in P \Leftrightarrow \mu(v_i \dots v_j) \in P$. This "language embedding" is precisely capturing the induced substructure ordering over the interpreted graphs, hence we could have a more fined grained (and hopefully decidable) characterisation of classes that are of linear bounded clique width and wqo. However, none of our proof scheme worked, and every conjecture was countered by explorations of monoids via Minisat.

This direction of research has been paused.

2 TOPOLOGICAL PRESERVATION THEOREMS

In Section 1.3, we saw that preservation theorems over a downwards closed set of finite structures can be rewritten as the fact that sentences preserved under extensions have finitely many minimal models. We use our Cheatsheet in Table 2 to translate it in a topological setting:

1. A sentence preserved under extensions is a sentence defining an upwards closed set, which is translated to an open set.
2. A sentence having finitely many minimal models is translated to a sentence defining a compact set.

Recall that a set E is *definable* in X whenever there exists a first-order sentence φ such that $E = [\![\varphi]\!]_X$. Moreover, in a topological space (X, τ), we write $\mathcal{K}^\circ(X)$ to denote the compact and open sets of X. As a consequence, preservation under extensions can be rephrased as $\tau_{\subseteq_i} \cap [\![\mathsf{FO}]\!]_X \subseteq \mathcal{K}^\circ(X)$.

However, this condition does not take into account a peculiar property of existential sentences: they define a *basis* of the topology $\tau_{\subseteq_i}$, i.e. every open set in $\tau_{\subseteq_i}$ is a union of sets defined by existential sentences. Moreover, existential sentences are compact, hence we have $\tau_{\subseteq_i} = \langle \mathcal{K}^\circ(X) \rangle$. This is the motivation behind the definition of logically presented pre-spectral spaces.

Definition 2.1 (Logically presented pre-spectral space)**.** Let $X \subseteq \mathrm{Fin}(\sigma)$ be a **set** of structures and τ be a topology over X. The triple $\langle X, \tau, \mathsf{FO}[\sigma] \rangle$ is a *lpps* if $\tau = \langle \mathcal{K}^\circ(X) \rangle$ and $\tau \cap [\![\mathsf{FO}]\!]_X \subseteq \mathcal{K}^\circ(X)$.

In a lpps, we automatically have the equality $\tau \cap [\![\mathsf{FO}]\!]_X = \mathcal{K}^\circ(X)$, hence $\mathcal{K}^\circ(X)$ is a bounded sublattice of $\wp(X)$. Moreover, the underlying topological space (X, τ) is pre-spectral [see. 14], explaining part of the notion's name. This connection with spectral spaces allows leveraging structural properties of the category of spectral spaces, such as the existence of products or projective limits [14, Section 2.2 and 2.3].

We generalise the notion of lpps to replace $[\![\mathsf{FO}]\!]_X$ by a bounded sublattice of $\wp(X)$. This allows us to express in full generality the fundamental property of lpps.

Theorem 2.2 (Fundamental Property)**.** *Let τ be a topology on X, $\mathcal{L}$ a bounded sublattice of $\wp(X)$, and $\mathcal{L}'$ a sublattice of $\mathcal{L}$. The following are equivalent:*

1. $\mathcal{L} \cap \tau \subseteq \mathcal{L}' \subseteq \mathcal{K}^\circ(X)$.
2. $\langle X, \tau, \mathcal{L} \rangle$ *is a lpps and* $\mathcal{L}'$ *is a basis of* τ.

Let us immediately translate this result in terms of preservation under extensions. Given a hereditary set of structures X with the Alexandroff topology from the extension ordering $\subseteq_i$, Theorem 2.2 states that it is equivalent that (a) FO-definable upwards-closed sets for $\subseteq_i$ are both EFO-definable and have finitely many $\subseteq_i$-minimal models, (b) FO-definable upwards-closed sets have finitely many $\subseteq_i$-minimal models, upwards-closed sets with finitely many $\subseteq_i$-minimal models generate all upwards-closed sets, and EFO-definable sets generate all upwards-closed sets. In this specific case, this amounts to the folklore result stating that preservation under extensions holds if and only if upwards-closed FO-definable subsets have finitely many $\subseteq_i$-minimal models.

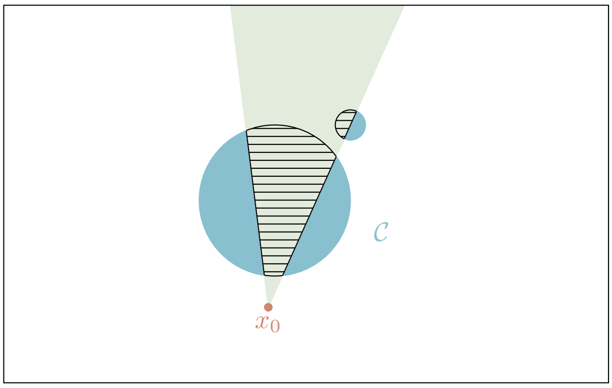
Note that for preservation theorems in the literature, one always has that the fragment of interest is a basis of the topology, but for some classes of structures, this fragment may not always define compact sets, moreover, compact sets might not be those with "finitely many minimal models" when the ambiant space is not downwards closed. Let us reuse our example of the class of finite cycles and preservation under extensions as an illustration of the power of our topological notions.

Example 2.3 (The case of cycles)**.** In the case of Cycles the class of finite cycles, it is an easy check that $\tau_{\subseteq_i}$ is the discrete topology, and as a consequence $\mathcal{K}^\circ(\mathsf{Cycles})$ consists solely of the finite sets of finite cycles. In particular, there are some existential sentences having infinitely many models, hence that are not compact.

However, τ_C defined as the cofinite topology over Cycles has the same specialisation preorder as $\tau_{\subseteq_i}$. Moreover, we do have that $\langle \mathsf{Cycles}, \tau_C, \mathsf{FO} \rangle$ is a lpps because it is Noetherian. We conclude that $[\![\mathsf{FO}]\!]_{\mathsf{Cycles}} \cap \tau_C \subseteq [\![\mathsf{EFO}]\!]_{\mathsf{Cycles}} \cap \tau_C \subseteq \mathcal{K}^\circ(\mathsf{Cycles})$. As a sentence $\varphi \in \mathsf{FO}$ defines either a *finite set* or a *cofinite set* of Cycles, in both cases, it is expressible in EFO.

We argue that lpps captures the "well-behaved" preservation theorems, i.e. those obtained by topological arguments, and demonstrate this by exhibiting stability properties of those spaces. Using properties of pre-spectral spaces, we prove that if $\langle X, \tau, \mathsf{FO}[\sigma] \rangle$ is lpps and Y be a Boolean combination of compact-open subsets of X, then $\langle Y, \theta, \mathsf{FO}[\sigma] \rangle$ is a lpps, where θ be the topology induced by τ on Y. This behaviour distinguishes the property of "being lpps" from "having a preservation theorem", because we capture explicitly the notion of compactness, as demonstrated in Figure 11.

Moreover, mimicking the grammars for wqos Figure 4 and Noetherian spaces Figure 8, we built a grammar for building lpps as described in Figure 12.



Struct(σ)

Figure 11: The upwards closure $\uparrow x_0 \cap \mathcal{C}$ may have infinitely many minimal models in $\mathcal{C}$

$D ::= (X, \leq)$	wqo
$\mid (X, \tau)$	Noetherian
$\mid D_1 \cdot D_2$	inner product
$\mid \Sigma_{i=1}^{n} D_i$	finite disjoint sums
$\mid \Pi_{i=1}^{n} D_i$	finite products
$\mid D^*$	finite words, regular subword topology
$\mid T(D)$	finite trees, regular subtree topology
$\mid D_1 \rtimes D_2$	wreath product, D_1 ∞-Noetherian

Figure 12: Grammar for building lpps.

