First Order Preservation Theorems in Finite Model Theory : Locality, Topology, and Limit Constructions

Théorèmes de préservation pour la logique au premier ordre : localité, topologie et constructions limites.

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°580 sciences et technologies de l'information et de la communication (STIC)
Spécialité de doctorat : Informatique
Graduate School : Informatique et sciences du numérique. Référent : ENS Paris-Saclay

Thèse préparée au Laboratoire Méthodes Formelles (Université Paris-Saclay, CNRS, ENS Paris-Saclay), et à l'Institut de Recherche en Informatique Fondamentale (Université Paris-Cité, CNRS) sous la direction de **Jean Goubault-Larrecq**, Professeur des Universités, et le co-encadrement de **Sylvain Schmitz**, Professeur des Universités

Thèse soutenue à Paris, le 12 septembre 2023, par

Aliaume LOPEZ

Composition du jury

Membres du jury avec voix délibérative

Serenella Cerrito Professeure des Universités, Université Paris Saclay, Univ EVRY, France	Présidente
Anuj Dawar Professor, University of Cambridge	Rapporteur & Examinateur
Arnaud Durand Professeur des Universités, Université Paris-Cité	Examinateur
Patrice Ossona de Mendez Chargé de Recherche, EHESS	Rapporteur & Examinateur
Daniela Petrisan Maîtresse de Conférences, Université Paris-Cité	Examinatrice
Luca Reggio Senior Fellow Researcher, University College London	Examinateur

THESE DE DOCTORAT

NNT : 2023UPASG046

I would like to express my sincere gratitude to the following individuals and organizations who have supported me during my time as a PhD student.

It feels more natural to start with Sylvain and Jean, who did not just supervise me, but taught me, guided me, encouraged me, and even left me to my own choices, making me grow as a better researcher every step of the way. I can confidently say that I would not have gotten this far without them.

I also would like to thank the members of my thesis committee: Anuj, Arnaud, Daniela, Luca, Patrice, and Serenella. Not only they were the best fitted to evaluate my thesis (as illustrated by the numerous suggestions that significantly improved this document), but also met the complex requirements for a jury to be validated by the administration.

J'aimerais aussi remercier mes collègues et amis qui m'ont soutenu dans ce travail, ont toujours été à l'écoute, et m'ont encouragé dans les moments de doute qui accompagnent toute thèse. En particulier, j'éprouve une certaine mélancolie en me remémorant la parenthèse enchantée que le programme EIG m'a fait vire, dans une époque troublée par les différentes restrictions sanitaires, et l'accueil qui m'a été réservé par des personnes formidables : merci à Dominique, Julien et tout le quatrième étage ! Plus encore, merci à Pierre-Étienne d'avoir été mon compagnon sur le projet SIANCE, d'être devenu un ami cher, et de l'être resté après mon retour à la science académique. Cela ne me fait pas oublier les origines de ma thèse au LSV, en compagnie d'Émilie et Gabriel, avec lesquels j'ai commencé cette aventure entre mots croisés, discussions passionnées et cuisine maison. Au LSV j'ai bien sûr une pensée pour Simon, d'abord chargé de TD puis collègue. Dans la liste de mes bureaux, il en manque un, et ce n'est pas des moindres : je remercie tous les membres de l'IRIF d'avoir été si chaleureux, et d'avoir fourni un environnement si stimulant intellectuellement. Plus particulièrement, je remercie Thomas pour ses interventions régulières et toujours passionnantes, Pierre-Évariste pour avoir toujours répondu à mes questions, même les moins intéressantes, avec la plus grande précision et un soutien sans faille. Je ne sais pas quelle sera ma destinée dans le monde académique (ou ailleurs), mais je peux déjà affirmer que vous allez me manquer. Bien entendu, j'ai une pensée pour Mickaël pour son aide au déménagement, sa patience sans faille quand il s'agit de discuter de tooling*, et d'avoir accepté la lourde responsabilité de référent technique de ma soutenance. Finalement, j'aimerais aussi remercier Alexandra d'avoir été une source de force, de stabilité, de calme, et de restaurants qualitatifs depuis son arrivée.*

Durant ces années de doctorat, j'ai déménagé plusieurs fois, et c'est souvent l'occasion de mesurer la chance que l'on a d'avoir des amis aussi fiables et généreux que Sébastien, Lucas et Rémi. J'espère un jour vous rendre la pareille. À eux s'ajoutent Kostia et Claire, que je ne pourrais pas remercier assez d'aimer jouer à tetris avec les cartons. Pour mes amis chers à mon cœur, mais qui n'ont pas eu la chance de faire partie du club très sélectif des déménageurs de l'extrême, je pense à vous : de la primaire à l'ENS en passant par la prépa, je ne vous ai jamais oubliés, et j'espère ne jamais le faire. Je profite de cet encart pour remercier particulièrement Gaëtan, qui m'a montré que d'autres horizons étaient possibles, et n'a jamais hésité à m'embarquer dans ses aventures : j'espère que les vents te porterons aux sommets, et que tu t'y plairas.

Je déclare formellement Pablo comme le meilleur colocataire, ami et être humain avec qui j'aurais pu espérer partager une partie de ma vie. Merci.

Enfin, je voudrais remercier ma famille pour leur soutien sans faille, et particulièrement Hélène.

Cette thèse n'aurait pas existé sans la présence simultanée des individus mentionnés ci-dessus, mais aussi de tous ceux qui n'ont pas leur nom dans cette liste.

Résumé en Français

Cette thèse est dédiée à l'étude des théorèmes de préservation en logique du premier ordre et leur relativisation à des classes de structures finies. Les théorèmes de préservation font partie d'une famille de résultats de la logique classique qui sont découverts durant la seconde moitié du XXème siècle, et sont motivés par leur rôle de liant entre les propriétés syntaxiques des théories, et leur réalisation en termes de classes de modèles. Ces résultats s'inscrivent donc dans la mouvance de la *Model Theory* qui perdure aujourd'hui. L'exemple archétypique d'un tel théorème est celui de Łoś-Tarski [15, Corollary 3.2.5]. Ce dernier fait correspondre l'ensemble des formules closes φ dites *universelles* à l'ensemble des formules closes ψ qui sont *décroissantes*. Ces considérations théoriques trouvent un intérêt en informatique où la terminaison de certains algorithmes (comme l'algorithme de *Chase* en bases de données) se réduisent à des questions de définissabilité au premier ordre par l'application d'un théorème de préservation idoine.

Une autre raison pour laquelle ces théorèmes "simples," car conséquences de la compacité de la logique au premier ordre, sont encore étudiés en informatique est la suivante : les modèles en informatique sont souvent supposés finis car représentables sur une machine. Or, la transposition de la *Model Theory* aux modèles finis, ingénieusement nommée *Finite Model Theory* (théorie des modèles finis), plus adaptée aux besoins de l'informatique théorique, s'est vite heurtée à des problèmes majeurs : les théorèmes qui sous-tendent la première s'avèrent presque tous devenir faux une fois transposés dans la seconde. La théorie des modèles finis a par conséquent eu recours à des méthodes combinatoires (jeux, localité, décompositions, paramètres de densité) pour appréhender les phénomènes nouveaux qui apparaissent dans le cas fini. Au cours de cet effort de formalisation, certains théorèmes de la théorie des modèles *classique* sont longtemps restés dans un entre deux : si la preuve classique ne peux pas se traduire *verbatim* dans le cas fini, il n'est pas impossible qu'une autre technique de preuve, potentiellement plus combinatoire, puisse être utilisée pour le démontrer. Les théorèmes de préservation font office de cas d'école de théorèmes tombant dans cette catégorie, puisque leur preuve classique est très courte, et repose essentiellement sur des arguments de compacité (théorème qui ne relativise pas aux structures finies). Si certains ont vite été invalidés (le théorème de Łoś-Tarski a été invalidé dans le cas fini en 1959 par Tait), d'autres se sont finalement vu démontrés dans le cas fini (le théorème de préservation par homomorphismes a été démontré dans le cas fini en 2008 par Rossman).

Ainsi, cette thèse s'inscrit dans la lignée des recherches en théorie des modèles finis qui tentent de cartographier les théorèmes de préservation qui relativisent au cas fini. Cependant, plutôt que de se contenter d'étudier quel théorème résiste à la traduction au cas fini, cette thèse suit le modèle inverse (déjà présent dans d'autres travaux) : fixer un théorème de préservation, et cartographier les classes de modèles (finis) pour lesquelles ce théorème relativise. Cette question est intéressante, car elle permet de mieux comprendre et d'affiner les outils de la théorie des modèles finis, mais aussi, car elle correspond à une approche plus "pratique" de l'informatique : les modèles en informatique sont souvent contraints (ou modélisés avec des contraintes), et cela correspond à restreindre l'étude à une sous-classe de modèles.

L'objet principal de cette thèse est la proposition d'un cade théorique basé sur la topologie pour déterminer la relativisation des théorèmes de préservation sur des classes de structures finies. La première contribution est la généralisation des schémas de preuve basés sur la localité de la logique au premier ordre, et la mise en lumière du rôle joué par les formules existentielles locales. La seconde contribution est l'introduction des *espaces préspectraux logiquement présentés* qui sont des espaces topologiques généralisant les arguments de compacité sur lesquels reposent certains schémas de preuve. L'avantage de ces espaces est qu'ils possèdent des propriétés de stabilité qui permettent par l'étude de la relativisation de théorèmes de préservation par des méthodes *compositionnelles*. Une dernière contribution est l'étude d'un sous ensemble des espaces préspectraux logiquement présentés, les espaces Nœthériens, pour lesquels une propriété de stabilité par construction limite est démontrée.

Contents

Preliminaries 1.

Every good thesis in theoretical computer science starts with a small introduction on the importance of theoretical knowledge in some relevant practical system. This thesis will not pretend to do better, but at least will start from a real-life experience.

1.1. SIANCE and Science

After the first year of my PhD, I took time to go and see for myself what was the concrete, day-to-day application of databases, data ingestion, ontology creation and software deployment. This experience was made possible thanks to Etalab's Entrepreneur d'Intérêt Général project, and allowed me to work for the French "Autorité de Sûreté Nucléaire" as a fullstack developer.[1]

Figure 1.1.: The EIG logo.

If you are wondering how this might be even remotely tied to the title of this thesis, this is completely natural. The goal of this little section is to illustrate in a concrete situation the kind of theoretical objects that will be of interest in this manuscript. In particular, it should also justify that the problems that are studied are…interesting.

1: I must put some emphasis on the highly non-standard and fresh way the EIG program was managed by Etalab, and thank (again) Soizic Pénicaud for her amazing work. Unfortunately, this initiative (along with Etalab altogether) will be permanently shut down from 2024 and replaced with a more "corporate" funding of startups, focusing on high-impact projects rather than public service and development of numerical commons.

How does a safety authority works? In full generality, a safety authority is an organisation that will monitor the behaviour of a system. The authority reports whether the given system satisfies correctness properties, preferably in a quantitative and qualitative way. If possible, the authority can report whether the system is likely to violate some correctness property in the near future. Due to budgetary constraints, the authority should be much cheaper to run than the monitored system. The last mission of a safety authority is to build knowledge on the system, in the form of guidelines, scenarios, and improved control efficiency.

In the particular case of the Nuclear Safety Authority, the system consists in *everything that can emit radiations*. This includes hospitals, food factories, paper factories, dentists, and of course nuclear power plants. The information on the state of the system is mainly obtained via direct inspections (a team of inspectors goes on site), and "instructions" (data is given to the inspectors and is checked for consistency). In both cases, the constant flux of incoming data (internal reports, values of some indicators) is a key component in the inspection work.

Figure 1.2.: The ASN logo.

What is expected from computer scientists? Concretely, I was asked to build with my co-worker (Pierre-Étienne Devineau) a service that allows the inspectors to simplify their access to the "knowledge inside the ASN." This means improving the existing workflows (easy access to some documents, aggregate statistics to prepare evaluations, etc.) but also allowing completely new instruction procedures (that were discarded because of the manpower needed to perform them). A common dream was to be able to cross-check every inspection report ever written and discover previously unseen correlations between the employer/location/type of equipment and incident reports.[2]

2: Note that this is partially done in a human way by "remembering" previous inspections, and talking to coworkers at the coffee break.

Practically, the computer science part would be a three-step process: ingest data, analyse data, and expose a query language for the data. From a theoretical point of view, nothing has to be done; it suffices to build a database, use the robust SQL language and ask the data-scientist to add tags via suitable natural language processing. The actual work should be defining a robust pipeline architecture, continuous integration and deployment, a well-defined archive mechanism, proper documentation and upgrade plans, together with tutorials and presentations to the teams for which the product is designed.

What really happened? After 10 month of intermittent development, the real core of SIANCE had emerged: we had built a complete pipeline to cross-check data against several databases, and manually defined ontologies with the inspectors of the ASN. This situation is easy to explain a posteriori, but every discovery was made at the cost of a disappointing pre-production launch.

Partial Data. This should be obvious to anyone that ever filled an administrative form: most of the fields are left blank. This is a minor inconvenience, but it basically means that one has to build suitable *database views* to answer even basic queries[3].

3: How these views are constructed range from simple "select the non-null field" to a more complex ontology driven construction such as "if the inspection was made before this date, and the inspected site was a REP nuclear reactor, then select the value of field X, otherwise use Y."

External Inconsistencies. This should also not be surprising. However, this poses another level of inconsistencies when the primary keys of one dataset are built using a secondary dataset that does not guarantee consistency.

Internal Inconsistencies. This should be more surprising. It can be explained by the fact that some datasets are shared in the form of Excel files, which do not enforce the given integrity constraints. Conversely, some databases that were used enforced "strong" integrity constraints that would push users towards duplicating data in the face of a poorly designed interface.

Constraint Evolution. I assumed that we had a notion of what a consistent dataset would be, but it does not exist outside the collective wisdom of the inspectors. [4]

4: For instance, could the same industrial site not be owned by two different companies? Can a restaurant be subject to an inspection? Can a food factory be inspected? What about a paper factory? A soccer club? Legislation evolves, companies merge, sites get reassigned, and techniques change.

Data History. Because the consistency constraints evolve, old data would often be updated to fit the current state of knowledge. If a company is split in two, then one would retroactively check which part would have been concerned in previous inspections. This allows historical data to be queried and not simply discarded.

Dreaming of verification. This is the paragraph where I pretend that theoretical computer science is useful, so bear with me the minor gaps between theory and practice. Given a complex pipeline trying to address the problem, how does one statically check that it is well-behaved? Here are some properties that we would like:

- The consolidated database is consistent throughout the run
- For every ingested document, a set of simple constraints hold

- Check that predictions are only run when some criterion is met

Even the most trivial pipelines of the form ingest, then consolidate, then update, are hard to verify.[5] The theoretical reader might have some buzzwords in mind right now: ontology mediated query, consistent answering over incomplete databases, schema mappings. These will be presented in time. For now, this is the end of the short story, and the retained solution was simple: horn clauses to complete databases, represented via the all-mighty Excel Spreadsheet. These are simple to write, easy to understand by non-expert users, can be model checked, and above all, the result is *explainable*.

5: By hard, I mean checking whether such a pipeline models a non-trivial property is generally undecidable.

However, from this experience I learned that verification of systems that dynamically update some kind of database is both *necessary* (as designing such system is error-prone) and at the same time *not possible* with the current tooling in a reasonable coding environment.[6]

6: For a data-scientist, this basically means the Python programming language.

1.2. How to Read this Thesis

This thesis aims at providing a theoretical background that might eventually be used to study data-centric systems such as SIANCE. Bear in mind that there are no clear industrial strength solutions yet, and that this thesis is first and foremost of theoretical nature. As a consequence, even when algorithms are described, they remain far from any kind of practical implementation.

Furthermore, this research work takes place at the crossroads of several fields of computer science. This implies that most readers will, at some point, see definitions coming from an area of computer science they are not familiar with. To mitigate this, we will often (re-)introduce folklore results, sometimes in a non-canonical fashion, so that nobody ever gets lost.[7]

7: Even though this is obviously a vain attempt, it is worth trying.

A daring wager. This thesis was co-supervised by Jean Goubault-Larrecq and Sylvain Schmitz. This should already arise suspicions as their only common paper[52] is from 2016. However, both worked on the MPRI course on well-quasi-orderings [23], and separately work on this topic. However, we are going to bridge the "topological" enterprise of Jean [45] together with the more "combinatorial" areas of computer science that are well-quasi-orderings and database theory. This thesis will be at the (surprising?) intersection of Finite Model Theory, Non Hausdorff Topology, and Well-Quasi-Orderings.

[52] Goubault-Larrecq and Schmitz (2016)

[23]: Demeri, Finkel, Goubault-Larrecq, Schmitz and Schnoebelen (2012), 'Algorithmic Aspects of WQO Theory (MPRI course)'

[45]: Goubault-Larrecq (2013), *Non-Hausdorff Topology and Domain Theory*

Le fond et la forme. A lot of effort has been put in this document to ensure that it is both readable, pedagogical, can be printed, and leverages the benefits of the PDF file format. Non exhaustively, we will now discuss and demonstrate some features that might make reading this document more pleasant.

The use of the knowledge package allows the reader to immediately jump[8] to the definition of most of the terms used in this manuscript. Concretely, one can go to the definition of a relational structure, a topological space, or even a bad sequence. A complement of this for

8: Or even preview without jumping, using a recent version of the `evince`, `skim` or `sioyek`. Anyway, it is a good idea to look at the "go back to last position" key binding on your pdf viewer.

printed versions is the presence of an index at the end of the document on page 265.

Because this is quite helpful, the use of the knowledge package was also pushed into the mathematical definitions. For instance, it is possible to click on the different symbols of the following equation to jump to their corresponding definitions:[9]

9: And of course, the import symbols appear in a separate command index.

$$[\![\exists^{\geq k}_{r} x.P(x) \wedge Q(x)]\!]_{\mathsf{Fin}(\sigma)}$$

In parallel, the thesis template is based on the `kaobook` LaTeX template, which advocates for a two-third view, allowing to place comments, figures, and references in a large left or right margin. This not only simplifies the reading process, but also allows us to correctly cite all the authors in place.

Because this thesis aims at being self-contained (even though this is impossible), a lot of results are imported from the literature. To make a clear distinction between those imported statements and newly introduced ones, we will use the following convention: imported statements will not use italic, and they will not be coloured. On the contrary, newly introduced statements will use the italic shape, and will have colours according to the "level" of the statement. Furthermore, we will often use the margin to give proof sketches of results. All of this is illustrated hereafter.

Definition 1.2.1. This is a foreign definition that we import.

Definition 1.2.2. *This is a new definition.*

Lemma 1.2.3. *This is a new lemma.*

Proof. This is a normal proof ■

Proposition 1.2.4. *This is a new proposition.*

Proof Sketch. This is a side proof of the proposition. ■

Corollary 1.2.5. *This is a new corollary.*

Theorem 1.2.6. *This is a new theorem.*

Fact 1.2.7. This is a new fact.

Example 1.2.8. This is a new example.

Exercise 1.2.9. This is a new exercise.

Finally, because some of the background knowledge needed does not fit the narration of the document, we provide in appendix several "cheat-sheets" that list definitions and folklore results on different topics. These can be used in complement of the index and list of symbols. The available cheatsheets are listed hereafter:

However, the "cheatsheets" will not contain duplicates of the results that are presented in the body of the manuscript, and therefore cannot be read as separate entities.

As opposed to the "cheatsheets," the body of the document is split in Chapters that behave as self-contained entities, with their own bibliography, introduction, and concluding remarks.

The document contains three preliminary chapters: Chapter 1 (Preliminaries),[10] Chapter 2 (From Databases To First Order Logic and Back), and Chapter 3 (Preservation Theorems for First Order Queries). These should ideally be read before the rest of the document and in the recommended order.

10: You are currently reading this one.

The contributions are gathered in the following chapters: Chapter 4 (Locality and Preservation), Chapter 5 (A Local-to-Global Preservation Theorem), Chapter 6 (Logically Presented Spaces), Chapter 7 (Topology expanders and Noetherian Topologies), and Chapter 8 (Inductive Constructions). The reader is not forced to read those in a particular ordering, and a few pages are dedicated to the interdependencies and important results in each of these chapters (see page 63). However, some chapters may not be readable before having understood the preliminary chapters.

1.3. You Might Also Like...

I will now use the advertising space available in this document to talk about some of the projects that I think deserve attention.

The Knowledge Package. As every Ph.D student, I spend too much time refining my "tooling" and "writing setup." While most of these efforts lead to unreproducible scripts that are used only once, some of the ideas were actually beneficial in writing this document, and may even be usable by other human beings. Even though the original author of the package may disagree about the following presentation, I believe that the knowledge package of LaTeX is a useful *tool* to write long documents. Abstractly, it provides a name resolution service, paired with

Figure 1.3.: The knowledge package

custom commands that leverage this ability to call the same “entity” using different names.[11]

11: Yes, the knowledge package is a LaTeX DNS.

Concretely, it allows us to tag parts of the document’s text (or maths commands) as being “knowledges,” that can later on be referred to. As a best practice, one creates two extra files along with its text document: one is the DNS table for textual knowledges,[12] and the other one is the DNS table for mathematical knowledges.[13]

12: I like to call it `knowledges.kl`, but the file is really a plain `.tex` document.

13: I like to call this one `mathdefs.tex`, because it contains LaTeX macros outside the knowledge system.

There are three main benefits to using such a setup:

- A poor-man’s debugger,[14] telling you what notions are used, whether you are using a notion that is not defined in your document, or worse, that was defined several times.
- An easier way to create indexes.
- A document that is easier to read due to cross-references between notions at the sentence level, rather than cross-referencing between numbered blocks.

14: This information is outputted in a file called `name-of-the-document.diagnose`.

It may be particularly useful to use the knowledge-clustering program to have a programmatic help when adding knowledges to a document. Finally, for commands related to maths, I cannot recommend enough the `xparse` package that allows the writer to design easy-to-use commands. Hereafter you will find a short demonstration of the coding style used to write this thesis.

```
%% -- in the knowledges.kl file --
\knowledge{notion,
           index parent key=word orderings,
           index key=subword,
           index=subword ordering}
 | subword embedding
 | subword ordering
 | word embedding
 | Higman's word embedding
 | embedding@word
%% -- in the mathdefs.tex file --
\NewDocumentCommand{\DiagFormula}{ O{\exists\AFrag} m }{
    \kl[\DiagFormula]{\Delta_{#2}^{#1}}
}
\knowledge{\DiagFormula}{notion,
    index name=commands,
    index key=diagram formula,
    index={$\DiagFormula{\AModel}$
    -- diagram formula of
       $\AModel$ in fragment $\Frag$}
}
%% -- in the body of the document
We can refer to the \kl{subword embedding},
and to the \kl(word){embedding}, and
they all mean the same thing as
\kl{Higman's word embedding}. It is also
possible to use $\DiagFormula{A}$
command, or to use $\DiagFormual[\EFO]{A}$ to
specify the optional argument.
```

Some Thoughts About the PhD. This part is complete self-promotion, and mainly targeting French citizens, so you can feel free to skip it and go to the actual content of this thesis. For the readers that are interested however, a joint work with Gaëtan Douéneau-Tabot and Laurent Prosperi on the insertion of PhDs inside the French administration (excluding teaching and research positions) is quite fun to look at [29]. As a teaser, one over ten PhD students ends up being a civil servant that does not teach nor does academic research. This raises a question: what are they doing? [28]

[29]: Douéneau-Tabot, Lopez and Prosperi (2021), 'Recrutement et emploi des docteurs dans les administrations publiques'

[28]: Douéneau-Tabot and Lopez (2022), 'Rejoindre l'administration publique après un doctorat'

The "akl" project. It may be a little early to advertise this project, but it is mature enough to run on both Linux and Windows systems. The AKL program[15] allows PDF documents to (transparently) refer to other PDF documents, and works out-of-the box for any PDF. It was primarily designed to make the exploration of bibliographies enjoyable.

15: `https://github.com/AliaumeL/akl`

Combined with the knowledge package, it allows us to write "wikipedia-like" documents that refer to custom sections or results of other documents, which makes easier to create higher quality citations in research papers.

Part I.

Introduction

From Databases To First Order Logic and Back 2.

> For anybody schooled in modern logic, first-order logic can seem an entirely natural object of study, and its discovery inevitable.
>
> William Ewald

Outline of the chapter

This chapter serves as an opinionated introduction to a key technical element of this thesis, namely first order logic. The introduction is motivated by its relationship to database theory, which is used to provide a "computer science oriented" introduction to the field.

Goals of the chapter

At the end of the chapter, I hope to have convinced the reader that relational structures are a good abstraction for databases, and first order logic can be used both as a query language and specification language of database constraints. Specific semantic properties of relational structures and syntactic fragments of first order logic will also be pointed out as particularly relevant in practice.

Reading notes. This chapter introduces well known folklore results about databases and can safely be skipped. Most of the technical content stems from a well established connection between databases and (finite) model theory [31, 54, 68]. As a consequence, no definitions or theorems are new in this chapter.

2.1. Real Databases and Query Languages

Most of the current digital databases are based on the so-called relational model proposed by Codd in [17]. In this model, the central element is the data table, described by a fixed number of named columns, where rows represent the data. From this point of view, most of the world's digital databases are relational[1] because they are represented as sheets in spreadsheets. A very simple example of a tabular representation of data in two spreadsheets is given in Figure 2.1. Although no additional information is given, tables in Figure 2.1 are implicitly connected via the columns **ID** and **Student**. In the relational model, these connections can be specified: it is expected that every student **ID** is unique, and that one cannot grade an exam of a non-existing student.

[17]: Codd (1970), 'A relational model of data for large shared data banks'

1: This is not as great as it might sound, as these spreadsheets do not (easily) enforce database constraints, do not prevent data corruption, do not come with a reasonable query language, do not allow for transactional modifications, do not support concurrent access, and the list goes on forever...

Figure 2.1.: Two spreadsheets representing students and their grades in different exams. The table with columns **ID**, **Name** and **Surname** represents a relation that we will call HasName(id, name, surname), while the second table represents a relation that we will call HasGrade(exam, student, grade). Note that some names are misspelled on purpose in the tables.

ID	Name	Surname
1034	Pierre	Jacques
1010	Piere	Jacqes
1029	Frank	Paul
1030	Jeanne	Rosalie

Exam	Student	Grade
0	1034	10
1	1029	16
0	1027	11
1	1034	15

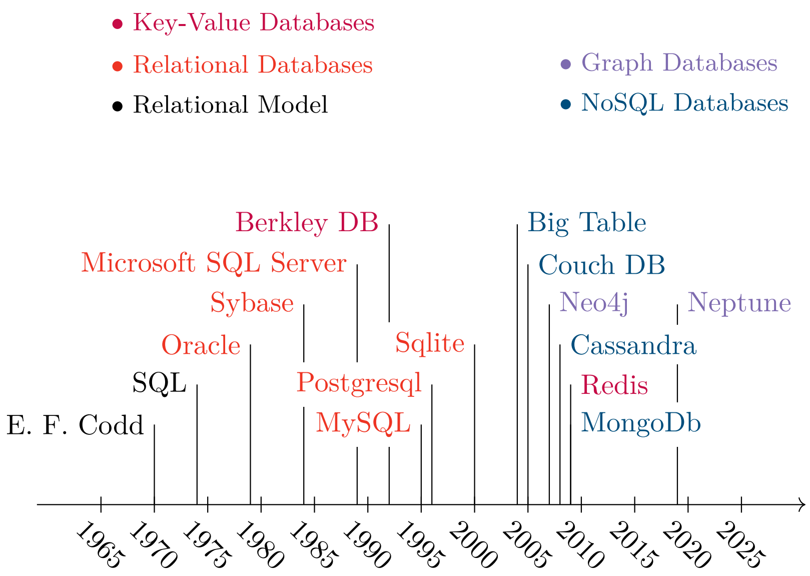


Figure 2.2.: Creation date of widely used commercial and open-source databases since 1970. Extracted from the different websites and Wikipedia pages of the projects.

Together with the stored data, one needs a query language to retrieve the desired information. The most famous query language is SQL, introduced in 1974 by Chamberlin et al. [13]. The SQL language is therefore 48 years old at the time of writing this thesis. Note that SQL coexists with other query languages. Tabular data stored in simple text files is typically queried using AWK (introduced in 1977), while data-scientists that use the Python programming language commonly import the Pandas library to manipulate data.[2]

[13]: Chamberlin and Boyce (1974), 'SEQUEL: A Structured English Query Language'

2: The library's primitives are so obviously related to SQL that the library Pandasql allows us to mix both syntaxes.

In the 21st century, so-called NoSQL databases were re-introduced as a reaction to the hegemony of the relational model, where NoSQL has to be read as "Not Only SQL." These databases answer the issues of scalability, distribution and highly nested structures of large data warehouses that flourished between 2000 and 2010, which can be seen in Figure 2.2. All of these models behave quite differently, and most of the differences come from the set of operations that are allowed, together with their respective expected cost.[3] However, they all fall into the umbrella of the term "database," because they store structured data that can then be queried.

3: This is actually blurred by the newly introduced GraphQL language.

In this thesis, the complexity issues together with the concrete data presentation will essentially be ignored. As a consequence, most of the aforementioned models will essentially look the same, and will all be translated to the generic framework of "First-order finite model theory," from which relational databases take their roots [92].[4]

[92]: Vianu (1997), 'Databases and finite-model theory'

4: For some applications, the relational model may not be the best abstraction. This is the case for "document oriented databases," or for "geographic information systems."

2.2. First Order Logic

It is difficult to introduce a domain that is central to computer science and mathematics. Logic is so crucial, and so deeply incorporated into the curriculum of the modern day computer scientist, that it is easy to forget that it is not as old as humanity itself. To counter this intuition, I strongly suggest the reading of [32] for a historical contextualisation of the birth of the core concepts in today's logic. As a first step towards first-order logic, let us introduce the abstract mathematical representation of databases.

[32]: Ewald (2019), 'The Emergence of First-Order Logic'

Definition 2.2.1. A *relational signature* σ is a non-empty collection $R_1, \ldots, R_n$ of distinct symbols together with a map arity associating to every symbol a *non-zero* natural number. When the collection of symbols is finite, we say that σ is a finite relational signature.

We will often just write $\{(R_1, n_1), \ldots, (R_k, n_k)\}$ for the relational signature with symbols $R_1, \ldots, R_k$, and arities $n_1, \ldots, n_k$.

Unless stated otherwise, signatures will always be assumed to be finite. To model the example database of Figure 2.1 consisting of two tables, one can define σ as two relations R_1 and R_2 of arity 3. In that sense, translating from a database to a relational signature consists in creating a relation symbol per table, and defining its arity to be the number of columns in the corresponding table. Now that we have defined the abstract notion of "table headers," let us introduce the analogue of a database.

Definition 2.2.2. A *relational structure* $\mathfrak{A}$ over a signature σ is a non-empty[5] set $A = \mathrm{dom}(\mathfrak{A})$ together with relations $R^{\mathfrak{A}} \subseteq A^k$ for every relation symbol $R \in \sigma$ of arity k.

The relational structure is said to be *finite* whenever $\mathrm{dom}(\mathfrak{A})$ is a finite set.

5: It is not only "zerology": empty models really do change the theory.

Because most databases considered in computer science are finite, we will give a particular name to the class[6] of finite relational structures: $\mathsf{Fin}(\sigma)$. However, nice mathematical theorems will in general consider arbitrary structures (that is, finite or infinite), whose class we denote by $\mathsf{Struct}(\sigma)$.

We will often use $\mathfrak{A}$ instead of $\mathrm{dom}(\mathfrak{A})$ or A when it is clear from the context that we are talking about the domain. For instance, we will write $a \in \mathfrak{A}$ rather than $a \in \mathrm{dom}(\mathfrak{A})$.

6: We do not wish to dwell on the technicalities related to the notion of classes and sets, and will generally assume that the cardinality of the domains of the considered structures are all bounded by some cardinal κ. This will not be a problem since we will either focus on finite structures up to isomorphisms, or the class of all structures, where the descending Löwenheim-Skolem theorem essentially shows that we can work with countable structures without getting worried.

One can translate a database into a relational structure by considering as domain A the set of all values appearing in the database, and placing them in the corresponding relations. We provide in Figure 2.3 a graphical representation of the relational structure shown in Figure 2.1.

As seen in the depiction from Figure 2.3, the specific column in which a value appears is forgotten in the translation and seen as irrelevant. This implicitly merges data from different tables and performs some kind of "join" operation. This behaviour exhibits several advantages: the mathematical description is simpler, and this naturally captures the so-called "graph databases."

Definition 2.2.3. Let $\sigma \stackrel{\text{def}}{=} \{E\}$ where $\mathsf{arity}(E) = 2$. The class of

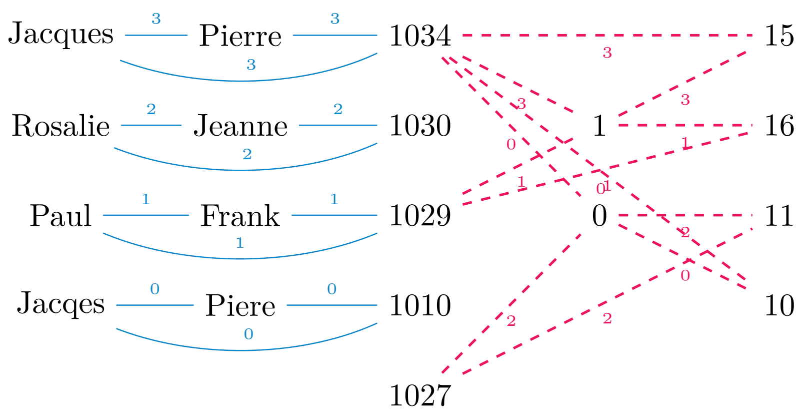


Figure 2.3.: A graphical representation of the relational structure associated with Figure 2.1. The relation HasName is depicted using the blue colour, and the relation HasGrade is depicted using the red colour, with tick and dashed edges. Relations are drawn as edges, together with a number to help realise that HasGrade$(1034, 1, 16)$ is not a valid relation, but HasGrade$(1029, 1, 16)$ is.

finite relational structures over σ is called the class of *directed graphs* and written DiGraphs.

The restriction of DiGraphs to structures where the relation E is symmetric and has no self-loop (i.e. $E(a, a)$ never holds) is called the class of *undirected graphs* FinGraphs.

Notice that mathematical structures can also be represented. In particular, the framework of relational structures and relational signatures is flexible enough to encompass classical mathematical structures such as the real numbers, together with more computer science oriented finite structures representing databases.

In a real database setting, the proposed model is unsatisfactory because it lacks a way to talk about constants. For instance, the number 10 that appears in Figure 2.3 should be considered as a “constant” that can be identified inside various relational structures. This issue will become more apparent when talking about the query language.

Example 2.2.4. The field of real numbers can be described over the signature $\{ (+, 3), (\times, 3), (\leq, 2) \}$. The structure $\mathfrak{A}$ with $\mathrm{dom}(\mathfrak{A}) \stackrel{\text{def}}{=} \mathbb{R}$, and relations defined via

- $(x, y, z) \in +^{\mathfrak{A}} \iff (x + y = z)$,
- $(x, y, z) \in \times^{\mathfrak{A}} \iff (x \times y = z)$,
- $(x, y) \in \leq^{\mathfrak{A}} \iff (x \leq y)$,

is an infinite (uncountable) relational structure.

One major issue in this setting is that it lacks a language to describe integrity constraints. This is problematic because not every relational structure is of interest: in mathematical terms, we want to add axioms. It turns out that the language used to query relational structures can also be used to specify constraints.

2.2.1. Choosing a Query Language

Up until this point, there was no particular reason to specify that we are interested in “first order logic,” and in fact, the term “logic” did not even appear apart from the title of the section. Note that we implicitly assume that equality is a relation symbol that is interpreted as the equality relation over the domain of the structure, unless explicitly stated otherwise.

In the upcoming Definition 2.2.5, we will use the notation $\vec{x}$ to denote a tuple $(x_1, \ldots, x_n)$ of elements, for some $n \in \mathbb{N}$. This notation will be overloaded in several ways. First, we allow ourselves to index the tuples with arbitrary sets. For instance, we will write $\vec{x} \stackrel{\text{def}}{=} (x_a)_{a \in \mathfrak{A}}$, to denote the function $\vec{x}\colon a \mapsto x_a$. Furthermore, we will implicitly cast tuples $\vec{x}$ into sets by considering the set of elements in the tuple when necessary. These notations will be particularly helpful when manipulating free variables.

Definition 2.2.5. Let σ be a finite relational signature, and $\mathbb{V}$ be a countable set of *free variables*. We define for every finite subset $\vec{x} \subseteq_{\mathsf{fin}} \mathbb{V}$ the *formulas* with *free variables* in $\vec{x}$ inductively as follows:

$$\begin{aligned} \varphi(\vec{x}) ::= {} & \varphi(\vec{x'}) \wedge \varphi(\vec{x''}) && \text{when: } \vec{x} = \vec{x'} \cup \vec{x''} \\ \mid {} & R(\iota(1), \ldots, \iota(n)) && \text{when: } R \in \sigma, \mathsf{arity}(R) = n, \iota \in \vec{x}^n \\ \mid {} & \neg\varphi(\vec{x}) \\ \mid {} & \exists y. \varphi(\vec{x} \cup \{ y \}) \\ \mid {} & \top \end{aligned}$$

We call $\mathsf{FO}[\sigma]$ the collection of *first order formulas*, and omit σ when the signature is irrelevant or clear from the context.

Furthermore, we define $\mathrm{fv}(\varphi)$ to be the minimal set $\vec{x}$ of free variables needed to define φ.

A first order formula is said to be a *closed formula* or a *first order sentence*, when it has no free variables. We often omit the empty set and write φ instead of $\varphi(\emptyset)$. Now that we have a language to talk about queries, it remains to provide a concrete execution model that evaluates a query over a given database.[7]

7: This is also referred to as Tarskian semantics.

Given a function $f\colon A \to B$, we write $f[a \mapsto b]$ for the function f' defined by $f'(x) = b$ when $x = a$, and $f'(x) = f(x)$ otherwise.

Definition 2.2.6. The *satisfaction relation* $\models$ is defined by induction on the formulas as follows: given $\mathfrak{A} \in \mathsf{Struct}(\sigma)$, $\vec{x} \subseteq_{\mathsf{fin}} \mathbb{V}$, $\upsilon\colon \vec{x} \to \mathfrak{A}$, and $\varphi \in \mathsf{FO}[\sigma]$

$$\begin{aligned} \mathfrak{A}, \upsilon \models R(\vec{y}) & \stackrel{\text{def}}{\iff} (\upsilon(y_1), \ldots, \upsilon(y_n)) \in R^{\mathfrak{A}} \\ \mathfrak{A}, \upsilon \models \psi_1(\vec{y}) \wedge \psi_2(\vec{z}) & \stackrel{\text{def}}{\iff} \mathfrak{A}, \upsilon \models \psi_1(\vec{y}) \text{ and } \mathfrak{A}, \upsilon \models \psi_2(\vec{z}) \\ \mathfrak{A}, \upsilon \models \neg\psi(\vec{y}) & \stackrel{\text{def}}{\iff} \mathfrak{A}, \upsilon \not\models \psi_1(\vec{y}) \\ \mathfrak{A}, \upsilon \models \exists z. \psi(\vec{y}z) & \stackrel{\text{def}}{\iff} \text{there exists } a \in \mathfrak{A}, \mathfrak{A}, \upsilon[z \mapsto a] \models \psi(\vec{y}z) \\ \mathfrak{A}, \upsilon \models \top & \quad \text{always holds} \quad . \end{aligned}$$

Fact 2.2.7. One can extend the language of first-order logic with disjunction $\psi_1(\vec{x}) \vee \psi_2(\vec{y}) \stackrel{\text{def}}{=} \neg(\psi_1(\vec{x}) \wedge \psi_2(\vec{y}))$, universal quantification $\forall y. \psi(\vec{x}y) \stackrel{\text{def}}{=} \neg\exists y. \neg\psi(\vec{x}y)$, "falsum" $\perp \stackrel{\text{def}}{=} \neg\top$, and implication $\psi_1(\vec{x}) \Rightarrow \psi_2(\vec{y}) \stackrel{\text{def}}{=} \neg a \vee b \neg\psi_1(\vec{x}) \vee \psi_2(\vec{y})$.

As promised, this query language can also be used to formalise the axioms ensuring the integrity constraints.

Example 2.2.8. Following up on the example of Figure 2.1, one could ask that every student is graded. This constraint is expressible in FO as $\forall s.\exists e.\exists g.\, \mathsf{HasGrade}(e, s, g)$.

One of the great successes of the relational model is that queries can be seen as building new relations.[8] This allows us to nest queries. Given a query $\varphi(\vec{x}) \in \mathsf{FO}[\sigma]$ and a relational structure $\mathfrak{A} \in \mathsf{Struct}(\sigma)$, one can construct a new relational structure $\mathfrak{A}' \in \mathsf{Struct}(\sigma \uplus \{ (S, |\vec{x}|) \})$, by simply defining $S^{\mathfrak{A}'} \stackrel{\text{def}}{=} \{ \vec{a} \colon \mathfrak{A}, x \mapsto a_x \models \varphi(\vec{x}) \}$. Let us generalise this construction to define how first order queries can be leveraged to build "first order programs," taking relational structures as input, and outputting relational structures.

8: This is called "views" in the database world, and they are crucial in real-life programs.

Note that we distinguish the syntax and semantics of first order interpretations, exactly as for first order formulas.

Definition 2.2.9. Let σ and σ' be two relational signatures. A (simple) *first order interpretation* I is given by:

- a domain formula $\varphi_{\mathsf{dom}}(x) \in \mathsf{FO}[\sigma]$
- a collection of first-order formulas $(\varphi_R)_{R \in \sigma'}$, where $\varphi_R(\vec{x}) \in \mathsf{FO}[\sigma]$ and has $|\vec{x}| = \mathsf{arity}(R)$ free variables.

Be careful that in the literature, "first order interpretations" are more general than what is described in Definition 2.2.9. We restricted our attention to first order interpretations that are parameter free, do not copy the input, do not restrict domain, and do not quotient the domain of their image.

Definition 2.2.10. Let σ and σ' be two relational signatures, let I be a first order interpretation, and $\mathfrak{A} \in \mathsf{Struct}(\sigma)$. The image $\mathsf{I}(\mathfrak{A})$ is a relational structure with domain $\mathrm{dom}(\mathsf{I}\mathfrak{A})$ defined as $\{ a \in \mathrm{dom}(\mathfrak{A}) \colon \mathfrak{A}, x \mapsto a \models \varphi_{\mathsf{dom}}(x) \}$, and relations $R^{\mathsf{I}(\mathfrak{A})}$ defined as the tuples of $\mathrm{dom}\, \mathsf{I}(\mathfrak{A})$ satisfying $\varphi_R(\vec{x})$.

In particular, any formula $\varphi(\vec{x}) \in \mathsf{FO}[\sigma]$ defines a first-order interpretation with domain $\varphi_{\mathsf{dom}}(x) \stackrel{\text{def}}{=} \top$, and a single relation.

Query Results
1034

Table 2.1.: The result of the query described Example 2.2.11 run over the database Figure 2.1.

Example 2.2.11. The following query selects elements x from Figure 2.1 such that x is the student ID of a student that had a grade of 10 at one of his exams.

$$\varphi(x) \stackrel{\text{def}}{=} \exists y.\exists z.\exists t.\, \mathsf{HasName}(x, y, z) \wedge \mathsf{HasGrade}(t, x, 10)$$

The result of the evaluation is depicted in Table 2.1.

Example 2.2.11 is not a real example. Indeed, the formula contains the term 10, which is a constant and does not appear in the syntax of formulas.

Let us now introduce a simple but crucial first order interpretation that is somehow canonical. It will serve both as an example, and will allow "graph" reasoning on databases in technical developments. Informally, the Gaifman graph of a relational structure is obtained by drawing it as in Figure 2.3, and forgetting the labels on the edges: that is, two elements are connected whenever they participate together in some tuple of some relation of the structure. An example is depicted in Figure 2.4.

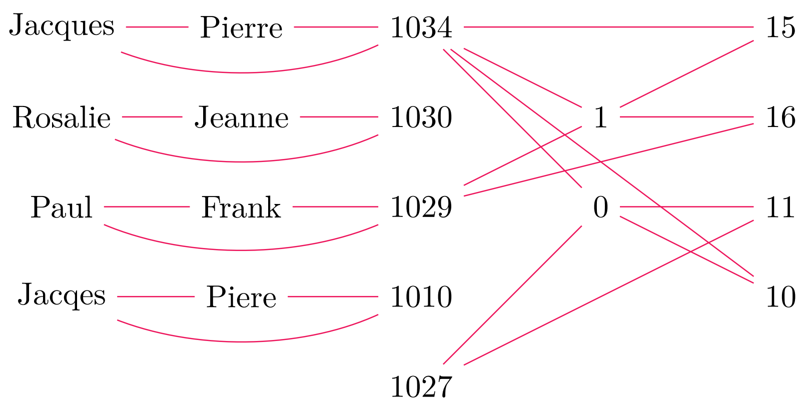


Figure 2.4.: The Gaifman graph of the relational structure described in Figure 2.3. Informally, we simply removed any extra information on the edges of the graphical representation.

Definition 2.2.12. Let σ be a finite relational signature. The *Gaifman graph* first order interpretation $\mathrm{Gaif}(\cdot)$ is defined via:

- $\varphi_{\mathsf{dom}}(x) \stackrel{\text{def}}{=} \top$,
- $\varphi_E(x, y) \stackrel{\text{def}}{=} \bigvee_{(R,n)\in\sigma} \exists z_1, \dots, z_n . R(z_1, \dots, z_n) \wedge \bigvee_{1\leq i,j\leq n} z_i = x \wedge z_j = y$.

Other choices of query languages. Now that we have a formal definition of first order logic, it is possible to compare it to other potential choices of query languages. The first candidate should be SQL, that was already mentioned. It is known that the two languages are equi-expressive[9] since [17].

9: Although one must be careful about this statement!

[17]: Codd (1970), 'A relational model of data for large shared data banks'

2.3. Incompleteness in Databases

The goal of this section is to introduce problems that, once formalised using first order logic, will have a direct connection with the title of this thesis, namely preservation theorems. This will also be the occasion to introduce useful fragments of first-order logic.

Before that, let us briefly recall that some decision problems are non-trivial (and in fact, undecidable) when considering first order logic. This will also be the occasion to illustrate the distinction that arises when considering arbitrary models ($\mathsf{Struct}(\sigma)$) or restricting our attention to finite models ($\mathsf{Fin}(\sigma)$).

Satisfiability. One natural question that arises when considering a database constraint $\varphi \in \mathsf{FO}[\sigma]$ is whether this constraint is satisfied by at least one database. This should guide the user that writes the constraints: it allows us to detect specification errors, but also to check whether the given specification ensures some desired property.

One would typically add a minimal constraint φ that is easy to state and to check, and then ask "do all databases satisfying φ also have the extra property ψ?." This specific query will be written $\varphi \models \psi$. Note that $\varphi \models \psi$ is the same as $\top \models \varphi \Rightarrow \psi$, also written as $\models \varphi \Rightarrow \psi$, or $\varphi \wedge \neg\psi \models \bot$.

Table 2.2.: Algorithmic properties of the decision problem regarding the entailment $\varphi \models \psi$, over $\mathsf{Struct}(\sigma)$ and $\mathsf{Fin}(\sigma)$.

	semi-decision	**co-semi-decision**
$\mathsf{Struct}(\sigma)$	✓[43]	✗[43]
$\mathsf{Fin}(\sigma)$	✗[91]	✓(folklore)

It turns out that checking if $\varphi \models \psi$ is undecidable, but (as promised) the nature of this roadblock differs in the finite case. Whether $\varphi \models_{\mathsf{Struct}(\sigma)} \psi$ is semi-decidable[10] but not co-semi-decidable [43]. On the contrary, when restricting our attention to finite models, $\varphi \models_{\mathsf{Fin}(\sigma)} \psi$ is co-semi-decidable, but not semi-decidable [91]. This is summarised in Table 2.2, and illustrates the discrepancy between the two models.

Let us seize the opportunity to introduce a new notation connected to $\models$: two formulas φ and ψ are equivalent on a class $\mathcal{C}$, which is written $\varphi \equiv_{\mathcal{C}} \psi$ whenever $\models_{\mathcal{C}} (\varphi \iff \psi)$ holds. One benefit of this notation is that it allows us to concretely state over which class of models this equivalence is considered, for instance $\varphi \equiv_{\mathsf{Fin}(\sigma)} \psi$ means that for all $\mathfrak{A} \in \mathsf{Fin}(\sigma)$, $\mathfrak{A} \models \varphi$ if and only if $\mathfrak{A} \models \psi$.

Recovering from bad states. Assuming that the given specification φ is not absurd, it might be that a given database of interest $\mathfrak{A}$ is such that $\mathfrak{A} \not\models \varphi$.[11] A key property that one wants to achieve in these settings is to somehow recover from this bad state. This can be done in two ways: one is to patch the model so that it satisfies the constraints, the other is to patch the query evaluation procedure to incorporate a notion of "invalid tuples."

We will restrict our attention to a useful specialisation of this database recovery setting, namely the case of "incomplete databases," where the problem is that some information is lacking in the model.

2.3.1. Incomplete Databases and Fragments

To represent "incompleteness" in databases, we will take the dual approach of describing how a database can be "completed." This will be done by describing which kind of maps between databases are considered as valid completions.[12] Different choices of completion can be chosen to model different real world situations, and we will restrict our attention to the three simplest forms: induced substructures, substructures, and homomorphisms.[13]

In a setting where the database at hand is incomplete, it is natural to adapt the query evaluation procedure to account for the potential missing information. This is not what we will do in this section. Instead, we will search syntactic criteria over sentences of $\mathsf{FO}[\sigma]$ that will imply that the usual evaluation procedure is compatible with the missing information.

It will be useful to refer to specific sets of sentences in $\mathsf{FO}[\sigma]$ as a whole when interested in a particular semantic property (such as, allowing the naïve evaluation to work). This is the motivation behind our definition of a fragment of first order logic.[14]

10: To keep the focus on logic and relational structures, we do not focus on the notions of decidability, semi-decidability, and co-semi-decidability. Here, we can understand semi-decidable as the existence of an effective method that given a formula φ and a formula ψ, answers 'YES' if $\varphi \models_{\mathsf{Struct}(\sigma)} \psi$, and can fail to terminate, or answer anything, if $\varphi \not\models_{\mathsf{Struct}(\sigma)} \psi$. A problem is co-semi-decidable when its complement is semi-decidable, and decidable when it is both semi-decidable and co-semi-decidable.

11: This can happen because $\mathfrak{A}$ comes from an external untrusted source, because of a bug in a program managing the database, or because $\mathfrak{A}$ is obtained by "consolidation" of several heterogeneous sources of information.

12: That is, we represent an incomplete database as the collection of its valid completions.

13: The naming convention may be counter-intuitive, so let us spell out what we mean by "completions." When $\mathfrak{A} \subseteq_i \mathfrak{B}$, we say that $\mathfrak{B}$ seen as a *completion* of $\mathfrak{B}$ for the induced substructure relation. The same holds for the relation $\subseteq$: if $\mathfrak{A} \subseteq \mathfrak{B}$, then $\mathfrak{B}$ is a *completion* of $\mathfrak{A}$ for the substructure relation.

14: Note that in the literature, this word is mostly used to refer to subsets of $\mathsf{FO}[\sigma]$. However, most of the fragments enjoy extra properties that we will actually need later on to develop our theory of preservation theorems.

One could technically write σ to specify that we see it as a fragment of $\mathsf{FO}[\sigma]$, but we find more pleasant to only state $\mathsf{F} \subseteq \mathsf{FO}[\sigma]$ to avoid cluttered notations.

Assuming that $x = y$ is a valid formula is the natural continuation of the hypothesis that the equality relation is always in the signature: it should always be possible to talk about equality of datum.

In Definition 2.3.1, we use the term α-renaming, but it will not be defined in this document. Informally, it consists in freely renaming bound variables without introducing name clashes.

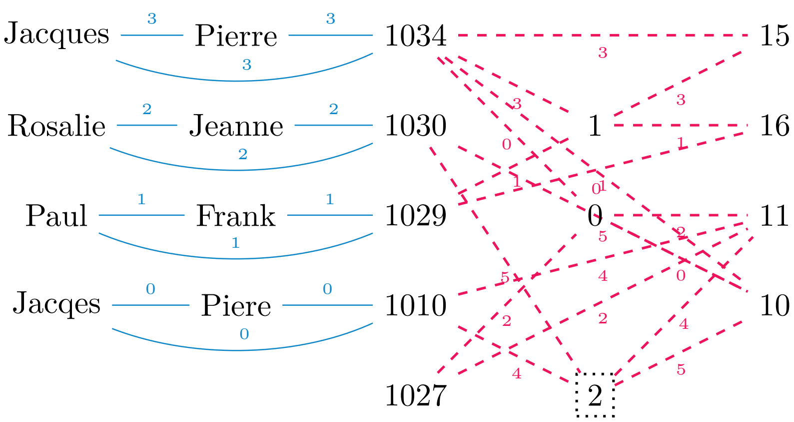


Figure 2.5.: Graphical representation of the extension of Figure 2.3 described in Example 2.3.2. The extra exam (number 2) is represented using a dotted box.

Definition 2.3.1. *Let σ be a relational signature (potentially infinite). A fragment* F *of* $\mathsf{FO}[\sigma]$ *is a subset of* $\mathsf{FO}[\sigma]$ *that contains the equality relation* $x = y$*, and that is stable under the following operations:*

1. *Finite conjunction:* $\varphi, \psi \in \mathsf{F}$ *implies that* $\varphi \land \psi \in \mathsf{F}$*.*
2. *Finite disjunction:* $\varphi, \psi \in \mathsf{F}$ *implies that* $\varphi \lor \psi \in \mathsf{F}$*.*
3. *Renaming:* $\varphi(\vec{x}) \in \mathsf{F}$ *implies that* $\varphi'(\vec{y}) \in \mathsf{F}$*, for any α-renaming* φ' *of* φ*, and (arbitrary) renaming of* $\vec{x}$ *as* $\vec{y}$ *using variables in* $\mathbb{V}$*.*

The extra assumptions are not problematic as one can always complete a subset $S \subseteq \mathsf{FO}[\sigma]$ into the smallest fragment containing S.

Induced Substructures. One way to think about incomplete databases is to consider that some elements of the domain might be missing. Concretely, in our example shown in Figure 2.3, we have a student (Marc Francis, with student id 1010) that has no grade. In this case, a reasonable completion ensuring that the database constraint of Example 2.2.8[15] holds, is to assume that there is a missing exam in our system.

15: Stating that every student should have a grade.

Example 2.3.2. One can complete the database Figure 2.3 by adding an exam 2 and two relations:

- $\mathsf{HasGrade}(2, 1010, 10)$,
- $\mathsf{HasGrade}(2, 1030, 11)$.

After this addition, the database satisfies the constraint that every student has a grade. A graphical representation is given in Figure 2.5.

Let us now provide a formal definition of what "adding missing elements" to our database means in terms of maps.

Definition 2.3.3. A database $\mathfrak{B}$ is an *extension* of a database $\mathfrak{A}$ when there exists an injective map $f\colon \mathsf{dom}(\mathfrak{A}) \to \mathsf{dom}(\mathfrak{B})$ such that for every relation symbol $R \in \sigma$, every tuple $\vec{a} \in \mathsf{dom}(\mathfrak{A})$:

$$\vec{a} \in R^{\mathfrak{A}} \text{ if and only if } f(\vec{a}) \in R^{\mathfrak{B}} \quad .$$

We call the map f an *embedding*, and we also say that $\mathfrak{A}$ is an *induced substructure* of $\mathfrak{B}$. We use the special notation $\mathfrak{A} \subseteq_i \mathfrak{B}$ for the quasi-ordering that the *induced substructure* relation defines.

Queries using universal quantifiers will not play well with this notion of completion: if there are missing elements, how can we say that "all the elements" satisfy something?

It turns out that prohibiting the use of the universal quantification[16] is sufficient to guarantee that the naïve evaluation algorithm works as expected. Before proving this fact, let us first consider an easier subset of formulas without any kind of quantification.

16: Beware that universal quantification is not really in the syntax of first order formulas, and arises as the interaction between the negation and the existential quantification. A formal definition will be given in Definition 2.3.7.

We will often write QF without specifying over which signature.

Definition 2.3.4. A first order formula $\varphi \in \mathsf{FO}[\sigma]$ is a *quantifier free formula* whenever it does not contain universal quantifiers $(\forall x)$ nor existential quantifiers $(\exists x)$. We write $\mathsf{QF}[\sigma]$ for the subset of *quantifier free formulas* of $\mathsf{FO}[\sigma]$.

Fact 2.3.5. Let σ be a finite relational signature, let $\varphi(\vec{x})$ be a quantifier free formula, let $\mathfrak{A}, \mathfrak{B} \in \mathsf{Struct}(\sigma)$, and let $f : \mathfrak{A} \to \mathfrak{B}$ be an embedding.

Then, whenever $\upsilon : \vec{x} \to \mathfrak{A}$ is such that $\mathfrak{A}, \upsilon \models \varphi(\vec{x})$, we have that $\mathfrak{B}, f \circ \upsilon \models \varphi(\vec{x})$.

We illustrate in Example 2.3.6 how Fact 2.3.5 concretely applies in a database setting.

In this example, we understand why constants could be a useful addition to the language. Indeed, the notion of embedding allows us to replace every 10 by a 11 and claim that this is the same database: in the relational model, only relations count. Adding constants, such as 10 to the query language would account to the *data* that is stored.

Example 2.3.6. Let $\varphi(s_1, s_2, e, g)$ be defined as the following query:

$$\varphi(s_1, s_2, e, g) \stackrel{\text{def}}{=} \mathsf{HasGrade}(e, s_1, g) \land \mathsf{HasGrade}(e, s_2, g) \quad .$$

This query asks whether the two students s_1 and s_2 had the same grade g at a given exam e. Observe that in Figure 2.3, no two students had the same grade at a given exam.

Assume that in some database $\mathfrak{A}$, there exists a quadruplet of elements (s_1, s_2, e, g) that satisfies φ. Then, for every embedding $f : \mathfrak{A} \to \mathfrak{B}$, the tuple $(f(s_1), f(s_2), f(e), f(g))$ satisfies φ.

Generalising Example 2.3.6 and Fact 2.3.5, one can ask "positive properties" stating the existence of some tuple of elements that satisfy a given quantifier free formula. Concretely, we are defining the existential fragment of first order logic.

Similarly, we define *universal formulas* as negations of existential formulas.

Definition 2.3.7. A first order formula is an *existential formula* whenever it can be written using the restricted syntax: $\land$, $\lor$, $\top$, $\bot$, $\neg R(\vec{x})$, $R(\vec{x})$, and $\exists x$. The collection of *existential formulas* is written $\mathsf{EFO}[\sigma]$.

It is worth noting that the collection of *existential sentences*[17] is as expressive as the fragment of existential closure of quantifier free formulas, that we write $\exists\mathsf{QF}$.[18] While it is true that Fact 2.3.5 generalises to EFO, we will simply state its consequence on $\exists\mathsf{QF}$, i.e., the sentences in EFO, because it avoids the use of free variables.

17: first order sentences that are also existential formulas

18: Beware that equi-expressiveness can hide other parameters, such as the minimal size of a sentence expressing a given property.

Formally, $\exists\mathsf{QF}$ is not a fragment. However, every sentence φ in the smallest fragment containing $\exists\mathsf{QF}$ is equivalent to an $\exists\mathsf{QF}$ sentence. To simplify notations and reasoning, we will talk about $\exists\mathsf{QF}$ as a fragment.

Fact 2.3.8. Let σ be a finite relational signature, let $\varphi \in \exists\mathsf{QF}$ be an existential sentence, and let $\mathfrak{A}, \mathfrak{B} \in \mathsf{Struct}(\sigma)$ be such that $\mathfrak{A} \subseteq_i \mathfrak{B}$. Then, $\mathfrak{A} \models \varphi$ implies $\mathfrak{B} \models \varphi$.

In general, Facts 2.3.5 and 2.3.8 state that a subset of formulas are compatible with a given notion of completion. Let us formally define this notion of compatibility, enabling us to later on compare these notions of completions.

Definition 2.3.9. A first order sentence φ is *preserved under* a relation $\leq \in \mathsf{Struct}(\sigma) \times \mathsf{Struct}(\sigma)$ if and only if for all $\mathfrak{A}, \mathfrak{B} \in \mathsf{Struct}(\sigma)$ such that $\mathfrak{A} \leq \mathfrak{B}$, $\mathfrak{A} \models \varphi$ implies $\mathfrak{B} \models \varphi$.

Using Definition 2.3.9, we can restate Fact 2.3.8 as follows: "existential sentences are preserved under extensions." Furthermore, this is the "correct" fragment of sentences because for every finite model $\mathfrak{A}$, there exists an existential sentence that characterises its extensions.

In the Fact 2.3.10, the sentence $\psi_{\mathfrak{A}}$ is essentially stating the existence of all the elements in $\mathrm{dom}(\mathfrak{A})$ and listing the relations (or lack thereof) between these elements.

Fact 2.3.10. Let σ be a finite relational signature, and $\mathfrak{A} \in \mathsf{Fin}(\sigma)$. There exists a first order formula $\psi_{\mathfrak{A}}((x_a)_{a\in\mathfrak{A}}) \in \mathsf{QF}[\sigma]$ such that the following are equivalent for all $\mathfrak{B} \in \mathsf{Struct}(\sigma)$ and $\nu \colon \vec{x} \to \mathfrak{B}$.

1. $\mathfrak{B}, \nu \models \psi_{\mathfrak{A}}$
2. $(\nu \circ (a \mapsto x_a)) \colon \mathfrak{A} \to \mathfrak{B}$ is an embedding of relational structures.

As a consequence, there exists a first order sentence $\theta_{\mathfrak{A}} \in \mathsf{EFO}$,[19] such that the following are equivalent for all $\mathfrak{B} \in \mathsf{Struct}(\sigma)$:

1. $\mathfrak{B} \models \theta_{\mathfrak{A}}$,
2. $\mathfrak{A} \subseteq_i \mathfrak{B}$.

19: Namely, $\theta_{\mathfrak{A}} = \exists\vec{x}.\psi_{\mathfrak{A}}$.

Substructures. 'Missing entries' is a natural framework, which motivates the use of extensions of structures. However, this framework fails to capture the most common case of incomplete data: databases where some rows are only partially filled.[20] Concretely, this means that we may not know all the elements present in the database, and furthermore may even not know all the facts about the elements that we do know about. Formally, this is the notion of a substructure.[21]

20: This will happen every time a field is set as optional in a form.

21: Note that in classical model theory textbooks, the notion of "substructure" actually refers to what is called induced substructure in this thesis.

Definition 2.3.11. A database $\mathfrak{A}$ is a *substructure* of a database $\mathfrak{B}$ when there exists an injective map $f \colon \mathrm{dom}(\mathfrak{A}) \to \mathrm{dom}(\mathfrak{B})$ such that for every relation symbol $R \in \sigma$, every tuple $\vec{a} \in R^{\mathfrak{A}}$, $f(\vec{a}) \in R^{\mathfrak{B}}$. We call such a map a *substructure embedding*, and write $\mathfrak{A} \subseteq \mathfrak{B}$ when such a map exists.

Continuing on our example Example 2.3.2, we can now find new ways to "repair" the database to satisfy our constraint, without adding any new exam!

Note that Example 2.3.12 builds a model that is not an extension of Figure 2.3.

Example 2.3.12. One can complete the database shown in Figure 2.3 by adding two relations:

- $\mathsf{HasGrade}(0, 1010, 11)$, and
- $\mathsf{HasGrade}(1, 1030, 11)$.

After this addition, the database satisfies the constraint that every student has a grade.

Quite naturally, the quantifier free formulas and existential sentences are now too powerful to be naïvely evaluated. For instance, the sentence stating that "there exists a student that has no grade" holds in Figure 2.3, but does not hold in the completion of Example 2.3.12, even though this is an existential sentence.

We define $\mathsf{PQF}^{\neq}[\sigma]$ for the subset of quantifier free formulas that do not use the negation operator $\neg$, except when checking the inequality of two variables via $x \neq y$, that is written $\neg(x = y)$ in our syntax.[22] Enriching these formulas by allowing the use of existential quantifiers, we obtain the subset $\mathsf{EPFO}^{\neq}[\sigma]$, which is as expressive as $\exists\mathsf{PQF}^{\neq}[\sigma]$ when restricted to first order sentences.

22: Formally, we allow formulas that are built using $\wedge$, $\vee$, $\top$, $\bot$, $R(\vec{x})$, and $x \neq y$.

Preventing the use of negations allows us to state an analogue of the preservation property described in Fact 2.3.8. As for EFO, an analogue of Fact 2.3.10 holds for $\mathsf{EPFO}^{\neq}$, showing how the fragment is large enough to describe all the possible completions of a finite model.

Fact 2.3.13. Let φ be a first order sentence. If $\varphi \in \mathsf{EPFO}^{\neq}$, then it is preserved under $\subseteq$.

Fact 2.3.14. Let σ be a finite relational signature, and $\mathfrak{A} \in \mathsf{Fin}(\sigma)$. There exists a first order formula $\psi_{\mathfrak{A}}((x_a)_{a\in\mathfrak{A}}) \in \mathsf{PQF}^{\neq}[\sigma]$ such that the following are equivalent for all $\mathfrak{B} \in \mathsf{Struct}(\sigma)$ and $\nu \colon \vec{x} \to \mathfrak{B}$.

1. $\mathfrak{B}, \nu \models \psi_{\mathfrak{A}}$
2. $(\nu \circ (a \mapsto x_a)) \colon \mathfrak{A} \to \mathfrak{B}$ is a substructure embedding of relational structures.

Furthermore, there exists a sentence $\theta_{\mathfrak{A}} \in \mathsf{EPFO}^{\neq}[\sigma]$,[23] such that the following are equivalent for all $\mathfrak{B} \in \mathsf{Struct}(\sigma)$:

1. $\mathfrak{B} \models \theta_{\mathfrak{A}}$
2. $\mathfrak{A} \subseteq \mathfrak{B}$.

23: Namely, $\theta_{\mathfrak{A}} = \exists \vec{x}.\psi_{\mathfrak{A}}$.

Homomorphisms. While the notion of substructure already allowed us to insert new relations to "complete" the databases, we argue that one useful completion method is still missing. Whenever data is filled in by human users (which happens quite often) there is a risk that the same entity is given different names in the database.[24] As a con-

24: Ironically, this typically happens when the input form tries to enforce uniqueness constraints, and users just invent new identifiers to get bypass the restriction and continue their day.

sequence, a reasonable additional operation is to merge elements rather than add new ones. Such an operation can be formally defined as a homomorphism.

Definition 2.3.15. A *homomorphism* between two relational structures $\mathfrak{A}$ and $\mathfrak{B}$ is a map $h\colon \mathrm{dom}(\mathfrak{A}) \to \mathrm{dom}(\mathfrak{B})$ such that for every relation symbol $R \in \sigma$, and every tuple $\vec{a} \in R^{\mathfrak{A}}$, $f(\vec{a}) \in R^{\mathfrak{B}}$.

We write $\mathfrak{A} \preceq_h \mathfrak{B}$ when such a map exists.

Continuing with our running example of Examples 2.3.2 and 2.3.12, let us illustrate how the ability to merge elements allows us to repair our database.

Example 2.3.16. One can complete the database shown in Figure 2.3 by merging “Jacques” with its misspelled counterpart “Jacqes,” merging “1034” with the identifier “1010”, and merging “Pierre“ with its misspelled counterpart “Piere.”

After this modification, the database satisfies the constraint that every student has a grade.

Note that Example 2.3.16 builds a model that is neither an extension of Figure 2.3, nor contains it as a substructure.

Remark that because we lack constants, we were already able to “rename” the stored data, even in the case of extensions. However, it was not possible to collapse several values into a single one, due to the presence of $\neq$.

Following what was done for induced substructure and substructures, we now want to characterise a suitable fragment of $\mathsf{FO}[\sigma]$ that corresponds to this new notion of incompleteness. Because completions can merge data, queries that use the non-equality test $x \neq y$ are probably badly behaved.[25]

25: Note that using “unique identifiers” also becomes problematic in a setting where data can be merged arbitrarily.

We define $\mathsf{PQF}[\sigma]$ for the subset of quantifier free formulas that do not use the negation operator $\neg$, nor the inequality $x \neq y$. As per usual, we introduce $\mathsf{EPFO}[\sigma]$ for the fragment of $\mathsf{FO}[\sigma]$ obtained when allowing to use existential quantifiers inside PQF formulas, and remark that $\exists\mathsf{PQF}[\sigma]$ is as expressive as the restriction of $\mathsf{EPFO}[\sigma]$ to sentences. Sentences in $\mathsf{EPFO}[\sigma]$ are called *existential positive sentences*.

The following two results should not be surprising, as they state the role of EPFO with respect to homomorphisms as analogue to EFO for induced substructures and $\mathsf{EPFO}^{\neq}$ for substructures.

Fact 2.3.17. Let φ be a first order sentence. If $\varphi \in \mathsf{EPFO}$, then it is preserved under $\preceq_h$.

Definition 2.3.18 [14]**.** Let σ be a finite relational signature, and $\mathfrak{A} \in \mathsf{Fin}(\sigma)$. There exists a first order formula $\psi_{\mathfrak{A}}((x_a)_{a\in\mathfrak{A}}) \in \mathsf{PQF}[\sigma]$ such that the following are equivalent for all $\mathfrak{B} \in \mathsf{Struct}(\sigma)$ and $\nu\colon \vec{x} \to \mathfrak{B}$.

1. $\mathfrak{B}, \nu \models \psi_{\mathfrak{A}}$
2. $(\nu \circ (a \mapsto x_a))\colon \mathfrak{A} \to \mathfrak{B}$ is a homomorphism of relational structures.

Furthermore, there exists a sentence $\theta_{\mathfrak{A}} \in \mathsf{EPFO}[\sigma]$,[26] such that the

26: Namely, $\theta_{\mathfrak{A}} = \exists \vec{x}.\psi_{\mathfrak{A}}$.

following are equivalent for all $\mathfrak{B} \in \mathsf{Struct}(\sigma)$:

1. $\mathfrak{B} \models \theta_{\mathfrak{A}}$
2. $\mathfrak{A} \preceq_h \mathfrak{B}$.

It turns out that PQF is exactly the fragment of "unions of conjunctive queries" that plays a key role in database theory as corresponding to the "SELECT-PROJECT-JOIN" SQL queries [2].

[2]: Abiteboul, Hull and Vianu (1995), *Foundations of Databases*

We will now spend a little more time explaining how Definition 2.3.18 introduces a nice complexity measure on finite relational structures. To the best of my knowledge, it was first noticed by Rossman that the formulas expressing the class of completions of a model can be inspected to extract a complexity measure for the structure [83]. Prior to the work of Rossman, an analysis of the quantifier rank needed to distinguish finite structures was carried out in [79], but this analysis is focused on separating finite structures, rather than describing embeddings.

[83]: Rossman (2008), 'Homomorphism preservation theorems'

[79]: Pikhurko and Verbitsky (2005), 'Descriptive complexity of finite structures: Saving the quantifier rank'

Recall that $\wedge$, $\forall$, and $\perp$ are not part of the syntax of first order formulas. We *deduce* from Definition 2.3.19, that $\mathrm{rk}(\forall x.\psi) = \mathrm{rk}(\neg\exists x.\neg\psi) = 1 + \mathrm{rk}(\psi)$.

Definition 2.3.19. The *quantifier rank* of a first order formula is defined inductively as follows

$$\begin{aligned}
\mathrm{rk}(\exists x.\psi) &\overset{\text{def}}{=} 1 + \mathrm{rk}(\psi) \\
\mathrm{rk}(\psi_1 \vee \psi_2) &\overset{\text{def}}{=} \max(\mathrm{rk}(\psi_1), \mathrm{rk}(\psi_2)) \\
\mathrm{rk}(\neg\psi) &\overset{\text{def}}{=} \mathrm{rk}(\psi) \\
\mathrm{rk}(R(\vec{x})) &\overset{\text{def}}{=} 0 \\
\mathrm{rk}(\top) &\overset{\text{def}}{=} 0 \quad .
\end{aligned}$$

Using this notion of quantifier rank, we can now look at the minimal quantifier rank that can be used to talk about a given (finite) structure. We use as a definition of tree depth a consequence of two lemmas in [83], because it highlights the role of diagram sentences.

[83]: Rossman (2008), 'Homomorphism preservation theorems'

Definition 2.3.20 [83, Lemmas 2.13 and 2.14]**.** Let σ be a finite relational signature, and let $\mathfrak{A} \in \mathsf{Fin}(\sigma)$. The *tree depth* of $\mathfrak{A}$, written $\mathrm{td}(\mathfrak{A})$ is defined as the minimum of $\mathrm{rk}(\psi)$ among all sentences $\psi \in \mathsf{EPFO}[\sigma]$ such that $\mathfrak{B} \models \psi$ if and only if $\mathfrak{A} \preceq_h \mathfrak{B}$ for all $\mathfrak{B} \in \mathsf{Struct}(\sigma)$.

Beware that in Definition 2.3.20, it is necessary to talk about EPFO and to avoid sentences in $\exists$PQF. The sentence $\exists x.\exists y.E_1(x,y) \wedge \exists z.E_2(x,z)$ is in EPFO and has quantifier rank 2, but any equivalent sentence in $\exists$PQF has quantifier rank at least 3.

This definition given in terms of logic can be translated in combinatorial terms.[27] Beware that the combinatorial definition is stated over undirected graphs, and only then transported to arbitrary structure via the Gaifman graph construction.

27: Which was the original definition of tree depth in the literature.

Definition 2.3.21 [78, Chapter 6]**.** The tree depth of a graph $\mathfrak{A} \in$ FinGraphs can be computed inductively as follows:

$$
\begin{aligned}
\mathrm{td}(\{\, a \,\}) &= 1 \\
\mathrm{td}(\mathfrak{A} \uplus \mathfrak{A}') &= \max(\mathrm{td}(\mathfrak{A}), \mathrm{td}(\mathfrak{A}')) && \text{if } \mathfrak{A} \text{ not connected} \\
\mathrm{td}(\mathfrak{A}) &= \min \{\, \mathrm{td}(\mathfrak{A} \setminus \{\, a \,\}) + 1 \colon a \in \mathfrak{A} \,\} && \text{otherwise}
\end{aligned}
$$

If σ is a finite relational signature, the tree depth of $\mathfrak{A} \in \mathsf{Fin}(\sigma)$ can be computed as $\mathrm{td}(\mathrm{Gaif}(\mathfrak{A}))$.

This notion of tree depth, has numerous theoretical applications [78]. For an example of tree depth computation, we show in Figure 2.6 how the computational description given in Definition 2.3.21 can be applied to produce a "tree decomposition" of Figure 2.3.

[78]: Nešetřil and Ossona de Mendez (2012), *Sparsity: Graphs, Structures, and Algorithms*

Wrapping up. We have introduced three different ways to allow "database repairs", associated with different scenarios, that we illustrate in Figure 2.7.

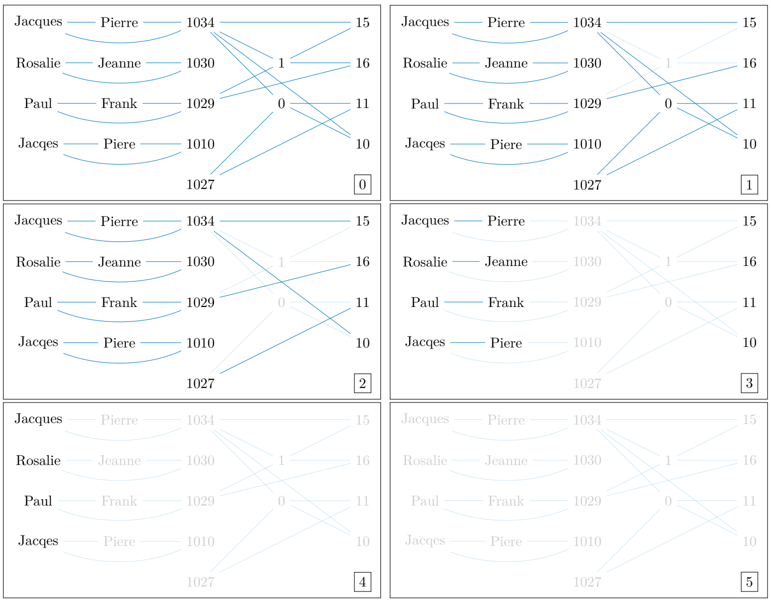


Figure 2.6.: Computation of the tree-depth of Figure 2.3. Every step removes one node per connected component, together with the relation these nodes belongs to.

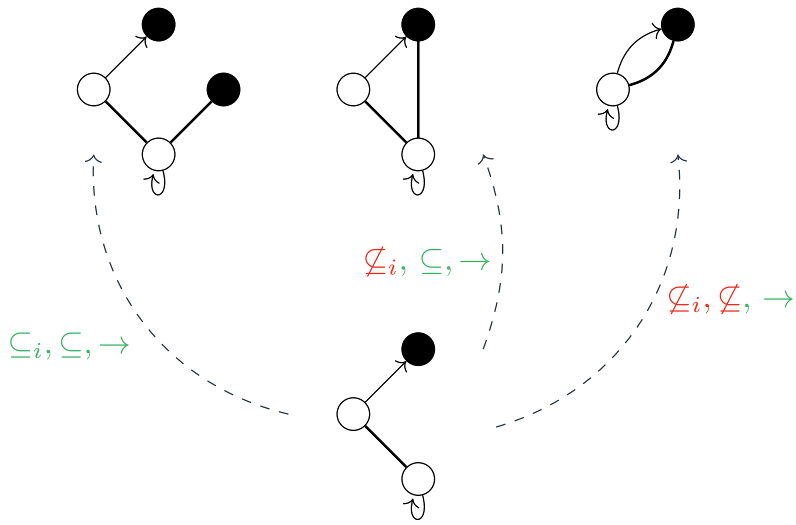


Figure 2.7.: Examples and non examples of completions (with respect to $\subseteq_i, \subseteq$ and $\preceq_h$) of a given model with two binary relations $-$ and $\to$, and one unary relation that colours vertices in black or not.

2.4. Discussion

Certain Answers. The approach of restricting the logic to be well-behaved with respect to the chosen notion of completion is not sound in general. Indeed, we only obtain an under-approximation of the valid answers when using the naïve evaluation, because of the presence of database constraints. Concretely, the running example of the chapter was to study completions (with respect to $\subseteq_i, \subseteq, \preceq_h$) of a given database *such that a constraint was satisfied*. In particular, it can happen that some sentence does not hold on the database, but must become true in every *valid* completion.

This is the setting of certain answers. Given a relational structure $\mathfrak{A}$, a formula $\varphi(\vec{x})$, and a type of morphisms over structures, one can build the set

$$\{ (\mathfrak{v} \colon \vec{x} \to \mathfrak{A}) \colon \forall \mathfrak{B}, \forall h \colon \mathfrak{A} \to \mathfrak{B}, \mathfrak{B}, h \circ \mathfrak{v} \models \varphi(\vec{x}) \} \quad .$$

These are the tuples that certainly satisfy the query, whatever data is missing in the database. By considering certain answers over $\varphi(\vec{x}) \wedge \Sigma$, where Σ is a fixed first order constraint, one obtains the answers over every correct completion of the database.

Note that even though the original query φ belongs to a fragment F, the new query $\varphi \wedge \Sigma$ will not *a priori*.

Chase-like Algorithms. One of the known algorithms that can be used to compute certain answers is known as *the chase algorithm*. Assuming that the constraints are of the form $\Sigma \stackrel{\text{def}}{=} \forall \vec{x}. \exists \vec{y}. \psi(\vec{x}) \Rightarrow \theta(\vec{xy})$, where both ψ and θ are in PQF, the (pseudo)-algorithm proceeds as follows:

1. Check if the current database $\mathfrak{A}$ satisfies Σ.
2. If not, there exists $\vec{a}$ in the database such that $\mathfrak{A}, \vec{x} \mapsto \vec{a} \models \forall \vec{y}. \psi(\vec{x}) \wedge \neg\theta(\vec{xy})$.
3. We update the model $\mathfrak{A}$ by merging it with a model of $\psi(\vec{x}) \wedge \theta(\vec{xy})$.
4. Go back to step 1.

Of course, this algorithm might not terminate, and there is a wide variability on how to perform the different steps. We refer the reader to [25] for a definition of the "core chase" and the notion of "chase-like" algorithm.

[25]: Deutsch, Nash and Remmel (2008), 'The chase revisited'

Completeness of the Fragments. A natural question is whether the provided fragments, which are guaranteed to be safely evaluated in the absence of database constraints are the "largest possible ones." This will be answered in the next chapter.

References

[2] Serge Abiteboul, Richard Hull and Victor Vianu. *Foundations of Databases*. Addison-Wesley, 1995 (cit. on p. 24).

[13] Donald D. Chamberlin and Raymond F. Boyce. 'SEQUEL: A Structured English Query Language'. In: (1974), pp. 249–264. DOI: 10.1145/800296.811515. URL: https://doi.org/10.1145/800296.811515 (cit. on p. 12).

[14] Ashok K. Chandra and Philip M. Merlin. 'Optimal implementation of conjunctive queries in relational databases'. In: *Proceedings of STOC'77*. 1977, pp. 77–90. DOI: 10.1145/800105.803397 (cit. on p. 23).

[17] Edgar F. Codd. 'A relational model of data for large shared data banks'. In: *FIXME Conference of the ACM* 13 (1970), pp. 377–387. DOI: 10.1145/362384.362685 (cit. on pp. 11, 17).

[25] Alin Deutsch, Alan Nash and Jeffrey B. Remmel. 'The chase revisited'. In: *Proceedings of PODS'08*. 2008, pp. 149–158. DOI: 10.1145/1376916.1376938 (cit. on pp. 27, 33, 35, 234, 235).

[31] Heinz-Dieter Ebbinghaus and Jörg Flum. *Finite Model Theory*. Springer, 1995 (cit. on p. 11).

[32] William Ewald. 'The Emergence of First-Order Logic'. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Metaphysics Research Lab, Stanford University, 2019. URL: https://plato.stanford.edu/archives/spr2019/entries/logic-firstorder-emergence/ (visited on 07/11/2022) (cit. on p. 13).

[43] Kurt Gödel. 'Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme'. In: *Monatshefte für Mathematik und Physik* 38 (1931), pp. 173–198. DOI: 10.1007/BF01700692 (cit. on p. 18).

[54] Erich Grädel, Phokion G Kolaitis, Leonid Libkin, Maarten Marx, Joel Spencer, Moshe Y Vardi, Yde Venema and Scott Weinstein. *Finite Model Theory and its Applications*. Springer, 2007 (cit. on p. 11).

[68] Leonid Libkin. *Elements of finite model theory*. Springer, 2012. DOI: 10.1007/978-3-662-07003-1 (cit. on pp. 11, 68, 239).

[78] Jaroslav Nešetřil and Patrice Ossona de Mendez. *Sparsity: Graphs, Structures, and Algorithms*. Springer, 2012. DOI: 10.1007/978-3-642-27875-4 (cit. on pp. 25, 120).

[79] Oleg Pikhurko and Oleg Verbitsky. 'Descriptive complexity of finite structures: Saving the quantifier rank'. In: *The Journal of Symbolic Logic* 70 (2005), pp. 419–450 (cit. on p. 24).

[83] Benjamin Rossman. 'Homomorphism preservation theorems'. In: *Journal of the ACM* 55 (2008), 15:1–15:53. DOI: 10.1145/1379759.1379763 (cit. on pp. i, 24, 35, 37, 38, 40, 42, 55, 74, 175, 177, 235).

[91] Boris Avraamovitch Trakhtenbrot. 'The impossibility of an algorithm for the decision problem for finite classes'. In: *Doklady Akademii Nauk SSSR. Novaya Seriya* 70 (1950), pp. 569–572. DOI: 10.1090/trans2/023 (cit. on p. 18).

[92] Victor Vianu. 'Databases and finite-model theory'. In: *Proceedings of a DIMACS Workshop on Descriptive Complexity and Finite Models*. Ed. by Neil Immerman and Phokion G. Kolaitis. Vol. 31. AMS, 1997, pp. 97–148 (cit. on p. 12).

Preservation Theorems for First Order Queries | 3.

Outline of the chapter

We leverage the language of Chapter 2 (From Databases To First Order Logic and Back) to introduce the main topic of this thesis and its connection with practical problems.

Goals of the chapter

At the end of this chapter, I hope you will be convinced that the relativisation of preservation theorems to classes of finite structures (actual databases) is a non-trivial problem that is worth studying.

Chapter 2 was devoted to the introduction of first order formulas and their semantics both in the finite and infinite case. We finished the chapter by matching fragments of $\mathsf{FO}[\sigma]$ with different notions of completions.[1] While these matches were justified by a two-way relationship,[2] a question remains: were the selected fragments "the largest ones" for which a naïve evaluation procedure would yield correct answers in the incomplete setting?

The typical answer to this last question takes the form of a "preservation theorem" that asserts the maximal expressiveness of a given fragment with respect to a notion of incomplete database. Historically, preservation theorems where studied during the 1950s by mathematicians in the field of Model Theory. We provide in Table 3.1 a list of the three preservation theorems corresponding to our previous notions of incompleteness, carefully selected[3] from [15, Section 5.2]. As an example, Table 3.1 asserts that every first order sentence that whose naïve evaluation is compatible with embeddings can be rewritten as an existential sentence.

1: Namely, the notions of completion associated with induced substructures, substructures, and homomorphisms.

2: One way, every sentence in the given fragment can be evaluated naïvely. The other way, the fact of being a completion of a given model is expressible in the given fragment, provided that the model is finite.

3: We omitted lesser known variants such as "dual Lyndon" or "strong onto homomorphism" preservation theorems for the purpose of a pedagogical introduction.

[15]: Chang and Keisler (1990), *Model Theory*

Organisation of the chapter. First, in Section 3.1, we will precisely define what a "preservation theorem" is beyond the list of examples given in Table 3.1. This precise definition will actually take the form of a generic result in Theorem 3.1.9.

Then, in Section 3.2, we will explore the relativisation of preservation theorems to classes of structures $\mathcal{C} \subseteq \mathsf{Struct}(\sigma)$, with a particular attention for the class $\mathsf{Fin}(\sigma)$ of finite structures. We will see how the

Table 3.1.: Three classical preservation theorems. For a given quasi-ordering $\leq$ and fragment F, a preservation theorem reads as follows: every first order sentence φ that is preserved under $\leq$ is equivalent to a first order sentence in the fragment F.

Preservation Theorem	Quasi-Ordering $\leq$	Fragment F
Homomorphism Preservation Theorem	$\preceq_h$	EPFO
Tarski-Lyndon	$\subseteq$	$\mathsf{EPFO}^{\neq}$
Łoś-Tarski	$\subseteq_i$	EFO

tools of classical Model Theory cannot apply in the finite, and explore a (tiny portion of) the tool set of Finite Model Theory.

Finally, in Section 3.3, we will bridge the notion of preservation theorem to more combinatorial and topological areas of computer science, namely the theory of well-quasi-orderings, and the notion of topological compactness.

3.1. Classical Preservation Theorems

The goal of this section is to move from a simple list of examples (Table 3.1) to an actual definition of a "preservation theorem," which is the central notion of this document.

In Table 3.1, the column considering a fragment of first-order logic is well understood.[4] However, it is unclear how to generically define "well-behaved" quasi orderings on $\mathsf{Struct}(\sigma)$. While those quasi-orderings were introduced in Chapter 2 (From Databases To First Order Logic and Back) in terms of morphisms of structures in a very simple and direct fashion, they actually come from a generic family of "embeddings" associated with a given fragment.

4: Recall that a fragment of first-order logic as a set $\mathsf{F} \subseteq \mathsf{FO}$, stable under finite disjunctions, finite conjunctions and renaming of free variables.

Definition 3.1.1. *Let $\mathfrak{A}, \mathfrak{B} \in \mathsf{Struct}(\sigma)$ be two structures, and F be a fragment of FO. An F-embedding $f\colon \mathfrak{A} \to_{\mathsf{F}} \mathfrak{B}$ is a map $f\colon \mathrm{dom}(\mathfrak{A}) \to \mathrm{dom}(\mathfrak{B})$ such that for every formula $\varphi(\vec{x}) \in \mathsf{F}$, every valuation $\nu\colon \vec{x} \to \mathfrak{A}$, $\mathfrak{A}, \nu \models \varphi(\vec{x})$ implies $\mathfrak{B}, f \circ \nu \models \varphi(\vec{x})$.*

Remark that the notion of QF-embedding can actually be replaced by the one of EFO-embedding, because the maps that respects QF also respects $\exists$QF by construction, and therefore respect EFO.

Example 3.1.2. A QF-*embedding* f between $\mathfrak{A}$ and $\mathfrak{B}$ is an injective map such that for every relation symbol $R \in \sigma$, for every $\vec{a} \in \mathfrak{A}^{\mathsf{arity}(R)}$, $\vec{a} \in R^{\mathfrak{A}}$ if and only if $f(\vec{a}) \in R^{\mathfrak{B}}$.

That is, a QF-*embedding* is a map witnessing that $\mathfrak{A}$ is an induced substructure of $\mathfrak{B}$.

Table 3.2.: Correspondence between fragments and usual quasi-orderings.

$\exists\mathsf{F}$	$\leq_{\mathsf{F}}$	**Order**
EFO	$\leq_{\mathsf{QF}}$	$\subseteq_i$
$\mathsf{EPFO}^{\neq}$	$\leq_{\mathsf{PQF}^{\neq}}$	$\subseteq$
EPFO	$\leq_{\mathsf{PQF}}$	$\preceq_h$
FO	$\leq_{\mathsf{FO}}$	$\to_e$

The introduction of F-embeddings naturally leads to a quasi-order over $\mathsf{Struct}(\sigma)$ by stating that $\mathfrak{A} \leq_{\mathsf{F}} \mathfrak{B}$ holds if and only if there exists an F-embedding from $\mathfrak{A}$ to $\mathfrak{B}$. Following Example 3.1.2, one notices that we actually have a clear correspondence between fragments $\{\mathsf{EFO}, \mathsf{EPFO}^{\neq}, \mathsf{EPFO}\}$ and their associated quasi-orderings in Table 3.1, which is witnessed by Table 3.2. We took the liberty of adding to this presentation the largest fragment, that is $\mathsf{FO}[\sigma]$ itself. This is meaningful because it demonstrates how the classical notion of "elementary embedding" naturally arises in this setting.

The preservation theorems listed in Table 3.1 are somehow incomparable with the "guarantees" that were given in Facts 2.3.10 and 2.3.14 and Definition 2.3.18, and asserted that a finite incomplete database was definable in these fragments. Let us restate, to be exhaustive, what this former completeness statement looked like.

Definition 3.1.3. Let σ be a finite relational signature, F be a fragment of $\mathsf{FO}[\sigma]$, and $\mathfrak{A} \in \mathsf{Struct}(\sigma)$. A *diagram sentence* for $\mathfrak{A}$ is a first order sentence $\Delta^{\mathsf{F}}_{\mathfrak{A}} \in \exists\mathsf{F}$ such that for all $\mathfrak{B} \in \mathsf{Struct}(\sigma)$

$$\mathfrak{B} \models \Delta^{\mathsf{F}}_{\mathfrak{A}} \text{ if and only if } \mathfrak{A} \leq_{\mathsf{F}} \mathfrak{B} \quad .$$

We will refine the notion of diagram sentence to consider free variables (hence the name, diagram "formula"). This other definition will only be used once in this manuscript, namely in the proof of Theorem 6.3.46.

Definition 3.1.4. Let σ be a finite relational signature, F be a fragment of $\mathsf{FO}[\sigma]$, and $\mathfrak{A} \in \mathsf{Struct}(\sigma)$. A *diagram formula* for $\mathfrak{A}$ is a first order formula $\mathring{\Delta}^{\mathsf{F}}_{\mathfrak{A}}((x_a)_{a\in\mathfrak{A}}) \in \mathsf{F}$ such that for all $\mathfrak{B} \in \mathsf{Struct}(\sigma)$

$$\mathfrak{B}, \nu \models \mathring{\Delta}^{\mathsf{F}}_{\mathfrak{A}} \text{ if and only if } (\nu \circ (a \mapsto x_a)) \colon \mathfrak{A} \to_{\mathsf{F}} \mathfrak{B} \quad .$$

Lemma 3.1.5. *Let* $\mathsf{F} \in \{ \mathsf{QF}, \mathsf{PQF}^{\neq}, \mathsf{PQF} \}$. *For every* $\mathfrak{A} \in \mathsf{Fin}(\sigma)$, *there exists a corresponding diagram sentence* $\Delta^{\exists\mathsf{F}}_{\mathfrak{A}} \in \exists\mathsf{F}$, *and diagram formula* $\mathring{\Delta}^{\mathsf{F}}_{\mathfrak{A}} \in \mathsf{F}$.

Proof. Combine Facts 2.3.10 and 2.3.14 and Definition 2.3.18. ■

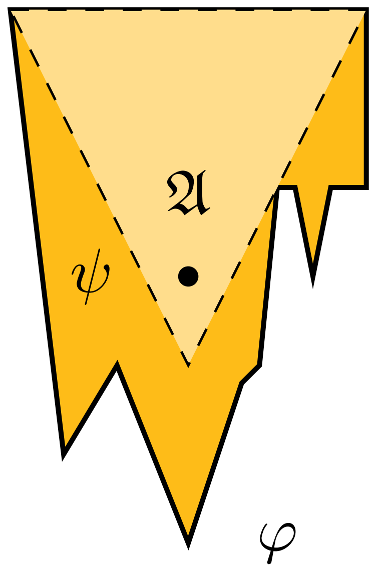


Figure 3.1.: A model $\mathfrak{A}$ of a first order sentence φ that is preserved under $\leq_{\mathsf{F}}$, and a sentence $\psi \in \exists\mathsf{F}$ in pale yellow that contains $\mathfrak{A}$ and is contained in φ.

We are going to provide a generic proof of the preservation theorems listed in Table 3.1 by showing that for every fragment F of first-order logic, sentences preserved under $\leq_{\mathsf{F}}$ over $\mathsf{Struct}(\sigma)$ are equivalent to sentences in $\exists\mathsf{F}$.

To that end, our first step is to provide a weak separation result: if a structure $\mathfrak{A}$ satisfies a first order sentence φ that is preserved under $\leq_{\mathsf{F}}$, there exists a sentence $\psi \in \exists\mathsf{F}$ that lies "between" $\mathfrak{A}$ and φ, which is depicted in Figure 3.1.

Up to this point, none of the fundamental theorems of first order logic were used. It is time to introduce one of the main tools of the classical Model Theory: the compactness theorem of first order logic. Unsurprisingly, the statement of Theorem 3.1.7 contains references to "theories," which we swiftly define hereafter.

Definition 3.1.6. A *first order theory* T is a collection of first order sentences. A structure $\mathfrak{A}$ satisfies a theory T, written $\mathfrak{A} \models \mathrm{T}$ whenever $\mathfrak{A} \models \varphi$ for all $\varphi \in \mathrm{T}$.

One concludes from Definition 3.1.6 that, semantically, a first order theory is an arbitrary conjunction of first order sentences. In general, such an arbitrary conjunction is not a first order sentence, and the following theorem states that, even though it is not, it behaves as a "limit" of sentences.

We did not define what a constant symbol is: it is a symbol s, and every structure $\mathfrak{A}$ should associate a unique element of the domain to the symbol s. In particular, this symbol can be leveraged in the logic, for instance stating that $\exists x.E(x, s)$.

Theorem 3.1.7 [15, Corollary 1.2.12]**.** Let σ be a (potentially infinite, and potentially with constant symbols) relational signature. The following properties are equivalent for a theory T:

1. There exists $\mathfrak{A} \in \mathsf{Struct}(\sigma)$ such that $\mathfrak{A} \models \mathrm{T}$.
2. For all $\mathrm{T}' \subseteq_{\mathsf{fin}} \mathrm{T}$, there exists $\mathfrak{A} \in \mathsf{Struct}(\sigma)$ such that $\mathfrak{A} \models \mathrm{T}'$.

This is called the *compactness theorem of first order logic*.

An equivalent formulation of the compactness theorem of first order logic can be given in terms of *consistency*. A theory T is said to be *consistent* whenever there exists $\mathfrak{A} \in \mathsf{Struct}(\sigma)$ such that $\mathfrak{A} \models \mathrm{T}$. In this setting, the compactness theorem of first order logic can be restated as: for all theory T that is not consistent, there exists a finite subset $\mathrm{T}' \subseteq_{\mathsf{fin}} \mathrm{T}$ that is not consistent.

Leveraging this theorem, we can prove our separation result, that is a simple generalisation of the first step in the proofs classical preservation theorems [15, Section 5.2].

We state the lemma as "foreign" even though we could not find a reference providing the result with this level of generality, probably because it is folklore.

Lemma 3.1.8. Let F be a fragment of FO, φ be a sentence preserved under F-embeddings, and $\mathfrak{A}$ be a countable model of φ. There exists a sentence $\psi \in \exists\mathsf{F}$ such that $\mathfrak{A} \models \psi$ and $\psi \models \varphi$.

Proof. This requires the exceptional addition of constants, which will not appear in the rest of this manuscript. Namely, we add a constant c_a for all $a \in \mathrm{dom}(\mathfrak{A})$. In this extended signature σ' defined as σ plus the constants $(c_a)_{a\in\mathrm{dom}(\mathfrak{A})}$, one can build the following theory:

$$\mathrm{T} \stackrel{\mathsf{def}}{=} \{\, \psi[x \mapsto c_{\nu(x)}] \colon \psi(\vec{x}) \in \mathsf{F} \wedge \mathfrak{A}, \nu \models \psi(\vec{x}) \,\}$$

That is, the theory composed of first order sentences over the signature σ' that are obtained by collecting the F formulas, and *naming* the variables using the constants c_a of the extended language. Notice that we transformed a first order formula into a first order sentence in this process using the constants.

Let us now argue that the theory $\mathrm{T} \cup \{\, \neg\varphi \,\}$ is inconsistent. Indeed, whenever $\mathfrak{B} \models \mathrm{T}$, we can build $\hat{\mathfrak{B}}$ for the model $\mathfrak{B}$ without the additional constants. Because $\mathfrak{B} \models \neg\varphi$, which does not contain the additional constants, $\hat{\mathfrak{B}} \models \neg\varphi$. Furthermore, the map h that sends $a \in \mathrm{dom}(\mathfrak{A})$ to the interpretation of the constant c_a in $\mathfrak{B}$ is well-defined from $\mathfrak{A}$ to $\hat{\mathfrak{B}}$, is an F-embedding by construction. As a consequence, $\mathfrak{A} \leq_{\mathsf{F}} \hat{\mathfrak{B}}$ and $\hat{\mathfrak{B}} \models \varphi$. This is absurd.

Using the compactness theorem of first-order logic, there exists a finite subset of $\mathrm{T}' \subseteq_{\mathsf{fin}} \mathrm{T}$ such that $\mathrm{T}' \cup \{\, \neg\varphi \,\}$ is inconsistent. Let us write $\psi \stackrel{\mathsf{def}}{=} \exists\vec{x}. \bigwedge_{\theta\in\mathrm{T}'} \theta[c_a \mapsto x_a]$, i.e., the existential closure of the conjunction of the formulas in T', obtained by transforming back the constants $(c_a)_{a\in\mathrm{dom}(\mathfrak{A})}$ into distinct variables $(x_a)_{a\in\mathrm{dom}(\mathfrak{A})}$.

Notice that $\mathfrak{A} \models \psi$. Furthermore, for all $\mathfrak{B}$ such that $\mathfrak{B} \models \psi$, there exists a valuation ν such that $\mathfrak{B}, \nu \models \bigwedge_{\theta\in T'} \theta[c_a \mapsto x_a]$. In particular, one can assign to the constant c_a the element $\nu(x_a)$ to obtain a structure $\hat{\mathfrak{B}}$ in the language extended with constants, such that $\hat{\mathfrak{B}} \models T'$. Finally, $\hat{\mathfrak{B}} \models \varphi$, which implies that $\mathfrak{B} \models \varphi$ since φ does not use these additional constants in the first place.

We have proven that there exists a sentence $\psi \in \exists\mathsf{F}$ such that $\mathfrak{A} \models \psi$ and $\psi \models \varphi$. ∎

Theorem 3.1.9. Let F be a fragment of FO. The F-*preservation theorem* states that every sentence φ preserved under F-embeddings is equivalent to an $\exists\mathsf{F}$ sentence over $\mathsf{Struct}(\sigma)$.

Beware that Theorem 3.1.9 does not provide any information on the equivalent sentence. In particular, parameters such as the quantifier rank can vastly differ between the original sentence and the equivalent one.

Proof. Let us consider for every countable model $\mathfrak{A}$ of φ a sentence $\psi_{\mathfrak{A}} \in \exists F$ such that $\mathfrak{A} \models \psi_{\mathfrak{A}}$ and $\psi_{\mathfrak{A}} \models \varphi$, as provided by Lemma 3.1.8.

Let us write $T \stackrel{\text{def}}{=} \{ \neg\psi_{\mathfrak{A}} : \mathfrak{A} \models \varphi \}$. It is clear that $T \cup \{ \varphi \}$ is inconsistent, hence there exists a finite subset $T' \subseteq_{\mathsf{fin}} T$ such that $T' \cup \{ \varphi \}$ is inconsistent. Therefore, $\varphi \equiv \bigvee_{\psi_{\mathfrak{A}} \in T'} \psi_{\mathfrak{A}}$. This last first order sentence is in $\exists\mathsf{F}$ because of the equivalence $(\exists \vec{x}.\psi_1(\vec{x})) \vee (\exists \vec{y}.\psi_2(\vec{y})) \equiv_{\mathcal{C}} \exists \vec{x}.\exists \vec{y}.\psi_1(\vec{x}) \vee \psi_2(\vec{y})$ up to α-renaming, and because F is closed under $\vee$. ∎

Notice that Theorem 3.1.9 does not provide a concrete characterisation of the F-embeddings. In order to claim that this result generalises Table 3.1, it remains to prove that the preorders correspond to their respective notions of F-embeddings. This was done for specific instances in Table 3.2.

An interesting example of preservation theorem that may seem out of the reach of Theorem 3.1.9 is the Chang-Łoś-Suszko Theorem. This preservation theorem is also important as an illustration of preservation theorems that have quantifier alternation, and is particularly meaningful in the database setting because the $\forall\exists\mathsf{QF}$ sentences[5] model most of the constraints that one can place on a database [25]. For instance, the constraint in Example 2.2.8 was of this shape.

5: That is, sentences of the form $\forall \vec{x}.\exists \vec{y}.\psi(\vec{x}\vec{y})$ where $\psi \in \mathsf{QF}$.

We will first formally state the Chang-Łoś-Suszko theorem, and then explain how a simple reformulation puts it on the tracks of Theorem 3.1.9. The Chang-Łoś-Suszko theorem considers "unions of chains" of models, that need to be carefully formalised.[6]

6: In particular, one cannot simply use the relation $\subseteq_i$, as it does not put any constraint on the domain of the relational structures.

Definition 3.1.10 ([15, p. 140]). *Let σ be a relational signature, α be an ordinal. A chain of models of length α is a family $(\mathfrak{A}_\beta)_{\beta<\alpha}$ of relational structures of $\mathsf{Struct}(\sigma)$, such that $\mathrm{dom}(\mathfrak{A}_\beta) \subseteq \mathrm{dom}(\mathfrak{A}_\gamma)$, and $R^{\mathfrak{A}_\beta} \subseteq R^{\mathfrak{A}_\gamma}$ for all $\gamma \leq \beta < \alpha$ and for every relation $R \in \sigma$.*

The union of a chain, written $\bigcup_{\beta<\alpha} \mathfrak{A}_\beta$ is the relational structure with domain the union $\bigcup_{\beta<\alpha} \mathrm{dom}(\mathfrak{A}_\beta)$, and relations $R^{\bigcup_{\beta<\alpha} \mathfrak{A}_\beta}$ defined as the union $\bigcup_{\beta<\alpha} R^{\mathfrak{A}_\beta}$.

The knowledgable reader might notice that sentences preserved under unions of chains are precisely those defining closed subsets in the d-topology associated with the ordering $\mathfrak{A} \leq \mathfrak{B}$ whenever $\mathrm{dom}(\mathfrak{A}) \subseteq \mathrm{dom}(\mathfrak{B})$ and $R^{\mathfrak{A}} \subseteq R^{\mathfrak{B}}$ for every relation symbol $R \in \sigma$.

Theorem 3.1.11 [15, Theorem 5.2.6]**.** A sentence ϕ is said to be *preserved under unions of chains* if $\bigcup_{\beta<\alpha} \mathfrak{A}_i \models \phi$ for every chain of models $(\mathfrak{A}_\beta)_{\beta<\alpha}$ of length α such that $\mathfrak{A}_\beta \models \phi$ for every $\beta < \alpha$. The *Chang-Łoś-Suszko* Theorem states that first-order sentences *preserved under unions of chains* are equivalent to $\forall\exists\mathsf{QF}$ sentences.

The reformulation that follows is heavily inspired by the results of [86], but we restate the results together with its proof using the newly introduced language to keep this document self-contained.

[86]: Sankaran, Adsul, Madan, Kamath and Chakraborty (2012), 'Preservation under substructures modulo bounded cores'

We will have another viewpoint in the conclusion of the thesis, were we will be interested in the ability to talk about free variables. This is meaningful because a sentence preserved under $\leq_{\forall\mathsf{QF}}$ corresponds to a preservation under $\leq_{\mathsf{QF}}$ over "pointed structures."

Lemma 3.1.12 [86, Theorem 2]**.** Let $\varphi \in \mathsf{FO}$ be a first order sentence. The following are equivalent:

1. φ is preserved under unions of chains,
2. $\neg\varphi$ is preserved under $\leq_{\forall\mathsf{QF}}$.

Proof. Let us first check that if $\varphi \in \forall\exists\mathsf{QF}$, then it is preserved under unions of chains. For that, consider a chain of models $(\mathfrak{A}_\beta)_{\beta<\alpha}$ for some ordinal α, such that $\mathfrak{A}_\beta \models \varphi$ for every $\beta < \alpha$. Let us define $\mathfrak{A} \stackrel{\text{def}}{=} \bigcup_{\beta<\alpha} \mathfrak{A}_\beta$, and prove that $\mathfrak{A} \models \varphi$. By assumption, $\varphi = \forall x_1 \dots x_k, \exists y_1, \dots, y_l, \theta(\vec{x}, \vec{y})$ where θ is quantifier-free.

Let $a_1, \dots, a_k \in \mathfrak{A}$, by definition of $\mathfrak{A}$ there exists $i_1, \dots, i_k < \alpha$ such that $a_1 \in \mathfrak{A}_{i_1}, \dots, a_k \in \mathfrak{A}_{i_k}$. In particular, all the elements belong to $\mathfrak{A}_i$, where $i \stackrel{\text{def}}{=} \max\{ i_1, \dots, i_k \}$. Now, $\mathfrak{A}_i \models \varphi$, hence there exists $b_1, \dots, b_l \in \mathfrak{A}_i$ such that $\mathfrak{A}_i, (x_j \mapsto a_j, y_j \mapsto b_j) \models \theta(\vec{x}, \vec{y})$. Because θ is quantifier-free and $\mathfrak{A}_i \subseteq_i \mathfrak{A}$, we conclude that $\mathfrak{A}, (x_j \mapsto a_j, y_j \mapsto b_j) \models \theta(\vec{x}, \vec{y})$, and we have proven that for all $\vec{a} \in \mathfrak{A}^k$, there exists $\vec{b} \in \mathfrak{A}^l$ such that θ is satisfied, that is, $\mathfrak{A} \models \varphi$.

Let φ be preserved under unions of chains. Thanks to the Chang-Łoś-Suszko theorem, there exists $\psi \in \forall\exists\mathsf{QF}$ such that $\varphi \equiv_{\mathsf{Struct}(\sigma)} \psi$. Therefore, $\neg\varphi \equiv_{\mathsf{Struct}(\sigma)} \neg\psi \in \exists\forall\mathsf{QF}$. This entails that $\neg\varphi$ is preserved under $\leq_{\forall\mathsf{QF}}$.

Conversely, let φ be a sentence such that $\neg\varphi$ is preserved under for $\leq_{\forall\mathsf{QF}}$. Leveraging Theorem 3.1.9, there exists $\psi \in \exists\forall\mathsf{QF}$ such that $\neg\varphi \equiv_{\mathsf{Struct}(\sigma)} \psi$. As a consequence, $\varphi \equiv_{\mathsf{Struct}(\sigma)} \neg\psi \in \forall\exists\mathsf{QF}$, which as mentioned in the beginning of the proof, is preserved under unions of chains. ∎

This nice presentation of preservation theorems hides two key points in database theory: the first one (that will not be considered in this manuscript) is that queries have parameters, the second one is that databases are finite structures, and therefore we need for the results to hold on the class of finite relational structures (or subclasses thereof).

3.2. Preservation Theorems do not Relativise

Because this section studies the relativisation of preservation theorems, let us first define what "relativisation" means in this context. Let

$\mathcal{C} \subseteq \mathsf{Struct}(\sigma)$ be a class of structures. The F-preservation theorem *relativises* to $\mathcal{C}$ when, for every sentence φ that is preserved under $\leq_{\mathsf{F}}$ when restricted to relational structures in $\mathcal{C}$, there exists a sentence $\psi \in \exists\mathsf{F}$ such that ψ is equivalent to ψ when restricted to $\mathcal{C}$, i.e., $\varphi \equiv_{\mathcal{C}} \psi$.

One might think that, because $\mathcal{C}$ contains fewer models than $\mathsf{Struct}(\sigma)$, preservation theorems always relativise. However, it is not the case for most of the fragments introduced so far, as witnessed in Table 3.3. As a matter of fact, understanding to what extent preservation theorems can be leveraged in the finite is a complex problem [81, 82], that remains an active field of study [22].

[81]: Rosen (1995), 'Finite model theory and finite variable logics'
[82]: Rosen (2002), 'Some aspects of model theory and finite structures'

This study is partially motivated by the implications of such results. For instance, the termination of the Chase has deep connection with the Homomorphism Preservation Theorem [25].

[25]: Deutsch, Nash and Remmel (2008), 'The chase revisited'

Name	**Relativises to $\mathsf{Fin}(\sigma)$**
H.P.T.	yes [83]
Tarski-Lyndon	no [5, Theorem 10.2]
Łoś-Tarski	no [22, 56, 90]
dual Lyndon	no [4, 89]
strong onto homomorphism	no [12]
Chang-Łoś-Suszko	no [86]

Table 3.3.: Classical preservation theorems and their relativisations to the finite case. We have taken the liberty of adding other classical preservation theorems to the list to better convey the feeling that the Homomorphism Preservation Theorem is an outlier.

Finite Model Theory strikes back. Let us discuss the negative results of Table 3.3 in the light of a simple example. Given a quasi ordering $\leq$ over a set X and a subset $E \subseteq \mathsf{X}$, we define $\downarrow_{\leq} E$ as the set $\{ m \in \mathsf{X} \colon \exists m' \in E, m \leq m' \}$, this is the *downward closure* of E in X with respect to $\leq$.

We are now going to introduce the main source of counterexamples in this manuscript: finite cycles. Let $\sigma \overset{\mathsf{def}}{=} \{ (E, 2) \}$, we define Cycles to be the class of finite undirected cycles in $\mathsf{Fin}(\sigma)$. Furthermore, we write C_n for the finite cycle with n vertices. Notice that $C_i \not\subseteq_i C_j$ whenever $i \neq j \geq 3$.

The following example is developed using rudimentary tools on purpose. We will have finer and finer explanation for the failure of the EFO-preservation theorem in this manuscript, but it is always useful to have a suboptimal proof using only basic knowledge.

Example 3.2.1. The Łoś-Tarski Theorem does not relativise to $\downarrow_{\subseteq_i}\mathsf{Cycles}$.

Proof. The sentence φ stating that every vertex is connected to at least 2 other vertices is expressible in $\mathsf{FO}[\sigma]$ as follows:

$$\varphi \overset{\mathsf{def}}{=} \forall x, \exists y_1 y_2, E(x, y_1) \wedge E(x, y_2) \wedge y_1 \neq y_2 \quad .$$

Observe that if $\mathfrak{A} \subseteq_i \mathfrak{B}$, $\mathfrak{A} \in \mathsf{Cycles}$ and $\mathfrak{B} \in \mathsf{Cycles}$, then $\mathfrak{A} = \mathfrak{B}$. Furthermore, if $\mathfrak{A} \models \varphi$ and $\mathfrak{A} \in \downarrow_{\subseteq_i}\mathsf{Cycles}$, then $\mathfrak{A} \in \mathsf{Cycles}$.

As a consequence, whenever $\mathfrak{B} \subseteq_i \mathfrak{A}$, $\mathfrak{B} \models \varphi$, and $\mathfrak{A} \in \downarrow_{\subseteq_i}\mathsf{Cycles}$, we have $\mathfrak{B} = \mathfrak{A}$, hence $\mathfrak{A} \models \varphi$. The sentence φ is therefore preserved under $\subseteq_i$ when restricted to the class $\downarrow_{\subseteq_i}\mathsf{Cycles}$.

Assume by contradiction that there exists an existential sentence ψ such that $\psi \equiv_{\downarrow_{\subseteq_i}\mathsf{Cycles}} \varphi$. Without loss of generality, one can assume that $\psi = \exists \vec{x}.\theta(\vec{x})$ where $\theta(\vec{x}) \in \mathsf{QF}$.

If $\mathfrak{A} \in \downarrow_{\subseteq_i}\mathsf{Cycles}$ and $\mathfrak{A} \models \psi$, then there exists a substructure $\mathfrak{A}' \subseteq_i \mathfrak{A}$ of size at most $|\vec{x}|$ that satisfies ψ: it suffices to restrict the domain of $\mathfrak{A}$ to the witnesses of the existential quantifiers of ψ. Remark that $\mathfrak{A}'$ remains in the same class $\downarrow_{\subseteq_i}\mathsf{Cycles}$.

As a consequence, the $\subseteq_i$-minimal models of ψ are of bounded size in $\downarrow_{\subseteq_i}\mathsf{Cycles}$, hence there can only be finitely many of them. This is absurd, because C_n is a $\subseteq_i$-minimal model of φ for all $n \in \mathbb{N}$. ■

While Example 3.2.1 is convincing, it remains unclear why the proof of Theorem 3.1.9 does not work in the finite. As per usual in computer science, the breaking change does not occur in the displayed proof, but in one of its dependencies: the compactness theorem of first order logic.

For a short illustration of the failure of the compactness theorem in the case of $\mathcal{C} \stackrel{\text{def}}{=} \downarrow_{\subseteq_i}\mathsf{Cycles}$, it suffices to consider the sentence ψ_n stating that there exists more than n elements. The theory $\{\, \psi_n : n \in \mathbb{N} \,\}$ has no model in $\mathcal{C}$,[7] but every finite subset has a model in $\mathcal{C}$.

7: And actually in $\mathsf{Fin}(\sigma)$.

Because we lost the main tool that was used twice to prove Theorem 3.1.9, one might now be surprised (but in the other direction this time) that the Homomorphism Preservation Theorem relativises to $\mathsf{Fin}(\sigma)$ in Table 3.3.

A new hope. There are numerous classes of structures for which preservation theorems relativise. And a first simple example is to consider classes $\mathcal{C} \subseteq_{\mathsf{fin}} \mathsf{Fin}(\sigma)$, i.e., finite classes of finite structures. Let us sketch the proof that preservation theorems naturally relativise to such classes.

A "simple cycle" is also called a chordless cycle.

Fact 3.2.2. Let σ be a finite relational signature, and $\mathcal{C} \subseteq_{\mathsf{fin}} \mathsf{Fin}(\sigma)$ be a finite set of finite structures. Then, the Łoś-Tarski Theorem relativises to $\mathcal{C}$.

Proof. Let $\varphi \in \mathsf{FO}[\sigma]$ be preserved under $\subseteq_i$ over $\mathcal{C}$. Let S be the set of models $\mathfrak{A} \in \mathcal{C}$ such that $\mathfrak{A} \models \varphi$.

Because $\mathcal{C}$ is finite, the sentence $\psi \stackrel{\text{def}}{=} \bigvee_{\mathfrak{A} \in S} \Delta^{\mathsf{EFO}}_{\mathfrak{A}}$ is an existential sentence. Furthermore, if $\mathfrak{B} \models \varphi$, then $\mathfrak{B} \in S$, hence $\mathfrak{B} \models \psi$.

Conversely, if $\mathfrak{B} \models \psi$, there exists $\mathfrak{A} \in S$ such that $\mathfrak{B} \models \Delta^{\mathsf{EFO}}_{\mathfrak{A}}$. As a consequence, $\mathfrak{A} \subseteq_i \mathfrak{B}$, and $\mathfrak{A} \models \varphi$, which implies that $\mathfrak{B} \models \varphi$ since φ is preserved under $\subseteq_i$. ■

The immediate consequence of Fact 3.2.2 is that whether a preservation theorem relativises or not to a given class $\mathcal{C}$ is a non-monotone property: preservation theorems are true for $\mathsf{Struct}(\sigma)$, fail to relativise $\mathsf{Fin}(\sigma)$ most of the time, and do relativise for finite classes of finite structures (at least for $\mathsf{F} \in \{\, \mathsf{EFO}, \mathsf{EPFO}^{\neq}, \mathsf{EPFO} \,\}$, where diagram sentences exist).

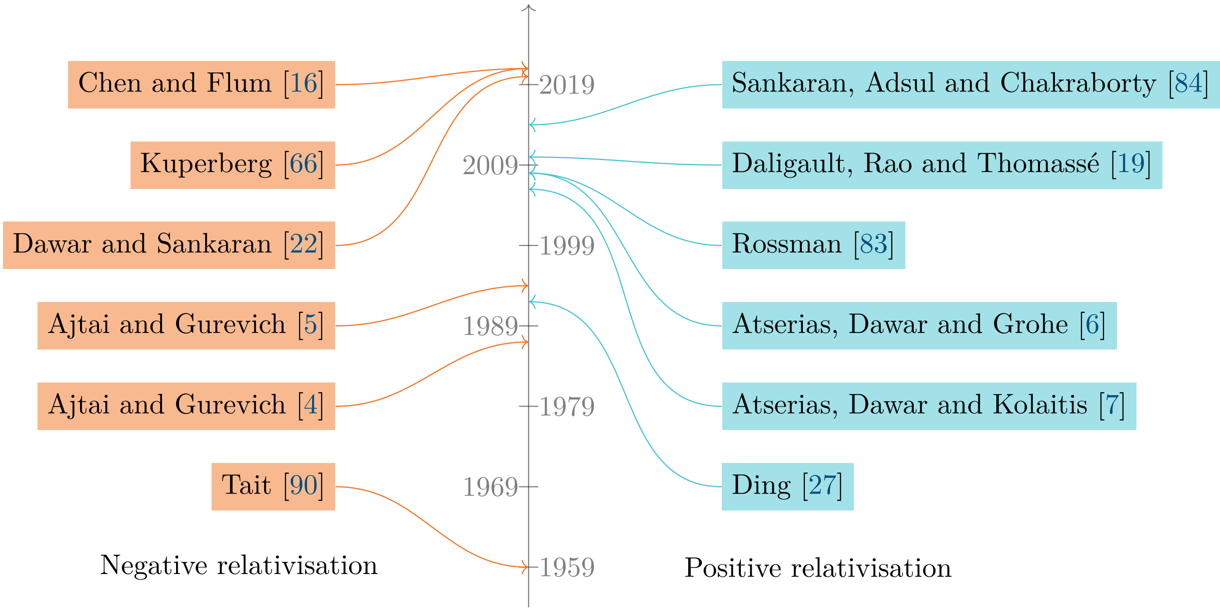


Figure 3.2.: A short history of preservation results considering $\mathsf{FO}[\sigma]$. Negative results are placed on the left-hand side of the timeline, while positive results are placed on its right-hand side.

Therefore, the landscape of relativisation will somehow look like an onion, with layers of classes trying to balance the two opposing variances in the statement of a preservation theorem: the more models you get, the less likely you are to be preserved under a given preorder, but the more complicated it is to produce an equivalent sentence. On that scale, preservation theorems over $\mathsf{Struct}(\sigma)$ hold because "it is very hard to be preserved under $\leq_{\mathsf{F}}$," while preservation theorems will relativise to finite classes of finite structures because "it is very easy to find equivalent sentences in $\exists\mathsf{F}$."

A duel of fates. The relativisation problem is non-trivial, and there has been a vast literature dedicated to understand both positive and negative results. We provide a brief and incomplete history of this line of research in Figure 3.2. It is quite interesting to notice that the authors of negative and positive examples are mostly disjoint, which can be explained by the difference in the required tool sets.

We will not focus too much on the details of Figure 3.2, because the specific statements will be introduced in a call-by-need fashion in the rest of the document.

Typical proof of non-relativisation. Let us go back to our first "simple proof" of non-relativisation that was done in Example 3.2.1. This is a prototypical example that will help us introduce the notions leveraged in the literature.

A key ingredient in Example 3.2.1 was the notion of "minimal model" of a given sentence φ over a class $\mathcal{C}$, that is, the models $\mathfrak{A} \in \mathcal{C}$ such that $\mathfrak{A} \models \varphi$ and such that for all $\mathfrak{B} \leq_{\mathsf{F}} \mathfrak{A}$, if $\mathfrak{B} \models \varphi$, then $\mathfrak{A} \leq_{\mathsf{F}} \mathfrak{B}$. We say that such a model is a F-*minimal model* of φ.[8]

8: It is a minimal element for $\leq_{\mathsf{F}}$ over $\mathcal{C}$.

Typically, sentences in $\exists\mathsf{F}$ will be described using finitely many of those F-minimal models, which provides a semantic counterpart to the otherwise syntactic presentation of the fragments.

To have a better grasp on the objects at hand, let us introduce the notation $[\![\varphi]\!]_{\mathcal{C}}$ to represent the collection of $\mathfrak{A} \in \mathcal{C}$ such that $\mathfrak{A} \models \varphi$. Using this notation, the set of $\leq_{\mathsf{F}}$-minimal models is simply $\min_{\leq_{\mathsf{F}}}[\![\varphi]\!]_{\mathcal{C}}$.[9]. Let us also say that a subset $S \subseteq \mathcal{C}$ is *first order definable* whenever there exists a sentence $\varphi \in \mathsf{FO}[\sigma]$ such that $S = [\![\varphi]\!]_{\mathcal{C}}$.

9: Beware that this set could be empty even though $[\![\varphi]\!]_{\mathcal{C}}$ is not, i.e., φ may not have minimal elements. This will however not happen for reasonable quasi-orders on $\mathsf{Fin}(\sigma)$.

A folklore result that is leveraged throughout both the positive and negative relativisation proofs is the following semantic characterisation.

Lemma 3.2.3 (folklore)**.** Let σ be a finite relational signature, $\mathsf{F} \in \{\mathsf{EFO}, \mathsf{EPFO}^{\neq}, \mathsf{EPFO}\}$, and $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ such that $\mathcal{C} = \downarrow_{\leq_{\mathsf{F}}}\mathcal{C}$. Then, for all $\varphi \in \mathsf{FO}[\sigma]$ the following are equivalent:

1. There exists $\psi \in \exists\mathsf{F}$ such that $[\![\psi]\!]_{\mathcal{C}} = [\![\varphi]\!]_{\mathcal{C}}$,
2. The set $\min_{\leq_{\mathsf{F}}}[\![\varphi]\!]_{\mathcal{C}}$ of $\leq_{\mathsf{F}}$-minimal models for φ is finite up to $\leq_{\mathsf{F}}$-equivalence.

Furthermore, Item 2 implies Item 1 even when $\mathcal{C}$ is not downwards closed.

Item 1 is another way of stating the equivalence $\varphi \equiv_{\mathcal{C}} \psi$.

It is necessary to consider finiteness up to the equivalence relation in Item 2, as the fragment $\mathsf{F} \stackrel{\text{def}}{=} \{\top, \bot\}$ will immediately illustrate.

In the light of Lemma 3.2.3, a typical proof of non-relativisation starts by considering a class $\mathcal{C}$ of finite structures that is *downwards closed*, i.e., such that $\mathcal{C} = \downarrow_{\leq_{\mathsf{F}}}\mathcal{C}$. This is precisely the reason why $\downarrow_{\subseteq_i}\mathsf{Cycles}$ was considered in Example 3.2.1.

Beware that in Lemma 3.2.3 the hypothesis that $\mathcal{C}$ is downwards closed is necessary for the implication Item 1 $\Rightarrow$ Item 2 to hold. This is illustrated in the following example.

Example 3.2.4. The sentence $\top$ is an existential sentence that has infinitely many non-equivalent $\subseteq_i$-minimal models in Cycles.

Because the orderings $\subseteq_i$, $\subseteq$, and $\preceq_h$ are abundantly used in the literature, specific names for their respective notions of "downwards closed classes." A class $\mathcal{C}$ that is downwards closed for $\preceq_h$ is called *co-homomorphism closed* in [83, Section 7.1.2 on page 50]. A class $\mathcal{C}$ that is downwards closed for $\subseteq$ is called *monotone*. Finally, a class $\mathcal{C}$ that is downwards closed for $\subseteq_i$ is called *hereditary*.

[83]: Rossman (2008), 'Homomorphism preservation theorems'

The next step to disprove relativisation over a given subclass $\mathcal{C}$ is to produce a sentence $\varphi \in \mathsf{FO}[\sigma]$ such that $[\![\varphi]\!]_{\mathcal{C}}$ has infinitely many non-equivalent minimal models, but is preserved under $\leq_{\mathsf{F}}$. In terms of subsets, this means finding a first order sentence φ such that $[\![\varphi]\!]_{\mathcal{C}}$ is *upwards closed*: $\uparrow_{\leq_{\mathsf{F}}}[\![\varphi]\!]_{\mathcal{C}} = [\![\varphi]\!]_{\mathcal{C}}$.[10]

10: We define the *upward closure* of a subset $S \subseteq \mathsf{X}$ with respect to a quasi-order $\leq$ similarly as the downward closure via: $\uparrow_{\leq} S \stackrel{\text{def}}{=} \{ m' \in \mathsf{X} \colon \exists m \in S. m \leq m' \}$.

Concretely, this search focuses on finding large subsets S of structures such that $\forall \mathfrak{A} \neq \mathfrak{A}' \in S$, $\mathfrak{A} \not\leq_{\mathsf{F}} \mathfrak{A}'$, and $\mathfrak{A}' \not\leq_{\mathsf{F}} \mathfrak{A}$. This kind of subsets is commonly known as an *antichain* for $\leq_{\mathsf{F}}$. We will often use the notation $a \perp_{\leq} b$ as a shorthand for $\neg(a \leq b) \wedge \neg(b \leq a)$.

Example 3.2.5. Whenever $i \neq j \geq 3$, $C_i \perp_{\subseteq_i} C_j$.

Let us provide a complete example of this technique by restating the original non-relativisation result of [90] where the different steps are clearly highlighted.

[90]: Tait (1959), 'A counterexample to a conjecture of Scott and Suppes'

Theorem 3.2.6 [90]**.** Let σ be the finite relational signature composed by two binary relation symbols $(R_1, 2)$, $(R_2, 2)$. The Łoś-Tarski Theorem does not relativise to $\mathsf{Fin}(\sigma)$.

Proof Scheme. We start by an informal construction, that you can skip if you don't like informal introductions. Let L_n be a finite total ordering with n elements, and LinOrd be the collection of such linear orderings. We will turn L_n into a cycle[11] for the relation R_2 by defining R_2 as the successor-or-equal relation associated with $\leq$, plus the edge connecting the maximal element of L_n with the least element of L_n.

11: Which is by now a recurrent source of counter examples.

Formally, let us define the FO-interpretation $\mathsf{I}\colon \mathsf{LinOrd} \to \mathsf{Fin}(\sigma)$ with domain $\varphi_{\mathsf{dom}}(x) \stackrel{\text{def}}{=} \top$, $\varphi_{R_1}(x,y) \stackrel{\text{def}}{=} x \leq y$, and $\varphi_{R_2}(x,y) \stackrel{\text{def}}{=} (x = y) \vee (y = x+1) \vee (\min y \wedge \max x)$.[12]

12: We have used $x = y \vee y = x + 1$ as a shorthand notation for the sentence $\forall z.(x \leq z \leq y) \Rightarrow (z = y \vee z = x)$. Similarly, we used $\min x$ to denote $\forall z.x \leq z$, and $\max y$ to denote $\forall z.z \leq y$.

It is immediate that the image $\mathrm{Im}(\mathsf{I})$ of LinOrd is an infinite antichain for $\subseteq_i$. Leveraging Lemma 3.2.3,[13] it suffices to propose a first order sentence $\varphi \in \mathsf{FO}[\sigma]$ such that $\uparrow_{\subseteq_i}\mathrm{Im}(\mathsf{I}) = [\![\varphi]\!]_{\mathsf{Fin}(\sigma)}$ to conclude the desired non-relativisation.[14]

13: Because $\mathsf{Fin}(\sigma)$ is downwards closed for $\subseteq_i$ inside $\mathsf{Struct}(\sigma)$.

14: We write $\uparrow_{\subseteq_i} S$ for the upward closure of S with respect to $\subseteq_i$, but did not specify in which "superset" the upward closure is considered. Here, we implicitly consider the upward closure with respect to $\subseteq_i$ inside the class $\mathsf{Fin}(\sigma)$.

A structure $\mathfrak{A} \in \mathsf{Fin}(\sigma)$ contains $\mathsf{I}(L_n)$ for some n as an induced substructure if and only if there exists $(x, y) \in \mathfrak{A}$, such that R_1 defines a total ordering $\leq$ between x and y, $R_2(y, x)$ holds in $\mathfrak{A}$, and R_2 is the successor-or-equal relation for element between x and y, except for the pair (x, y). All of this is expressible in $\mathsf{FO}[\sigma]$. ■

Theorem 3.2.6 serves as an illustration of a more general proof scheme, that has been instantiated with various levels of technicality: the correspondence between a semantic characterisation and a first order definable subset may be highly non-trivial.[15]

15: This will be the case in Subsection 4.3.1.

Furthermore, the technique can be generalised to not only obtain non-relativisation results, but also obtain undecidability properties [16, 66]. One way to obtain such theorems is to encode a grid-like structure instead of cycles, where the runs of a Turing machine can be encoded. In this setting, deciding whether a sentence φ is equivalent to a sentence ψ in the desired fragment $\exists\mathsf{F}$ amounts to checking the emptiness of $[\![\varphi]\!]_{\mathsf{Fin}(\sigma)}$, which corresponds to the Turing machine not having a finite run. This paragraph is purposely vague, because such techniques will be explained and developed in Chapter 4 (Locality and Preservation).

[16]: Chen and Flum (2021), 'Forbidden Induced Subgraphs and the Łoś-Tarski Theorem'
[66]: Kuperberg (2021), 'Positive First-order Logic on Words'

Typical proof of relativisation. Let us now turn our attention to the positive side of Figure 3.2. The typical proof in these case leverages the semantic characterisation using minimal models of Lemma 3.2.3, but also the expressiveness of the fragments in the finite described by Lemma 3.1.5.

Concretely, one proves that every first order sentence φ that is preserved under F-embeddings has the semantic property of having finitely many non $\leq_{\mathsf{F}}$-equivalent minimal models. Then, an equivalent sentence in $\exists\mathsf{F}$ is obtained as follows:

$$\varphi \equiv \bigvee_{\mathfrak{A} \in \min_{\leq_{\mathsf{F}}} [\![\varphi]\!]_{\mathcal{C}}} \Delta_{\mathfrak{A}}^{\exists\mathsf{F}} \quad .$$

Example 3.2.7. Let us define Paths to be the collection of finite undirected paths, and P_n to be a finite undirected path with n vertices for $n \geq 1$. Then, the Łoś-Tarski Theorem relativises to Paths.

Proof. Notice that $(\mathsf{Paths}, \subseteq_i)$ is order-isomorphic to $(\mathbb{N}, \leq)$. We use the proof scheme mentioned above: a sentence φ has at most one minimal model for $\subseteq_i$ because all paths are comparable. ∎

Now, the big question is of a combinatorial nature: how does one control the shape/number of $\leq_{\mathsf{F}}$-minimal models of a given first order sentence? There are three main answers in the literature:

- Understand the combinatorial nature of $\leq_{\mathsf{F}}$ (for a degenerate example, $\subseteq_i$ on Paths which was illustrated in Example 3.2.7) [19, 27].
- Understand the combinatorial nature of subsets definable $\mathsf{FO}[\sigma]$ (for a degenerate example, $\mathsf{FO}[\sigma]$ on finite classes of finite structures, which was illustrated in Fact 3.2.2) [6, 7].
- Understand the interplay between $\leq_{\mathsf{F}}$ and $\mathsf{FO}[\sigma]$ [83].

The typical combinatorial tool will be the notion of well-quasi-ordering which will be explored in Chapter 6 (Logically Presented Spaces), while the logic approach will usually leverage of the Gaifman Locality Theorem and will be explored in Chapter 4 (Locality and Preservation). For completeness, we also list the proof scheme developed by Rossman that lies at the frontier of the two extreme approaches. This interplay is materialised by an attention to the quantifier rank of the involved formulas.

[83]: Rossman (2008), 'Homomorphism preservation theorems'

One of the main ideas of [83] is to take a dual point of view and transform the existence of F-embeddings between structures into inclusion of their theories. This is materialised by the following Lemma 3.2.8, which then allows us to "stratify" embeddings.

For that, let us associate a theory to every structure as follows $\mathsf{Th}(\mathfrak{A})$ is defined as the set of all first order sentences $\varphi \in \mathsf{FO}[\sigma]$ such that $\mathfrak{A} \models \varphi$.

Lemma 3.2.8. *Let σ be a finite relational signature,* F *a fragment* *of* FO[σ]*, and* $\mathfrak{A}, \mathfrak{B} \in$ Fin(σ)*. Then,*

$$\mathfrak{A} \leq_{\mathsf{F}} \mathfrak{B} \iff \mathsf{Th}(\mathfrak{A}) \cap \exists\mathsf{F} \subseteq \mathsf{Th}(\mathfrak{B}) \cap \exists\mathsf{F}$$

Proof. The only hard part is going from a theory inclusion to build an embedding. Let $\vec{x} \stackrel{\text{def}}{=} (x_a)_{a \in \mathfrak{A}}$, and $\upsilon \colon x_a \mapsto a$. Because F is countable, let us enumerate its formulas $(\phi_i(\vec{y}_i))_{i \in \mathbb{N}} \in \mathsf{F}^{\mathbb{N}}$ such that $\vec{y}_i \subseteq \vec{x}$ and such that $\mathfrak{A}, \upsilon \models \phi_i$. Now, consider $\psi_i \stackrel{\text{def}}{=} \exists \vec{x}. \bigwedge_{j \leq i} \phi_i$. By definition, $\psi_i \in \exists\mathsf{F}$, and $\mathfrak{A} \models \psi_i$. Hence, $\psi_i \in \mathsf{Th}(\mathfrak{A}) \cap \exists\mathsf{F}$.

As a consequence, $\mathfrak{B} \models \psi_i$, and this provides a map[16] $h_i \colon \mathfrak{A} \to \mathfrak{B}$, that sends the existential witnesses of ψ_i in $\mathfrak{A}$ to their counterpart in $\mathfrak{B}$, and sends every other point of $\mathfrak{A}$ to some element in $\mathfrak{B}$ (this is possible because $\mathfrak{B}$ is non-empty).

16: Note that the existence of h_i as a *function* relies on the fact that $x_1 = x_2$ is expressible in every first order fragment. Otherwise, one would simply obtain a *correspondence* which is not enough to conclude.

Now, there are but finitely many maps from $\mathfrak{A}$ to $\mathfrak{B}$. As a consequence, there exists $i \in \mathbb{N}$ such that $h_i = h_j$ for infinitely many $j \in \mathbb{N}$. Let us prove that h_i is an F-embedding. Let $\theta(\vec{y})$ in F, and $\upsilon' \colon \vec{y} \to \mathfrak{A}$ be such that $\mathfrak{A}, \upsilon' \models \theta(\vec{y})$. Because free variables in formulas of F can be renamed, the formula $\theta' \stackrel{\text{def}}{=} \theta[y \mapsto x_{\upsilon'(y)}]$ belongs to F.[17] Then, $\mathfrak{A}, \upsilon \models \theta'$, whose free variables are in $\vec{x}$. As a consequence, there exists $k \in \mathbb{N}$ such that $\phi_k = \theta'(\vec{x})$. Furthermore, there exists a $j > k$ such that $h_i = h_j$. By construction of h_j, $\mathfrak{B}, h_j \circ \upsilon \models \phi_k$. Therefore, $\mathfrak{B}, h_i \circ \upsilon \models \theta[y \mapsto x_{\upsilon'(y)}]$, and we conclude that $\mathfrak{B}, h_i \circ \upsilon' \models \theta(\vec{y})$. We have proven that h_i is a F-embedding.

17: We use the fact that we can *arbitrarily* rename free variables of formulas in F. Indeed, it may be that $\upsilon'(y_1) = \upsilon'(y_2) = a$ for some $y_1 \neq y_2 \in \vec{y}$ and $a \in \mathfrak{A}$, in which case we are introducing a name clash by mapping both y_1 and y_2 to x_a.

We have proven that $\mathfrak{A} \leq_{\mathsf{F}} \mathfrak{B}$. ■

Lemma 3.2.8 allows us to refine the ordering $\leq_{\mathsf{F}}$ by considering theories up to a certain quantifier rank. Let us write $(\mathsf{F})^{\mathrm{rk} \leq q}$ for the restriction of F to first order formulas of quantifier rank at most q.

This leads to the definition of $\mathfrak{A} \leq_{\mathsf{F}}{}^{q} \mathfrak{B}$ as $\mathsf{Th}(\mathfrak{A}) \cap (\exists\mathsf{F})^{\mathrm{rk} \leq q} \subseteq \mathsf{Th}(\mathfrak{B}) \cap (\exists\mathsf{F})^{\mathrm{rk} \leq q}$, which is meaningful because of the following remark.

Lemma 3.2.9. *Let σ be a finite relational signature, $\mathcal{C} \subseteq$ Fin(σ) be a class of finite structures,* F *be a fragment of* FO[σ]*, a quantifier rank $q \in \mathbb{N}$, and φ be a first order sentence. The following are equivalent:*

- *The sentence φ is equivalent over $\mathcal{C}$ a finite positive Boolean combination of sentences in $(\exists\mathsf{F})^{\mathrm{rk} \leq q}$,*
- *The sentence φ is preserved under $\leq_{\mathsf{F}}{}^{q}$ over $\mathcal{C}$.*

Proof. If φ is equivalent to a sentence in $(\exists\mathsf{F})^{\mathrm{rk} \leq q}$, then in particular it is preserved under $\leq_{\mathsf{F}}{}^{q}$.

Conversely, assume that φ is preserved under $\leq_{\mathsf{F}}{}^{q}$. Let $\mathfrak{A} \in \mathcal{C}$ such that $\mathfrak{A} \models \varphi$. Notice that $(\exists\mathsf{F})^{\mathrm{rk} \leq q}$ contains finitely many non-equivalent sentences. Let us write $\psi_{\mathfrak{A}}$ for the (finite) conjunction of a choice of representatives of the equivalence classes of $\mathsf{Th}(\mathfrak{A}) \cap (\exists\mathsf{F})^{\mathrm{rk} \leq q}$. By construction, $\psi_{\mathfrak{A}}$ is a positive Boolean combination of sentences in $(\exists\mathsf{F})^{\mathrm{rk} \leq q}$. We argue that $\varphi \equiv_{\mathcal{C}} \bigvee_{\mathfrak{A} \models \varphi} \psi_{\mathfrak{A}}$, and that the disjunction is finite. The latter holds because there are finitely many non-equivalent sentences in $(\exists\mathsf{F})^{\mathrm{rk} \leq q}$, hence finitely many non-equivalent conjunctions of finitely many sentences in $(\exists\mathsf{F})^{\mathrm{rk} \leq q}$. Furthermore, if $\mathfrak{A} \models \varphi$, then

$\mathfrak{A} \models \psi_{\mathfrak{A}}$ by construction. Lastly, if $\mathfrak{B} \models \psi_{\mathfrak{A}}$ for some $\mathfrak{A} \models \varphi$, then $\mathsf{Th}(\mathfrak{A}) \cap (\exists\mathsf{F})^{\mathrm{rk} \leq q} \subseteq \mathsf{Th}(\mathfrak{B}) \cap (\exists\mathsf{F})^{\mathrm{rk} \leq q}$, and $\mathfrak{B} \models \varphi$. We have effectively rewritten φ as a finite positive Boolean combination of sentences in $(\exists\mathsf{F})^{\mathrm{rk} \leq q}$. ■

[83]: Rossman (2008), 'Homomorphism preservation theorems'

We now have the necessary tools to understand the main combinatorial lemma from [83] that relates $\leq_{\mathsf{F}}$ and the notion of quantifier rank in the specific case of $\mathsf{F} \stackrel{\text{def}}{=} \mathsf{EPFO}$.

$$\begin{tikzcd} \mathfrak{A} \arrow[r, "{\leq_{\mathsf{F}}}^{q}"] \arrow[d, dotted, "\leq_{\mathsf{F}}"'] & \mathfrak{B} \\ \mathfrak{A}' \arrow[r, dashed, "{\leq_{\mathsf{FO}}}^{q'}"'] & \mathfrak{B}' \arrow[u, dotted, "\leq_{\mathsf{F}}"'] \end{tikzcd}$$

Figure 3.3.: The combinatorial lemma of Rossman leading to the Homomorphism Preservation Theorem

> **Lemma 3.2.10** [83, Corollary 5.14]**.** Let σ be a finite relational signature, and $q' \in \mathbb{N}$. There exists a $q \in \mathbb{N}$, such that for all $\mathfrak{A}, \mathfrak{B} \in \mathsf{Fin}(\sigma)$ satisfying $\mathfrak{A} \,{\leq_{\mathsf{EPFO}}}^{q}\, \mathfrak{B}$, there exists $\mathfrak{A}', \mathfrak{B}' \in \mathsf{Fin}(\sigma)$ such that:
>
> 1. $\mathfrak{A} \leq_{\mathsf{EPFO}} \mathfrak{A}'$,
> 2. $\mathfrak{B}' \leq_{\mathsf{EPFO}} \mathfrak{B}$,
> 3. $\mathfrak{A}' \,{\leq_{\mathsf{FO}}}^{q'}\, \mathfrak{B}'$.

In the original paper [83, Corollary 5.14], the lemma is stated using the equivalence relations induced by $\leq_{\mathsf{F}}$, ${\leq_{\mathsf{F}}}^{q}$, and ${\leq_{\mathsf{FO}}}^{q}$ rather than the orderings themselves. However, it immediately implies Lemma 3.2.10.

A graphical depiction of Lemma 3.2.10 is given in Figure 3.3. Notice that it implies the relativisation of the Homomorphism Preservation Theorem to $\mathsf{Fin}(\sigma)$. Indeed, let φ be a first order sentence that is preserved under $\preceq_h$; since φ has a given quantifier rank $q' \stackrel{\text{def}}{=} \mathrm{rk}(\varphi)$, one can leverage Lemma 3.2.10 to conclude that φ is preserved under ${\leq_{\mathsf{EPFO}}}^{\mathrm{rk} \leq q}$: hence equivalent to some $\mathsf{EPFO}^{\mathrm{rk} \leq q}$ sentence by Lemma 3.2.9.

All of these three approaches are an invitation to a formalisation and generalisation. This is what we will try to explore in the upcoming section.

3.3. Order and Topology

This section connects the study of typical proofs carried out in Section 3.1 (Classical Preservation Theorems) and Section 3.2 (Preservation Theorems do not Relativise) to well established fields of computer science and mathematics: well-quasi-orderings and Noetherian spaces.

We will start by exploring the proof schemes that were described when dealing with classes of finite structures, and connect the combinatorial properties of $[\![\varphi]\!]$ with the notion of well-quasi-order.

Later, we will explore the meaning of the compactness theorem of first order logic in classes of structures $\mathcal{C} \subseteq \mathsf{Struct}(\sigma)$, which will lead us towards the language of topology.

3.3.1. The Finite Case: Well Quasi Orderings

Because the typical study carried out to understand whether a given preservation theorem relativises to a given class uses so-called "minimal models", we can ask ourselves what kind of properties this enforces on the ordering $\leq_{\mathsf{F}}$. We argue that this place us in a setting that is very similar to the one of well-quasi-orders that we introduce hereafter.

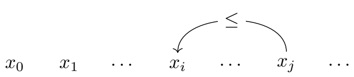


Figure 3.4.: Graphical representation of an increasing pair in a good sequence.

Definition 3.3.1. Let $(P, \leq)$ be a quasi-order. A sequence $(x_n)_{n\in\mathbb{N}} \in P^{\mathbb{N}}$ is *good* whenever there exists $i < j$ such that $x_i \leq x_j$.

A quasi-ordered set $(P, \leq)$ is *well-quasi-ordered* if every sequence $(x_n)_{n\in\mathbb{N}}$ of elements in P is good.

By calling a sequence *bad* whenever it is not good, well-quasi-orderings are equivalently defined as having no infinite bad sequences.

Example 3.3.2. The set $\mathbb{N}^2$, equipped with the pointwise ordering, is a well-quasi-order. This is illustrated in Figure 3.5, where we give a feel of why one cannot indefinitely choose elements that are not above the previous ones in $\mathbb{N}^2$.

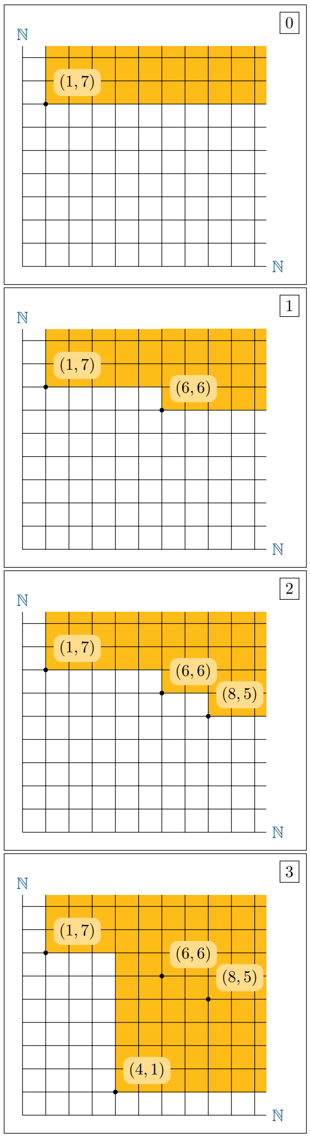


Figure 3.5.: Every infinite sequence in $\mathbb{N}^2$ with the pointwise ordering is good. We describe an example of a bad sequence in $\mathbb{N}^2$ and colour in orange the regions of $\mathbb{N}^2$ that are forbidden if one wants to continue building a bad sequence.

The theory of well-quasi-orders (abbreviated as wqo as of now), is a slight restriction of the maybe more familiar notion of well-founded orderings. When dealing with quasi-orders, it is not only important to control the decreasing sequences, but also to control sequences of elements that are incomparable, that is, antichains. The following Lemma 3.3.3 precisely states this connection.

Lemma 3.3.3. Let $(P, \leq)$ be a quasi-ordered set. The following are equivalent:

1. $(P, \leq)$ is a well-quasi-ordering,
2. The following infinite sequences do not occur in P: infinite antichains, and infinite decreasing sequences.

There is an immediate connection between Lemma 3.3.3 and the proof schemes to study the relativisation of preservation theorems to classes of finite structures. Recall that a key ingredient was the study of infinite antichains[18] and answering the question of whether or not such antichains were definable.

We can actually get closer to our original interests by leveraging another characterisation of well-quasi-orders in terms of upwards closed sets.

Lemma 3.3.4. Let $(P, \leq)$ be a quasi-order. The following are equivalent:

1. $(P, \leq)$ is a wqo,
2. For all upwards closed subsets H of P, there exists $S \subseteq_{\mathsf{fin}} H$ such that $\uparrow_{\leq} S = H$.

The above Lemma 3.3.4 is almost the statement that we used when studying preservation theorems via Lemma 3.2.3: sentences in $\exists\mathsf{F}$ as those being described via finitely many minimal models!

We have now a first "logic-free" way to prove the relativisation of preservation theorems: prove that the class $(\mathcal{C}, \leq_{\mathsf{F}})$ is a well-quasi-order.

18: For instance, the sequence of cycles for $\subseteq_i$.

Lemma 3.3.5 (folklore)**.** Let σ be a finite relational signature, $\mathsf{F} \in \{ \mathsf{EFO}, \mathsf{EPFO}^{\neq}, \mathsf{EPFO} \}$, and $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ be such that $(\mathcal{C}, \leq_{\mathsf{F}})$ is wqo. Then, the F-preservation theorem relativises to $\mathcal{C}$.

Proof. Let $\varphi \in \mathsf{FO}[\sigma]$ be preserved under $\leq_{\mathsf{F}}$ over $\mathcal{C}$. This precisely means that $[\![\varphi]\!]_{\mathcal{C}}$ is upwards closed in $\mathcal{C}$. Because $\mathcal{C}$ is a wqo, there exists $S \subseteq_{\mathsf{fin}} \mathcal{C}$ such that $[\![\varphi]\!]_{\mathcal{C}} = \uparrow_{\leq_{\mathsf{F}}} S$. Then, Lemma 3.2.3 provides a $\psi \in \exists\mathsf{F}$ such that $[\![\psi]\!]_{\mathcal{C}} = [\![\varphi]\!]_{\mathcal{C}}$, and we have concluded. ∎

Let us briefly place well-quasi-orderings in a historical context before continuing our journey. As a combinatorial tool, well-quasi-orderings appear frequently in varying fields of computer science, ranging from graph theory to number theory [19, 59, 64, 65]. Well-quasi-orderings have also been highly successful in proving the termination of verification algorithms, where one of their critical application is to the verification of infinite state transition systems, via the study of so-called Well-Structured Transition Systems (WSTS) [1, 35, 53].

[19]: Daligault, Rao and Thomassé (2010), 'Well-Quasi-Order of Relabel Functions'
[59]: Higman (1952), 'Ordering by divisibility in abstract algebras'
[64]: Kříž and Thomas (1990), 'On well-quasi-ordering finite structures with labels'
[65]: Kruskal (1972), 'The theory of well-quasi-ordering: A frequently discovered concept'
[1]: Abdulla, Čerāns, Jonsson and Tsay (1996), 'General decidability theorems for infinite-state systems'
[35]: Finkel and Schnoebelen (2001), 'Well-structured transition systems everywhere!'
[53]: Goubault-Larrecq, Seisenberger, Selivanov and Weiermann (2016), 'Well Quasi-Orders in Computer Science (Dagstuhl Seminar 16031)'

Classes of models that are wqo. Let us now explore to which extent Lemma 3.3.5 explains the successes in proving the relativisation of preservation theorems in the finite.[19]

19: Because the lemma only provides a sufficient condition, it cannot be used to explain negative relativisation results.

Remark that for $\leq \in \{ \subseteq_i, \subseteq \}$, the set $(\mathcal{C}, \leq)$ is well-founded as soon as $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$. Therefore, the question of being a wqo boils down to the behaviour of antichains for these quasi-orders.

The specific case of finite graphs has been investigated by Ding already in 1992, where he provided a characterisation of monotone classes of graphs that are wqo for $\subseteq_i$ (resp. $\subseteq$) [27, Theorem 2.7]. However, the provided characterisation in terms of forbidden subgraphs does not work in the presence of colours or directed edges, and fails *a fortiori* for general relational structures.

[27]: Ding (1992), 'Subgraphs and well-quasi-ordering'

Let us formally define what we mean by "colouring" structures.[20] Let σ be a finite relational signature, L be a finite set of labels, and $\mathcal{C} \in \mathsf{Struct}(\sigma)$ be a class of structures. One can build the class $\mathsf{Lab}(L, \mathcal{C})$ of structures in $\mathcal{C}$ *labelled by* elements in L as a subset of $\mathsf{Struct}(\sigma \uplus L)$ obtained by freely assigning the unary predicates in L to elements in structures of $\mathcal{C}$.

20: For better compatibility with unary predicates, we consider "labels" rather than "colours" but the two are as expressive.

Example 3.3.6. The class Paths of finite paths is wqo for $\subseteq_i$, but $\mathsf{Lab}(L, \mathsf{Paths})$ is not as soon as $|L| \geq 1$.

Proof. The class $(\mathsf{Paths}, \subseteq_i)$ is isomorphic to $(\mathbb{N}, \leq)$ which is wqo. However, the sequence of paths with labelled endpoints is an infinite antichain for $\subseteq_i$ in $\mathsf{Lab}(\{ B \}, \mathsf{Paths})$, as explained in Figure 3.6. ∎

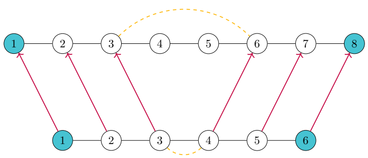


Figure 3.6.: Illustration of the impossibility to embed a path with 6 vertices into a path with 8 vertices when their endpoints are coloured. To build an embedding, one must send 1 and 6 of P_6 to the endpoints of P_8, and this leads to a contradiction because there will be at least one pair of consecutive nodes in P_6 that are not consecutive in P_8. We coloured the endpoints in blue, a potential embedding in thick red, and the contradiction arises from the presence/absence of the thick dashed yellow arcs.

In order to talk about signatures beyond finite graphs and unary predicates, let us leverage the notion of tree-depth that was already mentioned when defining $\subseteq_i$ in Definition 2.3.20.

Definition 3.3.7. Let σ be a finite relational signature. A class $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ of structures has *bounded tree-depth* if and only if there exists $m \in \mathbb{N}$ such that $\mathrm{td}(\mathfrak{A}) \leq m$ for all $\mathfrak{A}$ in $\mathcal{C}$.

A first result present in Figure 3.2 is [27, Theorem 2.6], that characterises monotone classes of finite graphs that are well-quasi-ordered for $\subseteq$ and $\subseteq_i$, when allowing colourings. We restate the theorem in terms of classes of finite structures which is an immediate generalisation.

[27]: Ding (1992), 'Subgraphs and well-quasi-ordering'

Let σ be a finite relational signature, L be a non-empty finite set of labels, and $\mathcal{C}$ be a monotone class of finite structures. Then, the following are equivalent:

Theorem 3.3.8 [27, Theorem 2.6]**.** The following are equivalent for a monotone class $\mathcal{C}$ of finite undirected graphs.

1. $\mathcal{C}$ has bounded tree depth,
2. $(\mathsf{Lab}(L, \mathcal{C}), \subseteq)$ is wqo,
3. $(\mathsf{Lab}(L, \mathcal{C}), \subseteq_i)$ is wqo.

There is a slight overloading of notations when talking about monotone classes of finite undirected graphs. In graph theory, a subgraph of an undirected graph cannot be a directed graph, while our definition of finite graphs as elements in $\mathsf{Fin}(\{ (E, 2) \})$ allows such behaviour. Hence, a monotone class of finite graphs must be empty using our definition, which is avoided by the common definition of a monotone class of undirected graphs as the restriction of a monotone class of directed graphs to those that are undirected. Note that there was no confusion about hereditary classes of undirected graphs, as a induced substructure of an undirected graph is undirected too.

While Theorem 3.3.8 only considers classes of undirected graphs, it can be lifted to monotone classes of finite structures over a finite relational signature, because of the lemma in [27, Lemma 2.5].

Fact 3.3.9. Let σ be a finite relational signature, let $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ be a class of finite structures, and L be a finite set of labels. If $\mathcal{C}$ has bounded tree-depth, then $(\mathsf{Lab}(L, \mathcal{C}), \subseteq_i)$ is a well-quasi-order.

We do not know if the following lemma is really folklore, but it is not really difficult to prove. However, at is it not central to the study carried out in this manuscript, we only give a brief intuition of the proof in the margin.

Lemma 3.3.10. Let σ be a finite relational signature, let $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ be a hereditary class of finite structures, and L be a finite set of labels with $|L| \geq 1$. The following are equivalent:

- $\mathcal{C}$ has bounded tree depth,
- $(\mathsf{Lab}(L, \mathcal{C}), \subseteq)$ is wqo,
- $(\mathsf{Lab}(L, \mathcal{C}), \subseteq_i)$ is wqo.

The formal proof of Lemma 3.3.10 does not fit in the margin. The only difficult implication in Lemma 3.3.10 is to go from an unbounded tree-depth to an antichain for $\subseteq$. The idea is that for a monotone class of structures with relations of arity at most k, one can prove by induction on n that a structure with tree-depth at least $k \times n$ must contain some sort of long "path of cliques," as a substructure for a good definition of "path" and "clique." Adding colours automatically leads to an infinite antichain.

This only studies monotone classes of structures, but recent results on hereditary classes have been provided by [19, Theorem 3], that provides a criterion for tree-generated classes of graphs to be wqo with respect to $\subseteq_i$. These results that connect the property of being wqo to structural properties of the class of graph (having bounded *something*) are recapitulated in Figure 3.7. This illustrates how the wqo approach quickly reaches its limits: while structural properties such as having bounded clique-width, or bounded tree-width are ordinarily indicators of a favourable ground for finite model theory to thrive, only degenerate

[19]: Daligault, Rao and Thomassé (2010), 'Well-Quasi-Order of Relabel Functions'

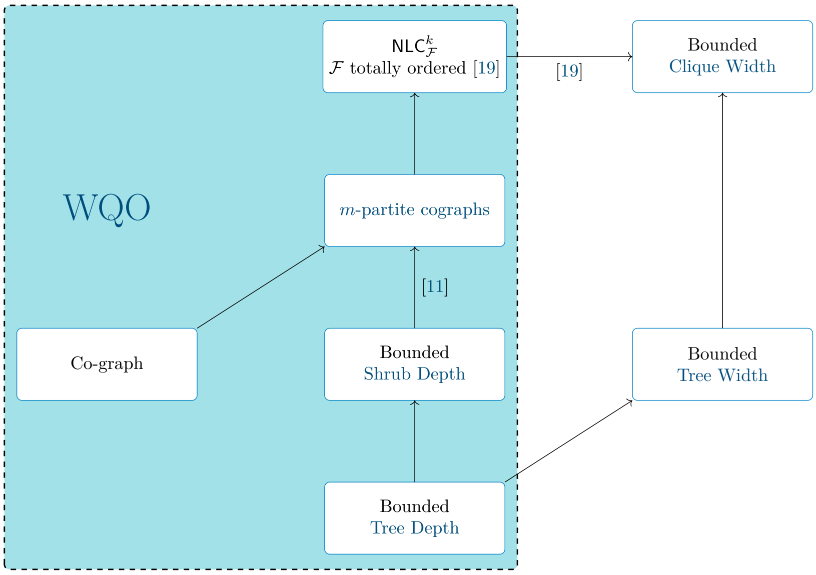


Figure 3.7.: Classes of (coloured) graphs that are structurally simple thanks to having tree decompositions of a specific shape. In green, the classes that are known to be well-quasi-ordered for $\subseteq_i$.

versions of these tree decompositions imply a property of well-quasi-ordering.

Another structural approach to the complexity of graph classes is the research program started by Nešetřil and Ossona de Mendez that studies the "sparsity" of the graph classes, with a taxonomy described in Figure 3.8. In this classification too, the realm of wqo is limited to a tiny subset of the classes of graphs that are considered as "simple."

Let us conclude this introduction to well-quasi-orderings by a useful definition that leads to two (unsolved) conjectures.

Definition 3.3.11. Let σ be a finite relational signature, and $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$. The class $\mathcal{C}$ is *∞-wqo* for $\subseteq_i$ if and only if $(\mathsf{Lab}(L, \mathcal{C}), \subseteq_i)$ is wqo for all finite L.

In its usual formulations with "colourings," the conjecture is formulated as the equivalence between being ∞-wqo and being "2-wqo."

Conjecture 3.3.12 (Pouzet's Conjecture)**.** A class $\mathcal{C}$ of finite graphs is ∞-wqo if and only if $(\mathsf{Lab}(\{ P \}, \mathcal{C}), \subseteq_i)$ is wqo, where P is an arbitrary symbol.

There are good reasons for me to believe that the above conjecture can be strengthened by considering classes of structures, and explicitly encoding the infinite antichain of Cycles. The conjecture is due to Sylvain Schmitz.

Conjecture 3.3.13 (Sylvain's Conjecture)**.** Let σ be a finite relational signature, and $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$. The class $\mathcal{C}$ is not ∞-wqo if and only if there exists an infinite subset $S \subseteq \mathsf{Cycles}$, a finite set L, and a first order interpretation $\mathsf{I}\colon \mathsf{Lab}(L, \mathcal{C}) \twoheadrightarrow S$ such that I is monotone.

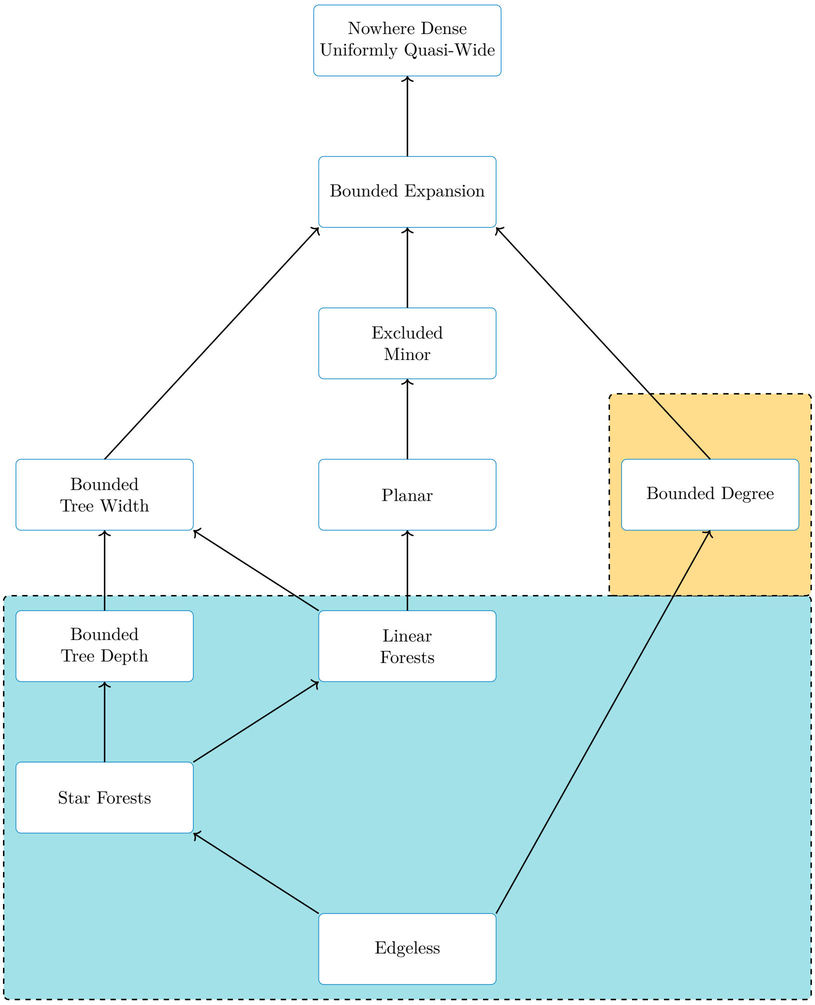


Figure 3.8.: What is known for wqo and preservation theorems with respect to the usual sparse classes that are hereditary. Beware that sparse "classes" depicted here are collections of classes of graphs. Arrows represent strict collection inclusion. Collections that are in the lower green rectangle contain classes that are well-quasi-ordered for $\subseteq_i$, and the others contain at least one class of graphs that is not well-quasi-ordered for $\subseteq_i$. Under the extra assumption that the classes are stable under disjoint unions, all the sparsity properties imply the relativisation of the Homomorphism Preservation Theorem [7, 20], and having bounded degree implies that the relativisation of the Łoś-Tarski Theorem [6].

3.3.2. The Infinite Case: Compactness.

Because of the limitations of "completely ignoring the logic" that the study of wqos imply, let us explore a different approach and try to get back the main tool that was missing and ask the following question: when does the compactness theorem of first order logic relativise to classes of finite structures?

We warn the reader that the topological approach will not provide more information than the one using well-quasi-orders at the end of this section. However, we will gain intuitions and a better vocabulary to go beyond what was done using wqos.

To that end, let us first *define* the notion of compactness, for which we need to introduce a few notions of basic topology.

> **Definition 3.3.14.** *A topological space is a pair (X, τ) where $\tau \subseteq \mathcal{P}(\mathsf{X})$, τ is stable under finite intersections and τ is stable under arbitrary unions.*

In particular, τ must contain both $\emptyset$ and X, the first one being the empty union, and the second one being the empty intersection inside the complete lattice $\mathcal{P}(\mathsf{X})$.

In a topological space (X, τ), the elements of τ are called *open subsets* while their complements are said to be *closed subsets*.

Perhaps the most striking difference between topological spaces and the previously encountered algebraic structures such as quasi-ordered sets, is the former's "second order" nature. However, this is particularly adapted to the study of first order logic, as a first order sentence defines a subset of $\mathcal{C}$ via the map $\varphi \mapsto [\![\varphi]\!]_{\mathcal{C}}$.

From a logic to a topology. It is tempting to define over a class[21] $\mathcal{C} \subseteq \mathsf{Struct}(\sigma)$ a topology defined as $\tau \stackrel{\text{def}}{=} [\![\mathsf{FO}[\sigma]]\!]_{\mathcal{C}}$.[22] This would be the "definable topology." Similarly, one could define, given a fragment F the topology $[\![\mathsf{F}]\!]_{\mathcal{C}}$ that contains fewer open subsets.[23]

21: Recall that we assume suitable cardinality restrictions so that the class is actually a set, and so is its powerset.

22: That is, the set of $[\![\varphi]\!]_{\mathcal{C}}$ where φ ranges over sentences of $\mathsf{FO}[\sigma]$.

23: Which is therefore a coarser topology

However, the above paragraph is nonsensical, as none of the claimed topologies are actual topologies. While it is true that $[\![\mathsf{FO}[\sigma]]\!]_{\mathcal{C}}$ is closed under finite intersections, finite unions, contains $\mathcal{C}$ and contains $\emptyset$, this set is in general not closed under arbitrary unions.

This is a common problem when defining a topology over a given set, which we tackle as follows. For every set X and collection of subsets $B \subseteq \mathcal{P}(\mathsf{X})$, one can construct the *topology generated by* B as the smallest topology on X containing B, which we write $\langle B \rangle_{\mathsf{topo}}$. This topology coincides with the one containing arbitrary unions of finite intersections of subsets in B. We say that B is a *subbasis* of τ when τ is the topology generated by B.[24]

24: When τ is obtained by only considering arbitrary unions of elements in B, we say that B is a *basis* of τ. But this will not play a role in this document.

Let us provide an example of topological space obtained through a subbasis that may be more familiar to the reader.

> **Example 3.3.15.** The set $\mathbb{R}$ of real numbers has a natural topology generated by the open intervals $]a,b[$ for $a < b \in \mathbb{R}$.
>
> In this topology, the interval $[a,b]$ is a closed subset, and the interval $]a,b[$ is an open subset. Furthermore, both $]a,b]$ and $[a,b[$ are neither closed subsets nor open subsets.

Now that we have given an example of a topology with a definition that does not come from a logic, let us settle on our next task: associating topological spaces to fragments of $\mathsf{FO}[\sigma]$.

> **Definition 3.3.16.** Let σ be a finite relational signature, $\mathcal{C}$ be a class of finite structures, and F be a fragment of $\mathsf{FO}[\sigma]$. We write $\langle \exists \mathsf{F} \rangle_{\mathsf{topo}}$ for the topology generated by the sets $[\![\varphi]\!]_{\mathcal{C}}$, where φ ranges over elements of $\exists \mathsf{F}$.
>
> The space $(\mathcal{C}, \langle \exists \mathsf{F} \rangle_{\mathsf{topo}})$ is therefore a topological space.

Remark that, *a priori*, the topology $\langle \exists \mathsf{F} \rangle_{\mathsf{topo}}$ carries more information than the simple quasi-ordering $\leq_{\mathsf{F}}$. In this setting, a sentence φ is preserved under F-embeddings whenever it defines an open subset:[25]

25: This is the topological translation of Lemma 3.2.8.

> **Remark 3.3.17.** Let σ be a finite relational signature, $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$, F be a fragment of $\mathsf{FO}[\sigma]$, and $\varphi \in \mathsf{FO}[\sigma]$. Then, the following are

equivalent:

1. φ is preserved under F-embeddings,
2. $[\![\varphi]\!]_{\mathcal{C}}$ is an open subset of $(\mathcal{C}, \langle \exists \mathsf{F} \rangle_{\mathsf{topo}})$.

It may help to think about the closed subsets in the topological space $(\mathsf{Struct}(\sigma), \langle \mathsf{FO} \rangle_{\mathsf{topo}})$: these are obtained as complements of open subsets, that is, arbitrary intersections of finite unions of subsets of the form $[\![\varphi]\!]$. In other words, the closed subsets of this topological spaces are precisely the subclasses of structures that can be described by first order theories, i.e., that have a first order axiomatisation.

We do not talk about sequences at all in this manuscript, therefore the following definition of compactness might not be the one that the reader is used to.

Compactness in topological spaces. One of the central topological properties that will arise in this thesis is the notion of "compactness." Informally, this is the topological generalisation of a finite subset in the following sense: if union $\bigcup_{i \in \mathbb{N}} S_i$ contains a finite set S, then there exists $n \in \mathbb{N}$ such that S is included in the finite union $\bigcup_{i<n} S_i$.

Definition 3.3.18. An *open cover* of a subset E in a topological space (X, τ) is a family $(U_i)_{i \in I}$ in τ^I such that $E \subseteq \bigcup_{i \in I} U_i$.

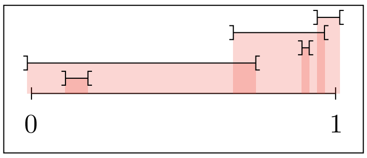


Figure 3.9.: Covering the unit interval $[0,1]$ using open intervals of the form $]a,b[$ with $a < b$.

Example 3.3.19. The collection of open intervals $(]-n,n[)_{n \in \mathbb{N}}$ is an open cover of $\mathbb{R}$ in its usual topology.

Definition 3.3.20. A subset K of a topological space (X, τ) is *compact* when for every open cover $(U_i)_{i \in I}$ of K, there exists a finite subset $J \subseteq_{\mathsf{fin}} I$ such that K is covered by $(U_j)_{j \in J}$.

We say that (X, τ) is *compact* when X is a *compact* subset of X.

Let us reuse our nice topology on $\mathbb{R}$ to have the first proof of compactness. This will illustrate the main technicalities that we face during those proofs.

Example 3.3.21. The set $\mathbb{R}$ with its usual topology (given in Example 3.3.15) is not compact, but $[0,1]$ is a compact subset of $\mathbb{R}$.

Proof. The sequence $(]-n,n[)_{n \in \mathbb{N}}$ is an open cover of $\mathbb{R}$. Assume by contradiction that there exists $I \subseteq_{\mathsf{fin}} \mathbb{N}$ is such that $(]-n,n[)_{n \in I}$ is a finite subcover of $\mathbb{R}$. The element $\max(I)$ belongs to $\mathbb{R}$ but is not in $\bigcup_{n \in I}]-n,n[$.

There are several proofs that $[0,1]$ is compact. The canonical way would be to use the notion of convergence, which does not fit well with the presentation of topological spaces we gave by avoiding the notion of convergence altogether. Let us provide a quick proof that only uses the Nested Intervals Theorem[26], and provide yet another proof leveraging Alexander's subbase lemma later on.[27]

26: See Lemma D.8.3 p. 255

27: One other proof might be more to the liking of order theoreticians since it connects the notion of compactness to the existence of least upper bounds.

Assume by contradiction that $[0,1]$ is not compact. In particular, there exists an open cover $(U_i)_{i\in I}$ of $[0,1]$ that has no finite subcover.[28]

28: Notice that U_i is an open subset, which may not be an open interval. Furthermore, the sequence is indexed by a set I that might not be $\mathbb{N}$.

Notice that $[0,1] = [0,1/2] \cup [1/2,0]$. At least one of them cannot be covered by finitely many elements in $(U_i)_{i\in I}$. Let us call this subset F_1, which is a closed subset, of the form $[a,b]$ with $b - a = 2^{-1}$. One can continue the process of dividing this closed interval in two to obtain a sequence $(F_n)_{n\geq 1}$ with the following properties:

- For all $n \geq 1$, F_n is not covered by finitely many elements in $(U_i)_{i\in I}$.
- $(F_n)_{n\geq 1}$ is a decreasing sequence of closed intervals,
- For all $n \geq 1$, F_n has length 2^{-n}.

Now, using the Nested Intervals Theorem, there exists $x \in \bigcap_{n\geq 1} F_n$. In particular, $x \in [0,1]$, and therefore there exists $i \in I$ such that $x \in U_i$. By definition of the topology on $\mathbb{R}$,[29] there exists $a < b \in \mathbb{R}$ such that $x \in]a,b[\subseteq U_i$. Because $\mathbb{R}$ is Archimedean, there exists $n \geq 1$ such that $2^{-n} < b - a$. Therefore, $x \in F_{n+1} \subseteq]a,b[\subseteq U_i$. This is absurd, because we have provided a finite subcover of F_n using exactly one element in $(U_i)_{i\in I}$. ■

29: This is the topology generated by the open intervals $]a,b[$ with $a < b$.

The proof of Example 3.3.21 used non elementary properties of the space of real numbers to overcome several issues. The first one was the presence of an arbitrarily large set of indices I, which was circumvented through the definition of an actual $\mathbb{N}$-indexed sequence of subsets. The second problem was the uncontrolled shape of open subsets that are unions of finite intersections of open intervals. This was tackled by selecting a specific point x, allowing us to only talk about the subbasis of the topology later on.

Let us now introduce a technical tool to avoid the complexities of Example 3.3.21 in more involved cases: Alexander's subbase lemma reduces the compactness of a given subset as a property that only depends on a given subbasis[30].

30: This lemma requires a non-trivial proof and uses the axiom of choice. Be aware that it is “overpowered” to simply prove the compactness of the unit interval.

Lemma 3.3.22 [45, Thm. 4.4.29]**.** Let (X, τ) be a topological space, B be a subbase of τ, and S be a subset of X. Then, the following are equivalent:

1. S is compact in X,
2. For any family $(U_i)_{i\in I}$ of open subsets in B that covers S, there exists a finite subset $J \subseteq_{\mathsf{fin}} I$ such that $(U_i)_{i\in J}$ covers S.

This is called *Alexander's subbase lemma*.

Let us demonstrate how Alexander's subbase lemma simplifies the proof of Example 3.3.21. While it is impossible to avoid using properties of the real line concerning the existence of infimum and supremum, the following proof does not require deep theorems and the only originality is in its choice of subbasis.

Example 3.3.23. The subset $[0,1]$ is compact in $\mathbb{R}$.

Proof. Let $(U_i)_{i \in I}$ be an open cover of $[0{,}1]$. Using Alexander's subbase lemma, one can assume without loss of generality that $U_i =]a_i{,}b_i[$ with $a_i < b_i$. One can even go further and assume that $U_i =]-\infty{,}b_i[$ or $U_i =]a_i, +\infty[$ since this is also a subbase of the topology.

Let us write I_- for the collection of $i \in I$ such that U_i is of the form $]-\infty{,}b_i[$, and I_+ for the collection of $i \in I$ such that U_i is of the form $]a_i, +\infty[$. Without loss of generality, one can assume that $0 < a_i < 1$ when $i \in I_+$ and $0 < b_i < 1$ when $i \in I_-$.[31]

31: Otherwise, the open subset does not intersect $[0{,}1]$ and can be safely discarded, or contains $[0{,}1]$, and we have extracted a finite subcover.

Consider $b \stackrel{\text{def}}{=} \sup\{\, b_i : i \in I_- \,\}$, and $a \stackrel{\text{def}}{=} \inf\{\, a_i : i \in I_+ \,\}$. This construction is depicted in Figure 3.10, where the drawing should already convince you that the proof is finished. We claim that $a < b$, which immediately entails (by definition of the supremum and infimum) the existence of $k \in I_-$ and $l \in I_+$ such that $[0{,}1] \subseteq U_k \cup U_l$, i.e., of a finite subcover (of size two).

Assume by contradiction that $a \geq b$. There is a point $x \in [b{,}a] \subseteq [0{,}1]$. This point must be covered by U_i for some $i \in I$. If $i \in I_+$, this contradicts the definition of a as an infimum. If $i \in I_-$, this contradicts the definition of b as a supremum. ∎

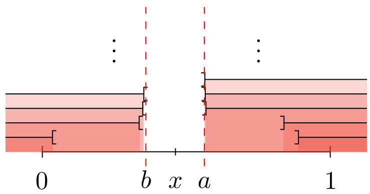


Figure 3.10.: Illustrating the construction of a, b and x in the proof of Example 3.3.23.

Now that we are acclimated to the notion of open cover, compactness and the technical tool of Alexander's subbase lemma, we are ready to recast the compactness theorem of first order logic into this topological setting.

> **Theorem 3.3.24.** Let σ be a relational signature. The topological space $(\mathsf{Struct}(\sigma), \langle \mathsf{FO}[\sigma] \rangle_{\mathsf{topo}})$ is compact.

The theorem is actually an equivalence: stating that the space $\mathsf{Struct}(\sigma)$ is compact directly implies the compactness of first order logic. Therefore, Theorem 3.3.24 should rather be understood as a formalisation that the "compactness of first order logic" really is a reformulation of the (more general) notion of topological compactness.

Proof. Thanks to Alexander's subbase lemma, it suffices to consider an open cover of $\mathsf{Struct}(\sigma)$ using first order sentences.[32] Let $(\varphi_i)_{i \in I}$ be a sequence of first order sentences such that $\mathsf{Struct}(\sigma) \subseteq \bigcup_{i \in I} [\![\varphi_i]\!]$. In particular, the theory $\mathrm{T} \stackrel{\text{def}}{=} \{\, \neg\varphi_i : i \in I \,\}$ has no model. Thanks to the compactness theorem of first order logic, there exists a finite subset $J \subseteq_{\mathsf{fin}} I$ such that $\{\, \neg\varphi_i : i \in J \,\}$ has no model. This proves that $\mathsf{Struct}(\sigma) \subseteq \bigcup_{i \in J} [\![\varphi_i]\!]$, which is a finite subcover. ∎

32: Formally, we consider subsets of $\mathsf{Struct}(\sigma)$ that are definable in $\mathsf{FO}[\sigma]$.

Preservation theorems and Noetherian Spaces. Let us now explore how the topological compactness can be leveraged to prove relativisation of preservation theorems to classes of structures.

The idea will be to replace the extra conditions placed on combinatorial lemmas using minimal models (see Lemma 3.2.3) by topological properties of the class of structures. At this point, we have reached a level of abstraction that makes the following Lemma 3.3.25 almost trivial.

> **Lemma 3.3.25.** Let σ be a finite relational signature, $\mathcal{C} \subseteq \mathsf{Struct}(\sigma)$, F be a fragment of $\mathsf{FO}[\sigma]$, and $\varphi \in \mathsf{FO}[\sigma]$ be such that $[\![\varphi]\!]_{\mathcal{C}}$ is a compact open subset of $(\mathcal{C}, \langle \exists\mathsf{F} \rangle_{\mathsf{topo}})$. Then, there exists a sentence $\psi \in \exists\mathsf{F}$ such that $\varphi \equiv_{\mathcal{C}} \psi$.

Proof. Because $[\![\varphi]\!]_{\mathcal{C}}$ is an open subset and $\exists\mathsf{F}$ is a subbasis stable under finite intersections, there exists a sequence $(\psi_i)_{i\in I}$ of sentences in $\exists\mathsf{F}$ such that

$$[\![\varphi]\!]_{\mathcal{C}} = \bigcup_{i\in I} [\![\psi_i]\!]_{\mathcal{C}} \quad .$$

Because $[\![\varphi]\!]_{\mathcal{C}}$ is compact, there exists a finite subset $J \subseteq_{\mathsf{fin}} I$ such that $(\psi_i)_{i\in J}$ defines an open cover of $[\![\varphi_i]\!]_{\mathcal{C}}$. In particular, we have $\varphi \equiv_{\mathcal{C}} \bigvee_{i\in J} \psi_i$, the latter being an $\exists\mathsf{F}$ sentence. ■

Because Remark 3.3.17 translates "φ is preserved under F-embeddings" into "φ is an open subset of $\langle\exists\mathsf{F}\rangle_{\mathsf{topo}}$", Lemma 3.3.25 can be applied to prove relativisation as soon as one controls the compactness of open subsets that can be defined in $\mathsf{Struct}(\sigma)$ using first order sentences.

In particular, the collection of the subsets that are both open and compact will be at the heart of our investigations.

Definition 3.3.26. Given a topological space (X, τ), we write $\mathcal{K}^{\circ}(\mathsf{X})$ for the set of *compact open* subsets of X.

As in the case of well-quasi-orderings, a working technique is to forget that the subset was definable, and ask for the compactness of *all subsets*.

Definition 3.3.27 [45, Definition 9.7.1, Proposition 9.7.7]**.** A space (X, τ) is *Noetherian* whenever every subset of X is compact.

Equivalently, a space (X, τ) is Noetherian whenever $\mathcal{K}^{\circ}(\mathsf{X}) = \tau$.

We immediately obtain the variant of Lemma 3.3.5 in this topological setting, stating that Noetherian spaces induce preservation theorems. While it does not seem to appear in the literature, it is an immediate consequence of known results hence appears as an external result.

Lemma 3.3.28. Let σ be a finite relational signature, F be a fragment of $\mathsf{FO}[\sigma]$, and $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ such that $(\mathcal{C}, \langle\exists\mathsf{F}\rangle_{\mathsf{topo}})$ is Noetherian. Then, the F-preservation theorem relativises to $\mathcal{C}$.

Proof. Let $\varphi \in \mathsf{FO}[\sigma]$ be preserved under $\leq_{\mathsf{F}}$ over $\mathcal{C}$. This precisely means that $[\![\varphi]\!]_{\mathcal{C}}$ is an open subset via Remark 3.3.17. Because the latter is Noetherian, $[\![\varphi]\!]_{\mathcal{C}}$ is also compact. Leveraging Lemma 3.3.25, there exists $\psi \in \exists\mathsf{F}$ such that $\varphi \equiv_{\mathcal{C}} \psi$, and we have concluded. ■

We saw in the case of wqos that the assumption was not a reasonable one in most cases. There is some hope that Lemma 3.3.28 applies more broadly because it does not require structural properties such as being downwards closed for a suitable quasi order.[33] This illusion will not resist a historical perspective on Noetherian spaces.