Specific Closure Properties. In addition to these algebraic constructions, we studied the logical closure of a set, which only makes sense in the case of sets of structures. If $X \subseteq Z \subseteq$ Struct(σ) then one can build the *logical closure* of X in Z as the closure of X in the topology generated by the definable subsets of Z. In practise, this means that we "complete" X to add structures in Z indistinguishable from the point of view of the logic. We proved the unsurprising theorem that if a set Y is stuck between X and the logical closure of X in some ambiant space Z, then $\langle X, \tau, \mathsf{FO} \rangle$ is a lpps if and only if $\langle Y, \tau, \mathsf{FO} \rangle$ is, and definable bases of X are definable bases of Y. This can be used to prove that countable disjoint unions of finite structures enjoy preservation under extensions.

Finally, we explained Rossman's proof of the H.P.T. in the finite through a lemma allowing to "transfer" projectives limits of pre-spectral spaces to lpps.

3 POSITIVE NORMAL FORMS

Gaifman Normal Forms. In our approach to study preservation under extensions, we developed an interpolant between arbitrary first-order sentences and existential sentences in the form of *existential local sentences*: these are sentences of the form $\exists \vec{x}.\psi(\vec{x})$ where ψ is a r-local formula around $\vec{x}$.

Our first contribution is to prove that existential local sentences are exactly those that can be put in Gaifman Normal Form without using negations, hereafter called *positive Gaifman normal form*. This result relies on careful combinatorial description of the neighbourhoods in a model and therefore relativises to arbitrary classes $\mathcal{C}$ of structures.

Theorem 2.4 (Positive Locality). *Let $\mathcal{C} \subseteq \mathrm{Struct}(\sigma)$ be a class of structures and $\varphi \in \mathrm{FO}[\sigma]$ be a first-order sentence. The sentence φ is equivalent over $\mathcal{C}$ to an existential local sentence if and only if it is equivalent over $\mathcal{C}$ to a positive Gaifman normal form.*

As every existential sentence is existential local, every proof of preservation under extensions can be factored as going from sentences preserved under extensions to a positive Gaifman normal forms and then to existential sentences. This puts the light on technical parts common to proofs of preservation under extensions [3, 2]. Given our proof factorisation of preservation under extensions, we turned our attention to proving that sentences preserved under extensions can always be rewritten in positive Gaifman normal form. Proving such a result would reduce preservation under extensions to a syntactic transformation from basic local sentences to existential sentences, and usual sparsity or model-theoretic properties of the considered classes can be used to tackle this last part.

Positive Preservation. Following this program, we characterised the semantic property corresponding to positive Gaifman normal forms, through a finer notion of embeddings than induced substructures, namely *local elementary embeddings*, that is, embeddings $f : A \to B$ such that for all local formula $\psi(\vec{x})$, for all $\vec{a} \in A$, $A, \vec{a} \models \psi(\vec{x})$ if and only if $B, f(\vec{a}) \models \psi(\vec{x})$. By restricting the quantifier rank q of sentences ψ, their locality radii r, and the number of free variables k, this provides *(r, q, k)-local elementary embeddings*.

Theorem 2.5 (Local preservation). *Let φ be a sentence in $\mathrm{FO}[\sigma]$. The following properties are equivalent over the class of all structures $\mathrm{Struct}(\sigma)$.*

(a) The sentence φ is equivalent to an existential local sentence.
(b) There exist $r, q, k \in \mathbb{N}$ such that φ is preserved under (r, q, k)-local elementary embeddings.
(c) The sentence φ is preserved under local elementary embeddings.

Unfortunately, Theorem 2.5 follows the usual proof scheme of classical model theory, relying on the compactness theorem of first-order logic. Early on, we noticed that *preservation under local elementary embeddings* —sentences preserved under local elementary embeddings are expressible as existential local sentences— boils down to preservation under disjoint unions —sentences preserved under disjoint unions are equivalent to existential local sentences— in the finite, where φ is *preserved under disjoint unions* over $\mathcal{C}$ whenever for all $A, B \in \mathcal{C}$, $A \models \varphi$ implies $A \uplus B \models \varphi$.

After trying combinatorial methods having similar flavour to Theorem 2.4, we found a (quite-involved) counter-example to Theorem 2.5 in the finite. This counter-example is built on the idea of Tait [47] in the case of the preservation under extensions and relativisation properties of lpps. We use as relational signature $\sigma \triangleq \{(\leq, 2), (S, 2), (E, 2)\}$, and define structures $O_m + \cdots + O_n$ with $2 \leq m \leq n$ as described in Figure 13 for $n = 5$ and $m = 2$. We define then $\mathcal{C}_{bad}$ to be the class of finite disjoint unions of such structures structures. It is easy to see that Theorem 2.5 does not relativise to $\mathcal{C}_{bad}$, and using stability properties of lpps this lifts to the whole class $\mathrm{Fin}(\sigma)$ of finite structures.

Theorem 2.6 (Counter example over $\mathrm{Fin}(\sigma)$). *There exists a sentence φ preserved under disjoint unions over $\mathrm{Fin}(\sigma)$ but not equivalent to an existential local sentence over $\mathrm{Fin}(\sigma)$.*

After building this "grid-like" counter-example, it was now possible to prove the undecidability of a number of natural questions on existential local sentences and preservation under local elementary embeddings.

Theorem 2.7 (Undecidability). *It is not possible, given a sentence φ preserved under disjoint unions over $\mathrm{Fin}(\sigma)$, to decide whether it has it is equivalent to an existential local sentence.*

Neither is it possible, given a sentence φ, to decide whether or not it is preserved under disjoint unions over $\mathrm{Fin}(\sigma)$.

Theorem 2.8 (Uncomputable equivalence). *There is no algorithm that given a sentence φ that is equivalent to an existential local sentence over $\mathrm{Fin}(\sigma)$ computes such a sentence.*

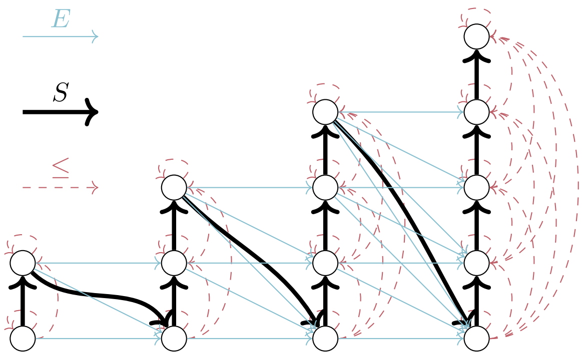


Figure 13: The structure $O_2 + O_3 + O_4 + O_5$

Cartography of Parameters. Because we can stratify local elementary embeddings using parameters (r, q, k), we asked ourselves whether these stronger preservation properties imply our existential local syntactic form. Though a systematic study, we mapped in Figure 14 regions of the "preorder-cube" implying positive Gaifman normal form over $\mathcal{C}$. One particularly interesting case is $k = \infty$ and $r, q = 0$, where the preorder coincides with induced substructures, whenever the class is hereditary and *closed under disjoint unions*, that is, such that $A \uplus B \in \mathcal{C}$ if $A, B \in \mathcal{C}$. For this particular case, we actually use monadic second order logic to deal with the possible intersections of neighbourhoods, and believe that this proof scheme has an interest of its own.