33: That is, being a hereditary class, a monotone class or a co-homomorphism closed class of structures.

34: We redirect the reader to Figure 3.11 for a short reminder on the usual properties of rings.

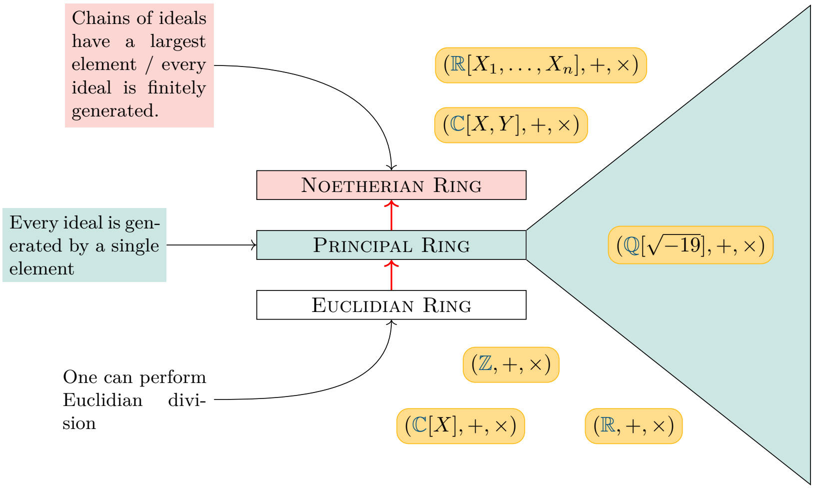


Figure 3.11.: Strict property inclusion for the usual notion of (commutative) ring. We refer the reader to Appendix B for a brief overview of ring properties.

Noetherian spaces and verification. Noetherian spaces arise from the corresponding notion of Noetherian Ring in classical algebra,[34] the idea to use them in the verification of transition system is mostly due to [44].

[44]: Goubault-Larrecq (2010), 'Noetherian Spaces in Verification'

In the mindset of verifying infinite state transition systems, Noetherian spaces are seen as a topological generalisation of well-quasi-orderings, that solves two orthogonal limitations of the latter.

- The first one is that Noetherian spaces are stable under the powerset construction[35] [45, Exercise 9.7.14], which fails for wqos [80], and drove the introduction of a *restriction* of wqos called *better quasi orders* [75].
- The second property is the ability to consider state spaces that do not have interesting quasi-ordering, such as sets of complex vectors $\mathbb{C}^n$. Typical verification tools try to leverage polynomial invariants of such systems [60], where a natural topological presentation arises in the shape of the Zariski topology ...which is Noetherian precisely because $\mathbb{C}[X_1, \ldots, X_n]$ is a Noetherian ring.

35: That is, the powerset of a Noetherian space is a Noetherian space in a suitable topology.

[45]: Goubault-Larrecq (2013), *Non-Hausdorff Topology and Domain Theory*

[80]: Rado (1954), 'Partial well-ordering of sets of vectors'

[75]: Milner (1985), 'Basic wqo-and bqo-theory'

[60]: Hrushovski, Ouaknine, Pouly and Worrell (2018), 'Polynomial Invariants for Affine Programs'

With this brief historical explanation, let us revisit the relativisation of Łoś-Tarski to the class Cycles of finite cycles using the topological tools that we developed. This will partially answer a natural question that was left untouched: what does it mean for a topology to *generalise* a quasi-order? Answering this question will in fact demonstrate how Noetherian spaces are as limited as wqos when talking about preservation theorems.

> **Example 3.3.29.** Every first order sentence that excludes finitely many cycles is equivalent (over Cycles) to an existential sentence.

Proof. Recall that Cycles forms an infinite antichain for $\subseteq_i$. As a consequence, every subset is upwards closed, $(\text{Cycles}, \subseteq_i)$ is not a wqo, and techniques based on minimal models will not really help us here.

However, one can place over Cycles the topology composed of co-finite subsets of Cycles, i.e., the co-finite topology[36].

36: See Definition D.2.1 p. 252

This topology is Noetherian almost by definition. And in particular, every first order sentence that excludes finitely many cycles defines a compact open subset of Cycles with the co-finite topology.

It is an easy check that every co-finite subset of Cycles can be described as an existential sentence.[37] As a consequence, every first order sentence that excludes finitely many cycles is equivalent (over Cycles) to an existential sentence. ■

37: Hint: consider the upward closure of P_n for a sufficiently large n.

Example 3.3.29 was an excuse to introduce a Noetherian topology that does not come from a logic or an ordering. In fact, it will also serve as an example of the connection between Noetherian topologies and well-quasi-orderings.

Given a topology τ over a set X, one can construct the *specialisation preorder* $\leq_\tau$ over τ by defining:

$$x \leq_\tau y \iff \forall U \in \tau, x \in U \Rightarrow y \in U \quad . \tag{3.1}$$

The topology of Example 3.3.29 *generalises* the quasi-order $\subseteq_i$ in the sense that $\subseteq_i$ is the specialisation preorder of the co-finite topology over Cycles.

Notice that the topology $\langle \mathsf{EFO} \rangle_{\mathsf{topo}}$ also has $\subseteq_i$ for specialisation preorder over Cycles. This is a general pattern: a lot of topologies encode the same preorder. Let us briefly clarify the situation hereafter:

Lemma 3.3.30. Let X be a set, and $\leq$ be a quasi-order on X. The collection of topologies over X with $\leq$ as specialisation preorder

1. Is a non-empty complete lattice for inclusion.
2. Has a maximum for inclusion: the *Alexandroff topology* that we write $\mathsf{Alex}(\leq)$, and is obtained by collecting the $\leq$-upwards closed subsets of X.
3. Has a minimum for inclusion: the *Upper topology* that we write $\mathsf{Upper}(\leq)$, and is generated by the sets $\mathsf{X} \setminus \downarrow_{\leq} x$ when $x \in \mathsf{X}$.

Using the vocabulary of Lemma 3.3.30, one can better understand the difference between the co-finite topology over Cycles and the topology $\langle \mathsf{EFO} \rangle_{\mathsf{topo}}$. The former is the Upper topology of $\subseteq_i$ and is Noetherian, while the latter is the Alexandroff topology associated with $\subseteq_i$ and is not Noetherian.

It is known that $\leq$ is a wqo if and only if $\mathsf{Alex}(\leq)$ is Noetherian [45, Proposition 9.7.17]. Since the connection[38] between preservation theorems and Noetherian spaces leverages $\langle \exists \mathsf{F} \rangle_{\mathsf{topo}}$, which precisely generates the Alexandroff topology of $\leq_{\mathsf{F}}$,[39] we are no better off using Noetherian spaces rather than well-quasi-orderings.

38: That is, the Lemma 3.3.28 that states the relativisation of the F-preservation theorem whenever the space is Noetherian.

39: This is the topological analogue of Lemma 3.2.8!

The angry reader might think that all of this subsection on Noetherian spaces and topologies is therefore unnecessary and has been a waste of time. This is only true up to Chapter 6 (Logically Presented Spaces), where we leverage the intimate conviction that compactness is the real key to understand preservation theorems in a more meaningful way. In the meantime, it will make the plan of this manuscript readable.

3.4. Discussion

State of the art. Because the chapter was focused on giving a general definition for preservation theorems, and illustrating the various proof scheme associated with their relativisation, space was lacking for a "big picture" overview of what preservation theorems are known to relativise (or not).

As already mentioned, most preservation theorems (except for the Homomorphism Preservation Theorem) do not relativise to $\mathsf{Fin}(\sigma)$ (see Table 3.3). The actual statement of [83] is that the Homomorphism Preservation Theorem relativises to every co-homomorphism closed subset of $\mathsf{Fin}(\sigma)$. An orthogonal result [7] later refined [20] states that the Homomorphism Preservation Theorem relativises to monotone classes of finite structures $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ such that $\mathcal{C}$ is *closed under disjoint unions*[40] and nowhere dense [20, Theorem 9].

For induced substructures, the landscape is a bit darker. It is known that the Łoś-Tarski theorem relativises to hereditary classes $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ that are closed under disjoint unions and have bounded degree[41] [6, Theorem 6]. The result can be extended to the class of *all* finite structures of tree width less than $k \in \mathbb{N}$ [6, Theorem 8], but it fails for some class of planar graphs even though it has bounded tree width, is monotone and closed under disjoint unions [6, Section 6].

The study of classes of graphs (or structures) that are well-quasi-ordered for $\subseteq_i$ or $\preceq_h$ provide orthogonal results. For instance, the class of structures of tree-depth bounded by 2 is hereditary, closed under disjoint union, but contains structures of unbounded degree: we have seen that this implies the relativisation of the Łoś-Tarski Theorem, and clearly escapes the combinatorial results of the previous paragraph. Similarly, the Homomorphism Preservation Theorem (trivially) relativises to the class of cliques, that is not sparse.

This summary should give the intuition of a landscape with a structure that is not well understood. Recall that the relativisation of a preservation theorem is not a property that transfers to arbitrary subsets (even assuming that those are suitably closed under substructures and disjoint unions), while all the aforementioned properties are, except for the full class of structures with bounded tree width.

40: That is, $\mathfrak{A} \uplus \mathfrak{B} \in \mathcal{C}$ for all $\mathfrak{A} \in \mathcal{C}$ and $\mathfrak{B} \in \mathsf{Fin}(\sigma)$

41: There exists $d \in \mathbb{N}$ such that the maximal degree of vertices $\mathrm{Gaif}(\mathfrak{A})$ is bounded by d for all $\mathfrak{A}$ in $\mathcal{C}$.

Other logics. We have chosen to focus on the queries written in first order logic, but it is not a requirement to talk about preservation theorems. For instance, Datalog and first-order logics with infinite disjunctions are encountered in the literature [81] and [7, Section 7]. We argue that the topological presentation of preservation theorems given in Chapter 6 (Logically Presented Spaces) is (almost) agnostic to the underlying notion of logic, and that the other chapters are either focused on the locality of first order logic, or the study of Noetherian spaces, hence do not apply to other kind of logical frameworks. This explains the opinionated presentation of preservation theorems that has been given.

However, there are interesting questions that are being answered as this manuscript is getting written. For instance, when restricting fragments of first-order logic that are not built using the notion of quantifier rank, but by restricting the number of used variables [10].

[10]: Bova and Chen (2019), 'How Many Variables are Needed to Express an Existential Positive Query?'

Stone Duality. It is worth noting that the short introduction to topology and its relationship with logic does come from a long line of results and is not anecdotal in any way. We redirect the reader to the book of [61], or a dedicated chapter in the more recent book [26, Chapter 3]. While it is a comparative advantage to know some of this theory, it is not necessary to understand the technical developments in this manuscript.

[61]: Johnstone (1982), *Stone Spaces*

[26]: Dickmann, Schwartz and Tressl (2019), *Spectral Spaces*

Sequential Compactness. Even though the real line was used to illustrate the basic notions of compactness, we chose to avoid the notion of converging sequences. Most (if not all) of the spaces present in this thesis will have a countable set of open subsets, and the definition of compactness via converging sequences or open covers will coincide. However, the theorems of Chapter 6 can be applied to general topological spaces, where the two notions may differ.

References

[1] Parosh Aziz Abdulla, Karlis Čerāns, Bengt Jonsson and Yih-Kuen Tsay. 'General decidability theorems for infinite-state systems'. In: *Proceedings of LICS'96*. IEEE, 1996, pp. 313–321. DOI: 10.1109/LICS.1996.561359 (cit. on p. 44).

[4] Miklós Ajtai and Yuri Gurevich. 'Monotone versus positive'. In: *Journal of the ACM* 34 (1987), pp. 1004–1015. DOI: 10.1145/31846.31852 (cit. on pp. 35, 37, 113).

[5] Miklós Ajtai and Yuri Gurevich. 'Datalog vs first-order logic'. In: *Journal of Computer and System Sciences* 49 (1994), pp. 562–588. DOI: 10.1016/S0022-0000(05)80071-6 (cit. on pp. 35, 37, 113).

[6] Albert Atserias, Anuj Dawar and Martin Grohe. 'Preservation under extensions on well-behaved finite structures'. In: *SIAM Journal on Computing* 38 (2008), pp. 1364–1381. DOI: 10.1137/060658709 (cit. on pp. 37, 40, 47, 55, 64, 67, 68, 74, 97, 98, 117, 121, 127).

[7] Albert Atserias, Anuj Dawar and Phokion G. Kolaitis. 'On preservation under homomorphisms and unions of conjunctive queries'. In: *Journal of the ACM* 53 (2006), pp. 208–237. DOI: 10.1145/1131342.1131344 (cit. on pp. 37, 40, 47, 55, 74).

[10] Simone Bova and Hubie Chen. 'How Many Variables are Needed to Express an Existential Positive Query?' In: vol. 63. 2019, pp. 1573–1594. DOI: 10.1007/s00224-018-9884-z. URL: https://doi.org/10.1007/s00224-018-9884-z (visited on 06/06/2023) (cit. on p. 56).

[11] Robert Brignall, Michael Engen and Vincent Vatter. 'A Counterexample Regarding Labelled Well-Quasi-Ordering'. In: *Graphs and Combinatorics* 34 (2018), pp. 1395–1409. DOI: 10.1007/s00373-018-1962-0 (cit. on p. 46).

[12] Florent Capelli, Arnaud Durand, Amélie Gheerbrant and Cristina Sirangelo. *A simple counter example of Lyndon's Theorem in the finite*. 2020 (cit. on p. 35).

[15] Chen Chung Chang and H. Jerome Keisler. *Model Theory*. Vol. 73. Elsevier, 1990 (cit. on pp. i, 29, 32–34, 68).

[16] Yijia Chen and Jörg Flum. 'Forbidden Induced Subgraphs and the Łoś-Tarski Theorem'. In: *2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*. 2021, pp. 1–13. DOI: 10.1109/LICS52264.2021.9470742 (cit. on pp. 37, 39, 98, 113).

[19] Jean Daligault, Michael Rao and Stéphan Thomassé. 'Well-Quasi-Order of Relabel Functions'. In: *Order* 27 (2010), pp. 301–315. DOI: 10.1007/s11083-010-9174-0. URL: https://doi.org/10.1007/s11083-010-9174-0 (cit. on pp. 37, 40, 44–46, 179, 235).

[20] Anuj Dawar. 'Homomorphism preservation on quasi-wide classes'. In: *Journal of Computer and System Sciences* 76 (2010), pp. 324–332. DOI: 10.1016/j.jcss.2009.10.005 (cit. on pp. 47, 55, 74).

[22] Anuj Dawar and Abhisekh Sankaran. 'Extension Preservation in the Finite and Prefix Classes of First Order Logic'. In: *29th EACSL Annual Conference on Computer Science Logic (CSL 2021)*. Ed. by Christel Baier and Jean Goubault-Larrecq. Vol. 183. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2021, 18:1–18:13. DOI: 10.4230/LIPIcs.CSL.2021.18. URL: https://drops.dagstuhl.de/opus/volltexte/2021/13452 (cit. on pp. 35, 37).

[25] Alin Deutsch, Alan Nash and Jeffrey B. Remmel. 'The chase revisited'. In: *Proceedings of PODS'08*. 2008, pp. 149–158. DOI: 10.1145/1376916.1376938 (cit. on pp. 27, 33, 35, 234, 235).

[26] Max Dickmann, Niels Schwartz and Marcus Tressl. *Spectral Spaces*. Vol. 35. Cambridge University Press, 2019. DOI: 10.1017/9781316543870.003 (cit. on pp. 56, 146–148, 174, 239, 253).

[27] Guoli Ding. 'Subgraphs and well-quasi-ordering'. In: *Journal of Graph Theory* 16 (1992), pp. 489–502. DOI: 10.1002/jgt.3190160509 (cit. on pp. 37, 40, 44, 45, 122).

[35] Alain Finkel and Philippe Schnoebelen. 'Well-structured transition systems everywhere!' In: *Theoretical Computer Science* 256 (2001), pp. 63–92. DOI: 10.1016/S0304-3975(00)00102-X. URL: https://linkinghub.elsevier.com/retrieve/pii/S030439750000102X (visited on 19/06/2023) (cit. on p. 44).

[44] Jean Goubault-Larrecq. 'Noetherian Spaces in Verification'. In: *Automata, Languages and Programming*. Ed. by Samson Abramsky, Cyril Gavoille, Claude Kirchner, Friedhelm Meyer auf der Heide and Paul G. Spirakis. Springer Berlin Heidelberg, 2010, pp. 2–21. DOI: 10.1007/978-3-642-14162-1_2 (cit. on pp. 53, 139, 178).

[45] Jean Goubault-Larrecq. *Non-Hausdorff Topology and Domain Theory*. Vol. 22. Cambridge University Press, 2013. DOI: 10.1017/cbo9781139524438 (cit. on pp. 3, 50, 52–54, 139, 159, 178, 179, 185, 186, 192, 193, 203, 208, 224, 239, 254, 255).

[53] Jean Goubault-Larrecq, Monika Seisenberger, Victor L. Selivanov and Andreas Weiermann. 'Well Quasi-Orders in Computer Science (Dagstuhl Seminar 16031)'. In: *Dagstuhl Reports* 6 (2016), pp. 69–98. DOI: 10.4230/DagRep.6.1.69. URL: https://doi.org/10.4230/DagRep.6.1.69 (cit. on pp. 44, 178).

[56] Yuri Gurevich. 'Toward logic tailored for computational complexity'. In: *Computation and Proof Theory, Proceedings of LC'84*. Vol. 1104. Springer, 1984, pp. 175–216. DOI: 10.1007/BFb0099486 (cit. on p. 35).

[59] Graham Higman. 'Ordering by divisibility in abstract algebras'. In: *Proceedings of the London Mathematical Society* 3 (1952), pp. 326–336. DOI: 10.1112/plms/s3-2.1.326 (cit. on pp. 44, 158, 178, 181).

[60] Ehud Hrushovski, Joël Ouaknine, Amaury Pouly and James Worrell. 'Polynomial Invariants for Affine Programs'. In: (2018), pp. 530–539. DOI: 10.1145/3209108.3209142. URL: https://dl.acm.org/doi/10.1145/3209108.3209142 (cit. on p. 53).

[61] Peter T Johnstone. *Stone Spaces*. Vol. 3. Cambridge university press, 1982 (cit. on p. 56).

[64] Igor Kříž and Robin Thomas. 'On well-quasi-ordering finite structures with labels'. In: *Graphs and Combinatorics* 6 (1990), pp. 41–49 (cit. on p. 44).

[65] Joseph B. Kruskal. 'The theory of well-quasi-ordering: A frequently discovered concept'. In: *Journal of Combinatorial Theory, Series A* 13 (1972), pp. 297–305. DOI: 10.1016/0097-3165(72)90063-5 (cit. on pp. 44, 158, 178, 179, 181).

[66] Denis Kuperberg. '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, 2021, pp. 1–13. DOI: 10.1109/LICS52264.2021.9470602 (cit. on pp. 37, 39, 68, 98, 113, 137).

[70] Aliaume Lopez. 'Preservation Theorems Through the Lens of Topology'. In: *Proceedings of CSL'21*. Vol. 183. 2021, 32:1–32:17. DOI: 10.4230/LIPIcs.CSL.2021.32 (cit. on pp. 63, 138, 148, 174, 175).

[71] Aliaume Lopez. 'When Locality Meets Preservation'. In: *LICS'22* (2022), pp. 1–14. DOI: 10.1145/3531130.3532498. URL: https://doi.org/10.1145/3531130.3532498 (cit. on pp. 63, 67, 117).

[72] Aliaume Lopez. 'Fixed Points and Noetherian Topologies'. In: *FOSSACS'23* (2023), pp. 1–14. DOI: 10.1145/3531130.3532498. URL: https://doi.org/10.1145/3531130.3532498 (cit. on pp. 63, 180, 191).

[75] Eric C Milner. 'Basic wqo-and bqo-theory'. In: *Graphs and order*. Springer, 1985, pp. 487–502 (cit. on p. 53).

[77] Jaroslav Nešetřil and Patrice Ossona de Mendez. 'On nowhere dense graphs'. In: *European Journal of Combinatorics* 32 (2011), pp. 600–617. DOI: 10.1016/j.ejc.2011.01.006. URL: https://www.sciencedirect.com/science/article/pii/S0195669811000151 (cit. on pp. 46, 121, 259).

[80] R. Rado. 'Partial well-ordering of sets of vectors'. In: *Mathematika* 1 (1954), pp. 89–95. DOI: 10.1112/S0025579300000565 (cit. on p. 53).

[81] Eric Rosen. 'Finite model theory and finite variable logics'. PhD thesis. 1995 (cit. on pp. 35, 55).

[82] Eric Rosen. 'Some aspects of model theory and finite structures'. In: *Bulletin of Symbolic Logic* 8 (2002), pp. 380–403. DOI: 10.2178/bsl/1182353894 (cit. on p. 35).

[83] Benjamin Rossman. 'Homomorphism preservation theorems'. In: *Journal of the ACM* 55 (2008), 15:1–15:53. DOI: 10.1145/1379759.1379763 (cit. on pp. i, 24, 35, 37, 38, 40, 42, 55, 74, 175, 177, 235).

[84] Abhisekh Sankaran, Bharat Adsul and Supratik Chakraborty. 'A generalization of the Łoś-Tarski preservation theorem over classes of finite structures'. In: *International Symposium on Mathematical Foundations of Computer Science*. 2014, pp. 474–485. DOI: 10.1007/978-3-662-44522-8_40 (cit. on p. 37).

[86] Abhisekh Sankaran, Bharat Adsul, Vivek Madan, Pritish Kamath and Supratik Chakraborty. 'Preservation under substructures modulo bounded cores'. In: *International Workshop on Logic, Language, Information, and Computation*. 2012, pp. 291–305. DOI: 10.1007/978-3-642-32621-9_22 (cit. on pp. 34, 35).

[89] Alexei P. Stolboushkin. 'Finitely monotone properties'. In: (1995), pp. 324–330. DOI: 10.1109/LICS.1995.523267 (cit. on pp. 35, 113, 137).

[90] William W. Tait. 'A counterexample to a conjecture of Scott and Suppes'. In: *Journal of Symbolic Logic* 24 (1959), pp. 15–16. DOI: 10.2307/2964569 (cit. on pp. i, 35, 37, 39, 97, 100, 113, 137).

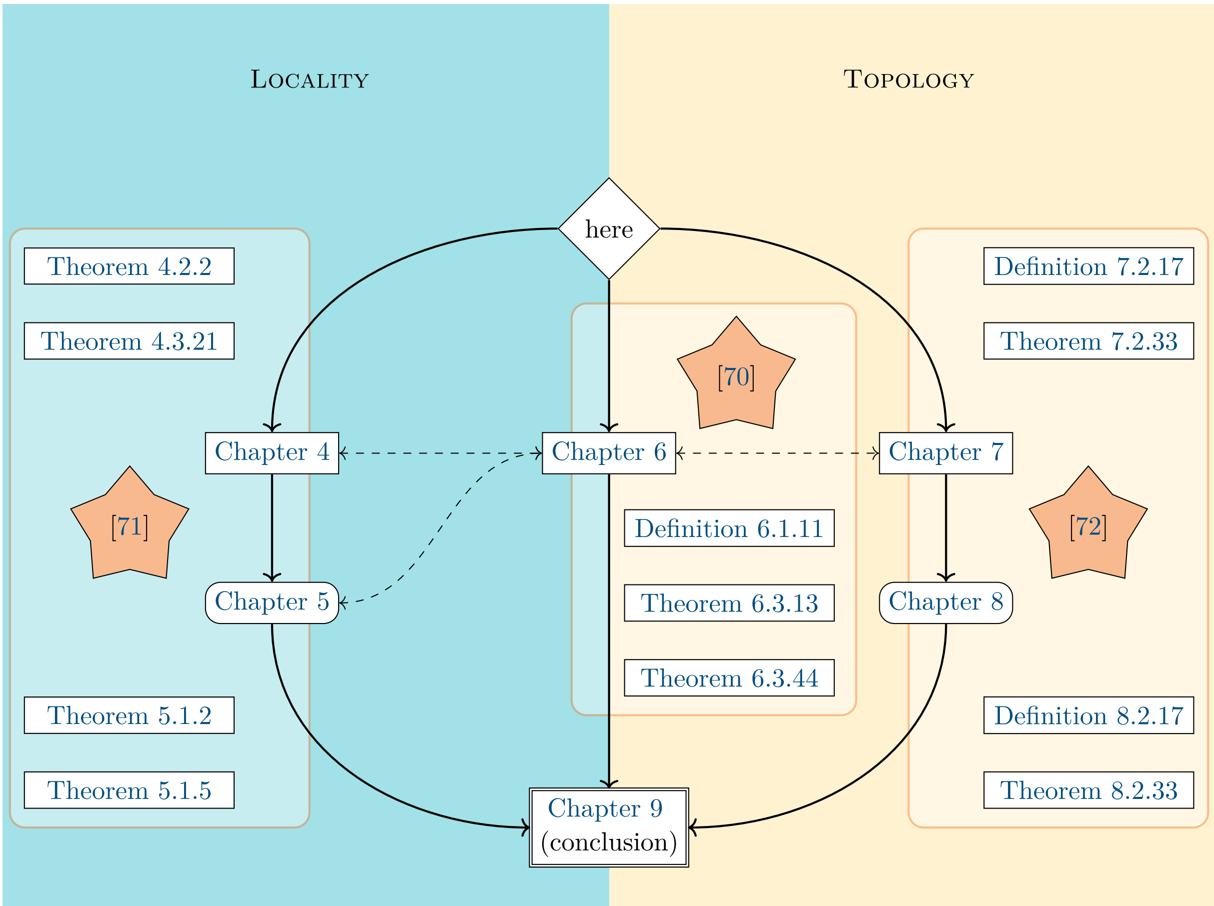
Part II.

Preservation Theorems in Finite Model Theory

Layout of the Contributions. We provide hereafter a layout of the manuscript that may be helpful to skim the document, and prevents from reading the chapters in an order that does not respect the dependencies. Even though every chapter is self-contained enough to be printed and read independently of the others, the reading process will be smoother if one follows the inter-chapter dependencies.

The Hasse diagram of the inter-chapter dependencies is represented in bold directed arrows. Because the document tries to not repeat itself, there are also weak bidirectional interaction between chapters that are represented in dashed arrows. Furthermore, we selected for each chapter definitions and theorems that are both crucial and representative of the chapter's technical content. We also distinguished introduction chapters (with squared corners), that will be more theoretical, from application chapters (with rounded corners) that have more practical considerations. Finally, we made a separation between chapters that will mostly employ topological tools (Chapters 6 to 8) and those that will focus on the locality of first order logic (Chapters 4 to 5).

On the background layer of the diagram, one can also find boxes that group results and chapters together, following the three separate paths from this page to the conclusion: these correspond to the three published paper this manuscript is based upon, for which the references are placed inside stars with an orange background. Beware that the definitions and theorems presented in this manuscript can be formulated in a way that does not correspond to the associated paper. This is because the chosen ordering (from Chapter 4 to Chapter 8) does not respect the chronological ordering of the papers, that would lead to the sequence: 6–4–5–7–8. Moreover, the order of publication of the chapters does not reflect the order in which the main ideas were developed, which would most likely lead to the sequence: 6–7–5–4–8.



Content of the Chapters. Let us now give a brief description of each chapter, together with a justification of the distinguished definitions and results that were selected in the layout of the manuscript.

In Chapter 4 (Locality and Preservation), we try pushing the existing approaches to proving the relativisation of preservation theorems in the finite to the extreme. This means understanding how the locality of first-order logic can play a role when the compactness theorem fails. Because this chapter focuses on the existing tools, no interesting new definition is distinguished. However, two main results are obtained:

1. In Theorem 4.2.2, we introduce a positive variant of the Gaifman normal form, and prove that it is equivalent (over any class of structures) to the fragment of existential local sentences.
   This theorem has its own interest as a new variation around Gaifman locality. However, it is distinguished because it provides an intuition on the proof schemes developed when studying relativisation properties: half of the work is often to remove negations from a Gaifman normal form, which precisely corresponds to building an existential local sentence. Only then, in a second proof, one moves from an existential local sentence (in $\exists\mathsf{FOLoc}$) to a sentence of the desired fragment (in $\exists\mathsf{F}$).
2. In Theorem 4.3.21, we prove that there exists a first order sentence $\varphi \in \mathsf{FO}[\sigma]$ that is preserved under FOLoc-embeddings over $\mathsf{Fin}(\sigma)$, but that is not equivalent to an $\exists\mathsf{FOLoc}$ sentence.
   This result has two sides. On the one hand, it continues the long tradition of non-relativisation proofs in the finite, by treating the case of the FOLoc-preservation theorem. On the other hand, it means that Theorem 4.2.2 is not enough to explain why some relativisations occur in the finite, and more precisely, that the following line of reasoning does not hold: a sentence φ that is preserved under QF-embeddings is *a fortiori* preserved under FOLoc-embeddings, hence equivalent to an existential local sentence, and an existential local sentence preserved under QF-embeddings must be equivalent to an existential sentence.

In Chapter 5 (A Local-to-Global Preservation Theorem), we will continue the study of preservation theorems via locality that started with Chapter 4 (Locality and Preservation). While the latter left a bittersweet aftertaste, the goal here is to demonstrate that, in one specific configuration, one can leverage the locality of first order logic to prove that the Łoś-Tarski Theorem relativises to a class $\mathcal{C}$ of finite structures.

1. In Theorem 5.1.2, we prove that under mild assumptions on a class $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$, the Łoś-Tarski Theorem relativises to $\mathcal{C}$ if and only if it locally relativises to $\mathcal{C}$ (for a suitable definition of "locally satisfying a property" that we leave out in this high level description).
   This theorem is a way to leverage the knowledge gained on existential local sentences in the previous chapter to provide a suitable explanation of some relativisation results in the literature (for instance, [6]). Interestingly enough, the proof of Theorem 5.1.2 *almost* follows the one that was invalidated by Theorem 4.3.21, at the price of extra assumptions and a deeper understanding of the induced substructure ordering.
2. In Theorem 5.1.5, we illustrate how Theorem 5.1.2 actually leads to new relativisation results for the Łoś-Tarski Theorem.
   While this theorem is simply a collection of examples showing the strict implication of properties, it is distinguished because it proves the validity of the "local approach" to relativisation results. The result is powerful enough to go way beyond what is known in the literature for hereditary classes of finite structures closed under disjoint unions.

In Chapter 6 (Logically Presented Spaces), we introduce an algebraic and topological approach to preservation theorems, by pairing the definability notions from logic together with the topological notions of Noetherian spaces. This sparks the study of triples $(\mathsf{X}, \tau, \mathcal{B})$ where τ is a topology and $\mathcal{B}$ a Boolean subalgebra of $\mathcal{P}(\mathsf{X})$. The main motivation for this abstract study is to find a way around the non-compositionality of preservation theorems and the complexity of proving their relativisation.

1. In Definition 6.1.11, we provide the topological counterpart to preservation theorems, namely logically presented pre-spectral spaces (abbreviated as lpps).
   This definition is distinguished because it provides a new way to think about preservation theorems in terms of topological compactness, in a way that connects them more clearly to Noetherian spaces (which was already done in the introduction), but also to more exotic spaces, such as spectral spaces. Furthermore, it abstracts the considered logic(s) as a Boolean subalgebra of the class of structures considered, opening the door to the study of non-FO preservation theorems, without any change in definitions.
2. In Theorem 6.3.13, we prove that the product of two such spaces remains an lpps.
   This result is distinguished because it is a first non-trivial compositional result about lpps (and therefore preservation theorems). Furthermore, with a bit of work, one *compositionally* obtains classes that are out of reach from the local approach advocated in Chapters 4 and 5.
3. In Theorem 6.3.44, we go beyond simple products and tackle the case of coloured classes of structures. Namely, we prove that if $\mathcal{C}$ is a ∞-wqo, and X is a lpps, then $\mathcal{C} \rtimes_{\mathsf{F}} \mathsf{X}$ is a lpps.
   This result is distinguished because it pushes the proof scheme of Theorem 6.3.13 to its limits and provide a truly powerful compositional technique. Furthermore, it is important to notice that it provides a positive answer to the lpps-variant of a conjecture by Pouzet, asking whether ∞-wqo (that is, being wqo for any colouring with finitely many colours) implies "wqo-wqo" (that is, being wqo when coloured by any wqo).

In Chapter 7 (Topology expanders and Noetherian Topologies), we try to understand what makes the limit constructions of lpps so complicated. In the continuity of the algebraic and topological approach that started in "Chapter 6 (Logically Presented Spaces)", we want to provide a systematic way to produce limits of pre-spectral spaces. It turns out that the question is already non-trivial when restricted to the subclass of Noetherian spaces, and this chapter focuses on this simpler setting.

1. In Definition 7.2.17, we define topology expanders as "well-behaved topology constructor."
   This definition is distinguished because it is an abstraction that captures all the examples and non-examples of limit constructions in Noetherian spaces that I know of. Furthermore, the interesting part of the definition is side-stepping from the study of inductive construction of spaces, to inductive constructions of a topology over a given (fixed) space. This change of perspective is what makes the definition sound, and gives such a flexibility to topology expanders.
2. In Theorem 7.2.33, we prove that the least fixed point of a topology expander is always Noetherian.
   This theorem is distinguished because it demonstrates the correctness of Definition 7.2.17. Furthermore, its proof "generalises" as much as possible what is known as the *topological minimal bad sequence argument* (a "meta" proof scheme that usually has to be instantiated whenever a new result is needed), and wraps it into a theorem with a simple interface: all that is needed is a topology expander.

In Chapter 8 (Inductive Constructions), we will leverage the master Theorem 7.2.33 from Chapter 7 (Topology expanders and Noetherian Topologies) to construct families of Noetherian spaces, and in particular recover the known Noetherian topologies over finite words, finite trees, infinite words, ordinal length words using Theorem 7.2.33. Furthermore, we will provide a generic way to build a topology expander from a suitable endofunctor, effectively answering to the original question of inductively defined *spaces* (and not only *topologies*).

1. In Definition 8.2.17, we define the divisibility expander associated with a (suitable) analytic functor over Set.
   This definition is distinguished because it bridges the framework developed in Chapter 7 (Topology expanders and Noetherian Topologies), with the original motivation of inductively defining Noetherian spaces.
2. In Theorem 8.2.33, we prove that the divisibility expander not only leads to the correct limit topology on selected examples, but is a (correct) topological generalisation of a whole theory over wqos. Namely, we prove a tight correspondence between the divisibility ordering over an inductively defined wqo, and the divisibility topology obtained as the least fixed point of the associated divisibility expander.
   This result is distinguished because it grounds the framework of topology expanders: it correctly generalises what can be done over wqos. Furthermore, it provides an answer to a “lack of canonicity” that arises from the topological setting: over finite words, there is no finest topology that is Noetherian, but our framework provides one that is *naturally* derived from the inductive definition, and *correctly generalises* the subword ordering. In this specific example, it provides a justification as to why the subword topology can be called “the topological generalisation of” the subword ordering.

Locality and Preservation 4.

Outline of the chapter

The goal of this chapter is to push the existing approaches to proving the relativisation of preservation theorems in the finite to the extreme. This means understanding how the locality of first-order logic can play a role when the compactness theorem fails.

Goals of the chapter

At the end of this chapter, I hope you will be convinced that there is a deep connection between locality and preservation under extensions, that is embodied by the fragment of existential local sentences.

Genesis. This chapter is stating results that are mostly coming from [71]. This research started during the pandemic, by looking at a paper from Atserias, Dawar and Grohe studying the relativisation of the Łoś-Tarski Theorem to classes of finite structures with bounded degree [6, Theorem 4.3]. The technical proof was mysterious, but made apparent the need for a construction using disjoint unions of structures, the preference for hereditary classes, and some kind of elimination of negations in the Gaifman normal form. This started the search for a positive variant of the Gaifman normal form, with the hope that the original proof of Atserias, Dawar and Grohe could be factorised as follows:

[6]: Atserias, Dawar and Grohe (2008), 'Preservation under extensions on well-behaved finite structures'

1. first order sentences that are preserved under extensions can be put in *positive* Gaifman normal form, and
2. first order sentences in *positive* Gaifman normal form that are preserved under extensions (on well-behaved classes of finite structures) are equivalent to existential sentences (over the same class).

The benefits from this separation of concerns between a locality argument (item 1) and a combinatorial argument (item 2) would be multiple. We expected that the two lemmas/theorems should have simpler (or at least clearer proofs), and that both lemmas could be generalised to obtain the relativisation of the Łoś-Tarski theorem to new classes of structures that are "less well-behaved."

Even though an existential Gaifman normal form had already been introduced [55], it did not correspond to my needs to factor out the proof of Atserias, Dawar and Grohe. It was hinted in [87] that the fragment corresponding to the notion of "positive" Gaifman normal form should be existential local sentences, but the formal statement was lacking.

[55]: Grohe and Wöhrle (2004), 'An existential locality theorem'

[87]: Schwentick and Barthelmann (1999), 'Local normal forms for first-order logic with applications to games and automata'

Eventually, the equivalence between existential local sentences and the positive Gaifman normal form (see Theorem 4.2.2) was proven in [71]. This reignited the hope to find a generic "localisation procedure" for preservation theorems. Unfortunately, the proof techniques employed

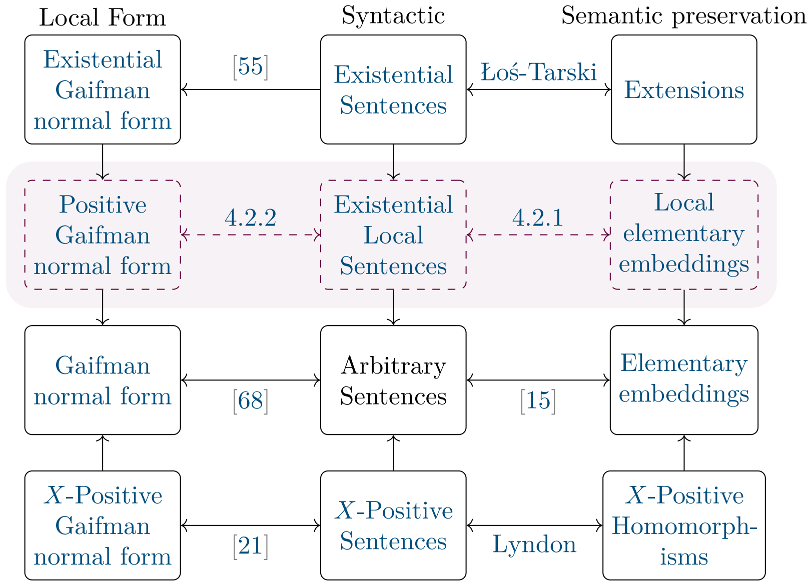


Figure 4.1.: Comparison of the expressiveness of different fragments of $\mathsf{FO}[\sigma]$ over *general structures*. Local forms correspond to variants of the Gaifman normal form, while syntactic fragments correspond to restrictions on the syntax of sentences. Finally, "semantic preservation" correspond to the kind of embeddings that characterises the fragment through a preservation theorem. Dashed boxes correspond to the notions introduced in this chapter, double ended arrows represent equivalence, while single headed arrows represent strict implication of properties. For instance, having an existential Gaifman normal form strictly implies having a positive Gaifman normal form, which strictly implies having a Gaifman normal form, the latter being always true. Similarly, preservation under extensions strictly implies preservation under local elementary embeddings. Finally, notice that being equivalent to an existential sentence strictly implies having an equivalent existential Gaifman normal form, which is the only strict implication in terms of horizontal boxes.

[66]: Kuperberg (2021), 'Positive First-order Logic on Words'

1: and providing undecidability of the associated decision problem

by [66] to prove that Lyndon's Positivity Theorem fails for finite words[1] can be generalised to conclude that Item 1 cannot be proven without extra assumptions, which put the research program to a (temporary) stop.

Contributions. The main contribution of this chapter is the correspondence between existential local sentences and the positive Gaifman normal form stated in Theorem 4.2.2, together with the light shed on the central role of the class of existential local sentences in preservation theorems, as shown in Figure 4.1. The role of disjoint unions in the original proof of [6] is The presence of existential local sentences, for which the quasi-order $\leq_{\mathsf{F}}$ property is that their set of models is closed under disjoint unions, justifies the presence of disjoint unions in [6].

The second main contribution is the *quantitative* study of the relativisation of Figure 4.1 in the finite, via a parametrised family of quasi-orders $\Rrightarrow_q^{r,k}$. In particular, we prove that relativisation fails in non-trivial cases, except for one…That is the core of Chapter 5 (A Local-to-Global Preservation Theorem). The failure results are accompanied by undecidability properties, definitely closing the door to a generic factorisation of the proofs into a first part using a positive locality result, and a second part relying on combinatorics of the local neighbourhoods of the considered structures.

We start the chapter with a gentle (and self contained) introduction to the locality of first-order logic in Section 4.1. The curious reader might look at the logic cheatsheet in Chapter E (Logic Cheat Sheet). Most of this section can be found (in some way) in a classical book on the subject such as [68]. The added value of this section is the introduction of notations that will be handy throughout the document, together

[68]: Libkin (2012), *Elements of finite model theory*

with the direct connection with preservation theorems, namely via Example 4.1.16.

In Section 4.2, we refine our understanding of the locality of first-order logic to study the connection between existential local sentences, and a positive version of the Gaifman normal form, culminating at Theorem 4.2.2. Aside from the main theorem, we also use an elegant formalisation of the compositional techniques for first-order logic[2] in terms of "typed formulas", based on discussions with Thomas Colcombet.[3] This formalism will be of particular use when studying the composition of preservation theorems in Chapter 6 (Logically Presented Spaces).

2: Understand, Feferman-Vaught like theorems

3: To the best of my knowledge, this is the first time such a presentation has been written down formally.

In Section 4.3, we study the preservation theorem associated with existential local sentences, which provides in Figure 4.1 an intermediate "order free" layer between the Łoś-Tarski Theorem and arbitrary formulas. However, this preservation theorem does not relativise to the finite (see Theorem 4.3.21), which puts a stop to our original idea of a "localisation procedure." However, a meticulous study of the parametrised version of the FOLoc-preservation theorem in the finite will open the door to a relativisation in a very specific case, which is at the heart of Chapter 5 (A Local-to-Global Preservation Theorem).

4.1. Locality of first order logic

Recall that the Gaifman graph[4] $\mathrm{Gaif}(\mathfrak{A})$ of a relational structure $\mathfrak{A}$ over a finite relational signature σ has vertices $\mathfrak{A}$ and an edge (a, b) whenever both a and b appear in a relation of $\mathfrak{A}$. For an example, see Figure 2.1 and Figure 2.3.

4: See Definition 2.2.12 p. 17

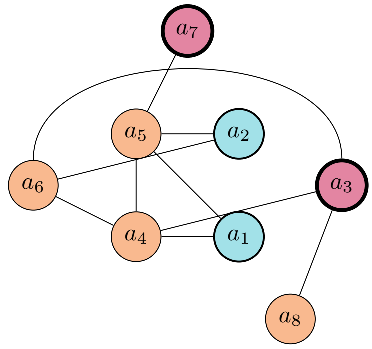


Figure 4.2.: A finite graph with two selected nodes (in red ■), two nodes at shortest path distance 1 of the selected nodes (in orange ■), and two nodes at shortest path distance 2 of the selected nodes (in blue ■).

> **Definition 4.1.1.** Let G be an undirected graph. The *shortest path distance* $\mathrm{d_{spd}}(a, b)$ between two nodes $(a, b) \in G$ is the minimum length of a path from a to b in G, and $+\infty$ if such a path does not exist.

This notion of distance can be lifted to relational structures by defining for a structure $\mathfrak{A}$ the distance $\mathrm{d}_{\mathfrak{A}}(a, b)$ as the distance $\mathrm{d_{spd}}(a, b)$ in the Gaifman graph $\mathrm{Gaif}(\mathfrak{A})$ of $\mathfrak{A}$. This notion of distance allows us to define "closed balls" inside graphs and structures, that are described by a centre a and a radius $r > 0$. Formally, the *r-local neighbourhood* of an element $a \in \mathfrak{A}$, written $\mathcal{N}_{\mathfrak{A}}(a, r)$, is the set $\{ b \in \mathfrak{A} : \mathrm{d}_{\mathfrak{A}}(a, b) \leq r \}$ of elements of $\mathfrak{A}$ that are at distance at most r from a. We generalise the notion to tuples $\vec{a}$ of elements by letting $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r) \overset{\text{def}}{=} \bigcup_{a \in \vec{a}} \mathcal{N}_{\mathfrak{A}}(a, r)$. For a graphical representation, see Figure 4.2 and Table 4.1. Beware that this union is not required to be disjoint.

Table 4.1.: The different neighbourhoods around the two selected points a_7, a_3 in the graph G described in Figure 4.2.

r	$\mathcal{N}_G(a_7 a_3, r)$
0	$\{ a_7, a_3 \}$
1	$\{ a_7, a_3, a_4, a_5, a_6 \}$
2	$\{ a_7, a_3, a_4, a_5, a_6, a_1, a_2 \}$

Slightly abusing notations, we identify the set $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r)$ of $\mathrm{dom}(\mathfrak{A})$ with the corresponding induced substructure of $\mathfrak{A}$, so that $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r) \subseteq_i \mathfrak{A}$. To describe the local behaviour of a given structure at a given radius r using tuples of at most k elements, one can build the set $\mathsf{Local}(\mathfrak{A}, r, k)$ defined as $\{ \mathcal{N}_{\mathfrak{A}}(\vec{a}, r) : \vec{a} \in \mathfrak{A}^k \}$. For instance, we have computed $\mathsf{Local}(G, 2, 1)$ for the graph G of Figure 4.2 in Table 4.2. We will also write $\mathsf{Local}(\mathcal{C}, r, k)$ for the set $\bigcup_{\mathfrak{A} \in \mathcal{C}} \mathsf{Local}(\mathfrak{A}, r, k)$.

Table 4.2.: The list of 2-local neighbourhoods collected in $\mathsf{Local}(G, 2, 1)$, where the graph G described in Figure 4.2.

Neighbourhoods
$\{ a_3, a_4, a_6, a_8 \}$
$\{ a_1, a_2, a_4, a_5, a_7 \}$
$\{ a_1, a_2, a_3, a_4, a_5, a_6, a_8 \}$
$\{ a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8 \}$
$\{ a_1, a_2, a_3, a_4, a_5, a_6, a_7 \}$

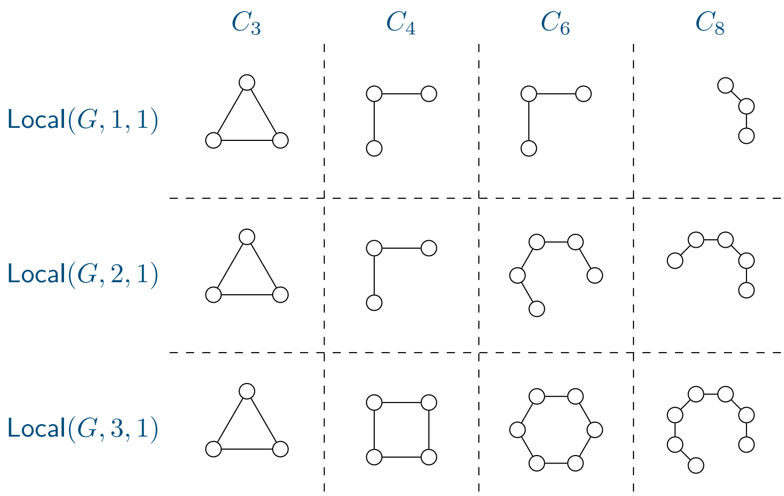


Figure 4.3.: Local neighbourhoods Local$(G, r, 1)$ of cycles C_3, C_4, C_6 and C_8, computed for $r \in \{1, 2, 3\}$. Notice that Local$(C_n, r, 1)$ is a singleton for all $n \geq 3$, $r \geq 1$.

First, let us notice that if $\mathfrak{A}$ is a finite structure and the parameters k and r are large enough, one can always find the structure $\mathfrak{A}$ inside Local$(\mathfrak{A}, r, k)$. Let us briefly note that $\mathfrak{A}$ can either be obtained using $r = 0$ and $k = |\mathfrak{A}|$, or using $r = |\mathfrak{A}|$ and k equal to the number of connected components in Gaif$(\mathfrak{A})$, i.e., $\mathfrak{A} \in$ Local$(\mathfrak{A}, 0, |\mathfrak{A}|)$ and $\mathfrak{A} \in$ Local$(\mathfrak{A}, |\mathfrak{A}|, k)$ where k is the number of connected components in Gaif$(\mathfrak{A})$. While this is not yet of importance, we will see later on that this has an impact in terms of the expressiveness of (local) first order sentences.

Let us illustrate how this local behaviour applies to various classes of structures. For that, we recall that C_n is a finite undirected simple cycle with n vertices, P_n is a finite undirected path with n vertices, and their collection is respectively written Cycles and Paths. To add new examples in terms of undirected graphs, we write K_n to denote an undirected clique of n vertices, and collect the cliques in a set Cliques.

Example 4.1.2. Let $n > 3$, $r > 0$ and $k > 0$. Then, Local(C_n, r, k) contains a cycle if and only if $(2r + 1) \times k \geq n$.

Example 4.1.3. Let $r > 0$ and $k > 0$, and $m, n > (2r + 1) \times k$. Then, Local$(C_m, r, k) =$ Local(C_n, r, k).

Example 4.1.4. Let $r > 0$ and $k > 0$, and $m > n > 0$. Then, the sets Local(K_m, r, k) and Local(K_n, r, k) are disjoint.

Example 4.1.4 shows that even though cliques exhibit very regular behaviours in terms of neighbourhoods, our approach to locality (that considers the local neighbourhoods inside the structures) separates all cliques even with fixed parameters. This should be understood as the fact that local neighbourhoods over approximates the complexity of the structures. To further demonstrate this fact, remark that the class of cliques is in (quantifier-free definable) bijection with the class of independent sets. Let us write, I_n for the (undirected) graph with n

vertices and no edges. While independent sets have the same first-order complexity as cliques, their neighbourhoods behave quite differently as shown below.

Example 4.1.5. Let $r > 0$ and $k > 0$, and $m > n > 0$. Then, the local behaviours satisfy $\mathsf{Local}(I_m, r, k) \supseteq \mathsf{Local}(I_n, r, k)$.

Rather than changing the definition of locality so that cliques and independent sets have the same local behaviour,[5] let us take the more pragmatic approach of restricting our attention to a so-called "local" fragment of first order logic.

5: Which while non-trivial, has been studied in terms of differential games, see [38].

Definition 4.1.6. A first-order formula $\varphi(\vec{x})$ is an *r-local formula* if its evaluation over a structure $\mathfrak{A}$ and a tuple $\vec{a} \in \mathfrak{A}$ only depends on the r-local neighbourhood of $\vec{a}$ in $\mathfrak{A}$. Formally, given $\mathcal{C} \subseteq \mathsf{Struct}(\sigma)$, a first-order formula $\varphi(\vec{x})$ is *r-local* if and only if the following holds:

$$\forall \mathfrak{A} \in \mathcal{C}, \forall \upsilon \colon \vec{x} \to \mathfrak{A}, \mathfrak{A}, \upsilon \models \varphi \text{ if and only if } \mathcal{N}_{\mathfrak{A}}(\upsilon(\vec{x}), r), \upsilon \models \varphi \quad .$$

In order to gain a little intuition about what r-local formulas can express, let us provide examples and non-examples of local formulas.

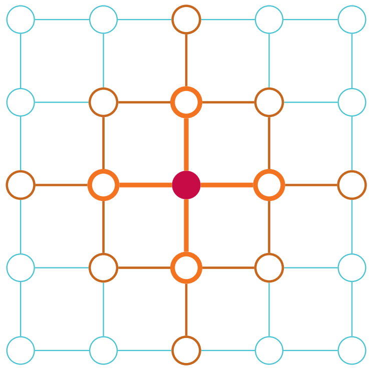


Figure 4.4.: Illustration of the 2-local neighbourhood of a point inside a grid. On this picture, one can understand that the first order formula $\forall y.E(x, y)$ holds at a given point $x \mapsto a$ if and only if the 2-local neighbourhood of a is a grid.

Example 4.1.7. Let Grids be the class of finite grids. Then, the formula $\forall y.E(x, y)$ is a 2-local formula around x over Grids.

Example 4.1.8. The $\exists y.R(x, y)$ is a 1-local formula, while $\forall y.R(x, y)$ is not a local formula. Any r-local sentence (formula with no free variables) is equivalent to either $\top$ or $\bot$.

In the particular case of $r = 0$, r-local formulas are equivalent to quantifier free formulas. As a general rule, one can always syntactically ensure that a given formula is r-local, by suitably restricting its quantifiers. Indeed, one can define for any fixed r the predicate $\mathrm{d}(a, b) \leq r$ in FO, whose semantic is that $\mathfrak{A}, \upsilon \models \mathrm{d}(a, b) \leq r$ if and only if $\mathrm{d}_{\mathfrak{A}}(a, b) \leq r$, using $\log_2 \lceil r \rceil$ quantifiers. As a consequence, whenever σ is a finite relational signature, one can construct the first order formula $y \in \mathcal{N}(\vec{x}, r)$, such that for all $(\mathfrak{A}, \upsilon)$, $\mathfrak{A}, \upsilon \models y \in \mathcal{N}(\vec{x}, r)$ if and only if $\upsilon(y) \in \mathcal{N}_{\mathfrak{A}}(\upsilon(\vec{x}), r)$.

Definition 4.1.9. Let σ be a finite relational signature and $r \geq 0$. Then given φ with free variables included in $\vec{x}$, one defines $|\varphi|^r_{\vec{x}}$ inductively as follows:

- $|\exists y.\varphi|^r_{\vec{x}} \stackrel{\text{def}}{=} \exists y \in \mathcal{N}(\vec{x}, r).|\varphi|^r_{\vec{x}}$;
- $|\varphi \wedge \psi|^r_{\vec{x}} \stackrel{\text{def}}{=} |\varphi|^r_{\vec{x}} \wedge |\varphi|^r_{\vec{x}}$;
- $|\neg\varphi|^r_{\vec{x}} \stackrel{\text{def}}{=} \neg|\varphi|^r_{\vec{x}}$;
- $|R(\vec{y})|^r_{\vec{x}} \stackrel{\text{def}}{=} R(\vec{y})$.

It is an easy check that for all $\varphi(\vec{x}) \in \mathsf{FO}$, for all $(\mathfrak{A}, \upsilon)$, one has $\mathfrak{A}, \upsilon \models |\varphi|^r_{\vec{x}}$ if and only if $\mathcal{N}_{\mathfrak{A}}(\upsilon(\vec{x}), r), \upsilon \models \varphi(\vec{x})$. As a consequence, whenever $\varphi(\vec{x})$ is a r-local formula, then $|\varphi|^r_{\vec{x}}$ is equivalent to φ. However, the transformation may increase the quantifier rank, as the subformula $y \in \mathcal{N}(\vec{x}, r)$ has a non-zero quantifier rank (if $r > 1$).

Example 4.1.10. Let $\varphi \stackrel{\text{def}}{=} \exists x.\forall y.E(x,y)$. Then,

$$|\varphi|^8_z = \exists x \in \mathcal{N}(z,8).\forall y \in \mathcal{N}(z,8).E(x,y) \quad .$$

Because every quantifier in a transformed formula is guarded, one might think that 1-local formulas are related to the guarded fragment of first order logic. However, the two fragments differ, because $\forall y \in \mathcal{N}(\vec{x}, r)$ unfolds to $\forall y \in \bigcup_{x \in \vec{x}} \mathcal{N}(x,r)$ while extensions of the guarded fragment consider only quantifications of the form $\forall y \in \bigcap_{x \in \vec{x}} \mathcal{N}(x,r)$, which is a stronger syntactic restriction.

We are almost ready to state the main theorem of this section, namely the locality of first order logic, a combinatorial tool that will be used to replace compactness arguments when dealing with classes of finite structures. Notice that Example 4.1.8 hints at a major issue when attempting to decompose first order sentences into *local* ones: locality only makes sense when free variables are involved.

To tackle this issue, let us introduce the notion of basic local sentences due to Gaifman. For that, let us write $\mathrm{indep}^{\mathcal{N}}(\vec{x},r)$ as a shorthand for the formula $\bigwedge_{1 \leq i \neq j \leq |\vec{x}|} \mathcal{N}(x_i, r) \cap \mathcal{N}(x_j, r) = \emptyset$, which is easily expressed using the predicates $\mathrm{d}(x_i, x_j) > 2r$. This formula expresses the fact that the vector $\vec{x}$ is composed of points that are "far enough" from each other.

Using this construction, let us now define a variant of the existential quantifier that includes a notion of repetition and distance: we define $\exists^{\geq n}_r x.\psi(x)$ as a notation for $\exists x_1, \ldots, x_n.\mathrm{indep}^{\mathcal{N}}(\vec{x},r) \land \bigwedge_{i=1}^n \psi(x_i)$. In plain English, this states that there exists n points, whose r-local neighbourhoods do not intersect, and such that each of these points satisfies ψ.

Definition 4.1.11 [37]**.** An *r-basic local first order sentence* is of the form $\exists^{\geq n}_r x.\psi(x)$, where $\psi(x)$ is an r-local formula with a single free variable x.

Example 4.1.12. Let $\sigma \stackrel{\text{def}}{=} \{ (E, 2) \}$. The sentence stating that one can find two disjoint 2-local neighbourhoods N_1, N_2 and two points $x_1 \in N_1, x_2 \in N_2$ such that $N_1 = \mathcal{N}_{N_1}(x_1, 1)$ and $N_2 = \mathcal{N}_{N_2}(x_2, 1)$, is a basic local first order sentence that can be written:

$$\exists^{\geq 2}_2 x.\forall y \in \mathcal{N}(x,2).E(x,y) \quad .$$

Example 4.1.13 uses the fact that basic local sentences are compatible with disjoint unions of structures. This will actually be a complete description of their specialisation preorder over $\mathsf{Fin}(\sigma)$, as we will prove in Proposition 4.3.10.

Example 4.1.13. The sentence $\forall x.\forall y.E(x,y)$ cannot be rewritten as a basic local sentence.

Proof Hint. Notice that if φ is a basic local sentence and $\mathfrak{A} \models \varphi$, then for all $\mathfrak{B} \in \mathsf{Struct}(\sigma)$, $\mathfrak{A} \uplus \mathfrak{B} \models \varphi$. ∎

The celebrated Gaifman Locality Theorem states the locality of first order logic: the evaluation of a first-order formula only depends on the first-order properties of the local neighbourhoods of that structure, which we formally restate below.

Theorem 4.1.14 ([37]). *Every first order formula is equivalent to a Boolean combination of basic local sentences and local formulas.*

This is the Gaifman Locality Theorem.

One can also restate the Gaifman Locality Theorem as the fact that every first order formula has a Gaifman normal form.

Even though Theorem 4.1.14 does not prove the uniqueness of the Boolean combination of basic local sentences and local formulas, it is usually said that such sentences are in *Gaifman normal form* (although, this is not a normal form).

One can add restrictions on the shape of basic local sentences and local formulas to build variants of the Gaifman normal form. Some examples can be found in the leftmost column of Figure 4.1.

Corollary 4.1.15. *For all sentences $\phi \in \mathsf{FO}[\sigma]$, there exists $r, k \geq 0$ such that for all structures $\mathfrak{A}, \mathfrak{B} \in \mathsf{Struct}(\sigma)$, if $\mathsf{Local}(\mathfrak{A}, r, k) = \mathsf{Local}(\mathfrak{B}, r, k)$, then $\mathfrak{A} \models \phi$ if and only if $\mathfrak{B} \models \phi$.*

Beware that the “converse” of Corollary 4.1.15 is not true. This is because the local neighbourhoods of a structure can have arbitrary size, and might not be first-order definable at a given $r, k \geq 0$. Formally, there exists a pair of structures $\mathfrak{A}, \mathfrak{B}$ such that $\mathfrak{A}$ and $\mathfrak{B}$ satisfy the same sentences but do not have the same local neighbourhoods.

Notice that one can extend Corollary 4.1.15 by considering first-order formulas with free variables, which requires to extend the construction of r-local neighbourhoods to structures with a valuation. In order to keep the notations simple and because such case will not appear, we will restrict ourselves to the current phrasing of Corollary 4.1.15.

Let us illustrate how the locality of first-order logic can be used to prove the relativisation of preservation theorems to classes of finite structures. Recall that Cycles is the class of undirected finite cycles.

Example 4.1.16. The Łoś-Tarski Theorem relativises to Cycles.

Proof. Let $\phi \in \mathsf{FO}$ be a sentence preserved under extensions over Cycles. Thanks to Corollary 4.1.15, there exists $r, k \geq 0$ such that whether $\mathfrak{A} \models \phi$ holds is determined by $\mathsf{Local}(\mathfrak{A}, r, k)$. Recall from Example 4.1.3 that if $m, n > 2r \times k$, then the local neighbourhoods of the two cycles are identical, i.e., $\mathsf{Local}(C_n, r, k) = \mathsf{Local}(C_m, r, k)$. Let us build the equivalent existential sentence by case analysis, using the diagram sentences $\Delta^{\mathsf{EFO}}_{\mathfrak{A}}$ that exists for EFO thanks to Lemma 3.1.5.

- If there exists $n > 2r \times k$ such that $C_n \models \phi$. Then for all $n > 2r \times k$, $C_n \models \phi$. Hence, ϕ is equivalent over Cycles to

$$\bigvee_{C_n \models \phi \wedge n \leq 2r \times k} \Delta^{\mathsf{EFO}}_{C_n} \vee \exists^{\geq 2r \times k + 1}_{0} x. \top \quad .$$

- Otherwise, for all $n > 2r \times k$, $C_n \models \neg\phi$. Hence, ϕ is equivalent over Cycles to

$$\bigvee_{C_n \models \phi \wedge n \leq 2r \times k} \Delta^{\mathsf{EFO}}_{C_n}$$

■

The technique presented is far from being an ad-hoc one, as it was already discussed in Chapter 3 (Preservation Theorems for First Order Queries). However, let us notice that we avoided the study of minimal models and the corresponding Lemma 3.2.3: here there are infinitely many non-equivalent minimal models! Note that using locality to study the combinatorial properties of minimal models is at the core of numerous proofs of relativisation [6, 7, 20, 57].

[6]: Atserias, Dawar and Grohe (2008), 'Preservation under extensions on well-behaved finite structures'
[7]: Atserias, Dawar and Kolaitis (2006), 'On preservation under homomorphisms and unions of conjunctive queries'
[20]: Dawar (2010), 'Homomorphism preservation on quasi-wide classes'
[57]: Harwath, Heimberg and Schweikardt (2015), 'Preservation and decomposition theorems for bounded degree structures'

Quite noticeably, Rossman's proof of the relativisation of the Homomorphism Preservation Theorem to the finite does not explicitly use Theorem 4.1.14, and builds upon custom Ehrenfeuch-Fraïsé games [83].

4.2. A Positive Locality Theorem

The astute reader might recall from Chapter 3 (Preservation Theorems for First Order Queries) and particularly in Section 3.1 that whenever a fragment of first order logic is given, a corresponding notion of embedding can be defined, and one has a corresponding preservation theorem.

Let us follow this pattern and write FOLoc for the fragment of local formulas, and FOLoc^r for the fragment of r-local formulas. Specialising Definition 3.1.1, let us define an *r-local elementary embedding* to be a map $h\colon \mathfrak{A} \to \mathfrak{B}$ such that for all $\phi \in \mathsf{FOLoc}^r$, $\mathfrak{A}, \upsilon \models \phi$ implies $\mathfrak{A}, h \circ \upsilon \models \phi$. Similarly, a *local elementary embedding* is a map preserving FOLoc. Thanks to Theorem 3.1.9, every first order sentence preserved under local elementary embedding is equivalent over $\mathsf{Struct}(\sigma)$ to an *existential local sentence*, that is, a sentence in $\exists\mathsf{FOLoc}$. Concretely, a sentence $\phi \in \exists\mathsf{FOLoc}$ is of the form $\exists \vec{x}.\psi(\vec{x})$ where ψ is a local formula[6][7].

6: We similarly define *universal local sentences* as sentences in $\forall\mathsf{FOLoc}$, that are also the negations of *existential local sentences*.

7: As for $\exists\mathsf{QF}$, the set of existential local sentences does not form a fragment of first order logic. We implicitly equate $\exists\mathsf{FOLoc}$ with the smallest fragment. This is valid because every sentence in the closure of $\exists\mathsf{FOLoc}$ is equivalent to some sentence in $\exists\mathsf{FOLoc}$.

The local elementary embeddings are the "local" counterpart to the usual notion of *elementary embedding* in Model Theory: a map $h\colon \mathfrak{A} \to \mathfrak{B}$ is an *elementary embedding* if and only if for all $\varphi(\vec{x}) \in \mathsf{FO}[\sigma]$, $\upsilon\colon \vec{x} \to \mathfrak{A}$, $\mathfrak{A}, \upsilon \models \varphi(\vec{x})$ implies that $\mathfrak{B}, h \circ \upsilon \models \varphi(\vec{x})$. By construction, *elementary embeddings* are the F-embeddings where F is choosen to be the full collection of first order formulas FO. In particular, every first order sentence is preserved under *elementary embeddings*.

In order to be able to refer to the specialisation of Theorem 3.1.9 in the case of existential local sentences, let us write the corollary hereafter.

Corollary 4.2.1. *Let φ be a sentence in $\mathsf{FO}[\sigma]$. The following properties are equivalent over the class $\mathsf{Struct}(\sigma)$ of all structures.*

1. *φ is equivalent to an existential local sentence.*
2. *φ is preserved under local elementary embeddings.*

The similarity between Corollary 4.2.1 and Theorem 4.1.14 is not a coincidence. However, it is not yet clear to see how the two relate, since one talks about existential local sentences and the other one

Table 4.3.: Variations around locality and sentences in first order logic.

Name	Shape	Hypothesis
Existential local	$\exists\vec{x}.\psi(\vec{x})$	ψ r-local formula
Almost basic local	$\exists\vec{x}.\mathrm{indep}^{\mathcal{N}}(\vec{x},r) \wedge \psi(\vec{x})$	ψ r-local formula
Asymmetric basic local	$\exists\vec{x}.\mathrm{indep}^{\mathcal{N}}(\vec{x},r) \wedge \bigwedge_i \psi_i(x_i)$	$(\psi_i)_{1\leq i\leq\lvert\vec{x}\rvert}$ r-local formulas
Basic local	$\exists\vec{x}.\mathrm{indep}^{\mathcal{N}}(\vec{x},r) \wedge \bigwedge_i \psi(x_i)$	ψ r-local formula

considers the more constrained basic local sentences. Let us state the main theorem of this section.

Theorem 4.2.2. *Let $\mathcal{C} \subseteq \mathsf{Struct}(\sigma)$ be a class of structures and $\varphi \in \mathsf{FO}[\sigma]$ be a first-order sentence. Then, the following properties are equivalent:*

1. *φ is equivalent over $\mathcal{C}$ to an existential local sentence*
2. *φ is equivalent over $\mathcal{C}$ to a positive Boolean combination of basic local sentences, which we call a* positive Gaifman normal form.

As a consequence of Theorems 4.1.14 and 4.2.2, every first-order sentence is equivalent to a finite conjunction of implications of the form $\psi_i \Rightarrow \varphi_i$ where both ψ_i and φ_i are existential local sentences. Notice that this can be seen as a variation of the normal form proposed by [87], who proved that every first order sentence is equivalent to a sentence of the shape $\exists\vec{x}.\forall y.\psi(x,y)$ where ψ is a local formula. From this point of view, the preservation under local elementary embeddings "removes the need for universal quantifications."

[87]: Schwentick and Barthelmann (1999), 'Local normal forms for first-order logic with applications to games and automata'

The rest of this section is devoted to proving Theorem 4.2.2, through a gradual understanding of the expressive power of local formulas. As basic local sentences form a subset of existential local sentences, the only difficulty is converting an existential local sentence into a positive Boolean combination of basic local sentences. We split this transformation into intermediate syntactic steps as described in Table 4.3. These intermediate steps are motivated as follows:

- Translating from existential local sentences to almost basic local sentences amounts to "separating" the variables into disjoint neighbourhoods;
- Translating from almost basic local sentences to asymmetric basic local sentences amounts to "separating" the properties checked on the variable, based on the fact that they are far away;
- Translating from asymmetric basic local sentences to basic local sentences amounts to "uniformizing" the local properties so that the same one is asked for every variable.

Exploring Table 4.3 might seem daunting, but the introduced fragments arise quite naturally. For instance, asymmetric basic local sentences were already introduced by Grohe and Wöhrle in order to construct an existential Gaifman normal form see [55, Theorem 6]. Because they have similar definitions, we will first help the reader distinguish their respective "expressive power" in Figure 4.5.

[55]: Grohe and Wöhrle (2004), 'An existential locality theorem'

Beware that in [55], “existential local sentences” do not refer to the existential closure of a local sentence, but to what is defined here as an existential basic local sentence. We take the time to introduce existential basic local sentences because of an analogue to Theorem 4.2.2 has been proven by Grohe and Wöhrle, which we restate in Theorem 4.2.4. Please note however that Theorem 4.2.4 is not an equivalence, and it is unclear whether it could be used to prove Theorem 4.2.2 (and vice-versa).

Definition 4.2.3 [55, Section 3]**.** Let σ be a finite relational signature. A first order sentence $\varphi \in \mathsf{FO}[\sigma]$ is an *existential Basic local sentence* when it is of the form $\exists^{\geq k}_{r} x.\psi(x)$ where ψ is an existential formula that is also r-local. I.e., φ is of the form:

$$\exists\vec{x}.\mathrm{indep}^{\mathcal{N}}(\vec{x},r) \wedge \bigwedge_i \psi(x_i) \quad .$$

Theorem 4.2.4 [55, Theorem 2]**.** Let σ be a finite relational signature. Every existential sentence $\varphi \in \mathsf{FO}[\sigma]$ is equivalent to a positive Boolean combination of existential basic local sentences.

Such a positive Boolean combination of existential basic local sentences is called an *existential Gaifman normal form* of φ.

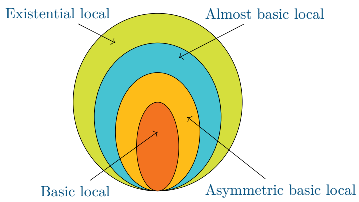


Figure 4.5.: Strict inclusions between the different set of local sentences described in Table 4.3. Witnessed by examples Examples 4.2.5 to 4.2.7.

Example 4.2.5. Let $\mathcal{C}$ be the class of coloured graphs, let $r \geq 2$. Let $\psi(x) \stackrel{\text{def}}{=} \forall x \in \mathcal{N}(x,r).\mathrm{green}(x)$ and $\theta(x) \stackrel{\text{def}}{=} \forall x \in \mathcal{N}(x,r).\mathrm{blue}(x)$. The sentence

$$\exists x_1, x_2.\mathrm{indep}^{\mathcal{N}}(\vec{x},r) \wedge \psi(x_1) \wedge \theta(x_2)$$

is asymmetric basic local but not basic local.

Example 4.2.6. Let $\mathcal{C}$ be the class of coloured graphs, let $r \geq 2$. Let $\psi(x) \stackrel{\text{def}}{=} \forall x \in \mathcal{N}(x,r).\mathrm{green}(x)$ and $\theta(x) \stackrel{\text{def}}{=} \forall x \in \mathcal{N}(x,r).\mathrm{blue}(x)$. The sentence

$$\exists x_1, x_2.\mathrm{indep}^{\mathcal{N}}(\vec{x},r) \wedge (\psi(x_1) \wedge \psi(x_2)) \vee (\theta(x_1) \wedge \theta(x_2))$$

is almost basic local but not asymmetric basic local.

Example 4.2.7. Let $\mathcal{C}$ be the class of coloured graphs, let $r \geq 2$. Let $\psi(x) \stackrel{\text{def}}{=} \forall x \in \mathcal{N}(x,r).\mathrm{green}(x)$ and $\theta(x) \stackrel{\text{def}}{=} \forall x \in \mathcal{N}(x,r).\mathrm{blue}(x)$. The sentence

$$\exists x_1, x_2.\forall y \in \mathcal{N}(x_1 x_2, r).(E(x_1,y) \wedge \theta(y)) \iff (E(x_2,y) \wedge \psi(y))$$

is existential local but not almost basic local.

Remark that Theorem 4.2.2 claims that positive Boolean combinations of existential local sentences are as expressive as positive Boolean combinations of basic local sentences, hence that the "hierarchy" of local variations introduced Table 4.3 and depicted in Figure 4.5 collapses when considering their closure under finite disjunctions and finite conjunctions.

As noticed in the above examples, the variation of expressive power essentially depend on the repartition of the local neighbourhoods in a given structure. To perform the combinatorial study of "possible overlaps" of neighbourhoods, we first prove the following lemma.[8] We encourage the reader to look at Figure 4.6 while reading Lemma 4.2.8 to understand the basic idea, which is to turn a tuple of points $\vec{a}$, into a tuple $\vec{b}$ of points that are guaranteed to have non-intersecting neighbourhoods.

8: Which holds in any given metric space, but is stated inside a given structure for clarity.

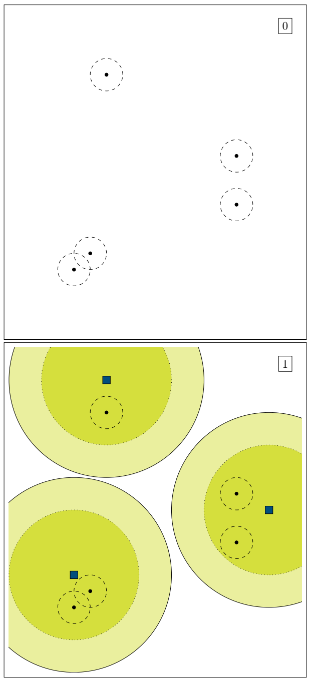


Figure 4.6.: Input (slide 0) and output (slide 1) of Lemma 4.2.8. Black nodes are the input vector $\vec{a}$, with their r-neighbourhoods represented as dashed circles. Square blue nodes represent the output vector $\vec{b}$ with their R-neighbourhoods represented as green circles, and $3R$-neighbourhoods represented as light-green circles.

Lemma 4.2.8. *For every $k, r \geq 0$, for every structure $\mathfrak{A} \in \mathsf{Struct}(\sigma)$ and vector $\vec{a} \in \mathfrak{A}^{\leq k}$, there exists a vector $\vec{b} \in \mathfrak{A}^{\leq k}$ and a radius $r \leq R \leq 4^k r$ such that:*

1. *The R-local neighbourhood of $\vec{b}$ contains the r-local neighbourhood of $\vec{a}$: $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r) \subseteq \mathcal{N}_{\mathfrak{A}}(\vec{b}, R)$,*
2. *The R-local neighbourhoods of elements in $\vec{b}$ are not intersecting: $\forall b \neq b' \in \vec{b}, \mathcal{N}_{\mathfrak{A}}(b, 3R) \cap \mathcal{N}_{\mathfrak{A}}(b', 3R) = \emptyset$.*

Proof. We proceed by induction over k.

- When $k \leq 1$ it suffices to take $R \stackrel{\text{def}}{=} r$ and for every vector $\vec{a} \in \mathfrak{A}^{\leq k}$ build $\vec{b} \stackrel{\text{def}}{=} \vec{a}$ and notice that $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r) = \mathcal{N}_{\mathfrak{A}}(\vec{b}, R)$.
- When $k \geq 2$, we proceed by a simple case analysis
  1. Either the balls $\mathcal{N}_{\mathfrak{A}}(a, 3r)$ are pairwise disjoint when a ranges over $\vec{a}$; in which case it suffices to consider $\vec{b} \stackrel{\text{def}}{=} \vec{a}$ and $R \stackrel{\text{def}}{=} r$ to conclude.
  2. Or at least two of the balls $\mathcal{N}_{\mathfrak{A}}(a, 3r)$ intersect when a ranges over $\vec{a}$, and we can assume without loss of generality that the neighbourhoods of a_1 and a_2 at radius $3r$ intersect.
     Let us consider $c \in \mathcal{N}_{\mathfrak{A}}(a_1, 3r) \cap \mathcal{N}_{\mathfrak{A}}(a_2, 3r)$. Define $\vec{c} \stackrel{\text{def}}{=} (c, a_3, \dots, a_k)$, this vector is of size 1 when $k = 2$.
     Because $d(a_1, c) \leq 3r$ and $d(a_2, c) \leq 3r$, $\mathcal{N}_{\mathfrak{A}}(a_1 a_2, r) \subseteq \mathcal{N}_{\mathfrak{A}}(c, 4r)$.
     By induction hypothesis, there exists a radius $4r \leq R \leq 4^{k-1}(4r)$ and a vector $\vec{b} \in \mathfrak{A}^{\leq k-1}$ such that $\mathcal{N}_{\mathfrak{A}}(\vec{c}, 4r) \subseteq \mathcal{N}_{\mathfrak{A}}(\vec{b}, R)$ and $\forall b \neq b' \in \vec{b}, \mathcal{N}_{\mathfrak{A}}(b, 3R) \cap \mathcal{N}_{\mathfrak{A}}(b', 3R) = \emptyset$.
     Since $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r) \subseteq \mathcal{N}_{\mathfrak{A}}(\vec{c}, 4r)$ and $r \leq 4r \leq R \leq 4^k r$, the statement holds. ∎

4.2.1. From Existential Local Sentences to Asymmetric Basic Local Sentences

In this section, we will do a first step towards proving Theorem 4.2.2 by showing that existential local sentences are equivalent to disjunctions of asymmetric basic local sentences. This transformation is split into two; in Lemma 4.2.9 on the following page, we perform the first step of conversion by rewriting existential local sentences into almost basic local sentences, leveraging the study of neighbourhoods intersections done in Lemma 4.2.8. In Lemma 4.2.20 on page 83, we continue by transforming an almost basic local sentence into a disjunction of asymmetric basic local sentences. The second step is more involved and relies on the Feferman-Vaught technique, which is a very convenient tool in finite model theory that will also come in handy in Chapter 6 (Logically Presented Spaces).

Lemma 4.2.9. *Let σ be a finite relational signature, and $\varphi(\vec{x})$ be an r-local formula. There exist $1 \leq n \leq 4^{|\vec{x}|}r$ and $\psi_1, \ldots, \psi_n$ almost basic local sentences such that $\exists \vec{x}.\varphi(\vec{x})$ is equivalent to the disjunction $\bigvee_{1\leq i\leq n} \psi_i$ over $\mathsf{Struct}(\sigma)$.*

Proof. Let us define $\Delta \stackrel{\text{def}}{=} \{ (k,R) \colon 0 \leq k \leq |\vec{x}| \wedge r \leq R \leq 4^{|\vec{x}|}r \}$, and for each $(k,R) \in \Delta$,

$$\psi_{(k,R)} \stackrel{\text{def}}{=} \exists b_1, \ldots, b_k.\mathrm{indep}^{\mathcal{N}}(\vec{b},R) \wedge \exists \vec{x} \in \mathcal{N}(\vec{b},R).\varphi(\vec{x}) \tag{4.1}$$

To conclude it suffices to prove that $\exists \vec{x}.\varphi(\vec{x})$ is equivalent to

$$\Psi \stackrel{\text{def}}{=} \bigvee_{(k,R)\in\Delta} \psi_{(k,R)} \quad .$$

- Assume that $\mathfrak{A} \in \mathsf{Struct}(\sigma)$ satisfies $\exists \vec{x}.\varphi(\vec{x})$. Then there exists $\vec{a} \in A^{|\vec{x}|}$ such that $\mathcal{N}_{\mathfrak{A}}(\vec{a},r) \models \exists \vec{x}.\varphi(\vec{x})$ since φ is r-local around $\vec{x}$. Using Lemma 4.2.8, there exists a size $0 \leq k \leq |\vec{a}|$ a radius $r \leq R \leq 4^{|\vec{a}|}r$, and a vector $\vec{b} \in \mathfrak{A}^k$ such that $\mathcal{N}_{\mathfrak{A}}(\vec{a},r) \subseteq \mathcal{N}_{\mathfrak{A}}(\vec{b},R)$ and the balls of radius $3R$ around the points of $\vec{b}$ do not intersect. In particular, $\biguplus_{b\in\vec{b}} \mathcal{N}_{\mathfrak{A}}(b,R) \models \exists \vec{x}.\varphi(\vec{x})$ since witnesses in $\mathcal{N}_{\mathfrak{A}}(\vec{a},r)$ can still be found and φ is r-local. This proves that $\mathfrak{A} \models \psi_{(k,R)}$ hence that $\mathfrak{A} \models \Psi$.
- Assume that $\mathfrak{A} \in \mathsf{Struct}(\sigma)$ satisfies Ψ. Then there exists (k,R) such that $\mathfrak{A} \models \psi_{k,R}$ thus proving that there exists $\vec{b} \in \mathfrak{A}^k$ such that $\biguplus_{b\in\vec{b}} \mathcal{N}_{\mathfrak{A}}(b,R) \models \exists \vec{x}.\varphi(\vec{x})$. Since $r \leq R$ and φ is r-local this proves $\mathfrak{A} \models \exists \vec{x}.\varphi(\vec{x})$. ■

This proves that even though existential local sentences and almost basic local sentences do not have the same expressive power, disjunctions of such sentences are equi-expressive.

Transforming almost basic local sentences into positive Boolean combination of asymmetric basic local sentences requires a bit more work, and typically leverages what is known as *Feferman-Vaught style compositional techniques* [33, 74]. This technique/meta-theorem reduces the evaluation of a sentence over a complex structure,[9] into an evaluation of (more-complex) sentences on simpler parts.[10] It is quite clear how such theorems will transform almost basic local sentences into positive Boolean combinations of asymmetric basic local sentences.

[33]: Feferman and Vaught (1959), 'The first order properties of products of algebraic systems'

[74]: Makowsky (2004), 'Algorithmic uses of the Feferman–Vaught Theorem'

9: For instance, a disjoint union of neighbourhoods.

10: Typically, the connected components of a disjoint union.

We present hereafter a suitably adapted version of the Feferman-Vaught technique, following a syntactic proof due to Thomas Colcombet. This syntactic proof avoids most of the problems arising from semantic ones, and is general enough so that it can be reused later when studying compositional properties of preservation theorems in Chapter 6 (Logically Presented Spaces).

The rough idea of the syntactic approach is to use a stricter language of "typed formulas" (Definition 4.2.10) to keep track of where variables are taken inside a complex structure. One can transform usual sentences into typed formulas preserving the semantics (Definition 4.2.14 and Lemma 4.2.17), and conversely discard the type annotations to turn a typed formula into a regular one. The main advantage of typed formulas is that they are easily rewritten into "monotyped formulas" (Lemma 4.2.19). Because types keep track of the components of a structure, this proves that one can transform a formula into a positive Boolean combination of formulas that only consider one connected component. A graphical representation of the proof scheme is given in Figure 4.7.

$$\begin{tikzcd}
\mathsf{FO} & & \mathbb{T}\mathsf{FO} \arrow[ll, "\text{Lemma 4.2.13}"'] \\
\mathsf{FO} \arrow[u, dashed, "\text{Lemma 4.2.20}"] \arrow[rr, "\text{Lemma 4.2.17}", "\mathrm{conv}^{\mathbb{T}}(\Gamma, \cdot)"'] & & \mathbb{T}\mathsf{FO} \arrow[u, "\text{Lemma 4.2.19}"']
\end{tikzcd}$$

Figure 4.7.: The proof scheme of Lemma 4.2.20.

Definition 4.2.10. *Let $\mathbb{T}$ be a non-empty set, and σ be a relational signature. The $\mathbb{T}$-typed formulas are defined inductively as follows:*

$$\begin{aligned}
\tau := {} & x : T && \textit{when } T \in \mathbb{T} \\
\mid {} & R(x_1 : T, \dots, x_n : T) && \textit{when } R \in \sigma, T \in \mathbb{T} \\
\mid {} & \exists x : T.\tau && \textit{when } T \in \mathbb{T} \\
\mid {} & \forall x : T.\tau && \textit{when } T \in \mathbb{T} \\
\mid {} & \tau \wedge \tau \mid \tau \vee \tau \mid \neg\tau \mid \top &&
\end{aligned}$$

In this new language of $\mathbb{T}$-typed formulas, a relation can only be introduced using variables of the same type $T \in \mathbb{T}$, as shown in Example 4.2.11.

Example 4.2.11. The formula $\exists z : T_2.R(x : T_1, y : T_1) \wedge P(z : T_3)$ is a typed formula. While $\exists z : T_2.R(x : T_1, z : T_2)$ is not a typed formula.

Let us now define the semantics associated with $\mathbb{T}$-typed formulas. The idea is to follow the usual evaluation of first order sentences[11] with added "runtime type checking" when evaluating existential quantifiers. To perform (at runtime) that a variable $x : T$ is correctly typed, we define the evaluation over typed structures, namely structures $\mathfrak{A} \in \mathsf{Struct}(\sigma)$ together with a *function* $\rho\colon \mathfrak{A} \to \mathbb{T}$.[12]

11: That was defined inductively in Definition 2.2.6.

12: This typed logic can actually be encoded in a subset of MSO, by considering types as (set) free variables that partition the structure. In this setting, $\exists x : T.\psi$ corresponds to the MSO formula $\exists x.x \in T \wedge \psi$. However, we prefer the syntactic approach because it makes the proof of Lemma 4.2.19 easier to follow.

Definition 4.2.12. *The $\mathbb{T}$-typed satisfaction relation $\models^{\mathbb{T}}$ is defined by induction on the $\mathbb{T}$-typed formulas follows: given $\mathfrak{A} \in \mathsf{Struct}(\sigma)$,*

$\vec{x} \subseteq_{\mathsf{fin}} \mathbb{V}$, $\upsilon \colon \vec{x} \to \mathfrak{A}$, $\rho \colon \mathfrak{A} \to \mathbb{T}$, *and* $\varphi \in \mathsf{FO}[\sigma]$

$$
\begin{aligned}
\mathfrak{A}, \rho, \upsilon \models^{\mathbb{T}} R(y_1 : T, \dots, y_n : T) &\overset{def}{\iff} \rho(y_1) = \dots = \rho(y_n) = T \text{ and} \\
&\phantom{\overset{def}{\iff}} (\upsilon(y_1), \dots, \upsilon(y_n)) \in R^{\mathfrak{A}} \\
\mathfrak{A}, \rho, \upsilon \models^{\mathbb{T}} \psi_1(\vec{y}) \wedge \psi_2(\vec{z}) &\overset{def}{\iff} \mathfrak{A}, \rho, \upsilon \models^{\mathbb{T}} \psi_1(\vec{y}) \text{ and} \\
&\phantom{\overset{def}{\iff}} \mathfrak{A}, \rho, \upsilon \models^{\mathbb{T}} \psi_2(\vec{z}) \\
\mathfrak{A}, \rho, \upsilon \models^{\mathbb{T}} \neg\psi(\vec{y}) &\overset{def}{\iff} \mathfrak{A}, \rho, \upsilon \not\models^{\mathbb{T}} \psi_1(\vec{y}) \\
\mathfrak{A}, \rho, \upsilon \models^{\mathbb{T}} \exists z : T.\psi(\vec{y}z) &\overset{def}{\iff} \text{there exists } a \in \mathfrak{A}, \\
&\phantom{\overset{def}{\iff}} \text{such that } \rho(a) = T, \text{ and} \\
&\phantom{\overset{def}{\iff}} \mathfrak{A}, \rho, \upsilon[z \mapsto a] \models^{\mathbb{T}} \psi(\vec{y}z) \\
\mathfrak{A}, \rho, \upsilon \models^{\mathbb{T}} \top &\overset{def}{\iff} \text{always} \quad .
\end{aligned}
$$

One can convert a typed formula into an untyped one by removing all type annotations, which we write $\mathrm{U}\mathbb{T}(\psi)$. In general, this conversion is "lossy," but let us remark that, when we restrict our attention to a universe with a single type T, the $\mathbb{T}$-typed satisfiability and the usual notion of satisfiability coincide.

Lemma 4.2.13. *For all $\mathfrak{A} \in \mathsf{Fin}(\sigma)$, for all sets $\mathbb{T}$, for all ρ maps every element to a single type $T \in \mathbb{T}$, for all $\mathbb{T}$-typed formulas ψ where all variables have the same type T, the following are equivalent:*

1. $\mathfrak{A}, \rho, \upsilon \models^{\mathbb{T}} \psi$
2. $\mathfrak{A}, \upsilon \models \mathrm{U}\mathbb{T}(\psi)$.

Proof. We prove the result by induction on the formula ψ. Because the two semantics have the same definition except for the case of the existential quantification, we only treat this specific case.

Assume that $\psi = \exists x : T'.\theta$. Because every variable in ψ has type T, we conclude that $T = T'$. Furthermore, we know that $\rho(a) = T$ for all $a \in \mathfrak{A}$. Now, remark that one can use the induction hypothesis on $\mathfrak{A}$, ρ, $\upsilon[z \mapsto a]$, and θ for all $a \in \mathfrak{A}$.

$$
\begin{aligned}
\mathfrak{A}, \rho, \upsilon \models^{\mathbb{T}} \exists z : T.\theta(\vec{y}z) &\overset{\text{def}}{\iff} \text{there exists } a \in \mathfrak{A}, \\
&\phantom{\overset{\text{def}}{\iff}} \text{such that } \rho(a) = T, \text{ and} \\
&\phantom{\overset{\text{def}}{\iff}} \mathfrak{A}, \rho, \upsilon[z \mapsto a] \models^{\mathbb{T}} \theta(\vec{y}z) \\
&\iff \text{there exists } a \in \mathfrak{A}, \\
& \mathfrak{A}, \rho, \upsilon[z \mapsto a] \models^{\mathbb{T}} \theta(\vec{y}z) \\
&\iff \text{there exists } a \in \mathfrak{A}, \\
& \mathfrak{A}, \upsilon[z \mapsto a] \models \mathrm{U}\mathbb{T}(\theta(\vec{y}z)) \\
&\overset{\text{def}}{\iff} \mathfrak{A}, \upsilon \models \exists z.\, \mathrm{U}\mathbb{T}(\theta(\vec{y}z)) \\
&\iff \mathfrak{A}, \upsilon \models \mathrm{U}\mathbb{T}(\exists z : T.\theta(\vec{y}z)) \quad . \qquad \blacksquare
\end{aligned}
$$

Lemma 4.2.13 proves that it is possible to erase typing annotation of *monotyped formulas*, that is formulas using a single type T. Let us now provide a way to go the other direction and *add type annotations* to an untyped formula. This conversion will use *$\mathbb{T}$-typing environments* Γ, the static counterpart to the "runtime environment" ρ, that is, functions from (finite subsets of) free variables to $\mathbb{T}$.

Beware that we must assume that $\mathbb{T}$ is finite for the conversion below to be well-defined. This is not a problem for this section, but will become an obstacle when applying the Feferman-Vaught technique to linear orderings in Lemma 6.3.35.

Definition 4.2.14. *Let $\mathbb{T}$ be a non-empty finite set, ψ be a first-order formula, and Γ be a typing environment. We define the $\mathbb{T}$-typed formula $\mathrm{conv}^{\mathbb{T}}(\Gamma, \psi)$ by induction on ψ as follows:*

1. $\mathrm{conv}^{\mathbb{T}}(\Gamma, \exists x.\psi) \stackrel{\text{def}}{=} \bigvee_{T \in \mathbb{T}} \exists x : T.\mathrm{conv}^{\mathbb{T}}(\Gamma[x \mapsto T], \psi)$,
2. $\mathrm{conv}^{\mathbb{T}}(\Gamma, \psi_1 \vee \psi_2) \stackrel{\text{def}}{=} \mathrm{conv}^{\mathbb{T}}(\Gamma, \psi_1) \vee \mathrm{conv}^{\mathbb{T}}(\Gamma, \psi_2)$,
3. $\mathrm{conv}^{\mathbb{T}}(\Gamma, \neg\psi) \stackrel{\text{def}}{=} \neg\mathrm{conv}^{\mathbb{T}}(\Gamma, \psi)$,
4. $\mathrm{conv}^{\mathbb{T}}(\Gamma, \top) \stackrel{\text{def}}{=} \top$,
5. $\mathrm{conv}^{\mathbb{T}}(\Gamma, R(y_1, \dots, y_n)) \stackrel{\text{def}}{=} R(y_1 : \Gamma(y_1), \dots, y_n : \Gamma(y_n))$ *when* $\Gamma(y_1) = \cdots = \Gamma(y_n)$, *and* $\perp$ *otherwise.*

Notice that Definition 4.2.14 not only adds types, but also performs a kind of "partial evaluation," by statically replacing relations with $\perp$ whenever there is a type mismatch.

We illustrate an example of conversion in Example 4.2.15, where we highlight that atomic formulas containing different types are mapped to $\perp$, which may change the semantics. The purpose of Lemma 4.2.17 is to show that a simple restriction on the choice of typing environment ensures that the translation is faithful.

Example 4.2.15. Let $\varphi(x, y) \stackrel{\text{def}}{=} \exists z.E(x, z) \wedge E(y, z)$, $\mathbb{T} \stackrel{\text{def}}{=} \{ T_1, T_2 \}$, and Γ be such that $\Gamma(x) = T_1 \neq T_2 = \Gamma(y)$. Then,

$$\mathrm{conv}^{\mathbb{T}}(\Gamma, \varphi) = \vee \begin{cases} \exists z : T_1.E(x : T_1, z : T_1) \wedge \perp \\ \exists z : T_2.\perp \wedge E(y : T_2, z : T_2) \end{cases} .$$

Note that for all $\mathfrak{A} \in \mathsf{Struct}(\sigma)$, $\upsilon \colon \{ x, y \} \to \mathfrak{A}$, and $\rho \colon \mathfrak{A} \to \mathbb{T}$, $\mathfrak{A}, \rho, \upsilon \not\models^{\mathbb{T}} \mathrm{conv}^{\mathbb{T}}(\Gamma, \varphi)$. This proves that $\mathrm{conv}^{\mathbb{T}}(\Gamma, \varphi)$ is not satisfiable, even though φ was satisfiable.

Because it is going to be crucial later on, let us highlight right now that the conversion $\mathrm{conv}^{\mathbb{T}}(\Gamma, \varphi)$ does not create new free variables, and assigns the "correct" type to the free variables of φ.

Fact 4.2.16. Let $\varphi(\vec{x}) \in \mathsf{FO}[\sigma]$ be a first order formula, $\mathbb{T}$ be a non-empty finite set of types, and Γ be a $\mathbb{T}$-typing environment. Then, $\mathrm{conv}^{\mathbb{T}}(\Gamma, \varphi)$ has free variables $\vec{x}$, the type of $\vec{x}$ is precisely $\Gamma(x)$, and the quantifier rank of $\mathrm{U}\mathbb{T}(\mathrm{conv}^{\mathbb{T}}(\Gamma, \varphi))$ is at most the quantifier rank of φ.

In Lemma 4.2.17, the hypothesis that $\Gamma(x) = \rho(\upsilon(x))$ ensures that the "static type" corresponds to the "runtime type" of the variable, which is a compatibility condition between Γ and ρ. This condition is crucially used when comparing the typed and untyped existential quantifications. On the other hand, the condition regarding distances is there to justify that the conversion is sound when it replaces $R(y_1, \dots, y_n)$ by $\perp$ in case of a type mismatch.

Lemma 4.2.17. *Let σ be a relational signature, $\mathbb{T}$ be a non-empty finite set, and $\psi(\vec{x})$ be a first-order formula. For all $\mathbb{T}$-typing environments Γ, for all $\mathfrak{A} \in \mathsf{Struct}(\sigma)$, for all $\rho \colon \mathfrak{A} \to \mathbb{T}$, for all $\upsilon \colon \vec{x} \to \mathfrak{A}$, such that $\mathrm{d}_{\mathfrak{A}}(a, b) \leq 1$ implies $\rho(a) = \rho(b)$ for all $a, b \in \mathfrak{A}$, and $\Gamma(x) = \rho(\upsilon(x))$ for all $x \in \vec{x}$, the following are equivalent:*

- $\mathfrak{A}, \upsilon \models \psi$,
- $\mathfrak{A}, \rho, \upsilon \models^{\mathbb{T}} \mathrm{conv}^{\mathbb{T}}(\Gamma, \psi)$.

Proof. We prove the result by induction on the (untyped) formula ψ.

The case of disjunction, negation, and the formula $\top$ are immediate, and we focus on the two difficult cases: the base case of a relation $R(y_1, \dots, y_n)$ and the existential quantification.

For the base case, let $\psi = R(y_1, \dots, y_n)$. Assume that $\mathfrak{A}, \upsilon \models \psi$. Then, $d_{\mathfrak{A}}(\upsilon(y_i), \upsilon(y_j)) \leq 1$ for all $1 \leq i, j \leq n$. In particular, there exists $T \in \mathbb{T}$ such that $T = \Gamma(y_i) = \rho(\upsilon(y_i)) = \rho(\upsilon(y_j)) = \Gamma(y_j)$ for all $1 \leq i, j \leq n$. Therefore, $\mathrm{conv}^{\mathbb{T}}(\Gamma, \psi) = R(y_1 : T, \dots, y_n : T)$. Finally, $\mathfrak{A}, \rho, \upsilon \models^{\mathbb{T}} R(y_1 : T, \dots, y_n : T)$, because $(\upsilon(y_1), \dots, \upsilon(y_n)) \in R^{\mathfrak{A}}$.

Conversely, assume that $\mathfrak{A}, \rho, \upsilon \models^{\mathbb{T}} \mathrm{conv}^{\mathbb{T}}(\Gamma, \psi)$. Then, $\mathrm{conv}^{\mathbb{T}}(\Gamma, \psi) \neq \bot$, and in particular there exists a type $T \in \mathbb{T}$ such that $\Gamma(y_1) = \dots = T(y_n) = T$, and $\mathrm{conv}^{\mathbb{T}}(\Gamma, \psi) = R(y_1 : T, \dots, y_n : T)$. By definition of the typed satisfaction relation, this means that $(\upsilon(y_1), \dots, \upsilon(y_n)) \in R^{\mathfrak{A}}$, and therefore that $\mathfrak{A}, \upsilon \models \psi$.

Let us now focus on the case of the existential quantification. That is, we assume that $\psi = \exists y. \theta(\vec{x}, y)$. In this case, we have that

$$\mathrm{conv}^{\mathbb{T}}(\Gamma, \psi) = \bigvee_{T \in \mathbb{T}} \exists y : T. \mathrm{conv}^{\mathbb{T}}(\Gamma[y \mapsto T], \theta) \quad .$$

Remark that for all $a \in \mathfrak{A}$, one can apply the induction hypothesis on the formula θ, with $\upsilon[y \mapsto a]$, ρ, and $\Gamma[y \mapsto \rho(a)]$, to conclude that

$$\left(\mathfrak{A}, \rho, \upsilon[y \mapsto a] \models^{\mathbb{T}} \mathrm{conv}^{\mathbb{T}}(\Gamma[y \mapsto \rho(a)], \theta)\right) \iff (\mathfrak{A}, \upsilon[y \mapsto a] \models \theta) \quad .$$

Now, we obtained the desired equivalence as follows:

$$\begin{aligned}
\mathfrak{A}, \rho, \upsilon \models^{\mathbb{T}} \mathrm{conv}^{\mathbb{T}}(\Gamma, \psi) \iff & \exists T \in \mathbb{T}, \exists a \in \mathfrak{A}, \\
& \rho(a) = T, \text{ and} \\
& \mathfrak{A}, \rho, \upsilon[y \mapsto a] \models^{\mathbb{T}} \mathrm{conv}^{\mathbb{T}}(\Gamma[y \mapsto T], \theta) \\
\iff & \exists a \in \mathfrak{A}, \\
& \mathfrak{A}, \rho, \upsilon[y \mapsto a] \models^{\mathbb{T}} \mathrm{conv}^{\mathbb{T}}(\Gamma[y \mapsto \rho(a)], \theta) \\
\iff & \exists a \in \mathfrak{A}, \\
& \mathfrak{A}, \upsilon[y \mapsto a] \models \theta \\
\iff & \mathfrak{A}, \upsilon \models \exists y. \theta \\
\iff & \mathfrak{A}, \upsilon \models \psi \quad .
\end{aligned}$$

■

Corollary 4.2.18. *Let σ be a relational signature, $r > 0$, $\vec{x} = (x_1, \dots, x_n)$ be a finite tuple of variables. Let us define $\mathbb{T}$ to be the set $\{ T_1, \dots, T_n \}$. For all models $\mathfrak{A} \in \mathsf{Struct}(\sigma)$ and valuation $\upsilon \colon \vec{x} \to \mathfrak{A}$ such that $\mathfrak{A}, \upsilon \models \mathrm{indep}^{\mathcal{N}}(\vec{x}, r)$, for all $\psi(\vec{x}) \in \mathsf{FO}[\sigma]$, the following are equivalent*

1. *$\mathcal{N}_{\mathfrak{A}}(\upsilon(\vec{x}), r), \upsilon \models \psi(\vec{x})$*
2. *$\mathcal{N}_{\mathfrak{A}}(\upsilon(\vec{x}), r), \rho, \upsilon \models^{\mathbb{T}} \mathrm{conv}^{\mathbb{T}}(x_i \mapsto T_i, \psi(\vec{x}))$.*

Where ρ maps $a \in \mathcal{N}_{\mathfrak{A}}(\upsilon(\vec{x}), r)$ to T_i, where $1 \leq i \leq n$ is such that $a \in \mathcal{N}_{\mathfrak{A}}(\upsilon(x_i), r)$.

Now that we have studied the translation between ordinary formulas and typed ones, let us prove the main structural property of typed formulas that is the reason for their introduction in the first place: one can separate types.

When we say that ψ and φ are equivalent typed formulas, we mean that both have the same set $\vec{x}$ of free variables, and that for all $\mathfrak{A} \in \mathsf{Struct}(\sigma)$, for all $\rho\colon \mathfrak{A} \to \mathbb{T}$, for all $\upsilon\colon \vec{x} \to \mathfrak{A}$, $\mathfrak{A}, \rho, \upsilon \models^{\mathbb{T}} \varphi$ if and only if $\mathfrak{A}, \rho, \upsilon \models^{\mathbb{T}} \psi$.

Lemma 4.2.19. *Let σ be a relational signature, and let $\mathbb{T}$ be a set. Every $\mathbb{T}$-typed formula is equivalent over $\mathsf{Struct}(\sigma)$ to a positive Boolean combination of monotyped formulas, i.e., formulas using a single type $T \in \mathbb{T}$.*

Furthermore, the quantifier rank of the formulas obtained in the process is bounded by the original quantifier rank of the typed formula.

Proof Sketch. We prove the result by induction on the typed formula. The cases of disjunction, negation, and $\top$ are easily handled by induction. Furthermore, for the base case of relational symbols, a typed formula $R(y_1 : T, \dots, y_n : T)$ is always a monotyped formula by definition. As a consequence, the only interesting case is the existential quantification.

Let $\psi = \exists x : T.\theta$. By induction hypothesis, θ can be rewritten as a finite positive Boolean combination of monotyped formulas. Notice that $\exists x : T.(\theta_1 \wedge \theta_2)$ is equivalent to $(\exists x : T.\theta_1) \wedge \theta_2$ whenever θ_2 does not contain the type T. Similarly, $\exists x : T.(\theta_1 \vee \theta_2)$ is equivalent to $(\exists x : T.\theta_1) \vee \theta_2$ whenever θ_2 does not contain the type T.[13]

13: This is because a sentence that does not use the type T cannot contain the variable x of type T, and its evaluation does not depend on the value given to x.

As a consequence, ψ is equivalent to a positive Boolean combination of monotyped formulas, where the $\exists x : T.$ quantification is applied to the monotyped formulas with type T, and the others are left untouched. ∎

Now, we are ready to state the first conversion lemma, that transforms an existential local sentence into a positive Boolean combination of asymmetric basic local sentences. This is done by following the proof scheme of Figure 4.7, which starts by "typing" the local formula inside the existential local sentence, uses the combinatorial tools developed on typed sentences to provide monotyped formulas, which are then turned into regular untyped formulas by erasing the type annotations.

Lemma 4.2.20. *Every almost basic local sentence is equivalent to a finite disjunction of asymmetric basic local sentences.*

Proof. Let $r > 0$. Let $\varphi \overset{\mathsf{def}}{=} \exists \vec{x}.\mathrm{indep}^{\mathcal{N}}(\vec{x},r) \wedge \psi(\vec{x})$, where ψ is a r-local formula around $\vec{x}$, and let $\mathbb{T} \overset{\mathsf{def}}{=} \{\, T_x : x \in \vec{x} \,\}$.

Let $\mathfrak{A} \in \mathsf{Struct}(\sigma)$, and $\upsilon\colon \vec{x} \to \mathfrak{A}$ such that $\mathfrak{A}, \upsilon \models \mathrm{indep}^{\mathcal{N}}(\vec{x},r)$. Let $\rho\colon \mathcal{N}_{\mathfrak{A}}(\upsilon(\vec{x}), r) \to \mathbb{T}$ that associates with a the unique T_x such that $a \in \mathcal{N}_{\mathfrak{A}}(\upsilon(x), r)$. Thanks to Corollary 4.2.18, we obtain the following equivalence:

$$\begin{aligned} & (\mathcal{N}_{\mathfrak{A}}(\upsilon(\vec{x}), r), \upsilon \models \psi(\vec{x})) \\ \iff & (\mathcal{N}_{\mathfrak{A}}(\upsilon(\vec{x}), r), \rho, \upsilon \models^{\mathbb{T}} \mathrm{conv}^{\mathbb{T}}(x \mapsto T_x, \psi(\vec{x}))) \end{aligned} \tag{4.2}$$

We are now going to use Lemma 4.2.19 to rewrite the $\mathbb{T}$-typed formula $\mathrm{conv}^{\mathbb{T}}(x \mapsto T_x, \psi(\vec{x}))$ into a positive Boolean combination of monotyped formulas. Because every free variable of ψ is assigned a different type in Γ, and the "type conversion" $\mathrm{conv}^{\mathbb{T}}(,)$ respects this property (see Fact 4.2.16), these monotyped formulas have at most one free variable! It is now possible to write:

$$\mathrm{conv}^{\mathbb{T}}(x \mapsto T_x, \psi(\vec{x})) \equiv \bigvee_{i=1}^{n} \bigwedge_{x \in \vec{x}} \tau_{n,x}(x : T_x) \tag{4.3}$$

By definition, $\psi_{i,x} \stackrel{\text{def}}{=} |\mathrm{U}\mathbb{T}(\tau_{i,x}(x : T_x))|_x^r$ is an r-local formula around its single free variable x.

To conclude, it suffices to prove that $\exists\vec{x}.\mathrm{indep}^{\mathcal{N}}(\vec{x},r) \wedge \psi(\vec{x})$ is equivalent over $\mathsf{Struct}(\sigma)$ to the following disjunction of asymmetric basic local sentences:

$$\bigvee_{i=1}^{n} \exists\vec{x}.\mathrm{indep}^{\mathcal{N}}(\vec{x},r) \wedge \bigwedge_{x \in \vec{x}} \psi_{i,x}(x) \quad .$$

To obtain such an equivalence, let us fix $\vec{x}$ such that $\mathfrak{A}, \upsilon \models \mathrm{indep}^{\mathcal{N}}(\vec{x},r)$ and combine the results gathered so far as follows:

$$\begin{aligned}
&\mathfrak{A}, \upsilon \models \psi(\vec{x}) \\
&\iff \mathcal{N}_{\mathfrak{A}}(\upsilon(\vec{x}), r), \upsilon \models \psi(\vec{x}) && r\text{-local} \\
&\iff \mathcal{N}_{\mathfrak{A}}(\upsilon(\vec{x}), r), \rho, \upsilon \models^{\mathbb{T}} \mathrm{conv}^{\mathbb{T}}(x \mapsto T_x, \psi(\vec{x})) && (4.2) \\
&\iff \mathcal{N}_{\mathfrak{A}}(\upsilon(\vec{x}), r), \rho, \upsilon \models^{\mathbb{T}} \bigvee_{i=1}^{n} \bigwedge_{x \in \vec{x}} \tau_{i,x}(x : T_x) && (4.3) \\
&\iff \bigvee_{i=1}^{n} \bigwedge_{x \in \vec{x}} (\mathcal{N}_{\mathfrak{A}}(\upsilon(\vec{x}), r), \rho, \upsilon \models^{\mathbb{T}} \tau_{i,x}(x : T_x)) && (\models^{\mathbb{T}}) \\
&\iff \bigvee_{i=1}^{n} \bigwedge_{x \in \vec{x}} (\mathcal{N}_{\mathfrak{A}}(\upsilon(x), r), \rho, \upsilon \models^{\mathbb{T}} \tau_{i,x}(x : T_x)) && (\models^{\mathbb{T}}) \\
&\iff \bigvee_{i=1}^{n} \bigwedge_{x \in \vec{x}} (\mathcal{N}_{\mathfrak{A}}(\upsilon(x), r), \upsilon \models \mathrm{U}\mathbb{T}(\tau_{i,x}(x : T_x))) && (4.2.13) \\
&\iff \bigvee_{i=1}^{n} \bigwedge_{x \in \vec{x}} (\mathfrak{A}, \upsilon \models |\mathrm{U}\mathbb{T}(\tau_{i,x}(x : T_x))|_x^r) && (|\cdot|) \\
&\iff \bigvee_{i=1}^{n} \bigwedge_{x \in \vec{x}} (\mathfrak{A}, \upsilon \models \psi_{i,x}) && (\psi_{i,x}) \\
&\iff \bigvee_{i=1}^{n} \left[\mathfrak{A}, \upsilon \models \left(\bigwedge_{x \in \vec{x}} \psi_{i,x}\right)\right] && (\models)
\end{aligned}$$

We conclude by noticing that $\mathfrak{A} \models \exists\vec{x}.\mathrm{indep}^{\mathcal{N}}(\vec{x},r) \wedge \psi(\vec{x})$ if and only if there exists υ such that $\mathfrak{A}, \upsilon \models \mathrm{indep}^{\mathcal{N}}(\vec{x},r)$, and $\mathfrak{A}, \upsilon \models \psi$, which is equivalent to stating that $\mathfrak{A}, \upsilon \models \mathrm{indep}^{\mathcal{N}}(\vec{x},r)$, and that there exists $1 \leq i \leq n$, such that $\mathfrak{A}, \upsilon \models \bigwedge_{x \in \vec{x}} \psi_{i,x}(x)$. Hence, $\mathfrak{A} \models \exists\vec{x}.\mathrm{indep}^{\mathcal{N}}(\vec{x},r) \wedge \psi(\vec{x})$ if and only if $\mathfrak{A} \models \bigvee_{1 \leq i \leq n} \exists\vec{x}.\mathrm{indep}^{\mathcal{N}}(\vec{x},r) \wedge \bigwedge_{x \in \vec{x}} \psi_{i,x}(x)$. ■

The combination of Lemma 4.2.20 and Lemma 4.2.9 states that every existential local sentence is equivalent to a finite disjunction of asymmetric basic local sentences. This was previously known for existential sentences[14] as a consequence of the existential Gaifman normal form

14: i.e., for existential 0-local sentences.

introduced in [55, Theorem 6].

4.2.2. From Asymmetric Basic Local to Basic Local Sentences

In order to prove that every existential local sentence is equivalent to a positive Boolean combination of basic local sentences (Theorem 4.2.2), it suffices to prove it for asymmetric basic local sentences thanks to Lemmas 4.2.9 and 4.2.20. This setting is a generalisation of [55, Lemma 4] that deals with existential sentences.

Removing variables. The first step of our construction is to notice that if a local property appears frequently in a structure, then it can be selected independently of other local properties. This was already noticed in [55] and is a fairly standard argument. To make the above statement precise, let us first formally define how to remove a local property from an asymmetric basic local sentence.

Definition 4.2.21. Let $\varphi \overset{\text{def}}{=} \exists \vec{x}.\mathrm{indep}^{\mathcal{N}}(\vec{x},r) \wedge \bigwedge_{x \in \vec{x}} \psi_x(x)$ be an asymmetric basic local sentence. For each $x \in \vec{x}$, the sentence $\varphi \ / \ x$ is defined as $\exists \vec{y}.\mathrm{indep}^{\mathcal{N}}(\vec{y},r) \wedge \bigwedge_{y \in \vec{y}} \psi_y(y)$, where $\vec{y} \overset{\text{def}}{=} \vec{x} \setminus \{ x \}$.

Example 4.2.22. Let ψ_1 and ψ_2 be two r-local formulas, and $\varphi \overset{\text{def}}{=} \exists x_1, x_2.\mathrm{indep}^{\mathcal{N}}(x_1 x_2,r) \wedge \psi_1(x_1) \wedge \psi_2(x_2)$. Then, $\varphi \ / \ x_1 = \exists x_2.\psi_2(x_2)$, and $\varphi \ / \ x_2 = \exists x_1.\psi_1(x_1)$.

It is immediate that the sentences obtained by removing a local property to check are weaker than the original sentence. Let us state this fact formally so that we can refer to it later on.

Fact 4.2.23. Let φ be an asymmetric basic local sentence with existential variables $\vec{x}$. Let $\mathfrak{A} \in \mathsf{Struct}(\sigma)$ such that $\mathfrak{A} \models \varphi$. Then, $\mathfrak{A} \models \varphi \ / \ x$ for all $x \in \vec{x}$.

We now have all the tools needed to formally state what was hinted at earlier: frequent properties can be checked independently of the rest in an asymmetric basic local sentence. We restate the proof of the following result for readability.

Lemma 4.2.24 [55, Step 1 of Lemma 4]**.** Let φ be an asymmetric basic local sentence of the following form:

$$\exists x_1, \dots, x_k.\mathrm{indep}^{\mathcal{N}}(\vec{x},r) \wedge \bigwedge_{x \in \vec{x}} \psi_x(x) \quad .$$

Let $\mathfrak{A} \in \mathsf{Struct}(\sigma)$ and $x \in \vec{x}$ be such that $\mathfrak{A} \models \varphi \ / \ x$ and $\mathfrak{A} \models \exists^{\geq k}_{2r} y.\psi_x(y)$. Then, $\mathfrak{A} \models \varphi$.

Proof. Let us write $\vec{z} \stackrel{\text{def}}{=} \vec{x} \setminus \{x\}$. Because $\mathfrak{A} \models \varphi \mathbin{/} x$, there exists a valuation $\upsilon\colon \vec{z} \to \mathfrak{A}$, such that the points in $\upsilon(\vec{z})$ have non-intersecting r-neighbourhoods in $\mathfrak{A}$, and $\mathcal{N}_{\mathfrak{A}}(\upsilon(y), r), \upsilon \models \psi_y(y)$ for all $y \in \vec{z}$. As a consequence, we only need to find a point $a \in \mathfrak{A}$ such that $\mathcal{N}_{\mathfrak{A}}(a, r) \cap \mathcal{N}_{\mathfrak{A}}(\upsilon(\vec{z}), r) = \emptyset$ and $\mathfrak{A}, y \mapsto a \models \psi_x(y)$ to conclude that $\mathfrak{A} \models \varphi$.

Because $\mathfrak{A} \models \exists^{\geq k}_{2r} y.\psi_x(y)$, there exists k points $a_1, \dots, a_k = \vec{a}$ in $\mathfrak{A}$ whose $2r$-neighbourhoods do not intersect, and such that $\mathfrak{A}, y \mapsto a_i \models \psi_x(y)$ for all $1 \leq i \leq k$. Let us show that at least one of them is far from $\upsilon(\vec{z})$.

Assume by contradiction that for all $a \in \vec{a}$, the point a is "too close" to the vector $\upsilon(\vec{z})$, that is, $\mathcal{N}_{\mathfrak{A}}(a, r) \cap \mathcal{N}_{\mathfrak{A}}(\upsilon(\vec{z}), r) \neq \emptyset$. Since $|\vec{a}| = k$ and $|\upsilon(\vec{z})| = k - 1$, there exists a pair of points $a_1 \neq a_2 \in \vec{a}$ and one variable $y \in \vec{z}$ such that, $\mathcal{N}_{\mathfrak{A}}(a_1, r) \cap \mathcal{N}_{\mathfrak{A}}(\upsilon(y), r) \neq \emptyset$ and $\mathcal{N}_{\mathfrak{A}}(a_2, r) \cap \mathcal{N}_{\mathfrak{A}}(\upsilon(y), r) \neq \emptyset$. As a consequence, $\upsilon(y) \in \mathcal{N}_{\mathfrak{A}}(a_1, 2r) \cap \mathcal{N}_{\mathfrak{A}}(a_2, 2r)$, and this contradicts the assumption on the vector $\vec{a}$. ■

Let us now take a step back and explain the proof scheme that we will employ in the upcoming Lemma 4.2.38 that rewrites asymmetric basic local sentences into positive Boolean combination of basic local sentences. The proof will be done by induction on the number of outer existential quantifiers as follows: given an asymmetric basic local sentence φ with outer existentially quantified variables $\vec{x}$, we will define $\varphi' \stackrel{\text{def}}{=} \bigvee_{x \in \vec{x}} \left[\varphi \mathbin{/} x \wedge \exists^{\geq k}_{2r} y.\psi_x(y)\right]$. Thanks to Fact 4.2.23, $\varphi' \Rightarrow \varphi$, and we can leverage the induction hypothesis to rewrite $\varphi \mathbin{/} x$ as a positive Boolean combination of basic local sentences. To conclude, it suffices to rewrite $\varphi \wedge \neg\varphi'$ as a positive Boolean combination of basic local sentences, which is easier because its models have a peculiar property: for all $\mathfrak{A} \models \varphi \wedge \neg\varphi'$, the set $W \stackrel{\text{def}}{=} \{ a \in \mathfrak{A} \colon \exists x \in \vec{x}. \mathfrak{A}, x \mapsto a \models \psi_x(x) \}$ is sparse, in that it is not possible to find more than $k(k-1)$ points in W whose $2r$-neighbourhoods do not intersect.

For clarity, let us call r-local formulas with exactly one free variable *r-local properties*.

Fact 4.2.25. Let $k, r > 0$, $\psi_1(x), \dots, \psi_k(x)$ be r-local properties, and $\mathfrak{A}$ be such that for all $1 \leq i \leq k$, $\mathfrak{A} \not\models \exists^{\geq k}_{2r} x.\psi_i(x)$. Then, the set $W \stackrel{\text{def}}{=} \{ a \in \mathfrak{A} \colon \exists x \in \vec{x}. \mathfrak{A}, x \mapsto a \models \psi_x(x) \}$ has the following property:

$$\begin{aligned} &\forall \vec{a} \in \mathfrak{A}^{k(k-1)+1}, \\ &\exists 1 \leq l \neq p \leq k(k-1)+1, \\ &\qquad \mathcal{N}_{\mathfrak{A}}(a_l, 2r) \cap \mathcal{N}_{\mathfrak{A}}(a_p, 2r) \neq \emptyset \quad . \end{aligned}$$

Abstractly, we will therefore be interested, given a finite (non-empty) set P_r of r-local properties, in the possible ways that the set $W \stackrel{\text{def}}{=} \{ a \in \mathfrak{A} \colon \mathfrak{A}, x \mapsto a \models \bigvee_{p \in \mathsf{P}_r} p(x) \}$ can be found in a structure $\mathfrak{A} \in \mathsf{Struct}(\sigma)$ where $|W|$ is guaranteed to be bounded by some K (thanks to Fact 4.2.25). Intuitively, there are only finitely many possible ways that a structure $\mathfrak{A}$ can represent a given set of witnesses W, and the rest of this section is devoted to formally proving this intuition.

Template Graphs. In order to represent sets of witnesses and their local relationship inside a given structure, we propose the following abstraction.

Definition 4.2.26. *Let $r, R \geq 0$, $\mathfrak{A} \in \mathsf{Struct}(\sigma)$, and P_r be a non-empty finite set of r-local properties. The (R, P_r)-template graph $\mathsf{G}^R_{\mathsf{P}_r}(\mathfrak{A})$ of $\mathfrak{A}$ is an edge-labelled coloured graph whose vertices are the elements of $\mathfrak{A}$ that satisfy at least one property in P_r. It has labelled edges $(u, v, \mathrm{d}_{\mathfrak{A}}(u, v))$ whenever $\mathrm{d}_{\mathfrak{A}}(u, v) \leq R$, and each vertex a is coloured by the set of $q_r(x) \in \mathsf{P}_r$ such that $\mathfrak{A}, x \mapsto a \models q_r(x)$.*

In a template graph, an edge has exactly one label, but a vertex can have multiple colours. We write $\mathsf{props}^{\mathsf{P}_r}(v)$ for the collection of r-local properties that colour the vertex v, which happens to always be non-empty.

Example 4.2.27. Let $\sigma \stackrel{\text{def}}{=} \{ (E, 2); (P, 1) \}$, $r = R = 2$, $\psi_1(x) \stackrel{\text{def}}{=} \exists y \in \mathcal{N}(x, 2).E(x, y) \wedge P(y)$, and $\psi_2(x) \stackrel{\text{def}}{=} \forall y \in \mathcal{N}(x, 2).\neg P(y)$. Let $\mathsf{P}_2 \stackrel{\text{def}}{=} \{ \psi_1, \psi_2 \}$. Let G be the relational structure in the first frame of Figure 4.8. Then, the associated template graph $\mathsf{G}^2_{\mathsf{P}_2}(G)$ is described in the last frame of Figure 4.8.

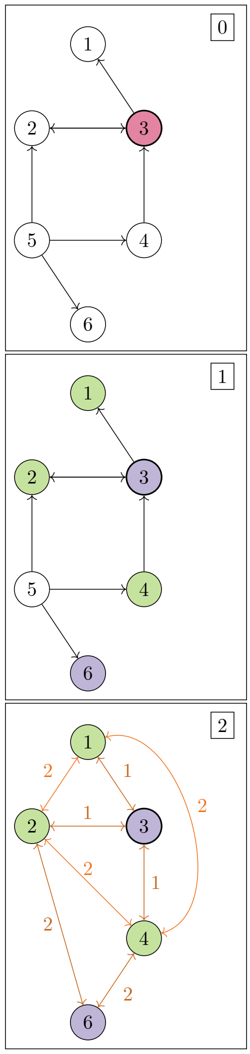


Figure 4.8.: An example of relational structure G in the first frame, and associated template graph $\mathsf{G}^2_{\{ \psi_1, \psi_2 \}}(G)$ in the last frame. Black directed arrows represent the relation $E(x, y)$, while the red colour represents the relation $P(x)$. Nodes are coloured in green when they satisfy ψ_1, and in violet when they satisfy ψ_2 of Example 4.2.27. The labelled edges of the template graph are represented by orange bidirectional arrows labelled with the distance d.

Because we will be interested in structures where the set of witnesses to the local properties is sparse, we will restrict our attention to template graphs that have a bounded number of vertices.

Definition 4.2.28. *Let $K \in \mathbb{N}$, $R, r > 0$, and P_r be a finite set of r-local properties. Let us define $\mathsf{Graphs}^{\mathsf{P}_r}_R(K)$ to be the set of (R, P_r)-template graphs having at most K vertices.*

Given a structure $\mathfrak{A}$, checking whether it satisfies some asymmetric basic local sentence amounts to computing the associated template graph and finding a tuple of vertices that have the right colouring and respect the distances imposed by the sentence. Therefore, one can collect, given an asymmetric basic local sentence φ the template graphs of models of φ. Notice that given a template graph of a model of φ, one "add new local properties" while remaining a witness that φ holds. This is the motivation for the following definition.

Definition 4.2.29. *Let $K \in \mathbb{N}$, $R, r > 0$, and P_r be a finite set of r-local properties. We order template graphs $G, G' \in \mathsf{Graphs}^{\mathsf{P}_r}_R(K)$ using $G \leq G'$ if and only if there exists a embedding $h \colon G \to G'$ between the underlying graphs respecting the edge labels, and such that $\mathsf{props}^{\mathsf{P}_r}(v) \subseteq \mathsf{props}^{\mathsf{P}_r}(h(v))$ for all $v \in G$.*

Given a structure $\mathfrak{A}$ and a tuple $\vec{a} \in \mathfrak{A}$, we can build the template graph $\mathsf{G}^{\mathsf{P}_r}_R(\mathfrak{A})$ intersected with the vertex set $\vec{a}$. This is $\mathsf{G}^{\mathsf{P}_r}_R(\mathfrak{A}, \vec{a})$.

Definition 4.2.30. *Let $\mathfrak{A}$ be a structure. A template graph G is represented by a vector $\vec{a} \in \mathfrak{A}$ whenever $G \leq \mathsf{G}_R^{\mathsf{P}_r}(\mathfrak{A}, \vec{a})$.*

We refer to Figure 4.9 for an example of a template graph represented by a given vector. Let us put emphasis on the fact that we are allowed to add vertices and properties when representing a graph.

Our first lemma after that long series of definitions is that template graphs can be represented in first-order logic via a kind of diagram sentence. That is, there is a local formula checking whether a given template graph can be represented using some tuple in a given neighbourhood of a structure.

Because the collection of template graphs of bounded size is finite, we can build a sentence stating that we contain at least one of these patterns.

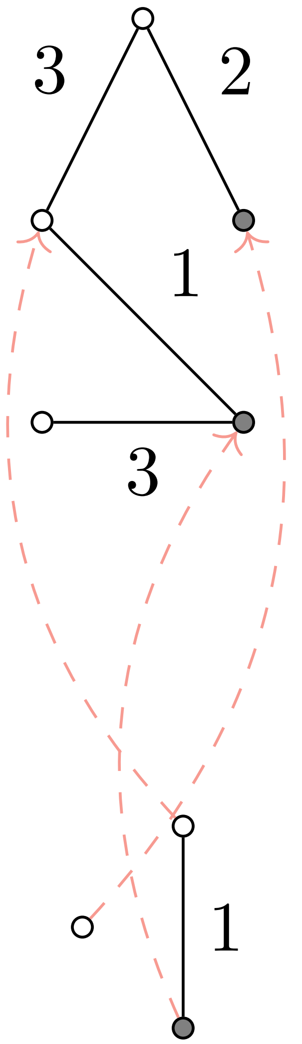


Figure 4.9.: A template graph (below) that is represented inside a bigger graph (above).

Lemma 4.2.31. *Let $K, R, r > 0$, P_r be a finite set of r-local properties, and $G \in \mathsf{Graphs}_R^K(\mathsf{P}_r)$. There exists an $(r + R)$-local formula $\mathsf{Diagram}_G^{\mathsf{P}_r,R}(x)$ such that, for all $\mathfrak{A} \in \mathsf{Struct}(\sigma)$, and for all $a \in \mathfrak{A}$ the following are equivalent:*

1. $\mathfrak{A}, x \mapsto a \models \mathsf{Diagram}_G^{\mathsf{P}_r,R}(x)$;
2. $\exists \vec{b} \in \mathcal{N}_{\mathfrak{A}}(a, R), G \leq \mathsf{G}_R^{\mathsf{P}_r}(\mathfrak{A}, \vec{b})$, *and* $\mathcal{N}_{\mathfrak{A}}(\vec{b}, r) \subseteq \mathcal{N}_{\mathfrak{A}}(a, R)$.

Proof. Let us define the formula as follows:

$$\begin{aligned}
\mathsf{Diagram}_G^{\mathsf{P}_r,R}(x) \stackrel{\text{def}}{=} \exists v_1, \dots, v_{\mathrm{V}(G)} \in \mathcal{N}(x, R).& \bigwedge_{(v_i,v_j,h)\in \mathrm{E}(G)} d(v_i, v_j) = h \\
&\wedge \bigwedge_{v_i \in \mathrm{V}(G)} \bigwedge_{p \in \mathsf{props}^{\mathsf{P}_r}(v_i)} p(v_i) \\
&\wedge \bigwedge_{v_i \in \mathrm{V}(G)} \mathcal{N}(v_i, r) \subseteq \mathcal{N}(x, R)\,.
\end{aligned}$$

It is immediate that this is an $(R + r)$-local formula, moreover, the stated equivalence holds by definition. ∎

It may be time to reflect on what has been done so far. Recall that we are trying to provide a combinatorial description of the ways in which witnesses of an asymmetric basic local sentence can be found in a structure, provided that these witnesses are “sparse.” We built template graphs to abstract the relationship between these witnesses inside real structures, and designed ways to talk about these templates using local formulas. The only remaining problem is that basic local sentences can only count the repetition of some local pattern, and a template graph can be composed of different connected components. We can tackle this problem by studying how the template graph of a union of a tuple $\vec{a}\vec{b}$ of points can be obtained by combining the template graphs of $\vec{a}$ and $\vec{b}$, building towards an inductive construction of the template graphs.

In general, it is not immediately clear how $\mathsf{G}_R^{\mathsf{P}_r}(\mathfrak{A},\vec{a})$ and $\mathsf{G}_R^{\mathsf{P}_r}(\mathfrak{A},\vec{b})$ can be used to recover the template graph $\mathsf{G}_R^{\mathsf{P}_r}(\mathfrak{A},\vec{a}\vec{b})$. However, the construction is quite simple if we assume that $\vec{a}$ and $\vec{b}$ are "far enough" apart.

Fact 4.2.32. Let $r, R > 0$, P_r be a non-empty finite set of r-local properties, $\mathfrak{A} \in \mathsf{Struct}(\sigma)$, and $\vec{a}, \vec{b}$ be vectors of elements of $\mathfrak{A}$, such that $\mathcal{N}_{\mathfrak{A}}(\vec{a}, R)$ and $\mathcal{N}_{\mathfrak{A}}(\vec{b}, R)$ do not intersect. Then,

$$\mathsf{G}_R^{\mathsf{P}_r}(\mathfrak{A},\vec{a}\vec{b}) = \mathsf{G}_R^{\mathsf{P}_r}(\mathfrak{A},\vec{a}) \uplus \mathsf{G}_R^{\mathsf{P}_r}(\mathfrak{A},\vec{b}) \quad .$$

The problem is now to state (using local formulas) that the template graphs that we extract in a given structure are far enough apart, so that we can use Fact 4.2.32 later on. To that end, let us introduce a sentence that does exactly this check in Definition 4.2.33. We will prove that this definition has the intended semantics, in Lemma 4.2.34.

Definition 4.2.33. *Let $r, R > 0$, and P_r be a non-empty finite set of r-local properties. Let us define the security cylinder formula as the following $\max(3R, R + r)$-local formula:*

$$\begin{aligned}\mathsf{Security}_R^{\mathsf{P}_r}(x) \stackrel{\text{def}}{=} &\forall y \in \mathcal{N}(x, 3R). \\ &\left(\bigvee_{p \in \mathsf{P}_r} p(y)\right) \Rightarrow \mathcal{N}(y, r) \subseteq \mathcal{N}(x, R)\,.\end{aligned}$$

Lemma 4.2.34. *Let $R, r > 0$, P_r be a non-empty finite set of r-local properties, and $G \in \mathsf{Graphs}_{\mathsf{P}_r}^{K}(R)$. The following implication holds over $\mathsf{Struct}(\sigma)$:*

$$\left(\mathsf{Diagram}_G^{\mathsf{P}_r,R}(x) \wedge \mathsf{Security}_R^{\mathsf{P}_r}(y) \wedge \mathrm{d}(x,y) \leq 2R\right) \Rightarrow \mathsf{Diagram}_G^{\mathsf{P}_r,R}(y)$$

Proof. Let $\mathfrak{A} \in \mathsf{Struct}(\sigma)$ and $\upsilon \colon \{x, y\} \to \mathfrak{A}$ be a valuation such that $\mathfrak{A}, \upsilon \models \mathsf{Diagram}_G^{\mathsf{P}_r,R}(x) \wedge \mathsf{Security}_R^{\mathsf{P}_r}(y) \wedge \mathrm{d}(x,y) \leq 2R$.

By Lemma 4.2.31, there exists a vector $\vec{b} \in \mathcal{N}_{\mathfrak{A}}(\upsilon(x), R)$ such that $G \leq \mathsf{G}_R^{\mathsf{P}_r}(\mathfrak{A}, \vec{b})$, and $\mathcal{N}_{\mathfrak{A}}(\vec{b}, r) \subseteq \mathcal{N}_{\mathfrak{A}}(\upsilon(x), R)$.

In particular, every point of $\vec{b}$ satisfies at least one property $p \in \mathsf{P}_r$. As $\mathrm{d}_{\mathfrak{A}}(\upsilon(x), \upsilon(y)) \leq 2R$, $\vec{b} \in \mathcal{N}_A(\upsilon(y), 3R)$.

Because, $\mathfrak{A}, \upsilon \models \mathsf{Security}_R^{\mathsf{P}_r}(y)$, we conclude that $\mathcal{N}_{\mathfrak{A}}(\vec{b}, r) \subseteq \mathcal{N}_{\mathfrak{A}}(\upsilon(y), R)$.

Using Lemma 4.2.31, this time in the converse direction, we conclude that $\mathfrak{A}, \upsilon \models \mathsf{Diagram}_G^{\mathsf{P}_r,R}(y)$. ■

Putting it all together. We have partially solved the problem of the local presence of template graphs in a given structure, but basic local sentences have another restriction: we can only ask for the repetition of the *same* local property. Therefore, we have to "encode" the presence of some template graph as the repetition of identical local properties.

To that end, we will consider finite multisets of template graphs. The collection of finite multisets with at most p elements over a set X is written $\mathsf{M}_p^\diamond(X)$ and is a subset of functions $M \colon X \to \mathbb{N}$ with finite support, and such that $\sum_{x \in X} M(x) \leq p$. We identify $\{ a \}$ with the finite multiset M that maps a to 1 and every other element to 0, and write $M_1 + M_2$ for the pointwise addition of the two multisets.

It is not strictly necessary to use $\mathcal{P}_{\mathsf{fin}}(\Delta)$ in the Definition 4.2.35 because the latter is finite, but the notation puts emphasis on the fact that everyting we manipulate is finite.

Definition 4.2.35. *Let $R, r, k, p > 0$, P_r be a finite non-empty set of r-local properties, $\Delta \subseteq \mathsf{Graphs}^k_{\mathsf{P}_r}(R)$, and $M \in \mathsf{M}_p^\diamond(\mathcal{P}_{\mathsf{fin}}(\Delta))$ be a finite multiset with at most p elements. We define*

$$\Theta_R^M \stackrel{\mathsf{def}}{=} \bigwedge_{S \in \mathcal{P}_{\mathsf{fin}}(\Delta)} \exists^{\geq M(S)}_{3R} x, \mathsf{Security}_R^{\mathsf{P}_r}(x) \wedge \bigwedge_{G \in S} \mathsf{Diagram}_G^{\mathsf{P}_r, R}(x) \wedge \wedge \bigwedge_{G' \in \Delta \setminus S} \neg \mathsf{Diagram}_{G'}^{\mathsf{P}_r, R}(x) \,.$$

Remark 4.2.36. The sentence Θ_R^M is a conjunction of basic local sentences.

The intuition behind Definition 4.2.35 is that we want to find enough witnesses of the subsets $S \subseteq \Delta$, in completely independent areas of the structure.

Imagine that a structure satisfies Θ_R^M for some multiset, then it must represent certain template graphs. Let us formally describe this set.

The pattern $\{ S \} + M'$ may have several matches in Definition 4.2.37, however we claim that the order in which the sets S are selected does not change the output of obt.

Definition 4.2.37. *Let $R, r, k, p > 0$, P_r be a finite non-empty set of r-local properties, $\Delta \subseteq \mathsf{Graphs}^k_{\mathsf{P}_r}(R)$, and $M \in \mathsf{M}_p^\diamond(\mathcal{P}_{\mathsf{fin}}(\Delta))$ be a finite multiset with at most p elements. The set $\mathrm{obt}(M)$ of graphs obtainable from M is defined inductively as follows:*

- $\mathrm{obt}(\emptyset) \stackrel{\mathsf{def}}{=} \emptyset$*, and*
- $\mathrm{obt}(\{ S \} + M')$ *is defined as the collection of graphs G, G', and $G \uplus G'$ for $G \in S$, and $G' \in \mathrm{obt}(M')$.*

The next lemma finishes the conversion from existential local sentences to sentences in positive Gaifman normal form. The main technicality is proving that Definition 4.2.37 correctly represents the template graphs that are found inside the structures.

Lemma 4.2.38. *Every asymmetric basic local sentence is equivalent to a positive Boolean combination of basic local sentences.*

Proof. Let φ be of the form $\exists x_1, \ldots, x_k . \mathrm{indep}^{\mathcal{N}}(\vec{x}, r) \wedge \bigwedge_{i=1}^k \psi_i(x_i)$ where $(\psi_i)_{1 \leq i \leq k}$ is a sequence of r-local formulas. We prove by induction over k that φ is equivalent to a positive Boolean combination of basic local sentences. When $k \leq 1$, φ is already a basic local sentence hence we assume $k \geq 2$.

For $1 \leq i \leq k$, we apply the induction hypothesis on $\varphi \,/\, x_i$, which has fewer existentially quantified variables, and call $\overline{\varphi \,/\, i}$ the obtained positive Boolean combination of basic local sentences.

Let us define $\mathsf{P}_r \stackrel{\text{def}}{=} \{\, \psi_i \colon 1 \leq i \leq k \,\}$ which is a non-empty finite set of r-local properties, and $\Delta \stackrel{\text{def}}{=} \mathsf{Graphs}^{\mathsf{P}_r}_{4^{k^2}6r}(k)$.

Let us define $\mathbb{M}_V$ as the set of multisets $M \in \mathsf{M}^{\diamond}(\mathcal{P}_{\mathsf{fin}}(\Delta))k(k-1)$ such that $\mathrm{obt}(M)$ contains some template graph G, which itself contains vertices $v_1, \ldots, v_k$, at pairwise distance greater than $2r$, that are such that $\psi_i \in \mathsf{props}^{\mathsf{P}_r}(v_i)$ for all $1 \leq i \leq k$.

That is, $\mathbb{M}_V$ is the collection of multisets from which one can obtain a template graph that can then be used as a witness that $\mathfrak{A} \models \varphi$.

Let us prove that φ is equivalent to the positive Boolean combination of basic local sentences Ψ defined as

$$\Psi \stackrel{\text{def}}{=} \underbrace{\left(\bigvee_{1 \leq i \leq k} \overline{\varphi \,/\, i} \land \exists^{\geq k}_{2r} x.\psi_i(x) \right)}_{\stackrel{\text{def}}{=} \Psi_1} \lor \underbrace{\left(\bigvee_{M \in \mathbb{M}_V} \bigvee_{6r \leq R \leq 4^{k^2}6r} \Theta^M_R \right)}_{\stackrel{\text{def}}{=} \Psi_2} .$$

Direction $\varphi \Rightarrow \Psi$: Assume that $\mathfrak{A} \models \varphi$. Using Fact 4.2.23, $\mathfrak{A} \models \varphi \,/\, x_i$ for all $1 \leq i \leq k$ and by the induction hypothesis this proves that $A \models \overline{\varphi \,/\, i}$ for all $1 \leq i \leq k$.

If $\mathfrak{A} \models \exists^{\geq k}_{2r} x.\psi_i(x)$ for some $1 \leq i \leq k$, then $\mathfrak{A} \models \Psi$ and we are done.

Otherwise, $\mathfrak{A} \not\models \exists^{\geq k}_{2r} x.\psi_i(x)$ for all $1 \leq i \leq k$, and the remaining goal is to show that for such a structure, $\mathfrak{A} \models \Psi_2$.

In such a structure $\mathfrak{A}$, let us call W the set of elements in $\mathfrak{A}$ that satisfy at least one ψ_i. It is not possible to find more than $k(k-1)$ points in W whose neighbourhoods of radius $2r$ do not intersect, thanks to Fact 4.2.25. This implies that there exists a vector $\vec{c}$ in A of size at most $k(k-1)$ such that $W \subseteq \mathcal{N}_{\mathfrak{A}}(\vec{c}, 6r)$.[15]

Using Lemma 4.2.8 over $\vec{c}$ and $6r$ one obtains a radius $6r \leq R \leq 4^{k^2}6r$ and a vector $\vec{b}$ such that $\mathcal{N}_A(\vec{c}, 6r) \subseteq \mathcal{N}_A(\vec{b}, R)$ and the neighbourhoods of radius $3R$ around the points in $\vec{b}$ do not intersect.

15: It suffices to consider a set $\vec{c}$ of maximal size W with non-intersecting $2r$-neighbourhoods. We know that $|\vec{c}| \leq k(k-1)$, and that for all $a \in W$, $\mathcal{N}_{\mathfrak{A}}(a, 2r) \cap \mathcal{N}_{\mathfrak{A}}(\vec{c}, 2r) \neq \emptyset$. As a consequence, $\mathcal{N}_{\mathfrak{A}}(a, 2r) \subseteq \mathcal{N}_{\mathfrak{A}}(\vec{c}, 6r)$.

Let $b \in \vec{b}$. Let us define S_b as the set of template graphs $\mathsf{G}^{\mathsf{P}_r}_R(\mathfrak{A}, \vec{a})$ where $\vec{a}$ ranges over the tuples of k points inside $W \cap \mathcal{N}_{\mathfrak{A}}(b, R)$.

Notice if $a \in W$, then $\mathcal{N}_{\mathfrak{A}}(a, 2r) \subseteq \mathcal{N}_{\mathfrak{A}}(\vec{c}, 6r) \subseteq \mathcal{N}_{\mathfrak{A}}(\vec{b}, R)$. As a consequence, the vector $\vec{b}$ can be used to *partition* the elements in W. Furthermore, for all $b \in \vec{b}$, $S_b \in \mathcal{P}_{\mathsf{fin}}(\Delta)$.

Let us now write $M_{\vec{b}}$ for the multiset obtained by collecting all the S_b when b ranges over elements in $\vec{b}$. By construction, $\mathfrak{A} \models \Theta^{M_{\vec{b}}}_R$.[16]

16: Let us briefly sketch why in the margin. For every $S \in \mathcal{P}_{\mathsf{fin}}(\Delta)$, either $S \notin M_{\vec{b}}$, and there is nothing to check, or $S \in M_{\vec{b}}$, and we have proven that there exists precisely enough points in $\vec{b}$ for which $S_b = S$, and with non-intersecting $3R$-neighbourhoods. Notice that by definition of S_b, the only template graphs that occur in the neighbourhood of b are precisely those in S_b. The only property left to check is that b has a proper security cylinder, but this is precisely how the vector $\vec{b}$ was constructed: every template graph is in the $6r$-neighbourhood of some $c \in \vec{c}$, which is itself included in the R-neighbourhood of some $b \in \vec{b}$.

Let us now prove that $M_{\vec{b}} \in \mathbb{M}_V$. Because $\mathfrak{A} \models \varphi$, there exists a valuation $\upsilon \colon \vec{x} \to \mathfrak{A}$, representing a tuple $\vec{a} \stackrel{\text{def}}{=} \upsilon(\vec{x})$ of points at pairwise distance greater than $2r$ such that $\mathfrak{A}, \upsilon \models \bigwedge_{i=1}^{k} \psi_i(x_i)$. Let us now prove that $\mathsf{G}^{\mathsf{P}_r}_R(\mathfrak{A}, \vec{a}) \in \mathrm{obt}(M_{\vec{b}})$, which will prove that $M_{\vec{b}} \in \mathbb{M}_V$.

First, $\vec{a} \in W$, hence $\vec{a} \subseteq \mathcal{N}_{\mathfrak{A}}(\vec{b}, R)$. Because the $3R$-neighbourhoods of elements in $\vec{b}$ do not intersect, one can leverage Fact 4.2.32 and notice that $\mathsf{G}^{\mathsf{P}_r}_R(\mathfrak{A}, \vec{a}) = \biguplus_{b \in \vec{b}} \mathsf{G}^{\mathsf{P}_r}_R(\mathfrak{A}, \vec{a} \cap \mathcal{N}_{\mathfrak{A}}(b, R))$, and conclude that this template graph belongs to $\mathrm{obt}(M_{\vec{b}})$.

We have proven that $\mathfrak{A} \models \Theta_R^{S_{\vec{b}}}$ with $S_{\vec{b}} \in \mathbb{M}_V$ and $6r \leq R \leq 4^{k^2} 6r$, hence $\mathfrak{A} \models \Psi_2$. In particular, $\mathfrak{A} \models \Psi$.

Direction $\Psi \Rightarrow \varphi$: Assume conversely that $\mathfrak{A} \models \Psi = \Psi_1 \vee \Psi_2$. If $\mathfrak{A} \models \Psi_1$, then some local property appears frequently in $\mathfrak{A}$ thanks to the induction hypothesis, and $\mathfrak{A} \models \varphi$ via Lemma 4.2.24.

Otherwise, $\mathfrak{A} \models \Psi_2$, hence $\mathfrak{A} \models \Theta_R^M$ for some $M \in \mathbb{M}_V$ and $6r \leq R \leq 4^{k^2} 6r$.

Because $M \in \mathbb{M}_V$, there exists a template graph $G \in \mathrm{obt}(M)$ having k vertices $v_1, \ldots, v_k$ at pairwise distance greater than $2r$ such that $\psi_i \in \mathsf{props}^{\mathsf{P}_r}(v_i)$ for all $1 \leq i \leq k$. To conclude that $\mathfrak{A} \models \varphi$, it therefore suffices to find some vector $\vec{a} \in \mathfrak{A}$ such that $G \leq \mathsf{G}_R^{\mathsf{P}_r}(\mathfrak{A}, \vec{a})$. Indeed, this provides a map $h \colon G \to \mathfrak{A}$ such that for all $1 \leq i \leq k$, $\mathfrak{A}, x \mapsto h(v_i) \models \psi_i(x)$, and $\mathrm{d}_{\mathfrak{A}}(h(v_i), h(v_j)) > 2r$ whenever $i \neq j$, i.e., $\mathfrak{A} \models \varphi$.

To find such a representative vector, let us prove a stronger result: for all $G \in M$, there exists $\vec{a} \in \mathfrak{A}$ such that $G \leq \mathsf{G}_R^{\mathsf{P}_r}(\mathfrak{A}, \vec{a})$.

Let us write $M = \{S_1, \ldots, S_n\}$ where each S_i is repeated $m_i \leq k(k-1)$ times. By definition of Θ_R^M, there exists a sequence of points $b_i^j \in A$ with $1 \leq i \leq n, 1 \leq j \leq m_i$ such that

$$\begin{aligned}
\mathfrak{A}, x \mapsto b_i^j &\models \mathsf{Security}_R^{\mathsf{P}_r}(x) \\
\mathfrak{A}, x \mapsto b_i^j &\models \bigwedge_{G \in S_i} \mathsf{Diagram}_G^{\mathsf{P}_r, R}(x) \\
\mathfrak{A}, x \mapsto b_i^j &\models \bigwedge_{G \notin S_i} \neg\mathsf{Diagram}_G^{\mathsf{P}_r, R}(x) \quad .
\end{aligned}$$

Furthermore, for all fixed $1 \leq i \leq n$ the points b_i^j for $1 \leq j \leq m_i$ are at pairwise distance greater than $6R$.

Let us now prove that the points b_i^j are at pairwise distance greater than $2R$ when i varies. Assume by contradiction that there exists i, j and i', j' such that $i \neq i'$ and $d\left(b_i^j, b_{i'}^{j'}\right) \leq 2R$. Because of Lemma 4.2.34, for all $G \in S_i \cup S_{i'}$, $\mathfrak{A}, x \mapsto b_i^j, y \mapsto b_{i'}^{j'} \models \mathsf{Diagram}_G^{\mathsf{P}_r, R}(x) \iff \mathsf{Diagram}_G^{\mathsf{P}_r, R}(y)$. This clearly implies $S_i = S_{i'}$, which is absurd.

Now that we have the guarantee that the elements b_i^j are far enough from each other, let us leverage the definition of $\mathrm{obt}(M)$ and Fact 4.2.32 to find a representative for every template graph if $\mathrm{obt}(M)$.

Formally, we prove by induction on the definition of $\mathrm{obt}(M)$, that for all $G \in \mathrm{obt}(M)$, there exists $\vec{a} \in \mathfrak{A}$ such that $G \leq \mathsf{G}_R^{\mathsf{P}_r}(\mathfrak{A}, \vec{a})$. The base case being trivial and the induction hypothesis holds because the R-neighbourhoods around the points b_i^j do not intersect, hence Fact 4.2.32 applies. ■

Having completed the last step of our translation from existential local sentences to positive Boolean combinations of basic local sentences, let us restate the main theorem of Section 4.2 (A Positive Locality Theorem), namely Theorem 4.2.2 on page 75.

Theorem 4.2.2. *Let $\mathcal{C} \subseteq \mathsf{Struct}(\sigma)$ be a class of structures and $\varphi \in \mathsf{FO}[\sigma]$ be a first-order sentence. Then, the following properties are equivalent:*

1. *φ is equivalent over $\mathcal{C}$ to an existential local sentence*
2. *φ is equivalent over $\mathcal{C}$ to a positive Boolean combination of basic local sentences, which we call a* positive Gaifman normal form.

Proof. This follows from Lemmas 4.2.9, 4.2.20 and 4.2.38 ∎

Discussion. As mentioned at the beginning of the section, the combinatorics involved in proving Theorem 4.2.2 are similar to the existential Gaifman normal form stated in Theorem 4.2.4 and introduced in [55, Theorem 2]. Let us recall that an existential basic local sentence is a sentence of the form $\exists_r^{\geq n} x.\psi(x)$ for some $\psi(x)$ that is both an existential formula and an r-local formula.[17]

[55]: Grohe and Wöhrle (2004), 'An existential locality theorem'

17: Remark that if $\psi(\vec{x}) \in \mathsf{EFO}$, then its localisation $|\psi|_r^{\vec{x}} \in \mathsf{EFO}$ too.

Since an existential sentence is in particular an existential local sentence, one would hope that a connection between Theorem 4.2.2 and Theorem 4.2.4 would exist. However, one cannot deduce the existence of the existential Gaifman normal form from a simple application of Theorem 4.2.2, since Lemma 4.2.38 does not produce existential basic local sentences when given an existential sentence as input. Already, the sentence $\mathsf{Diagram}_R^{\mathsf{P}_r,G}$ introduced in Lemma 4.2.31 contains a universal quantifier, and this is also true of $\mathsf{Security}$ and Θ_R^M respectively introduced in Definitions 4.2.33 and 4.2.35

Conversely, it does not seem that the existence of an existential Gaifman normal form for existential sentences implies Theorem 4.2.2 in any meaningful way[18]. Furthermore, we would like to point out that there are existential basic local sentences that are clearly not expressible using existential sentences[19].

18: That is, apart from noticing that both are true, hence equivalent statements.

19: For instance, the sentence $\exists_2^{\geq 2} x.\top$ stating that two points are at distance greater than 2.

4.3. An Existential Local Preservation Theorem

As shown in Figure 4.1 the two results of Corollary 4.2.1 and Theorem 4.2.2 presented in Section 4.2 (A Positive Locality Theorem) are fitting a nice picture where a semantic property (preservation under local elementary embeddings) coincide with both a syntactic fragment (existential local sentences) and a variant of the Gaifman normal form (positive Gaifman normal form).

Because Theorem 4.2.2 holds on any class $\mathcal{C}$ of structures, and because basic local sentences are quite simple, one might be hoping to prove that Corollary 4.2.1 relativises to at least $\mathsf{Fin}(\sigma)$. As a direct application, one would obtain a new way to prove that preservation theorems for classes of finite structures, using this positive Gaifman normal form as a starting point.

Recall that we write $\mathsf{Local}(\mathfrak{A}, r, k)$ for the collection of r-neighbourhoods of tuples of length at most k $\mathfrak{A}$. Furthermore, the notation is extended to classes of structures, so that $\mathsf{Local}(\mathcal{C}, r, k)$ is the collection of r-neighbourhoods of tuples of length at most k inside some structure $\mathfrak{A} \in \mathcal{C}$.

Before going further into the analysis of Corollary 4.2.1 in the finite, let us describe the semantic properties of existential local sentences. The main idea is that existential local sentences check the presence of a certain r-local neighbourhood inside structures. Because an existential local sentence is written in FO, it cannot check the isomorphism of such local neighbourhoods. Therefore, we will not be interested in the local behaviours $\mathsf{Local}(\mathfrak{A}, r, k)$, but their equivalence classes up to formulas of a certain quantifier rank.

Definition 4.3.1. Let σ be a finite relational signature, $\mathfrak{A}$ be a structure, and $\vec{a}$ be a tuple of elements from $\mathfrak{A}$. Let us write $\mathsf{tp}^q_{\mathfrak{A}}(\vec{a}, r)$ for the *(q, r)-local type* of $\vec{a}$, that is, the set of all formulas of quantifier rank at most q with at most $|\vec{a}|$ free variables that are r-local, and such that $\mathfrak{A}, \vec{x} \mapsto \vec{a} \models \varphi(\vec{x})$.

Note that there are only finitely many possible local types for a given pair $(q, r) \in \mathbb{N}^2$ and a given number of variables $|\vec{a}|$. Therefore, given a structure $\mathfrak{A} \in \mathsf{Struct}(\sigma)$, one can describe the local behaviour of $\mathfrak{A}$ using the following finite set of local-types:

$$\mathsf{Types}^{q,k}_r(\mathfrak{A}) \stackrel{\text{def}}{=} \{ \mathsf{tp}^q_{\mathfrak{A}}(\vec{a}, r) : \vec{a} \in \mathfrak{A}^{\leq k} \} . \tag{4.4}$$

We will now reformulate the notion of r-local elementary embedding, that is the semantic property of existential local sentences in terms of (q, r)-local types as follows.

Definition 4.3.2. *Let $\mathfrak{A}, \mathfrak{B} \in \mathsf{Struct}(\sigma)$, $r, q, k \geq 0$. We say that $\mathfrak{A} \Rrightarrow^{r,k}_q \mathfrak{B}$ if and only if for every $\vec{a} \in \mathfrak{A}$, there exists a tuple $\vec{b} \in \mathfrak{B}$ such that $\vec{a}$ and $\vec{b}$ have the same the (q, r)-local types, i.e., $\mathsf{Types}^{q,k}_r(\mathfrak{A}) \subseteq \mathsf{Types}^{q,k}_r(\mathfrak{B})$.*

We generalise the definition to unbounded parameters r, q, k as follows:

$$\Rrightarrow^{\infty,\infty}_{\infty} \stackrel{\text{def}}{=} \bigcap_{r\geq 0} \bigcap_{q\geq 0} \bigcap_{k\geq 0} \Rrightarrow^{r,k}_q . \tag{4.5}$$

As a sanity check, existential local sentences are preserved under $\Rrightarrow^{\infty,\infty}_{\infty}$, and one can relate the quantifier rank, locality radius, and number of outer existential variables of a given existential local sentence to the least $k, q, r \geq 0$ such that said sentence is preserved under $\Rrightarrow^{r,k}_q$. This is formally stated in Fact 4.3.3.

Fact 4.3.3. Let $\mathcal{C} \subseteq \mathsf{Struct}(\sigma)$. Let $r, q, k \geq 0$, and $\phi \in \mathsf{FO}[\sigma]$. The following are equivalent:

1. ϕ is equivalent (over $\mathcal{C}$) to an existential local sentence $\exists \vec{x}. \psi(x)$ where $|\vec{x}| \leq k$, ψ has quantifier rank at most q, and ψ is r-local;
2. ϕ is preserved (over $\mathcal{C}$) under $\Rrightarrow^{r,k}_q$, i.e., whenever $\mathfrak{A} \models \phi$ and $\mathfrak{A} \Rrightarrow^{r,k}_q \mathfrak{B}$, then $\mathfrak{B} \models \phi$.

To get a better grasp of this family of quasi-orders, let us first note a few facts and play a bit with the definition over carefully selected examples. First, Fact 4.3.4 demonstrates that increasing the parameters unsurprisingly refines the quasi-orders, i.e., that $\Rrightarrow_q^{r,k}$ gets finer as r, q, k increase. Second, we show in Example 4.3.7 that the relation $\mathfrak{A} \Rrightarrow_q^{r,k} \mathfrak{B}$ is indeed strictly coarser than checking the inclusion $\mathsf{Local}(\mathfrak{A}, r, k) \subseteq \mathsf{Local}(\mathfrak{B}, r, k)$. Finally, we illustrate in Fact 4.3.6 the connection between our local quasi-orders and disjoint unions of finite structures.

Fact 4.3.4. Let $\mathcal{C} \subseteq \mathsf{Struct}(\sigma)$, and let $r \leq r'$, $q \leq q'$, and $k \leq k'$ be integers. Then, $\Rrightarrow_{q'}^{r',k'} \subseteq \Rrightarrow_q^{r,k}$ as relations over $\mathcal{C}$.

Proof. Assume that $\mathfrak{A} \Rrightarrow_{q'}^{r',k'} \mathfrak{B}$. Let $\mathsf{tp}_{\mathfrak{A}}^q(\vec{a}, r) \in \mathsf{Types}_r^{q,k}(\mathfrak{A})$. In particular, $\vec{a} \in \mathfrak{A}$ has size $k \leq k'$, hence $\mathsf{tp}_{\mathfrak{A}}^{q'}(\vec{a}, r') \in \mathsf{Types}_{r'}^{q',k'}(\mathfrak{A})$. By assumption, this provides a vector $\vec{b} \in \mathfrak{B}$ of length at most k' such that $\mathsf{tp}_{\mathfrak{A}}^{q'}(\vec{a}, r') = \mathsf{tp}_{\mathfrak{B}}^{q'}\left(\vec{b}, r'\right)$.

Now, it is quite clear that it implies $|\vec{a}| = |\vec{b}| = k$. Furthermore, $\mathsf{tp}_{\mathfrak{A}}^q(\vec{a}, r)$ is completely determined[20] by $\mathsf{tp}_{\mathfrak{A}}^{q'}(\vec{a}, r')$ (and similarly for $\vec{b}$). As a consequence, $\mathsf{tp}_{\mathfrak{A}}^q(\vec{a}, r) = \mathsf{tp}_{\mathfrak{B}}^q\left(\vec{b}, r\right)$. ■

20: as the set of all r-local formulas with quantifier rank at most q in $\mathsf{tp}_{\mathfrak{A}}^{q'}(\vec{a}, r')$.

Example 4.3.5. Let $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$, and $r, q \in \mathbb{N}$. Then, $\Rrightarrow_q^{0,\infty}$, $\Rrightarrow_0^{r,\infty}$, and $\subseteq_i$ are the same relations over $\mathcal{C}$.

Proof. Let us first prove that the locality based ordering imply $\subseteq_i$. Leveraging Lemma 3.2.8, it suffices to prove that if $\mathfrak{A} \Rrightarrow_0^{0,\infty} \mathfrak{B}$, then $\mathsf{Th}(\mathfrak{A}) \cap \mathsf{EFO} \subseteq \mathsf{Th}(\mathfrak{B}) \cap \mathsf{EFO}$.

Let us consider $\varphi \in \mathsf{Th}(\mathfrak{A}) \cap \mathsf{EFO}$. Then, $\varphi = \exists \vec{x}.\psi(\vec{x})$ where ψ is quantifier-free. In particular, there is a tuple $\vec{a} \in \mathfrak{A}$ that witnesses the truth of φ. Because $\mathfrak{A} \Rrightarrow_0^{0,|\vec{a}|} \mathfrak{B}$, there exists $\vec{b} \in \mathfrak{B}$ such that $\mathsf{tp}_{\mathfrak{A}}^0(\vec{a}, 0) = \mathsf{tp}_{\mathfrak{B}}^0\left(\vec{b}, 0\right)$.

In particular, $\psi(\vec{x}) \in \mathsf{tp}_{\mathfrak{B}}^0\left(\vec{b}, 0\right)$, and we conclude that $\mathfrak{B} \models \varphi$.

For the converse inclusion, assume that $\mathfrak{A} \subseteq_i \mathfrak{B}$, and in particular, let us call $h\colon \mathfrak{A} \to \mathfrak{B}$ the associated QF-embedding. Let us prove that $\mathfrak{A} \Rrightarrow_0^{r,\infty} \mathfrak{B}$, and that $\mathfrak{A} \Rrightarrow_q^{0,\infty} \mathfrak{B}$. Let $\mathsf{tp}_{\mathfrak{A}}^q(\vec{a}, 0) \in \mathsf{Types}_0^{q,|\vec{a}|}(\mathfrak{A})$. We claim that $\mathsf{tp}_{\mathfrak{A}}^q(\vec{a}, 0) = \mathsf{tp}_{\mathfrak{B}}^q(h(\vec{a}), 0)$.

This holds because a 0-local formula is equivalent to an existential formula. Similarly, a quantifier-free r-local formula is in fact, just a quantifier-free formula. ■

Fact 4.3.6. Let $\mathfrak{A}, \mathfrak{B} \in \mathsf{Fin}(\sigma)$. The following are equivalent:

1. $\mathfrak{A} \Rrightarrow_\infty^{\infty,\infty} \mathfrak{B}$;
2. For all $r, k \in \mathbb{N}$, $\mathsf{Local}(\mathfrak{A}, r, k) \subseteq \mathsf{Local}(\mathfrak{B}, r, k)$;
3. There exists $\mathfrak{C} \in \mathsf{Fin}(\sigma)$, such that $\mathfrak{A} \uplus \mathfrak{C} = \mathfrak{B}$.

Proof. Let us first prove that Item 3 implies Item 1. Let $\vec{a} \in \mathfrak{A}^k$, then $\vec{a} \in \mathfrak{B} = \mathfrak{A} \uplus \mathfrak{C}$. Furthermore, it is an easy check that $\mathsf{tp}^q_{\mathfrak{A}}(\vec{a}, r) = \mathsf{tp}^q_{\mathfrak{A} \uplus \mathfrak{C}}(\vec{a}, r)$. Indeed,[21] given an r-local formula $\varphi(\vec{x})$ the following are equivalent:

- $\mathfrak{A}, \vec{x} \mapsto \vec{a} \models \varphi(\vec{x})$,
- $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r), \vec{x} \mapsto \vec{a} \models \varphi(\vec{x})$,
- $\mathcal{N}_{\mathfrak{A} \uplus \mathfrak{C}}(\vec{a}, r), \vec{x} \mapsto \vec{a} \models \varphi(\vec{x})$, and
- $\mathfrak{A} \uplus \mathfrak{C}, \vec{x} \mapsto \vec{a} \models \varphi(\vec{x})$.

21: The following computations are only restating formally that a local formula only depends on the local behaviour of a structure, and that the local behaviour of the disjoint union is the disjoint union of the local behaviours.

Now, let us prove that Item 1 implies Item 2. Let $\vec{a} \in \mathfrak{A}^k$. Because $\mathfrak{A}$ is finite, so is $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r)$. Hence, one can construct[22] a first order sentence φ such that $\mathfrak{B} \models \varphi$ if and only if $\mathfrak{B} = \mathcal{N}_{\mathfrak{A}}(\vec{a}, r)$. By definition, $\mathfrak{A}, \vec{x} \mapsto \vec{a} \models |\varphi|^r_{\vec{x}}$. Hence, $|\varphi|^r_{\vec{x}} \in \mathsf{tp}^q_{\mathfrak{A}}(\vec{a}, r)$ for some $q \in \mathbb{N}$. Because $\mathfrak{A} \Rrightarrow^{\infty,\infty}_{\infty} \mathfrak{B}$, there exists a vector $\vec{b} \in \mathfrak{B}^k$ such that $\mathsf{tp}^q_{\mathfrak{A}}(\vec{a}, r) = \mathsf{tp}^q_{\mathfrak{B}}\left(\vec{b}, r\right)$. In particular, $|\varphi|^r_{\vec{x}} \in \mathsf{tp}^q_{\mathfrak{B}}\left(\vec{b}, r\right)$. This proves that $\mathcal{N}_{\mathfrak{B}}(\vec{b}, r) = \mathcal{N}_{\mathfrak{A}}(\vec{a}, r)$, and we have concluded.

22: This is a diagram sentence for the full FO fragment.

Finally, let us prove that Item 2 implies Item 3. Because $\mathfrak{A}$ is finite, $\mathfrak{A} \in \mathsf{Local}(\mathfrak{A}, |\mathfrak{A}|, |\mathfrak{A}| + 1)$. By assumption, there exists a vector $\vec{b} \in \mathfrak{B}$ such that $\mathcal{N}_{\mathfrak{B}}(\vec{b}, |\mathfrak{A}| + 1) = \mathfrak{A}$. Let us conclude by showing that $\mathfrak{B} = \mathcal{N}_{\mathfrak{B}}(\vec{b}, |\mathfrak{A}| + 1) \uplus \left[\mathfrak{B} \setminus \mathcal{N}_{\mathfrak{B}}(\vec{b}, |\mathfrak{A}| + 1)\right]$. Assume by contradiction that there exists $c_1 \in \mathfrak{B} \setminus \mathcal{N}_{\mathfrak{B}}(\vec{b}, |\mathfrak{A}| + 1)$, and $c_2 \in \mathcal{N}_{\mathfrak{B}}(\vec{b}, |\mathfrak{A}| + 1)$, such that (c_1, c_2) are neighbours in $\mathfrak{B}$. Notice that every point of $\mathcal{N}_{\mathfrak{B}}(\vec{b}, |\mathfrak{A}| + 1)$ is actually at distance at most $|\mathfrak{A}|$ from $\vec{b}$ (because this holds for $\mathfrak{A}$ itself). Hence, $c_2 \in \mathcal{N}_{\mathfrak{B}}(\vec{b}, |\mathfrak{A}|)$. This proves that $c_1 \in \mathcal{N}_{\mathfrak{B}}(\vec{b}, |\mathfrak{A}| + 1)$, but this is absurd. ∎

Example 4.3.7. Let $n, m \in \mathbb{N}$. Recall that $P_n \in \mathsf{Paths}$ is a path of size n, and $C_m \in \mathsf{Cycles}$ is a cycle of size m. Assume moreover that $m \leq k(2r+1) < n$ and $k < m$. Then, $\mathsf{Local}(C_m, r, k) \not\subseteq \mathsf{Local}(P_n, r, k)$. However, $C_m \Rrightarrow^{r,k}_{0} P_n$.

Proof Sketch. Let us notice that whenever $m \leq k(2r+1)$, C_m is inside $\mathsf{Local}(C_m, r, k)$. As $\mathsf{Local}(P_n, r, k)$ contains only finite paths, this proves the non-inclusion of local neighbourhoods.

To prove that $C_m \Rrightarrow^{r,k}_{0} P_n$, let us consider a vector $\vec{a}$ of k points in C_m. The local type $\mathsf{tp}^0_{C_m}(\vec{a}, r)$ is the collection of quantifier free formulas with k variables that hold in $\vec{a}$, which essentially describes a finite disjoint unions of paths. Because $k < m \leq k(2r+1) < n$, it is an easy check that $\mathsf{tp}^0_{C_m}(\vec{a}, r) \in \mathsf{Types}^{q,k}_{r}(P_n)$. ∎

Let us summarise all the results about infinite structures that we have obtained so far in the following proposition, which refines Figure 4.1.

Proposition 4.3.8. *Let φ be a sentence in $\mathsf{FO}[\sigma]$. The following properties are equivalent over the class $\mathsf{Struct}(\sigma)$ of all structures.*

1. *φ is equivalent to a positive Boolean combination of basic local sentences.*
2. *φ is equivalent to an existential local sentence.*

3. *There exists $r, q, k \in \mathbb{N}$ such that φ is preserved under $\Rrightarrow_q^{r,k}$.*
4. *φ is preserved under $\Rrightarrow_\infty^{\infty,\infty}$.*
5. *φ is preserved under local elementary embeddings.*

Proof. The equivalence between Item 1 and Item 2 is exactly Theorem 4.2.2. The equivalence between Items 2 and 5 follows from Corollary 4.2.1. The equivalence between Items 2 and 3 follows from Fact 4.3.3. Finally, remark that Item 3 implies Item 4, which itself implies Item 5, hence all of the above are equivalent to Item 2. ■

The connection between Proposition 4.3.8 and preservation theorems[23] might seem tenuous. Let us explain why the two are deeply connected in the case of the Łoś-Tarski Theorem. As proven by Tait, the Łoś-Tarski Theorem does not relativise to $\mathsf{Fin}(\sigma)$. However, a combinatorial proof of Atserias, Dawar and Grohe shows that for hereditary classes of finite structures $\mathcal{C}$ that have bounded degree, and that are closed under disjoint unions, the Łoś-Tarski theorem relativises to $\mathcal{C}$. As per our Example 4.1.16, the proof uses the Gaifman Locality Theorem, and one of the main obstacles to overcome is the presence of negations in this normal form. If Proposition 4.3.8 were to relativise in the finite, this would greatly simplify the proof of this result, and potentially allows us to weaken the hypotheses.

23: See Theorem 3.1.9 p. 33

First, the fine semantic characterisations of the existential local sentences and positive Gaifman normal forms remain true in the finite, as depicted in Proposition 4.3.9 hereafter. Moreover, the coarse semantic characterisations of existential local sentences are intimately related to disjoint unions, as shown right after in Proposition 4.3.10.

Proposition 4.3.9. *Let φ be a sentence in $\mathsf{FO}[\sigma]$, and $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$. The following properties are equivalent over $\mathcal{C}$:*

1. *φ is equivalent to a positive Boolean combination of basic local sentences.*
2. *φ is equivalent to an existential local sentence.*
3. *There exists $r, q, k \in \mathbb{N}$ such that φ is preserved under $\Rrightarrow_q^{r,k}$.*

Proof. The equivalence between the first two items follows from Theorem 4.2.2, while the equivalence between the last two items follows from Fact 4.3.3. ■

Proposition 4.3.10. *Let φ be a sentence in $\mathsf{FO}[\sigma]$, and $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$. The following properties are equivalent over $\mathcal{C}$:*

1. *φ is preserved under local elementary embeddings;*
2. *φ is preserved under $\Rrightarrow_\infty^{\infty,\infty}$.*
3. *φ is preserved under disjoint unions, i.e., if $\mathfrak{A} \models \varphi$ and $\mathfrak{A} \uplus \mathfrak{C} \in \mathcal{C}$, then $\mathfrak{A} \uplus \mathfrak{C} \models \varphi$;*

Proof. The equivalence between the last two items is simply Fact 4.3.6. The equivalence between the first two items follows from Lemma 3.2.8 and the fact that if $\mathfrak{A} \Rrightarrow_\infty^{\infty,\infty} \mathfrak{B}$, then $\mathsf{Th}(\mathfrak{A}) \cap \mathsf{EFO} \subseteq \mathsf{Th}(\mathfrak{B}) \cap \mathsf{EFO}$. ■

However, Proposition 4.3.8 does not relativise to finite structures: we show in Subsection 4.3.1 that Items 2 and 5 of Proposition 4.3.8 are not equivalent in the finite.

[6]: Atserias, Dawar and Grohe (2008), 'Preservation under extensions on well-behaved finite structures'

This seemingly puts an end to the attempted generalisation of [6], and naturally leads to a decision problem: given a sentence φ preserved under disjoint unions over $\mathsf{Fin}(\sigma)$, can it be rewritten as an existential local sentence? We prove in Subsection 4.3.2 that this problem, and two other associated decision problems, are undecidable. This follows the line of negative results on preservation theorems in the finite [16, 66].

In order to understand why the proof of [6] works, while the FOLoc-preservation theorem does not relativise, we then carry out a fine-grained analysis of the local quasi-orders $\Rrightarrow_q^{r,k}$. This analysis is performed in Subsection 4.3.3 and classifies the tuple of parameters (r, q, k) such that first-order sentences preserved under $\Rrightarrow_q^{r,k}$ over $\mathsf{Fin}(\sigma)$ are equivalent to an existential local sentence. It turns out that perseverance is a great quality, as this section conclude with an open question: what happens for $\Rrightarrow_0^{0,\infty}$? A question interesting enough that Chapter 5 (A Local-to-Global Preservation Theorem) is devoted to answering it.

4.3.1. Failure in the finite

In this subsection, our goal is to prove that Items 2 and 5 of Proposition 4.3.8 are not equivalent in the finite. This is what will be shown in Theorem 4.3.21 on page 106.

The main idea will be to use Fact 4.3.6, which characterises preservation under local elementary embeddings as preservation under disjoint unions in the finite.

To that end, let us define $\mathsf{2CC}$ as the set of finite structures whose Gaifman graph has at least two connected components (i.e., is disconnected).

The proof scheme will be the following:

1. First, provide a generic non-relativisation proof over classes $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$, under the assumption that $\mathsf{2CC}$ is first-order definable over the class $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$.[24] This is Lemma 4.3.12 on page 100.
2. Then, show that one can provide a finite list of universal local sentences that axiomatise a class $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$, where $\mathsf{2CC}$ is easily seen to be first-order definable. These are Lemmas 4.3.16, 4.3.18 and 4.3.19 on pages 101–103.
3. Finally, lift this non-equivalence to $\mathsf{Fin}(\sigma)$. This is handled in Lemma 4.3.20 on page 105

24: A set S is first order definable over a class $\mathcal{C}$ when $S \subseteq \mathcal{C}$, and there exists a sentence $\varphi \in \mathsf{FO}[\sigma]$ such that $[\![\varphi]\!]_{\mathcal{C}} = S$, i.e., $[\![\varphi]\!] \cap \mathcal{C} = S \cap \mathcal{C} = S$.

A Generic Counter Example. Let $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ where $\mathsf{2CC}$ is first-order definable, and assume that a unary predicate B is available in σ. It is then possible to define

$$\phi_{\mathrm{bad}} \stackrel{\text{def}}{=} [\forall x. \neg B(x)] \vee \mathsf{2CC} \quad .$$

By definition, ϕ_{bad} is preserved under disjoint unions over $\mathcal{C}$.

What remains to be proven is that whenever the class $\mathcal{C}$ contains large enough structures, ϕ_{bad} cannot be expressed using an existential local sentence. The intuition behind this statement is that such an existential local sentence will not distinguish between a large connected component with one B node (not satisfying ϕ_{bad}) and two connected components with one B node (satisfying ϕ_{bad}). A prototypical example built using finite path is presented in Example 4.3.11 to understand the proof in a very simple setting.

In order to build examples and non-examples, we will have to produce classes of finite structures that are hereditary, but also closed under disjoint unions. For the former, one can always consider the downward closure $\downarrow_{\leq}\mathcal{C}$, but we lack a similar closure operation for disjoint unions. Let us therefore introduce $\uplus\mathcal{C}$ for the closure of $\mathcal{C}$ under disjoint unions inside $\mathsf{Fin}(\sigma)$. That is, $\uplus\mathcal{C}$ is the collection of finite disjoint unions of structures in $\mathcal{C}$.

Recall that we introduced the notion of labelling of structures in Subsection 3.3.1. The notation $\mathsf{Lab}(L, \mathcal{C})$ stands for the structures of $\mathcal{C}$ freely coloured using elements of L as unary predicates.

Example 4.3.11. Let $\mathsf{Lab}(\{B\}, \uplus\mathsf{Paths})$ be the class of disjoint unions of finite $\{B, \neg B\}$-labelled paths. Then, $\mathsf{2CC}$ is first-order definable, ϕ_{bad} is preserved under disjoint unions, and ϕ_{bad} is not expressible as an existential local sentence over $\mathsf{Lab}(\{B\}, \uplus\mathsf{Paths})$.

Proof. One can detect the presence of two connected components using the fact that paths have at most two vertices of degree below 2, through the disjunction of the following sentences:

- Two distinct paths:

$$\exists x_1, x_2, x_3, x_4. \bigwedge_{1\leq i\neq j\leq 4} x_i \neq x_j \wedge \bigwedge_{i=1}^{4} (\deg(x_i) = 1) \quad ,$$

- Two isolated vertices:

$$\exists x_1, x_2.(x_1 \neq x_2 \wedge \deg(x_1) = \deg(x_2) = 0) \quad ,$$

- One path and one isolated vertex:

$$\begin{aligned} &\exists x_1, x_2, x_3. \\ &x_1 \neq x_2 \wedge x_2 \neq x_3 \wedge x_3 \neq x_1 \wedge \\ &\deg(x_1) = \deg(x_2) = 1 \wedge \deg(x_3) = 0 \quad . \end{aligned}$$

Assume by contradiction that there exists a sentence $\psi = \exists\vec{x}.\theta(x)$ where $\theta(x)$ is an r-local sentence of quantifier rank q such that ϕ_{bad} is equivalent to ψ over $\mathsf{Lab}(\{B\}, \uplus\mathsf{Paths})$. Consider the family of structures $P_k^{\neg B}$ that are paths of length k coloured by $\neg B$. It is clear that $P_k^{\neg B} \models \phi_{\mathrm{bad}}$ for all $k \geq 1$. Consider $k > |\vec{x}| \cdot (2r+1)$. Since $P_k^{\neg B} \models \psi$, there exists a valuation $\upsilon\colon \vec{x} \to P_k^{\neg B}$, such that $\mathcal{N}_{P_k^{\neg B}}(\upsilon(\vec{x}), r), \upsilon \models \theta(x)$ and $\mathcal{N}_{P_k^{\neg B}}(\upsilon(\vec{a}), r) \subsetneq P_k^{\neg B}$. Consider a point $b \in P_k^{\neg B}$ that is not in $\mathcal{N}_{P_k^{\neg B}}(\upsilon(\vec{x}), r)$, and build P_k^B as a copy of $P_k^{\neg B}$, except that b is coloured by B instead of $\neg B$. The structure P_k^B belongs to $\mathsf{Lab}(\{B\}, \uplus\mathsf{Paths})$, and does not satisfy ϕ_{bad}. However, $P_k^B, \upsilon \models \theta(\vec{x})$, thus $P_k^B \models \psi$, which is absurd. ■

It is not hard to generalise the proof of Example 4.3.11 to other classes of structures, provided that they contain some structures "large enough" to not be covered by a finite number of local neighbourhoods.

Lemma 4.3.12. *Let σ be a finite relational signature, $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ be a class of finite structures. Assume moreover that 2CC is first-order definable over $\mathcal{C}$, and that for all $r, k \in \mathbb{N}$, there exists a structure $\mathfrak{A} \in \mathcal{C}$ such that for all $\vec{a} \in \mathfrak{A}^{\leq k}$, $\mathfrak{A} \setminus \mathcal{N}_{\mathfrak{A}}(\vec{a}, r) \neq \emptyset$, Then,*

(i) ϕ_{bad} *is first-order definable over* $\mathsf{Lab}(\{B\}, \mathcal{C})$*,*
(ii) ϕ_{bad} *is preserved under disjoint unions over* $\mathsf{Lab}(\{B\}, \mathcal{C})$*,*
(iii) and ϕ_{bad} *is not equivalent to any existential local sentence over* $\mathsf{Lab}(\{B\}, \mathcal{C})$*.*

Proof. The first two items follow from the definition of ϕ_{bad}.

For the third point, assume by contradiction that there exists a sentence $\psi = \exists \vec{x}.\theta(x)$ where $\theta(x)$ is an r-local sentence such that ϕ_{bad} is equivalent to ψ over $\mathsf{Lab}(\{B\}, \mathcal{C})$.

Consider $\mathfrak{A} \in \mathcal{C}$ such that for all $\vec{a} \in \mathfrak{A}^{\leq k}$, $\mathfrak{A} \setminus \mathcal{N}_{\mathfrak{A}}(\vec{a}, r) \neq \emptyset$. Let us write $\mathfrak{A}^{\neg B}$ to be the model $\mathfrak{A}$ where all elements are coloured with $\neg B$. By construction, $\mathfrak{A}^{\neg B}$ satisfies ϕ_{bad}. Since $\mathfrak{A}^{\neg B} \models \psi$, there exists a valuation $\upsilon \colon \vec{x} \to \mathfrak{A}^{\neg B}$, such that $\mathcal{N}_{\mathfrak{A}^{\neg B}}(\upsilon(\vec{x}), r), \upsilon \models \theta(x)$.

By construction, there exists a point $b \in \mathfrak{A}^{\neg B} \setminus \mathcal{N}_{\mathfrak{A}^{\neg b}}(\upsilon(\vec{x}), r)$. Defining $\mathfrak{A}^{B}$ to be the copy of $\mathfrak{A}$ where all points except b are coloured with $\neg B$, we obtain a structure that satisfies ψ by construction, but does not satisfy ϕ_{bad}, which is absurd. ■

Axiomatising A Bad Subclass. Note that the class $\uplus$Paths disjoint unions of finite paths has no finite axiomatisation in $\mathsf{Fin}(\sigma)$. Thus, Example 4.3.11 cannot be taken as a class to follow Item 2 of the proof scheme outlined on page 98.

As a workaround, we find some inspiration in counter example provided by [90] in the case of Łoś-Tarski's preservation theorem, and leverage the idea given by the class of finite paths.[25] Before providing a concrete axiomatisation, let us semantically define the class of interest. The basic building bloc generalises paths by considering finite total orderings in Definition 4.3.13. This total ordering will allow us to define complex sentences, but all nodes are at distance at most one. To avoid this problem, we introduce in Definition 4.3.14 the "glueing" of such total orders, so that the new structures exhibit complex neighbourhoods.

[90]: Tait (1959), 'A counterexample to a conjecture of Scott and Suppes'

25: Recall that the counter example of Tait was restated in Theorem 3.2.6 on page 39.

Definition 4.3.13. *Let $\sigma \stackrel{\text{def}}{=} \{(\leq, 2), (S, 2), (E, 2)\}$. Define O_n to be the structure with domain $\{1, \ldots, n\}$, S interpreted as the successor relation, $\leq$ as the usual ordering of natural numbers, and E the empty relation.*

Definition 4.3.14. *Given a strictly increasing sequence of natural numbers $2 \leq n_1 < \cdots < n_k$ (with $k \geq 1$) one can build a structure denoted by $\langle O_{n_1} + \cdots + O_{n_k} \rangle$ by extending the disjoint union $\biguplus_{1 \leq i \leq k} O_{n_i}$ with new relations. We add the relations $S(a,b)$ whenever a is the last element of O_{n_i} and b the first of $O_{n_{i+1}}$, and $E(a,b)$ whenever $a \in O_{n_i}$, $b \in O_{n_{i+1}}$ and b is below a when interpreted as integers.*

A graphical representation of O_4 is given in Figure 4.10, and similarly a representation of $\langle O_2 + \cdots + O_5 \rangle$ is given in Figure 4.11.