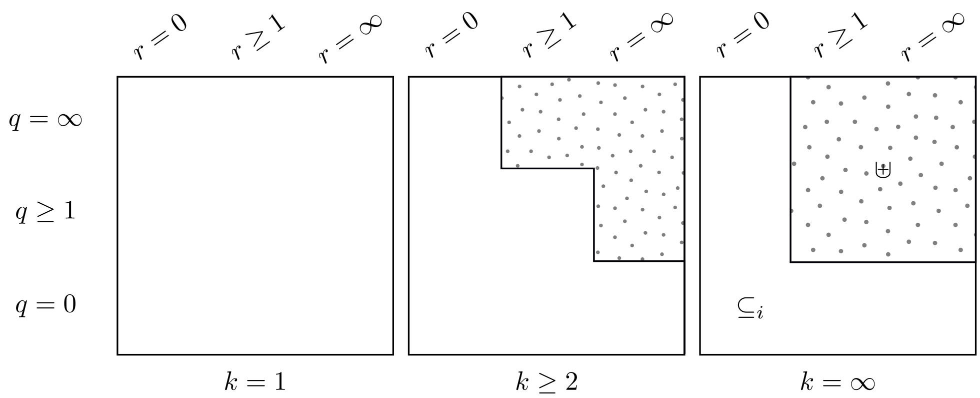


Figure 14: Parameters (r, q, k) leading to an existential local normal form (white), those with a counter example (dots) over hereditary classes closed under disjoint unions.

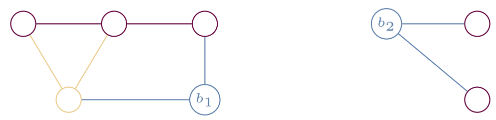
4 LOCALISABLE CLASSES OF STRUCTURES

We proved in Figure 14 that sentences preserved under extensions have a positive Gaifman normal form on classes of structures that are hereditary and closed under disjoint unions, which was the first step of our "preservation through locality" program. The second step ouf our program is to leverage this positive Gaifman normal forms to build preservation under extensions theorems. To that end, we place ourselves in the situation of *localisable* classes of structures, that is, classes $\mathcal{C}$ where every sentence φ preserved under extensions is equivalent over $\mathcal{C}$ to some existential local sentence ψ.

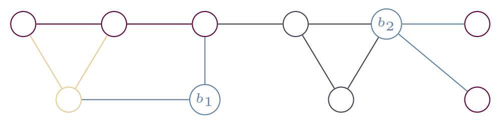
Let us restate explicitly that from Figure 14, we know that if $\mathcal{C} \subseteq \mathrm{Fin}(\sigma)$ is hereditary and closed under disjoint unions then it is localisable, in particular, the class $\mathrm{Fin}(\sigma)$ is localisable.

Surprisingly, given the general non relativisation properties of preservation theorems, we were able to relate preservation under extensions over $\mathcal{C}$ and preservation under extensions for a stratification of $\mathcal{C}$ using neighbourhoods. More precisely, given a class $\mathcal{C}$, we defined $\mathrm{Balls}(\mathcal{C}, r, k)$ as the set of all $\mathcal{N}_A(\vec{a}, r)$ for $A \in \mathcal{C}$ and $\vec{a} \in A^k$. This is a stratification because the increasing union $\bigcup_{r \geq 0} \bigcup_{k \geq 1} \mathrm{Balls}(\mathcal{C}, r, k)$ equals $\mathcal{C}$ whenever $\mathcal{C}$ is hereditary. One key argument in our proof is that we can extend an induced substructure $A \subseteq_i B \in \mathrm{Balls}(\mathcal{C}, r, k)$ with at most $r \times k \times |A|$ points to ensure that $A \subseteq_i A' \subseteq_i B$ with $A' \in \mathrm{Balls}(\mathcal{C}, r, k)$, see Figure 15.

The induced substructure $A \subseteq_i B$.



The extended substructure $A' \in \mathrm{Balls}(\mathcal{C}, 2, 2)$ is centered around (b_1, b_2).



The whole structure B is centered around (b_1, b_2).

Figure 15: Extracting a structure in $\mathrm{Balls}(\mathcal{C}, 2, 2)$ from an induced substructure A of a larger structure B.

Theorem 2.9 (Local preservation under extensions)**.** *Let $\mathcal{D}$ be a hereditary localisable class of finite structures. Preservation under extensions holds over $\mathcal{D}$ if and only if preservation under extensions holds over* $\mathrm{Balls}(\mathcal{D}, r, k)$ *for all $r \geq 0$ and $k \geq 1$.*

Corollary 2.10 (Local preservation under extensions)**.** *Let $\mathcal{D}$ be a hereditary class of finite structures closed under disjoint unions. Preservation under extensions holds over $\mathcal{D}$ if and only if preservation under extensions holds over* $\mathrm{Balls}(\mathcal{D}, r, k)$ *for all $r \geq 0$ and $k \geq 1$.*

This completes our two-step approach to preservation under extensions, and allows us to strictly generalise previously known classes of structures enjoying preservation under extensions. As an example, one of the best known results was that hereditary classes of finite structures closed under disjoint unions and wide enjoy preservation under extensions [2, Theorem 4.3]. This is captured in our program as hereditary wide classes $\mathcal{C}$ are exactly those where $\mathrm{Balls}(\mathcal{C}, r, k)$ is *finite* for every $r \geq 0, k \geq 1$, aka *locally finite classes*. As finite classes enjoy preservation under extensions, Corollary 2.10 implies that locally finite classes do, and thus recovers Atserias, Dawar, and Grohe's result [2, Theorem 4.3].

By *localising* properties known to imply preservation under extensions, we harvest from Theorem 2.9 new structural properties implying preservation under extensions. Examples of such properties, along with their respective implications, are given in Figure 16.

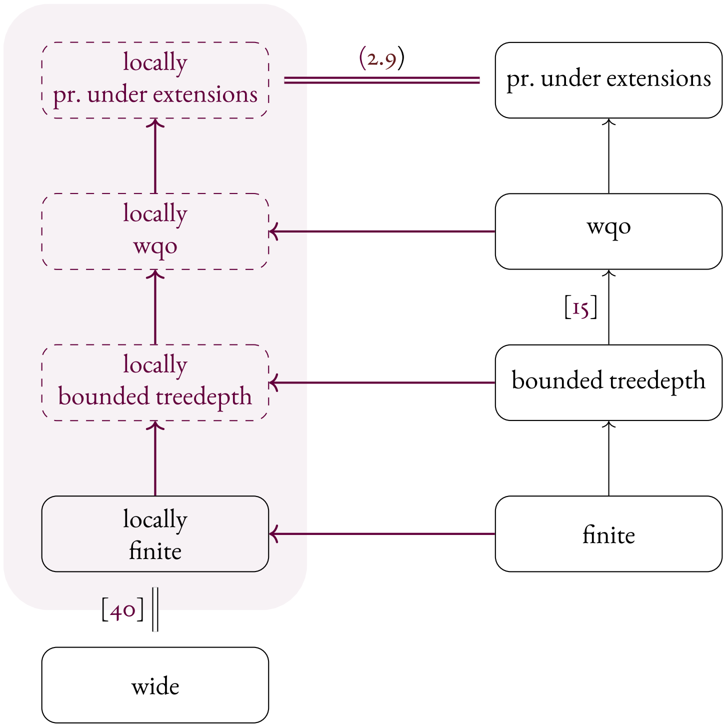


Figure 16: Implications of properties for localisable hereditary classes of finite structures. Arrows are strict implications, thick arrows are new results. Boxes are properties over classes of finite structures, and dashed ones are new.

Following the idea that wqos can be generated on inductive datatypes by adding the transitive closure of a notion of substructure ordering, we study how one can iteratively build Noetherian topologies on a given set. However, as opposed to the constructions of the divisibility topology, we will not change the underlying space during our construction. This can be understood as keeping the same underlying space and refining the resolution of the space, rather than building bigger and bigger (high resolution) parts of the space.

Iterative Topology Refinement. Given a set X, one can consider functions F from topologies over X to topologies over X. If F is monotone and preserves Noetherian topologies, we call F a *refinement function*. Naturally, one can build the *least fixed point* of a refinement function, and it is obtained by iterating (transfinitely) the function F starting from the trivial topology on X. Recall that given a limit ordinal α, the *limit topology* $F^{\alpha}(\tau)$ is defined as the join over all topologies $F^{\beta}(\tau)$ for $\beta < \alpha$.

Example 2.11. For instance, one can consider $X = \Sigma^*$ where (Σ, θ) is a finite topological space, and define $F(\tau)$ as generated by the set $\{\{av \mid a \in U, v \in V\} \mid U \in \theta, V \in \tau\}$. It is an easy check that F is a refinement function, but its least fixed point is the discrete topology over Σ^* whenever θ is the discrete topology over Σ (see Figure 17).

As the discrete topology over Σ^* is not Noetherian, this proves that one cannot, in general, iterate refinement functions.

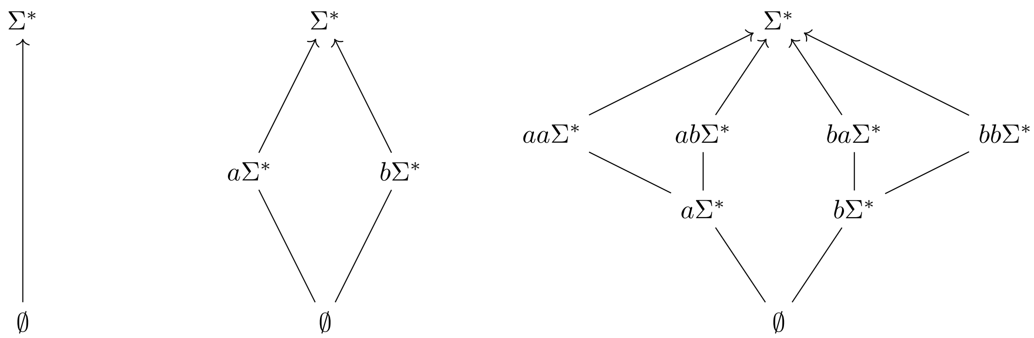


Figure 17: Iterating F over Σ^*. On the left the trivial topology, followed by F, and on the right F^2.

However, we discovered that a simple requirement on the refinement function F allows to use a minimal bad sequence argument (adapted to topological spaces). This requirement is roughly that the refinement function "respects subsets". Let (X, τ) be a topological space and H be a subset of X. We call the *subset restriction* $\tau \mid H$ the topology generated by the opens $U \cap H$ where U ranges in τ. See Figure 18 for a visual understanding of this construction.

Example 2.12. It might be useful to consider what happens in the case of an Alexandroff topology to interpret this notion on quasi-orders. Let us write $H \subseteq X, \tau = \tau_{\leq}$ over X and compute $\mathrm{spec}(\tau \mid H)$.

$$
\begin{aligned}
x \ \mathrm{spec}(\tau \mid H)\ y &\iff \forall U \in \tau \mid H, x \in U \implies y \in U \\
&\iff \forall U \in \tau, x \in U \cap H \implies y \in U \cap H \\
&\iff \forall z \in X, x \in \uparrow z \cap H \implies y \in \uparrow z \cap H \\
&\iff \begin{cases} x \leq y \land x \in H \land y \in H \\ x \notin H \end{cases}
\end{aligned}
$$

We call *topology expanders* the refinement functions that respects subsets, i.e. such that for every Noetherian topology τ satisfying $\tau \subseteq F(\tau)$, for all closed set H in τ, $F(\tau) \mid H \subseteq F(\tau \mid H) \mid H$.

Example 2.13. In the realm of quasi-orders, this condition can be translated as follows: for all $x, y \in H$, $xF(\leq_H)y$ implies $xF(\leq)y$, where $u \leq_H v$ if $u \leq v$ and $u, v \in H$ or $y \notin H$. This is interpreted as the fact that to compute the refinement of the quasi-order over a closed subset, it suffices to refine over this subset, i.e. no outside information is needed by F.

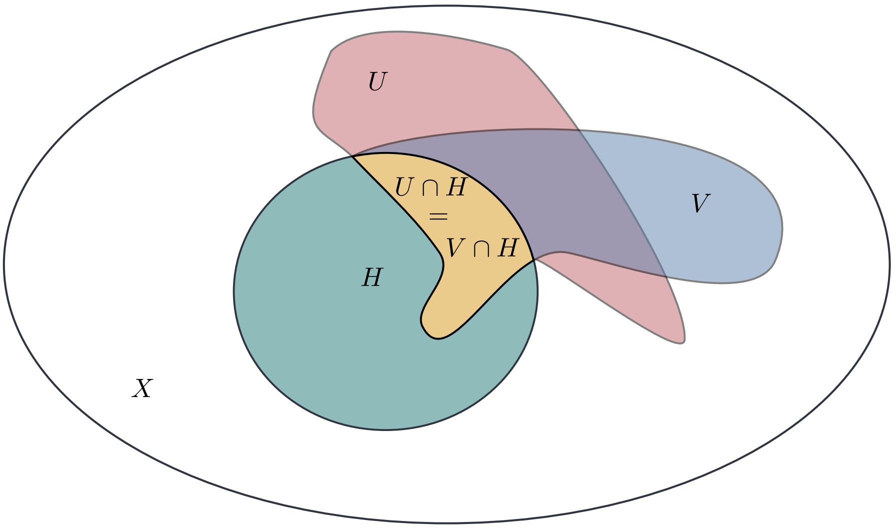


Figure 18: Subset Restriction. Whenever H is closed, $W = U \cup H^c$ is the largest open set satisfying $U \cap H = W \cap H$.

Theorem 2.14 (Limit topologies are Noetherian)**.** *Let α be a limit ordinal and F be a topology expander. If $F^{\beta}(\tau)$ is Noetherian for all $\beta < \alpha$ then $F^{\alpha}(\tau)$ is Noetherian.*

Corollary 2.15. *Let F be a topology expander. The least fixed point of F is a Noetherian topology.*

Note that this topological setting allows getting rid of category theory altogether and focus on the main minimal bad sequence argument. Moreover, one can define topology expanders over sets that are not usual inductive datatypes, such as infinite words. Let us write $X^{<\omega}$ for the words of size at most ω, and define the following topology expander, given a topology θ on X: $F(\tau)$ is generated by the open sets $\{wav \mid u \in \Sigma^*, a \in U, v \in V\}$ and $\{w \mid \forall i, \exists j > i, w_i \in U\}$, for $U \in \theta$ and $V \in \tau$. It turns out that F is a topology expander and that its least fixed point is the regular subword topology over infinite words, as shown by Goubault-Larrecq and Lopez [26].

Conjecture 2.16. One can obtain the same theorem by redefining $\tau \mid H$ via $\{U \cup H^c \mid U \in \tau\}$, which has a nicer interpretation in terms of quasi-orders: $u \leq\mid Hv$ if and only if $u \leq v$ or $v \notin H$.

Divisibility Topologies. We are now working in a categorical setting. Let us write Set for the category of sets with functions, Top for the category of topological spaces with continuous maps. If Corollary 2.15 can be applied to sets that are not defined inductively, it can also serve as a way to build *canonical* topologies over inductive datatypes. For that, we will consider "type constructors" as analytic endofunctors of Set, that are lifted to Top to put topologies on the generated constructions.

Example 2.17. Let G' be the functor sending a topological space (X, τ) to X^* with the regular subword topology. Then G' is the lift of G mapping X to X^* in Set to Top. Moreover, G is analytic.

Let G^T : Top $\to$ Top be a lifting of an analytic functor G, and $(\mu G, \delta)$ an initial algebra of G. There exists a notion of substructure ordering over μG, written $\sqsubseteq$. For instance, if one defines finite words over Σ as the least fixed point of $G(X) \triangleq 1 + \Sigma \times X$, the substructure ordering is the usual suffix ordering, described in Figure 19.