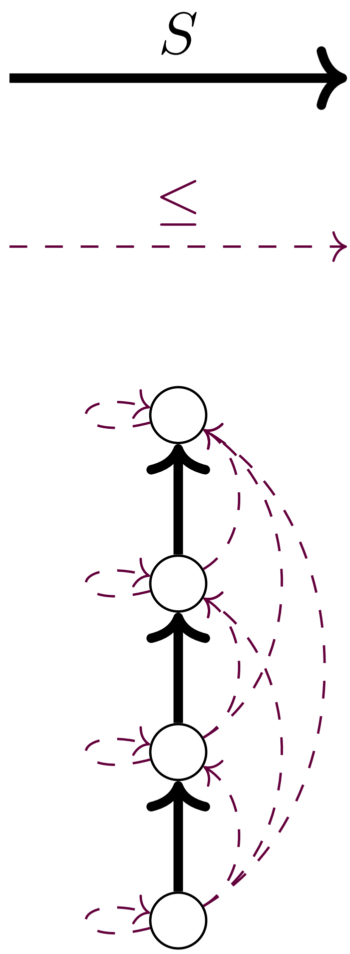


Figure 4.10.: The structure O_4.

Definition 4.3.15. *We define then $\mathcal{C}_{\mathrm{ord}}$ to be the class of finite disjoint unions of structures of the form $\langle O_m + \cdots + O_n \rangle$, with $2 \leq m \leq n$.*

Because $\mathcal{C}_{\mathrm{ord}}$ essentially behaves as a union of paths, it is quite easy to define $\mathsf{2CC}$, following the same intuitions as for unions of finite paths.

Lemma 4.3.16. *The property $\mathsf{2CC}$ is first-order definable in $\mathcal{C}_{\mathrm{ord}}$.*

Proof. Define $\mathsf{2CC} \overset{\mathsf{def}}{=} \exists x_1, x_2. x_1 \neq x_2 \wedge \forall y. \neg S(y, x_1) \wedge \neg S(y, x_2)$. Remark that this definition is not only in FO, but also existential local, preserved under $\Rrightarrow_1^{\infty,2}$, and preserved under $\Rrightarrow_{\infty}^{1,2}$. This is not important right now, but will be crucial in Subsection 4.3.3 on page 109. ■

Corollary 4.3.17. *There exists a first order sentence that is preserved under disjoint unions over $\mathcal{C}_{\mathrm{ord}}$ but not equivalent to any existential local sentence over $\mathcal{C}_{\mathrm{ord}}$.*

Proof. It suffices to apply Lemma 4.3.12 to conclude. For that, let $k, r \in \mathbb{N}$. Define $r' \overset{\mathsf{def}}{=} 2k \cdot (2r+1) + 2$. Then, the structure $\langle O_2 +$

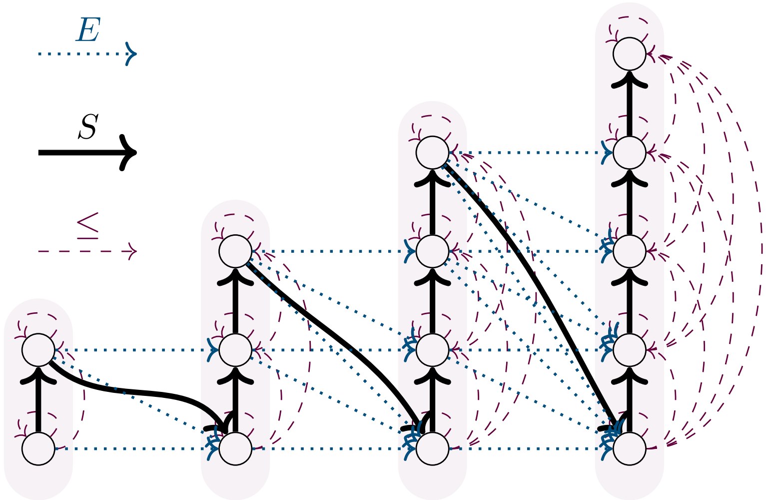


Figure 4.11.: The structure $\langle O_2 + \cdots + O_5 \rangle$.

$\cdots + O_{r'}\rangle$ satisfies that for every $\vec{a}$ of size at most k, $\langle O_2 + \cdots + O_{r'}\rangle \setminus \mathcal{N}_{\langle O_2+\cdots+O_{r'}\rangle}(\vec{a}, r) \neq \emptyset$. ■

Let us now propose an axiomatisation of $\mathcal{C}_{\mathrm{ord}}$. To simplify notations, let us consider for a structure $\mathfrak{A} \in \mathsf{Fin}(\sigma)$ the structure $(\mathfrak{A}, \leq)$ to be the structure $\mathfrak{A}$ without the relations S and E. Using this convention, a *$\leq$-component of a structure* $\mathfrak{A}$ is defined as a connected component of the Gaifman Graph $\mathrm{Gaif}(\mathfrak{A}, \leq)$.

We introduce in Figure 4.12 a list $\mathcal{A}_{\mathrm{ord}}$ of axioms that characterises $\mathcal{C}_{\mathrm{ord}}$. The goal of the rest of this section is to prove that $[\![\mathcal{A}_{\mathrm{ord}}]\!]_{\mathsf{Fin}(\sigma)} = \mathcal{C}_{\mathrm{ord}}$.

We start by noticing that $\leq$-components are total orderings, i.e., that $\leq$ is transitive, reflexive and antisymmetric (Axioms 1 to 3). Moreover, the $\leq$-components have size at least 2 (Axiom 4). Furthermore, S is an injective partial function without fixed points (Axioms 5 to 7). Also, S and $\leq$ cannot conflict (Axiom 8). There exists a proto-induction principle via (Axiom 9).

In order to characterise how $\leq$-components are connected via S and E, we notice the following properties. First, edges E can be factorised inside $(\leq)(S \setminus \leq)(\leq)$ (Axiom 10). Pre-images through E form a suffix of the ordering defined by S inside a $\leq$-component (Axiom 11). Images through E form a prefix of the S ordering inside a $\leq$-component (Axiom 12). Images through E are strictly increasing subsets (Axiom 13). Pre-images through E are strictly decreasing (Axiom 14). The last element for S of a $\leq$-component cannot be obtained as an image through E (Axiom 15). The relation $(\leq)(S \setminus \leq)$ is included in E (Axiom 16).

The following Lemma 4.3.18 is a simple check that we only witnessed true axioms and is quite boring. We propose that its proof is skipped and that the reader moves directly to the more interesting converse inclusion in Lemma 4.3.19 on the facing page.

Lemma 4.3.18. *The following inclusion holds:* $\mathcal{C}_{\mathrm{ord}} \subseteq [\![\mathcal{A}_{\mathrm{ord}}]\!]_{\mathsf{Fin}(\sigma)}$.

Proof. Before listing the axioms in $\mathcal{A}_{\mathrm{ord}}$, notice that it suffices to prove that connected structures in $\mathcal{C}_{\mathrm{ord}}$ satisfies $\mathcal{A}_{\mathrm{ord}}$ as they are closed under disjoint unions of models. This is because the universal and existential quantifications in every axiom are guarded by the presence of a relation.

Let $\mathfrak{A}$ be a connected structure in $\mathcal{C}_{\mathrm{ord}}$, that is $\mathfrak{A} = \langle O_{n_1} + \cdots + O_{n_k}\rangle$ with $2 \leq n_1 < \cdots < n_k$.

1. Axioms 1 to 3 hold over each O_i, hence they remain true over their disjoint unions.
2. Axiom 4 holds because for all $1 \leq i \leq k$, $2 \leq n_k$.
3. Axioms 5 to 7 hold for every O_{n_i}. Furthermore, the maximal element of O_{n_i} is connected through S to the minimal element of $O_{n_{i+1}}$. Hence, the properties are true for $\mathfrak{A}$.
4. Axioms 8 and 9 hold because S is the successor relation over each O_{n_i}.

$$\text{Ax. 1}\ \frac{a \leq b \qquad b \leq c}{a \leq c} \qquad \text{Ax. 2}\ \frac{}{a \leq a} \qquad \text{Ax. 3}\ \frac{a \leq b \qquad b \leq a}{a = b}$$

$$\text{Ax. 4}\ \frac{}{\exists b.(a \neq b) \wedge (a \leq b \vee b \leq a)} \qquad \text{Ax. 5}\ \frac{S(a,b) \qquad S(a,c)}{b = c}$$

$$\text{Ax. 6}\ \frac{S(a,c) \qquad S(b,c)}{a = b} \qquad \text{Ax. 7}\ \frac{}{\neg S(a,a)} \qquad \text{Ax. 8}\ \frac{}{\neg(S(a,b) \wedge b \leq a)}$$

$$\text{Ax. 9}\ \frac{a \neq b \qquad a \leq b}{\exists c.S(a,c) \wedge a \leq c \wedge c \leq b}$$

$$\text{Ax. 10}\ \frac{E(a,b)}{\exists c_1, c_2.a \leq c_1 \wedge S(c_1,c_2) \wedge \neg(c_1 \leq c_2) \wedge c_2 \leq b}$$

$$\text{Ax. 11}\ \frac{a \leq b \qquad E(a,c) \qquad S(a,b)}{E(b,c)} \qquad \text{Ax. 12}\ \frac{b \leq c \qquad E(a,c) \qquad S(b,c)}{E(a,b)}$$

$$\text{Ax. 13}\ \frac{S(a,b) \qquad a \leq b \qquad \exists c, E(b,c)}{\exists c.E(b,c) \wedge \neg E(a,b)} \qquad \text{Ax. 14}\ \frac{S(a,b) \qquad a \leq b \qquad \exists c.E(c,a)}{\exists c.E(c,a) \wedge \neg E(c,b)}$$

$$\text{Ax. 15}\ \frac{E(a,b)}{\exists c.S(b,c) \wedge b \leq c} \qquad \text{Ax. 16}\ \frac{}{\exists c.a \leq c \wedge S(c,b) \wedge \neg(c \leq b) \Rightarrow E(a,b)}$$

Figure 4.12.: The axioms of $\mathcal{A}_{\text{ord}}$. For the axioms to be clearer, they are written as inference rules, and outer universal quantifications are omitted. For instance, Axiom 2 is representing the universal local sentence $\forall a.a \leq a$. We grouped the axioms to be easily understandable. For instance, Axioms 1 to 3 state that $\leq$ is a transitive, antisymmetric, reflexive relation, while Axioms 5 to 7 state that S is an injective function without fixed points. However, this construction implies that some axioms are redundant: Axiom 7 is a consequence of Axiom 2 and Axiom 8.

5. Axiom 10 holds because edges E between O_{n_i} and O_{n_j} only appear if $j = i + 1$ and the maximal element of O_{n_i} for $\leq$ is connected through S to the minimal element of O_{n_j} for $\leq$.
6. Axioms 11 to 15 hold because edges $E(a,b)$ between O_{n_i} and O_{n_j} exists if and only if $a \leq b$ when considered as integers.
7. Axiom 16 holds because edges $E(a,b)$ between O_{n_i} and O_{n_j} exists if and only if $a \leq b$ when considered as integers and O_{n_i} is of size n_i. ■

Now that we have checked that our axiomatisation $\mathcal{A}_{\text{ord}}$ is sound, let us prove that it is complete, that is, the converse inclusion holds.

We invite the reader to consider the figure Figure 4.11 to better follow the line of reasoning that will be taken.

Lemma 4.3.19. *Let $\mathfrak{A} \in [\![\mathcal{A}_{\text{ord}}]\!]_{\mathsf{Fin}(\sigma)}$. The following properties hold:*

1. *If $\mathfrak{B}$ is a $\leq$-component of $\mathfrak{A}$ then the substructure induced by $\mathfrak{B}$ in $\mathfrak{A}$ is isomorphic to a total ordering of size greater than two with no E relations, i.e., to some O_m with $m \geq 2$;*
2. *If $\mathfrak{B}_1$ and $\mathfrak{B}_2$ are two $\leq$-components of $\mathfrak{A}$ that are connected in $\mathfrak{A}$ with the relation S, either the last element of $\mathfrak{B}_1$ is connected to the first one of $\mathfrak{B}_2$ or the last element of $\mathfrak{B}_2$ is connected to the first one of $\mathfrak{B}_1$;*
3. *If $\mathfrak{B}_1$ and $\mathfrak{B}_2$ are two $\leq$-components of $\mathfrak{A}$ connected through the relation E, then $\mathfrak{B}_1$ and $\mathfrak{B}_2$ are connected through the relation S;*

4. *If $\mathfrak{B}_1$ and $\mathfrak{B}_2$ are two $\leq$-components of $\mathfrak{A}$ connected through the relation S, the function $f : a \mapsto \max_{\leq} \{ d : E(a, d) \}$ is a $\leq$-strictly increasing non-surjective function from $\mathfrak{B}_1$ to $\mathfrak{B}_2$, mapping the $\leq$-minimal element of $\mathfrak{B}_1$ to the $\leq$-minimal of $\mathfrak{B}_2$ satisfying $f(S(a)) = S(f(a))$.*
5. *The connected components of $\mathrm{Gaif}(\mathfrak{A})$ are in $\mathcal{C}_{\mathrm{ord}}$.*

In particular, $\mathfrak{A} \in \mathcal{C}_{\mathrm{ord}}$.

Proof. Let $\mathfrak{A}$ be a structure in $[\![\mathcal{A}_{\mathrm{ord}}]\!]_{\mathsf{Fin}(\sigma)}$, without loss of generality, assume that the Gaifman graph of $\mathfrak{A}$ has a single connected component.

Let us first prove Item 1. Leveraging the "proto-induction axiom" Axiom 9, one can use the fact that S is a partial injective function without fixed points (Axioms 5 to 7) to prove that whenever $a \leq b$ in $\mathfrak{A}$, there exists $k \in \mathbb{N}$ such that $S^k(a) = b$. Moreover, Axiom 8 implies that for all $0 \leq i \leq k$, $a \leq S^i(a) \leq b$.

As a consequence, if $a \leq b_1$ and $a \leq b_2$, there exists $0 \leq k$ and $0 \leq l$ such that $S^k(a) = b_1$ and $S^l(a) = b_2$. Without loss of generality $k \leq l$ and $b_2 = S^{k-l}(b_1)$ and $b_1 \leq b_2$. Similarly, if $a_1 \leq b$ and $a_2 \leq b$ there exists $0 \leq k, l$ such that $S^k(a_1) = b$ and $S^l(a_2) = b$

Combined with Axioms 1 to 3, this proves that $\leq$-components of $\mathfrak{A}$ are totally ordered for $\leq$, and the partial function S restricted to these $\leq$-components is the successor relation for $\leq$. Furthermore, Axiom 4 proves that $\leq$-components have size at least 2.

Finally, assume by contradiction that some $\leq$-component of $\mathfrak{A}$ contains a relation $E(a, b)$. Notice that Axiom 10 provides c_1, c_2 such that $a \leq c_1$, $S(c_1, c_2)$, $c_2 \leq b$, and $\neg(c_1 \leq c_1)$. This is absurd since the $\leq$-component is a total ordering, where $S(c_1, c_2)$ and $c_2 \leq c_1$ are in conflict.

Let us now prove that Item 2 holds. Assume $\mathfrak{B}_1$ and $\mathfrak{B}_2$ are two $\leq$-components of $\mathfrak{A}$ connected through the relation S. Because S is a partial injective function (Axioms 5 to 7) and is included in $\leq$ inside each component, this can only happen by connecting an element that has no $\leq$-successor to an element that has no $\leq$-predecessor. As a consequence, it is only possible to connect the last element of $\mathfrak{B}_1$ to the first one of $\mathfrak{B}_2$ or vice-versa.

We are now ready to prove Item 3. Assume that $\mathfrak{B}_1$ and $\mathfrak{B}_2$ are two $\leq$-components of $\mathfrak{A}$ connected through the relation E. By definition, there exists $a \in B_1$ and $b \in B_2$ such that $E(a, b)$ holds. Axiom 10 provides c_1 and c_2 such that $a \leq c_1$, $S(c_1, c_2)$, $c_2 \leq b$ and $\neg(c_1 \leq c_2)$. As a consequence c_1 is in $\mathfrak{B}_1$ (connected to a through $\leq$), $c_2 \in \mathfrak{B}_2$ (connected to b through $\leq$) and therefore $\mathfrak{B}_1$ is connected to $\mathfrak{B}_2$ through the relation S.

Let us continue and prove Item 4. Assume $\mathfrak{B}_1$ and $\mathfrak{B}_2$ are two $\leq$-components of $\mathfrak{A}$ connected through the relation S. The function $g : a \mapsto \{ b : E(a, b) \}$ from $\mathfrak{B}_1$ to $\mathcal{P}_{\mathsf{fin}}(\mathfrak{B}_2)$ is well-defined since we proved that there cannot be edges E outside of $\mathfrak{B}_2$. Remark that $\mathfrak{B}_2$ is a finite total ordering with respect to $\leq$ and the image of g is non-empty thanks to Axiom 16. Therefore, the function f is well-defined.

Whenever $a, b, \in \mathfrak{B}_1$ and $S(a,b)$ holds, then $g(a) \subseteq g(b)$ by Axiom 11 and thus $g(a) \subsetneq g(b)$ by Axiom 13. Furthermore, Axiom 12 that $g(a)$ is $\leq$-downwards closed in $\mathfrak{B}_2$. As a consequence, f must be strictly increasing.

Similarly, by Axioms 11 and 14, if $S(a,b)$ in $\mathfrak{B}_1$, then $|g(a)| + 1 = |g(b)|$ and as a consequence $S(f(a)) = f(b)$. Indeed, $g(a) \subsetneq g(b)$; by contradiction, assume that there exists $c \neq d$ that are both in $g(b) \setminus g(a)$. Since $g(a)$ and $g(b)$ are downwards closed, we can assume without loss of generality that $S(c,d)$ holds. Then Axiom 14 tells us that there exists e such that $E(e,c)$ holds but not $E(e,d)$. Because $\neg(b \leq e)$, and $\neg(e \leq a)$, this contradicts the assumption that b is the successor of a.

Finally, property Axiom 15 states that g does not cover some last element of $\mathfrak{B}_2$, which must be the $\leq$-max element by Axiom 12. In particular, f is not surjective.

Let us now conclude by showing Item 5. Recall that relations E can only appear between two $\leq$-components $\mathfrak{B}_1$ and $\mathfrak{B}_2$. We have proven that $\mathfrak{B}_1, \mathfrak{B}_2$ are respectively isomorphic to O_{m_1} and O_{m_2} for some $m_1, m_2 \geq 2$. Let us write 0_1 and 0_2 their respective $\leq$-minimal elements. We showed that $f(S^k(0_1)) = S^k(0_2)$ whenever $S^k(0_1)$ is in $\mathfrak{B}_1$. Moreover, if $a = S^k(0_1) \in \mathfrak{B}_1$ then $\{ d \colon E(a,d) \} = \downarrow_{\leq} f(a) = \{ S^l(0_2) \colon l \leq k \}$. As a consequence, $|\mathfrak{B}_2| = m_2 > |\mathfrak{B}_1| = m_1$.

We have proven that a connected component of $\mathfrak{A}$ is of the form $\langle O_{m_1} + \cdots + O_{m_k} \rangle$, where k is the number of $\leq$-components of $\mathfrak{A}$, and $m_i \geq 2$ for all $1 \leq i \leq k$. Therefore, $\mathfrak{A} \in \mathcal{C}_{\mathrm{ord}}$. ■

Lifting Axiomatised Subclasses. Assume that $\mathcal{C}_{\mathrm{ord}}$ is a class of finite structures defined over $\mathsf{Fin}(\sigma)$ through a finite axiomatisation $\mathcal{A}_{\mathrm{ord}}$ using universal local sentences[26]. The following fact allows us to lift arguments over $\mathcal{C}_{\mathrm{ord}}$ to $\mathsf{Fin}(\sigma)$.

Lemma 4.3.20 (Relativisation to $\mathcal{C}_{\mathrm{ord}}$). *Let $\varphi \in \mathsf{FO}$. The sentence φ is equivalent over $[\![\mathcal{A}_{\mathrm{ord}}]\!]_{\mathsf{Fin}(\sigma)}$ to an existential local sentence if and only if $\mathcal{A}_{\mathrm{ord}} \Rightarrow \varphi$ is equivalent to an existential local sentence over $\mathsf{Fin}(\sigma)$.*

Similarly, the sentence φ is preserved under disjoint unions over $[\![\mathcal{A}_{\mathrm{ord}}]\!]_{\mathsf{Fin}(\sigma)}$ if and only if $\mathcal{A}_{\mathrm{ord}} \Rightarrow \varphi$ is preserved under disjoint unions over $\mathsf{Fin}(\sigma)$.

Proof. It suffices to notice that $\neg(\mathcal{A}_{\mathrm{ord}})$ is an existential local sentence, hence is preserved under disjoint unions.[27]

As a consequence, if φ is equivalent over $[\![\mathcal{A}_{\mathrm{ord}}]\!]_{\mathsf{Fin}(\sigma)}$ to an existential local sentence ψ, then $\varphi \vee \neg\mathcal{A}_{\mathrm{ord}}$ is equivalent over $\mathsf{Fin}(\sigma)$ to $\psi \vee \neg\mathcal{A}_{\mathrm{ord}}$ which is existential local. Conversely, if $\varphi \vee \neg\mathcal{A}_{\mathrm{ord}}$ is equivalent to an existential local sentence ψ, then φ is equivalent to ψ over $[\![\mathcal{A}_{\mathrm{ord}}]\!]_{\mathsf{Fin}(\sigma)}$.

The same arguments work for preservation under disjoint unions. ■

26: Recall that a universal local sentence is the negation of an existential local sentence

Lemma 4.3.20 follows a more general pattern that will be described in Chapter 6 (Logically Presented Spaces), where the relationship between classes satisfying preservation theorems is investigated in depth. In particular, one can generalise the lemma to various fragmets F of FO and quasi-orderings $\leq$ over $\mathsf{Struct}(\sigma)$.

27: Recall that on finite structures, being preserved under disjoint unions amounts to being preserved under local elementary embeddings, which is always true for an existential local sentence (see Propositions 4.3.8 and 4.3.10).

Through Lemmas 4.3.18 and 4.3.19 we learn that $\mathcal{C}_{\mathrm{ord}}$ is definable using finitely many universal local sentences. We can lift the counter example provided in Lemma 4.3.12 using Lemma 4.3.20.

Theorem 4.3.21. *Let* $\sigma \stackrel{\text{def}}{=} \{ (B,1); (S,2); (E,2); (\leq, 2) \}$. *Then, the* FOLoc-*preservation theorem (re-introduced in Corollary 4.2.1) does not relativise to* $\mathsf{Fin}(\sigma)$.

4.3.2. Undecidability

We have proven in Theorem 4.3.21 that deciding whether a sentence preserved under local elementary embeddings is equivalent to an existential local sentence is a non-trivial problem in $\mathsf{Fin}(\sigma)$[28]. The goal of this subsection is to strengthen this result by proving that this problem is actually undecidable, which will be done in Theorem 4.3.27 on page 108.

28: That is, we have positive and negative instances.

Because Theorem 4.3.27 is proven by directly encoding the runs of a Turing machine, it can actually be slightly modified to conclude that two other decision problems are undecidable, respectively that the semantic property of being preserved under disjoint unions is undecidable (see Theorem 4.3.28), and that there exists no algorithm that, given a sentence φ produces an equivalent existential local sentence ψ, even under the promise that such a ψ exists (see Theorem 4.3.29).

Coding a Universal Turing Machine. Let us fix for the remainder of this section a Universal Turing Machine $\mathcal{U}$ over an alphabet Σ and with control states Q. We extend the signature of $\mathcal{C}_{\mathrm{ord}}$ with unary predicates $(q,1)$ for $q \in Q$, $(P_a, 1)$ for $a \in \Sigma$, $(P_\$, 1)$ and $(P_\Box, 1)$. This allows the encoding of configurations[29] of $\mathcal{U}$ inside the $\leq$-components of structures in $[\![\mathcal{A}_{\mathrm{ord}}]\!]_{\mathsf{Fin}(\sigma)}$. Without loss of generality, we assume that this Universal Turing Machine accepts only on a specific state $q_f^a \in Q$ and rejects only on a specific state $q_f^r \in Q$, those two states being the only ones with no possible forward transitions.

29: We will represent configurations as words over the alphabet Σ starting with the symbol \$, padded to the right with blank symbols $\Box$. The head of the machine is represented by a unary predicate q, that uniquely describes both the position of the head on the machine tape, together with its current state.

Given a structure $\mathfrak{A} \in \mathcal{C}_{\mathrm{ord}}$ and $a \in \mathfrak{A}$, we call $\mathrm{C}^{\leq}(a)$ the $\leq$-component of $a \in \mathfrak{A}$. This notation is handy because $\mathrm{C}^{\leq}(a)$ will also be the configuration of the encoded Turing Machine to which a belongs.

Since $\mathrm{C}^{\leq}(a)$ is exactly of the shape O_m for some $m \geq 2$, it is a total ordering of m elements with the successor relation. Because it is easily seen that one can check whether a total order represents a valid configuration, the following is left as an exercise.

Fact 4.3.22. There exists a 1-local formula $\theta_C(x)$ such that for all $\mathfrak{A} \in \mathcal{C}_{\mathrm{ord}}$, for all $a \in \mathfrak{A}$, $\mathfrak{A}, x \mapsto a \models \theta_C(x)$ if and only if $\mathrm{C}^{\leq}(a)$ represents a valid configuration of $\mathcal{U}$.

The only difficulty in representing runs of the machine $\mathcal{U}$ is to map positions from one $\leq$-component to its successor. To that end, we exploit the first-order definability of the function $f\colon a \mapsto \max_{\leq} \{ d\colon E(a,d) \}$ that links $\leq$-components.

Lemma 4.3.23. *There exists a 1-local formula $\theta_T^{\rightarrow}(x,y)$ such that for every structure $\mathfrak{A} \in \mathcal{C}_{\mathrm{ord}}$ and points $a,b \in \mathfrak{A}$ the following two properties are equivalent:*

1. *The $\leq$-components $\mathrm{C}^{\leq}(a)$ and $\mathrm{C}^{\leq}(b)$ are connected via the relation S, $\mathrm{C}^{\leq}(a)$ represents a valid configuration of $\mathcal{U}$, $\mathrm{C}^{\leq}(b)$ represents a valid configuration of $\mathcal{U}$, and the transition from $\mathrm{C}^{\leq}(a)$ to $\mathrm{C}^{\leq}(b)$ is a valid transition of $\mathcal{U}$.*
2. *$\mathfrak{A}, (a \mapsto x, b \mapsto y) \models \theta_T^{\rightarrow}(x,y)$.*

Proof Sketch. The only technical issue lies in relating the positions in the configuration $\mathrm{C}^{\leq}(a)$ to positions in the configuration $\mathrm{C}^{\leq}(b)$, which is done through the use of the function $f\colon a \mapsto \max_{\leq} \{ d\colon E(a,d) \}$, which is first-order definable as a 1-local formula as follows:

$$\phi_f(x,y) \stackrel{\text{def}}{=} E(x,y) \land \forall z. y < z \Rightarrow \neg E(x,z) \quad .$$

As a shorthand, let us write $z \in \mathrm{C}^{\leq}(x)$ instead of $z \leq x \lor x \leq z$.

Using the function f, one can for instance assert that every letter except those near the current position in the tape are left unchanged. To that end, one can first write a formula stating that the position of the head of the Turing Machine in $\mathrm{C}^{\leq}(x)$ is not close to x itself as follows:

$$\phi_Q(x) \stackrel{\text{def}}{=} \forall z.(z \in \mathrm{C}^{\leq}(x) \land (S(x,z) \lor S(z,x) \lor x = z)) \Rightarrow \bigwedge_{q \in Q} \neg q(x)$$

Using ϕ_f and ϕ_Q it is easy to write a formula stating that letters far from the head of the Turing Machine $\mathcal{U}$ are unchanged in a transition:

$$\begin{aligned}\phi(x,y) \stackrel{\text{def}}{=} &\forall z. z \in \mathrm{C}^{\leq}(x) \land \phi_Q(z) \\ &\Rightarrow \exists z'. z' \in \mathrm{C}^{\leq}(y) \land \phi_f(z,z') \land \bigwedge_{a \in \Sigma \cup \{\$, \square\}} P_a(z) \Leftrightarrow P_a(z')\end{aligned}$$

We leave the other checks as an exercise, as all of them follow the same pattern. ■

Following the ideas of Lemma 4.3.23, we leave as an exercise for the reader the definition of the following formulas: θ_I^w, θ_N, θ_F in Facts 4.3.24 to 4.3.26.

Fact 4.3.24. There exists a 1-local formula $\theta_N(x)$ such that for all $\mathfrak{A} \in \mathcal{C}_{\mathrm{ord}}$ and $a \in \mathfrak{A}$, $\mathfrak{A}, x \mapsto a \models \theta_N(x)$ if and only if $\mathrm{C}^{\leq}(a)$ has no S-successor in $\mathfrak{A}$.

Fact 4.3.25. Given a word $w \in \Sigma^*$ there exists a 1-local formula $\theta_I^w(x)$ such that for all $\mathfrak{A} \in \mathcal{C}_{\mathrm{ord}}$ and $a \in \mathfrak{A}$, $\mathfrak{A}, x \mapsto a \models \theta_I^w(x)$ if and only if $\mathsf{C}^{\leq}(a)$ is the initial configuration of $\mathcal{U}$ on the word w and has no S-predecessor in $\mathfrak{A}$.

Fact 4.3.26. There exists a 1-local formula $\theta_F(x)$ such that for $\mathfrak{A} \in \mathcal{C}_{\mathrm{ord}}$ and $a \in \mathfrak{A}$, $\mathfrak{A}, x \mapsto a \models \theta_F(x)$ if and only if $\mathsf{C}^{\leq}(a)$ is a final configuration of $\mathcal{U}$ and has no S-successor in $\mathfrak{A}$.

Leveraging the Coding System. We now are able to simply use the coding system developed above to deduce that most of the problems we may be interested in are actually undecidable. To that end, let us first recall that in Lemma 4.3.16, we showed not only that $\mathsf{2CC}$ is first-order definable in $\mathcal{C}_{\mathrm{ord}}$, but also that it is definable as an existential local sentence.

Theorem 4.3.27. *It is in general not possible to decide, given a sentence φ, whether there exists an equivalent existential local sentence over $\mathsf{Fin}(\sigma)$, even under the promise that φ is preserved under disjoint unions.*

Proof. Without loss of generality, thanks to Lemma 4.3.20, we only work over $\mathcal{C}_{\mathrm{ord}}$, and we reduce from the halting problem.

Let $\mathcal{T}$ be a Turing Machine and $\langle\mathcal{T}\rangle$ be its code in the alphabet of the Universal Turing Machine $\mathcal{U}$. Let $\varphi_{\mathcal{T}}$ be defined as the following existential local sentence: $\exists x.\theta_I^{\langle\mathcal{T}\rangle}(x) \wedge \forall x, y.S(x,y) \wedge \neg(x \leq y) \Rightarrow \theta_T^{\rightrightarrows}(x,y)$. We consider the sentence[30] $\varphi \stackrel{\text{def}}{=} \varphi_{\mathcal{T}} \vee \mathsf{2CC}$ that is closed under disjoint unions over $\mathcal{C}_{\mathrm{ord}}$. This sentence is computable from the data $\langle\mathcal{T}\rangle$.

30: A more understandable definition is $\varphi \stackrel{\text{def}}{=} \neg\mathsf{2CC} \Rightarrow \varphi_{\mathcal{T}}$, but the one given clearly shows that it is preserved under disjoint unions.

Assume that $\mathcal{T}$ halts. Then, there exists a bound k for the run of the universal Turing Machine $\mathcal{U}$. Given a size $n \in \mathbb{N}$, we define $\varphi_{\mathcal{T}}^n$ to be the following existential local sentence: $\exists x_1, \dots, x_n.\theta_I^{\langle\mathcal{T}\rangle}(x_1) \wedge \theta_T^{\rightrightarrows}(x_1, x_2) \wedge \dots \wedge \theta_T^{\rightrightarrows}(x_{n-1}, x_n) \wedge \theta_N(x_n)$. It is a routine check that φ is equivalent to the following sentence over $\mathcal{C}_{\mathrm{ord}}$:

$$\varphi' \stackrel{\text{def}}{=} \mathsf{2CC} \vee \bigvee_{1 \leq n \leq k} \varphi_{\mathcal{T}}^n$$

Assume that $\mathcal{T}$ does not halt. The universal Turing Machine $\mathcal{U}$ does not halt on the word $\langle\mathcal{T}\rangle$. Assume by contradiction that φ is equivalent to a sentence $\psi \stackrel{\text{def}}{=} \exists x_1, \dots, x_k.\theta$ where θ is r-local. Find a run of size greater than $k \cdot (2r+1)$ and evaluate ψ. It cannot look at all the configurations simultaneously; change the state of the one not seen: it still satisfies ψ but this is no longer a run of $\mathcal{U}$, which is absurd. ∎

An analogous proof allows us to conclude that the semantic property of closure under disjoint unions is also undecidable. Theorem 4.3.27 is then strengthened to prove that equivalent existential sentences are uncomputable in general.

Theorem 4.3.28. *It not possible, given a sentence φ, to decide whether or not it is preserved under disjoint unions over* $\mathsf{Fin}(\sigma)$.

Proof. Without loss of generality, thanks to Lemma 4.3.20, we only work over $\mathcal{C}_{\mathrm{ord}}$, and we reduce from the halting problem. Consider the sentence $\varphi_{\mathcal{T}}$ defined as in the proof of Theorem 4.3.27: $\exists x.\theta_I^{\langle\mathcal{T}\rangle}(x) \wedge \forall x, y.S(x,y) \wedge \neg(x \leq y) \Rightarrow \theta_{\mathcal{T}}^{\rightrightarrows}(x,y)$, and let now $\varphi \overset{\mathsf{def}}{=} \varphi_{\mathcal{T}} \wedge \neg 2\mathsf{CC}$.

Assume that $\mathcal{T}$ halts. There exists exactly one model of φ, the unique run of the universal Turing machine $\mathcal{U}$, and this contradicts preservation under disjoint unions.

Assume that $\mathcal{T}$ does not halt. The sentence φ has no finite model and is therefore (trivially) preserved under disjoint unions.

Hence, the sentence φ is preserved under disjoint unions if and only if $\mathcal{T}$ does not halt. ∎

Theorem 4.3.29. *There is no algorithm which, given a sentence φ that is equivalent to some existential local sentence over* $\mathsf{Fin}(\sigma)$, *computes such a sentence.*

Proof. Without loss of generality thanks to Lemma 4.3.20, we only work over $\mathcal{C}_{\mathrm{ord}}$, and we reduce from the halting problem. Let $\mathcal{T}$ be a Turing Machine, $\varphi_{\mathcal{T}} \overset{\mathsf{def}}{=} \exists x.\theta_I^{\langle\mathcal{T}\rangle}(x) \wedge \forall x, y.S(x,y) \wedge \neg(x \leq y) \Rightarrow \theta_{\mathcal{T}}^{\rightrightarrows}(x,y)$, and $\varphi \overset{\mathsf{def}}{=} \varphi_{\mathcal{T}} \vee 2\mathsf{CC}$, as it was done in Theorem 4.3.27.

Assume that $\mathcal{T}$ halts. Then φ is equivalent to an existential local sentence $\varphi' \vee 2\mathsf{CC}$, as observed in the proof of Theorem 4.3.27. Assume that $\mathcal{T}$ does not halt. Then φ is equivalent to $2\mathsf{CC}$, which is existential local. We have proven that φ is equivalent to an existential local sentence in all cases.

Assume by contradiction that there exists an algorithm computing an existential local sentence μ that is equivalent to $\varphi_{\mathcal{T}} \vee 2\mathsf{CC}$ over $\mathcal{C}_{\mathrm{ord}}$. Then, one can use μ to decide whether $\mathcal{T}$ halts. Let us write k for the number of existential quantifiers of μ and r for the locality radius of its inner formula. If the sentence μ accepts the coding of a run of size greater than $2k \cdot (2r+1)$, then it accepts structures that are not coding runs of $\mathcal{U}$, and μ cannot be equivalent to φ. As a consequence, if $\mathcal{T}$ terminates, it must terminate in at most $2k \cdot (2r+1)$ steps, which is decidable. ∎

4.3.3. Generalisation to Weaker Quasi-Orders

Although the FOLoc-preservation theorem[31] fails in the finite (as demonstrated in Theorem 4.3.21), recall that the refinement of the quasi-orders into $\Rrightarrow_q^{r,k}$ with finite parameters allowed to somehow get around this issue in Proposition 4.3.9. Namely, a sentence preserved under $\Rrightarrow_q^{r,k}$ for some finite $r, q, k \in \mathbb{N}$ is always equivalent to an existential local sentence over $\mathsf{Fin}(\sigma)$.

31: That was restated in Corollary 4.2.1

Hence, it is natural to ask for $0 \leq r, q, k \leq \infty$, whether sentences preserved under $\Rrightarrow_q^{r,k}$ over $\mathsf{Fin}(\sigma)$ are always equivalent to existential local sentences. Proposition 4.3.9 answers this question positively for $0 \leq r, q, k < \infty$, and Theorem 4.3.21 answers negatively for $r = q = k = \infty$.

The main idea is that whenever a counter-example sentence can be build, it is because of our previous counter example $[\![\mathcal{A}_{\mathrm{ord}}]\!]_{\mathsf{Fin}(\sigma)}$. We provide in Figure 4.13 a panel of the existence of an existential local form for different values for r, q and k over the class of finite structures $\mathsf{Fin}(\sigma)$. This picture is obtained by compiling the earlier results of Proposition 4.3.9 and Theorem 4.3.21, with the yet to be proven results of Lemmas 4.3.30, 4.3.31 and 4.3.33. Furthermore, one particular instance of Figure 4.13 is postponed to the next chapter, Chapter 5 (A Local-to-Global Preservation Theorem), because its proof tells a self-contained story about induced substructures.

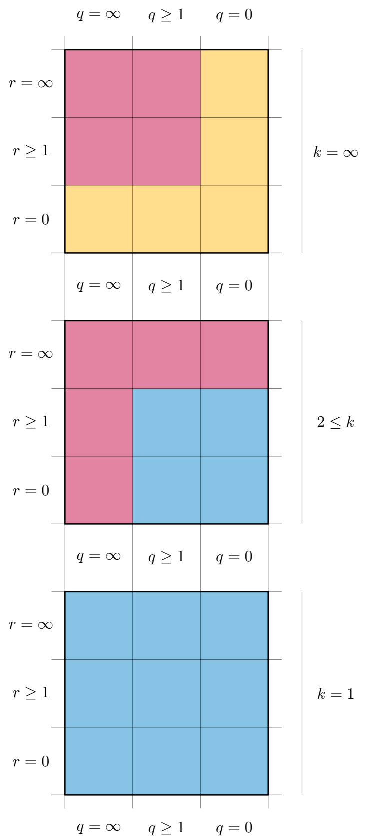


Figure 4.13.: The preservation cube: studying whether formulas preserved under $\Rrightarrow_q^{r,k}$ are equivalent over $\mathsf{Fin}(\sigma)$ to existential local sentences. In red, the answer is 'NO', in blue, the answer is 'YES'. In yellow, the answer is 'YES', but the proof is postponed to Chapter 5 (A Local-to-Global Preservation Theorem).

We will first prove that if one allows arbitrary radius or arbitrary quantifier rank, then the counter example that was built for existential local sentences over $\mathsf{Fin}(\sigma)$ using the class $\mathcal{C}_{\mathrm{ord}}$ and the sentence $\mathsf{2CC}$ can be re-encoded in this seemingly less powerful fragments. Typically, we will list the axioms of $\mathcal{A}_{\mathrm{ord}}$ and see that they can be suitably rewritten.

In the specific case of $k = 1$, such methods will not apply because the quasi-orders $\Rrightarrow_q^{r,1}$ cannot distinguish a structure A from $A \uplus A$, even when $r = q = \infty$. And in fact, every sentence that is preserved under $\Rrightarrow_q^{r,1}$ over $\mathsf{Fin}(\sigma)$ is equivalent to an existential local sentence over $\mathsf{Fin}(\sigma)$, as proven in Lemma 4.3.33.

Lemma 4.3.30. *Let $\sigma \stackrel{\text{def}}{=} \{(\leq, 2), (S, 2), (E, 2), (B, 1)\}$. There exists a sentence ϕ_{bad} preserved under $\Rrightarrow_q^{\infty,k}$ over $\mathsf{Fin}(\sigma)$ for $k \geq 2$ and $q \geq 1$ that is not equivalent to an existential local sentence over $\mathsf{Fin}(\sigma)$.*

Proof. Using Fact 4.3.4, it suffices to consider the case where $k = 2$ and $q = 1$ to conclude.

A first check is that the axiomatisation of $\mathcal{C}_{\mathrm{ord}}$ that was given in Figure 4.12 can be rewritten as a conjunction of sentences of the form $\forall x_1, \forall x_2, Qy \in \mathcal{N}(x_1 x_2, 1), \theta(x_1, x_2, y)$ where θ is quantifier-free, and $Q \in \{\exists, \forall\}$. This is true for all but one axiom that needs to be rewritten carefully: Axiom 10, that states that the relation E "factorises" through $(\leq)(S)(\leq)$. For that particular axiom, we can equivalently state the following "compatibility" axiom:

$$\forall x, \forall y, [E(x, y) \wedge (\forall z. x \leq z \Rightarrow z = x) \wedge (\forall z. y \leq z \Rightarrow z = y)] \Rightarrow S(x, y)$$

Let us briefly sketch the proof that this new formulation implies (together with the other axioms) Axiom 10. The main idea is that whenever $E(a, b)$ holds, one can leverage axioms Axioms 11 and 12 to conclude that the $\leq$-maximal element a' in the $\leq$-component of a is connected through E to the $\leq$-minimal element b' in the $\leq$-component of b. That is, $E(a', b')$ holds. The new axiom then proves that $S(a', b')$ holds, and in particular, $a \leq a' \; S \; b' \leq b$. Conversely, this new sentence clearly holds on structures of $\mathcal{C}_{\mathrm{ord}}$.

This proves that $\mathcal{C}_{\mathrm{ord}}$ is definable and downwards closed for $\Rrightarrow_1^{1,2}$. In particular, sentences preserved under $\Rrightarrow_1^{\infty,2}$ over $\mathcal{C}_{\mathrm{ord}}$ are preserved under $\Rrightarrow_1^{\infty,2}$ over $\mathsf{Fin}(\sigma)$, which is a strengthening of Lemma 4.3.20.

Using the same syntactical analysis, $\mathsf{2CC}$ is preserved under $\Rrightarrow_1^{\infty,2}$ over $\mathcal{C}_{\mathrm{ord}}$. As a consequence it is an easy check that ϕ_{bad}[32] is preserved under $\Rrightarrow_1^{\infty,2}$ over $\mathcal{C}_{\mathrm{ord}}$. Moreover, it suffices to leverage Lemma 4.3.12 to conclude that ϕ_{bad} cannot be defined as an existential local sentence over $\mathcal{C}_{\mathrm{ord}}$. ■

32: Recall that $\phi_{\mathrm{bad}} \overset{\mathsf{def}}{=} \forall x.\neg B(x) \vee \mathsf{2CC}$

Lemma 4.3.31. *Let* $\sigma \overset{\mathsf{def}}{=} \{(\leq, 2), (S, 2), (E, 2), (B, 1)\}$. *For every* $r \geq 1, k \geq 2$, *there exists a sentence* φ *preserved under* $\Rrightarrow_\infty^{r,k}$ *over* $\mathsf{Fin}(\sigma)$ *but not equivalent to an existential local sentence over* $\mathsf{Fin}(\sigma)$.

Proof. Using Fact 4.3.4 it suffices to consider the case $r = 1$ and $k = 2$. The proof follows the same pattern as Lemma 4.3.30, we enumerate the axioms from $\mathcal{A}_{\mathrm{ord}}$ and notice that they are of the form $\forall x.\theta(x)$ where $\theta(x)$ is a 1-local formula. As a consequence, sentences preserved under $\Rrightarrow_\infty^{1,2}$ over $\mathcal{C}_{\mathrm{ord}}$ are preserved under $\Rrightarrow_\infty^{1,2}$ over $\mathsf{Fin}(\sigma)$, which is a strengthening of Lemma 4.3.20.

Moreover, $\mathsf{2CC}$[33] is preserved under $\Rrightarrow_\infty^{1,2}$ using a simple syntactical analysis. We now check that ϕ_{bad} is preserved under $\Rrightarrow_\infty^{1,2}$ over $\mathcal{C}_{\mathrm{ord}}$. Let $\mathfrak{A}, \mathfrak{B} \in \mathcal{C}_{\mathrm{ord}}$ such that $\mathfrak{A} \models \phi_{\mathrm{bad}}$ and $\mathfrak{A} \Rrightarrow_\infty^{1,2} \mathfrak{B}$.

33: As defined in Lemma 4.3.16.

- Let us first examine the case where A has a single connected component.
  Let $a \in \mathfrak{A}$; since $\mathcal{N}_{\mathfrak{A}}(a, 1)$ is finite, there exists a 1-local formula $\psi_a(x)$ of quantifier rank less than $|\mathcal{N}_{\mathfrak{A}}(a, 1)| + 1$ such that $\mathfrak{B}, x \mapsto b \models \psi_a(x)$ if and only if $\mathcal{N}_{\mathfrak{A}}(a, 1)$ is isomorphic to $\mathcal{N}_{\mathfrak{B}}(b, 1)$. In particular, if C is a $\leq$-component of $\mathfrak{A}$, it is of radius at most 1 and there exists C' a $\leq$-component of $\mathfrak{B}$ isomorphic to C. Moreover, the 1-neighbourhood of a $\leq$-component contains the previous and next $\leq$-components for S.
  If $\mathfrak{B}$ has a single connected component, then two distinct $\leq$-components in $\mathfrak{B}$ must have distinct sizes. Using the fact that the components of $\mathfrak{A}$ are all found in $\mathfrak{B}$ and that their relative position is preserved, this proves that $\mathfrak{B}$ contains exactly the same $\leq$-components as $\mathfrak{A}$, which only happens if $\mathfrak{A} = \mathfrak{B}$.
  If $\mathfrak{B}$ has at least two connected components, then it satisfies $\mathsf{2CC}$ which implies ϕ_{bad}.
- In the case where $\mathfrak{A}$ has two connected components, $\mathfrak{A} \models \mathsf{2CC}$ but then $\mathfrak{B} \models \mathsf{2CC}$ and $\mathfrak{B} \models \phi_{\mathrm{bad}}$.

Moreover, we know from Lemma 4.3.12 that ϕ_{bad} cannot be defined as an existential local sentence over $\mathcal{C}_{\mathrm{ord}}$. ■

Before studying the case $k = 1$, let us describe the behaviour of $\Rrightarrow_q^{r,k}$ with respect to disjoint unions. In particular, we prove that for any fixed k, the quasi-order cannot distinguish between more than k copies of the same structure.

Lemma 4.3.32. *Let σ be a finite relational signature. Let $\mathfrak{A}, \mathfrak{B} \in \mathsf{Fin}(\sigma)$. For all $0 \leq r, q \leq \infty$ and $1 \leq k < \infty$, $\mathfrak{A} \Rrightarrow_q^{r,k} \mathfrak{A} \uplus \mathfrak{B}$ and $\biguplus_{i=1}^{k+n} \mathfrak{A} \Rrightarrow_q^{r,k} \biguplus_{i=1}^{k} \mathfrak{A}$ for all $n \in \mathbb{N}$.*

Proof. Let us prove that $\mathfrak{A} \Rrightarrow_q^{r,k} \mathfrak{A} \uplus \mathfrak{B}$. Consider a vector $\vec{a} \in \mathfrak{A}^k$; it is clear that this vector appears as-is in $\mathfrak{A} \uplus \mathfrak{B}$ and since the union is disjoint, thus $\mathcal{N}_{\mathfrak{A} \uplus \mathfrak{B}}(\vec{a}, r) = \mathcal{N}_{\mathfrak{A}}(\vec{a}, r)$. In particular, $\mathsf{tp}^q_{\mathfrak{A}}(\vec{a}, r) = \mathsf{tp}^q_{\mathfrak{A} \uplus \mathfrak{B}}(\vec{a}, r)$ for all $q \geq 0$.

Let us write $\mathfrak{A}^{\uplus k} \stackrel{\text{def}}{=} \biguplus_{i=1}^{k} \mathfrak{A}$ and $\mathfrak{A}^{\uplus k+n} \stackrel{\text{def}}{=} \biguplus_{i=1}^{k+n} \mathfrak{A}$. Remark that the previous statement shows $\mathfrak{A}^{\uplus k} \Rrightarrow_q^{r,k} \mathfrak{A}^{\uplus k+n}$. Let us now prove that $\mathfrak{A}^{\uplus k+n} \Rrightarrow_q^{r,k} \mathfrak{A}^{\uplus k}$. Consider a vector $\vec{a} \in (\mathfrak{A}^{\uplus k+n})^k$; this vector has elements in at most k copies of $\mathfrak{A}$, hence one can select one copy of $\mathfrak{A}$ in $\mathfrak{A}^{\uplus k}$ for each of those and consider the exact same elements in those copies. As the unions are disjoint, the obtained neighbourhoods are isomorphic to those in $\mathfrak{A}^{\uplus k+n}$ and in particular share the same local types. ■

Lemma 4.3.33. *Let σ be a finite relational signature. For every $0 \leq r, q \leq \infty$, for every sentence φ preserved under $\Rrightarrow_q^{r,1}$ over $\mathsf{Fin}(\sigma)$, there exists an existential local sentence ψ that is equivalent to φ over $\mathsf{Fin}(\sigma)$.*

Proof. Without loss of generality thanks to Fact 4.3.4 we consider a sentence φ preserved under $\Rrightarrow_\infty^{\infty,1}$. Let us prove that φ is preserved under $\Rrightarrow_q^{r,k}$ where all parameters are finite. This property combined with Proposition 4.3.9 will prove that φ is equivalent to an existential local sentence.

Thanks to Gaifman's Locality Theorem, we can assume that φ is a Boolean combination of the following basic local sentences for $1 \leq i \leq n$: $\theta_i = \exists^{\geq k_i}_{r_i} x.\psi_i(x)$, where each $\psi_i(x)$ is an r_i-local formula of quantifier rank q_i. Define $r \stackrel{\text{def}}{=} \max\{ r_i : 1 \leq i \leq n \}$, $q \stackrel{\text{def}}{=} \max\{ q_i : 1 \leq i \leq n \}$ and $k \stackrel{\text{def}}{=} \max\{ k_i : 1 \leq i \leq n \}$.

Let $\mathfrak{A}, \mathfrak{B}$ be two finite structures such that $\mathfrak{A} \models \varphi$ and $\mathfrak{A} \Rrightarrow_q^{r,k} \mathfrak{B}$. Our goal is to prove that $\mathfrak{B} \models \varphi$. Let us write $\mathfrak{B}^{\uplus k}$ for the disjoint union of k copies of $\mathfrak{B}$, as in the previous proof.

Let us show that $\mathfrak{A} \uplus \mathfrak{B}^{\uplus k} \models \varphi$ if and only if $\mathfrak{B}^{\uplus k} \models \varphi$. To that end, let us fix $1 \leq i \leq n$ and prove that $\mathfrak{A} \uplus \mathfrak{B}^{\uplus k} \models \exists^{\geq k_i}_{r_i} x.\psi_i(x)$ if and only if $\mathfrak{B}^{\uplus k} \models \exists^{\geq k_i}_{r_i} x.\psi_i(x)$.

- Assume that $\mathfrak{A} \uplus \mathfrak{B}^{\uplus k} \models \exists^{\geq k_i}_{r_i} x.\psi_i(x)$, there exists a vector $\vec{c}$ of witnesses of ψ at pairwise distance greater than $2r_i$ in $\mathfrak{A} \uplus \mathfrak{B}^{\uplus k}$. If $\vec{c}$ lies in $\mathfrak{B}^{\uplus k}$ then we conclude, otherwise some element of $\vec{c}$ lies in $\mathfrak{A}$. In particular, $\mathfrak{A} \models \exists x.\psi_i(x)$, which is an existential r-local sentence of quantifier rank at most q. Since $\mathfrak{A} \Rrightarrow_q^{r,k} \mathfrak{B}$, we know that $\mathfrak{B} \models \exists x.\psi_i(x)$ and as $k_i \leq k$, $\mathfrak{B}^{\uplus k} \models \exists^{\geq k_i}_{r_i} x.\psi_i(x)$.
- Conversely, whenever $\mathfrak{B}^{\uplus k} \models \exists^{\geq k_i}_{r_i} x.\psi_i(x)$ the structure $\mathfrak{A} \uplus \mathfrak{B}^{\uplus k}$ satisfies $\exists^{\geq k_i}_{r_i} x.\psi(x)$ as basic local sentences are preserved under disjoint unions (see Propositions 4.3.8 and 4.3.10).

Since $\mathfrak{A} \Rrightarrow_{\infty}^{\infty,1} \mathfrak{A} \uplus \mathfrak{B} \Rrightarrow_{\infty}^{\infty,1} \mathfrak{A} \uplus \mathfrak{B}^{\uplus k}$ (by Lemma 4.3.32), and since $\mathfrak{A} \models \varphi$, we know that $\mathfrak{A} \uplus \mathfrak{B}^{\uplus k} \models \varphi$. This implies that $\mathfrak{B}^{\uplus k} \models \varphi$, and we deduce from Lemma 4.3.32 that $\mathfrak{B}^{\uplus k}$ is $\Rrightarrow_{\infty}^{\infty,1}$-equivalent to $\mathfrak{B}$. In particular $\mathfrak{B}^{\uplus k} \models \varphi$ implies $\mathfrak{B} \models \varphi$. ■

4.4. Discussion

Positive locality and existential sentences. The positive locality theorem in Theorem 4.2.2 can be thought of as a “local” variant of the existential Gaifman normal form of [55], hence can find similar applications in the optimisation of query evaluation. We stress that the two results are not comparable, even though they use similar combinatorial tools[34]. Furthermore, Theorem 4.2.2 is actually a theorem stating the equi-expressiveness of two fragments of FO, while the existential Gaifman normal form clearly allows building formulas that are not expressible in EFO.

[55]: Grohe and Wöhrle (2004), ‘An existential locality theorem’

34: That is, the combinatorics behind the intersection patterns of local neighbourhoods

Failure in the finite. The proofs of undecidability and failure in the finite case closely follows the generic methods that have been employed for other preservation theorems [4, 5, 16, 66, 89, 90]. These results are incomparable, and there still lacks a “meta proof” that generalises the ad-hoc study that is performed. It could be that, on the opposite of well-quasi-orderings and Noetherian spaces that imply the relativisation, there is a simple property that implies the non-relativisation. Furthermore, the connection between relativisation and effective procedures remains unclear: experimentally the two are related, but there is no theoretical statement linking the two properties.

Relativisation in the finite. As for usual preservation theorems, a natural question after the failure to relativise the FOLoc-preservation theorem to $\mathsf{Fin}(\sigma)$ is to find natural subclasses where the theorem relativises. This is still an open question, but note hat this relativisation fails even in the simple case of disjoint unions of coloured paths (see Example 4.3.11).

References

[4] Miklós Ajtai and Yuri Gurevich. 'Monotone versus positive'. In: *Journal of the ACM* 34 (1987), pp. 1004–1015. DOI: 10.1145/31846.31852 (cit. on pp. 35, 37, 113).

[5] Miklós Ajtai and Yuri Gurevich. 'Datalog vs first-order logic'. In: *Journal of Computer and System Sciences* 49 (1994), pp. 562–588. DOI: 10.1016/S0022-0000(05)80071-6 (cit. on pp. 35, 37, 113).

[6] Albert Atserias, Anuj Dawar and Martin Grohe. 'Preservation under extensions on well-behaved finite structures'. In: *SIAM Journal on Computing* 38 (2008), pp. 1364–1381. DOI: 10.1137/060658709 (cit. on pp. 37, 40, 47, 55, 64, 67, 68, 74, 97, 98, 117, 121, 127).

[7] Albert Atserias, Anuj Dawar and Phokion G. Kolaitis. 'On preservation under homomorphisms and unions of conjunctive queries'. In: *Journal of the ACM* 53 (2006), pp. 208–237. DOI: 10.1145/1131342.1131344 (cit. on pp. 37, 40, 47, 55, 74).

[15] Chen Chung Chang and H. Jerome Keisler. *Model Theory*. Vol. 73. Elsevier, 1990 (cit. on pp. i, 29, 32–34, 68).

[16] Yijia Chen and Jörg Flum. 'Forbidden Induced Subgraphs and the Łoś-Tarski Theorem'. In: *2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*. 2021, pp. 1–13. DOI: 10.1109/LICS52264.2021.9470742 (cit. on pp. 37, 39, 98, 113).

[20] Anuj Dawar. 'Homomorphism preservation on quasi-wide classes'. In: *Journal of Computer and System Sciences* 76 (2010), pp. 324–332. DOI: 10.1016/j.jcss.2009.10.005 (cit. on pp. 47, 55, 74).

[21] Anuj Dawar, Martin Grohe, Stephan Kreutzer and Nicole Schweikardt. 'Approximation schemes for first-order definable optimisation problems'. In: *21st Annual IEEE Symposium on Logic in Computer Science (LICS'06)*. 2006, pp. 411–420. DOI: 10.1109/lics.2006.13 (cit. on pp. 68, 130, 131).

[33] Solomon Feferman and Robert Vaught. 'The first order properties of products of algebraic systems'. In: *Fundamenta Mathematicae* 47 (1959), pp. 57–103. DOI: 10.4064/fm-47-1-57-103 (cit. on pp. 78, 155).

[37] Haim Gaifman. 'On local and non-local properties'. In: *Proceedings of the Herbrand Symposium*. Vol. 107. Elsevier, 1982, pp. 105–135. DOI: 10.1016/S0049-237X(08)71879-2 (cit. on pp. 72, 73).

[38] Jakub Gajarsky, Maximilian Gorsky and Stephan Kreutzer. 'Differential Games, Locality, and Model Checking for FO Logic of Graphs'. In: 216 (2022). Ed. by Florin Manea and Alex Simpson, 22:1–22:18. DOI: 10.4230/LIPIcs.CSL.2022.22 (cit. on pp. 71, 233).

[55] Martin Grohe and Stefan Wöhrle. 'An existential locality theorem'. In: *Annals of Pure and Applied Logic* 129 (2004), pp. 131–148. DOI: 10.1016/j.apal.2004.01.005 (cit. on pp. 67, 68, 75, 76, 85, 93, 113).

[57] Frederik Harwath, Lucas Heimberg and Nicole Schweikardt. 'Preservation and decomposition theorems for bounded degree structures'. In: *Logical Methods in Computer Science* Volume 11, Issue 4 (2015). DOI: 10.2168/LMCS-11(4:17)2015. URL: https://lmcs.episciences.org/1618 (cit. on p. 74).

[66] Denis Kuperberg. '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, 2021, pp. 1–13. DOI: 10.1109/LICS52264.2021.9470602 (cit. on pp. 37, 39, 68, 98, 113, 137).

[68] Leonid Libkin. *Elements of finite model theory*. Springer, 2012. DOI: 10.1007/978-3-662-07003-1 (cit. on pp. 11, 68, 239).

[71] Aliaume Lopez. 'When Locality Meets Preservation'. In: *LICS'22* (2022), pp. 1–14. DOI: 10.1145/3531130.3532498. URL: https://doi.org/10.1145/3531130.3532498 (cit. on pp. 63, 67, 117).

[74] Johann A. Makowsky. 'Algorithmic uses of the Feferman–Vaught Theorem'. In: *Annals of Pure and Applied Logic* 126 (2004), pp. 159–213. DOI: 10.1016/j.apal.2003.11.002 (cit. on pp. 78, 155).

[83] Benjamin Rossman. 'Homomorphism preservation theorems'. In: *Journal of the ACM* 55 (2008), 15:1–15:53. DOI: `10.1145/1379759.1379763` (cit. on pp. i, 24, 35, 37, 38, 40, 42, 55, 74, 175, 177, 235).

[87] Thomas Schwentick and Klaus Barthelmann. 'Local normal forms for first-order logic with applications to games and automata'. In: *Discrete Mathematics and Theoretical Computer Science* 3 (1999), pp. 109–124. DOI: `10.46298/dmtcs.254`. URL: `https://www.dmtcs.org/dmtcs-ojs/index.php/dmtcs/article/view/101/234.html` (cit. on pp. 67, 75).

[89] Alexei P. Stolboushkin. 'Finitely monotone properties'. In: (1995), pp. 324–330. DOI: `10.1109/LICS.1995.523267` (cit. on pp. 35, 113, 137).

[90] William W. Tait. 'A counterexample to a conjecture of Scott and Suppes'. In: *Journal of Symbolic Logic* 24 (1959), pp. 15–16. DOI: `10.2307/2964569` (cit. on pp. i, 35, 37, 39, 97, 100, 113, 137).

A Local-to-Global Preservation Theorem

5.

Outline of the chapter

In this chapter, we will continue the study of preservation theorems via locality that started with Chapter 4 (Locality and Preservation). While the latter left a bittersweet aftertaste, the goal here is to demonstrate that, in one specific configuration, one can leverage the locality of first order logic to prove that the Łoś-Tarski Theorem relativises to a class $\mathcal{C}$ of finite structures.

Goals of the chapter

At the end of the chapter, I hope you will be convinced that the "local to global" guide to preservation is a powerful technique that can be applied to relativise the Łoś-Tarski Theorem whenever the local behaviours of the structure are "simpler" than their global one. This also provides information about when this technique does not apply: when the local neighbourhoods are as complex as the structures themselves.

Genesis. This chapter is the original goal that led to the study of Chapter 4 (Locality and Preservation). Recall that the interest in the local behaviour of first-order logic was sparked in Chapter 3 (Preservation Theorems for First Order Queries) by the relativisation of the Łoś-Tarski Theorem to hereditary classes of structures of bounded degree that are closed under disjoint unions [6]. Also recall that Subsection 4.3.1 and Subsection 4.3.3 put a stop to the original idea of a generic proof scheme that starts by applying a positive locality result (see Figure 4.13). We will see that the infinitesimal space that was not tackled[1] is enough to circumvent the failure of the FOLoc-preservation theorem in the finite. The resulting Theorem 5.1.2 was already stated in [71], but its "generic" counterpart Theorem 5.3.1 is an addition made while writing this thesis.

[6]: Atserias, Dawar and Grohe (2008), 'Preservation under extensions on well-behaved finite structures'

1: That is, when $\Rrightarrow_q^{r,k}$ is precisely $\subseteq_i$.

Contributions. The goal of this section is to prove the following "local to global" theorem, generalising the previously known results over hereditary class of structures that are closed under disjoint unions [6].

[6]: Atserias, Dawar and Grohe (2008), 'Preservation under extensions on well-behaved finite structures'

Theorem 5.1.2. *Let $\mathcal{C}$ be a hereditary class of finite structures closed under disjoint unions. The Łoś-Tarski Theorem relativises to $\mathcal{C}$ if and only if the Łoś-Tarski Theorem relativises to* $\mathsf{Local}(\mathcal{C}, r, k)$ *for all $r \geq 0$ and $k \geq 1$.*

While the whole section is organised around Łoś-Tarski's Theorem and induced substructures, we have in fact a more general construction for fragments of EFO.

> **Theorem 5.3.1.** *Let $\mathcal{C}$ be a hereditary class of finite structures closed under disjoint unions, and $\mathsf{F} \in \{\mathsf{EFO}, \mathsf{EPFO}^{\neq}, \mathsf{EPFO}\}$ be a fragment. The F-preservation theorem relativises to $\mathcal{C}$ if and only if the F-preservation theorem relativises to $\mathsf{Local}(\mathcal{C}, r, k)$ for all $r \geq 0$ and $k \geq 1$.*

Even though Subsection 4.3.1 essentially closed the door to such a proof scheme we will actually leverage Chapter 4 (Locality and Preservation), and in particular the finitary version of the FOLoc-preservation theorem (Proposition 4.3.9) to prove Theorems 5.1.2 and 5.3.1.

As hinted in Figure 4.13, produced during the study of the "local preorders" $\Rrightarrow_q^{r,k}$ in Subsection 4.3.3 (Generalisation to Weaker Quasi-Orders), the specific case of $k = \infty$, $r = 0 \vee q = 0$ has a wildly different behaviour than the rest of the "preorder cube." In particular, it is quite easy to notice that $\Rrightarrow_q^{0,\infty} = \Rrightarrow_0^{r,\infty} = \subseteq_i$ for $0 \leq r, q < \infty$ (see Example 4.3.5).

We define a class $\mathcal{C}$ to be *localisable* whenever first-order sentences preserved under $\subseteq_i$ (over $\mathcal{C}$) are equivalent (over $\mathcal{C}$) to existential local sentences. In particular, every class $\mathcal{C}$ over which the Łoś-Tarski Theorem relativises is localisable, since existential sentences are specific existential local sentences.

The first section of this chapter is devoted to proving Theorem 5.1.2, under the assumption that some classes are localisable. This is not a complicated task, and we will see that it actually encompasses most of the previously known relativisations in the finite, and even provides new examples of such classes (see Figure 5.2).

We left the proof of Theorem 5.2.5 in a separate section, after demonstrating its use because the proof is tedious and involved. Hence, it is best to know "what we are fighting for" to keep some motivation.

The second section of this chapter is devoted to proving that $\mathsf{Fin}(\sigma)$ is actually localisable. We actually obtain a stronger result: every hereditary class of finite structures closed under disjoint unions is localisable.

> **Theorem 5.2.5.** *Let $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ be a hereditary class of finite structures closed under disjoint unions. Then, $\mathcal{C}$ is localisable.*

5.1. Leveraging Locality

Because we assume Theorem 5.2.5, it suffices in order to prove Theorem 5.1.2 to show that every existential local sentence φ that is preserved under extensions is equivalent to an existential sentence.

Let us now explain why the hypotheses of Theorem 5.1.2 appear in the first place. Because existential local sentences talk about local neighbourhoods, and $\mathsf{Local}(\mathcal{C}, r, k)$ explicitly appears in the statement of the theorem, it is very convenient to notice the following fact: if $\mathcal{C}$ is hereditary, then $\mathsf{Local}(\mathcal{C}, r, k) \subseteq \mathcal{C}$ for all $r, k \in \mathbb{N}$.[2]

2: Beware that even though $\mathcal{C}$ is hereditary, the localised classes $\mathsf{Local}(\mathcal{C}, r, k)$ are not hereditary in general. A simple way to observe this fact is that an element in $\mathsf{Local}(\mathcal{C}, r, 1)$ has a single connected component in its Gaifman graph, a property that is in general not stable when removing vertices.

> **Lemma 5.1.1.** *Let $\mathcal{C}$ be a hereditary class of finite structures, the*

> *following properties are equivalent:*
>
> 1. *The Łoś-Tarski Theorem relativises to* $\mathsf{Local}(\mathcal{C}, r, k)$ *for all* $r, k \geq 0$.
> 2. *Existential local sentences preserved under extensions over* $\mathcal{C}$ *are equivalent (over* $\mathcal{C}$*) to existential sentences.*

Proof. Let us first prove that Item 1 implies Item 2. Let $\varphi = \exists \vec{x}.\psi(\vec{x})$ be an existential local sentence preserved under extensions over $\mathcal{C}$. Let us write $k \stackrel{\text{def}}{=} |\vec{x}|$, and let r be the locality radius of ψ.

Because $\mathsf{Local}(\mathcal{C}, r, k) \subseteq \mathcal{C}$, the sentence φ is also preserved under extensions over $\mathsf{Local}(\mathcal{C}, r, k)$. Hence, by Item 1, there exists $\theta \in \mathsf{EFO}$ such that $\varphi \equiv_{\mathsf{Local}(\mathcal{C},r,k)} \theta$. Because, $\mathcal{C}$ is hereditary, Lemma 3.2.3 provides a finite set of structures $M \in \mathcal{P}_{\mathsf{fin}}(\mathcal{C})$ such that $[\![\theta]\!]_{\mathcal{C}} = \uparrow_{\subseteq_i} M$.

Let us define $n_0 \stackrel{\text{def}}{=} \max \{ |\mathfrak{A}| : \mathfrak{A} \in M \}$. We are going to prove that $\subseteq_i$-minimal models of φ are of size bounded by $n_0 + k + r \times n_0$. Thanks to Lemma 3.2.3, this will provide a sentence $\theta' \in \mathsf{EFO}$ such that $\varphi \equiv_{\mathcal{C}} \theta'$ and will allow us to conclude.

To that end, let us consider a $\subseteq_i$-minimal model $\mathfrak{B}$ of φ in $\mathcal{C}$. Since $\mathfrak{B} \models \varphi$, there exists a valuation $\nu \colon \vec{x} \to \mathfrak{B}$, such that $\mathfrak{B}, \nu \models \psi(\vec{x})$. Because ψ is an r-local formula, this implies that $\mathcal{N}_{\mathfrak{B}}(\nu(\vec{x}), r), \nu \models \psi(\vec{x})$, hence that $\mathcal{N}_{\mathfrak{B}}(\nu(\vec{x}), r) \models \varphi$. Because $\mathcal{N}_{\mathfrak{B}}(\nu(\vec{x}), r) \subseteq_i \mathfrak{B}$ and the latter is an $\subseteq_i$-minimal model of φ, we conclude that $\mathfrak{B} = \mathcal{N}_{\mathfrak{B}}(\nu(\vec{x}), r)$, i.e., that $\mathfrak{B} \in \mathsf{Local}(\mathcal{C}, r, k)$.

As $\varphi \equiv_{\mathsf{Local}(\mathcal{C},r,k)} \theta$, we conclude that $\mathfrak{B} \models \theta$. Since $[\![\theta]\!]_{\mathcal{C}} = \uparrow_{\subseteq_i} M$, there exists $\mathfrak{A} \in M$ such that $\mathfrak{A} \subseteq_i \mathfrak{B}$. Let us write $h \colon \mathfrak{A} \to_{\mathsf{QF}} \mathfrak{B}$ for the QF-embedding witnessing that $\mathfrak{A} \subseteq_i \mathfrak{B}$.

We will now define an induced substructure $\mathfrak{B}' \subseteq_i \mathfrak{B}$, by restricting the domain of $\mathfrak{B}$ to the union of the following sets: $h(\mathfrak{A})$ (of size at most n_0), $\nu(\vec{x})$ (of size at most k), and (just) enough points so that every element of $h(\mathfrak{A})$ is at distance at most r of $\nu(\vec{x})$ inside $\mathfrak{B}'$. The latter is possible because $\mathfrak{B} = \mathcal{N}_{\mathfrak{B}}(\nu(\vec{x}), r)$, hence one can add to $\mathfrak{B}'$ a path of at most r vertices connecting every element in $h(\mathfrak{A})$ to some element of $\nu(\vec{x})$. As a consequence, one can assume that $|\mathfrak{B}'| \leq n_0 + k + r \times n_0$. We refer the reader to Figure 5.1 for a graphical representation of this construction.

Notice that by definition, $\mathfrak{B}'$ satisfies the following properties:

1. $\mathfrak{B}' \subseteq_i \mathfrak{B}$,
2. $\mathfrak{B}' = \mathcal{N}_{\mathfrak{B}'}(\nu(\vec{x}), r)$, hence $\mathfrak{B}' \in \mathsf{Local}(\mathcal{C}, r, k)$,
3. $h \colon \mathfrak{A} \to_{\mathsf{QF}} \mathfrak{B}'$ is a QF-embedding, and $\mathfrak{A} \models \varphi$, hence $\mathfrak{B}' \models \varphi$.

Because $\mathfrak{B}' \subseteq_i \mathfrak{B}$ and $\mathfrak{B}$ is a $\subseteq_i$-minimal model of φ, the above properties entail $\mathfrak{B}' = \mathfrak{B}$. In particular, we have proven that $|\mathfrak{B}| \leq n_0 + k + r \times n_0$.

Let us now turn our attention to the converse implication from Item 2 to Item 1. Let φ be a sentence preserved under extensions over the class $\mathsf{Local}(\mathcal{C}, r, k)$. Let us recall that $|\varphi|^r_{\vec{x}}$ is the r-local formula obtained by relativising the quantifications in φ to the r-neighbourhood of $\vec{x}$, that are fresh variables here.

We are going to prove that the sentence $\psi \stackrel{\text{def}}{=} \exists\vec{x}.|\varphi|_{\vec{x}}^{r}$ is preserved under extensions over $\mathcal{C}$. Let $\mathfrak{A}, \mathfrak{B} \in \mathcal{C}$ such that $\mathfrak{A} \models \psi$ and $\mathfrak{A} \subseteq_i \mathfrak{B}$. By definition, there exists a map $h \colon \mathfrak{A} \to_{\mathsf{QF}} \mathfrak{B}$ that is a QF-embedding. Since $\mathfrak{A} \models \psi$, there exists $\vec{a} \in \mathfrak{A}$ such that $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r) \models \varphi$. Notice that $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r) \subseteq_i \mathcal{N}_{\mathfrak{B}}(h(\vec{b}), r)$ (via the same map h). Because φ is preserved under extensions over $\mathsf{Local}(\mathcal{C}, r, k)$, we conclude that $\mathcal{N}_{\mathfrak{B}}(h(\vec{b}), r) \models \varphi$. By construction, this means that $\mathfrak{B} \models \psi$.

Thanks to the hypothesis (Item 2), ψ is equivalent to an existential sentence θ over $\mathcal{C}$. Remark that φ is equivalent to ψ over $\mathsf{Local}(\mathcal{C}, r, k)$. As a consequence, φ is equivalent to the existential sentence θ over $\mathsf{Local}(\mathcal{C}, r, k)$. ■

We are now ready to prove Theorem 5.1.2, admitting for now that Theorem 5.2.5 (recalled below) is proven.

Theorem 5.2.5. *Let $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ be a hereditary class of finite structures closed under disjoint unions. Then, $\mathcal{C}$ is localisable.*

Theorem 5.1.2. *Let $\mathcal{C}$ be a hereditary class of finite structures closed under disjoint unions. The Łoś-Tarski Theorem relativises to $\mathcal{C}$ if and only if the Łoś-Tarski Theorem relativises to $\mathsf{Local}(\mathcal{C}, r, k)$ for all $r \geq 0$ and $k \geq 1$.*

Proof. Assume that the Łoś-Tarski Theorem relativises $\mathsf{Local}(\mathcal{C}, r, k)$ for $r, k \geq 0$. Let φ be a sentence preserved under extensions over $\mathcal{C}$. Because $\mathcal{C}$ is hereditary and closed under disjoint unions, we can apply Theorem 5.2.5, and $\mathcal{C}$ is localisable. Hence, φ is equivalent to an existential local sentence ψ over $\mathcal{C}$. Thanks to Lemma 5.1.1, ψ is equivalent to an existential sentence, and we have concluded.

Conversely, assume that preservation under extensions relativises to $\mathcal{C}$. In particular, existential local sentences preserved under extensions are equivalent over $\mathcal{C}$ to existential sentences. Thanks to Lemma 5.1.1, this proves that $\mathsf{Local}(\mathcal{C}, r, k)$ satisfies preservation under extensions for $r, k \geq 0$. ■

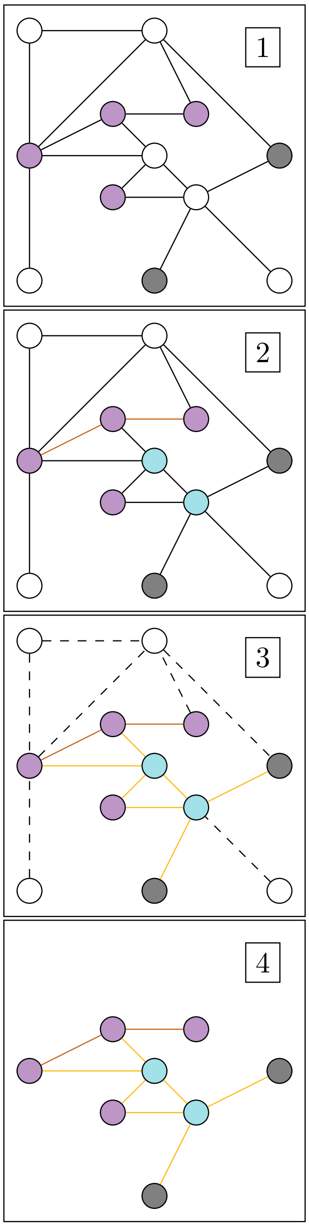


Figure 5.1.: Graphical representation of the induced substructure extraction done in Lemma 5.1.1. The first frame represents a finite structure with a violet induced substructure, and two selected centres. The second frame describes how to add intermediary points to ensure that there exists a path from the centres to the induced substructure. The third and fourth frame exhibit the extraction of the "enriched" induced substructure.

The Power of Locality. The rest of this section is an illustration of the power of Theorem 5.1.2. This demonstration is done by first showing how previously known results are (easy) consequences of our theorem, and how they can be (easily) generalised to construct new classes of finite structures where the Łoś-Tarski Theorem relativises. Finally, we also will discuss the two hypotheses of Theorem 5.1.2, and prove that both are necessary for the theorem to hold.

As a first hint that most of the heavy-lifting has already been done, let us point out that the spaces $\mathsf{Local}(\mathcal{C}, r, k)$ already appear (although implicitly) in previous studies of preservation theorems. This notion of locality is itself well-studied inside the realm of so-called *sparse* structures [78]. To be precise, let us focus on one specific notion of sparsity: that of wide classes of structures.

Definition 5.1.3 [6, Definition 4.1]**.** A class $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ is *wide* whenever there exists a map $\rho\colon \mathbb{N}^2 \to \mathbb{N}$, such that for all $r, m \in \mathbb{N}^2$, for all $\mathfrak{A} \in \mathcal{C}$ of size greater than $\rho(r,m)$, there exists $\vec{a} \in \mathfrak{A}^m$ such that $\mathfrak{A}, \vec{x} \mapsto \vec{a} \models \mathrm{indep}^{\mathcal{N}}(\vec{a},r)$.

Alternatively, one can think of a wide class $\mathcal{C}$ as one where the sentences of the form $\exists \vec{x}.\mathrm{indep}^{\mathcal{N}}(\vec{x},r)$ define co-finite subsets of $\mathcal{C}$. Borrowing the vocabulary of Section 4.2, the basic local sentences $\exists^{\geq r}_{k} x.\top$ define co-finite sets.[3]

3: In particular, for every r-local formula ψ, $\neg(\exists^{\geq k}_{r} x.\psi(x)) \Rightarrow \exists^{\geq k}_{r} x.\neg\psi(x)$, for all but finitely many elements in a wide class $\mathcal{C}$.

Leveraging this notion of wideness, Atserias, Dawar and Grohe prove that the Łoś-Tarski Theorem relativises to wide, hereditary, classes $\mathcal{C}$ that are closed under disjoint unions [6, Theorem 4.3].

We argue that this is a very simple consequence of Theorem 5.1.2, because hereditary wide classes are exactly those that are locally finite.

Exercise 5.1.4 ([77, Theorem 5.1])**.** For a hereditary class $\mathcal{C}$ the following properties are equivalent:

- $\mathcal{C}$ is wide,
- $\mathsf{Local}(\mathcal{C}, r, k)$ is finite for all $r, k \geq 1$,
- $\mathcal{C}$ has bounded degree.

Recall that we introduced the notion of degree of an element in a relational structure as its degree in the Gaifman graph of said structure. Equivalently, the degree of a point $a \in \mathfrak{A}$ is precisely $|\mathcal{N}_{\mathfrak{A}}(a,1)|$. This may provide some intuition on the relationship between the last two items of Exercise 5.1.4.

Over a finite set of finite models, every sentence φ is equivalent to an existential sentence as observed in Fact 3.2.2, hence the Łoś-Tarski Theorem relativises trivially. Leveraging Exercise 5.1.4, if $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ is a class that is hereditary, wide, and closed under disjoint unions, then for all $r, k \in \mathbb{N}$, the Łoś-Tarski Theorem relativises (trivially) to $\mathsf{Local}(\mathcal{C}, r, k)$. We recover Atserias, Dawar and Grohe's result by applying Theorem 5.1.2.

We can generalise this proof scheme as follows. Given a property $P \in \mathcal{P}(\mathsf{Fin}(\sigma))$, we say that a class $\mathcal{C}$ *locally* satisfies P when, for all $r, k \in \mathbb{N}$, $\mathsf{Local}(\mathcal{C}, r, k) \in P$

Theorem 5.1.5. *Let $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ be a hereditary class of finite structures closed under disjoint unions. The following properties imply that the Łoś-Tarski Theorem relativises to $\mathcal{C}$.*

1. *$\mathcal{C}$ is locally finite (i.e., wide).*
2. *$\mathcal{C}$ has locally bounded tree-depth.*
3. *$\mathcal{C}$ is locally wqo with respect to $\subseteq_i$.*
4. *The Łoś-Tarski Theorem locally relativises to $\mathcal{C}$.*

Moreover, Item 1 $\Rightarrow$ Item 2 $\Rightarrow$ Item 3 $\Rightarrow$ Item 4, and these implications are strict.

The remaining of this section is devoted to proving Theorem 5.1.5, together with its relation with previously known properties as illustrated in Figure 5.2. In particular, the examples prove that we strictly generalise previously known properties implying preservation under extensions. As no logic is involved in the generation of these new classes, we

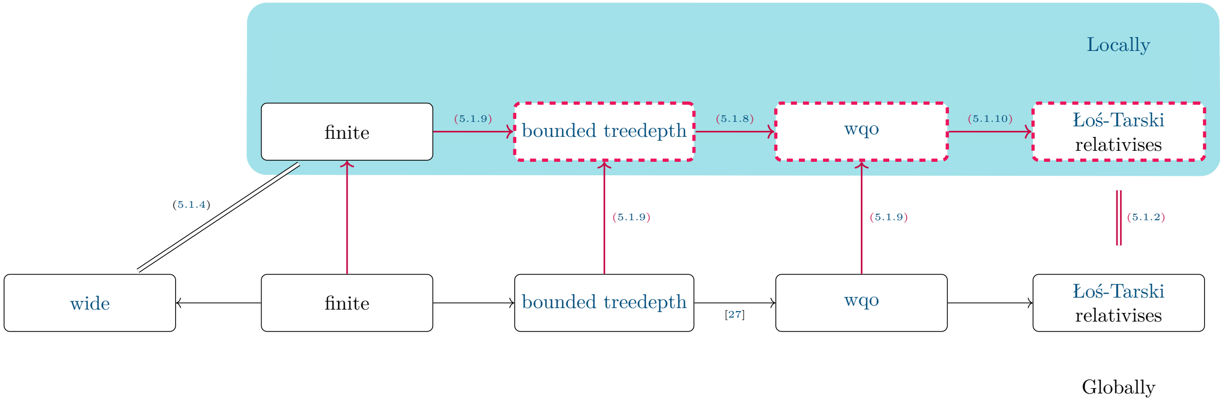


Figure 5.2.: Implications of properties over classes $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ that are hereditary and closed under disjoint unions. Arrows represent strict inclusions of properties and dashed boxes are the new properties introduced in this paper.

effectively *decoupled* our proofs of preservation theorems in a *locality argument* followed by a *combinatorial argument.*

Note that working with $\mathsf{Local}(\mathcal{C}, r, k)$ rather than $\mathsf{Local}(\mathcal{C}, r, 1)$ is a somewhat uncommon way to localise properties. Thankfully, for properties that are well-behaved with respect to disjoint unions,[4] the localisation using neighbourhoods centred around a single point or several ones will coincide; examples of such properties include wideness, exclusion of a minor, or bounded clique-width. The following proposition illustrates this point in the case of bounded tree-depth.

4: Namely, if the property P is satisfied for all classes $\mathsf{Local}(\mathcal{C}, r, 1)$, where $r \in \mathbb{N}$, then it is satisfied for all classes $\mathsf{Local}(\mathcal{C}, r, k)$, where $r, k \in \mathbb{N}$.

Recall here that the tree-depth $\mathrm{td}(G)$ is defined inductively as described in Definition 2.3.20.

Lemma 5.1.6. *A class $\mathcal{C} \subseteq \mathsf{Struct}(\sigma)$ has locally bounded tree-depth if and only if $\exists \rho \colon \mathbb{N} \to \mathbb{N}, \forall \mathfrak{A} \in \mathcal{C}, \forall a \in \mathfrak{A}, \forall r \geq 1, \mathrm{td}(\mathcal{N}_{\mathfrak{A}}(a, r)) \leq \rho(r)$. That is, if and only if $\mathsf{Local}(\mathcal{C}, r, 1)$ has bounded tree-depth for all $r \geq 1$.*

Proof. Assuming that for all k there is a bound on the tree-depth of elements of $\mathsf{Local}(\mathcal{C}, r, k)$, we define $\rho(r)$ to be the maximum of $\mathrm{td}(\mathfrak{A})$ for $\mathfrak{A}$ in $\mathsf{Local}(\mathcal{C}, r, 1)$.

Conversely, assume that there exists an increasing function ρ such that $\mathrm{td}(\mathfrak{A}) \leq \rho(r)$ for every $\mathfrak{A} \in \mathsf{Local}(\mathcal{C}, r, 1)$; a simple induction on A shows that $\mathrm{td}(\mathfrak{A}) \leq \rho(r \times (2k+1))$ for every $\mathfrak{A} \in \mathsf{Local}(\mathcal{C}, r, k)$:

- ▶ If $\mathfrak{A} \in \mathsf{Local}(\mathcal{C}, r, k)$ can be written as $\mathfrak{A}_1 \uplus \mathfrak{A}_2$, then $\mathrm{td}(\mathfrak{A})$ is exactly $\max(\mathrm{td}(\mathfrak{A}_1), \mathrm{td}(\mathfrak{A}_2))$, and we conclude by induction hypothesis since ρ is increasing.
- ▶ If $\mathfrak{A}$ is totally connected and in $\mathsf{Local}(\mathcal{C}, r, k)$, then it is included in a ball of radius $r \times (2k+1)$ hence the tree-depth is bounded by $\rho(r \times (2k+1))$. ∎

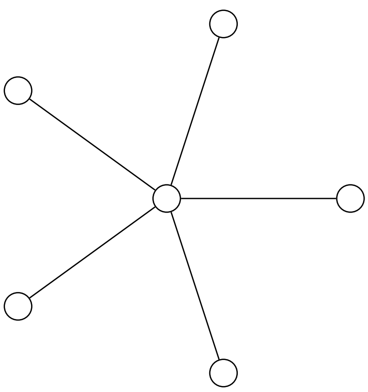


Figure 5.3.: An example of a star in Stars.

Let Stars be the class of finite stars, i.e., connected undirected graphs with at most one node of degree greater than 1. See Figure 5.3 for a graphical representation.

Example 5.1.7. The class $\downarrow_{\subseteq_i} \uplus\mathsf{Stars}$ has bounded tree-depth, locally bounded tree-depth, but is neither finite nor locally finite.

Proof. It is clear that a star has tree-depth at most 2. Furthermore, the tree-depth is monotone with respect to induced substructures, and a morphism with respect to disjoint unions, i.e., $\mathrm{td}(\mathfrak{A} \uplus \mathfrak{B}) = \max(\mathrm{td}(\mathfrak{A}), \mathrm{td}(\mathfrak{B}))$. As a consequence, $\mathrm{td}(\downarrow_{\subseteq_i} \uplus\mathsf{Stars}) \leq 2$. For the same reasons, the class has locally bounded tree-depth, because of the inclusion $\mathsf{Local}(\mathcal{C}, r, k) \subseteq \mathcal{C}$. However, the class is not locally finite, because it has unbounded degree. And it is clearly not finite. ■

Example 5.1.8. Consider the class $\uplus\mathsf{Cliques}$ of finite disjoint unions of undirected cliques. This class is locally wqo, but is not of locally bounded tree-depth.

Proof. Notice that $\mathsf{Local}(\uplus\mathsf{Cliques}, r, k) \subseteq \uplus\mathsf{Cliques}$ which is a wqo for $\subseteq_i$, as it is isomorphic to $\mathsf{M}^\diamond(\mathbb{N})$ with the multiset ordering, which is a wqo thanks to Table 6.1. Now, it is quite clear that for all $n, r, k \geq 1$, $\mathrm{td}(K_n) = n$, and $K_n \in \mathsf{Local}(\uplus\mathsf{Cliques}, r, k)$. As a consequence, $\uplus\mathsf{Cliques}$ is not locally of bounded tree-depth. ■

Let $n \in \mathbb{N}$. A diamond D_n is a cycle C_n extended with two new points a, b that are connected to every node in the cycle via a path of length n. We call $\mathcal{D}$ the class of all D_n for $n \geq 3$. We provide in Figure 5.4 a graphical representation of D_4.

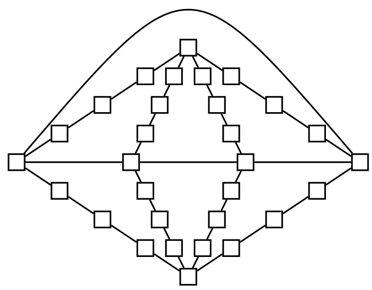


Figure 5.4.: The diamond $D_4 \in \mathcal{D}$.

Example 5.1.9. Consider the class $\downarrow_{\subseteq_i} \uplus\mathcal{D}$ of induced substructures of disjoint unions of diamonds. This class has locally bounded tree-depth, but is neither wqo nor locally finite.

Proof. The class is not wqo because $(D_n)_{n>4}$ is an infinite bad sequence. Assume by contradiction that $D_n \subseteq_i D_m$ for $4 < n < m$. Then, the two outer nodes a, b of D_n are of degree greater than 4, while every other node in the graph are of degree at most 4. Therefore, the two extra outer nodes of D_n must be mapped to the two extra outer nodes of D_m. Similarly, the element C_n over which D_n has all its nodes of degree exactly 4, while the rest of the nodes have degree either greater than 4, or exactly 2. Hence, the inner cycle C_n must be mapped to the cycle C_m, which is not possible when $n < m$.

However, the class is locally of bounded tree-depth, because we add paths of increasing length to separate the cycle from the extra outer nodes. Informally, for every fixed r, $\mathcal{N}_{D_n}(a, r)$ is either a star, or a tree with radius at most r, when n is large enough. In both cases, because the radius is fixed, we obtain a bound on the tree-depth.

Furthermore, it should be clear that the class is not locally finite, as it contains Stars, which is not. ■

Let $n \in \mathbb{N}$, and σ be a finite relational signature. We define Δ^n_{deg} as the class of finite structures of degree bounded by n.

To introduce Example 5.1.10, which separates classes that are locally wqo and classes where the Łoś-Tarski Theorem locally relativises, the simplest way is to use the notions of first order interpretations, logical maps, and the closure properties of logically presented pre-spectral spaces (lpps). These notions are formally introduced in Chapter 6 (Logically Presented Spaces), and we think that it is a good time to illustrate how these abstract notions are actually helpful in concrete cases.

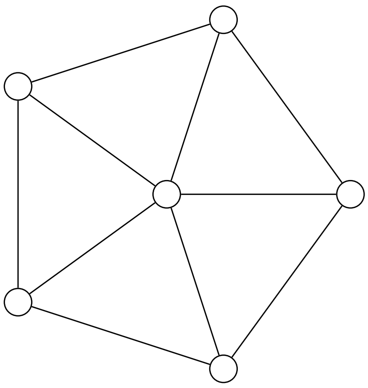


Figure 5.5.: An example of a "wheel" used in Example 5.1.10.