We define the *divisibility topology* on μG as the least fixed point of $F_{\Diamond}$ that maps a topology τ to the topology generated by $\{\uparrow_{\sqsubseteq} \delta(U) \mid U \text{ open in } G^T(\mu G, \tau)\}$. If we suppose that G^T preserves embeddings, then $F_{\Diamond}$ is a topology expander on the initial algebra μG and the divisibility topology is Noetherian thanks to Corollary 2.15. We provide in the case of finite words, the computation of $F_{\Diamond}(\{\emptyset, \Sigma^*\})$ in Figure 20; notice how this topology differs from Figure 17.

As a sanity check, we prove that the divisibility topology is the Alexandroff topology of the divisibility preorder from Hasegawa, Freund. Moreover, we can prove that the regular subword topology and the regular subtree topology are the divisibility topologies of their inductive constructors.

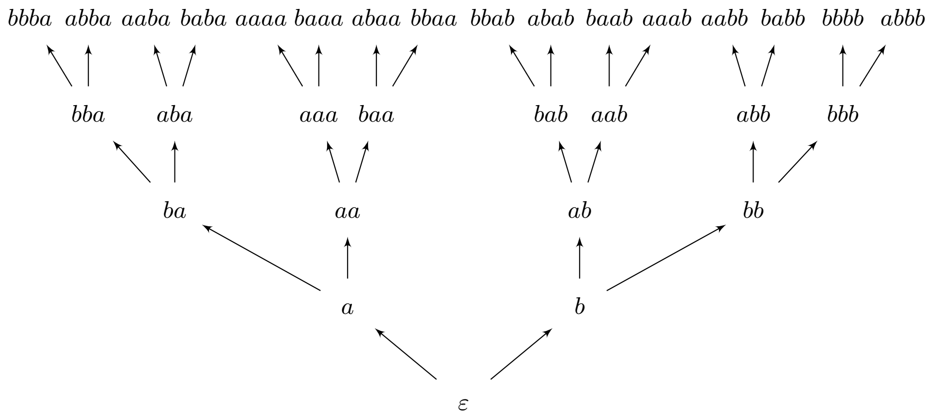


Figure 19: Substructure ordering $\sqsubseteq$ on words of length at most 4.

Theorem 2.18. *Let G^T be the lift of an analytic functor respecting Alexandroff topologies, Noetherian spaces, and embeddings. The divisibility topology of μG is the Alexandroff topology of the divisibility preorder of μG, which is a well-quasi-ordering.*

$$
\begin{tikzcd}
\top \\
\\
\bot \arrow[uu]
\end{tikzcd}
\qquad
\begin{tikzcd}
 & \top & \\
\{a\} \times \top \arrow[ur] & \{\varepsilon\} \arrow[u] & \{b\} \times \top \arrow[ul] \\
 & \bot \arrow[ul] \arrow[u] \arrow[ur] &
\end{tikzcd}
\qquad
\begin{tikzcd}
 & \uparrow_{\sqsubseteq}\{\varepsilon\} = \top & \\
\uparrow_{\sqsubseteq}\{a\} \times \top \arrow[ur] & & \uparrow_{\sqsubseteq}\{b\} \times \top \arrow[ul] \\
 & \bot \arrow[ul] \arrow[ur] &
\end{tikzcd}
$$

$$
\begin{tikzcd}
 & \top & \\
\{a, aa, ba, aaa, aba, baa, bba\} \arrow[ur] & & \{b, bb, ab, bbb, bab, abb, aab\} \arrow[ul] \\
 & \bot \arrow[ul] \arrow[ur] &
\end{tikzcd}
$$

Figure 20: One iteration of $F_{\Diamond}$ over the set of words of length at most 3.

6 TRANSFINITE WORDS

The Regular Subword Topology. We extend the Nœtherian topology introduced by Goubault-Larrecq [24] to transfinite words in Goubault-Larrecq, Halfon, and Lopez [25] by defining the regular subword topology over transfinite words. The *regular subword topology* on transfinite words is defined by the means of its closed sets as in Figure 21.

$$\frac{C \text{ closed in } \theta \qquad \beta \leq \alpha}{C^{<\beta} \text{ closed}} \qquad\qquad \frac{P_1 \text{ closed} \qquad \cdots \qquad P_n \text{ closed}}{P_1 \cdots P_n \text{ closed}}$$

Figure 21: Closed sets of the regular subword topology over $X^{<\alpha}$.

After proving that this topology is Noetherian, we compute concretely the specialisation preorder of the regular subword topology by first characterising the closures of words. A word w is said to be *topologically indecomposable* when every factorisation $w = uv$ with $v \neq \varepsilon$ satisfies $\overline{w} = \overline{v}$. Because $X^{<\alpha}$ is Noetherian in its regular subword topology, we know that every word is a finite product of topologically indecomposable words.

Our next step is proving that $\overline{w_1 \ldots w_n} = \overline{w_1} \ldots \overline{w_n}$ when the words w_i are indecomposable. Moreover, $\overline{w} = (\overline{\text{letters}(w)})^{<|w|+1}$ when w is topologically indecomposable. As a consequence, we characterised the closure of words.

This can be used to prove that the specialisation preorder of the regular subword topology is not the embeddings of transfinite words, by considering for instance $w \triangleq (a^\omega b)^\omega$, whose closure is $\{a, b\}^{<\omega^2+1}$ which is definitely not equal to $\downarrow w$. As this topology fails to have the correct specialisation preorder, even when assuming that the specialisation preorder of X is better-quasi-ordered, we try to extend the range of closed sets.

Recurrent Subword Topology. It is possible to strengthen the topology by allowing the rule $P^{<\beta}$ in our inductive definition of closed sets. This provides a topology expander over transfinite words, and therefore a Noetherian topology called the *recurrent subword topology*, strictly finer than the regular subword topology. For the exact same reasons as for the regular subword topology, the specialisation preorder cannot be the word embedding in general, but we conjecture that this holds when we start from a better-quasi-order.

FUTURE WORK

III

1 TOPOLOGICAL INTERPRETATIONS

We have yet to produce a proper interpretation of our localisation theorems in terms of topological spaces. For instance, the fact that a class $\mathcal{C}$ is localisable cannot be interpreted topologically –assuming that $\mathcal{C}$ is a **set**– as the fact that $\langle \mathcal{C}, \tau_{\text{basic}}, \mathsf{FO} \rangle$ is a logically presented pre-spectral space, where τ_{basic} is the topology generated using basic local sentences. Indeed, basic local sentences may not define compact sets in τ_{basic}.

2 INDUCTIVE CONSTRUCTIONS OF PRESERVATION THEOREMS

We hope that the technology of minimal bad sequences in the topological setting better adapts to lpps and that the proof schemes of divisibility topologies allows building automatic preservation theorems over classes of structures, given an algebra over the class.

3 CHARACTERISING THE TREE EXPRESSIONS TOPOLOGY.

We have yet to compute concretely the specialisation preorder of the recurrent subword topology. We hope that we can reuse our proof techniques developed in the case of the regular subword topology, in particular the notion of topologically indecomposable words.

Conjecture 3.1. If the specialisaton preorder on X is BQO then we get back on $X^{<\alpha}$ the embedding preorder as a specialisation preorder.

4 FOLLOW UP ON INDUCTIVE CONSTRUCTIONS

We extended the divisibility preorders to divisibility topologies, but some hypothesis of the former are not needed in the latter, namely, that the underlying Set-functor is analytic. Indeed, it is crucially used by both Hasegawa, Freund in their minimal bad sequence argument to guarantee that the support of an element is finite. In Noetherian spaces, we do not need such a restriction, as the usual powerset construction preserves Noetherian spaces. We can therefore study the notion of "quasi-analytic functor", where the finiteness of the supports is dropped.

We also hope that we can adapt our proof scheme to arbitrary algebras of a given functor, to justify the canonicity of the regular subword topology or the recurrent subword topology as emerging from reasonable algebras over infinite words.

We have yet to design an approach to compute the "ordinal invariants" of the topologies generated by topology expanders, that could give us automatic upper bounds on the constructions.

5 SPARSITY, MODEL THEORY, AND COMPACTNESS

There has been recent development in the theory of sparsity [39, 20, 5, 4, 16, 46, 27, 21] that is getting closer to describe combinatorially classes of structures for which model checking is fixed-parameter tractable.

The techniques used may have applications to the study of preservation theorems, particularly the characterisation of first-order transductions of sparse classes [39, 20]. Sylvain Schmitz proposed the following conjecture, following the intuitions from these results.

Conjecture 3.2. A class $\mathcal{C}$ enjoys preservation under extensions if and only if one cannot *existentially* transduce arbitrarily large cycles from $\mathcal{C}$.

Moreover, the study of sparsity has been boosted by incorporating elements from finite model theory such as stable classes and dependent classes, for which suitable regularity lemmas have been proven [see for example 10, 9]. We informally conjecture that stable classes may be localisable in some sense, because one cannot build large grid-like structures as in Theorem 2.7.

6 WELL QUASI-ORDERS DEFINED VIA SEMIGROUPS

Recent development in the semigroup community [32] allows characterising ordered semigroups for which the composition preorder is wqo, leading to a proper algorithm to detect such semigroups. This reignites the hope of characterising *regular languages* for which the composition preorder is wqo.

7 Ordinal invariants of well-quasi-orders

In collaboration with Simon Halfon, Philippe Schnoebelen, and Isa Vialard, I started investigating the behaviour of the finite powerset construction in wqos through its "ordinal invariants". This is connected to the study of divisibility preorders as they provide families attaining our lower bounds.

8 Query counting

In collaboration with Thomas Colcombet and Gaëtan Douéneau, I study formulas through their "growth behaviour" when counting the number of tuples that satisfies said formula in a structure. Over words, it seems that a polynomial growth of an MSO-formula implies that it can be rewritten in FO. Over arbitrary structures, we hope to prove a similar result, using radically different techniques.

REFERENCES

[1] Miklós Ajtai and Yuri Gurevich. 1994. Datalog vs first-order logic. *J. Comput. System Sci.* 49, 3 (1994), 562–588. https://doi.org/10.1016/S0022-0000(05)80071-6

[2] Albert Atserias, Anuj Dawar, and Martin Grohe. 2008. Preservation under extensions on well-behaved finite structures. *SIAM J. Comput.* 38, 4 (2008), 1364–1381. https://doi.org/10.1137/060658709

[3] Albert Atserias, Anuj Dawar, and Phokion G. Kolaitis. 2006. On preservation under homomorphisms and unions of conjunctive queries. *J. ACM* 53, 2 (2006), 208–237. https://doi.org/10.1145/1131342.1131344

[4] Édouard Bonnet, Colin Geniet, Eun Jung Kim, Stéphan Thomassé, and Rémi Watrigant. 2021. Twin-width II: small classes. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 1977–1996.

[5] Édouard Bonnet, Eun Jung Kim, Stéphan Thomassé, and Rémi Watrigant. 2020. Twin-width I: tractable FO model checking. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 601–612.

[6] Walter Bucher, Andrzej Ehrenfeucht, and David Haussler. 1985. On total regulators generated by derivation relations. In *International Colloquium on Automata, Languages, and Programming*. Springer, 71–79.

[7] Chen Chung Chang and H. Jerome Keisler. 1990. *Model Theory*. Studies in Logic and the Foundations of Mathematics, Vol. 73. Elsevier.

[8] Yijia Chen and Jörg Flum. 2021. Forbidden Induced Subgraphs and the Łoś-Tarski Theorem. In *2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*. 1–13. https://doi.org/10.1109/LICS52264.2021.9470742

[9] Artem Chernikov, David Galvin, and Sergei Starchenko. 2020. Cutting lemma and Zarankiewicz's problem in distal structures. *Selecta Mathematica* 26, 2 (2020), 1–27.

[10] Artem Chernikov and Sergei Starchenko. 2018. Regularity lemma for distal structures. *Journal of the European Mathematical Society* 20, 10 (2018), 2437–2466.

[11] Jean Daligault, Michael Rao, and Stéphan Thomassé. 2010. Well-Quasi-Order of Relabel Functions. *Order* 27, 3 (01 Nov 2010), 301–315. https://doi.org/10.1007/s11083-010-9174-0

[12] Anuj Dawar and Abhisekh Sankaran. 2021. Extension Preservation in the Finite and Prefix Classes of First Order Logic. In *29th EACSL Annual Conference on Computer Science Logic (CSL 2021) (Leibniz International Proceedings in Informatics (LIPIcs))*, Christel Baier and Jean Goubault-Larrecq (Eds.), Vol. 183. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 18:1–18:13. https://doi.org/10.4230/LIPIcs.CSL.2021.18

[13] Nachum Dershowitz and Iddo Tzameret. 2003. Gap Embedding for Well-Quasi-Orderings. *Electronic Notes in Theoretical Computer Science* 84 (2003), 80 – 90. https://doi.org/10.1016/S1571-0661(04)80846-6 WoLLIC'2003, 10th Workshop on Logic, Language, Information and Computation.

[14] Max Dickmann, Niels Schwartz, and Marcus Tressl. 2019. *Spectral Spaces*. New Mathematical Monographs, Vol. 35. Cambridge University Press.

[15] Guoli Ding. 1992. Subgraphs and well-quasi-ordering. *Journal of Graph Theory* 16 (1992), 489–502. https://doi.org/10.1002/jgt.3190160509

[16] Jan Dreier, Jakub Gajarský, Yiting Jiang, Patrice Ossona de Mendez, and Jean-Florent Raymond. 2022. Twin-width and generalized coloring numbers. *Discrete Mathematics* 345, 3 (2022), 112746.

[17] Andrzej Ehrenfeucht. 1982. On Regularity of Context-Free Languages; CU. (1982).

[18] Anton Freund. 2020. From Kruskal's theorem to Friedman's gap condition. *Mathematical Structures in Computer Science* 30, 8 (2020), 952–975.

[19] Haim Gaifman. 1982. On local and non-local properties. In *Proceedings of the Herbrand Symposium*. Studies in Logic and the Foundations of Mathematics, Vol. 107. Elsevier, 105–135. https://doi.org/10.1016/S0049-237X(08)71879-2

[20] Jakub Gajarskỳ, Stephan Kreutzer, Jaroslav Nešetřil, Patrice Ossona de Mendez, Michał Pilipczuk, Sebastian Siebertz, and Szymon Toruńczyk. 2018. First-order interpretations of bounded expansion classes. *arXiv preprint arXiv:1810.02389* (2018).

[21] Robert Ganian, Petr Hlinený, Jaroslav Nešetřil, Jan Obdrzálek, and Patrice Ossona de Mendez. 2017. Shrub-depth: Capturing Height of Dense Graphs. *CoRR* abs/1707.00359 (2017). arXiv:1707.00359 http://arxiv.org/abs/1707.00359

[22] Jean Goubault-Larrecq. 2007. On Noetherian spaces. In *Proceedings of LICS'07*. 453–462. https://doi.org/10.1109/LICS.2007.34

[23] Jean Goubault-Larrecq. 2013. *Non-Hausdorff Topology and Domain Theory*. New Mathematical Monographs, Vol. 22. Cambridge University Press.