Example 5.1.10. Let $\sigma \stackrel{\text{def}}{=} \{ (E, 2); (c, 1) \}$, and let I be a first order interpretation of $\mathsf{Fin}(\sigma)$ in $\mathsf{Fin}(\{ (E, 2) \})$ defined by $\varphi_E(x, y) \stackrel{\text{def}}{=} c(x) \vee c(y) \vee E(x, y)$. The Łoś-Tarski Theorem locally relativises to $\mathsf{I}(\Delta^2_{\mathrm{deg}})$, but the latter is not locally wqo.

Proof. The class $\mathsf{I}(\Delta^2_{\mathrm{deg}})$ is not locally well-quasi-ordered because it contains the class of "wheels": elements of Cycles with an extra point connected to all the nodes (see Figure 5.5). Similarly to Cycles, the class of wheels is an infinite antichain. Furthermore, every wheel is the 1-local neighbourhood of one of its points, that is, every wheel belongs to $\mathsf{Local}(\mathsf{I}(\Delta^2_{\mathrm{deg}}), 1, 1)$. As a consequence, $\mathsf{I}(\Delta^2_{\mathrm{deg}})$ is not locally wqo.

To avoid using the machinery of logically presented pre-spectral spaces, one could apply the following reasoning. Let φ be a first order sentence that is preserved under extensions over $\mathsf{I}(\Delta^2_{\mathrm{deg}})$. Because I is a (simple) first order interpretation, there exists a first-order sentence $\psi \in \mathsf{FO}[\sigma]$ such that for all $\mathfrak{A} \in \mathsf{Fin}(\sigma)$, $\mathfrak{A} \models \psi$ if and only if $\mathsf{I}(\mathfrak{A}) \models \varphi$. It is an easy check that ψ is preserved under extensions over Δ^2_{deg}, since $\mathfrak{A} \models \psi$ and $\mathfrak{A} \subseteq_i \mathfrak{B}$ implies that $\mathsf{I}(\mathfrak{A}) \models \varphi$ and $\mathsf{I}(\mathfrak{A}) \subseteq_i \mathsf{I}(\mathfrak{B})$. We know that the Łoś-Tarski theorem relativises to Δ^2_{deg}, hence that there exists an existential sentence θ such that $[\![\theta]\!]_{\Delta^2_{\mathrm{deg}}} = [\![\psi]\!]_{\Delta^2_{\mathrm{deg}}}$. However, Δ^2_{deg} is a downwards closed subset of $\mathsf{Fin}(\sigma)$, and as a consequence, $[\![\theta]\!]_{\Delta^2_{\mathrm{deg}}}$ is the upward closure (in Δ^2_{deg}) of a *finite set* of non-equivalent $\subseteq_i$-minimal models. We conclude that $\mathsf{I}([\![\theta]\!]_{\Delta^2_{\mathrm{deg}}})$ also has finitely many non-equivalent $\subseteq_i$-minimal models, hence is equivalent (over $\mathsf{I}(\Delta^2_{\mathrm{deg}})$) to the finite union of their respective diagram formulas.

We argue that the proof is easier to understand (and reproduce or adapt) using the theory of lpps. We start by noticing that Δ^2_{deg} is hereditary, and leverage Lemma 6.1.16 to conclude that $\langle\!\langle \Delta^2_{\mathrm{deg}}, \langle \mathsf{EFO} \rangle_{\mathsf{topo}}, \mathsf{FO} \rangle\!\rangle$ is a logically presented pre-spectral space. Now, the map I is a logical map, hence we conclude that $\langle\!\langle \mathsf{I}(\Delta^2_{\mathrm{deg}}), \langle \mathsf{EFO} \rangle_{\mathsf{topo}}, \mathsf{FO} \rangle\!\rangle$ is a lpps via Lemma 6.3.3. Because the latter is also a hereditary class, we conclude (again using Lemma 6.1.16) that the Łoś-Tarski Theorem relativises to $\mathsf{I}(\Delta^2_{\mathrm{deg}})$. Thanks to Theorem 5.1.2, this is the same as stating that the Łoś-Tarski Theorem locally relativises to $\mathsf{I}(\Delta^2_{\mathrm{deg}})$. ∎

We can now state formally the proof of Theorem 5.1.5 by gathering all the above examples.

Proof of Theorem 5.1.5. The implications are consequences of the Theorem 5.1.2. The implications are strict thanks to Examples 5.1.7, 5.1.8 and 5.1.10. ∎

Furthermore, notice that, as explained in Figure 5.2, locally wqo strictly generalises wqo (see Example 5.1.9), which strictly generalises locally bounded treedepth (see Example 5.1.8). Notice that locally bounded treedepth already strictly generalises wqo via Example 5.1.9.

To conclude our study, let us now provide examples that go outside the realm of hereditary classes of structures closed under disjoint unions. This will illustrate how both hypotheses of Theorem 5.1.2 are necessary.

> **Example 5.1.11.** Let $\mathcal{C} \stackrel{\text{def}}{=} \downarrow_{\subseteq_i} \mathsf{Cycles}$. Then, $\mathcal{C}$ is hereditary, the Łoś-Tarski Theorem locally relativises to $\mathcal{C}$, $\mathcal{C}$ is not closed under disjoint unions, and the Łoś-Tarski Theorem does not relativise to $\mathcal{C}$.

Proof. We already know from Example 3.2.1 that the Łoś-Tarski Theorem does not relativise to $\mathcal{C}$, let us prove that it locally does.

Let $r, k \geq 0$ and consider $\mathsf{Local}(\mathcal{C}, r, k)$. Using Example 4.1.2, there exists $N(r,k) \in \mathbb{N}$ such that $\mathsf{Local}(\mathcal{C}, r, k) \subseteq \mathsf{Paths} \cup \{ C_n : 3 \leq n \leq N \}$, the latter being well-quasi-ordered by $\subseteq_i$. In particular, Lemma 3.3.5 shows that the Łoś-Tarski Theorem relativises to $\mathsf{Local}(\mathcal{C}, r, k)$. ∎

> **Example 5.1.12.** Let $\mathcal{C}$ be the class of finite disjoint unions of black and white coloured paths with black endpoints. Then, $\mathcal{C}$ is closed under disjoint unions, the Łoś-Tarski Theorem locally relativises to $\mathcal{C}$, $\mathcal{C}$ is not hereditary, and the Łoś-Tarski Theorem does not relativise to $\mathcal{C}$.

Proof. It is clear that $\mathcal{C}$ is closed under disjoint unions and not hereditary. Let us prove that the Łoś-Tarski Theorem does not relativise to $\mathcal{C}$.

The sentence φ defined in Example 4.3.11 stating that either the structure has two distinct connected components, or all the nodes of the (unique) path are black is a first order sentence. Furthermore, this sentence is preserved under $\subseteq_i$, because one cannot add a white coloured node to a structure without creating a new connected component.[5] The proof of Example 4.3.11 states that such a sentence cannot be rewritten as an existential local sentence, hence *a fortiori* cannot be rewritten as an existential sentence.

5: This is the purpose of the restrictions that endpoints of paths are black coloured vertices.

To conclude that the Łoś-Tarski Theorem locally relativises to $\mathcal{C}$, let us remark that $\mathsf{Local}(\mathcal{C}, r, k)$ is finite for all $r, k \geq 0$, and in particular is a well-quasi-ordering. We conclude similarly to Example 5.1.11. ∎

5.2. The Missing Square

Theorem 5.1.2 is a quite powerful and handy tool to study the relativisation of preservation theorems. However, its proof relied on Theorem 5.2.5, which remains to be proven. This is also filling the gap in Figure 4.13 that was left open at the end of Chapter 4 (Locality and Preservation).

Recall that a class $\mathcal{C}$ is localisable whenever sentences $\varphi \in \mathsf{FO}$ preserved under $\subseteq_i$ are equivalent to existential local sentences over $\mathcal{C}$ (see page 118).

Theorem 5.2.5. *Let $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ be a hereditary class of finite structures closed under disjoint unions. Then, $\mathcal{C}$ is localisable.*

As usual when proving such theorems, we will be interested in the $\subseteq_i$-minimal models of a sentence φ. To simplify[6] the study of $\min_{\subseteq_i}[\![\varphi]\!]_{\mathcal{C}}$, and as suggested in the introductory Chapter 3 (Preservation Theorems for First Order Queries), we restrict our attention to classes $\mathcal{C}$ that are downwards closed for $\subseteq_i$, also known as hereditary classes, so that $\min_{\subseteq_i}[\![\varphi]\!]_{\mathcal{C}} = \min_{\subseteq_i}[\![\varphi]\!]_{\mathsf{Fin}(\sigma)} \cap \mathcal{C}$.

6: Notice that this is not only a simplification, but also a necessary condition for Theorem 5.1.2 to hold, as proven in Example 5.1.11.

Lemma 5.2.1. *Let $\mathcal{C}$ be a hereditary class of finite structures, and $\varphi \in \mathsf{FO}$ be preserved under extensions over $\mathcal{C}$. The following are equivalent:*

1. *There exists $\psi \in \exists\mathsf{FOLoc}$ such that $\varphi \equiv_{\mathcal{C}} \psi$.*
2. *There exists $r, k \in \mathbb{N}$, such that $\min_{\subseteq_i}[\![\varphi]\!]_{\mathcal{C}} \subseteq \mathsf{Local}(\mathcal{C}, r, k)$.*

Proof. Let us first prove the direction $\Rightarrow$. If $\varphi \equiv_{\mathcal{C}} \exists\vec{x}.\tau(\vec{x})$ where $\tau(x)$ is an r-local formula (for some $r \geq 0$), then a $\subseteq_i$-minimal model $\mathfrak{A} \in \mathcal{C}$ of φ necessarily contains a vector $\vec{a}$ such that $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r), \vec{a} \models \tau$. This shows that $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r) \models \varphi$ where $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r) \subseteq_i \mathfrak{A}$ by definition. Since $\mathcal{C}$ is hereditary, $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r) \in \mathcal{C}$, thus $\mathfrak{A} = \mathcal{N}_{\mathfrak{A}}(\vec{a}, r)$ by minimality and $\mathfrak{A} \in \mathsf{Local}(\mathcal{C}, r, |\vec{x}|)$.

Let us now prove the direction $\Leftarrow$. Assume that the $\subseteq_i$-minimal models of φ are all found in $\mathsf{Local}(\mathcal{C}, r, k)$ for some $r, k \in \mathbb{N}$. Let q be the quantifier rank of φ. We are going to show that φ is preserved under $\Rrightarrow_{qr}^{r,k}$ over $\mathcal{C}$ and deduce by Proposition 4.3.9 that φ is equivalent to some existential local sentence over $\mathcal{C}$.

Let $\mathfrak{A} \models \varphi$ and $\mathfrak{A} \Rrightarrow_{qr}^{r,k} \mathfrak{B}$. Since $\mathfrak{A} \models \varphi$, there exists a $\subseteq_i$-minimal model $\mathfrak{A}_0 \in \mathsf{Local}(\mathcal{C}, r, k)$ with $\mathfrak{A}_0 \subseteq_i \mathfrak{A}$. Let $\vec{a} \in \mathfrak{A}^k$ be the centers of the balls of radius r in $\mathfrak{A}$ that contain $\mathfrak{A}_0$. Since $\mathfrak{A} \Rrightarrow_{qr}^{r,k} \mathfrak{B}$ there exists a vector $\vec{b} \in \mathfrak{B}^k$ such that $\mathsf{tp}_{\mathfrak{A}}^{qr}(\vec{a}, r) = \mathsf{tp}_{\mathfrak{B}}^{qr}\left(\vec{b}, r\right)$.

Notice that $\mathfrak{A}_0 \subseteq_i \mathcal{N}_{\mathfrak{A}}(\vec{a}, r)$, hence $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r) \models \varphi$ since φ is preserved under extensions. Thus,[7] $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r), \vec{x} \mapsto \vec{a} \models |\varphi|_r^{\vec{x}}$, the latter being an r-local formula, with k free variables, and of quantifier rank qr.[8] This shows that $\mathcal{N}_{\mathfrak{B}}(\vec{b}, r), \vec{x} \mapsto \vec{b} \models |\varphi|_r^{\vec{x}}$. To conclude, observe that this entails $\mathcal{N}_{\mathfrak{B}}(\vec{b}, r) \models \varphi$, hence $\mathfrak{B} \models \varphi$ since $\mathcal{N}_{\mathfrak{B}}(\vec{b}, r) \subseteq_i \mathfrak{B}$ and φ is preserved under extensions. ∎

7: Recall that by definition, $\mathfrak{A}, \vec{x} \mapsto \vec{a} \models |\varphi|_r^{\vec{x}}$ if and only if $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r), \vec{x} \mapsto \vec{a} \models \varphi$. Furthermore, notice that $\mathcal{N}_{\mathcal{N}_{\mathfrak{A}}(\vec{a},r)}(\vec{a}, r) = \mathcal{N}_{\mathfrak{A}}(\vec{a}, r)$.

8: The definition of $|\varphi|_r^{\vec{x}}$ requires that the free variables of φ belong to $\vec{x}$, which is the case here because φ does not have any.

We now provide the reader with alternative reading paths. The first one is to trust that the following Lemma 5.2.2 holds because one can extract it from the proof of a different result, namely the relativisation of the Łoś-Tarski Theorem to hereditary classes of finite structures that are closed under disjoint unions [6, Theorem 4.3]. If this is the case, you can read the next paragraph, and jump directly to Section 5.3 on page 133. If you do not trust that a one paragraph explanation is enough to convince you, you can still read the upcoming paragraph, but you will take combinatorial road once again for five to six pages.

The proof of preservation under extensions over some specific classes provided by [6, Theorem 4.3] is done by contradiction, using the fact that $\subseteq_i$-minimal models that are "large enough" must contain large "scattered sets of points." Forgetting about the size of the structure, this actually proves that $\subseteq_i$-minimal models are in some $\mathsf{Local}(\mathcal{C}, r, k)$ for well-chosen[9] parameters $r, k \in \mathbb{N}$.

9: We invite the interested reader to check in [6] how these parameters are actually chosen.

Lemma 5.2.2. Let $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ be a hereditary class of finite structures closed under disjoint unions and $\varphi \in \mathsf{FO}[\sigma]$ be a sentence preserved under extensions over $\mathcal{C}$.

There exist R, K such that $\min_{\subseteq_i} [\![\varphi]\!]_{\mathcal{C}} \subseteq \mathsf{Local}(\mathcal{C}, R, K)$.

Even though we stated Lemma 5.2.2 as a "foreign result", it cannot be found *as is* and must be (carefully) extracted from the proof of [6, Theorem 4.3]. For that reason, and for completeness, we provide hereafter a standalone proof of this lemma, that explicitly studies the distribution of types in the structures. Because the proof method differs from the original one, it is also interesting on its own.

Before getting into the technicalities of a formal proof, let us explain why the "naïve" approach does not work. The simplest approach would be to consider a formula φ in Gaifman normal form, and a $\subseteq_i$-minimal model $\mathfrak{A}$ of φ. The structure $\mathfrak{A}$ models a conjunction of (potential negations of) sentences of the form $\exists_r^{\geq k} x.\psi(x)$ where $\psi(x)$ is an r-local formula. Considering all the basic local sentences that appear positively in the conjunction, one can build a vector $\vec{a} \in \mathfrak{A}$ containing the witnesses of the outer existential quantifications. The main hope being that $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r) \models \varphi$, as this would imply, by minimality of $\mathfrak{A}$, that $\mathfrak{A} = \mathcal{N}_{\mathfrak{A}}(\vec{a}, r)$. By letting $R = r$ and $K = |\vec{a}|$, which is bounded independently of $\mathfrak{A}$, we would conclude.

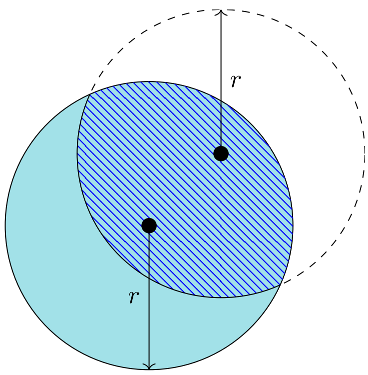


Figure 5.6.: Illustration that the local neighbourhoods of a local neighbourhood exhibit new behaviour.

Unfortunately the structure $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r)$ does not satisfy φ in general. The crucial issue comes from *intersections* of local neighbourhoods: there are new local neighbourhoods appearing in $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r)$, as depicted in Figure 5.6. In equational terms, the r-local neighbourhoods of $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r)$ are written $\mathcal{N}_{\mathcal{N}_{\mathfrak{A}}(\vec{a},r)}(\vec{b}, r) = \mathcal{N}_{\mathfrak{A}}(\vec{a}, r) \cap \mathcal{N}_{\mathfrak{A}}(b, r)$, i.e., as an intersection of two r-local neighbourhoods of $\mathfrak{A}$. This is problematic because φ contains basic local sentences that appear negatively, and the fact that $\mathfrak{A}$ does not contain some local behaviour does not transport to $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r)$.

To tackle this issue and represent the intersections inside our formulas, we temporarily leave the realm of first-order logic and consider MSO *local types*, written $\mathsf{mtp}^q_{\mathfrak{A}}(\vec{a}, r)$. There are finitely many MSO *local types* up to logical equivalence at a given quantifier rank and locality radius.

The definition of $\mathsf{MSO}[\sigma]$, its associated satisfaction relation, and the notion of r-local MSO formula are defined in Appendix E.2.

We update our *type-collector* function accordingly through

$$\mathsf{MTypes}_r^{q,k}(\mathfrak{A}) \stackrel{\mathsf{def}}{=} \{\, \mathsf{mtp}_{\mathfrak{A}}^q(\vec{a}, r) \colon \vec{a} \in \mathfrak{A}^k \,\}\,. \tag{5.1}$$

As for first-order local types, MSO local types are enough to characterise the natural preorder associated to *existential local* MSO-*sentences*, that is, sentences of the form $\exists \vec{x}.\theta(\vec{x})$, where $\theta(\vec{x})$ is an r-local MSO formula[10] around $\vec{x}$. Before going through the proof of Lemma 5.2.2 we translate the main properties of local types and existential local sentences to MSO-local types and existential local MSO-sentences. We can therefore write $\mathfrak{A} \Rrightarrow_{q,\mathsf{MSO}}^{r,k} \mathfrak{B}$ to mean $\mathsf{MTypes}_r^{q,k}(\mathfrak{A}) \subseteq \mathsf{MTypes}_r^{q,k}(\mathfrak{B})$.

10: See Definition E.2.3 p. 258

Exercise 5.2.3. For all structures $\mathfrak{A}, \mathfrak{B}$, $r, q, k \in \mathbb{N}$, the following are equivalent

- $\mathsf{MTypes}_r^{q,k}(\mathfrak{A}) \subseteq \mathsf{MTypes}_r^{q,k}(\mathfrak{B})$.
- For all $\theta(\vec{x})$ r-local MSO-formula, $\mathfrak{A} \models \exists \vec{x}.\theta(\vec{x})$ implies $\mathfrak{B} \models \exists \vec{x}.\theta(\vec{x})$.

Lemma 5.2.4. *Let $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ be a class of finite structures, and $\varphi \in \mathsf{FO}$. The following are equivalent*

1. *There exists an existential local MSO sentence ψ such that $\psi \equiv_{\mathcal{C}} \varphi$.*
2. *There exists $r, q, k \in \mathbb{N}$ such that φ is preserved under $\Rrightarrow_{q,\mathsf{MSO}}^{r,k}$.*

Proof Sketch. It is the same proof as Proposition 4.3.9 when replacing $\mathsf{Types}_r^{q,k}$ with $\mathsf{MTypes}_r^{q,k}$, leveraging Exercise 5.2.3. ∎

In order to prove Lemma 5.2.2, we are going to slightly change our proof scheme, and use Lemma 5.2.4. The main idea will be that using MSO-local types prevents the problems associated with intersections of local neighbourhoods, because these intersections can be captured[11] by MSO sentences. This will be the following Lemma 5.2.10 on page 131, that we preemptively state below.

11: It will not be possible to encode the real intersection, but find neighbourhoods whose intersection have the same first-order theory up to some quantifier rank q.

Lemma 5.2.10. *Let $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ be a hereditary class closed under disjoint unions, and $\varphi \in \mathsf{FO}[\sigma]$ a sentence preserved under extensions over $\mathcal{C}$.*

There exists R, Q, K such that φ is preserved under $\Rrightarrow_{Q,\mathsf{MSO}}^{R,K}$ over $\mathcal{C}$.

Lemma 5.2.10 allows us to prove that sentences preserved under extensions have their minimal models in $\mathsf{Local}(\mathcal{C}, r, k)$ for some $r, k \geq 0$ whenever $\mathcal{C}$ is hereditary and closed under disjoint unions.

Lemma 5.2.2. Let $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ be a hereditary class of finite structures closed under disjoint unions and $\varphi \in \mathsf{FO}[\sigma]$ be a sentence preserved under extensions over $\mathcal{C}$.

There exist R, K such that $\min_{\subseteq_i} [\![\varphi]\!]_{\mathcal{C}} \subseteq \mathsf{Local}(\mathcal{C}, R, K)$.

Proof. Applying Lemma 5.2.10 provides R, Q, K such that φ is preserved over $\mathcal{C}$ under (R, Q, K) MSO types. Using Lemma 5.2.4, φ is equivalent over $\mathcal{C}$ to a sentence $\psi = \exists \vec{x}.\theta(\vec{x})$ where $\theta(\vec{x})$ is an r-local MSO-formula.

Consider $\mathfrak{A}$ a $\subseteq_i$-minimal model of ψ in $\mathcal{C}$. As $\mathfrak{A} \models \psi$, there exists a vector $\vec{a} \in \mathfrak{A}^{|\vec{x}|}$ such that $\mathfrak{A}, \vec{x} \mapsto \vec{a} \models \theta(\vec{x})$. Because θ is an r-local MSO-formula, this proves that $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r), \vec{x} \mapsto \vec{a} \models \theta(\vec{x})$. As a consequence, $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r) \models \psi$. Moreover, $\mathcal{N}_{\mathfrak{A}}(\vec{a}, r) \in \mathcal{C}$ because the latter is hereditary. The minimality of $\mathfrak{A}$ proves that $\mathfrak{A} = \mathcal{N}_{\mathfrak{A}}(\vec{a}, r)$, thus, that $\mathfrak{A} \in \mathsf{Local}(\mathcal{C}, r, |\vec{x}|)$. ■

Theorem 5.2.5. *Let $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ be a hereditary class of finite structures closed under disjoint unions. Then, $\mathcal{C}$ is localisable.*

Proof. Consider a sentence φ preserved under local elementary embeddings over $\mathcal{C}$. Using Lemma 5.2.2 its minimal models are in some $\mathsf{Local}(\mathcal{C}, r, k)$, and using Lemma 5.2.1 this provides an equivalent existential local sentence over $\mathcal{C}$. ■

Corollary 5.2.6. *A sentence φ preserved under extensions over $\mathsf{Fin}(\sigma)$ is equivalent over $\mathsf{Fin}(\sigma)$ to an existential local sentence.*

We have postponed the combinatorial core of the proof of Theorem 5.2.5, and in the Lemma 5.2.10 on page 131, which remains to be proven. This combinatorial lemma can be decomposed into two steps. First, we will analyse the "distribution of MSO-local types" in a structure, not unlike what was done in Section 4.2 and in particular in Lemma 4.2.24 on page 85. Then, we leverage this analysis to gain a better understanding of what $\mathfrak{A} \Rrightarrow^{r,k}_{q,\mathsf{MSO}} \mathfrak{B}$ means, and in particular prove that first-order sentences preserved under extensions must exhibit some regular behaviour with respect to this preorder.

As in Lemma 4.2.24, we are interested in the points of a structure that satisfy a given r-local property, except this time we have MSO formulas. For simplicity, we will not search the points that satisfy a given r-local MSO formula, but a "full" description of the formulas satisfied up to some quantifier rank. To that end, let us formally define what it means for a vector to "realise" an MSO (q, r)-type.

Definition 5.2.7. Let $\mathfrak{A} \in \mathsf{Struct}(\sigma)$, $\mathbf{T}^{\mathsf{MSO}}$ be an MSO (q, r)-type with a single free variable x. The set of *realisations of the type* $\mathbf{T}^{\mathsf{MSO}}$ is the set of $a \in \mathfrak{A}$ such that $\mathsf{mtp}^q_{\mathfrak{A}}(a, r) = \mathbf{T}^{\mathsf{MSO}}$. We write this subset of $\mathfrak{A}$ as follows:

$$\mathsf{real}^{\mathfrak{A}}_{\mathbf{T}^{\mathsf{MSO}}} \stackrel{\text{def}}{=} \{ a \in \mathfrak{A} \colon \mathsf{mtp}^q_{\mathfrak{A}}(a, r) = \mathbf{T}^{\mathsf{MSO}} \} \quad .$$

Because we are interested in the spatial distribution of realisations of a given type, and in particular want to find a small number of points to represent $\mathsf{MTypes}^{q,k}_r(\mathfrak{A})$, let us define what it means for a $\mathbf{T}^{\mathsf{MSO}}$ to be "covered."

Definition 5.2.8. Let $\mathfrak{A} \in \mathsf{Struct}(\sigma)$, $\mathbf{T}^{\mathsf{MSO}}$ be an MSO (q,r)-type with a single free variable x, $R \in \mathbb{N}$, and S be a subset of $\mathfrak{A}$. The type $\mathbf{T}^{\mathsf{MSO}}$ is *R-covered* by S when $\mathcal{N}_{\mathfrak{A}}(\mathsf{real}^{\mathfrak{A}}_{\mathbf{T}^{\mathsf{MSO}}}, r) \subseteq \mathcal{N}_{\mathfrak{A}}(S, R)$.

[21]: Dawar, Grohe, Kreutzer and Schweikardt (2006), 'Approximation schemes for first-order definable optimisation problems'

The following lemma is a uniform version of the one given by [21, Lemma 8]. It can be thought as a generalisation of the technique from Lemma 4.2.8 on page 77 to describe the spatial distribution of points of interest in a given structure. The intuition being that one can find (uniformly bounded) subsets of generic points and covering points, so that every MSO-local type is either found in large quantities (hence generic) or is rare enough (hence can be covered). We hope that the picture Figure 5.7 will help understand the overall idea in a simple case.

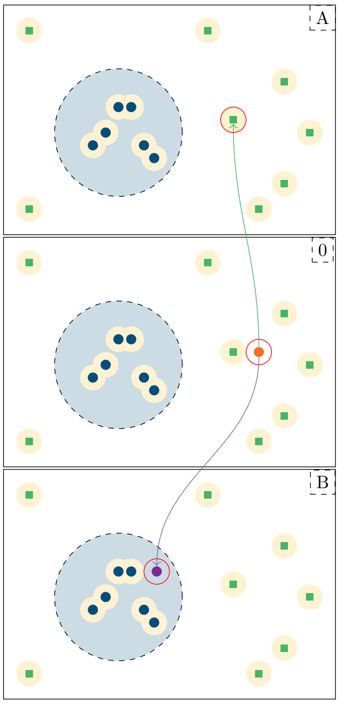


Figure 5.7.: Illustration of Lemma 5.2.9 with parameter $k = 1$. We have depicted the set of covering points in blue, with their r local neighbourhoods in beige. The R local neighbourhood of the covering points is in light blue. The collection of generic points is represented using the green squares, with their r-local neighbourhoods also in beige. The three pictures illustrate how, starting from frame 0 with a selected point (in orange), one can either find a corresponding generic point (in frame A), or a corresponding point that is R-covered by the covering points (in frame B).

Lemma 5.2.9. *Let σ be a finite relational signature, $r, q, k \geq 0$. Then, there exists a maximal size $K_m \geq k$ and maximal radius $R_m \geq r$ such that for all structures $\mathfrak{A} \in \mathsf{Struct}(\sigma)$ there exists $r \leq R \leq R_m$, a covering subset $\mathfrak{C}^{\mathfrak{A}} \subseteq \mathfrak{A}$ and a generic subset $\mathfrak{G}^{\mathfrak{A}} \subseteq \mathfrak{A}$ satisfying the following properties*

1. *Both $\mathfrak{C}^{\mathfrak{A}}$ and $\mathfrak{G}^{\mathfrak{A}}$ have size at most K_m.*
2. *The sets $\mathfrak{C}^{\mathfrak{A}}$ and $\mathfrak{G}^{\mathfrak{A}}$ are non-interferring, i.e.,*

$$\mathcal{N}_{\mathfrak{A}}(\mathfrak{C}^{\mathfrak{A}}, R) \cap \mathcal{N}_{\mathfrak{A}}(\mathfrak{G}^{\mathfrak{A}}, R) = \emptyset \quad .$$

3. *Elements in $\mathfrak{G}^{\mathfrak{A}}$ are independent, i.e.,*

$$\mathfrak{A}, \vec{x} \mapsto \mathfrak{G}^{\mathfrak{A}} \models \mathrm{indep}^{\mathcal{N}}(\vec{x}, R) \quad .$$

4. *For every $a \in \mathfrak{A}$, if $\mathsf{mtp}^{q}_{\mathfrak{A}}(a, r)$ is not R-covered by $\mathfrak{C}^{\mathfrak{A}}$, then $|\mathsf{real}^{\mathfrak{A}}_{\mathsf{mtp}^{q}_{\mathfrak{A}}(a,r)} \cap \mathfrak{G}^{\mathfrak{A}}| \geq k$.*

Proof. Let $r, q, k \geq 0$ be natural numbers and consider Q the number of different MSO (q,r)-local types with 1 free variable. Define $K_m \stackrel{\text{def}}{=} Q \times Q \times k$ and $R_m \stackrel{\text{def}}{=} 3^{Q+1} r$.

Let $\mathfrak{A} \in \mathsf{Struct}(\sigma)$ be a structure and consider $T \stackrel{\text{def}}{=} \mathsf{MTypes}^{q,1}_{r}(\mathfrak{A})$. By definition, $|T| \leq Q$. We construct iteratively for $i \leq |T|$ two sets $S_i \subseteq T$ and $C_i \subseteq A$ such that every type of S_i is $3^i r$-covered by C_i, $|C_i| \leq i \times Q \times k$, and examine whether the following property holds at each step i:

$$\begin{aligned} P(i) \stackrel{\text{def}}{=} \exists G \subseteq \mathfrak{A}, \forall u, v \in G, \mathrm{d}_{\mathfrak{A}}(u,v) > 2 \times 3^i r \\ \wedge \forall u \in G, \mathrm{d}_{\mathfrak{A}}(u, C_i) > 2 \times 3^i r \\ \wedge \forall \mathbf{T}^{\mathsf{MSO}} \in T \setminus S_i, |\mathsf{real}^{\mathfrak{A}}_{\mathbf{T}^{\mathsf{MSO}}} \cap G| \geq k \end{aligned}$$

The rationale is that if $P(i)$ holds for some $0 \leq i \leq Q$, it defines a subset $G_i \subseteq \mathfrak{A}$. By extracting only k witnesses per type in $T \setminus S_i$ from such a set G_i, we can construct $\mathfrak{G}^{\mathfrak{A}}$ of size at most $Q \times k \leq K_m$. Defining $\mathfrak{C}^{\mathfrak{A}} \stackrel{\text{def}}{=} C_i$, $R \stackrel{\text{def}}{=} 3^i r$, we have concluded, because the size of $\mathfrak{C}^{\mathfrak{A}}$ is bounded by $i \times Q \times k \leq K_m$.

We start with $S_0 \stackrel{\text{def}}{=} \emptyset, C_0 \stackrel{\text{def}}{=} \emptyset$. Assume that S_i and C_i have been defined so that $P(i)$ does not hold; we want to construct S_{i+1} and C_{i+1}. To that end, we enumerate the types in $T \setminus S_i$ in a sequence $(t_p)_{1 \leq p \leq |T \setminus S_i|}$. Using this sequence, we construct iteratively a set G_i^j of size at most $Q \times k$ such that points of G_i^j are at pairwise distance greater than $2 \times 3^i r$ and at distance greater than $2 \times 3^i r$ from C_i as follows. Let $G_i^0 \stackrel{\text{def}}{=} \emptyset$, and construct G_i^{j+1} by selecting the first type t_p that has fewer than k realisations in G_i^j; this is possible because $P(i)$ does not hold. If there exists a point $a \in \mathfrak{A}$ at distance greater than $2 \times 3^i r$ from G_i^j and C_i and of type t_p, we can add it to G_i^j to build G_i^{j+1}. Otherwise, every choice of point $a \in \mathfrak{A}$ of type t_p is at distance at most $2 \times 3^i r$ from $C_i \cup G_i^j$; in this case let $C_{i+1} \stackrel{\text{def}}{=} C_i \cup G_i^j$ and $S_{i+1} \stackrel{\text{def}}{=} S_i \cup \{t_p\}$, the hypothesis on S_i is that every type in S_i is $3^i r$-covered by C_i, and we showed that t_p was $3^{i+1} r$-covered by $C_i \cup G_i^j = C_{i+1}$.

The above process must terminate because the set S_i is strictly increasing and of size bounded by Q. By definition, this means that $P(i)$ holds for some $i \leq Q$, and we have concluded. ■

Following the terminology used by Dawar, Grohe, Kreutzer and Schweikardt, the MSO local types realised by $\mathfrak{C}^{\mathfrak{A}}$ can be thought of as *rare*, while those realised by $\mathfrak{G}^{\mathfrak{A}}$ can be thought of as *frequent*.

Lemma 5.2.10. *Let $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ be a hereditary class closed under disjoint unions, and $\varphi \in \mathsf{FO}[\sigma]$ a sentence preserved under extensions over $\mathcal{C}$.*

There exists R, Q, K such that φ is preserved under $\Rrightarrow_{Q,\mathsf{MSO}}^{R,K}$ over $\mathcal{C}$.

Proof. Consider a first order sentence φ that is preserved under $\subseteq_i$ over $\mathcal{C}$. As a first step, we write φ in Gaifman normal form and collect $\theta_1, \dots, \theta_\ell$ the basic local sentences appearing in this normal form. The sentences θ_i are of the form $\exists_{r_i}^{\geq k_i} x.\psi_i(x)$ where ψ_i is an r_i-local formula of quantifier rank q_i.

Let $r \stackrel{\text{def}}{=} \max\{ r_i : 1 \leq i \leq \ell \}$, $q \stackrel{\text{def}}{=} \max\{ q_i : 1 \leq i \leq \ell \}$, and $k \stackrel{\text{def}}{=} \max\{ k_i : 1 \leq i \leq \ell \}$.

We use Lemma 5.2.9 over the tuple $(2r, 2 \times k \times \ell, q+1)$ to obtain numbers $2r \leq R_m$ and $k \leq K_m$. Define $K \stackrel{\text{def}}{=} 2K_m$, $R \stackrel{\text{def}}{=} 2R_m$, and $Q \stackrel{\text{def}}{=} 2R_m + k + q + 1 + \max\{ \mathrm{rk}(\theta_i) : 1 \leq i \leq \ell \}$. Our goal is to prove that φ is preserved under $\Rrightarrow_{Q,\mathsf{MSO}}^{R,K}$.

Let $\mathfrak{A} \in \mathsf{Fin}(\sigma)$ such that $\mathfrak{A} \models \varphi$. Leveraging Lemma 5.2.9, we know that there exists subsets $\mathfrak{C}^{\mathfrak{A}}$ and $\mathfrak{G}^{\mathfrak{A}}$, respectively the covering subset and the generic subset for the MSO $(2r, q+1)$-local types of $\mathfrak{A}$. Recall that both $\mathfrak{C}^{\mathfrak{A}}$ and $\mathfrak{G}^{\mathfrak{A}}$ are of size bounded by K_m, and such that we have a control over the R'-local behaviour of these two sets, for some $2r \leq R' \leq R_m$.

Let us explore the distribution of our r-local properties $\psi_i(x)$ for $1 \leq i \leq \ell$ inside the sets $\mathfrak{C}^{\mathfrak{A}}$ and $\mathfrak{G}^{\mathfrak{A}}$. To that end, we write

$$I_f \stackrel{\text{def}}{=} \{ 1 \leq j \leq \ell : \exists a \in \mathfrak{G}^{\mathfrak{A}}, \mathfrak{A}, x \mapsto a \models \psi_j(x) \}$$

In order to deal with the (now infamous) intersections of neighbourhoods, let us write, given a set variable X, the MSO-formula $\psi_i^X(x)$ as the relativisation of $\psi_i(x)$ to elements of X. This allows us to collect the r-local properties that *may become true* provided that we have a suitable intersection.

$$I_m \overset{\text{def}}{=} \{ 1 \leq j \leq l \colon \exists a \in \mathfrak{G}^{\mathfrak{A}}, \mathfrak{A}, x \mapsto a \models \exists X. \psi_j^X(x) \}$$

Assume now that $\mathfrak{A} \Rrightarrow_{Q,\mathsf{MSO}}^{R,K} \mathfrak{B}$. By definition, there exists sets[12] $C^{\mathfrak{B}}$ and $G^{\mathfrak{B}}$ in $\mathfrak{B}$ such that

$$\mathsf{mtp}_{\mathfrak{A}}^{Q}\left(\mathfrak{C}^{\mathfrak{A}} \cup \mathfrak{G}^{\mathfrak{A}}, R\right) = \mathsf{mtp}_{\mathfrak{B}}^{Q}\left(C^{\mathfrak{B}} \cup G^{\mathfrak{B}}, R\right). \tag{5.2}$$

12: We do not use the notation $\mathfrak{C}^{\mathfrak{B}}$ and $\mathfrak{G}^{\mathfrak{B}}$ because $C^{\mathfrak{B}}$ and $G^{\mathfrak{B}}$ do not have the properties stated in Lemma 5.2.9: there may be MSO local types realised in $\mathfrak{B}$ that are neither covered by $C^{\mathfrak{B}}$ nor frequent in $G^{\mathfrak{B}}$. However, every MSO local types realised in $\mathfrak{B}$ that was *also* realised in $\mathfrak{A}$ is either covered by $C^{\mathfrak{B}}$ or frequently appears in $G^{\mathfrak{B}}$.

Because we have chosen Q large enough to check distances up to R, we can transport most of the properties of $\mathfrak{C}^{\mathfrak{A}}$ and $\mathfrak{G}^{\mathfrak{A}}$ to $C^{\mathfrak{B}}$ and $G^{\mathfrak{B}}$. It is immediate that $\mathfrak{C}^{\mathfrak{A}}$ and $C^{\mathfrak{B}}$ have the same size, and the same also holds for $\mathfrak{G}^{\mathfrak{A}}$ and $G^{\mathfrak{B}}$. Furthermore, $\mathfrak{B}, \vec{x} \mapsto G^{\mathfrak{B}} \models \mathrm{indep}^{\mathcal{N}}(\vec{x}, R')$, and $\mathcal{N}_{\mathfrak{B}}(G^{\mathfrak{B}}, R') \cap \mathcal{N}_{\mathfrak{B}}(C^{\mathfrak{B}}, R') = \emptyset$. However, it is not possible to transport the "covering property" from $\mathfrak{A}$ to $\mathfrak{B}$ directly, as the latter might have more MSO-local types than $\mathfrak{A}$.

Let us define $E \overset{\text{def}}{=} \mathfrak{B} \setminus \mathcal{N}_{\mathfrak{B}}(C^{\mathfrak{B}}, R')$. Notice that $\mathcal{N}_{\mathfrak{B}}(G^{\mathfrak{B}}, r) \subseteq E$ because $2r \leq R' \leq R$.

Given $i \in I_m$, one can choose k distinct elements in $G^{\mathfrak{B}}$ such that $\mathfrak{B}, x \mapsto b \models \exists X. \psi_i^X(x)$; let us call this vector $\vec{b}_i^m$. Similarly, given $i \in I_f$, one can choose k distinct elements in $G^{\mathfrak{B}}$ such that $\mathfrak{B}, x \mapsto b \models \psi_i^X(x)$, let us call this vector $\vec{b}_i^f$.

Without loss of generality[13] since types in $G^{\mathfrak{B}}$ have at least $2 \times k \times l$ realisations, we can assume that $\vec{b}_i^f$ and $\vec{b}_j^m$ are disjoint for all choices of i, j.

13: Recall that elements in $\mathfrak{G}^{\mathfrak{A}}$ are not covered by $\mathfrak{C}^{\mathfrak{A}}$ by construction. Hence, every MSO-local type that is realised by an element in $\mathfrak{G}^{\mathfrak{A}}$ (hence, in $G^{\mathfrak{B}}$), must appear frequently in $\mathfrak{G}^{\mathfrak{A}}$, (hence, frequently in $G^{\mathfrak{B}}$).

Let $i \in I_m$ and $b \in \vec{b}_i^m$; there exists a subset $F_b \subseteq \mathcal{N}_{\mathfrak{B}}(b, r)$, such that $b \in F_b$, and $\mathcal{N}_B(b, r) \cap F_b, x \mapsto b \models \psi_i(x)$. Let us build E' as the structure E where the complements of the sets F_b have been removed:

$$E' \overset{\text{def}}{=} E \setminus \left(\bigcup_{i \in I_m} \bigcup_{b \in \vec{b}_i^m} [\mathcal{N}_{\mathfrak{B}}(b, r) \setminus F_b] \right) \quad .$$

We assert that for every $1 \leq i \leq l$ the following properties are equivalent:

1. $\mathfrak{A} \uplus E' \models \theta_i$,
2. $\mathcal{N}_{\mathfrak{A}}(\mathfrak{C}^{\mathfrak{A}}, R') \uplus E' \models \theta_i$, and
3. $\mathcal{N}_{\mathfrak{B}}(C^{\mathfrak{B}}, R') \uplus E' \models \theta_i$.

Since $\mathsf{mtp}_{\mathfrak{A}}^{Q}\left(\mathfrak{C}^{\mathfrak{A}}, R\right) = \mathsf{mtp}_{\mathfrak{B}}^{Q}\left(C^{\mathfrak{B}}, R\right)$, and $r \leq R' \leq R_m \leq R$, it is clear that $\mathsf{tp}_{\mathfrak{A}}^{Q}\left(\mathfrak{C}^{\mathfrak{A}}, R'\right)$ and $\mathsf{tp}_{\mathfrak{B}}^{Q}\left(C^{\mathfrak{B}}, R'\right)$ are equal. Let $1 \leq i \leq l$. Since, $Q \geq \mathrm{rk}(\theta_i)$, this implies $\mathcal{N}_{\mathfrak{A}}(\mathfrak{C}^{\mathfrak{A}}, R') \uplus E' \models \theta_i$ if and only if $\mathcal{N}_{\mathfrak{B}}(C^{\mathfrak{B}}, R') \uplus E' \models \theta_i$.[14]

14: This is actually the consequence of a stronger statement: $\mathcal{N}_{\mathfrak{A}}(\mathfrak{C}^{\mathfrak{A}}, R') \uplus E'$ satisfies the same first order sentences of quantifier rank at most Q as $\mathcal{N}_{\mathfrak{B}}(C^{\mathfrak{B}}, R') \uplus E'$. This holds because the first order theory of a disjoint union is uniquely determined by the first order theory of its parts (even at a fixed quantifier rank).

Let us now prove the equivalence between the first two items.

- Assume $\mathfrak{A} \uplus E' \models \theta_i$. Let $S \overset{\text{def}}{=} \{ a \in \mathfrak{A} \colon \mathfrak{A}, x \mapsto a \models \psi_i(x) \}$.

- Assume that $\mathcal{N}_{\mathfrak{A}}(S, r) \subseteq \mathcal{N}_{\mathfrak{A}}(\mathfrak{C}^{\mathfrak{A}}, R')$. Then, $\mathcal{N}_{\mathfrak{A}}(\mathfrak{C}^{\mathfrak{A}}, R') \uplus E' \models \theta_i$.
- Otherwise, there exists $a \in \mathfrak{A}$ such that $\mathfrak{A}, x \mapsto a \models \psi_i(x)$ and $\mathcal{N}_{\mathfrak{A}}(a, r) \not\subseteq \mathcal{N}_{\mathfrak{A}}(\mathfrak{C}^{\mathfrak{A}}, R')$. By definition of $\mathfrak{C}^{\mathfrak{A}}$, this proves that $i \in I_f$. As a consequence, the vector $\vec{b}_i^f$ of size k is defined and belongs[15] to $G^{\mathfrak{B}} \cap E'$. In particular, $E' \models \theta_i$,[16] therefore $\mathcal{N}_{\mathfrak{A}}(\mathfrak{C}^{\mathfrak{A}}, R') \uplus E' \models \theta_i$.

▶ Assume that $\mathcal{N}_{\mathfrak{A}}(\mathfrak{C}^{\mathfrak{A}}, R') \uplus E' \models \theta_i$. Let us define S as the set of witnesses for the property $\psi_i(x)$ in $\mathcal{N}_{\mathfrak{A}}(\mathfrak{C}^{\mathfrak{A}}, R')$, which amounts formally to: $S \stackrel{\text{def}}{=} \{ a \in \mathfrak{A} \colon \mathcal{N}_{\mathfrak{A}}(\mathfrak{C}^{\mathfrak{A}}, R'), x \mapsto a \models \psi_i(x) \}$.

- If $\mathcal{N}_{\mathfrak{A}}(S, r) \cap \mathcal{N}_{\mathfrak{A}}(\mathfrak{C}^{\mathfrak{A}}, R') = \mathcal{N}_{\mathfrak{A}}(S, r)$. Then, that $\mathfrak{A} \uplus E' \models \theta_i$.
- Otherwise, there exists $a \in S$ such that $\mathcal{N}_{\mathfrak{A}}(a, r)$ is a proper subset of $\mathcal{N}_{\mathfrak{A}}(\mathfrak{C}^{\mathfrak{A}}, R')$ and such that $\mathcal{N}_{\mathfrak{A}}(a, r) \cap \mathcal{N}_{\mathfrak{A}}(\mathfrak{C}^{\mathfrak{A}}, R'), x \mapsto a \models \psi_i(x)$. In particular, $\mathcal{N}_{\mathfrak{A}}(a, r), x \mapsto a \models \exists X. \psi_i^X(x)$. By definition of $\mathfrak{C}^{\mathfrak{A}}$, this proves that $i \in I_m$.
  Given $b \in \vec{b}_i^m$, it holds that $E', x \mapsto b \models \psi_i(x)$ because $\mathcal{N}_{\mathfrak{B}}(b, r) \cap F_b, x \mapsto b \models \psi_i(x)$ and $\mathcal{N}_{E'}(b, r) = \mathcal{N}_{\mathfrak{B}}(b, r) \cap F_b$. As a consequence, $E' \models \theta_i$, and we conclude that $\mathfrak{A} \uplus E' \models \theta_i$.

We can now conclude by playing a simple ping-pong argument. Since $\mathfrak{A} \models \varphi$, and $\mathfrak{A} \subseteq_i \mathfrak{A} \uplus E'$, we know that $\mathfrak{A} \uplus E' \models \varphi$. Moreover, the equivalences above assert that $\mathcal{N}_{\mathfrak{B}}(C^{\mathfrak{B}}, R') \uplus E' \models \varphi$. Remark that $\mathcal{N}_{\mathfrak{B}}(C^{\mathfrak{B}}, R') \uplus E' \subseteq_i \mathfrak{B}$, hence $\mathfrak{B} \models \varphi$. ■

15: They belong to $G^{\mathfrak{B}}$ by construction, and the important thing to notice is that they are still present in E'. This is because elements of $G^{\mathfrak{B}}$ are far away, and in particular the vectors $\vec{b}_j^m$ are not close to $\vec{b}_i^f$: in particular, E' contains $\mathcal{N}_{\mathfrak{B}}(\vec{b}_i^f, r)$.

16: This holds because we found $|\vec{b}_i^f|$ elements that are far enough from each other, and that all share the same MSO-local type. Because $\psi_i(x)$ belongs to the MSO-local type of a, we have found enough pairwise distant witnesses to conclude that $E' \models \theta_i$.

The use of disjoint unions was crucial in the construction, and removing the assumption that $\mathcal{C}$ is closed under this operation provides counter-examples to Lemma 5.2.10. We refine Example 5.1.11 hereafter.

Example 5.2.11. The sentence $\varphi \stackrel{\text{def}}{=} \forall x. \deg(x) = 2$ is preserved under extensions over $\downarrow_{\subseteq_i} \mathsf{Cycles}$ but is not equivalent to an existential local sentence over $\downarrow_{\subseteq_i} \mathsf{Cycles}$.

Proof Sketch. We use Lemma 5.2.1: for every r, k, the cycle C_{2rk+1} is a $\subseteq_i$-minimal model of φ in $\downarrow_{\subseteq_i} \mathsf{Cycles}$ that does not belong to $\mathsf{Local}(\downarrow_{\subseteq_i} \mathsf{Cycles}, r, k)$. ■

Remark 5.2.12. Some classes $\mathcal{C} \subsetneq \mathsf{Fin}(\sigma)$ are localisable but not closed under disjoint unions. It is the case for finite classes of finite structures.

5.3. Discussion

Beyond existential sentences? While there is no clear adaptation of Theorem 5.1.2 to arbitrary fragments of FO, one can check that for some subfragments of EFO, the analogue theorems holds.

Theorem 5.3.1. *Let $\mathcal{C}$ be a hereditary class of finite structures* *closed under disjoint unions, and* $\mathsf{F} \in \{\mathsf{EFO}, \mathsf{EPFO}^{\neq}, \mathsf{EPFO}\}$ *be a fragment. The* F*-preservation theorem relativises to* $\mathcal{C}$ *if and only if the* F*-preservation theorem relativises to* $\mathsf{Local}(\mathcal{C}, r, k)$ *for all* $r \geq 0$ *and* $k \geq 1$*.*

Proof Sketch. This is because Lemma 5.1.1 is trivially adapted to any fragment $\mathsf{F} \in \{\mathsf{EFO}, \mathsf{EPFO}^{\neq}, \mathsf{EPFO}\}$. Indeed, what was needed is essentially the existence of suitable diagram sentences. We restate the claim hereafter.

Let $\mathcal{C}$ be a hereditary class of finite structures, and let F be one of EFO, $\mathsf{EPFO}^{\neq}$, or EPFO. Then, the following are equivalent:

- The F-preservation theorem relativises to $\mathsf{Local}(\mathcal{C}, r, k)$ for all $r, k \geq 0$,
- Existential local sentences that are preserved under $\leq_{\mathsf{F}}$ (over $\mathcal{C}$) are equivalent (over $\mathcal{C}$) to $\exists\mathsf{F}$ sentences.

The proof is obtained by considering the proof of Lemma 5.1.1 verbatim, and replacing $\subseteq_i$ with $\leq_{\mathsf{F}}$. This is possible because Lemma 3.2.3 applies to all $\mathsf{F} \in \{\mathsf{EFO}, \mathsf{EPFO}^{\neq}, \mathsf{EPFO}\}$. ∎

However, this is not quite satisfactory: we would expect Theorem 5.3.1 to hold for all $\mathsf{F} \subseteq \mathsf{EFO}$. The only roadblock is the usage of Lemma 3.2.3, which itself relies on the existence of suitable diagram sentences given by Lemma 3.1.5. This leads to the following conjectures.

Conjecture 5.3.2 (Diagram Formulas Below EFO)**.** Let σ be a finite relational signature, and $\mathsf{F} \subseteq \mathsf{EFO}$ be a fragment. Then, for all $\mathfrak{A} \in \mathsf{Fin}(\sigma)$, the diagram sentence $\Delta_{\mathfrak{A}}^{\exists\mathsf{F}} \in \exists\mathsf{F}$ exists.

Conjecture 5.3.3 (Diagram Formulas)**.** Let σ be a finite relational signature, and $\mathsf{F} \subseteq \mathsf{FO}[\sigma]$ be a fragment. Then, for all $\mathfrak{A} \in \mathsf{Fin}(\sigma)$, the diagram sentence $\Delta_{\mathfrak{A}}^{\exists\mathsf{F}} \in \exists\mathsf{F}$ exists.

Limitations. Notice that Theorems 5.1.2 and 5.3.1 have intrinsic limitations. They only consider hereditary classes that are closed under disjoint unions. Moreover, their application “suppose” the fact that it is easier to prove that a preservation theorem locally relativises, than that it globally does. In the presence of dense classes such as Example 5.1.10, this is simply not true.

References

[6] Albert Atserias, Anuj Dawar and Martin Grohe. 'Preservation under extensions on well-behaved finite structures'. In: *SIAM Journal on Computing* 38 (2008), pp. 1364–1381. DOI: 10.1137/060658709 (cit. on pp. 37, 40, 47, 55, 64, 67, 68, 74, 97, 98, 117, 121, 127).

[21] Anuj Dawar, Martin Grohe, Stephan Kreutzer and Nicole Schweikardt. 'Approximation schemes for first-order definable optimisation problems'. In: *21st Annual IEEE Symposium on Logic in Computer Science (LICS'06)*. 2006, pp. 411–420. DOI: 10.1109/lics.2006.13 (cit. on pp. 68, 130, 131).

[27] Guoli Ding. 'Subgraphs and well-quasi-ordering'. In: *Journal of Graph Theory* 16 (1992), pp. 489–502. DOI: 10.1002/jgt.3190160509 (cit. on pp. 37, 40, 44, 45, 122).

[71] Aliaume Lopez. 'When Locality Meets Preservation'. In: *LICS'22* (2022), pp. 1–14. DOI: 10.1145/3531130.3532498. URL: https://doi.org/10.1145/3531130.3532498 (cit. on pp. 63, 67, 117).

[77] Jaroslav Nešetřil and Patrice Ossona de Mendez. 'On nowhere dense graphs'. In: *European Journal of Combinatorics* 32 (2011), pp. 600–617. DOI: 10.1016/j.ejc.2011.01.006. URL: https://www.sciencedirect.com/science/article/pii/S0195669811000151 (cit. on pp. 46, 121, 259).

[78] Jaroslav Nešetřil and Patrice Ossona de Mendez. *Sparsity: Graphs, Structures, and Algorithms*. Springer, 2012. DOI: 10.1007/978-3-642-27875-4 (cit. on pp. 25, 120).

Logically Presented Spaces 6.

Outline of the chapter

This chapter introduces an algebraic and topological approach to preservation theorems, by pairing the idea of definability coming from logic together with the idea of compactness coming from topology. This sparks the study of triples $(\mathsf{X}, \tau, \mathcal{B})$ where τ is a topology and $\mathcal{B}$ a Boolean subalgebra of $\mathcal{P}(\mathsf{X})$. The main motivation for this abstract study is to find a way around the non-compositionality of preservation theorems and the complexity of proving their relativisation.

Goals of the chapter

At the end of the chapter, I hope you will be convinced that logically presented pre-spectral spaces are good candidates for abstracting preservation theorems. In many cases, the two notion coincide, but the former enjoys appealing compositional properties akin to those of well-quasi-orders and Noetherian spaces.

Genesis. In general, stability properties of preservation theorems are not studied for themselves in the literature.[1] However, some simple properties are often used implicitly, especially when proving that preservation theorems do not relativise in the finite [66, 89, 90]. This is already what happened in Subsection 4.3.1, where we leveraged stability properties to simplify the proofs. This motivated the search for a “generic stability argument”, which lead us to the introduction of logically presented pre-spectral spaces[2] (lpps).

1: Except for [85, Chapter 10] that lists compositional properties in the case of the Chang-Łoś-Suszko Theorem. However, this study is tied to a specific preservation theorem.

2: See Definition 6.1.11 p. 143

At first, our goal was to connect preservation theorems with the theory of Noetherian spaces and well-quasi-orderings. However, these two theories do not correctly model preservation theorems, and this has lead to the introduction of a new category of topological spaces: logically presented spaces. Later on, a worthwhile connection with spectral spaces[3] was uncovered, which gave us the impression that some theorems could be directly taken from the theory developed around these topological spaces. As for wqos and Noetherian spaces, the connection with preservation theorems remained too coarse for our purposes, but gave us intuition about what kind of positive and negative results were to be expected.

3: See Definition 6.2.3 p. 147

Once a proper definition was given, the natural question was whether the stability properties of related spaces (wqos, Noetherian spaces, spectral spaces) could be generalised to this new notion. We answered this in the positive, which allows us to prove relativisation of preservation theorems by “compositional techniques”: instead of providing an ad-hoc proof that a preservation theorem relativises to a class $\mathcal{C}$, one can now interpret $\mathcal{C}$ as structures constructed from simpler parts where such a relativisation was known to hold.

The first formal appearance of lpps followed from the research that was done in [70], which the current chapter mostly follows. There are two main differences that should be highlighted: in [70], there was a focus on using quasi-orderings rather than topologies, and the definition of lpps has been adapted to remove the need for the rather uncommon notion of "diagram basis."

[70]: Lopez (2021), 'Preservation Theorems Through the Lens of Topology'

Contributions. In Section 6.1, we will introduce the topological object that will capture preservation theorems formally. We split the definition in two parts: first we define logically presented spaces[4] (lps), which are topological spaces with a notion of definable subsets, and then we adapt the notion of preservation theorem to this setting and define the subclass of logically presented pre-spectral spaces[5] (lpps). This barbaric name was given due to its connection with spectral spaces, and will be explained in Section 6.2, where we study how lpps are a generalisation of Noetherian spaces, but also a variant of spectral spaces.

4: See Definition 6.1.1 p. 140

5: See Definition 6.1.11 p. 143

Finally, Section 6.3 is devoted to the study of "stability properties" of lpps. In particular, we will give sufficient conditions for a subset to be an lpps in the induced topology as show in Lemma 6.3.2, which together with the ability to use logical maps[6] Lemma 6.3.3, provides us with a basic toolbox to study preservation theorems in the stricter setting of lpps.

6: See Definition 6.1.8 p. 142

We continue the study of the stability properties of lpps by trying to recover analogues of the algebraic constructions that preserve well-quasi-orderings or Noetherian spaces. This is a worthwhile effort because this offers a "compositional" way to build lpps, which is one of the reasons for the success of wqos and Noetherian spaces, and is an approach that was otherwise lacking in the literature. Let us briefly recall in Tables 6.1 and 6.2 the usual operations that can be used to build complex wqos and Noetherian spaces out of simpler ones.

In the case of lpps, we prove that lpps can be built using algebraic operations (products, sums, colourings), morphisms (logical maps), and subsets. Notice that in this last section, results are split into pairings of a topological theorem and a logical theorem. For instance, the stability under products is first a topological result in Theorem 6.3.13, and then turned into a result about FO in Lemma 6.3.16. This is a strength of this approach, as one can separately talk about what are the "good" definable subsets for a given space, and then separately prove that FO-definable subsets are indeed of the correct shape.

Note that the "logical proofs" of Section 6.3 are mostly obtained via composition theorem for first order logic, also known as, Feferman-Vaught style compositional techniques. We believe that this illustrates why the syntactic presentation of these techniques is a valuable contribution.

Constructor	Syntax	Quasi-order
Finite set	Σ	equality
Well founded set	P	$\leq$
Disjoint sum	$W_1 + W_2$	co-product ordering
Product	$W_1 \times W_2$	product ordering
Finite words	$W^\star$	subword embedding
Finite multisets	$\mathsf{M}^\diamond(W)$	multiset embedding
Finite sets	$\mathsf{P}_\mathsf{f}(W)$	Hoare quasi-ordering
Finite trees	$\mathsf{T}(W)$	tree embedding

Table 6.1.: An algebra of well-quasi-orders [23]. This table is meant as an illustration of the prolific stability properties of well-quasi-orderings. For every operation, we not only provide the construction of the new elements, but also the new preorder. To gain space and because most of these will not be used in the document, we deferred the definition to the corresponding cheatsheet Chapter C (Well-Quasi-Orderings Cheat Sheet).

Constructor	Syntax	Topology
Well-quasi-orders	P	Alexandroff topology
Complex vectors	$\mathbb{C}^k$	Zariski topology
Disjoint sum	$\mathsf{X}_1 + \mathsf{X}_2$	co-product topology
Product	$\mathsf{X}_1 \times \mathsf{X}_2$	product topology
Finite words	$\mathsf{X}^\star$	subword topology
Finite trees	$\mathsf{T}(\mathsf{X})$	tree topology
Finite multisets	$\mathsf{M}^\diamond(\mathsf{X})$	multiset topology
Transfinite words	$\mathsf{X}^{<\alpha}$	regular transfinite subword topology
Powerset	$\mathcal{P}(\mathsf{X})$	lower Vietoris

Table 6.2.: An algebra of Noetherian spaces [see 44, 45, 47]. This table is meant as an illustration of the prolific stability properties of Noetherian spaces, which mimics (and extends) the stability of well-quasi-orderings. As for wqos, we have to define the topologies placed on the newly constructed spaces, and to gain space, we deferred their definition to the corresponding cheatsheet Chapter D (Topology Cheat Sheet).

Table 6.3.: Algebraic operations preserving lpps. Arbitrary topological spaces are written X, and classes of structures $\mathcal{C}$. For each constructor, there is one theorem stating that the resulting space is an lpps, and one that states that the resulting space with the Boolean subalgebra FO is an lpps.

Operation	Symbol	Hypothesis	Topology	FO-version
sum	$\mathsf{X} + \mathsf{X}'$	-	Lemma 6.3.8	Lemma 6.3.15
product	$\mathsf{X} \times \mathsf{X}'$	-	Theorem 6.3.13	Lemma 6.3.16
inner product	$\mathcal{C} \otimes \mathcal{C}'$	-	n/a	Exercise 6.3.20
finite words	$\mathsf{X}^\star$	-	Theorem 6.3.33	Corollary 6.3.36
wreath product	$\mathcal{C} \rtimes_\mathsf{F} \mathsf{X}'$	$\mathcal{C}$ is ∞-wqo	Theorem 6.3.44	Theorem 6.3.46

6.1. Topological Preservation Theorems

7: See Theorem 3.1.9 p. 33

Recall that a preservation theorem[7] for a fragment F of FO states for all sentences $\varphi \in \mathsf{FO}$ that are preserved under $\leq_{\mathsf{F}}$, there exists a sentence $\psi \in \exists\mathsf{F}$ that is equivalent to φ. The typical proof of this result relies on the compactness theorem of first order logic.

To provide an abstract counterpart to a preservation theorem in terms of purely topological properties, we want to represent both the logic FO and the fragment F. Given a class $\mathcal{C}$, the logic FO will be interpreted as a Boolean subalgebra of $(\mathcal{P}(\mathcal{C}), \emptyset, \mathcal{C}, \subseteq)$, and the fragment F will be interpreted as a topology over $\mathcal{C}$. The motivation behind this difference of treatment is that "first order definability" should be closed under negation, while "being preserved under $\leq_{\mathsf{F}}$" typically amounts to an infinite disjunction of facts in F, the latter being stable under finite intersections, but not under negations.

Concretely, we are interested in state spaces that can simultaneously be viewed as topological spaces and as spaces with a notion of definable set, through the interpretation of first-order sentences.

Definition 6.1.1. *A logically presented space (abbreviated as lps) is a triple $\langle\langle \mathsf{X}, \tau, \mathcal{B} \rangle\rangle$, where*

1. *τ is a topology on X,*
2. *$\mathcal{B}$ is a Boolean subalgebra of $(\mathcal{P}(\mathsf{X}), \emptyset, \mathsf{X}, \subseteq)$,*
3. *$\tau = \langle \tau \cap \mathcal{B} \rangle_{\mathsf{topo}}$.*

The name logically presented comes from the fact that $\tau \cap \mathcal{B}$ is a sub-basis of τ, which is stable under finite intersections and unions, and will in our examples be the collection of first order definable subsets of X (which makes sense when X is a class of structures).

The running example of logically presented space is the class of finite structures $\mathsf{Fin}(\sigma)$ together with the Alexandroff topology $\mathsf{Alex}(\subseteq_i)$, where the Boolean subalgebra is just $[\![\mathsf{FO}]\!]_{\mathsf{Fin}(\sigma)}$. Following this intuition, we call sets in τ open subsets, and those in $\mathcal{B}$ are *definable*.

Example 6.1.2. Let σ be a finite relational signature. The following is a logically presented space: $\langle\langle \mathsf{Fin}(\sigma), \mathsf{Alex}(\subseteq_i), [\![\mathsf{FO}]\!]_{\mathsf{Fin}(\sigma)} \rangle\rangle$.

Proof Sketch. Recall from Lemma 3.1.5 that the $\subseteq_i$-upward closure of a finite structure $\mathfrak{A}$ is definable in $\exists\mathsf{QF}$. As a consequence, the upward closure of $\mathfrak{A}$ is first order definable inside $\mathcal{C}$. ∎

To ease the notations, we will often write $\langle\langle \mathsf{Fin}(\sigma), \mathsf{Alex}(\subseteq_i), \mathsf{FO} \rangle\rangle$ when there is no ambiguity that first order sentences are interpreted over $\mathsf{Fin}(\sigma)$ to build a Boolean subalgebra of $\mathcal{P}(\mathsf{Fin}(\sigma))$.

In order to provide an example that is not explicitly tailored to tackle preservation theorems, let us see in Example 6.1.3 how the real line can be seen both from a topological perspective, and from an algebraic one.

Example 6.1.3. Let $(\mathbb{R}, \tau)$ be the real line with its usual topology. Let $\mathcal{B} \stackrel{\text{def}}{=} \langle\{ [a,b] : a \leq b \in \mathbb{R} \}\rangle_{\mathsf{bool}}$. Then, $\langle\!\langle \mathsf{X}, \tau, \mathcal{B} \rangle\!\rangle$ is an lps.

Proof. Recall that the usual topology[8] of the real line is generated by the open intervals $]a,b[$, for $a < b \in \mathbb{R}$. To conclude that $\langle\!\langle \mathsf{X}, \tau, \mathcal{B} \rangle\!\rangle$ is an lps, it suffices to notice that for $a < b \in \mathbb{R}$, $]a,b[\in \mathcal{B}$.

8: The topology of the real line was introduced in Example 3.3.15.

To that end, let us remark that for $a < b \in \mathbb{R}$, the following equality holds:

$$]a,b[= \underbrace{[a,b] \setminus ([a,a] \cup [b,b])}_{\in \mathcal{B}} \quad . \qquad \blacksquare$$

Because we are interested in the relativisation of preservation theorems, most of the time we will start from a fragment F of FO and a class $\mathcal{C} \subseteq \mathsf{Struct}(\sigma)$, in order to build the lps $\langle\!\langle \mathcal{C}, \langle \exists \mathsf{F} \rangle_{\mathsf{topo}}, \mathsf{FO} \rangle\!\rangle$. As a sanity check, let us prove that this effectively generalises in a topological way the setting of preservation theorems in the finite.

Fact 6.1.4. Let $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$, and F be a fragment of FO. Then, the space $\langle\!\langle \mathcal{C}, \langle \exists \mathsf{F} \rangle_{\mathsf{topo}}, \mathsf{FO} \rangle\!\rangle$ is a logically presented space.

Proof. We only have to check that $\langle \exists \mathsf{F} \rangle_{\mathsf{topo}} = \langle \langle \exists \mathsf{F} \rangle_{\mathsf{topo}} \cap \mathsf{FO} \rangle_{\mathsf{topo}}$. This is an immediate consequence of the inclusion $\exists \mathsf{F} \subseteq \mathsf{FO}$. $\blacksquare$

There is a discrepancy when talking about $\langle \exists \mathsf{F} \rangle_{\mathsf{topo}}$ instead of $\mathsf{Alex}(\leq_{\mathsf{F}})$, but for the fragments of interest, the two coincide.

Lemma 6.1.5. *Let $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$, $\mathsf{F} \in \{ \mathsf{EFO}, \mathsf{EPFO}^{\neq}, \mathsf{EPFO} \}$ be a fragment of FO. The following two logically presented spaces are equal:*

1. $\langle\!\langle \mathsf{Fin}(\sigma), \langle \exists \mathsf{F} \rangle_{\mathsf{topo}}, \mathsf{FO} \rangle\!\rangle$,
2. $\langle\!\langle \mathsf{Fin}(\sigma), \mathsf{Alex}(\leq_{\mathsf{F}}), \mathsf{FO} \rangle\!\rangle$.

Proof. Let $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$, F be a fragment of FO. Then, $\mathsf{Alex}(\leq_{\mathsf{F}}) \subseteq \langle \exists \mathsf{F} \rangle_{\mathsf{topo}}$ because one can define the upward closure of a given structure using a diagram formula (see Lemma 3.1.5).

Conversely, a sentence $\varphi \in \exists \mathsf{F}$ is always preserved under $\leq_{\mathsf{F}}$, and in particular, $[\![\varphi]\!]_{\mathcal{C}}$ is upwards closed for $\leq_{\mathsf{F}}$.

We have proven that $\langle \exists \mathsf{F} \rangle_{\mathsf{topo}} = \mathsf{Alex}(\leq_{\mathsf{F}})$. $\blacksquare$

When considering finite structures, it might seem that introducing the $\exists \mathsf{F}$ construction overcomplicates the presentation. Let us argue that the presentation $\langle\!\langle \mathsf{Fin}(\sigma), \langle \exists \mathsf{F} \rangle_{\mathsf{topo}}, \mathsf{FO} \rangle\!\rangle$ is truly the one found when studying preservation theorems. We prove in Example 6.1.6 that Item 2 does not yield an lps over $\mathsf{Struct}(\sigma)$, while Fact 6.1.4 shows that Item 1 does.

Example 6.1.6. The space $\langle\!\langle \mathsf{Struct}(\sigma), \mathsf{Alex}(\subseteq_i), \mathsf{FO}\rangle\!\rangle$ is not a logically presented space.

Proof. Thanks to the Łoś-Tarski Theorem, $\mathsf{Alex}(\subseteq_i) \cap [\![\mathsf{FO}]\!] = [\![\mathsf{EFO}]\!]$. Remark that $[\![\mathsf{EFO}]\!]$ is stable under finite unions and finite intersections, and that for all non-empty open subset $U \in [\![\mathsf{EFO}]\!]$, $\mathsf{Fin}(\sigma) \cap U \neq \emptyset$, i.e., every existential positive sentence that has a model has a *finite* model. If the space were to be logically presented, the latter would imply that $\mathsf{Fin}(\sigma)$ is dense in $\mathsf{Struct}(\sigma)$. However, the $\subseteq_i$ upwards closed subset $\mathsf{Struct}(\sigma) \setminus \mathsf{Fin}(\sigma)$ of "truly infinite structures" is an open subset of $\mathsf{Alex}(\subseteq_i)$ that does not intersect $\mathsf{Fin}(\sigma)$ by definition: this is absurd. ∎

A first key observation is that the property of being logically presented is naturally preserved under subspaces as witnessed by the following proposition, which has to be compared to the fact that preservation theorems in general do not relativise (see Table 3.3 on page 35).

Recall that the induced topology written $\tau_{|\mathsf{X}}$ is the collection of all $U \cap \mathsf{X}$ when U ranges in τ, see Definition D.2.5.

Fact 6.1.7. Let $\langle\!\langle \mathsf{X}, \tau, \mathcal{B}\rangle\!\rangle$ be an lps, and $\mathsf{Y} \subseteq \mathsf{X}$. Then, Y endowed with the induced topology $\tau_{|\mathsf{Y}}$ and the induced Boolean algebra $\mathcal{B}_{|\mathsf{Y}}$ is a logically presented space.

Proof Sketch. Let $U \in \tau_{|\mathsf{Y}}$. By definition, there exists $V \in \tau$ such that $U = V \cap \mathsf{Y}$. Because X is an lps, $V \in \langle \tau \cap \mathcal{B}\rangle_{\mathsf{topo}}$. As a consequence, $U \in \langle \tau_{|\mathsf{Y}} \cap \mathcal{B}_{|\mathsf{Y}}\rangle_{\mathsf{topo}}$. ∎

The category of logically presented spaces. The language of category theory is well-adapted to the goal of stating stability properties of logically presented spaces. However, to keep minimal theoretical dependencies, we will (almost always) state our theorems without requiring a deep knowledge of category theory and commutative diagrams.

Let us now define the natural morphisms between logically presented spaces, which allows us to talk about the category LPS whose objects are lps and whose morphisms are logical maps.

The notion of logical map deals with the topology and the Boolean subalgebra separately, and this is crucial. Asking that $f^{-1}(U) \in \tau \cap \mathcal{B}$ whenever $U \in \tau' \cap \mathcal{B}'$, only states that f is continuous.

Definition 6.1.8. *Let $\langle\!\langle \mathsf{X}, \tau, \mathcal{B}\rangle\!\rangle$ and $\langle\!\langle \mathsf{X}', \tau', \mathcal{B}'\rangle\!\rangle$ be two lps. A function $f : \mathsf{X} \to \mathsf{X}'$ is a logical map whenever*

1. *For all $U \in \tau'$, $f^{-1}(U) \in \tau$;*
2. *For all $D \in \mathcal{B}'$, $f^{-1}(D) \in \mathcal{B}$.*

While this is not the main focus of this chapter, let us state the following exercises for category theory enthusiasts.

Exercise 6.1.9 (Beginner Category)**.** The category LPS has all products and all co-products.

Exercise 6.1.10 (Advanced Category)**.** The category LPS is topological, and is not cartesian closed.

Adding compactness. In our definition of a logically presented space (Definition 6.1.1) we do not capture the full statement of a preservation theorem. While it is possible to use this framework to express that a given definable subset is "open," we lack a theorem stating that such definable open subsets have a strong semantic property.[9] In the light of the canonical proof of preservation theorems using the compactness theorem of first order logic, we will ask for *definable open subsets* to be compact.

9: For preservation theorems, this strong semantic property was "being expressible in a fragment of FO."

Recall that $\mathcal{K}^\circ(\mathsf{X})$ denotes the set of compact open subsets of X.

Definition 6.1.11. *A logically presented pre-spectral space is a logically presented space $\langle\!\langle \mathsf{X}, \tau, \mathcal{B} \rangle\!\rangle$ such that the open definable subsets are precisely the compact open subsets:*

$$\mathcal{B} \cap \tau = \mathcal{K}^\circ(\mathsf{X}) \quad . \tag{6.1}$$

Remark that the inclusion $\mathcal{K}^\circ(\mathsf{X}) \subseteq \mathcal{B} \cap \tau$ always holds for a lps. This is because, $\mathcal{B} \cap \tau$ generates the topology τ. As a consequence, and for the rest of this document, we will only check the interesting inclusion $\mathcal{K}^\circ(\mathsf{X}) \supseteq \mathcal{B} \cap \tau$ when proving that an lps is an lpps.

Let us immediately justify the above definition by relating it to preservation theorems via Theorem 6.1.12. This can be seen as a "correctness theorem" for our modelling of preservation theorems.

Theorem 6.1.12. *Let $\mathcal{C} \subseteq \mathsf{Struct}(\sigma)$, let F be a fragment of FO. The following are equivalent:*

1. *The F-preservation theorem relativises to $\mathcal{C}$, and $[\![\exists\mathsf{F}]\!]_{\mathcal{C}} \subseteq \mathcal{K}^\circ(\mathcal{C}, \langle\exists\mathsf{F}\rangle_{\mathsf{topo}})$,*
2. *The space $\langle\!\langle \mathcal{C}, \langle\exists\mathsf{F}\rangle_{\mathsf{topo}}, \mathsf{FO} \rangle\!\rangle$ is an lpps, and $\langle\exists\mathsf{F}\rangle_{\mathsf{topo}} \cap [\![\mathsf{FO}]\!]_{\mathcal{C}} = \mathsf{Alex}(\leq_{\mathsf{F}}) \cap [\![\mathsf{FO}]\!]_{\mathcal{C}}$.*

The extra assumption of Item 1 in Theorem 6.1.12 states that $\exists\mathsf{F}$ sentences are "reasonable." In the specific case where $\exists\mathsf{F} = \mathsf{EFO}$, and $\mathcal{C} = \mathsf{Fin}(\sigma)$, $\mathcal{K}^\circ(\mathsf{Fin}(\sigma), \langle\mathsf{EFO}\rangle_{\mathsf{topo}})$ is exactly the set of $\subseteq_i$-upward closures of finitely many structures, and the extra assumption boils down to asking that an existential sentence has finitely many $\subseteq_i$-minimal models.

The extra assumption of Item 2 in Theorem 6.1.12 states that $\exists\mathsf{F}$ can express suitable diagram sentences. Namely, it states that the upward closure of a structure is expressible using $\exists\mathsf{F}$.

Proof. Assume that the F-preservation theorem relativises to $\mathcal{C}$ and that $[\![\exists\mathsf{F}]\!]_{\mathcal{C}} \subseteq \mathcal{K}^\circ(\mathcal{C}, \langle\exists\mathsf{F}\rangle_{\mathsf{topo}})$. Then, $\mathsf{Alex}(\leq_{\mathsf{F}}) \cap [\![\mathsf{FO}]\!]_{\mathcal{C}} = \langle\exists\mathsf{F}\rangle_{\mathsf{topo}} \cap [\![\mathsf{FO}]\!]_{\mathcal{C}} = [\![\exists\mathsf{F}]\!]_{\mathcal{C}}$. Furthermore, $\langle\exists\mathsf{F}\rangle_{\mathsf{topo}} \cap [\![\mathsf{FO}]\!]_{\mathcal{C}} \subseteq \mathcal{K}^\circ(\mathcal{C}, \langle\exists\mathsf{F}\rangle_{\mathsf{topo}})$, and therefore $\langle\!\langle \mathcal{C}, \langle\exists\mathsf{F}\rangle_{\mathsf{topo}}, \mathsf{FO} \rangle\!\rangle$ is a logically presented pre-spectral space.

Assume that the space $\langle\!\langle \mathcal{C}, \langle\exists\mathsf{F}\rangle_{\mathsf{topo}}, \mathsf{FO} \rangle\!\rangle$ is a logically presented pre-spectral space and $\langle\exists\mathsf{F}\rangle_{\mathsf{topo}} \cap [\![\mathsf{FO}]\!]_{\mathcal{C}} = \mathsf{Alex}(\leq_{\mathsf{F}}) \cap [\![\mathsf{FO}]\!]_{\mathcal{C}}$. Remark that $[\![\varphi]\!]_{\mathcal{C}}$ is a compact subset of $\mathcal{C}$. Hence, we can leverage Lemma 3.3.25 to conclude the existence of a sentence $\psi \in \exists\mathsf{F}$ that is equivalent to φ over $\mathcal{C}$. ■

Notice that in this theorem, both Items 1 and 2 have "extra assumptions." These are needed because every lpps is a compact topological space (Lemma 6.1.13). In the case of spaces of structures, this compactness is a stronger property than simply asking for a preservation theorem to hold, as demonstrated in Example 6.1.14. Conversely, preservation theorems such as introduced in Theorem 3.1.9 talk about F-embeddings between structures, which is not a property that can be described *as is* in topological terms.

Lemma 6.1.13. *Every logically presented pre-spectral space is compact.*

Proof. Because $\mathcal{B}$ is a Boolean subalgebra of $\mathcal{P}(\mathsf{X})$, $\mathsf{X} \in \mathcal{B}$. Moreover, X is always an open subset of X, hence $\mathsf{X} \in \mathcal{K}^\circ(\mathsf{X})$ and is compact. ■

Recall that Cycles were already used as counter examples in Example 5.1.11 on page 125.

Example 6.1.14. The Łoś-Tarski Theorem relativises to the class Cycles of finite cycles, but $\langle\!\langle \mathsf{Cycles}, \langle \mathsf{EFO} \rangle_{\mathsf{topo}}, \mathsf{FO} \rangle\!\rangle$ is not an lpps.

Proof. Notice that over Cycles, $\langle \mathsf{EFO} \rangle_{\mathsf{topo}}$,[10] is precisely the discrete topology. Indeed, C_i and C_j are not comparable for $\subseteq_i$ whenever $i \neq j$.[11]

Assume by contradiction that $\langle\!\langle \mathsf{Cycles}, \langle \mathsf{EFO} \rangle_{\mathsf{topo}}, \mathsf{FO} \rangle\!\rangle$ is an lpps. Thanks to Lemma 6.1.13, we conclude that Cycles is a compact subset. Because the compact subsets of the discrete topology are always finite subsets,[12] we conclude that Cycles is a finite subset, which is absurd. ■

10: Which equals $\mathsf{Alex}(\subseteq_i)$ over finite structures

11: The fact that Cycles forms an antichain for $\subseteq_i$ was already at the heart of Example 3.2.1.

12: To convince yourself, consider the following open cover of a compact subset K: $(\{ x \})_{x \in K}$.

The extra assumption of Item 2 is very often satisfied for the usual classes and fragments considered. These were already stated in a sequence of lemmas: Lemma 3.1.8 and Lemmas 3.1.5 and 3.2.8. By combining those results with the correctness theorem (Theorem 6.1.12), we obtain the following correspondences Lemmas 6.1.15 and 6.1.16.

Lemma 6.1.15. *Let σ be a finite relational signature, and F be a fragment of FO. Then, $\langle\!\langle \mathsf{Struct}(\sigma), \langle \exists \mathsf{F} \rangle_{\mathsf{topo}}, \mathsf{FO} \rangle\!\rangle$ is an lpps.*

Proof. We use Theorem 6.1.12 in the direction "Item 1 implies Item 2."

We know that the F-preservation theorem[13] "relativises" to $\mathsf{Struct}(\sigma)$, and the fact that $\exists\mathsf{F}$ sentences define compact open subsets follows immediately from the compactness theorem of first order logic. ■

13: See Theorem 3.1.9 p. 33

Lemma 6.1.16. *Let σ be a finite relational signature, F be one of EFO, $\mathsf{EPFO}^{\neq}$, or EPFO, and let $\mathcal{C}$ be a $\leq_{\mathsf{F}}$-downwards closed subset of $\mathsf{Fin}(\sigma)$. Then, the following are equivalent:*

1. *$\langle\!\langle \mathcal{C}, \langle \exists \mathsf{F} \rangle_{\mathsf{topo}}, \mathsf{FO} \rangle\!\rangle$ is an lpps,*
2. *The F-preservation theorem relativises to $\mathcal{C}$.*

Proof. We first use the equivalence between Item 1 and Item 2 in Theorem 6.1.12. This is almost the statement that we want to prove, except it has two extra assumptions:

- $\langle \exists \mathsf{F} \rangle_{\mathsf{topo}} \cap [\![\mathsf{FO}]\!]_{\mathcal{C}} = \mathsf{Alex}(\leq_{\mathsf{F}}) \cap [\![\mathsf{FO}]\!]_{\mathcal{C}}$,
- $[\![\exists \mathsf{F}]\!]_{\mathcal{C}} \subseteq \mathcal{K}^{\circ}(\mathcal{C}, \langle \exists \mathsf{F} \rangle_{\mathsf{topo}})$.

These two properties holds over $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$, as soon as it is $\leq_{\mathsf{F}}$-downwards closed, and $\mathsf{F} \in \{ \mathsf{EFO}, \mathsf{EPFO}^{\neq}, \mathsf{EPFO} \}$.

Indeed, the first item is implied by the existence of diagram sentences,[14] effectively proving that the upward closure of a point $\mathfrak{A}$ is definable in $\exists\mathsf{F}$.[15]

14: See Lemma 3.1.5.

15: Remark that we even have a stronger statement in this case, namely that $\mathsf{Alex}(\leq_{\mathsf{F}}) = \langle \exists \mathsf{F} \rangle_{\mathsf{topo}}$.

The second item is a consequence of the fact that $\exists\mathsf{F}$ sentences have finitely many non-equivalent $\leq_{\mathsf{F}}$-minimal models. This is true for $\exists\mathsf{F} \in \{ \mathsf{EFO}, \mathsf{EPFO}^{\neq}, \mathsf{EPFO} \}$ because a sentence $\varphi \in \exists\mathsf{F}$ can be rewritten in the form $\exists \vec{x}. \psi(\vec{x})$ where $\psi(\vec{x})$ is quantifier-free. As a consequence, whenever $\mathfrak{A} \models \varphi$, there exists $\mathfrak{B} \subseteq_i \mathfrak{A}$[16] with $|\mathfrak{B}| \leq |\vec{x}|$ such that

16: Which implies $\mathfrak{B} \leq_{\mathsf{F}} \mathfrak{A}$!

$\mathfrak{B} \models \varphi$. We have proven that $\leq_{\mathsf{F}}$-minimal models of φ can always be assumed to have size bounded by $|\vec{x}|$, and therefore[17] that φ has finitely many (non-equivalent) $\leq_{\mathsf{F}}$-minimal models. In particular, $[\![\varphi]\!]_{\mathcal{C}}$ is a compact subset in $\mathsf{Alex}(\leq_{\mathsf{F}})$. ■

17: Recall that the relational signature σ is finite.

We do not see the extra assumptions in Theorem 6.1.12 as a weakness. In fact, these mild restrictions will ensure that lpps exhibit more regular behaviours than simple preservation theorems, and this is at the core of the stability properties that will be derived in Section 6.3 (Stability Properties).

As an example, let us come back over Example 6.1.14 in the light of the recently proven Lemma 6.1.16: we derive that while the EFO-preservation theorem relativises to the class Cycles, it must not relativise to $\downarrow_{\subseteq_i}\mathsf{Cycles}$. From this little adventure, we learn two things. First, that the extra assumption that the class $\mathcal{C}$ is downwards closed in Lemma 6.1.16 cannot be bypassed by considering its downward closure. Second, that Example 6.1.14 is an example of "badly behaved" relativisation of Łoś-Tarski's preservation theorem, which does not happen for topological reasons.[18]

18: This last statement is purposely vague. The intuition that we want to give is that the relativisation of the Łoś-Tarski theorem to Cycles has more to do with the fact that FO, EFO, and $\mathsf{EPFO}^{\neq}$ have the same expressiveness over Cycles, than it has to do with topological notions such as compactness.

The ability to independently consider the topology and the Boolean subalgebra independently over a fixed space X gives an extra flexibility. For instance, given a class $\mathcal{C}$ and a fragment $\mathsf{F} \subseteq \mathsf{FO}$, one can study the set of Boolean subalgebras $\mathcal{B} \subseteq \mathcal{P}(\mathcal{C})$ such that $\langle\!\langle \mathcal{C}, \langle \exists\mathsf{F}\rangle_{\mathsf{topo}}, \mathcal{B} \rangle\!\rangle$ is an lpps. The idea behind this reasoning considering larger algebras $\mathcal{B}$ lead to stronger preservation theorems. This approach quickly sees its limits, as shown in Example 6.1.17.

Example 6.1.17. Let $\mathcal{B} \subseteq \mathcal{P}(\mathsf{Cycles})$. Then, $\langle\!\langle \mathsf{Cycles}, \langle \mathsf{EFO}\rangle_{\mathsf{topo}}, \mathcal{B} \rangle\!\rangle$ is not an lpps.

Proof Sketch. This holds for the same reason as Example 6.1.14, i.e., the space $(\mathsf{Cycles}, \langle \mathsf{EFO}\rangle_{\mathsf{topo}})$ is not compact, hence cannot be an lpps. ■

In order to state categorical results about lpps later on, let us introduce the category LPPS of logically presented pre-spectral spaces with logical maps, which is a full subcategory of LPS. This notation will not be central in the results, but we hope that it can provide valuable insight to people that are already versed in category theory, because some results will feel more natural in this language.

6.2. Connection with Other Spaces

The goal of this section is to explore the connections between lpps and topological spaces, with the hope that these connections allow transporting theorems from general topology to lpps, which in turns can be transformed into properties describing the (non-)relativisation of preservation theorems.

In Subsection 6.2.1, we will continue the work that was started in Section 3.3 and start by analysing how well-quasi-orderings and Noetherian spaces fit the lpps narrative.

Then, in Subsection 6.2.2, we will turn our attention to a weakening of Noetherian spaces that gave the name to lpps: spectral spaces.

6.2.1. Noetherian Spaces and WQO

Recall that Noetherian spaces and well-quasi-orderings were introduced as sufficient conditions for preservation theorems to relativise, respectively in Lemma 3.3.28 and Lemma 3.3.5. It is therefore natural that lpps generalises both, which we will prove hereafter.

Let us first see that the definition of Noetherian spaces can be framed into the realm of lpps by simply considering the Boolean subalgebra of $\mathcal{P}(\mathsf{X})$ to be $\mathcal{P}(\mathsf{X})$ itself.

Lemma 6.2.1. *Let* (X, τ) *be a topological space. The following are equivalent*

1. (X, τ) *is a Noetherian space;*
2. $\langle\!\langle \mathsf{X}, \tau, \mathcal{P}(\mathsf{X}) \rangle\!\rangle$ *is an lpps.*

Proof. In this case, being an lpps amounts to checking the equation $\tau \cap \mathcal{P}(\mathsf{X}) = \mathcal{K}^\circ(\mathsf{X})$, i.e., that every open subset of X is compact, which is equivalent to being a Noetherian space. ■

This has a natural consequence in terms of well-quasi-orderings, because of the connection between the Alexandroff topology and the specialisation preorder.

Corollary 6.2.2. *Let* $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$, $\mathsf{F} \in \{ \mathsf{EFO}, \mathsf{EPFO}^{\neq}, \mathsf{EPFO} \}$. *Then, the following are equivalent*

1. $\langle\!\langle \mathcal{C}, \langle \mathsf{F} \rangle_{\mathsf{topo}}, \mathcal{P}(\mathcal{C}) \rangle\!\rangle$ *is an lpps,*
2. $(\mathcal{C}, \leq_{\mathsf{F}})$ *is a wqo.*

The link between Noetherian spaces, well-quasi-orders and lpps is bittersweet. On the one hand, it allows us to import such spaces *as is* in our framework, and on the other hand, it tells us that those spaces will never be enough to understand when and why preservation theorems relativise.

6.2.2. Spectral Spaces

In search for topological spaces that have weaker properties than being Noetherian, two main families of spaces come to mind. Spectral spaces were introduced by M. H. Stone in the 1930s to study the duals of distributive lattices. They have a deep connection with logic, as noted in [26, Chapter 10]. A second, more general pattern that has been introduced by Kelly in 1963 is the notion of *bitopological space* [63]. These resemble logically presented spaces because they are built as triples $(\mathsf{X}, \tau^+, \tau_-)$ where τ^+ and τ_- are two topologies over X. For clarity and space, we will not review the relationship between bitopological spaces and logically presented spaces, as it mostly follows the one for the spectral spaces.

[26]: Dickmann, Schwartz and Tressl (2019), *Spectral Spaces*

[63]: Kelly (1963), 'Bitopological Spaces'

For this thesis to be self-contained, we restate the definition of a spectral space, following the one given in [26]. However, we claim that understanding every point in the definition is not mandatory to read the rest of this section, and therefore that the details of Definition 6.2.3 can be skipped.

A space X is T_0 whenever every pair of distinct points $x \neq y$ is topologically distinguishable, i.e., there exists an open set U containing x but not y (or vice-versa).

A space X is sober whenever for every non-empty closed and irreducible subset $H \subseteq \mathsf{X}$, there is a unique point $x \in \mathsf{X}$ such that $H = \overline{x}$. You can look at Chapter D (Topology Cheat Sheet) if this feels like random words.

Definition 6.2.3 [26, Definition 1.1.5]**.** A space (X, τ) is *spectral* whenever the following properties are satisfied

1. X is compact and T_0.
2. $\mathcal{K}^\circ(\mathsf{X})$ is a basis of τ.
3. $\mathcal{K}^\circ(\mathsf{X})$ is closed under finite intersections.
4. X is a sober space.

The definition of Definition 6.2.3 is strikingly similar to the one of Definition 6.1.11. We formalise their relationship in Lemma 6.2.4.

Spectral spaces are sober and T_0, which is not always the case of spaces considered up to this point. However, given a topological space (X, τ), one can always build a sober version $\mathcal{S}(\mathsf{X}, \tau)$ using the sobrification functor[19], whose topology does not change, but where points are added and collapsed in a clever way. Similarly, if a topological space (X, τ) is not T_0, one can always build the T_0-quotient of X, obtained by considering the same topology, but collapsing indistinguishable points. As a consequence, these two notions can be safely "discarded."