[24] Jean Goubault-Larrecq. 2021. Infinitary Noetherian Constructions I. Infinite Words. (2021). arXiv:math.GN/1910.10054

[25] Jean Goubault-Larrecq, Simon Halfon, and Aliaume Lopez. 2022. Infinitary Noetherian Constructions II. Transfinite Words and the Regular Subword Topology. (2022). arXiv:math.GN/2202.05047

[26] Jean Goubault-Larrecq and Aliaume Lopez. 2020. On the word topology, and beyond. https://projects.lsv.ens-cachan.fr/topology/?page_id=2735. (2020).

[27] Martin Grohe, Stephan Kreutzer, and Sebastian Siebertz. 2017. Deciding first-order properties of nowhere dense graphs. *J. ACM* 64, 3 (2017), 17. https://doi.org/10.1145/3051095

[28] Frederik Harwath, Lucas Heimberg, and Nicole Schweikardt. 2015. Preservation and decomposition theorems for bounded degree structures. *Logical Methods in Computer Science* Volume 11, Issue 4 (Dec. 2015). https://doi.org/10.2168/LMCS-11(4:17)2015

[29] Ryu Hasegawa. 2002. Two applications of analytic functors. *Theoretical Computer Science* 272, 1 (2002), 113–175. https://doi.org/10.1016/S0304-3975(00)00349-2 Theories of Types and Proofs 1997.

[30] André Joyal. 1986. Foncteurs analytiques et espèces de structures. In *Combinatoire énumérative*, Gilbert Labelle and Pierre Leroux (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 126–159.

[31] Daniel Kirsten. 2002. The finite power problem revisited. *Information processing letters* 84, 6 (2002), 291–294.

[32] Ondvrej Kl'ima and Jonatan Kolegar. 2022. Well quasi-orders arising from finite ordered semigroups. *ArXiv* abs/2203.06535 (2022).

[33] Phokion G. Kolaitis. 2007. Reflections on finite model theory. In *Proceedings of LICS'07*. 257–269. https://doi.org/10.1109/LICS.2007.39

[34] Michal Kunc. 2005. Regular solutions of language inequalities and well quasi-orders. *Theoretical Computer Science* 348, 2-3 (2005), 277–293.

[35] Denis Kuperberg. 2021. Positive First-order Logic on Words. In *36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2021, Rome, Italy, June 29 - July 2, 2021*. IEEE, 1–13. https://doi.org/10.1109/LICS52264.2021.9470602

[36] Jerzy Łoś. 1955. On the extending of models (I). *Fundamenta Mathematicae* 42, 1 (1955), 38–54. https://doi.org/10.4064/fm-42-1-38-54

[37] Roger C. Lyndon. 1959. Properties preserved under homomorphism. *Pacific J. Math.* 9, 1 (1959), 143–154. https://doi.org/10.2140/pjm.1959.9.143

[38] Eric C Milner. 1985. Basic wqo-and bqo-theory. In *Graphs and order*. Springer, 487–502.

[39] Jaroslav Nesetril, Patrice Ossona de Mendez, and Sebastian Siebertz. 2022. Structural Properties of the First-Order Transduction Quasiorder. In *30th EACSL Annual Conference on Computer Science Logic, CSL 2022, February 14-19, 2022, Göttingen, Germany (Virtual Conference) (LIPIcs)*, Florin Manea and Alex Simpson (Eds.), Vol. 216. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 31:1–31:16. https://doi.org/10.4230/LIPIcs.CSL.2022.31

[40] Jaroslav Nešetřil and Patrice Ossona de Mendez. 2010. First order properties on nowhere dense structures. *Journal of Symbolic Logic* 75, 3 (2010), 868–887. https://doi.org/10.2178/jsl/1278682204

[41] Maurice Pouzet. 1972. Un bel ordre d'abritement et ses rapports avec les bornes d'une multirelation. *CR Acad. Sci. Paris Sér. AB* 274 (1972), A1677–A1680.

[42] R. Rado. 1954. Partial well-ordering of sets of vectors. *Mathematika* 1, 2 (1954), 89–95. https://doi.org/10.1112/S0025579300000565

[43] Eric Rosen. 2002. Some aspects of model theory and finite structures. *Bulletin of Symbolic Logic* 8, 3 (2002), 380–403. https://doi.org/10.2178/bsl/1182353894

[44] Benjamin Rossman. 2008. Homomorphism preservation theorems. *J. ACM* 55, 3 (2008), 15:1–15:53. https://doi.org/10.1145/1379759.1379763

[45] Benjamin Rossman. 2016. An improved homomorphism preservation theorem from lower bounds in circuit complexity. *ACM SIGLOG News* 3, 4 (2016), 33–46. https://doi.org/10.1145/3026744.3026746

[46] Nicole Schweikardt, Luc Segoufin, and Alexandre Vigny. 2018. Enumeration for FO queries over nowhere dense graphs. In *Proceedings of PODS'18*. 151–163. https://doi.org/10.1145/3196959.3196971

[47] William W. Tait. 1959. A counterexample to a conjecture of Scott and Suppes. *Journal of Symbolic Logic* 24, 1 (1959), 15–16. https://doi.org/10.2307/2964569

[48] Alfred Tarski. 1954. Contributions to the Theory of Models. I. *Indagationes Mathematicae (Proceedings)* 57 (1954), 572–581. https://doi.org/10.1016/S1385-7258(54)50074-0

[49] Iddo Tzameret. 2002. Kruskal-Friedman Gap Embedding Theorems over Well-Quasi-Orderings. (2002). http://www.cs.tau.ac.il/thesis/thesis/Tzameret-Iddo-MSc-Thesis.pdf

OUT OF CONTEXT DEFINITIONS

I

To keep the main text readable, some definitions have been omitted and are now recalled here.

1 TOPOLOGY

Definition A.1 (Topological Closure). The *closure* $\overline{E}$ of a set E in a topological space (X, τ) is the smallest closed set containing E.

Definition A.2 (Continuous Function). A function $f: X \to Y$ is *continuous* whenever $f^{-1}(U)$ is open in X for every open U of Y.

Definition A.3 (Embeddings). A map $f: A \to B$ is an *embedding* when every open set of A is the pre-image though f of an open set of B.

Definition A.4 (Metric space). A space (X, τ) space is *metric* whenever there exists a distance $d: X \times X \to \mathbb{R}_+$ such that τ is generated using the open balls of d, that is, $\{x \in X \mid d(x, y) < r\}$ for $yinX$ and $r \in \mathbb{R}_+^*$.

Definition A.5 (Lower Vietoris Topology). Given a topological space (X, τ), the *lower Vietoris topology* over $\wp(X)$ is generated by the sets $\Diamond U \triangleq \{E \mid E \cap U \neq \emptyset\}$ for $U \in \tau$.

Definition A.6 (Regular subtree topology). Given a topological space (Σ, θ) the *regular subtree topology* over $T(\Sigma)$ is generated by

$$\frac{U \in \theta \qquad U_1, \dots, U_n \text{ opens}}{\Diamond U \times [U_1, \dots, U_n] \text{ open}}$$

Where $t \in \Diamond U \times [U_1, \dots, U_n]$ if there exists a subtree of t with the root labelled by an element in U and a subset of its children respectively in $U_1, \dots, U_n$.

Definition A.7 (Irreducible Set). A closed subset F of a topological space X is *irreducible* whenever F is non-empty and is not the disjoint union of two non-empty closed sets.

Definition A.8 (Sober Space). A space X is *sober* whenever any irreducible closed subset F is the closure of exactly one point $x \in X$.

Definition A.9 ([23, Definition 8.2.17]). The *sobrification* $\mathcal{S}(X)$ of a topological space (X, τ) is the set of irreducible closed sets of X, and the topology is generated by the sets $\Diamond U \triangleq \{F \in \mathcal{S}(X) \mid F \cap U \neq \emptyset\}$ where U is an open set of X. It can be shown that this construction leads to a sober space, is idempotent up to homeomorphism, and constructs the *free* sober space over X [23, Theorem 8.2.44].

Definition A.10 (Pre spectral space). A *pre-spectral space* is a topological space (X, τ) such that τ is generated by $\mathcal{K}^\circ(X)$, and $\mathcal{K}^\circ(X)$ is stable under finite unions and finite intersections.

Definition A.11 (Spectral space). A *spectral space* is a pre-spectral space that is sober [14, Definition 1.1.5].

2 GRAPHS AND STRUCTURES

Definition A.12 (Graph Minor). A graph G is a *graph minor* of a graph H whenever one can go from H to G using the following rules

(i) One can remove vertices from H.
(ii) One can remove edges from H.
(iii) One can contract edges, i.e. merge two nodes that are on the same edge.

When G is a graph minor of H, we write $G \leq_{\text{minor}} H$.

Definition A.13 (Wide structures). A class of structures $\mathcal{C}$ is *wide* when there exists $\rho: \mathbb{N}^2 \to \mathbb{N}$ such that for all $r, m \in \mathbb{N}^2$, for all $A \in \mathcal{C}$ of size greater than $\rho(n, m)$, there exists a (r, m)-scattered set in A, i.e. a set of at least m points at pairwise distance greater than $2r$.

Definition A.14 (Letters). Given a word $w: \alpha \to \Sigma$ where α is an ordinal, we write $\text{letters}(w)$ for the set $\{w(\beta) \mid \beta < \alpha\}$.

3 ORDERS

Definition A.15 (Complete Lattice)**.** A poset $(L, \leq)$ is a *complete lattice* whenever every family $E \subseteq L$ has a supremum and an infimum.

Definition A.16 (Gap Embedding)**.** For finite words with letters in $\mathbb{N}$, the *gap embedding* is the composition ordering of the monoid $(\mathbb{N}, \max)$.

Definition A.17 (Better Quasi Order)**.** A poset $(X, \leq)$ is a *BQO* whenever for every ordinal α, $X^{<\alpha}$ is a wqo for the embedding on words.

4 ANALYTIC FUNCTORS

We will avoid as much as possible the use of complex machinery related to analytic functors, and use as a definition an equivalent characterisation given by Hasegawa [29, Theorem 1.6]. For an introduction to analytic functors and combinatorial species, we redirect the reader to Joyal [30]. We assume basic understanding of *functors*, *endofunctors* and let C be an arbitrary category. We write $\mathsf{Hom}(A, B)$ for the morphisms between A and B in C.

Definition A.18 (Category of elements)**.** Given G an endofunctor of Set, the *category of elements* $\mathsf{el}(G)$ has as objects pairs (E, a) with $a \in G(E)$, and as morphisms between (E, a) and (E', a') maps $f \colon E \to E'$ such that $G_f(a) = a'$.

As an intuition to the unfamiliar reader, an element (E, a) in $\mathsf{el}(G)$ is a witness that a can be produced through G by using elements of E. Morphisms of elements are witnessing how relations between elements of $G(E)$ and $G(E')$ arise from relations between E and E'. It is quite natural to define the notion of a "smallest" set of elements E such that a can be found in $G(E)$ as a notion of *support*. We replace this definition of "smallest" by the better behaved notion of *transitive object*.

Definition A.19 (Transitive object)**.** A *transitive object* in a category C is an object X satisfying the following two conditions for every object A of C:

- $\mathsf{Hom}(X, A)$ is non-empty;
- The right action of $\mathsf{Aut}(X)$ on $\mathsf{Hom}(X, A)$ by composition is transitive.

Recall that $\mathsf{Aut}(A)$ are the *automorphisms* of A, i.e., maps $g \colon A \to A$ such that there exists $p \colon A \to A$ and $gp = pg = \mathsf{Id}_A$. A transitive object A is an object such that automorphisms of A completely describe morphisms with domain A. Indeed, transitivity means that any morphism $f \in \mathsf{Hom}(A, B)$ can be mapped to any morphism $g \in \mathsf{Hom}(A, B)$ by pre-composing with an automorphism $h \in \mathsf{Aut}(A)$. As an immediate consequence, $\mathsf{Hom}(A, A) = \mathsf{Aut}(A)$ when A is a transitive object, moreover, two transitive objects A and B are isomorphic: the non emptiness conditions give maps $f \in \mathsf{Hom}(A, B)$ and $g \in \mathsf{Hom}(B, A)$, but $fg \in \mathsf{Aut}(A)$ and $gf \in \mathsf{Aut}(B)$, hence $g(fg)^{-1}$ and $(gf)^{-1}g$ are respectively right and left inverses of f, meaning that f is invertible. This is better seen in Figure 22.

$$\begin{tikzcd}
A \arrow[loop above, "fg"] \arrow[loop below, "(fg)^{-1}"] \arrow[r, shift left, "f"] & B \arrow[l, shift left, "g"] \arrow[loop above, "gf"] \arrow[loop below, "(gf)^{-1}"]
\end{tikzcd}$$

Figure 22: Morphisms between two transitive objects in a category C.

In a category, a transitive object can be thought of as a generalisation of the notion of *initial object*, that is, an object A such that $\mathsf{Hom}(A, B)$ contains exactly one element for every object B, in this case $\mathsf{Aut}(A) = \mathsf{Id}_A$. In the category Set of sets and functions, the empty set is an initial object hence a transitive object.

Given an object A in a category C, one can build the *slice category* C/A whose objects are elements of $\mathsf{Hom}(B, A)$ when B ranges over objects of C and morphisms between $c_1 \in \mathsf{Hom}(B_1, A)$ and $c_2 \in \mathsf{Hom}(B_2, A)$ are maps $f \colon B_1 \to B_2$ such that $c_2 \circ f = c_1$. This notion of slice category can be combined with the one of transitive object to build so-called "weak normal forms".

Definition A.20 (Weak Normal Form)**.** A *weak normal form* of an object A in a category C is a transitive object in C/A.

A category C has the *weak normal form property* whenever every object A has a weak normal form. We are now ready to formulate a definition of analytic functors through the existence of weak normal forms for objects in their category of elements.

Definition A.21 (Analytic functor)**.** An endofunctor G of Set is an *analytic functor* whenever its category of elements $\mathsf{el}(G)$ has the weak normal form property. Moreover; X is a finite set for every weak normal form $f \in \mathsf{Hom}((X, x), (Y, y))$ in $\mathsf{el}(G)/(Y, y)$.

5 CATEGORY THEORY

Definition A.22 (F-Algebra)**.** If C is a category and $F\colon \mathsf{C} \to \mathsf{C}$ is an endofunctor of C then a F-algebra is a tuple (A, α).

Definition A.23 (Morphism of F-algebras)**.** A morphism f of F-algebras (A, α) and (B, β) is a map $f \in \mathsf{Hom}(A, B)$ such that the following diagram commutes

$$\begin{tikzcd} F(A) \arrow[r, "\alpha"] \arrow[d, "F_f"] & A \arrow[d, "f"] \\ F(B) \arrow[r, "\beta"] & B \end{tikzcd}$$

Definition A.24 (Initial Algebra)**.** An *initial algebra* is an initial object in the category of F-algebras for a given endofunctor F.

Definition A.25 (Lift)**.** An endofunctor G^T of Top is a *lift* of an endofunctor G of Set if the following diagram commutes, where U is the forgetful functor

$$\begin{tikzcd} \mathsf{Top} \arrow[r, "G^T"] \arrow[d, "U"] & \mathsf{Top} \arrow[d, "U"] \\ \mathsf{Set} \arrow[r, "G"] & \mathsf{Set} \end{tikzcd}$$

6 AUTOMATA THEORY

Definition A.26 (Monoid)**.** A *monoid* is a set M equipped with a binary operation $\cdot$ from M^2 to M that is associative and has a neutral element $e \in M$.