19: See Definition D.5.4 p. 254

Lemma 6.2.4. *Let (X, τ) be a topological space and $\mathcal{B}$ be a Boolean subalgebra of $\mathcal{P}(\mathsf{X})$. The following are equivalent:*

1. *$\langle\langle \mathsf{X}, \tau, \mathcal{B} \rangle\rangle$ is an lpps;*
2. *$\mathcal{S}(\mathsf{X}, \tau)$ is a spectral space and $\mathcal{K}^\circ(\mathsf{X}) = \tau \cap \mathcal{B}$.*

Proof. If $\langle\langle \mathsf{X}, \tau, \mathcal{B} \rangle\rangle$ is an lpps. Then its sobrification is both T_0 and sober. By assumption, $\mathcal{K}^\circ(\mathsf{X}) = \tau \cap \mathcal{B}$ and $\tau = \langle \tau \cap \mathcal{B} \rangle_{\mathsf{topo}}$. As a consequence, the resulting space is compact, and has a subbasis of compact open subsets, which forms a sublattice of $\mathcal{P}(\mathsf{X})$. Conversely, one concludes that $\tau \cap \mathcal{B}$ is a subbasis of τ, composed of compact open subsets, hence the space is an lpps. ■

One nice application of Lemma 6.2.4 is to transform a spectral space (X, τ) into an lpps, by noticing that $\tau \cap \langle \mathcal{K}^\circ(\mathsf{X}) \rangle_{\mathsf{bool}} = \mathcal{K}^\circ(\mathsf{X})$. As a consequence, $\langle\langle \mathsf{X}, \tau, \langle \mathcal{K}^\circ(\mathsf{X}) \rangle_{\mathsf{bool}} \rangle\rangle$ is an lpps, whenever (X, τ) is a spectral space. Conversely, given an lpps $\langle\langle \mathsf{X}, \tau, \mathcal{B} \rangle\rangle$, $\langle\langle \mathsf{X}, \tau, \langle \mathcal{K}^\circ(\mathsf{X}) \rangle_{\mathsf{bool}} \rangle\rangle$ is always an lpps, with a potentially smaller Boolean subalgebra of definable subsets.

In order to build an intuition, let us see what it means in terms of preservation theorems. The space $\langle\langle \mathsf{Struct}(\sigma), \langle \mathsf{EFO} \rangle_{\mathsf{topo}}, \mathsf{FO} \rangle\rangle$ is an lpps thanks to Lemma 6.1.15. In particular, $\langle\langle \mathsf{Struct}(\sigma), \langle \mathsf{EFO} \rangle_{\mathsf{topo}}, \langle \mathsf{EFO} \rangle_{\mathsf{bool}} \rangle\rangle$ is an lpps. Using Theorem 6.1.12, let us compare these two spaces:

- the first being an lpps boils down to stating that first-order sentences preserved under extensions are equivalent to existential sentences[20];

20: the Łoś-Tarski Theorem

- while the second one states that Boolean combinations of existential sentences preserved under extensions are equivalent to existential sentences, which is a weaker statement.[21]

21: This also illustrates how distinguishing the topology and the logic is needed to perform such fined grained analysis.

While the sobrification functor and the T_0-quotient mostly behave nicely with the usual topological constructions, one must be careful when using theorems about spectral spaces.

The space $\mathsf{Struct}(\sigma)$ is neither T_0 nor sober, but a "fun" non-mandatory exercise, shows that the T_0 quotient of $\mathsf{Struct}(\sigma)$ is a spectral space where the compact open subsets are the existential sentences.

Exercise 6.2.5 (see [70]). The T_0 quotient of $(\mathsf{Struct}(\sigma), \langle \mathsf{EFO} \rangle_{\mathsf{topo}})$ is a sober space.

One reason to study the relationship between lpps and spectral spaces is that relativisation is well-understood for the latter. Indeed, not all subsets of a spectral space are spectral spaces with the induced topology, and we will see a complete characterisation of the spectral subspaces in terms of pro-constructible subsets in Theorem 6.2.8.

Before stating the theorem, we have to spend a little time to properly define what a spectral subspace is. As noted in the introduction of [26, Section 2.1], it is not "enough" that the subset with the induced topology is a spectral space; for the notion to make sense one ought to relate their compact open subsets.

Definition 6.2.6 [26, Definition 2.1.1]**.** Let (X, τ) be a spectral space. A subset $S \subseteq \mathsf{X}$ is a *spectral subspace* of X when $(S, \tau_{|S})$ is a spectral space, and $\mathcal{K}^\circ(\mathsf{X})_{|S} \subseteq \mathcal{K}^\circ(S)$.

To simplify the notation and make the connection with logic apparent, let us write $V \Rightarrow U$ as a shorthand for $U \cup (\mathsf{X} \setminus V)$. Using this notation, we are ready to define the pro-constructible subsets.

Definition 6.2.7 [26, Definition 1.3.11]**.** Let (X, τ) be a topological space. A subset $S \subseteq \mathsf{X}$ is *pro-constructible* whenever there exists a family $(U_i, V_i)_{i \in I}$ of compact open subsets of X, such that

$$S = \bigcap_{i \in I} (U_i \Rightarrow V_i) \quad .$$

Recalling the connection described in Lemma 6.2.4, a spectral space (X, τ) describes in particular an lpps $\langle\!\langle \mathsf{X}, \tau, \langle \mathcal{K}^\circ(\mathsf{X}) \rangle_{\mathsf{bool}} \rangle\!\rangle$. Thinking in terms of preservation theorems, where $\mathcal{K}^\circ(\mathsf{X})$ comes from first-order definable sets, pro-constructible subsets can be interpreted as specific kinds of first-order theories. This intuition will be helpful in the upcoming examples.

Theorem 6.2.8 [26, Theorem 2.3.11]**.** Let (X, τ) be a spectral space, and $S \subseteq \mathsf{X}$ be a subset of X. The following are equivalent:

1. S is a spectral subspace of X;
2. S is pro-constructible.

Because of Theorem 6.2.8, it might be tempting to state that the problem of relativisation of preservation theorems to classes of structures is now closed. However, Example 6.2.9 provides an example of a subset that is an lpps and is not pro-constructible, while Example 6.2.10 provides an example of a pro-constructible subset that does not induce an lpps.

Both Examples 6.2.9 and 6.2.10 are possible because of two crucial differences between spectral spaces and lpps: the first one is the separation axioms (sober, T_0), and the second one is the fact that spectral spaces "do not care" about the algebra of definable subsets, and only ever deals with the Boolean subalgebra $\langle \mathcal{K}^\circ(\mathsf{X}) \rangle_{\mathsf{bool}}$.

Recall that Δ^2_{deg} denotes the class of finite graphs of degree bounded by 2. This class is hereditary, closed under disjoint unions, and locally finite.

Example 6.2.9. The space $\langle\!\langle \mathsf{Struct}(\sigma), \langle \mathsf{EFO} \rangle_{\mathsf{topo}}, \mathsf{FO} \rangle\!\rangle$ is an lpps, and $\langle\!\langle \Delta^2_{\text{deg}}, \langle \mathsf{EFO} \rangle_{\mathsf{topo}}, \mathsf{FO} \rangle\!\rangle$ is an lpps, but the latter is not a pro-constructible subset of the former when seen as topological spaces.

Proof. The two spaces are lpps. Indeed, $\mathsf{Struct}(\sigma)$ is an lpps thanks to Lemma 6.1.15, while Δ^2_{deg} is lpps because it is locally finite, hence one can leverage Theorem 5.1.2 to conclude that the Łoś-Tarski Theorem relativises to $\mathcal{C}$, and conclude using Lemma 6.1.16.

Assume by contradiction that Δ^2_{deg} is obtained as an intersection

$$H \stackrel{\text{def}}{=} \bigcap_{i \in I} (U_i \Rightarrow V_i) \quad ,$$

where U_i and V_i are compact open subsets of $\mathsf{Struct}(\sigma)$.[22]

Let $\mathfrak{A} \in \Delta^2_{\text{deg}}$ (which is non-empty), and let us define by induction on $k \in \mathbb{N}$ $\mathfrak{A}_0 \stackrel{\text{def}}{=} \mathfrak{A}$ and $\mathfrak{A}_{k+1} \stackrel{\text{def}}{=} \mathfrak{A} \uplus \mathfrak{A}_k$. Finally, let us define $\mathfrak{A}_\omega \stackrel{\text{def}}{=} \biguplus_{i \in \omega} \mathfrak{A}$. Notice that for all $k \in \mathbb{N}$, $\mathfrak{A}_k \in \Delta^2_{\text{deg}}$, but that $\mathfrak{A}_\omega \notin \Delta^2_{\text{deg}}$ because it is an infinite structure.[23]

Let us prove that $\mathfrak{A}_\omega \in H$, which will lead to a contradiction. For that, let us consider $i \in I$, and prove that $\mathfrak{A}_\omega \in U_i \Rightarrow V_i$ by a simple case analysis.

- If there exists a $k \in \mathbb{N}$ such that $\mathfrak{A}_k \in V_i$, then $\mathfrak{A}_\omega \in V_i$ because the latter is upwards closed for $\subseteq_i$ and $\mathfrak{A}_k \subseteq_i \mathfrak{A}_\omega$.
- Otherwise, for all $k \in \mathbb{N}$, $\mathfrak{A}_k \notin V_i$. Because $\mathfrak{A}_k \in H$, this implies that $\mathfrak{A}_k \notin U_i$ for all $k \in \mathbb{N}$.
  Recall that the ambient space is an lpps, and in particular that U_i can be seen as the set of models of an existential first order sentence $\exists \vec{x}.\theta(\vec{x})$, where θ is quantifier-free. In particular, we just have to prove that $\mathfrak{A}_\omega \models \forall \vec{x}.\neg\theta(\vec{x})$ to conclude that $\mathfrak{A}_\omega \notin U_i$.
  Now, consider a valuation $\upsilon \colon \vec{x} \to \mathfrak{A}_\omega$, it takes values in finitely many of the copies of $\mathfrak{A}$, hence can be seen as a function $\upsilon \colon \vec{x} \to \mathfrak{A}_p$ for some large enough $p \in \mathbb{N}$. Because $\mathfrak{A}_p \models \forall \vec{x}.\neg\theta(\vec{x})$, we

22: An alternative and shorter proof of the contradiction is to notice that H is in fact a first order theory that has finite models of arbitrary size, hence must have an infinite model, which is absurd because Δ^2_{deg} is composed only of finite models. To perform the shift from "finite models of arbitrary size" to "infinite model," it suffices to use the compactness theorem of first order logic. We have chosen to provide a more combinatorial proof that relies on the syntactic shape of the sentences to avoid the use of the compactness theorem of first order logic.

23: This is true because $|\mathfrak{A}| > 0$, since a model is always non-empty.

conclude that $\mathfrak{A}_p, \upsilon \models \neg\theta(\vec{x})$. Since the formula θ is quantifier-free, it entails that $\mathfrak{A}_\omega, \upsilon \models \neg\theta(\vec{x})$.
We have proven that $\mathfrak{A}_\omega \notin U_i$. ■

Example 6.2.10. The subset $\mathsf{Cycles} \subseteq \Delta^2_{\mathrm{deg}}$ is a pro-constructible subset of $(\Delta^2_{\mathrm{deg}}, \mathsf{Alex}(\subseteq_i))$, but $\langle\langle \mathsf{Cycles}, \mathsf{Alex}(\subseteq_i), \mathsf{FO} \rangle\rangle$ is not an lpps.

Proof. Remark that $\downarrow_{\subseteq_i} C_n$ is (by definition) a closed subset of Δ^2_{deg}, and that it is definable in $\mathsf{FO}[\sigma]$, because $\downarrow_{\subseteq_i} C_n$ is a finite subset of $\mathsf{Fin}(\sigma)$.

As a consequence, the following subsets are open and definable for all $n \geq 3$:
$$U_n \stackrel{\text{def}}{=} \left(\Delta^2_{\mathrm{deg}} \setminus \downarrow_{\subseteq_i} C_n\right) \cup \uparrow_{\subseteq_i} C_n \quad .$$

Because $\langle\langle \Delta^2_{\mathrm{deg}}, \mathsf{Alex}(\subseteq_i), \mathsf{FO} \rangle\rangle$ is an lpps,[24] we conclude that U_n is a compact open subset for every $n \geq 3$.

24: This was stated in Example 6.2.9.

Furthermore, let us notice that for all $n \geq 3$, $p \geq 1$, $V_{n,p} \stackrel{\text{def}}{=} \uparrow_{\subseteq_i} (C_n \uplus P_p)$ is a definable open subset of Δ^2_{deg}. As before, we conclude that these subsets are compact open subsets.

Let us now remark that Cycles can be defined as an infinite intersection inside Δ^2_{deg} as follows:
$$\mathsf{Cycles} = \bigcap_{n \geq 3} U_n \cap \bigcap_{n \geq 3, p \geq 1} [V_{n,p} \Rightarrow \emptyset] \quad .$$

Hence, Cycles is a pro-constructible subset of Δ^2_{deg} when the latter is seen as a topological space. However, we have seen in Example 6.1.17 $\langle\langle \mathsf{Cycles}, \mathsf{Alex}(\subseteq_i), \mathsf{FO} \rangle\rangle$ is not an lpps. ■

6.3. Stability Properties

This section is devoted to the study of stability properties of lpps. The goal is to obtain an algebra akin to the one for wqos or Noetherian spaces.[25]

25: These were listed in Table 6.1 and in Table 6.2.

Unsurprisingly, we will first focus on the behaviour of lpps with respect to subspaces, as it has direct connections with the relativisation of preservation theorems. The study of subspaces of lpps was already used in Lemma 4.3.20 on page 105. We will then turn our attention to morphisms of lpps, which are of particular interest since they too were already at play in Example 5.1.10 on page 124.[26]

26: These results were used in Chapter 5 (A Local-to-Global Preservation Theorem). While their use was not *stricto sensu* necessary at that time, it severed the dual purpose of providing short proofs and showcasing the stability properties introduced in this chapter.

Then, we will complete our algebraic study by providing a list of operations preserving lpps. Some of these are quite usual (disjoint sums, product spaces), while some have a more involved definition and crucially rely on the fact that we consider spaces of finite structures (the wreath product).

Overall, the stability properties proven in this section enable us to reason compositionally about preservation theorems, something which was lacking in the literature. Not only does this provide for a nice framework to infer or refute preservation theorems, but it allows us to go beyond the relativisation results proven using locality techniques.

6.3.1. Subspaces and Morphisms

Because one of the main issues with preservation theorems is the lack of relativisation to subsets, let us explore what happens in the case of logically presented pre-spectral spaces.

Recall from Lemma 6.1.13 that lpps are compact spaces. As a consequence, for a subspace to be an lpps, a necessary condition is that it is compact. However, we will see in Example 6.3.1 that this condition is not sufficient, as it was already hinted at in Example 6.2.10.

Example 6.3.1. The class $\downarrow_{\subseteq_i}\mathsf{Cycles}$ is a closed and compact subset of the class Δ^2_{deg} of finite graphs of degree bounded by 2. As remarked in Example 6.2.9, the latter is an lpps, while the former is not.

A simple sufficient condition for a subset to define an lpps is that it is obtained as a compact open subset, or the complement of a compact open subset. Furthermore, the property of being a sub-lpps is stable under finite unions and intersections, which automatically leads to the following lemma.

Lemma 6.3.2. *Let $\langle\!\langle \mathsf{X}, \tau, \mathcal{B} \rangle\!\rangle$ be an lpps, and $Y \in \langle \mathcal{K}^\circ(\mathsf{X}) \rangle_{\mathsf{bool}}$. Then Y with the induced topology and the induced Boolean algebra is an lpps.*

Proof. This is because the intersection of a closed subset and a compact subset is always compact, and because $\mathcal{K}^\circ(\mathsf{X})$ is stable under finite intersections. ■

Let us notice by "upcycling" Example 6.2.9, that some subspaces are lpps and do not belong to $\langle \mathcal{K}^\circ(\mathsf{X}) \rangle_{\mathsf{bool}}$.

Lemma 6.3.3. *Let $\langle\!\langle \mathsf{X}, \tau, \mathcal{B} \rangle\!\rangle$ and $\langle\!\langle \mathsf{X}', \tau', \mathcal{B}' \rangle\!\rangle$ be two lps, and Let f be a surjective logical map*[27] *from $\langle\!\langle \mathsf{X}, \tau, \mathcal{B} \rangle\!\rangle$ to $\langle\!\langle \mathsf{X}', \tau', \mathcal{B}' \rangle\!\rangle$. If X is an lpps, then so is X'.*

27: See Definition 6.1.8 p. 142

Proof. Let $U \in \mathcal{B}' \cap \tau'$. Then, $f^{-1}(U) \in \mathcal{B} \cap \tau = \mathcal{K}^\circ(\mathsf{X})$. Because f is continuous, $f(f^{-1}(U))$ is compact. Furthermore, $f(f^{-1}(U)) = U$ because f is surjective. ■

A concrete example of logical maps is obtained via first-order interpretations, which is a formalism already used in Example 5.1.10 on page 124. Recall that (simple) first order interpretations were defined along with the evaluation of queries in Definition 2.2.9.[28]

28: In particular, first order interpretations define *functions* between classes of structures.

Fact 6.3.4. A first-order interpretation I between $\langle\langle \mathcal{C}, \tau, \mathsf{FO}[\sigma] \rangle\rangle$ and $\langle\langle \mathcal{C}', \tau', \mathsf{FO}[\sigma'] \rangle\rangle$ is a logical map if and only if it is continuous.

Example 6.3.5. The first order interpretation from finite graphs to finite graphs defined using $\varphi_{\mathsf{dom}}(x) \stackrel{\text{def}}{=} \top$ and $\varphi_E(x,y) \stackrel{\text{def}}{=} E(x,y) \vee \exists z.E(x,z) \wedge E(z,y)$ is a logical map when both spaces are endowed with $\mathsf{Alex}(\subseteq_i)$.

6.3.2. Closure under Finitary Operations

As promised, we are now going to slowly but surely prove the compositional results listed in Table 6.3, starting with the simple constructors that are finite sums and finite products of spaces.

In order to be readable even without background knowledge in category theory, we introduce hereafter concrete definitions for the disjoint sums and products of spaces, even though these have abstract categorical definitions.

The sum topology and sum algebra are respectively defined in Definition D.2.3 and Definition B.3.5. Both of them are generated by the (arbitrary) unions of open subsets (resp. definable subsets) of the summands.

Definition 6.3.6. *Let $\langle\langle \mathsf{X}_i, \tau_i, \mathcal{B}_i \rangle\rangle$ be a family of lpps for $i \in I$. Then, the sum is defined as*

$$\sum_{i\in I} \langle\langle \mathsf{X}_i, \tau_i, \mathcal{B}_i \rangle\rangle \stackrel{\text{def}}{=} \langle\langle \sum_{i\in I} \mathsf{X}_i, \sum_{i\in I} \tau_i, \sum_{i\in I} \mathcal{B}_i \rangle\rangle \quad .$$

In the hope that a categorical presentation provides some insight, let us state formally that this is what the reader expects from a disjoint sum.

Fact 6.3.7. That is, the *sum* is the co-product in the category LPS.

We are now ready to use a compositional technique to build new lpps for the first time.

Lemma 6.3.8. *Finite sums of lpps are lpps.*

Proof. By definition, the sum is an lps. It remains to prove that definable open subsets are compact. This holds because a definable open subset U of $\sum_{1\leq i\leq n} \mathsf{X}_i$ is a disjoint union of definable open subsets over each of the X_i's. Hence, U is a finite union of compact open subsets,[29] hence is compact (see Lemma D.4.3). ■

29: We use the fact that an open subset (resp. compact subset) of a summand is an open subset (resp. compact subset) in the sum.

Corollary 6.3.9. *Finite co-products exists in LPPS and are computed as in LPS.*

Although the category LPS has all co-products, the restriction to finite sums in Lemma 6.3.8 is necessary, as witnessed by the following counter-example.

Example 6.3.10. Let $X \overset{\text{def}}{=} \langle\!\langle 1, \tau_{\mathsf{disc}}, \mathcal{P}(1) \rangle\!\rangle$. The space X is an lpps, but $\sum_{i \in \mathbb{N}} X$ is not an lpps.

Proof Sketch. Use Lemma 6.1.13, and the fact that $\sum_{i \in \mathbb{N}} X$ is $\mathbb{N}$ with the discrete topology, hence not compact. ∎

Let us now turn our attention to products of lpps.

Definition 6.3.11. *Let $\langle\!\langle \mathsf{X}_i, \tau_i, \mathcal{B}_i \rangle\!\rangle$ be a sequence of lpps for $i \in I$. Then, the product is defined as*

$$\prod_{i \in I} \langle\!\langle \mathsf{X}_i, \tau_i, \mathcal{B}_i \rangle\!\rangle \overset{\text{def}}{=} \langle\!\langle \prod_{i \in I} \mathsf{X}_i, \prod_{i \in I} \tau_i, \prod_{i \in I} \mathcal{B}_i \rangle\!\rangle \quad .$$

We deferred the definition of the product topology and product algebra respectively to Definition D.2.2 and Definition B.3.4. These are generated by products of open subsets (resp. definable subsets) that have *finitely many* components that are not the whole space. That is, the subbasic sets are of the form $\prod_{i \in I} A_i$, such that $A_i = \mathsf{X}_i$ for all but finitely many indices $i \in I$, and such that A_i is an open subset of X_i for all $i \in I$ (in the case of the product topology), or such that A_i is a definable subset of X_i (in the case of the product algebra).

Unsurprisingly, we have (again) defined the categorical product in the category LPS.

Fact 6.3.12. The *product* is the product in the category LPS.

The following theorem stating that the category LPPS has arbitrary products is non-trivial. In particular, it is not a direct consequence of the connection between lpps and spectral spaces that was described in Lemma 6.2.4.

Theorem 6.3.13. *The product of a family of lpps is an lpps.*

Proof. Let us first do the proof in the case of a product of two lpps $\langle\!\langle \mathsf{X}, \tau_{\mathsf{X}}, \mathcal{B}_{\mathsf{X}} \rangle\!\rangle$ and $\langle\!\langle \mathsf{Y}, \tau_{\mathsf{Y}}, \mathcal{B}_{\mathsf{Y}} \rangle\!\rangle$.

It is clear that $\mathsf{X} \times \mathsf{Y}$ is a logically presented space. The only thing left to check is that definable open subsets of $\mathsf{X} \times \mathsf{Y}$ are compact.

Let U be a definable open subset of $\mathsf{X} \times \mathsf{Y}$. Because it is a definable subset,[30] there exists a finite sequence $\left(D^{\mathsf{X}}_{i,j}\right)_{1 \leq i,j \leq n}$ of definable subsets of X, and a finite sequence $\left(D^{\mathsf{Y}}_{i,j}\right)_{1 \leq i,j \leq n}$ of definable subsets of Y, such that

$$U = \bigcap_{1 \leq i \leq n} \bigcup_{1 \leq j \leq n} D^{\mathsf{X}}_{i,j} \times D^{\mathsf{Y}}_{i,j} \quad . \tag{6.2}$$

30: Recall that the product algebra is defined precisely in terms of finite Boolean combinations. Furthermore, $(\mathsf{X} \times \mathsf{Y}) \setminus (D^{\mathsf{X}} \times D^{\mathsf{Y}})$ can be rewritten as $(\mathsf{X} \setminus D^{\mathsf{X}}) \times \mathsf{Y} \cup \mathsf{X} \times (\mathsf{Y} \setminus D^{\mathsf{Y}})$, which explains the lack of negations in Equation 6.2.

Let us define over the set X the following equivalence relation: $x \equiv_{\mathsf{X}} x'$ if and only if $x \in D^{\mathsf{X}}_{i,j} \iff x' \in D^{\mathsf{X}}_{i,j}$ for all $1 \leq i,j \leq n$, that is, the equivalence relation of finite index induced by the sets $\left(D^{\mathsf{X}}_{i,j}\right)_{1 \leq i,j \leq n}$. We define $y \equiv_{\mathsf{Y}} y'$ similarly over Y.

Because of Equation 6.2, we conclude that U is saturated for the equivalence relation $\equiv_{\mathsf{X}} \times \equiv_{\mathsf{Y}}$.

Let $(x, y) \in U$. Because U is an open subset in the product topology, there exist $V_{\mathsf{X}}^{x,y} \in \tau_{\mathsf{X}}$, and $V_{\mathsf{Y}}^{x,y} \in \tau_{\mathsf{Y}}$, such that $(x, y) \in V_{\mathsf{X}}^{x,y} \times V_{\mathsf{Y}}^{x,y} \subseteq U$. Our goal is to show that one can always assume that $V_{\mathsf{X}}^{x,y} \times V_{\mathsf{Y}}^{x,y}$ are also saturated for $\equiv_{\mathsf{X}} \times \equiv_{\mathsf{Y}}$. This will allows us to conclude that U is compact as follows:

1. Rewrite U as the union of its open neighbourhoods
$$U = \bigcup_{(x,y)\in U} V_{\mathsf{X}}^{x,y} \times V_{\mathsf{Y}}^{x,y} \quad .$$
2. Because $\equiv_{\mathsf{X}} \times \equiv_{\mathsf{Y}}$ is of finite index, this union is actually finite.
3. Because $V_{\mathsf{X}}^{x,y}$ is saturated for $\equiv_{\mathsf{X}}$, it is in particular a definable subset of X, hence it is compact. Similarly, $V_{\mathsf{Y}}^{x,y}$ is also compact.
4. Finally, we have shown that U is a finite union of products of compact subsets, hence U is compact via the Tychonoff theorem and Lemma D.4.3.

Let $(x, y) \in U$, and let us consider (using Zorn's Lemma) $V_{\mathsf{X}}^{x,y} \in \tau_{\mathsf{X}}$ and $V_{\mathsf{Y}}^{x,y} \in \tau_{\mathsf{Y}}$, maximal for inclusion, such that $(x, y) \in V_{\mathsf{X}}^{x,y} \times V_{\mathsf{Y}}^{x,y} \subseteq U$. We argue that $V_{\mathsf{X}}^{x,y} \times V_{\mathsf{Y}}^{x,y}$ is saturated for $\equiv_{\mathsf{X}} \times \equiv_{\mathsf{X}}$. Assume by contradiction that it is not, and without loss of generality, that there exists $y \in V_{\mathsf{Y}}^{x,y}$, and $y' \notin V_{\mathsf{Y}}^{x,y}$ such that $y \equiv_{\mathsf{Y}} y'$.

Let $F \subseteq_{\mathsf{fin}} V_{\mathsf{X}}^{x,y}$ be a choice of representatives for every $\equiv_{\mathsf{X}}$-equivalence class present in $V_{\mathsf{X}}^{x,y}$. That is, for all $x' \in V_{\mathsf{X}}^{x,y}$, there exists a unique $x'' \in F$ such that $x' \equiv_{\mathsf{X}} x''$.

For every $x'' \in F$, $(x'', y) \in U$. Because U is saturated for $\equiv_{\mathsf{X}} \times \equiv_{\mathsf{Y}}$ and $y \equiv_{\mathsf{Y}} y'$, $(x'', y') \in U$ too. Therefore, there exists $V_{\mathsf{X}}^{x'',y'} \in \tau_{\mathsf{X}}$, $V_{\mathsf{Y}}^{x'',y'} \in \tau_{\mathsf{Y}}$ such that $(x'', y') \in V_{\mathsf{X}}^{x'',y'} \times V_{\mathsf{Y}}^{x'',y'} \subseteq U$.

Let $W \stackrel{\mathsf{def}}{=} \bigcap_{x''\in F} V_{\mathsf{Y}}^{x'',y'}$. By construction, $y' \in W$, and because the intersection is finite, W is an open subset of Y.

We obtain $V_{\mathsf{X}}^{x,y} \times (V_{\mathsf{Y}}^{x,y} \cup W)$ contains (x, y), and strictly contains $V_{\mathsf{X}}^{x,y} \times V_{\mathsf{Y}}^{x,y}$. Let us now derive a contradiction by showing that $V_{\mathsf{X}}^{x,y} \times (V_{\mathsf{Y}}^{x,y} \cup W) \subseteq U$. Note that the only inclusion to prove is actually that $V_{\mathsf{X}}^{x,y} \times W \subseteq U$. For that, let us consider $(a, b) \in V_{\mathsf{X}}^{x,y} \times W$. By definition of F, there exists $x'' \in F$ such that $a \equiv_{\mathsf{X}} x''$. Because $b \in W$, in particular $b \in V_{\mathsf{Y}}^{x'',y'}$, and by definition of the latter $(x'', b) \in U$. Because U is saturated for $\equiv_{\mathsf{X}} \times \equiv_{\mathsf{Y}}$, and $a \equiv_{\mathsf{X}} x''$, we conclude that $(a, b) \in U$.

Note that we can obtain finitary products by iterating the construction.[31]

31: This induction is sound because the empty product is the lps $\langle\langle \emptyset, \{\emptyset\}, \{\emptyset\} \rangle\rangle$ which is an lpps, and the product of lpps is associative.

For infinite products, it suffices to notice that a definable open subset is not the entire space on finitely many components, and therefore the infinite case reduces to the finite one. Let us formalise this intuition hereafter.

Let $(\mathsf{X}_i)_{i\in I}$, $(\tau_i)_{i\in I}$ and $(\mathcal{B}_i)_{i\in I}$ be such that
$$\forall i \in I, \langle\langle \mathsf{X}_i, \tau_i, \mathcal{B}_i \rangle\rangle \text{ is an lpps} \quad .$$

Let D be a definable open subset of the product of the family of lpps above. By definition of the product algebra,[32] there exists $n \in \mathbb{N}$, and

32: As for the case of the product of two spaces, we use the fact that the complement of a product having finitely many components that are not the entire space is a finite union of products of such sets.

a family $(D^i_j)_{i \in I, 1 \leq j \leq n}$, such that D is saturated for the following equivalence relation:

$$(x \equiv y) \iff \forall i \in I, \forall 1 \leq j \leq n, \underbrace{\left[x_i \in D^i_j \iff y_i \in D^i_j\right]}_{\overset{\text{def}}{=} (x_i \equiv_i y_i)} \quad .$$

Remark that the relation $\equiv$ is of finite index because $D^i_j \neq \mathsf{X}_i$ for finitely many $i \in I$ and $1 \leq j \leq n$. Let us define J as a finite subset of I as follows:

$$J \overset{\text{def}}{=} \{ i \in I : \exists 1 \leq j \leq n, D^i_j \neq \mathsf{X}_i \} \quad .$$

This allows us to define D' as the projection of D over $\prod_{i \in J} \mathsf{X}_i$, that is:

$$D' \overset{\text{def}}{=} \{ (x_i)_{i \in J} : x \in D \} \quad .$$

Now, D' is a definable open subset of $\prod_{i \in J} \mathsf{X}_i$. Indeed, the image of an open subset through a projection map remains open, and D' is saturated for the product of the relations $\prod_{i \in J} \equiv_i$, hence is definable. Since we have proven above that a *finite* product of lpps is an lpps, we conclude that D' is a compact open subset.

To conclude that D is a compact open subset, it suffices to notice that X_i is compact for all $i \in I$, and therefore that we can rewrite D as a product of compact spaces, which is compact thanks to the Tychonoff theorem:[33]

$$D \simeq \prod_{i \in (I \setminus J)} \mathsf{X}_i \times D' \qquad \blacksquare$$

33: We use the symbol $\simeq$ in the equation because one has to reorder the components of the product to obtain the desired equality, but considering the product as written felt clearer to me.

Corollary 6.3.14. *The category* LPPS *has all products, and they are computed as in* LPS*.*

The statements of Lemma 6.3.8 and Theorem 6.3.13 are not quite satisfactory to study preservation theorems, as the Boolean subalgebra that appears is not FO. The purpose of the upcoming Lemmas 6.3.15 and 6.3.16 is to prove that in the specific cases where the original Boolean subalgebra were defined using first order logic, the resulting ones are too.

Before getting back to concrete preservation theorems, let us first prevent any “wishful thinking” from the reader inclined in category theory: even though LPPS has all products, it lacks equalisers, and we will provide in Subsection 6.3.4 on page 174 examples of limits that are missing in LPPS.

The following Lemmas 6.3.15 and 6.3.16 are obtained via Feferman-Vaught style compositional techniques [33, 74], which were introduced in Chapter 4 (Locality and Preservation).[34]

34: This was the key ingredient in the proof of Lemma 4.2.9.

We write $\tau +_t \theta$ for the sum topology in order to distinguish it from the sum of two lpps in the equations. Similarly, we use $\mathcal{B} +_b \mathcal{B}'$ for the sum Boolean algebra.

Lemma 6.3.15. *Let* $X \overset{\text{def}}{=} \langle\!\langle \mathcal{C}, \tau, \mathsf{FO}[\sigma] \rangle\!\rangle$ *and* $X' \overset{\text{def}}{=} \langle\!\langle \mathcal{C}', \tau', \mathsf{FO}[\sigma'] \rangle\!\rangle$ *be two lps. Then,* $X + X'$ *is isomorphic to* $\langle\!\langle \mathcal{C} \uplus \mathcal{C}', \tau +_t \tau', \mathsf{FO}[\sigma''] \rangle\!\rangle$, *where* $\sigma'' \overset{\text{def}}{=} \sigma \uplus \sigma' \{ (L, 1), (R, 1) \}$, *and* L, R *are unary predicates used to distinguish the left-hand and the right-hand part of the disjoint union.*

Proof Sketch. Notice that the first inclusion $\mathsf{FO}[\sigma] +_b \mathsf{FO}[\sigma'] \subseteq \mathsf{FO}[\sigma'']$ is trivial. For the converse inclusion, it suffices to remark that every sentence φ is equivalent over $\mathcal{C} \uplus \mathcal{C}'$ to a disjunction $(\exists x. L(x) \wedge \varphi_1) \vee (\exists x. R(x) \wedge \varphi_2)$ where $\varphi_1 \in \mathsf{FO}[\sigma]$ and $\varphi_2 \in \mathsf{FO}[\sigma']$. ■

Lemma 6.3.16 is the use of the Feferman-Vaught technique that is the closest match to its "original" formulation in terms of the logic on ultra-products of spaces.

Lemma 6.3.16. *Let* $X \overset{\text{def}}{=} \langle\!\langle \mathcal{C}, \tau, \mathsf{FO}[\sigma] \rangle\!\rangle$ *and* $X' \overset{\text{def}}{=} \langle\!\langle \mathcal{C}', \tau', \mathsf{FO}[\sigma'] \rangle\!\rangle$ *be two lps. Then,* $X \times X'$ *is isomorphic to* $\langle\!\langle \{ \mathfrak{A} \uplus \mathfrak{B} : \mathfrak{A} \in \mathcal{C}, \mathfrak{B} \in \mathcal{C}' \}, \tau \times_t \tau', \mathsf{FO}[\tau] \rangle\!\rangle$, *where* $\tau \overset{\text{def}}{=} \sigma \uplus \sigma' \{ (L, 1), (R, 1) \}$, *and* L, R *are unary predicates used to distinguish the left-hand and the right-hand part in the disjoint unions.*

Proof Sketch. Notice that the first inclusion $\mathsf{FO}[\sigma] \times_b \mathsf{FO}[\sigma'] \subseteq \mathsf{FO}[\sigma \uplus \sigma' \uplus \{ (L, 1), (R, 1) \}]$ is trivial.

Furthermore, given a first-order sentence $\psi(\vec{x}) \in \mathsf{FO}[\tau]$, we can use the Feferman-Vaught technique with $\mathbb{T} \overset{\text{def}}{=} \{ T_L, T_R \}$. Then, Lemma 4.2.17 actually proves that for all $\mathfrak{A}, \upsilon : \vec{x} \to \mathfrak{A}$, for all $\rho : \mathfrak{A} \to \mathbb{T}$, for all $\mathbb{T}$-typing environment $\Gamma : \vec{x} \to \mathbb{T}$ such that $\mathfrak{A}, \upsilon \models L(x)$ if and only if $\rho(\upsilon(x)) = T_L$, $\mathfrak{A}, \upsilon \models R(x)$ if and only if $\rho(\upsilon(x)) = T_R$, and $\rho(\upsilon(x)) = \Gamma(x)$ for all $x \in \vec{x}$, then $\mathfrak{A}, \upsilon \models \psi(\vec{x})$ if and only if $\mathfrak{A}, \rho\upsilon \models^{\mathbb{T}} \mathrm{conv}^{\mathbb{T}}(\Gamma, \psi)$.

As a consequence of Lemma 4.2.19, every $\mathbb{T}$-typed formula is equivalent to a finite Boolean combination of monotyped formulas. In particular, to check whether $\mathfrak{A} \uplus \mathfrak{A}' \in \mathcal{C} \times \mathcal{C}'$ satisfies φ, it suffices to evaluate the T_L properties on $\mathfrak{A}$, and the T_R properties on $\mathfrak{A}'$. This is clearly definable in $\mathsf{FO}[\sigma] \times_b \mathsf{FO}[\sigma']$. ■

Let us now demonstrate how the two composition theorems that we have just presented, on finite sums and on products allows us to study actual preservation theorems.

Example 6.3.17. Let $\mathcal{C} \overset{\text{def}}{=} \{ \mathfrak{A} \uplus \mathfrak{B} : \mathfrak{A} \in \mathsf{Cliques}, \mathfrak{B} \in \mathsf{Paths} \}$. Then, the Łoś-Tarski Theorem relativises to $\mathcal{C}$.

Proof Sketch. Thanks to Lemma 6.3.16, it suffices to prove that "erasing" the colours distinguishing the left-hand side and right-hand side of the disjoint unions preserves the property of being an lpps. Formally, we want to prove that the map $\mathsf{I} : \mathsf{Cliques} \times \mathsf{Paths} \twoheadrightarrow \mathcal{C}$ is a logical map and then leverage Lemma 6.3.3. Because it already is defined as a first order interpretation, it suffices to check that it is monotone thanks to Fact 6.3.4. The latter is immediate because removing unary predicates preserves the property of being an induced substructure. ■

There is another natural operation on structures that is not yet considered in our study: the cartesian product. Beware that $\mathcal{C} \times \mathcal{C}'$ is a class of pairs, or equivalently a class of disjoint unions of structures in $\mathcal{C}$ and $\mathcal{C}'$. We now want to build the class of $\mathfrak{A} \times \mathfrak{B}$, where $\mathfrak{A} \in \mathcal{C}$ and $\mathfrak{B} \in \mathcal{C}'$.

Definition 6.3.18. *Let $\mathcal{C} \subseteq \mathsf{Struct}(\sigma)$ and $\mathcal{C}' \subseteq \mathsf{Struct}(\sigma')$ be two classes of structures. The inner product $\mathcal{C} \otimes \mathcal{C}'$ is the class of products $\mathfrak{A} \times \mathfrak{B}$ for $\mathfrak{A} \in \mathcal{C}$ and $\mathfrak{B} \in \mathcal{C}'$.*

An element $\mathfrak{A} \times \mathfrak{B} \in \mathcal{C} \otimes \mathcal{C}'$ is equivalently described as a structure in $\mathsf{Struct}(\sigma \uplus \sigma')$, such that

- *a tuple $((a_i, b_i))_{1 \leq i \leq n}$ belongs to a $R^{\mathfrak{A} \times \mathfrak{B}}$ for relation $(R, n) \in \sigma$ if and only if $(a_i)_{1 \leq i \leq n}$ belongs to $R^{\mathfrak{A}}$, and*
- *a tuple $((a_i, b_i))_{1 \leq i \leq n}$ belongs to a $R'^{\mathfrak{A} \times \mathfrak{B}}$ for relation $(R', n) \in \sigma'$ if and only if $(b_i)_{1 \leq i \leq n}$ belongs to $R'^{\mathfrak{B}}$.*

In this setting, the equality relation $=_\sigma$ and $=_{\sigma'}$ become equivalence relations on the product structure $\mathfrak{A} \times \mathfrak{B}$.

In order to define the inner product of lpps, we need to construct topologies and Boolean subalgebras over the inner product. Unsurprisingly, one defines the *inner product* $\tau \otimes \tau' \stackrel{\text{def}}{=} \langle \{ U \otimes V : U \in \tau, V \in \tau' \} \rangle_{\mathsf{topo}}$ and the *inner product* $\mathcal{B} \otimes \mathcal{B}' \stackrel{\text{def}}{=} \langle \{ D \otimes D' : D \in \tau, D' \in \tau' \} \rangle_{\mathsf{bool}}$. This allows us to define the *inner product* of two lpps as follows: $\langle\langle \mathcal{C}, \tau, \mathcal{B} \rangle\rangle \otimes \langle\langle \mathcal{C}', \tau', \mathcal{B}' \rangle\rangle$ as $\langle\langle \mathcal{C} \otimes \mathcal{C}', \tau \otimes \tau', \mathcal{B} \otimes \mathcal{B}' \rangle\rangle$.

Lemma 6.3.19. *Let $\mathsf{X} \stackrel{\text{def}}{=} \langle\langle \mathcal{C}, \tau, \mathcal{B} \rangle\rangle$ and $\mathsf{X}' \stackrel{\text{def}}{=} \langle\langle \mathcal{C}', \tau', \mathcal{B}' \rangle\rangle$ be two lpps. Then, $\mathsf{X} \otimes \mathsf{X}'$ is an lpps.*

Proof. Let us define $f : \mathsf{X} \times \mathsf{X}' \twoheadrightarrow \mathsf{X} \otimes \mathsf{X}'$ defined by $f : (\mathfrak{A}, \mathfrak{B}) \mapsto \mathfrak{A} \times \mathfrak{B}$. Clearly f is surjective, and it is an easy check that it is a logical map. We conclude by Lemma 6.3.3. ■

To save some space and because the following is another application of the same proof scheme using Feferman-Vaught techniques, we leave the following property as an exercise.[35]

35: The idea behind Exercise 6.3.20 is that the first order theory of a product $\mathfrak{A} \times \mathfrak{B}$ is uniquely determined by the first order theories of $\mathfrak{A}$ and $\mathfrak{B}$.
We write $\vee_t$ for the join of two topologies, to distinguish it from the symbol $\vee$ that is already used quite a lot in this manuscript.

Exercise 6.3.20. Let $\langle\langle \mathcal{C}, \langle \exists \mathsf{F} \rangle_{\mathsf{topo}}, \mathsf{FO}[\sigma] \rangle\rangle$ and $\langle\langle \mathcal{C}', \langle \exists \mathsf{F}' \rangle_{\mathsf{topo}}, \mathsf{FO}[\sigma'] \rangle\rangle$ be two lpps. Then, $\langle\langle \mathcal{C} \otimes \mathcal{C}', \langle \exists \mathsf{F} \rangle_{\mathsf{topo}} \vee_t \langle \exists \mathsf{F}' \rangle_{\mathsf{topo}}, \mathsf{FO}[\sigma \uplus \sigma'] \rangle\rangle$ is an lpps.

Example 6.3.21. The space $\mathcal{C} \stackrel{\text{def}}{=} \langle\langle \mathsf{Cliques} \otimes \mathsf{Paths}, \langle \mathsf{EFO} \rangle_{\mathsf{topo}}, \mathsf{FO} \rangle\rangle$ is an lpps.

Proof. We know using Exercise 6.3.20 that the following space, representing cartesian products of cliques and paths (with distinct edges E_1

for cliques and E_2 for paths) is an lpps: $\langle\!\langle \mathcal{C}', \tau', \mathcal{B}' \rangle\!\rangle$ where:

$$\begin{aligned}
\mathcal{C}' &\stackrel{\text{def}}{=} \mathsf{Cliques} \otimes \mathsf{Paths} \\
\tau' &\stackrel{\text{def}}{=} \langle \mathsf{EFO}[E_1] \rangle_{\mathsf{topo}} \vee_t \langle \mathsf{EFO}[E_2] \rangle_{\mathsf{topo}} \\
\mathcal{B}' &\stackrel{\text{def}}{=} \mathsf{FO}[\{ (E_1, 2), (E_2, 2) \}] \quad .
\end{aligned}$$

Now, the map $\mathsf{I} \colon \mathcal{C}' \twoheadrightarrow \mathcal{C}$ defined by $\varphi_E(x,y) \stackrel{\text{def}}{=} E_1(x,y) \vee E_2(x,y)$ is a monotone, surjective first order interpretation, and we conclude using Lemma 6.3.3. ■

Remark that we are now leaving the realm of what was possible to achieve using the "locality techniques" introduced in Chapter 4 (Locality and Preservation), as the r-local neighbourhoods inside structures belonging to the class introduced in Example 6.3.21 contain the whole structure when $r \geq 2$. In particular, it is not possible to leverage Theorem 5.1.2 to conclude that the Łoś-Tarski Theorem relativises to $\mathsf{Cliques} \otimes \mathsf{Paths}$.

6.3.3. Colouring Structures

We now turn our attention to more complex constructions, and leave the realm of "nice" categorical constructions to follow more rigorously the list of constructors provided for wqos and Noetherian spaces respectively in Table 6.1 and Table 6.2.

To that end, let us formally state the two stability results that we are going to use, namely the stability under finite words and finite trees, both in the case of well-quasi-orderings and Noetherian spaces. This forces us to not only state the lemmas, but actually give the definitions of the associated quasi-orderings and topologies.

We allow ourselves to write $|w|$ for the set $\{ 1, \dots, |w| \}$ when writing domains and co-domains of functions, as we believe this simplifies the reading process.

Definition 6.3.22. *A word u embeds in a word v, whenever there exists a strictly increasing map $h \colon |u| \to |v|$ such that $u_i \leq v_{h(i)}$ for $1 \leq i \leq |u|$. This relation is written $u \leq_{\mathsf{w}} v$.*

Implicitly in this document, trees are unranked, ordered, and finite.

Definition 6.3.23. *A tree t embeds in a tree t', whenever there exists a map h from the nodes of t to the nodes of t', such that $v \leq h(v)$ for every node v of t, and h respects the least common ancestor of nodes. This relation is written $t \leq_{\mathsf{t}} t'$.*

Theorem 6.3.24 [59]**.** Let $(W, \leq)$ be a well-quasi-order. Then, the set $(W^\star, \leq_{\mathsf{w}})$ is a well-quasi-order.

This is known as *Higman's Lemma*.

Theorem 6.3.25 [65]**.** Let $(W, \leq)$ be a well-quasi-order. Then, the set $(\mathsf{T}(W), \leq_{\mathsf{t}})$ is a well-quasi-order.

This is known as *Kruskal's Tree Theorem*.

Now that we have a definition of orders over trees and words, let us turn our attention to topologies. For simplicity, we only consider the subword topology over finite words and delay the definition of trees to Chapter 8 (Inductive Constructions). We write Σ^* for the set of finite words over Σ, and AB for the set of words uv with $u \in A$ and $v \in B$ whenever A and B are subsets of Σ^* for some alphabet Σ. Furthermore, we introduce $[U_1, \ldots, U_n]$ as a shorthand notation for the set $\Sigma^* U_1 \Sigma^* \ldots \Sigma^* U_n \Sigma^*$.

Definition 6.3.26 [45, Definition 9.7.26]**.** Given a topological space (Σ, τ), the space $\Sigma^\star$ of finite words over Σ can be endowed with the *subword topology* $\tau^{\star,t}$, generated by the open sets $[U_1, \ldots, U_n]$ when $U_i \in \tau$.

Theorem 6.3.27 *Topological Higman lemma* [45, Theorem 9.7.33]**.** Let (Σ, τ) be a topological space. (Σ, τ) is Noetherian if and only if $(\Sigma^\star, \tau^{\star,t})$ is Noetherian.

Our goal is to generalise Theorem 6.3.27 to lpps, i.e., to prove that if the space of finite words is a lpps assuming the space of letters is. This is the content of the upcoming Theorem 6.3.33.

Colouring finite words. We will first demonstrate how the ideas from Theorem 6.3.13 can be generalised to "constructors" where the underlying space is Noetherian. This requires us to refine our understanding of the subword topology, and in particular of its compact subsets.[36]

36: Beware that the subword topology is Noetherian if and only if the topology on the letters is. However, we are going to consider words with finite structures as letters, and not assume that the class of structures is Noetherian!

Lemma 6.3.28. *Let* (X, τ) *be a topological space, and* $(U_i)_{1 \leq i \leq n}$ *be a finite sequence of compact open subsets. Then,* $[U_1, \ldots, U_n]$ *is a compact open subset of* $\tau^{\star,t}$.

Proof. We use Alexander's subbase lemma, and consider an open cover $\left(V^i\right)_{i \in I}$ of $U \stackrel{\text{def}}{=} [U_1, \ldots, U_n]$, where $V^i \stackrel{\text{def}}{=} [V^i_1, \ldots, V^i_{n_i}]$ for all $i \in I$.

Let $w \in U$ of length n, there exists an $i(w) \in I$, and a strictly increasing map $\rho_w \colon \{1, \ldots, n_{i(w)}\} \to \{1, \ldots, n\}$, such that:

$$\forall 1 \leq j \leq n_{i(w)}, w_{\rho_w(j)} \in V_j^{i(w)} \quad .$$

Let us define $Z_w(p) \stackrel{\text{def}}{=} \bigcap_{j \in \rho_w^{-1}(\{p\})} V_j^{i(w)}$, which is an open subset of X. Equivalently, $Z_w(p)$ is defined as $V_j^{i(w)}$ if there exists j such that $p = \rho_w(j)$, and X otherwise.[37]

37: The advantage of the first definition is that it will generalise to maps ρ_w that are not injective in the upcoming section.

Then, we define $Z_w \stackrel{\text{def}}{=} [Z_w(1), \ldots, Z_w(n)]$, which is an open subset of $\mathsf{X}^\star$, and $Z_w^\times \stackrel{\text{def}}{=} Z_w(1) \times \cdots \times Z_w(n)$, which is an open subset of X^n in its product topology.

Our first claim is that for $1 \leq p \leq n$, $w_p \in Z_w(p)$. Which holds by construction. This implies that $w \in Z_w$, and that $(w_1, \ldots, w_n) \in Z_w^{\times}$.

Our second claim is that for all w of length n, we have $Z_w \subseteq V^{i(w)}$. Indeed, consider a word $u \in Z_w$. By definition, there exists a strictly increasing map $h \colon \{1, \ldots, n\} \to \{1, \ldots, |u|\}$ such that

$$\forall 1 \leq p \leq n, u_{h(p)} \in Z_w(p) \quad .$$

Now, remark that

$$\forall 1 \leq j \leq n_{i(w)}, u_{h(\rho_w(j))} \in Z_w(h(j)) \subseteq V_j^{i(w)} \quad .$$

Because $h \circ \rho_w$ is the composition of two strictly increasing maps, it is itself a strictly increasing map from $\{1, \ldots, n_{i(w)}\}$ to $\{1, \ldots, |u|\}$, and we have proven that $u \in V^{i(w)}$.

Now, the core of the proof is to notice that $U^{\times} \stackrel{\text{def}}{=} U_1 \times \cdots \times U_n$ is a compact subset of X^n thanks to the Tychonoff theorem, which can be applied since every U_p is compact for $1 \leq p \leq n$. Furthermore, $(Z_w^{\times})_{w \in U, |w| = n}$ is an open cover of $U^{\times}$. As a consequence, and there exists a finite set W of words of length n in U such that

$$U^{\times} \subseteq \bigcup_{w \in W} Z_w^{\times} \quad .$$

To conclude, let $u \in U$, it has a subword v of length n that is in U, hence the word v (seen as a tuple) belongs to $U^{\times}$. Therefore, v (seen as a tuple of length n) belongs to a $Z_w^{\times}$ for some $w \in W$. In particular, v belongs to Z_w (when seen as a word). Because Z_w is closed under insertion of letters,[38] we conclude that $u \in Z_w$ too. We have successfully extracted a finite subcover of $\left(V^i\right)_{i \in I}$:

$$U \subseteq \bigcup_{w \in W} Z_w \subseteq \bigcup_{w \in W} V^{i(w)} \quad . \qquad \blacksquare$$

38: Recall that $[A_1, \ldots, A_n]$ means $\mathsf{X}^* A_1 \mathsf{X}^* \cdots \mathsf{X}^* A_n \mathsf{X}^*$, which is upwards closed for $(=)_{\mathsf{w}}$, the subword ordering with equality on the letters.

As a consequence of Lemma 6.3.28, we conjecture that the following variation of the topological Higman lemma holds.

Conjecture 6.3.29. Let (X, τ) be a spectral space. Then, its sobrification $(\mathcal{S}(\mathsf{X}^\star), \tau^{\star,t})$ is a spectral space.

In the definition of an lpps, the topology and the Boolean subalgebra of definable subsets play an equally important role. In the case of finite words, the definable subsets (that is formally introduced in Definition 6.3.31) are basically a choice of finitely many definable properties over the letters, and then an arbitrary Boolean combination of these properties. In order to properly define the above notion, we let use first introduce the notion of *join* of Boolean subalgebras (of $\mathcal{P}(\mathsf{X})$).

Definition 6.3.30. *Let $\mathcal{B}$ and $\mathcal{B}'$ be two Boolean subalgebras of $\mathcal{P}(\mathsf{X})$. The join algebra $\mathcal{B} \vee_b \mathcal{B}'$ is the Boolean subalgebra of X generated by $\mathcal{B} \cup \mathcal{B}'$.*

> *For a family $(\mathcal{B}_i)_{i\in I}$ of Boolean subalgebras, the join algebra is generated by $\bigcup_{i\in I}\mathcal{B}_i$, and written $\bigvee_{i\in I}\mathcal{B}_i$.*

Furthermore, we introduce, when B is a Boolean subalgebra of some $\mathcal{P}(\mathsf{X})$, the construction $\langle B\rangle^{\mathsf{comp}}_{\mathsf{bool}}$ to denote the complete Boolean subalgebra generated by B, that is, arbitrary intersections of arbitrary unions of elements in B and their complements.[39][40]

> **Definition 6.3.31.** *Let $\mathcal{B}$ be a Boolean subalgebra of $\mathcal{P}(\mathsf{X})$. Then,*
>
> $$(\mathcal{B})^{\star,b} \stackrel{\mathsf{def}}{=} \bigvee_{\mathcal{W}\subseteq_{\mathsf{fin}}\mathcal{B}} \langle\{\, W_1\cdots W_n : n\in\mathbb{N}, W_i\in\mathcal{W}\,\}\rangle^{\mathsf{comp}}_{\mathsf{bool}} \quad . \tag{6.3}$$

The (limited) ability to consider arbitrary unions (and intersections) of elements in $(\mathcal{B})^{\star,\mathrm{b}}$ allows us to define subsets that clearly escape the realm of first order logic, as illustrated in Example 6.3.32.[41]

> **Example 6.3.32.** Let X be a set, and let $\mathcal{B}$ be a Boolean subalgebra of $\mathcal{P}(\mathsf{X})$. The set of words w such that $|w|$ is a prime number is definable in $\langle\!\langle \mathsf{X}^\star, \tau^{\star,\mathrm{t}}, (\mathcal{B})^{\star,\mathrm{b}}\rangle\!\rangle$.

Proof. This set can be written as the following union:

$$\bigcup_{p \text{ prime}} \underbrace{\mathsf{X}\cdots\mathsf{X}}_{p\text{ times}} \quad . \qquad \blacksquare$$

> **Theorem 6.3.33.** *Let $\langle\!\langle \mathsf{X}, \tau, \mathcal{B}\rangle\!\rangle$ is an lpps. Then, $\langle\!\langle \mathsf{X}^\star, \tau^{\star,t}, (\mathcal{B})^{\star,b}\rangle\!\rangle$ is an lpps.*

Proof. Let us first prove that $\langle\!\langle \mathsf{X}^\star, \tau^{\star,\mathrm{t}}, (\mathcal{B})^{\star,\mathrm{b}}\rangle\!\rangle$ is an lps. For that, consider a subbasic open subset $U \stackrel{\mathsf{def}}{=} [U_1,\ldots,U_n]$, where $U_i\in\tau$ for all $1\le i\le n$. Because $U_i\in\tau$, and $\langle\!\langle \mathsf{X}, \tau, \mathcal{B}\rangle\!\rangle$ is itself an lps, this provides us with a set J of indices and a family $D^i_j\in\mathcal{B}\cap\tau$ where $j\in J$ such that

$$\forall 1\le i\le n, U_i = \bigcup_{j\in J} D^i_j \quad .$$

Now, it suffices to unfold the definition of the syntax $[\cdots]$ to conclude that U is an arbitrary union of elements in $\tau^{\star,\mathrm{t}}\cap(\mathcal{B})^{\star,\mathrm{b}}$:

$$\begin{aligned} U &= [U_1,\ldots,U_n] \\ &= \mathsf{X}^* U_1 \mathsf{X}^*\cdots\mathsf{X}^* U_n \mathsf{X}^* \\ &= \bigcup_{j_1,\ldots,j_n\in J} \mathsf{X}^* D^1_{j_1}\mathsf{X}^*\cdots\mathsf{X}^* D^n_{j_n}\mathsf{X}^* \\ &= \bigcup_{j_1,\ldots,j_n\in J} \underbrace{\bigcup_{k_0,\ldots,k_{n+1}\in\mathbb{N}} \mathsf{X}^{k_0} D^1_{j_1}\mathsf{X}^{k_1}\cdots\mathsf{X}^{k_n} D^n_{j_n}\mathsf{X}^{k_{n+1}}}_{=[D^1_{j_1},\ldots,D^n_{j_n}]\in\tau^{\star,\mathrm{t}} \text{ and } \in(\mathcal{B})^{\star,\mathrm{b}}} \quad . \end{aligned}$$

We have proven that $\langle\tau^{\star,\mathrm{t}}\cap(\mathcal{B})^{\star,\mathrm{b}}\rangle_{\mathsf{topo}} = \tau^{\star,\mathrm{t}}$.

39: This definition is sound because unions distribute over intersections in $\mathcal{P}(\mathsf{X})$.

40: We restrict our attention to Boolean subalgebras to avoid using advanced tools such as the Dedekind–MacNeille completion.

Informally, the idea of Definition 6.3.31 is that a "definable subset" of finite words starts by fixing a finite set of properties that can be checked on the letters, and then can perform *arbitrary* Boolean combination of checks on the words using these properties.

41: However, the Boolean subalgebra defined in is *not* a complete Boolean subalgebra.

The proof that the space is an lpps is quite similar to the one of Theorem 6.3.13, with an extra twist at the end. As usual, the only thing to check is that definable open subsets of $\mathsf{X}^\star$ are compact.

Let U be a definable open subset. Because $(\mathcal{B})^{\star,\mathsf{b}}$ is defined as a *directed* union,[42] there exists a finite subset $\mathcal{W} \subseteq_{\mathsf{fin}} \mathcal{B}$ such that:

$$U \in \langle \{ W_1 \cdots W_n : n \in \mathbb{N}, W_i \in \mathcal{W} \} \rangle_{\mathsf{bool}}^{\mathsf{comp}} \quad .$$

42: For every $W_1, W_2 \subseteq_{\mathsf{fin}} \mathcal{B}$, there exists $W_3 \subseteq_{\mathsf{fin}} \mathcal{B}$ such that $W_1 \subseteq W_3$ and $W_2 \subseteq W_3$, namely $W_3 \stackrel{\mathsf{def}}{=} W_1 \cup W_2$.

Without loss of generality, we assume that $\mathcal{W} = \langle \mathcal{W} \rangle_{\mathsf{bool}}$, which is also a finite subset of $\mathcal{B}$. Let us define the equivalence relation $x \equiv_{\mathcal{W}} y$ on X to hold when for all $W \in \mathcal{W}$, $x \in W \iff y \in W$, that is, the equivalence relation of finite index generated by the sets in $\mathcal{W}$. Remark that for $u, v \in \mathsf{X}^\star$, such that for all $i \in I$, $u_i \equiv_{\mathcal{W}} v_i$, $u \in U$ if and only if $v \in U$.

Let us fix $m \in \mathbb{N}$ and consider (thanks to Zorn's Lemma) an open subset $[U_1, \ldots, U_m] \subseteq U$ that is maximal for inclusion among opens of this form.[43]

43: We have fixed m, hence the inclusion $[A_1, \ldots, A_m] \subseteq [B_1, \ldots, B_m]$ amounts to checking whether $A_i \subseteq B_i$ pointwise, in the case where none of the A_i's is the empty set.

Using the same arguments as for the product of two spaces in Theorem 6.3.13, we easily conclude that the sets U_i must be saturated for $\equiv_{\mathcal{W}}$.[44] As a consequence, they are definable open subsets, hence compact subsets of (X, τ).

44: Assume by contradiction that some U_i is not saturated. There exists $x \equiv_{\mathcal{W}} y$ such that $x \in U_i$ and $y \notin U_i$. We select finite sets of representatives F_j for the $\equiv_{\mathcal{W}}$-equivalence classes that intersect U_j for $1 \leq j \neq i \leq m$. Because U respects the $\equiv_{\mathcal{W}}$-equivalence, the words built using elements in F_j and y all belongs to U. The open neighbourhoods of these words provide open neighbourhoods around y. The union of U_i together with the intersection of neighbourhoods that were built above contradicts the maximality of U_i. This sketch will be formally developed in the generalised setting of wreath products, and in particular in the proof of Theorem 6.3.44.

Let us now unfix m. We know that U is included in the union over $m \in \mathbb{N}$ of the subsets $[U_1, \ldots, U_m]$ that are maximal for inclusion (among other subsets of the same shape). As a consequence, there exists a sequence $(I_m)_{m \in \mathbb{N}}$ of subsets of $\mathbb{N}$ and a sequence $(U_{i,m})_{m \in \mathbb{N}, i \in I_m}$ of elements in $\mathcal{K}^\circ(\tau)$ that are saturated for $\equiv_{\mathcal{W}}$, such that

$$U = \bigcup_{m \in \mathbb{N}} \bigcup_{i \in I_m} [U_{1,i} \ldots, U_{m,i}] \quad .$$

Recall that $\equiv_{\mathcal{W}}$ is of finite index, hence there are finitely many possible values for the sets $U_{m,i}$. Let us write $\Sigma \stackrel{\mathsf{def}}{=} \{ U_{m,i} : m \in \mathbb{N}, i \in I_m \}$. The topology $\langle \Sigma \rangle_{\mathsf{topo}}$ is Noetherian because Σ is finite. As a consequence of the topological Higman lemma, $(\mathsf{X}^\star, \langle \Sigma \rangle_{\mathsf{topo}}{}^{\star,\mathsf{t}})$ is Noetherian, and in particular, there exists $m_0 \in \mathbb{N}$ and a sequence $(J_m)_{m \leq m_0}$ of *finite* subsets of $\mathbb{N}$ such that

$$U = \bigcup_{m \in \mathbb{N}} \bigcup_{i \in I_m} [U_{1,i} \ldots, U_{m,i}]$$
$$U = \bigcup_{m \leq m_0} \bigcup_{i \in J_m} [U_{1,i} \ldots, U_{m,i}] \quad .$$

Furthermore, we can leverage Lemma 6.3.28, stating that $[U_1, \ldots, U_p] \in \mathcal{K}^\circ(\tau^{\star,\mathsf{t}})$ whenever $U_1, \ldots, U_p \in \mathcal{K}^\circ(\tau)$, to conclude that U is a finite union of compact open subsets. Hence, U is compact thanks to Lemma D.4.3. ■

We have used a deep theorem about the subword topology to get around the last part of the proof of Theorem 6.3.33. However, there exists a longer proof that avoids it and goes back to simple well-quasi-ordering arguments. Let us sketch this alternative proof hereafter, not only for completeness, but because it allows us to generalise the argument to wreath products.[45]

45: The notion of wreath product will be introduced in Definition 6.3.38.

Alternative Ending. Let us write $S \stackrel{\text{def}}{=} \{ [U_{m,i}, \dots, U_{m,i}] : m \in \mathbb{N}, i \in I_m \}$, and $\Sigma \stackrel{\text{def}}{=} \{ U_{i,m} : m \in \mathbb{N}, i \in I_m \}$.

Because elements in Σ are saturated for $\equiv_{\mathcal{W}}$, which is of finite index, Σ is a finite alphabet, and in particular a well-quasi-order when equipped with the equality preorder. Using Higman's Lemma, $(\Sigma^\star, (=)_{\mathsf{w}})$ is again wqo. The map $f : (\Sigma^\star, (=)_{\mathsf{w}}) \to (S, \supseteq)$ that sends w to $[w_1, \dots, w_n]$ is surjective and monotone. As a consequence, $(S, \supseteq)$ is a wqo, and S is the upward closure (for $\supseteq$) of finitely many elements $S_f \subseteq_{\mathsf{fin}} S$.

We have proven that

$$U = \bigcup_{V \in S_f} V \quad .$$

Hence, U is compact via Lemma D.4.3. ∎

As for the products and sums, we are not entirely satisfied by Theorem 6.3.33, because the Boolean subalgebra in the statement is not obtained by interpreting FO sentences over finite words. Before proving Lemma 6.3.35, which leverages the Feferman-Vaught technique to transform first order sentences over $\mathcal{C}^\star$ into elements of $(\mathsf{FO})^{\star,\mathrm{b}}$, we first need to explain how to interpret elements of $\mathcal{C}^\star$ as (finite) relational structures.

Definition 6.3.34. *Let σ be a finite relational signature, and $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$. We interpret a word $w \in \mathcal{C}^\star$ as the structure $\biguplus_{1 \leq i \leq |w|} w_i$ with extra relations $a \leq b$ whenever $a \in w_i$, $b \in w_j$, and $i \leq j$.*

Definition 6.3.34 allows us to evaluate first order formulas $\varphi \in \mathsf{FO}[\sigma \uplus \{ (\leq, 2) \}]$ over the words w in $\mathcal{C}^\star$. Remark that if two elements $a, b \in \mathrm{dom}(w)$ are in a relation R that is not $\leq$, then $a \leq b$ and $b \leq a$.

Lemma 6.3.35. *Let $\mathcal{C} \subseteq \mathsf{Struct}(\sigma)$ be a class of structures. For every $\varphi \in \mathsf{FO}[\sigma \uplus \{ (\leq, 2) \}]$, $[\![\varphi]\!]_{\mathcal{C}^\star} \in (\mathsf{FO}[\sigma])^{\star,b}$.*

Proof Sketch. To prove the desired result, we use the Feferman-Vaught technique, with a varying set of types, with $\mathbb{T}_n \stackrel{\text{def}}{=} L_n$. Let us introduce a variant $\text{w-conv}^{\mathbb{T}_n}(\cdot, \cdot)$ of the conversion operator $\text{conv}^{\mathbb{T}}(\Gamma, \varphi)$ by induction on the structure of ψ and for all $\Gamma : \mathrm{fv}(\psi) \to \mathbb{T}_n$ as follows:[46]

1. $\text{w-conv}^{\mathbb{T}_n}(\Gamma, \exists x.\psi) \stackrel{\text{def}}{=} \bigvee_{T \in \mathbb{T}_n} \exists x : T.\text{w-conv}^{\mathbb{T}_n}(\Gamma[x \mapsto T], \psi)$,
2. $\text{w-conv}^{\mathbb{T}_n}(\Gamma, \psi_1 \vee \psi_2) \stackrel{\text{def}}{=} \text{w-conv}^{\mathbb{T}_n}(\Gamma, \psi_1) \vee \text{w-conv}^{\mathbb{T}_n}(\Gamma, \psi_2)$,
3. $\text{w-conv}^{\mathbb{T}_n}(\Gamma, \neg\psi) \stackrel{\text{def}}{=} \neg\text{w-conv}^{\mathbb{T}_n}(\Gamma, \psi)$,
4. $\text{w-conv}^{\mathbb{T}_n}(\Gamma, \top) \stackrel{\text{def}}{=} \top$,
5. $\text{w-conv}^{\mathbb{T}_n}(\Gamma, x \leq y) \stackrel{\text{def}}{=} \top$ if $\Gamma(x) \leq \Gamma(y)$, and $\bot$ otherwise,
6. For $R \in \sigma$, $\text{w-conv}^{\mathbb{T}_n}(\Gamma, R(y_1, \dots, y_n)) \stackrel{\text{def}}{=} R(y_1 : \Gamma(y_1), \dots, y_n : \Gamma(y_n))$ when $\Gamma(y_1) = \dots = \Gamma(y_n)$, and $\bot$ otherwise.

46: The idea is that an element of a word $w \in \mathcal{C}^\star$ is a pair (i, a) where $1 \leq i \leq |w|$ and $a \in w_i$. We will give to an element (i, a) the "runtime type" i, which allows us to perform a partial evaluation of the sentences by statically evaluating the relation $\leq$.

47: The proof is done by induction, as for Lemma 4.2.17. The case of the existential quantification, is exactly the same, and the only thing to check is that replacing relations with $\bot$ and the ordering with the "ordering on types" is correct, which holds by definition of ρ, and because elements that are not in the same position in the word cannot be in any relation that is not $\leq$.

Now, we claim[47] that the following analogue of Lemma 4.2.17 holds for this new conversion. For all $n \in \mathbb{N}$, for all $\psi(\vec{x}) \in \mathsf{FO}[\sigma \uplus \{ (\leq, 2) \}]$, for all $\mathfrak{A} \in \mathcal{C}^\star$ of length n, for all valuation $\upsilon\colon \vec{x} \to \mathfrak{A}$, for all runtime typing $\rho\colon \mathfrak{A} \to \mathbb{T}_n$, for all $\mathbb{T}_n$-typing environments Γ, such that $\Gamma(x) = \rho(\upsilon(x))$ for all $x \in \vec{x}$, and such that $\rho(a) \leq \rho(b)$ if and only if $a \leq b$ in $\mathfrak{A}$, the following are equivalent:

- $\mathfrak{A}, \upsilon \models \psi(\vec{x})$,
- $\mathfrak{A}, \rho, \upsilon \models^{\mathbb{T}} \text{w-conv}^{\mathbb{T}_n}(\Gamma, \psi)$.

Applying the above claim to $\mathfrak{A}$, the empty valuation, and the function $\rho\colon \mathfrak{A} \to L_n$ that maps an element $a \in \mathfrak{A}$, to its position in the total ordering defined by $\leq$ inside $\mathfrak{A}$, we conclude that:

$$(\mathfrak{A}, \emptyset \models \varphi) \iff \left(\mathfrak{A}, \rho, \emptyset \models^{\mathbb{T}} \text{w-conv}^{\mathbb{T}_n}(\emptyset, \varphi)\right)$$

Applying Lemma 4.2.19 over $\text{w-conv}^{\mathbb{T}_n}(\emptyset, \varphi)$, we obtain a finite set S of monotyped formulas, such that a Boolean combination of formulas in S that is equivalent to $\text{w-conv}^{\mathbb{T}_n}(\emptyset, \varphi)$. Notice that a monotyped formula can only talk about one equivalence class for $\leq$ in the structure $\mathfrak{A}$. As a consequence, whether $\mathfrak{A} \models \varphi$ is uniquely determined by the evaluation of monotyped formulas on each of the $\leq$-equivalence classes of $\mathfrak{A}$.

Recall that Lemma 4.2.13 states that evaluating a monotyped formula θ with type $T \in \mathbb{T}$, over a structure $\mathfrak{A}$ with typing function $\rho\colon a \mapsto T$ is equivalent to evaluating $\mathrm{U}\mathbb{T}(\theta)$ over $\mathfrak{A}$.

Furthermore, Lemma 4.2.19 states that the monotyped formulas in S have quantifier rank at most $\mathrm{rk}(\varphi)$. As a consequence, $\{ \mathrm{U}\mathbb{T}(\theta) \colon \theta \in S \}$ is finite up to logical equivalence, and its size is bounded independently of the length n of $\mathfrak{A}$!

What we have proven is that there exists a finite subset S' of (usual) formulas, namely $\{ \mathrm{U}\mathbb{T}(\theta) \colon \theta \in S \}$, such that for all $\mathfrak{A} \in \mathcal{C}^\star$, their evaluation on the $\leq$-equivalence classes of $\mathfrak{A}$ uniquely determines whether $\mathfrak{A} \models \varphi$. Formally, for all $n \in \mathbb{N}$, we have a Boolean function ρ_n from $\{ 0, 1 \}^{\{ 1,\dots,n \} \times S'}$ to $\{ 0, 1 \}$, such that for all $\mathfrak{A}$ of length n in $\mathcal{C}^\star$, $\mathfrak{A} \models \varphi$ if and only if $\rho_n((\mathfrak{A}_i \models \psi)_{1 \leq i \leq n, \psi \in S'}) = 1$, where $\mathfrak{A}_i$ is the ith $\leq$-equivalence class in $\mathfrak{A}$.

This is clearly definable in $(\mathsf{FO}[\sigma])^{\star,\mathrm{b}}$, because it is an arbitrary union of elements $W_1 \cdots W_n$, where $n \in \mathbb{N}$, and $W_i \in \langle S' \rangle_{\mathsf{bool}}$, the latter being a finite Boolean subalgebra of $\mathsf{FO}[\sigma]$. ■

Corollary 6.3.36. *Let $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ be a class of structures, $\tau \in \mathrm{Top}(\mathcal{C})$ be a topology such that $\langle\!\langle \mathcal{C}, \tau, \mathsf{FO}[\sigma] \rangle\!\rangle$ is an lpps. Then, the following space of finite words is an lpps:*

$$\langle\!\langle \mathcal{C}^\star, \tau^{\star,t}, \mathsf{FO}[\sigma \uplus \{ \leq \}] \rangle\!\rangle \quad .$$

Proof. Using Theorem 6.3.33, it remains for us to prove two statements:

1. That $[\![\mathsf{FO}[\sigma \uplus \{ \leq \}]\!]_{\mathcal{C}^\star} \subseteq ([\![\mathsf{FO}[\sigma]]\!]_{\mathcal{C}})^{\star,\mathrm{b}}$, which is the content of Lemma 6.3.35, and

2. That $\langle\!\langle \mathcal{C}^\star, \tau^{\star,\mathrm{t}}, \mathsf{FO}[\sigma \uplus \{ \leq \}]\rangle\!\rangle$ is an lps.

For this last item, remark that for all sequence $(D_i)_{1 \leq i \leq n}$ of definable subsets of $\mathcal{C}$, $[D_1, \dots, D_n]$ is first order definable in $\mathcal{C}^\star$. Indeed, $D_i = [\![\theta_i]\!]_{\mathcal{C}}$ for some first order sentence $\theta_i \in \mathsf{FO}[\sigma]$, hence we can define

$$\psi \overset{\text{def}}{=} \exists x_1, \dots, x_n. \bigwedge_{1 \leq i < j \leq n} x_i < x_j \wedge \bigwedge_{1 \leq i \leq n} [\theta_i]_{|\equiv x_i} \quad ,$$

where $[\theta_i]_{|\equiv x_i}$ is the relativisation of θ_i to variables y such that $y \leq x_i \wedge x_i \leq y$.

It remains for us to prove that the topology $\tau^{\star,\mathrm{t}}$ is generated by the subsets of the form $[D_1, \dots, D_n]$, where D_i is a definable open subset of $\mathcal{C}$.

Let $[U_1, \dots, U_n]$ be a subbasic open subset of $\tau^{\star,\mathrm{t}}$. Because $\langle\!\langle \mathcal{C}, \tau, \mathsf{FO}\rangle\!\rangle$ is an lpps, for all $1 \leq i \leq n$, there exists a family $(\theta_{i,j})_{j \in J_i}$ such that $U_i = \bigcup_{j \in J_i} [\![\theta_{i,j}]\!]_{\mathcal{C}}$. Therefore,

$$[U_1, \dots, U_n] = \bigcup_{f \in \prod_{1 \leq i \leq n} J_i} \underbrace{[[\![\theta_{1,f(1)}]\!]_{\mathcal{C}}, \dots, [\![\theta_{1,f(n)}]\!]_{\mathcal{C}}]}_{\text{first order definable and open}} \quad . \qquad \blacksquare$$

Let us now discuss a little the tremendous power given by Corollary 6.3.36. For instance, $\Delta^2_{\mathrm{deg}}{}^\star$ is clearly not a class of finite structures having any kind on sparsity properties.[48] In particular, this class shows how the "structural" approach complements the approach using locality presented in Chapter 5 (A Local-to-Global Preservation Theorem), and its Theorem 5.1.2 on page 120.

48: Recall that Definition 6.3.34 interprets $\Delta^2_{\mathrm{deg}}{}^\star$ as "flattened" structures over $\sigma \uplus \{ (\leq, 2) \}$. Therefore, the Gaifman graph of an element in $\Delta^2_{\mathrm{deg}}{}^\star$ is always a clique.

Example 6.3.37. Let $\mathcal{C}$ be the class $\Delta^2_{\mathrm{deg}}{}^\star$ of finite words of finite graphs of degree bounded by 2. Then,

- The Łoś-Tarski Theorem relativises to $\mathcal{C}$,
- $\mathcal{C}$ is hereditary,
- $\mathcal{C}$ is not closed under disjoint unions,
- $\mathcal{C}$ is not of bounded degree,
- For all $r \geq 1$, $k \geq 1$, $\mathsf{Local}(\mathcal{C}, r, k) = \mathcal{C}$.

Proof. It is clear that $\Delta^2_{\mathrm{deg}}{}^\star$ does not have bounded degree, because of the order relation $\leq$. It is not closed under disjoint unions because the disjoint union of two total orderings is not a total ordering. The total ordering relation $\leq$ turns the Gaifman graph of elements in $\Delta^2_{\mathrm{deg}}{}^\star$ into cliques, hence the (non-trivial) local neighbourhoods are always the entire structure.

To prove that the Łoś-Tarski Theorem relativises to $\mathcal{C}$, we first use the Lemma 6.1.16, and notice that $\Delta^2_{\mathrm{deg}}{}^\star$ is a hereditary class of finite structures in $\mathsf{Fin}(\sigma \uplus \{ (\leq, 2) \})$. As a consequence, it suffices for us to demonstrate that $\langle\!\langle \mathcal{C}, \mathsf{Alex}(\subseteq_i), \mathsf{FO}[\sigma \uplus \{ (\leq, 2) \}\rangle\!\rangle$ is an lpps. Recall that Corollary 6.3.36, implies that $\langle\!\langle \mathcal{C}, \mathsf{Alex}(\subseteq_i)^{\star,\mathrm{t}}, \mathsf{FO}[\sigma \uplus \{ (\leq, 2) \}\rangle\!\rangle$ is an lpps. Hence, it suffices to prove that $\mathsf{Alex}(\subseteq_i)^{\star,\mathrm{t}} = \mathsf{Alex}(\subseteq_i)$.[49] To that end, let us consider $w \in \Delta^2_{\mathrm{deg}}{}^\star$ of length n. We claim that $[\uparrow_{\subseteq_i} w_1, \dots, \uparrow_{\subseteq_i} w_n]$ is the $\subseteq_i$-upward closure of w in $\mathcal{C}^\star$ seen as models.

49: This equation hides the fact that the signatures over which $\subseteq_i$ is defined varies...

This is because $w \subseteq_i w'$ holds if and only if there exists a map $h : w \to w'$ that is a QF-embedding. In particular, h remains a QF-embedding from w seen as $L_{|w|}$ to w' seen as $L_{|w'|}$, obtained by removing every relation from the structures except the order relation $\leq$. This proves that h is a strictly increasing map from $\{ 1, \dots, |w| \}$ to $\{ 1, \dots, |w'| \}$. Furthermore, h restricted to w_i remains a QF-embedding from w_i to $w'_{h(i)}$. In particular, we have proven that $w' \in [\uparrow_{\subseteq_i} w_1, \dots, \uparrow_{\subseteq_i} w_n]$. The converse inclusion is done similarly, combining QF-embeddings on every component into a global QF-embedding that also respects the ordering. ■

Colouring complex structures. We are now going to leverage what was done in the case of coloured finite words to colour arbitrary classes of structures. This is motivated by the fact that in Tables 6.1 and 6.2 several constructors are obtained via a suitable colouring of a finite class of structures (finite words, finite trees, finite multisets).

Definition 6.3.38. *Let $\mathcal{C}$ be a class of structures and X be a set. The* wreath product $\mathcal{C} \rtimes \mathsf{X}$ *is defined as the class of functions $f : \mathfrak{A} \to \mathsf{X}$ where $\mathfrak{A} \in \mathcal{C}$. Equivalently, it is defined as $\sum_{\mathfrak{A} \in \mathcal{C}} \prod_{a \in \mathfrak{A}} \mathsf{X}$.*

Informally, $\mathcal{C} \rtimes \mathsf{X}$ is the class of structures in $\mathcal{C}$ where elements are coloured by points in X. From this viewpoint, if $\mathcal{C} \subseteq \mathsf{Struct}(\sigma)$ and $\mathcal{C}' \subseteq \mathsf{Struct}(\sigma')$, then $\mathcal{C} \rtimes \mathcal{C}'$ can be understood as a kind of "substitution of structures," and we will provide an interpretation of these new structures in terms of structures over $\sigma \uplus \sigma'$ in Definition 6.3.45. Remark that Definition 6.3.38 encodes the lexicographic product when the functions f are chosen to be constant functions.

Before studying wreath products in terms of classes of structures, let us see how the notion of wreath product generalises the intuition that words are linear orders with colours in Example 6.3.39. Then, we will prove Theorem 6.3.44, that strictly generalises Theorem 6.3.33; and Theorem 6.3.46 that strictly generalises Corollary 6.3.36.

Example 6.3.39. Let LinOrd be the class of finite linear orders. Then $\mathcal{C}^\star$ is exactly $\mathsf{LinOrd} \rtimes \mathcal{C}$.

Before going to the formal definition, let us recall that the subword topology is generated by elements of the form $[U_1, \dots, U_n]$ which are themselves *words of open subsets*. We will take a similar approach when defining the topology on the wreath product: open subsets will be generated by maps from a finite structure to open subsets. However, this is only half of what needs to be defined: in the case of words, we had a corresponding "semantics", telling us which words belong to $V \overset{\text{def}}{=} [U_1, \dots, U_n]$. The informal idea to produce this semantics is that a word w belongs to an open subset whenever there is a map h from the positions in the open subset V, to the positions in the word w, such that $w_{h(i)} \in U_i$.

The general form $\mathcal{C} \rtimes \mathsf{X}$ where X is not a class of structures allows us to consider heterogeneous constructions, such as "linear orders coloured with polynomials" (or linear orders of polynomials if one prefers the idea that structures are substituted rather than coloured). Moreover, it provides a handy description of the topology that will be placed on this wreath product.

Suppose that (X, τ) is a topological space, that F is a fragment of FO, and $\mathcal{C} \subseteq \mathsf{Struct}(\sigma)$. We want to build a topology over $\mathcal{C} \rtimes \mathsf{X}$. The idea is that a subbasic open subset is defined by an element U of $\mathcal{C} \rtimes \tau$, which describes the following subset of $\mathcal{C} \rtimes \mathsf{X}$: $[\mathsf{F} \| U] \overset{\text{def}}{=} \{ f \in \mathcal{C} \rtimes \mathsf{X} \colon \exists h \colon \mathrm{dom}(U) \to \mathrm{dom}(f)$ F-embedding$, \forall a \in \mathrm{dom}(U), f(h(a)) \in$

$U(a)\,\}$. The *wreath subbasis* $\mathcal{C} \rtimes_{\mathsf{F}} \tau$ is the collection of the sets $[\mathsf{F}\|U]$ where U ranges in $\mathcal{C} \rtimes \tau$.

This allows us to define the *wreath topology* as $\langle \mathcal{C} \rtimes_{\mathsf{F}} \tau\rangle_{\mathsf{topo}}$, but also to define the *wreath algebra* as $\bigvee_{\mathcal{W}\subseteq_{\mathsf{fin}}\mathcal{B}} \langle \mathcal{C} \rtimes_{\mathsf{FO}} \mathcal{W}\rangle^{\mathsf{comp}}_{\mathsf{bool}}$.

Example 6.3.40. Let (X, τ) be a topological space. The spaces $(\mathsf{LinOrd} \rtimes \mathsf{X}, \langle \mathsf{LinOrd} \rtimes_{\mathsf{QF}} \tau\rangle_{\mathsf{topo}})$ and $(\mathsf{X}^\star, \tau^{\star,\mathsf{t}})$ are homeomorphic.

Proof Sketch. First of all, it is clear that words $w \in \mathsf{X}^\star$ are in one to one correspondence with elements of $\mathsf{LinOrd} \rtimes \mathsf{X}$ via the construction $f_w \colon L_{|w|} \to \mathsf{X}$ defined by $f_w(i) \stackrel{\mathsf{def}}{=} w_i$.

Now, let us consider a subbasic open subset $W \stackrel{\mathsf{def}}{=} [U_1, \ldots, U_n]$ of $\tau^{\star,\mathsf{t}}$. This corresponds to the open subset $f_W \colon L_n \to \tau$ that is defined by $f_W(i) \stackrel{\mathsf{def}}{=} U_i$. Conversely, a subbasic open subset $W \colon L_n \to \tau$ corresponds to an open subset $[W(1), \ldots, W(n)]$.

Let us check that the above bijective mapping preserves the semantics. In particular, we want to prove the following equivalence

$$w \in W \stackrel{\mathsf{def}}{=} [U_1, \ldots, U_n] \iff f_w \in [\mathsf{QF}\|f_W] \quad .$$

Recall that by definition of $[\cdots]$, $w \in W$ if and only if there exists a map $h \colon L_n \to L_{|w|}$ that is strictly increasing, and such that $w_{h(i)} \in W_i$ for all $1 \leq i \leq n$. Now, such a map h is precisely a QF-embedding of elements in LinOrd. Therefore, $w \in W$ exactly corresponds to $f_w \in [\mathsf{QF}\|f_W]$. ∎

Now that we have gained a little intuition on the behaviour of the wreath product, let us construct the associated logically presented pre-spectral space.

Definition 6.3.41. *Let* F *be a fragment of* $\mathsf{FO}[\sigma]$, $\mathcal{C} \subseteq \mathsf{Struct}(\sigma)$, *and* $\langle\!\langle \mathsf{X}, \tau, \mathcal{B}\rangle\!\rangle$ *be a logically presented space. The wreath product of the spaces* $\langle\!\langle \mathcal{C}, \langle \exists\mathsf{F}\rangle_{\mathsf{topo}}, \mathsf{FO}\rangle\!\rangle \rtimes_{\mathsf{F}} \langle\!\langle \mathsf{X}, \tau, \mathcal{B}\rangle\!\rangle$ *is defined as follows*

$$\langle\!\langle \mathcal{C} \rtimes \mathsf{X}, \langle \mathcal{C} \rtimes_{\mathsf{F}} \tau\rangle_{\mathsf{topo}}, \bigvee_{\mathcal{W}\subseteq_{\mathsf{fin}}\mathcal{B}} \langle \mathcal{C} \rtimes_{\mathsf{FO}} \mathcal{W}\rangle^{\mathsf{comp}}_{\mathsf{bool}} \rangle\!\rangle \quad .$$

As for the case of $\mathcal{C}^\star$, we first study how compact open subsets of the wreath topology can be generated. The following Lemma 6.3.42 is the generalisation of Lemma 6.3.28 to the general setting of wreath products.

Lemma 6.3.42. *Let* (X, τ) *be a topological space,* σ *be a finite relational signature,* $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$, F *be a fragment of* $\mathsf{FO}[\sigma]$, $\mathfrak{A} \in \mathcal{C}$, *and* $U \colon \mathfrak{A} \to \mathcal{K}^\circ(\mathsf{X})$. *Then,* $[\mathsf{F}\|U]$ *is a compact open subset of* $\langle \mathcal{C} \rtimes_{\mathsf{F}} \tau\rangle_{\mathsf{topo}}$.

Proof. This proof is the analogue of Lemma 6.3.28 in a more general setting. We use Alexander's subbase lemma and consider a family $(V_i\colon \mathfrak{A}_i \to \tau)_{i\in I}$ such that $([\mathsf{F}\|V_i])_{i\in I}$ forms an open cover of $[\mathsf{F}\|U]$.

Let $f\colon \mathfrak{A} \to \mathsf{X}$ be an element of $[\mathsf{F}\|U]$. Then, there exists $i(f) \in I$, and $h_f\colon \mathfrak{A}_{i(f)} \to_{\mathsf{F}} \mathfrak{A}$, such that for all $a \in \mathfrak{A}_{i(f)}$, $f(h_f(a)) \in V_{i(f)}(a)$.

We define $Z_f\colon \mathfrak{A} \to \tau$ as follows:

$$Z_f(a) \stackrel{\text{def}}{=} \bigcap_{a'\in h_f^{-1}(\{\, a \,\})} V_{i(f)}(a') \quad .$$

Let us define $Z_f^{\times}$ as the subset of $\mathsf{X}^{\mathfrak{A}}$ obtained via $\prod_{a\in\mathfrak{A}} Z_f(a)$. It is an open subset of $\mathsf{X}^{\mathfrak{A}}$ in the product topology.

We claim that $f \in [\mathsf{F}\|Z_f]$ which is an open subset of the wreath topology, because the identity map is always an F-embedding. Furthermore, for all $a \in \mathfrak{A}$, $f(a) \in Z_f(a)$ by definition. Hence, $f \in Z_f^{\times}$ when the former is seen as an element of $\mathsf{X}^{\mathfrak{A}}$.

Let $U^{\times} \stackrel{\text{def}}{=} \prod_{a\in\mathfrak{A}} U(a)$, which is a compact subset of $\mathsf{X}^{\mathfrak{A}}$ thanks to the Tychonoff theorem. We have proven that $U^{\times} \subseteq \bigcup_{f\in U^{\times}} Z_f^{\times}$. As a consequence, there exists a finite subset W of $U^{\times}$ such that

$$U^{\times} \subseteq \bigcup_{f\in W} Z_f^{\times} \quad .$$

Let us now conclude that $U \subseteq \bigcup_{f\in W} [\mathsf{F}\|V_{i(f)}]$. For that, consider $g\colon \mathfrak{B} \to \mathsf{X}$ that belongs to U. There exists $h\colon \mathfrak{A} \to_{\mathsf{F}} \mathfrak{B}$ such that for all $a \in \mathfrak{A}$, $g(h(a)) \in U(a)$. Now, $(g \circ h)\colon \mathfrak{A} \to \mathsf{X}$ is an element of $U^{\times}$, hence there exists an $f \in W$ such that $(g \circ h) \in Z_f^{\times}$. In particular, we obtain that

$$\forall a \in \mathfrak{A}, (g \circ h)(a) \in \bigcap_{a'\in h_f^{-1}(\{\, a \,\})} V_{i(f)}(a') \quad .$$

This proves that

$$\forall a' \in \mathfrak{A}_{i(f)}, g((h \circ h_f)(a')) \in V_{i(f)}(a') \quad .$$

Remark that both h and h_f are F-embeddings, hence so is $h \circ h_f$, and we have proven that $g \in [\mathsf{F}\|V_{i(f)}]$. ■

Recall that in the proof of Theorem 6.3.33, there were two alternative endings, one of which relied on the fact that finite words over a finite alphabet are well-quasi-ordered. The latter is equivalent to stating that finite linear orders are ∞-wqo with respect to $\subseteq_i$.[50] Now, let us generalise the notion of ∞-wqo to an arbitrary fragment F.[51]

50: Recall that the notion of ∞-wqo was introduced for $\subseteq_i$ on page 46.

51: Beware that in Definition 6.3.43, there is *a prori* no uniform way to transform a fragment F of $\mathsf{FO}[\sigma]$ into a fragment of $\mathsf{FO}[\sigma \uplus L]$ (with the exception of the identity map). It would be interesting to find a natural setting in which these unary extensions make sense.

Definition 6.3.43. *Let σ be a finite relational signature, $\mathcal{C}$ be a class of finite relational structures over σ, and F be a fragment of $\mathsf{FO}[\sigma]$. We say that $(\mathcal{C}, \leq_{\mathsf{F}})$ is ∞-wqo, or equivalently, that $\mathcal{C}$ is (F, ∞)-wqo, if for every finite subset L of labels, the class $\mathsf{Lab}(L, \mathcal{C})$ is a wqo for* *the following ordering:* $\mathfrak{A} \leq \mathfrak{B}$ *if and only if*

$$\begin{aligned}&\exists h \colon \mathfrak{A}_{|\sigma} \to_{\mathsf{F}} \mathfrak{B}_{|\sigma},\\ &\quad \forall \upsilon \colon \{\, x \,\} \to \mathfrak{A},\\ &\qquad \forall P \in L,\\ &\qquad\quad \mathfrak{A}, \upsilon \models P(x) \iff \mathfrak{A}, h \circ \upsilon \models P(x) \quad .\end{aligned}$$

Theorem 6.3.44. *Let* $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$, F *be a fragment of* FO *such that* $(\mathcal{C}, \leq_{\mathsf{F}})$ *is* ∞*-wqo, and* $\langle\!\langle \mathsf{X}, \tau, \mathcal{B} \rangle\!\rangle$ *be an lpps. Then,* $(\mathcal{C}, \langle \exists \mathsf{F} \rangle_{\mathsf{topo}}) \rtimes_{\mathsf{F}} \langle\!\langle \mathsf{X}, \tau, \mathcal{B} \rangle\!\rangle$ *is an lpps.*

52: This is exactly the same as the proof that the space of finite words is an lps, except for the notation that varies.

Proof. Let us first prove that space is a logically presented space.[52] For that, consider a subbasic open subset $[\mathsf{F} \| U]$, where $U \colon \mathfrak{A} \to \tau$ for some $\mathfrak{A} \in \mathcal{C}$. Because $\langle\!\langle \mathsf{X}, \tau, \mathcal{B} \rangle\!\rangle$ is an lps, for all $a \in \mathfrak{A}$, there exists a family $D_{i,a}$ of definable open subsets of X such that

$$U(a) = \bigcup_{i \in I_a} D_{i,a} \quad .$$

Let us now rewrite $[\mathsf{F} \| U]$ using the above decomposition in definable open subsets as follows:

$$\begin{aligned}(f \colon \mathfrak{B} \to \mathsf{X}) \in [\mathsf{F} \| U] &\iff \exists h \colon \mathfrak{A} \to_{\mathsf{F}} \mathfrak{B}, \forall a \in \mathfrak{A},\\ &\qquad f(h(a)) \in U(a)\\ &\iff \exists h \colon \mathfrak{A} \to_{\mathsf{F}} \mathfrak{B}, \forall a \in \mathfrak{A}, \exists i \in I_a,\\ &\qquad f(h(a)) \in D_{i,a}\\ &\iff \exists I \in \prod_{a \in \mathfrak{A}} I_a, \exists h \colon \mathfrak{A} \to_{\mathsf{F}} \mathfrak{B}, \forall a \in \mathfrak{A},\\ &\qquad f(h(a)) \in D_{I(a),a}\end{aligned}$$

Given $I \in \prod_{a \in \mathfrak{A}} I_a$, $f \colon \mathfrak{B} \to \mathsf{X}$, and $h \colon \mathfrak{A} \to_{\mathsf{F}} \mathfrak{B}$, we define $D_{I,h,f}$ as the set $[\mathsf{FO} \| b \mapsto \bigcap_{a \in h^{-1}(\{\, b \,\})} D_{I(a),a}]$. We conclude by noticing that the following equality holds:

$$[\mathsf{F} \| U] = \bigcup_{I \in \prod_{a \in \mathfrak{A}} I_a} \underbrace{\bigcup_{(f \colon \mathfrak{B} \to \mathsf{X}) \in [\mathsf{F} \| U]} \bigcup_{h \colon \mathfrak{A} \to_{\mathsf{F}} \mathfrak{B}} D_{I,h,f}}_{\text{definable and open}} \quad .$$

Indeed, if $g \colon \mathfrak{B} \to \mathsf{X}$ is such that $g \in [\mathsf{F} \| U]$, there exists $h \colon \mathfrak{A} \to_{\mathsf{F}} \mathfrak{B}$, such that $\forall a \in \mathfrak{A}, g(h(a)) \in U(a)$. In particular, there exists $I \in \prod_{a \in \mathfrak{A}} I_a$ such that $\forall a \in \mathfrak{A}, g(h(a)) \in D_{I(a),a}$. Remark that $\mathsf{Id} \colon \mathfrak{B} \to_{\mathsf{FO}} \mathfrak{B}$ is such that $\forall b \in \mathfrak{B}, g(\mathsf{Id}(b)) = g(b) \in \bigcap_{a \in h^{-1}(\{\, b \,\})} D_{I(a),a})$, which proves that $g \in [\mathsf{FO} \| D_{I,h,g}]$. Conversely, assume that there exists $I \in \prod_{a \in \mathfrak{A}} I_a$, $h \colon \mathfrak{A} \to_{\mathsf{F}} \mathfrak{B}'$, $f \colon \mathfrak{B}' \to \mathsf{X}$, such that $f \in U$ and $g \in [\mathsf{FO} \| D_{I,h,f}]$. The latter implies the existence of $h' \colon \mathfrak{B}' \to_{\mathsf{FO}} \mathfrak{B}$ such that $\forall b' \in \mathfrak{B}', g(h'(b')) \in \bigcap_{a \in h^{-1}(\{\, b' \,\})} D_{I(a),a}$. Remark that $(h' \circ h) \colon \mathfrak{A} \to \mathfrak{B}$ is an F-embedding, because F is a fragment of FO, and that

$$\forall a \in \mathfrak{A}, g(h'(h(a))) \in \left(\bigcap_{a' \in h^{-1}(\{ h(a) \})} D_{I(a'),a'} \right) \subseteq D_{I(a),a} \subseteq U(a) \quad .$$

We have proven that $g \in [\mathsf{F} \| U]$.

Now, let us prove that the space is an lpps. For that, consider a non-empty definable open subset U of the wreath product. We will do the same thing as for Theorems 6.3.13 and 6.3.33. As a consequence, we will skip the details and only highlight the key arguments.

By definition of the wreath product,[53] there exists $\mathcal{W} \subseteq_{\mathsf{fin}} \mathcal{B}$ such that U is an (arbitrary) Boolean combination of a family $[\mathsf{F} \| V_i]$ where $V_i \in \mathcal{C} \rtimes \mathcal{W}$ for $i \in I$.

53: Again, recall that the union of the Boolean subalgebras is *directed*.

Let us write $\equiv_{\mathcal{W}}$ for the equivalence relation of finite index that $\mathcal{W}$ induces over X. Notice that a subset $S \subseteq \mathsf{X}$ that is saturated for $\equiv_{\mathcal{W}}$ automatically belongs to $\langle \mathcal{W} \rangle_{\mathsf{bool}} \subseteq \mathcal{B}$.

Let us fix a structure $\mathfrak{A} \in \mathcal{C}$, and consider $T \in \mathcal{C} \rtimes \tau$ that is maximal such that $[\mathsf{F} \| T] \subseteq U$ and $\mathrm{dom}(T) = \mathfrak{A}$. This maximal element exists because U is a non-empty an open subset, and we fixed the domain to be $\mathfrak{A}$, so that we can apply Zorn's Lemma. Our goal is to prove that for all $a \in \mathfrak{A}$, $T(a)$ is saturated for $\equiv_{\mathcal{W}}$.

This is done by contradiction, using the same techniques as for the finite product. The key argument is that for all $\mathfrak{A} \in \mathcal{C}$, for all $f, g \colon \mathfrak{A} \to \mathsf{X}$ such that $\forall a \in \mathfrak{A}, f(a) \equiv_{\mathcal{W}} g(a)$, $f \in U$ if and only if $g \in U$. Because the proof was only sketched in the case of finite words, we provide a formal one hereafter.

Assume by contradiction that there exists $a_0 \in \mathfrak{A}$ such that $T(a_0)$ is not saturated for $\equiv_{\mathcal{W}}$. This provides us with a pair $x \equiv_{\mathcal{W}} y$ such that $x \in T(a_0)$ and $y \notin T(a_0)$. Let us consider a finite subset $F \subseteq_{\mathsf{fin}} \mathsf{X}$ of representatives for the $\equiv_{\mathcal{W}}$-equivalence classes. For all $z \in \mathsf{X}$, there exists a unique $\iota(z) \in F$ such that $z \equiv_{\mathcal{W}} \iota(z)$. Let us define $\mathcal{F}$ as the finite set of functions f from $\mathfrak{A}$ to X such that $f(a) \in \iota(T(a))$[54] for all $a \in \mathfrak{A} \setminus \{ a_0 \}$, and $f(a_0) \in \{ x, y \}$. Using the "saturation" property of U that was stated in the previous paragraph, we know that $\mathcal{F} \subseteq U$.

54: We use the notation $\iota(T(a))$ to denote the set of $\iota(z)$, for $z \in T(a)$.

For all $f \in \mathcal{F}$, there exists an open neighbourhood $f \in [\mathsf{F} \| V_f] \subseteq U$, for some $V_f \colon \mathfrak{A}_f \to \tau$. As a consequence, there exists a map $h_f \colon \mathfrak{A}_f \to_{\mathsf{F}} \mathfrak{A}$, such that for all $a' \in \mathfrak{A}_f$, $f(h_f(a')) \in V_f(a')$. Let us define

$$Z \stackrel{\mathsf{def}}{=} \bigcap_{f \in \mathcal{F}} \bigcap_{a' \in h_f^{-1}(\{ a_0 \})} V_f(a') \quad .$$

We claim that Z is an open subset of X, because it is a finite[55] intersection of open subsets. Furthermore, we claim that the map T' that maps $a \in \mathfrak{A} \setminus \{ a_0 \}$ to $T(a)$ and maps a_0 to $T(a_0) \cup Z$ is such that $[\mathsf{F} \| T'] \subseteq U$, which contradicts the maximality of T.[56]

55: The intersection is finite because $\mathfrak{A}$ itself is finite!

56: This is because $T(a_0) \subsetneq T(a_0) \cup Z$, because $y \in Z \setminus T(a_0)$.

Let us formally prove the claim that $[\mathsf{F} \| T'] \subseteq U$.[57] Consider $g \colon \mathfrak{B} \to \mathsf{X}$ such that $g \in [\mathsf{F} \| T']$. There exists $h \colon \mathfrak{A} \to_{\mathsf{F}} \mathfrak{B}$ such that for all $a \in \mathfrak{A}$, $g(h(a)) \in T'(a)$. Recall that for all $a \in \mathfrak{A} \setminus \{ a_0 \}$, $g(h(a)) \in T'(a) = T(a)$. Therefore, if $g(h(a_0)) \in T(a_0)$, then $g \in [\mathsf{F} \| T]$ and we conclude

57: This is the formal generalisation of the proof sketch that was given in Theorem 6.3.33.

that $g \in U$. Otherwise, we know that $g(h(a_0)) \in T'(a) \setminus T(a) = Z$. Let us define $g' \colon \mathfrak{B} \to \mathsf{X}$ as follows:[58]

58: In particular, $g'(h(a_0)) = g(h(a_0))$ in the following definition.

$$g'(b) \stackrel{\text{def}}{=} \begin{cases} \iota(g(b)) & \text{if } b \in \mathrm{Im}(h) \setminus \{ h(a_0) \} \\ g(b) & \text{otherwise} \end{cases} .$$

Because for all $b \in \mathfrak{B}$, $g'(b) \equiv_{\mathcal{W}} g(b)$, we know that $g \in U$ if and only if $g' \in U$.

Now, let us prove that $g' \in U$. For that, let us define $f \colon \mathfrak{A} \to \mathsf{X}$ as follows:

$$f(a) \stackrel{\text{def}}{=} \begin{cases} g'(h(a)) & \text{if } a \neq a_0 \\ y & \text{otherwise} \end{cases} .$$

Notice that $f \in \mathcal{F}$. Therefore, for all $a' \in \mathfrak{A}_f$, $f(h_f(a')) \in V_f(a')$. Furthermore, $f(h_f(a')) = g'(h(h_f(a')))$ if $h_f(a') \neq a_0$, and $f(h_f(a')) = y$ otherwise. As a consequence,

$$\forall a' \in \mathfrak{A}_f, h_f(a') \neq a_0 \Rightarrow g'(h(h_f(a')) \in V_f(a') \quad .$$

If $h_f(a') = a_0$, then $g'(h(h_f(a'))) = g'(h(a_0)) = g(h(a_0))$. Because $g(h(a_0)) \in Z \subseteq V_f(a')$, we conclude that $g'(h(h_f(a'))) \in V_f(a')$.

Since $h \circ h_f$ is the composition of two F-embeddings, it is itself an F-embedding, and we have proven that $g' \in [\mathsf{F} \| V_f] \subseteq U$.

Now, we know that for all $a \in \mathfrak{A}$, $T(a)$ is saturated for the $\equiv_{\mathcal{W}}$-equivalence relation. As mentioned earlier, this implies that $T(a) \in \langle \mathcal{W} \rangle_{\mathsf{bool}} \subseteq \mathcal{B}$, for all $a \in \mathfrak{A}$. Therefore, we can use the fact that $\langle\!\langle \mathsf{X}, \tau, \mathcal{B} \rangle\!\rangle$ is an lpps to conclude that $T(a) \in \mathcal{K}^\circ(\mathsf{X})$ for all $a \in \mathfrak{A}$, that is, $T \colon \mathfrak{A} \to \mathcal{K}^\circ(\mathsf{X})$. Thanks to Lemma 6.3.42, this implies that $[\mathsf{F} \| T]$ is a compact open subset of $\langle \mathcal{C} \rtimes_{\mathsf{F}} \tau \rangle_{\mathsf{topo}}$.

Finally, we have proven that U is a union of compact open subsets of the form $[\mathsf{F} \| T]$, where $\mathrm{Im}(T) \subseteq \mathcal{K}^\circ(\mathsf{X}) \cap \langle \mathcal{W} \rangle_{\mathsf{bool}}$. Because $\langle \mathcal{W} \rangle_{\mathsf{bool}}$ is a finite set, and $(\mathcal{C}, \leq_{\mathsf{F}})$ is ∞-wqo, we conclude that this union can be rewritten as a finite one, similarly to the alternative ending of the proof of Theorem 6.3.33 on page 163. Formally, have proven that there exists a family $(T_i \colon \mathfrak{A}_i \to \mathcal{K}^\circ(\mathsf{X}) \cap \langle \mathcal{W} \rangle_{\mathsf{bool}})_{i \in I}$ such that

$$U = \bigcup_{i \in I} [\mathsf{F} \| T_i] \quad .$$

Using the fact that $(\mathcal{C}, \leq_{\mathsf{F}})$ is ∞-wqo, there exists a finite subset $J \subseteq_{\mathsf{fin}} I$ such that for all $i \in I$, there exists $j \in J$, and a map $h \colon \mathfrak{A}_j \to_{\mathsf{F}} \mathfrak{A}_i$, such that $T_i(h(a)) = T_j(a)$ for all $a \in \mathfrak{A}_j$. This last property implies that for all $i \in I$, there exists $j \in J$ such that $[\mathsf{F} \| T_i] \subseteq [\mathsf{F} \| T_j]$.

As a consequence, U is a finite union of the compact open subsets $[\mathsf{F} \| T_j]$ for $j \in J$, hence is itself compact. ■

Again, while Theorem 6.3.44 is useful, it does not play well with preservation theorems, and one leverages the Feferman-Vaught decomposition technique to get back to first order logic. Before we introduce the corresponding theorem, let us define how elements of $\mathcal{C} \rtimes \mathcal{C}'$ are interpreted as relational structures.

Definition 6.3.45. *Let σ, σ' be two finite relational signatures, $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$, and $\mathcal{C}' \subseteq \mathsf{Fin}(\sigma')$. Let $f \in \mathcal{C} \rtimes \mathcal{C}'$, it is of the form $f\colon \mathfrak{A} \to \mathcal{C}'$ for some $\mathfrak{A} \in \mathcal{C}$. We interpret f as a relational structure over $\sigma \uplus \sigma'$, with domain $\{ (a,b)\colon a \in \mathfrak{A}, b \in f(a) \}$, and such that*

- *For all $(R,n) \in \sigma$, $((a_i,b_i))_{1 \leq i \leq n}$ belongs to R^f if and only if $(a_i)_{1\leq i\leq n}$ belongs to $R^{\mathfrak{A}}$,*
- *For all $(R',n) \in \sigma'$, $((a_i,b_i))_{1 \leq i \leq n}$ belongs to R'^f if and only if $a_i = a_j$ for all $1 \leq i,j \leq n$, and $(b_i)_{1\leq i\leq n}$ belongs to $R^{f(a)}$.*

In particular, the equality relation $(=) \in \sigma$, defines an equivalence relation $\equiv$ via $(a,b) \equiv (a',b')$ if and only if $a = a'$, while the equality $(=) \in \sigma'$ defines the usual notion of equality on elements of the domain.

The purpose of Definition 6.3.45 is to permit the evaluation of first order formulas over elements of $\mathcal{C} \rtimes \mathcal{C}'$.

Theorem 6.3.46. *Let $\mathcal{C}, \mathcal{C}'$ be two classes of finite structures, F and F' be two fragments of FO. Assume moreover that $(\mathcal{C}, \leq_{\mathsf{F}})$ is ∞-wqo, $\langle\!\langle \mathcal{C}', \langle \exists \mathsf{F}' \rangle_{\mathsf{topo}}, \mathsf{FO} \rangle\!\rangle$ is an lpps, and that F has diagram formulas. Then, $\langle\!\langle \mathcal{C} \rtimes \mathcal{C}', \langle \mathcal{C} \rtimes_{\mathsf{F}} \langle \exists \mathsf{F}' \rangle_{\mathsf{topo}} \rangle_{\mathsf{topo}}, \mathsf{FO} \rangle\!\rangle$ is an lpps.*

Proof Sketch. Let us first check that the space is a logically presented space. For that, we notice that for all $T\colon \mathfrak{A} \to \mathsf{FO}[\sigma']$, where $\mathfrak{A} \in \mathcal{C}$, there exists a first order sentence $\psi_T \in \mathsf{FO}[\sigma \uplus \sigma']$ such that $[\![\psi_T]\!] = [\mathsf{F} \| T]$. Namely,

$$\psi_T \overset{\mathsf{def}}{=} \exists\, (x_a)_{a \in \mathfrak{A}}\,,\ \bigwedge_{a \in \mathfrak{A}} [f(a)]_{|\equiv x_a} \wedge \mathring{\Delta}^{\mathsf{F}}_{\mathfrak{A}}((x_a)_{a\in\mathfrak{A}}) \quad ,$$

where $[f(a)]_{|\equiv x_a}$ is the first order sentence $f(a)$ with quantifiers relativised to the equivalence class of x_a. To conclude, notice that $\langle \mathcal{C} \rtimes_{\mathsf{F}} \langle \exists \mathsf{F}' \rangle_{\mathsf{topo}} \rangle_{\mathsf{topo}} = \langle \mathcal{C} \rtimes_{\mathsf{F}} \exists \mathsf{F}' \rangle_{\mathsf{topo}}$.

As a consequence of Theorem 6.3.44, it suffices to prove that, given a first-order sentence $\varphi \in \mathsf{FO}[\sigma \uplus \sigma']$, there exists finitely many first-order sentences $\psi_1, \dots, \psi_n \in \mathsf{FO}[\sigma']$ such that

$$[\![\varphi]\!] \in \langle \mathcal{C} \rtimes_{\mathsf{FO}} \{ \psi_1, \dots, \psi_n \} \rangle^{\mathsf{comp}}_{\mathsf{bool}} \quad .$$

To prove the desired result, we use the Feferman-Vaught technique, with a varying set $\mathbb{T}$ of types that will range over $\mathcal{C}$. Let us introduce a variant $\text{a-conv}^{\mathfrak{A}}(\cdot,\cdot)$ of the conversion operator $\text{conv}^{\mathbb{T}}(\Gamma, \varphi)$ by induction on the structure of ψ and for all $\Gamma\colon \mathsf{fv}(\psi) \to \mathfrak{A}$ as follows:

1. $\text{a-conv}^{\mathfrak{A}}(\Gamma, \exists x.\psi) \overset{\mathsf{def}}{=} \bigvee_{T \in \mathfrak{A}} \exists x : T.\text{a-conv}^{\mathfrak{A}}(\Gamma[x \mapsto T], \psi)$,
2. $\text{a-conv}^{\mathfrak{A}}(\Gamma, \psi_1 \vee \psi_2) \overset{\mathsf{def}}{=} \text{a-conv}^{\mathfrak{A}}(\Gamma, \psi_1) \vee \text{a-conv}^{\mathfrak{A}}(\Gamma, \psi_2)$,
3. $\text{a-conv}^{\mathfrak{A}}(\Gamma, \neg\psi) \overset{\mathsf{def}}{=} \neg\text{a-conv}^{\mathfrak{A}}(\Gamma, \psi)$,
4. $\text{a-conv}^{\mathfrak{A}}(\Gamma, \top) \overset{\mathsf{def}}{=} \top$,
5. For $(R,n) \in \sigma$, we define $\text{a-conv}^{\mathfrak{A}}(\Gamma, R(y_1, \dots, y_n)) \overset{\mathsf{def}}{=} \top$ if $(\Gamma(y_1), \dots, \Gamma(y_n)) \in R^{\mathfrak{A}}$, and $\bot$ otherwise,

6. For $(R', n) \in \sigma'$, we define $\text{a-conv}^{\mathfrak{A}}(\Gamma, R'(y_1, \dots, y_n)) \stackrel{\text{def}}{=} R(y_1 : \Gamma(y_1), \dots, y_n : \Gamma(y_n))$ when $\Gamma(y_1) = \dots = \Gamma(y_n)$, and $\bot$ otherwise.

Now, we claim that the following analogue of Lemmas 4.2.17 and 6.3.35 holds for this new conversion. For all $\mathfrak{A} \in \mathcal{C}$, for all $\psi(\vec{x}) \in \mathsf{FO}[\sigma \uplus \sigma']$, for all $f \in \mathcal{C} \rtimes \mathcal{C}'$ such that $\text{dom}(f) = \mathfrak{A}$, for all valuation $\upsilon : \vec{x} \to f$, for all runtime typing $\rho : f \to \mathfrak{A}$, for all $\mathfrak{A}$-typing environments Γ, such that $\Gamma(x) = \rho(\upsilon(x))$ for all $x \in \vec{x}$, and such that $\rho((a, b)) = a$ for all $(a, b) \in f$,[59] the following are equivalent:

- $f, \upsilon \models \psi(\vec{x})$,
- $f, \rho, \upsilon \models^{\mathbb{T}} \text{a-conv}^{\mathfrak{A}}(\Gamma, \psi)$.

The proof of the above claim follows by induction on the formula ψ. The inductive steps are all trivial, and the base cases are direct consequence of how formulas are evaluated on f (see Definition 6.3.45).

Applying the above claim to $f \in \mathcal{C} \rtimes \mathcal{C}'$ of the form $f : \mathfrak{A} \to \mathcal{C}'$, the empty valuation, and the function $\rho : f \to \mathfrak{A}$, that maps an element $(a, b) \in f$ (seen as a relational structure) to a, we conclude that:

$$(\mathfrak{A}, \emptyset \models \varphi) \iff \big(\mathfrak{A}, \rho, \emptyset \models^{\mathbb{T}} \text{a-conv}^{\mathfrak{A}}(\emptyset, \varphi)\big)$$

Let us now proceed exactly as we did in Lemma 6.3.35. Applying Lemma 4.2.19 over $\text{a-conv}^{\mathfrak{A}}(\emptyset, \varphi)$, we obtain a finite set S of monotyped formulas, such that a Boolean combination of formulas in S that is equivalent to $\text{a-conv}^{\mathfrak{A}}(\emptyset, \varphi)$. Notice that a monotyped formula can only talk about one equivalence class for $\equiv$[60] in the structure f. As a consequence, whether $f \models \varphi$ is uniquely determined by the evaluation of monotyped formulas on each of the $\equiv$-equivalence classes of f.

Recall that Lemma 4.2.13 states that evaluating a monotyped formula θ with type $T \in \mathbb{T}$, over a structure $\mathfrak{B}$ with typing function $\rho : a \mapsto T$ is equivalent to directly evaluating $\mathsf{U}\mathbb{T}(\theta)$ over $\mathfrak{B}$.

Furthermore, Lemma 4.2.19 states that the monotyped formulas in S have quantifier rank at most $\text{rk}(\varphi)$. As a consequence, $\{ \mathsf{U}\mathbb{T}(\theta) : \theta \in S \}$ is finite up to logical equivalence, and its size is bounded independently of the structure $\mathfrak{A}$!

What we have proven is that there exists a finite subset S' of (usual) formulas, namely $\{ \mathsf{U}\mathbb{T}(\theta) : \theta \in S \}$, such that for all $f \in \mathcal{C} \rtimes \mathcal{C}'$, their evaluation on the $\equiv$-equivalence classes of f uniquely determines whether $f \models \varphi$. Hence, we have proven that

$$\llbracket \varphi \rrbracket \in \langle \mathcal{C} \rtimes_{\mathsf{FO}} S' \rangle^{\mathsf{comp}}_{\mathsf{bool}} \quad . \qquad \blacksquare$$

While Theorem 6.3.46 might seem complicated to apply, let us prove that it encompasses very generic properties. As a first example, it immediately implies Corollary 6.3.36 when words are seen as coloured linear orders (see Example 6.3.39). Furthermore, replacing linear orders with finite trees, one obtains the analogue of Kruskal's tree theorem for lpps for free. Recall that the tree embedding relation $\leq_{\mathsf{t}}$ was defined in Definition 6.3.23, and that it is dependent on the ordering placed on

[59]: We see f as a relational structure!

[60]: Recall that in Definition 6.3.45, we introduced $\equiv$ for the equivalence relation defined by $(a, b) \equiv (a', b')$ if and only if $a = a'$.

It is also worth noting that in the upcoming chapters one will encounter the topological Kruskal theorem, that defines an ad-hoc topology over finite trees, which happens to be exactly the one described "naturally" in Example 6.3.47.

the labels of the trees, so that $(\subseteq_i)_{\mathsf{t}}$ is a quasi-order over trees where labels are compared using the induced substructure relation.

Example 6.3.47. Let σ be a finite relational signature, and $\mathcal{C} \subseteq \mathsf{Fin}(\sigma)$ be a class of finite structures such that $\langle\langle \mathcal{C}, \mathsf{Alex}(\subseteq_i), \mathsf{FO} \rangle\rangle$ is an lpps.

Then, $\langle\langle \mathsf{T}(\mathcal{C}), \mathsf{Alex}((\subseteq_i)_{\mathsf{t}}), \mathsf{FO} \rangle\rangle$ is an lpps.

Proof. We define Trees to be the class of finite trees with the ancestor relation and the ordering among children of a given node. It is an easy check that $\subseteq_i$ is precisely the tree embedding relation over elements of Trees.

By applying Theorem 6.3.46, we conclude that the following space is an lpps:

$$\langle\langle \mathsf{Trees} \rtimes \mathcal{C}, \langle \mathsf{Trees} \rtimes_{\mathsf{QF}} \langle \mathsf{EFO} \rangle_{\mathsf{topo}} \rangle_{\mathsf{topo}}, \mathsf{FO} \rangle\rangle$$

Recall that $\langle \mathsf{EFO} \rangle_{\mathsf{topo}}$ is precisely $\mathsf{Alex}(\subseteq_i)$ over $\mathcal{C}$, since the latter is composed of finite structures.

Now, it is clear (as it was for finite words in Example 6.3.39), that $\mathsf{T}(\mathcal{C})$ with the Alexandroff topology of $(\subseteq_i)_{\mathsf{t}}$ is precisely $\mathsf{Trees} \rtimes \mathcal{C}$ with the topology generated by $\mathsf{Trees} \rtimes_{\mathsf{QF}} \mathsf{Alex}(\subseteq_i)$.

As a consequence, we conclude that $\langle\langle \mathsf{T}(\mathcal{C}), (\subseteq_i)_{\mathsf{t}}, \mathsf{FO}] \rangle\rangle$ is an lpps. ■

6.3.4. Limit Constructions

One compelling reason to study projective limits of lpps is the following application to the proof that the Homomorphism Preservation Theorem relativises to the finite. While it is known that the category Spec of spectral spaces and spectral maps is closed under projective limits[61] [26, Corollary 2.3.8], unfortunately this fails for lpps.

61: We invite the reader to read the Chapter F (Category Theory Cheat Sheet) if notions of limits are not clear.

Example 6.3.48. Let $\mathsf{X}_n \stackrel{\text{def}}{=} \langle\langle \mathsf{Cycles}, \tau_n, \mathcal{P}(\mathsf{Cycles}) \rangle\rangle$ where τ_n is generated by subsets containing cycles of size at most n, and cofinite subsets. The system $\mathcal{F} \stackrel{\text{def}}{=} (\mathsf{Id}_{i,j} \colon \mathsf{X}_i \to \mathsf{X}_j)_{i \geq j \in \mathbb{N}}$ is a projective system in LPPS, which has no limit in LPPS.

Proof. It is an easy check that X_n is Noetherian for $n \in \mathbb{N}$. As a consequence, the maps $\mathsf{Id}_{i,j}$ are always logical maps and spectral maps. Assume by contradiction that some limit $\{ f_i \colon \mathsf{X} \to \mathsf{X}_i : i \in \mathbb{N} \}$ exists in LPPS. By assumption, $\mathsf{Id}_{i,j} \circ f_j = f_i$, hence $f_i = f_0$ for all $i \in \mathbb{N}$. In particular, $f_0^{-1}(\{ C_n \})$ is an open subset of X for all $n \in \mathbb{N}$. As a consequence, $f_0 \colon \mathsf{X} \to (\mathsf{Cycles}, \tau_{\mathsf{disc}})$ is continuous. However, X is compact, and $(\mathsf{Cycles}, \tau_{\mathsf{disc}})$ is not, which is absurd. ■

We provide here a sufficient condition for the projective limit in LPS to exist in LPPS.

Exercise 6.3.49 ([70, Lemma 7.2])**.** Let $\mathcal{F} \stackrel{\text{def}}{=} (f_{i,j} \colon \mathsf{X}_i \to \mathsf{X}_j)_{i \geq j \in \mathbb{N}}$

be a projective system in LPPS. Let (X, τ) be the limit of $\mathcal{F}$ in Top. If the limiting maps $f_i : \mathsf{X} \to \mathsf{X}_i$ are also spectral maps, then $\mathcal{K}^\circ(\mathsf{X}) = \bigcup_{i \in I} \{ f_i^{-1}(V) : V \in \mathcal{K}^\circ(\mathsf{X}_i) \}$, and $\langle\!\langle \mathsf{X}, \tau, \langle \mathcal{K}^\circ(\mathsf{X}) \rangle_{\mathsf{bool}} \rangle\!\rangle$ is an lpps.

Let us sketch the argument that was developed in [70, Section 7.2], showing how the Homomorphism Preservation Theorem relativises in the finite by rephrasing the core combinatorial arguments from [83].

Example 6.3.50. Let us define X to be $\langle\!\langle \mathsf{Fin}(\sigma), \mathsf{Alex}(\preceq_h), \mathsf{FO} \rangle\!\rangle$, and $\mathsf{X}_n \stackrel{\text{def}}{=} \langle\!\langle \mathsf{Fin}(\sigma), \langle \mathsf{EPFO}^n \rangle_{\mathsf{topo}}, \mathsf{FO}^n \rangle\!\rangle$, where FO^n is the set of first order sentence of quantifier rank at most n, and EPFO^n is the set of positive existential sentences of quantifier rank at most n.

Then, X is the limit of the projective system $(\mathsf{Id}_{i,j} : \mathsf{X}_i \to \mathsf{X}_j)_{i \geq j \in \mathbb{N}}$ in LPPS.

Proof Sketch. It is clear that X_n is a logically presented space for all $n \in \mathbb{N}$, and that the identity between X_n and X_m is a logical map whenever $n \geq m$. Furthermore, it is an lpps because $\langle \mathsf{EPFO}^n \rangle_{\mathsf{topo}}$ is a Noetherian topology as it contains finitely many open subsets. It is clear that $\mathsf{Id}_i : \mathsf{X} \to \mathsf{X}_i$ is a logical map, and in fact a spectral map. Moreover, the core combinatorial lemma of [83, Corollary 5.14] states that a definable open subset U of X is the pre-image of some definable open subset U' of some X_n. Leveraging Exercise 6.3.49, this proves that X is an lpps. ■

6.4. Discussion

Relationship with Locality. As discussed in Example 6.3.37, the compositional theorems on logically presented pre-spectral spaces allows us to go beyond what was possible using the locality-based techniques of Chapter 5 (A Local-to-Global Preservation Theorem). Conversely, lpps do not capture simple preservation theorems, such as Example 6.1.17, which are easily handled via locality. In that sense, the two approaches are orthogonal, and even complementary.

Wreath Products and Noetherian Spaces. Because finite words and finite trees are naturally represented using wreath products, we may wonder whether a generic argument applying to wreath products could encompass both the topological Higman lemma and topological Kruskal tree theorem. This remains an open question.

Furthermore, the limitation to finite structures seems less relevant in this setting, and it would be interesting to compare the wreath product representing infinite words (or even ordinal words) with the ad-hoc topologies that have been defined in the literature [47, 49].

References

[23] Stéphane Demeri, Alain Finkel, Jean Goubault-Larrecq, Sylvain Schmitz and Philippe Schnoebelen. 'Algorithmic Aspects of WQO Theory (MPRI course)'. 2012. URL: https://cel.archives-ouvertes.fr/cel-00727025 (cit. on pp. 3, 139, 178, 181, 239).

[26] Max Dickmann, Niels Schwartz and Marcus Tressl. *Spectral Spaces*. Vol. 35. Cambridge University Press, 2019. DOI: 10.1017/9781316543870.003 (cit. on pp. 56, 146–148, 174, 239, 253).

[33] Solomon Feferman and Robert Vaught. 'The first order properties of products of algebraic systems'. In: *Fundamenta Mathematicae* 47 (1959), pp. 57–103. DOI: 10.4064/fm-47-1-57-103 (cit. on pp. 78, 155).

[44] Jean Goubault-Larrecq. 'Noetherian Spaces in Verification'. In: *Automata, Languages and Programming*. Ed. by Samson Abramsky, Cyril Gavoille, Claude Kirchner, Friedhelm Meyer auf der Heide and Paul G. Spirakis. Springer Berlin Heidelberg, 2010, pp. 2–21. DOI: 10.1007/978-3-642-14162-1_2 (cit. on pp. 53, 139, 178).

[45] Jean Goubault-Larrecq. *Non-Hausdorff Topology and Domain Theory*. Vol. 22. Cambridge University Press, 2013. DOI: 10.1017/cbo9781139524438 (cit. on pp. 3, 50, 52–54, 139, 159, 178, 179, 185, 186, 192, 193, 203, 208, 224, 239, 254, 255).

[47] Jean Goubault-Larrecq. 'Infinitary Noetherian Constructions I. Infinite Words'. In: *Colloquium Mathematicum* 168 (2022), pp. 257–268. DOI: https://doi.org/10.4064/cm8077-4-2021 (cit. on pp. 139, 175, 178, 179, 209–211).

[49] Jean Goubault-Larrecq, Simon Halfon and Aliaume Lopez. *Infinitary Noetherian Constructions II. Transfinite Words and the Regular Subword Topology*. 2022 (cit. on pp. 175, 178, 179, 209, 211).

[59] Graham Higman. 'Ordering by divisibility in abstract algebras'. In: *Proceedings of the London Mathematical Society* 3 (1952), pp. 326–336. DOI: 10.1112/plms/s3-2.1.326 (cit. on pp. 44, 158, 178, 181).

[63] J. C. Kelly. 'Bitopological Spaces'. In: *Proceedings of the London Mathematical Society* s3-13 (1963), pp. 71–89. DOI: 10.1112/plms/s3-13.1.71. URL: https://onlinelibrary.wiley.com/doi/abs/10.1112/plms/s3-13.1.71 (visited on 15/11/2022) (cit. on p. 146).

[65] Joseph B. Kruskal. 'The theory of well-quasi-ordering: A frequently discovered concept'. In: *Journal of Combinatorial Theory, Series A* 13 (1972), pp. 297–305. DOI: 10.1016/0097-3165(72)90063-5 (cit. on pp. 44, 158, 178, 179, 181).

[66] Denis Kuperberg. '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, 2021, pp. 1–13. DOI: 10.1109/LICS52264.2021.9470602 (cit. on pp. 37, 39, 68, 98, 113, 137).

[70] Aliaume Lopez. 'Preservation Theorems Through the Lens of Topology'. In: *Proceedings of CSL'21*. Vol. 183. 2021, 32:1–32:17. DOI: 10.4230/LIPIcs.CSL.2021.32 (cit. on pp. 63, 138, 148, 174, 175).

[74] Johann A. Makowsky. 'Algorithmic uses of the Feferman–Vaught Theorem'. In: *Annals of Pure and Applied Logic* 126 (2004), pp. 159–213. DOI: 10.1016/j.apal.2003.11.002 (cit. on pp. 78, 155).

[83] Benjamin Rossman. 'Homomorphism preservation theorems'. In: *Journal of the ACM* 55 (2008), 15:1–15:53. DOI: 10.1145/1379759.1379763 (cit. on pp. i, 24, 35, 37, 38, 40, 42, 55, 74, 175, 177, 235).

[85] Abhisekh Sankaran, Bharat Adsul and Supratik Chakraborty. 'A generalization of the Łoś–Tarski preservation theorem'. In: *Annals of Pure and Applied Logic* 167 (2016), pp. 189–210 (cit. on pp. 137, 236).

[89] Alexei P. Stolboushkin. 'Finitely monotone properties'. In: (1995), pp. 324–330. DOI: 10.1109/LICS.1995.523267 (cit. on pp. 35, 113, 137).

[90] William W. Tait. 'A counterexample to a conjecture of Scott and Suppes'. In: *Journal of Symbolic Logic* 24 (1959), pp. 15–16. DOI: 10.2307/2964569 (cit. on pp. i, 35, 37, 39, 97, 100, 113, 137).

Topology expanders and Noetherian Topologies 7.

Outline of the chapter

In this chapter, we try to understand what makes the limit constructions of lpps so complicated. In the continuity of the algebraic and topological approach that started in "Chapter 6 (Logically Presented Spaces)", we want to provide a systematic way to produce limits of pre-spectral spaces.

Goals of the chapter

At the end of this chapter, I hope you will be convinced that the topological minimal bad sequence arguments generalise the usual proof schemes in wqos and Noetherian spaces. It should also be clear that topological minimal bad sequence arguments are not easily transported to lpps.

Genesis. One reason to study the limits of lpps is to tackle the Homomorphism Preservation Theorem that was proven to relativise in the finite using techniques that felt like iterative constructions [83]. However, this also offers a new way to consider spaces of structures that are defined inductively. For now, we have seen finite words as coloured linear orderings in Example 6.3.39, but it may be more natural to use their definition as a fixed-point construction when deciding which sets are definable.

[83]: Rossman (2008), 'Homomorphism preservation theorems'

To simplify the approach, the first step was to study how one could build the limits of lpps when the definable sets were fixed, and to further simplify, we actually considered $\mathcal{B} = \mathcal{P}(\mathsf{X})$, i.e., Noetherian spaces.

Apart from being a first step towards a better understanding of lpps, the study of inductively defined Noetherian spaces and wqos also follows from a long-lasting question. Since the M2 course on well-quasi-orderings, where the minimal bad sequence arguments [76] were introduced to me, I wanted to understand better the similarity between Higman's lemma and Kruskal's tree theorem. More precisely, I wanted to understand why they were not simple applications of a more generic theorem based on minimal bad sequences. Furthermore, the topological analogues of the subword embedding and tree embedding have involved[1] definitions, are proven to be Noetherian topologies via a topological variant of the minimal bad sequence argument. We refer to Figure 7.1 for a graphical depiction of this evolution.

[76]: Nash-Williams (1965), 'On well-quasi-ordering transfinite sequences'

1: And somehow arbitrary

The first idea that made the following chapter possible came in 2020 when the "canonicity" issues of Noetherian spaces where mitigated by focusing on the "destructors" associated with inductive constructions (words, trees) rather than their constructors [51, 69]. However, this "destructive approach" did not yield a generic theorem and was only

[51]: Goubault-Larrecq and Lopez (2020), 'On the word topology, and beyond'
[69]: Lopez (2020), *On the word topology, and beyond*

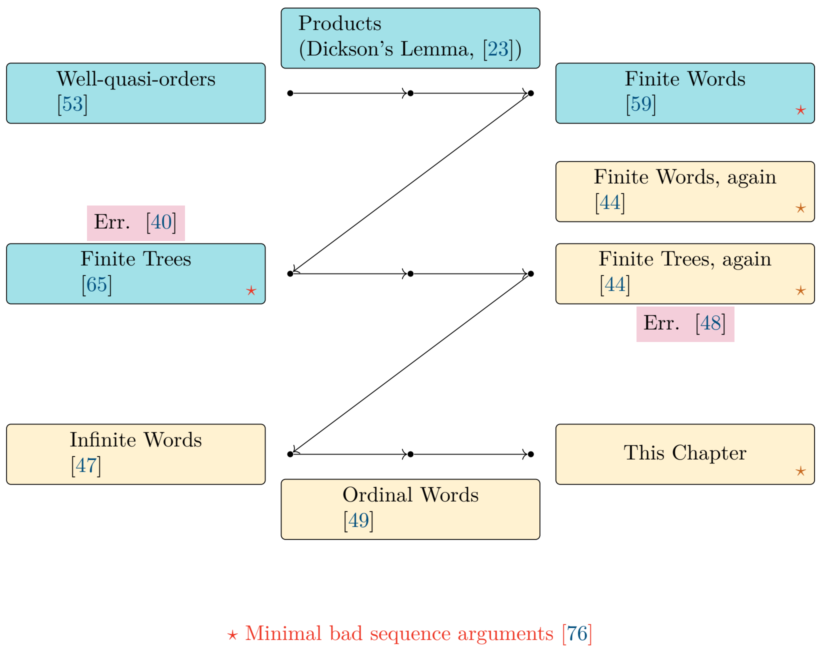


Figure 7.1.: Increasing complexity of the stability of wqos and Noetherian spaces under operations. Green boxes represent theorems about Noetherian topologies, stars indicate the use of a minimal bad sequence argument, and "Err." corresponds to erratums that correct flaws in the original proofs of the result.

able to justify *a posteriori* the definitions of the subword topology and the tree topology.

The second idea came from the background process of finding good restrictions to construct limits of Noetherian spaces. While these limits are not Noetherian in general, a nice restriction was found so that the topological minimal bad sequence argument applies. This is the core idea behind the definition of topology expanders, which is the main contribution in this chapter.

[23]: Demeri, Finkel, Goubault-Larrecq, Schmitz and Schnoebelen (2012), 'Algorithmic Aspects of WQO Theory (MPRI course)'
[45]: Goubault-Larrecq (2013), *Non-Hausdorff Topology and Domain Theory*

Why Inductive Types? As for well-quasi-orders, Noetherian spaces are stable under finite products and finite sums [23, 45]. While this is enough to describe the set of configurations of simple programs computing over natural numbers by endowing $\mathbb{N}^k$ with the product ordering, it is not sufficient to tackle complex data structures that are typically defined inductively, such as lists and trees. To make the above statement precise, let $\mathbb{1}$ be the singleton set, $A + B$ be the disjoint union of A and B, and $A \times B$ their Cartesian product.

Fact 7.0.1. Let Σ be an alphabet. The set Σ^* of finite words over Σ is precisely the least fixed point of $\mathrm{CWords}_\Sigma \colon X \mapsto \mathbb{1} + \Sigma \times X$.

Fact 7.0.2. Let Σ be an alphabet. The set of finite trees labelled by Σ is precisely the least fixed point of $\mathrm{CTrees}_\Sigma \colon X \mapsto \Sigma \times X^*$.

Note that these constructions can be generalised, and we will write $\mathsf{lfp}_X.F(X)$ to denotes the least fixed point of a constructor F whenever it exists.

[59]: Higman (1952), 'Ordering by divisibility in abstract algebras'

In the realm of well-quasi-orderings, the specific cases of finite words and finite trees are handled respectively via Higman's Lemma [59] and

Kruskal's Tree Theorem [65].[2] Let us recall that a word u embeds into a word w (written $u \leq_{\mathsf{w}} v$) whenever there exists a strictly increasing map $h\colon |w| \to |w'|$ such that $w_i \leq w_{h(i)}$ for $1 \leq i \leq |w|$ (see Definition 6.3.22). Similarly, a tree t embeds into a tree t' (written $t \leq_{\mathsf{t}} t'$) whenever there exists a map from nodes of t to nodes of t' respecting the least common ancestor relation, and respecting the ordering on the colours of the nodes (see Definition 6.3.23).

[65]: Kruskal (1972), 'The theory of well-quasi-ordering: A frequently discovered concept'

2: The two theorems were introduced in Theorem 6.3.24 and Theorem 6.3.25, in the study of wreath products.

Let us first remark that the ordering placed on finite words and finite trees do not seem to clearly derive from the inductive characterisations using CWords and CTrees, and correspond to the induced substructure relation when considering words and trees as relational structures over a suitable relational signature (see Example 6.3.39). Even though the definitions of $\leq_{\mathsf{w}}$ and $\leq_{\mathsf{t}}$ do not rely on the inductive nature of finite words and finite trees, proofs that they are well-quasi-orders[3] rely on a minimal bad sequence argument due to [76], where the inductive nature of words and trees is crucial to the well-foundedness of the argument.

3: Assuming, of course, that the alphabet itself is well-quasi-ordered.

[76]: Nash-Williams (1965), 'On well-quasi-ordering transfinite sequences'

However, minimal bad sequence arguments are quite subtle, and must be handled with care [40, 88]. In addition, this type of argument is not compositional and has to be slightly modified whenever a new inductive construction is desired. This type of careful manipulations can be found for instance in [19, 24]. This situation has been adapted by Goubault-Larrecq to the topological setting by proposing analogues of the word embedding and tree embedding, together with a proof that they preserve Noetherian spaces [45, Section 9.7]. However, both the definitions and the proofs have an increased complexity, as they rely on an adapted "topological minimal bad sequence argument" that appears to be even more subtle than the classical one [48, errata n. 26]. Moreover, the newly introduced topologies have involved definitions often relying on ad-hoc constructions.

[40]: Gallier (1997), 'Ann. Pure Appl. Logic: Erratum to "What's so special about Kruskal's Theorem and the ordinal Γ_0? A survey of some results in proof theory" [53 (1991) 199–260]'

[88]: Singh, Shuaibu and Ndayawo (2013), 'Simplified Proof of Kruskal's Tree Theorem'

[19]: Daligault, Rao and Thomassé (2010), 'Well-Quasi-Order of Relabel Functions'

[24]: Dershowitz and Tzameret (2003), 'Gap Embedding for Well-Quasi-Orderings'

[45]: Goubault-Larrecq (2013), *Non-Hausdorff Topology and Domain Theory*

[48]: Goubault-Larrecq (2022), 'Non-Hausdorff Topology and Domain Theory. Electronic supplements to the book – Errata'

Towards a Generic Framework In the realm of wqos, two generic fixed point constructions have been proposed to handle inductively defined datatypes [36, 58]. In these frameworks, $\mathsf{lfp}_X.F(X)$ is guaranteed to be a well-quasi-ordering provided that F is a "well-behaved functor" of quasi-orders. Both proposals, while relying on different categorical notions, successfully recover Higman's word embedding and Kruskal's tree embedding through their respective definitions as least fixed points. As a side effect, they reinforce the idea that these two quasi-orders are somehow canonical by demonstrating how they emerge from the inductive definitions themselves and not through an encoding into relational structures.

[36]: Freund (2020), 'From Kruskal's theorem to Friedman's gap condition'

[58]: Hasegawa (2002), 'Two applications of analytic functors'

For Noetherian spaces, no equivalent framework exists to build inductive data types, and the notions of "well-behaved" constructors from [36, 58] rule out the use of important Noetherian spaces, as they require that an element $a \in F(X)$ has been built using *finitely many* elements of X: while this is the case for finite words and finite trees, it does not hold for the arbitrary powerset. Moreover, there have been recent advances in placing Noetherian topologies over spaces that are not straightforwardly obtained through "well-behaved" definitions, such as infinite words or even ordinal length words [47, 49].

[47]: Goubault-Larrecq (2022), 'Infinitary Noetherian Constructions I. Infinite Words'

[49]: Goubault-Larrecq, Halfon and Lopez (2022), *Infinitary Noetherian Constructions II. Transfinite Words and the Regular Subword Topology*

[72]: Lopez (2023), 'Fixed Points and Noetherian Topologies'

Contributions The results of this section are mostly taken from [72]. In Section 7.1, we explain in depth the classical minimal bad sequence argument due to Nash-Williams. We reformulate it in a way that makes the similarities between the proof of Higman's Lemma and Kruskal's Tree Theorem apparent, and highlights the main difficulties one encounters when trying to generalise them.

Then, we follow in Section 7.2 a topological approach based on the strange idea of *fixing the carrier set* instead of simultaneously constructing the space and its topology. This leads to the definition of topology expander, and their associated Theorem 7.2.33 stating the existence of a least-fixed-point topology that is Noetherian.

We conclude the chapter by studying the structural properties of topology expanders in Section 7.3. Even though the study of Noetherian spaces was thought as a first step towards a similar theorem for logically presented pre-spectral spaces, we discuss at the end of the chapter why the current techniques cannot apply in this setting.

7.1. "Commentaire Comparé"

In order to build intuition and before introducing more involved arguments, let us restate the proofs of Higman's Lemma and Kruskal's tree theorem to highlight their similarities and point out key articulations. In order to compare the two proofs, only the case of finite words will be detailed in the main document, and the differences with the case of finite trees will be described in margin notes.

7.1.1. Inductively Defining the Ordering

$$\text{(init)}\ \frac{}{\vdash \varepsilon \leq_{\mathsf{w}} w}$$

$$\text{(incr)}\ \frac{\vdash u \leq_{\mathsf{w}} v \qquad a \leq b}{\vdash au \leq_{\mathsf{w}} bv}$$

$$\text{(disc)}\ \frac{\vdash u \leq_{\mathsf{w}} v}{\vdash u \leq_{\mathsf{w}} bv}$$

Figure 7.2.: An inductive characterisation of the subword ordering.

As a first step in most of the minimal bad sequence arguments, we will extract from the definition of $\leq_{\mathsf{w}}$ and $\leq_{\mathsf{t}}$ inductive characterisations. These will be crucial to the well-foundedness of our proof.

Lemma 7.1.1. *Let $(P, \leq)$ be a quasi-ordering, and $u, v \in P^\star$. Then, $u \leq_{\mathsf{w}} v$ if and only if there is a derivation of $\vdash u \leq_{\mathsf{w}} v$ in the proof system of Figure 7.2.*

Proof. It is clear that the rules in Figure 7.2 are admissible for $\leq_{\mathsf{w}}$, hence we only show that they are complete by induction on the length of u.

Whenever $|u| = 0$, then $u = \varepsilon$. In this case, $\vdash u \leq_{\mathsf{w}} v$ is derivable for all $v \in P^\star$, and there is a word embedding from u to every word $w \in P^\star$.

Let $u = au'$. Assume that $u \leq_{\mathsf{w}} v$, there exists a word embedding $h \colon u \to v$ such that $u_i \leq v_{h(i)}$. In particular, $a \leq v_{h(0)}$. Let us define $v' \stackrel{\text{def}}{=} v_{h(0)+1} \cdots v_{|v|-1}$. By induction, since $u' \leq_{\mathsf{w}} v'$ we have a proof that $\vdash u' \leq_{\mathsf{w}} v'$. Hence, $\vdash u \leq_{\mathsf{w}} v_{h(0)} v'$ is derivable. As a consequence, $\vdash u \leq_{\mathsf{w}} v$. ∎

The advantage of the inductive description is that it isolates a part of the definition of $\leq_{\mathsf{w}}$ that does not depend on the order $\leq$ placed over the letters, obtained by removing the rule (*incr*) from the description in Figure 7.2, and enforcing reflexivity, as detailed in Figure 7.3. We call this ordering the *suffix ordering*, because it is an easy check that $\vdash u \leq_{\mathsf{suf}} v$ if and only if u is a suffix of v in the usual sense.

$$\text{(init)}\ \frac{}{\vdash \varepsilon \leq_{\mathsf{suf}} w}$$

$$\text{(refl)}\ \frac{u = v}{\vdash u \leq_{\mathsf{suf}} v}$$

$$\text{(disc)}\ \frac{\vdash u \leq_{\mathsf{suf}} v}{\vdash u \leq_{\mathsf{suf}} bv}$$

Figure 7.3.: An inductive characterisation of the suffix ordering.

It is a simple exercise to adapt the inductive definitions given for $\leq_{\mathsf{suf}}$ and $\leq_{\mathsf{w}}$ to finite trees, as witnessed by Figures 7.4 and 7.5, respectively defining the tree embedding relation $\leq_{\mathsf{t}}$ and the *structural subtree* relation $\leq_{\mathsf{t\text{-}suf}}$.

$$\text{(incr)}\ \frac{f \leq g \qquad \vdash \vec{s}\ (\leq_{\mathsf{t}})_{\mathsf{w}}\ \vec{t}}{\vdash f(\vec{s}) \leq_{\mathsf{t}} g(\vec{t})}$$

$$\text{(disc)}\ \frac{\exists 1 \leq i \leq n. \vdash s \leq_{\mathsf{t}} t_i}{\vdash s \leq_{\mathsf{t}} f(t_1, \dots, t_n)}$$

Figure 7.4.: An inductive characterisation of the tree embedding.

Exercise 7.1.2. Let $(P, \leq)$ be a quasi-ordering, and $s, t \in \mathsf{T}(P)$. Then, $s \leq_{\mathsf{t}} t$ if and only if there is a derivation of $\vdash s \leq_{\mathsf{t}} t$ in the proof system of Figure 7.4.

It might seem that the introduction of the “structural orderings” ($\leq_{\mathsf{suf}}$,$\leq_{\mathsf{t\text{-}suf}}$), and the detour through an inductive definition of $\leq_{\mathsf{w}}$ are unnecessary. While not *stricto sensu* necessary, they prefigure in a simpler setting the abstract constructions to come. The inductive definition of $\leq_{\mathsf{suf}}$ follows the construction of $P^\star$ as $P^\star \simeq \mathrm{CWords}_P(P^\star)$, i.e., $P^\star \simeq 1 + P \times P^\star$. The suffix ordering can equivalently be defined using this equivalence by stating that $(a, u) \geq u$ for all $a \in P$ and $u \in P^\star$.

$$\text{(refl)}\ \frac{s = t}{\vdash s \leq_{\mathsf{t\text{-}suf}} t}$$

$$\text{(disc)}\ \frac{\exists 1 \leq i \leq n. \vdash s \leq_{\mathsf{t\text{-}suf}} t_i}{\vdash s \leq_{\mathsf{t\text{-}suf}} g(t_1, \dots, t_n)}$$

Figure 7.5.: An inductive characterisation of the subtree ordering.

Fact 7.1.3. Let $(P, \leq)$ be a quasi-order. The relations $\leq_{\mathsf{suf}}$ and $\leq_{\mathsf{t\text{-}suf}}$ are well-founded.

Lemma 7.1.4. *Let $(P, \leq)$ be a quasi-order. Then $\leq_{\mathsf{w}}\leq_{\mathsf{suf}}\ = \leq_{\mathsf{w}}$ and $\leq_{\mathsf{t}}\leq_{\mathsf{t\text{-}suf}}\ = \leq_{\mathsf{t}}$.*

Proof. Because $\leq_{\mathsf{w}}$ and $\leq_{\mathsf{t}}$ are transitive, it suffices to notice that $\leq_{\mathsf{suf}}\ \subseteq\ \leq_{\mathsf{w}}$ and $\leq_{\mathsf{t\text{-}suf}}\ \subseteq\ \leq_{\mathsf{t}}$, which follows from their respective inductive characterisations. ∎

7.1.2. The Minimal Bad Sequence Argument(s)

Now, let us prove Higman's Lemma following the traditional proofs that rely on Nash-Williams's *minimal bad sequence argument*. The proof that follows derives from [23, Lemma 1.9].

[23]: Demeri, Finkel, Goubault-Larrecq, Schmitz and Schnoebelen (2012), ‘Algorithmic Aspects of WQO Theory (MPRI course)’

Beware that the “combined proof” will prove Theorems 6.3.24 and 6.3.25 at the same time. This can be done because only minor variations are necessary to go from one to the other, and therefore can and will be written in the margin. The original proofs of these two results can be found in [59, 65].

[59]: Higman (1952), ‘Ordering by divisibility in abstract algebras’
[65]: Kruskal (1972), ‘The theory of well-quasi-ordering: A frequently discovered concept’

Theorem 7.1.5 Combined theorem**.** Let $(W, \leq)$ be a wqo. Then both $(W^\star, \leq_{\mathsf{w}})$ and $(\mathsf{T}(W), \leq_{\mathsf{t}})$ are wqo.

Combined proof. We are going to assume for contradiction that there exists a bad sequence in $W^\star$. Because $\leq_{\mathsf{w}}$ is defined inductively over words, if such a sequence exists, we can try to build one using the smallest possible words, an intuition that we now make precise.

For trees: We build a bad sequence $(t_i)_{i\in\mathbb{N}}$ that is minimal for $\leq_{\mathsf{t\text{-}suf}}$. The same arguments apply when replacing $\leq_{\mathsf{suf}}$ by $\leq_{\mathsf{t\text{-}suf}}$.

Using the axiom of choice, there exists a bad sequence $(w_i)_{i\in\mathbb{N}}$, such that every sequence of the form $w_0,\dots,w_n,v,\dots$ with $v <_{\mathsf{suf}} w_{n+1}$ is a good sequence. To construct this sequence, we start by selecting a minimal w_0 for $\leq_{\mathsf{suf}}$ starting a bad sequence, which exists because $\leq_{\mathsf{suf}}$ is well-founded, and because we have assumed that a bad sequence exists. We continue to build this sequence iteratively by selecting a minimal w_i for $\leq_{\mathsf{suf}}$ such that $w_0,\dots,w_i$ starts a bad sequence. Notice that at every finite step, the constructed sequence is bad. The infinite sequence remains bad because for every $i,j \in \mathbb{N}$ such that $i < j$, $w_0,\dots,w_j$ is a bad sequence, hence $\neg(w_i \leq w_j)$.

For trees: we construct the set $S \overset{\text{def}}{=} \{ s \in \mathsf{T}(W) \colon \exists i \in \mathbb{N}, s <_{\mathsf{t\text{-}suf}} t_i \}$.

Let us now define $S \overset{\text{def}}{=} \{ v \in W^\star \colon \exists i \in \mathbb{N}. v <_{\mathsf{suf}} w_i \}$. We are now claiming that $(S, \leq_{\mathsf{w}})$ is a wqo. For that, assume that $(v_i)_{i\in\mathbb{N}}$ is a bad sequence in S.

For trees: we name the sequence $(s_i)_{i\in\mathbb{N}}$.

By definition of S, there exists a map $\rho\colon \mathbb{N} \to \mathbb{N}$ such that $v_i <_{\mathsf{suf}} w_{\rho(i)}$. Without loss of generality (using Lemma C.0.5), one can extract from the sequence v_i a subsequence so that $(v_i)_{i\in\mathbb{N}}$ remains infinite, bad, and satisfies that ρ is non-decreasing.

For trees: let $o_i \overset{\text{def}}{=} t_i$ if $i < \rho(0)$, $o_i \overset{\text{def}}{=} s_{i-\rho(0)}$ otherwise.

Consider $u_i \overset{\text{def}}{=} w_i$ if $i < \rho(0)$ and $v_{i-\rho(0)}$ otherwise. Because $u_{\rho(0)} = v_0$ is a strict suffix of $w_{\rho(0)}$, the sequence $(u_i)_{i\in\mathbb{N}}$ is lexicographically smaller than $(w_i)_{i\in\mathbb{N}}$ for $\leq_{\mathsf{suf}}$, hence is a good sequence. As a consequence, there exists a pair $i < j$ such that $u_i \leq_{\mathsf{w}} u_j$. We now do a case analysis that can be visualised in Figure 7.6.

If $i < j < \rho(0)$: then $w_i = u_i \leq_{\mathsf{w}} u_j = w_j$, hence $(w_i)_{i\in\mathbb{N}}$ is a good sequence, which is absurd.

If $\rho(0) \leq i < j$: then $v_{i-\rho(0)} = u_i \leq_{\mathsf{w}} u_j = v_{j-\rho(0)}$. Therefore, $(v_i)_{i\in\mathbb{N}}$ is a good sequence, which is absurd.

If $i < \rho(0) \leq j$: then $w_i = u_i \leq_{\mathsf{w}} u_j = v_{j-\rho(0)}$. By definition of ρ, $v_{j-\rho(0)} \leq_{\mathsf{suf}} w_{\rho(j-\rho(0))}$. Furthermore, $i < \rho(0)$ implies $i < \rho(j - \rho(0))$ because ρ is non-decreasing. As a consequence, $w_i \leq_{\mathsf{w}}\leq_{\mathsf{suf}} w_{\rho(j-\rho(0))}$, but since $\leq_{\mathsf{w}}\leq_{\mathsf{suf}} = \leq_{\mathsf{w}}$ (see Lemma 7.1.4), we conclude that $(w_i)_{i\in\mathbb{N}}$ is a good sequence which is absurd.

For trees: we use the fact that $\leq_{\mathsf{t\text{-}suf}}\leq_{\mathsf{t}}$ corresponds to $\leq_{\mathsf{t}}$, using the same lemma.

For trees: we let $T \overset{\text{def}}{=} (W, \leq) \times (S^\star, (\leq_{\mathsf{t}})_{\mathsf{w}})$, following the inductive constructor $\mathrm{CTrees}_W(S)$.

For trees: we let $\delta_{\mathsf{t}}(f(\vec{t})) \overset{\text{def}}{=} (f, \vec{t})$.

For trees: We obtain that $t_i = f_i(t'_1,\dots,t'_n)$, $t_j = f_j(s'_1,\dots,s'_k)$, with $f_i \leq f_j$ and $\vec{t'} \ (\leq_{\mathsf{t}})_{\mathsf{w}} \ \vec{s'}$. The increment rule from Figure 7.4 allows us to conclude.

Now, because $(S, \leq_{\mathsf{w}})$ is wqo, one can build $T \overset{\text{def}}{=} \mathrm{CWords}_W(S) = 1 + (W, \leq) \times (S, \leq_{\mathsf{w}})$ with the sum / product ordering to form a new wqo (see Table 6.1). For every $i \in \mathbb{N}$, either $w_i = \varepsilon$, or $w_i = a_i w'_i$ with $a \in W$ and $w'_i \in S$. We let $\delta_{\mathsf{w}}(\varepsilon) \overset{\text{def}}{=} 1$ and $\delta_{\mathsf{w}}(au) \overset{\text{def}}{=} (a, u)$. Using this notation, $(\delta_{\mathsf{w}}(w_i))_{i\in\mathbb{N}}$ is an infinite sequence in T, hence there exists $i < j$ such that $\delta_{\mathsf{w}}(w_i) \leq_T \delta(w_j)$. This happens when $w_i = w_j = \varepsilon$, or when $w_i = a_i w'_i$, $w_j = a_j w'_j$ with $a_i \leq a_j$ and $w'_i \leq_{\mathsf{w}} w'_j$. The increment/init rules from Figure 7.2 then prove that $w_i \leq_{\mathsf{w}} w_j$. We have arrived at a contradiction by showing that $(w_i)_{i\in\mathbb{N}}$ is a good sequence. ■

As witnessed by the few number of differences in the margin, the proofs of Kruskal's tree theorem and Higman's lemma are structurally similar. Apart from the syntactic difference (words w versus trees t), there are two main variations in the proofs, and both appear at the very

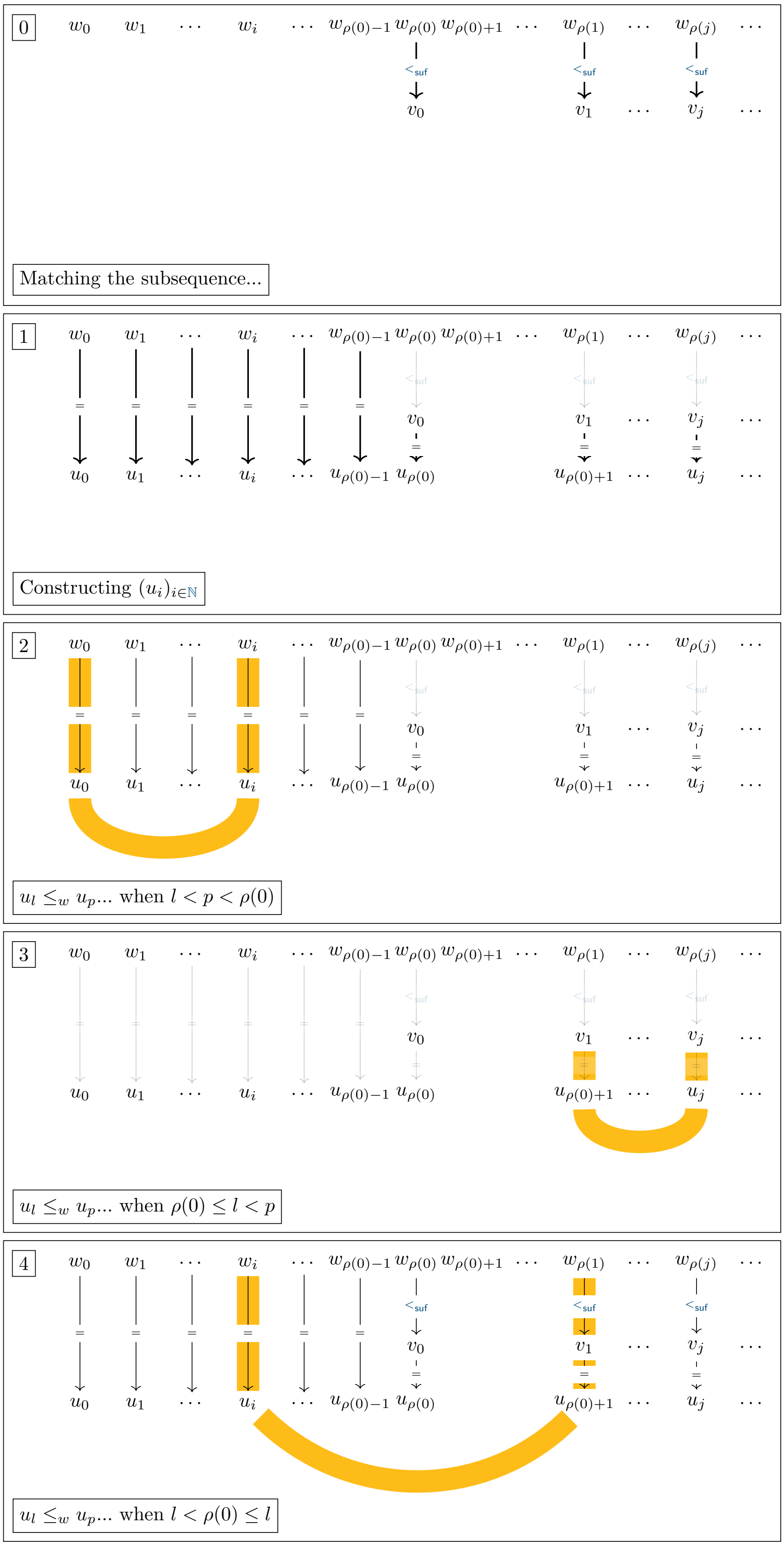


Figure 7.6.: Illustration of the merging process in Higman's Lemma.

end: we moved from CWords to CTrees, and consequently changed the “destructor function” from δ_{w} to δ_{t}. This combined presentation of the minimal bad sequence argument highlights that one only needs to tweak these two parameters to tackle new inductive constructions. Bear in mind however, that this tweaking will involve an inductive presentation of the preorder which can be difficult to obtain.

7.2. Inductive Constructions of Noetherian Spaces

Before we introduce the notion of refinement function, let us explain the philosophy behind our approach. Because the spaces of interest are inductively defined, the most natural way to define a topology/quasi-order is to do it *together* with the construction of the space. This is the “categorical” vision appearing in both [58] and [36]. A more practical approach is to assume that the space is already constructed, without any ordering or topology. On the upside, this greatly simplifies the statements and broadens the range of allowed operations;[4] while the downside is that this setting is too permissive to directly work with.[5] The problems of Subsection 7.2.1 will be corrected by adding some kind of “naturality condition” in Subsection 7.2.2, leading to the notion of topology expander[6].

4: As one can define ad-hoc topologies for a given space.

5: Understand, the constructions lack some “naturality.”

6: See Definition 7.2.17 p. 189

7.2.1. Topology Refinements

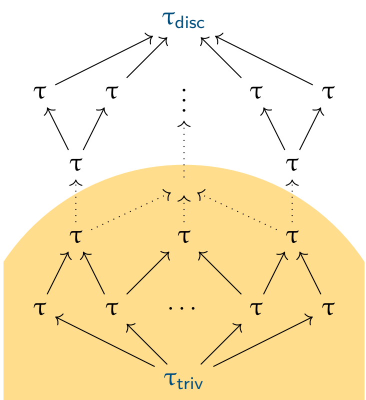


Figure 7.7.: The lattice of topologies over a given set X, where arrows denote inclusions and the coloured part covers the Noetherian topologies. As depicted, the join of two Noetherian topologies is Noetherian, and the collection of Noetherian topologies is downwards closed for inclusion.

Let X be a set, then the set $\mathrm{Top}(\mathsf{X})$ of all topologies over X is a complete lattice for inclusion. Remark that the subset $\mathrm{NTop}(\mathsf{X})$ of $\mathrm{Top}(\mathsf{X})$ composed of Noetherian topologies is a downwards closed subset of $\mathrm{Top}(\mathsf{X})$ for inclusion. In this lattice, the least element is the *trivial topology* $\tau_{\mathsf{triv}} \stackrel{\text{def}}{=} \{\emptyset, \mathsf{X}\}$, and the largest element is the *discrete topology* $\tau_{\mathsf{disc}} \stackrel{\text{def}}{=} \mathcal{P}(\mathsf{X})$. Thanks to the Knaster-Tarski fixed point theorem, every monotone function R mapping topologies over X to topologies over X has a least fixed point, which can be obtained by transfinitely iterating R from the trivial topology. Writing $\mathsf{lfp}_{\tau}.R(\tau)$ for the least fixed point of R, our goal is to provide sufficient conditions for $(\mathsf{X}, \mathsf{lfp}_{\tau}.R(\tau))$ to be Noetherian.

Definition 7.2.1. *A* refinement function *over a set X is a function R mapping topologies over X to topologies over X, such that*

- $\mathsf{R}(\tau)$ *is Noetherian whenever* τ *is, and*
- $\mathsf{R}(\tau) \subseteq \mathsf{R}(\tau')$ *when* $\tau \subseteq \tau'$.

As $(\mathsf{X}, \tau_{\mathsf{triv}})$ is always Noetherian, $(\mathsf{X}, \mathsf{R}^n(\tau_{\mathsf{triv}}))$ is Noetherian for all $n \in \mathbb{N}$ and refinement function R. However, it remains unclear whether the transfinite iterations needed to reach a fixed point preserve Noetherian spaces.

Recall that given a quasi-order $(P, \leq)$ and a set $E \subseteq P$, the upward closure of E is written $\uparrow_{\leq} E$, and is defined as the set of elements that are greater or equal than some element of E in P.

We demonstrate in Example 7.2.2 how to obtain the topology $\mathsf{Alex}(\leq)$ over $\mathbb{N}$ as a least fixed point of some simple refinement function.

Example 7.2.2 (Natural Numbers)**.** Over $X \stackrel{\text{def}}{=} \mathbb{N}$, one can define $\mathsf{Div}(\tau)$ as the collection of sets $\uparrow_{\leq}(U+1)$ for $U \in \tau$, plus $\mathbb{N}$ itself.[7] It is an easy check that $\mathsf{lfp}_{\tau}.\mathsf{Div}(\tau)$ is precisely $\mathsf{Alex}(\leq)$.

7: The empty set $\emptyset$ satisfies $\uparrow_{\leq}(\emptyset + 1) = \emptyset$ and does not have to be added explicitly.

Proof. Let us remark that $\mathsf{Div}(\tau_{\mathsf{triv}}) = \{\emptyset, \uparrow_{\leq}1, \mathbb{N}\}$, and similarly that $\mathsf{Div}^2(\tau_{\mathsf{triv}}) = \{\emptyset, \uparrow_{\leq}1, \uparrow_{\leq}2, \mathbb{N}\}$. More generally, one can prove by induction that for every $k \geq 0$, $\mathsf{Div}^k(\tau_{\mathsf{triv}}) = \{\emptyset, \uparrow_{\leq}1, \ldots, \uparrow_{\leq}k, \mathbb{N}\}$. As a consequence, $\mathsf{lfp}_{\tau}.\mathsf{Div}(\tau)$ is composed of every upwards-closed set of $\mathbb{N}$, i.e, is exactly $\mathsf{Alex}(\leq)$. ∎

In the particular case of Example 7.2.2, one deduces that the least fixed point is Noetherian from the fact that $(\mathbb{N}, \leq)$ is a well-quasi-order. However, not all refinement functions behave as nicely as in Example 7.2.2, and one can obtain non-Noetherian topologies via their least fixed points.

An Ill-behaved Example In this section, let us consider $\Sigma \stackrel{\text{def}}{=} \{a, b\}$ with the discrete topology, i.e., $\{\emptyset, \{a\}, \{b\}, \Sigma\}$. Let us now build the set $\Sigma^\star$ of finite words over Σ. Whenever U and V are subsets of $\Sigma^\star$, let us write UV for their concatenation, defined as $\{uv : u \in U, v \in V\}$. To construct an ill-behaved refinement function, we will associate with a topology τ the set $\{UV : U \in \{\emptyset, \{a\}, \{b\}, \Sigma\}, V \in \tau\}$. However, the latter fails to be a topology in general. This problem frequently appears in this paper, and is solved by considering the so-called generated topology.

Definition 7.2.3. *Let $\mathsf{R}_{\mathsf{pref}}$ be the function mapping a topology τ over $\Sigma^\star$ to the topology generated by the sets UV where $U \subseteq \Sigma$, and $V \in \tau$.*

We refer to Figure 7.8 for a graphical presentation of the first two iterations of the refinement function $\mathsf{R}_{\mathsf{pref}}$. For the sake of completeness, let us compute $\mathsf{lfp}_{\tau}.\,\mathsf{R}_{\mathsf{pref}}(\tau)$, which is the Alexandroff topology of the prefix ordering on words. Beware that the upcoming definition differs from what is called the "prefix topology" in the literature [34, 45, resp. Section 8 and Exercise 9.7.36].

[34]: Finkel and Goubault-Larrecq (2020), 'Forward analysis for WSTS, part I: completions'
[45]: Goubault-Larrecq (2013), *Non-Hausdorff Topology and Domain Theory*

Definition 7.2.4. *The prefix topology τ_{pref^*}, over $\Sigma^\star$ is generated by the following open sets: $U_1 \ldots U_n \Sigma^\star$, where $n \geq 0$ and $U_i \subseteq \Sigma$ for $1 \leq i \leq n$.*

Lemma 7.2.5. *The prefix topology over $\Sigma^\star$ is the least fixed point of $\mathsf{R}_{\mathsf{pref}}$.*

Proof. Consider a subbasic open set $W \in \mathsf{R}_{\mathsf{pref}}(\tau_{\mathrm{pref}^*})$. It is of the form UV with $U \subseteq \Sigma$ and $V \in \tau_{\mathrm{pref}^*}$. Hence, $UV \in \tau_{\mathrm{pref}^*}$. We have proven that, $\mathsf{R}_{\mathsf{pref}}(\tau_{\mathrm{pref}^*}) \subseteq \tau_{\mathrm{pref}^*}$.

Conversely, consider a subbasic open set $W \in \tau_{\mathrm{pref}^*}$. Either it is $\emptyset$, or $\Sigma^\star$, in which case it trivially belongs to $\mathsf{lfp}_\tau.\ \mathsf{R}_{\mathsf{pref}}(\tau)$, or it is of the form $U_1 \dots U_n \Sigma^\star$, with $U_i \subseteq \Sigma$ for $1 \leq i \leq n$, in which case one proves by induction over n that it belongs to ${\mathsf{R}_{\mathsf{pref}}}^n(\tau_{\mathsf{triv}})$. ■

Lemma 7.2.6. *The function* $\mathsf{R}_{\mathsf{pref}}$ *is a refinement function.*

Proof. It is an easy check that whenever $\tau \subseteq \tau'$, $\mathsf{R}_{\mathsf{pref}}(\tau) \subseteq \mathsf{R}_{\mathsf{pref}}(\tau')$. Now, assume that τ is Noetherian, it remains to prove that $\mathsf{R}_{\mathsf{pref}}(\tau)$ remains Noetherian. Consider a subset $E \subseteq \Sigma^\star$ and let us prove that E is compact in $\mathsf{R}_{\mathsf{pref}}(\tau)$.

For that, we consider an open cover $E \subseteq \bigcup_{i \in I} W_i$, where $W_i \in \mathsf{R}_{\mathsf{pref}}(\tau)$. Thanks to Alexander's subbase lemma, we can assume without loss of generality that W_i is a subbasic open set of $\mathsf{R}_{\mathsf{pref}}(\tau)$, that is, $W_i = U_i V_i$ with $U_i \subseteq \Sigma$ and $V_i \in \tau$.

Since $(\Sigma^\star, \tau) \times (\Sigma^\star, \tau)$ is Noetherian (see Lemma D.6.6), there exists a finite set $J \subseteq I$ such that $\bigcup_{i \in J} U_i \times V_i = \bigcup_{i \in I} U_i \times V_i$. This implies that $E \subseteq \bigcup_{i \in J} U_i V_i$, and provides a finite subcover of E. ■

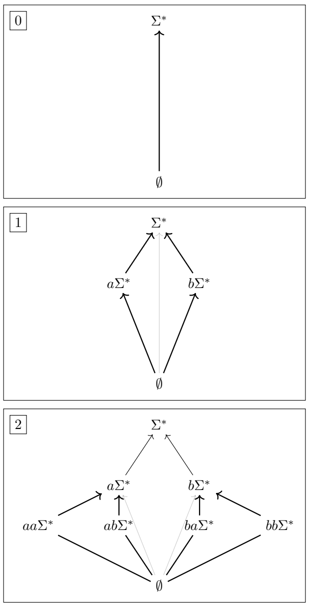


Figure 7.8.: Iterating $\mathsf{R}_{\mathsf{pref}}$ over $\Sigma^\star$. The cell i represents the lattice ${\mathsf{R}_{\mathsf{pref}}}^i(\tau_{\mathsf{triv}})$.

The sequence $\bigcup_{0 \leq i \leq k} a^i b \Sigma^*$, for $k \in \mathbb{N}$, is a strictly increasing sequence of opens. Therefore, the prefix topology is not Noetherian. The terms $a^i b \Sigma^*$ can be observed in Figure 7.8 as a diagonal of incomparable open sets.

Corollary 7.2.7. *The topology* $\mathsf{lfp}_\tau.\ \mathsf{R}_{\mathsf{pref}}(\tau)$ *is not Noetherian.*

The prefix topology is not Noetherian, even when starting from a finite alphabet. If you are reading this thesis linearly, this should be surprising, as in Chapter 6 (Logically Presented Spaces), we claimed that there was something known as the subword topology and that $(\mathsf{X}^\star, \tau^{\star,t})$ is Noetherian whenever (X, τ) is (see Table 6.2).

We already defined the subword topology in Definition 6.3.26, where we introduced the notation $[U_1, \dots, U_n]$ as a shorthand for the subset $\mathsf{X}^* U_1 \mathsf{X}^* \dots \mathsf{X}^* U_n \mathsf{X}^*$ of $\mathsf{X}^\star$. Recall that the subword topology was then defined as the coarsest topology over $\mathsf{X}^\star$ that contains the sets $[U_1, \dots, U_n]$ for $(U_i)_{1 \leq i \leq n}$ open subsets of X. Furthermore, we used the notation $\tau^{\star,t}$ to denote the subword topology. Finally, we recalled in Theorem 6.3.27 (following from [45, Theorem 9.7.33]) that the subword topology was Noetherian.

Although the subword topology might seem ad-hoc, it can be validated as a generalisation of the subword embedding because the subword topology of $\mathsf{Alex}(\leq)$ equals the Alexandroff topology of the subword ordering of $\leq$, for every quasi-order $\leq$ over Σ [45, Exercise 9.7.30]. Let us now reverse engineer a refinement function whose least fixed point is the subword topology.

Definition 7.2.8. *Let* (Σ, θ) *be a topological space. Let* $\mathsf{E}^\theta_{\mathrm{words}}$ *be* *defined as mapping a topology τ over $\Sigma^\star$ to the topology generated by the following sets:*

- $\uparrow_{\leq_w} UV$ *for* $U, V \in \tau$;
- *and* $\uparrow_{\leq_w} W$, *for* $W \in \theta$.

Lemma 7.2.9. *Let (Σ, θ) be a topological space. The subword topology over $\Sigma^\star$ is the least fixed point of $\mathsf{E}^{\theta}_{\mathsf{words}}$.*

Proof. First, we notice that the subword topology is stable under $\mathsf{E}^{\theta}_{\mathsf{words}}$. Then, we prove by induction on n shows that $[U_1, \dots, U_n]$ is open in the least fixed point of $\mathsf{E}^{\theta}_{\mathsf{words}}$. ∎

Lemma 7.2.10. *Let (Σ, θ) be a Noetherian topological space. The map $\mathsf{E}^{\theta}_{\mathsf{words}}$ is a refinement function over Σ.*

Proof. We leave the monotonicity of $\mathsf{E}^{\theta}_{\mathsf{words}}$ as an exercise and focus on the proof that $\mathsf{E}^{\theta}_{\mathsf{words}}(\tau)$ is Noetherian, whenever τ is. Thanks to Lemma D.6.6, it suffices to prove that the topology generated by the sets $\uparrow_{\leq_w} UV$ (U, V open in τ), and the topology generated by the sets $\uparrow_{\leq_w} W$ (W open in θ) are Noetherian.

Let $(\uparrow_{\leq_w} U_i V_i)_{i \in \mathbb{N}}$ be a sequence of open sets. Because Noetherian topologies are closed under products (see Lemma D.6.4), there exists k such that $\bigcup_{i \leq k} U_i \times V_i = \bigcup_{i \in \mathbb{N}} U_i \times V_i$. Hence, $\bigcup_{i \leq k} \uparrow_{\leq_w} U_i V_i = \bigcup_{i \in \mathbb{N}} \uparrow_{\leq_w} U_i V_i$

Let $\uparrow_{\leq_w} W_i$ be a sequence of open subsets. Because θ is Noetherian, there exists k such that $\bigcup_{i \leq k} W_i = \bigcup_{i \in \mathbb{N}} W_i$. As a consequence, we conclude that $\bigcup_{i \leq k} \uparrow_{\leq_w} W_i = \bigcup_{i \in \mathbb{N}} \uparrow_{\leq_w} W_i$. ∎

We have designed two refinement functions $\mathsf{R}_{\mathsf{pref}}$ and $\mathsf{E}^{\theta}_{\mathsf{words}}$ over $\Sigma^\star$. Fixing $\theta \overset{\text{def}}{=} \tau_{\mathsf{disc}}$, the least fixed point of $\mathsf{R}_{\mathsf{pref}}$ is not Noetherian while the least fixed point of $\mathsf{E}^{\theta}_{\mathsf{words}}$ is. We have depicted the result of iterating $\mathsf{E}^{\theta}_{\mathsf{words}}$ twice over the trivial topology in Figure 7.9. As opposed to $\mathsf{R}_{\mathsf{pref}}$, the "diagonal" elements are comparable for inclusion.

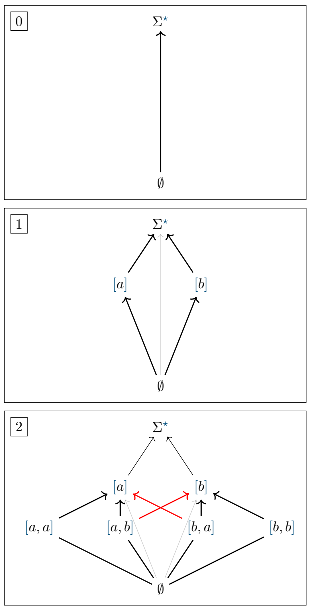


Figure 7.9.: Iterating $\mathsf{E}^{\theta}_{\mathsf{words}}$ over $\Sigma^\star$. The cell i represents the lattice ${\mathsf{E}^{\theta}_{\mathsf{words}}}^{i}(\tau_{\mathsf{triv}})$. Bold red arrows denote inclusions that were not present in the iteration of $\mathsf{R}_{\mathsf{pref}}$ (see Figure 7.8).

7.2.2. Topology Expanders

Before we give the definition of a topology expander, let us analyse the differences between the two seemingly similar refinement functions $\mathsf{R}_{\mathsf{pref}}$ and $\mathsf{E}_{\mathsf{words}}$. To that end, let us compare Figure 7.8 and Figure 7.9. The "new inclusions" appearing in the latter prevent the construction of an infinite increasing sequence of open sets. Conversely, $\mathsf{R}_{\mathsf{pref}}$ creates such infinite sequences by going from $a^i b \Sigma^\star$ to $a^{i+1} b \Sigma^\star$. Notice that $a^i b \Sigma^\star \cap a^{i+1} b \Sigma^\star = \emptyset$, while $[a^i b] \cap [a^{i+1} b] = [a^{i+1} b]$. In order to continue our analysis, we will need to talk about the way $\mathsf{R}_{\mathsf{pref}}$ and $\mathsf{E}_{\mathsf{words}}$ deal with these set intersections. For that, let us introduce the notion of subset restriction.

> **Definition 7.2.11.** *Let* (X, τ) *be a topological space and* H *be a closed subset of* X*. Define the* subset restriction $\tau{\downarrow}H$ *to be the topology generated by the opens* $U \cap H$ *where* U *ranges over* τ*.*

Exploring Subset Restrictions This definition is a way to lift the induced topology over H to the whole space X.[8] The knowledgeable reader might have noticed that we are actually trying to represent *subspaces* of X by a suitable change in its topology, and made the connection with the corresponding notions in the theory of "topology without point." For the sake of completeness, let us provide an alternative — but ill-behaved for our purposes — representation of a closed subset H inside X: $\tau{\uparrow}H \stackrel{\text{def}}{=} \{\emptyset\} \cup \{U \cup H^c : U \in \tau\}$. While it looks similar to $\tau{\downarrow}H$, both behave wildly differently. Indeed, $\tau{\uparrow}H \subseteq \tau$ while $\tau{\downarrow}H \not\subseteq \tau$ in general. Furthermore, the two topologies have different specialisation pre-orderings.

8: Recall that the induced topology over H is exactly $\{U \cap H : U \in \tau\}$ while $\tau{\downarrow}H = \{\mathsf{X}\} \cup \{U \cap H : U \in \tau\}$.

> **Lemma 7.2.12.** *Let* (X, τ) *be a topological space,* H *be a closed subset of* X*, and* $\leq$ *be the* specialisation preorder *of* τ*. Then, for all* $x, y \in \mathsf{X}$*,*
>
> - $x \leq_{\tau{\downarrow}H} y$ *if and only if* $x, y \in H \wedge x \leq_\tau y$ *or* $x \notin H$*.*
> - $x \leq_{\tau{\uparrow}H} y$ *if and only if* $x, y \in H \wedge x \leq_\tau y$ *or* $y \notin H$*.*

Proof. We only deal with the case of $\leq_{\tau{\downarrow}H}$, the other one is left as an exercise to the reader. Let $x, y \in \mathsf{X}$.

$$
\begin{aligned}
x \leq_{\tau{\downarrow}H} y &\iff \forall U \in \tau, x \in U \cap H \Rightarrow y \in U \cap H \\
&\iff \begin{cases} x \notin H \\ x \in H \wedge \forall U \in \tau, x \in U \Rightarrow y \in U \cap H \end{cases} \\
&\iff \begin{cases} x \notin H \\ x \in H \wedge y \in H \wedge \forall U \in \tau, x \in U \Rightarrow y \in U \end{cases} \\
&\iff \begin{cases} x \notin H \\ x, y \in H \wedge x \leq_\tau y \end{cases}
\end{aligned}
$$
■

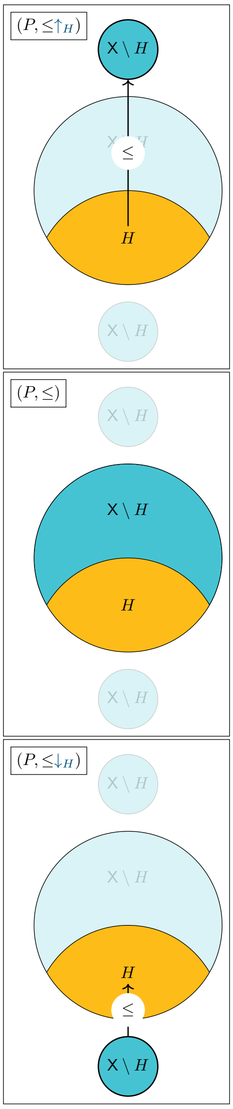


Figure 7.10.: Illustration of the action of the two possible restrictions over a quasi-order $(P, \leq)$.

Leveraging the correspondence between quasi-orders and topologies, we can transport the definition of subset restriction to quasi-orders in order to "better see" transformation that happened. For that, let us recall that a downwards closed subset H of a quasi order $(P, \leq)$ is a closed subset of $(P, \mathsf{Alex}(\leq))$. Hence, we can construct the functions $f_1 : (\leq, H) \mapsto \leq_{\mathsf{Alex}(\leq){\downarrow}H}$ and $f_2 : (\leq, H) \mapsto \leq_{\mathsf{Alex}(\leq){\uparrow}H}$. In order to simplify the notations, we will simply write these as $\leq{\downarrow}_H$ and $\leq{\uparrow}_H$ in the Figure 7.10.

As a final warning about the subset restriction, beware that even though $\tau{\downarrow}H$ encodes the topology induced by τ over H, the two spaces are not homeomorphic.

> **Example 7.2.13.** Let $\mathbb{R}$ be endowed with the usual metric to-

pology. The set $\{a\}$ is a closed set. The induced topology over $\{a\}$ is $\{\emptyset, \{a\}\}$. The subset restriction of the topology to $\{a\}$ is $\tau_a \overset{\mathsf{def}}{=} \{\emptyset, \{a\}, \mathbb{R}\}$. Clearly, $(\mathbb{R}, \tau_a)$ and $(\{a\}, \tau_{\mathsf{triv}})$ are not homeomorphic.

Refinement Functions and Subset Restrictions We now have appropriate tools to properly compare the behaviour of our two refinement functions $\mathsf{R}_{\mathsf{pref}}$ and $\mathsf{E}_{\mathrm{words}}$. We start by a simple fact about upwards closed subsets for the (scattered) word embedding.

Remark 7.2.14. Let (Σ, θ) be a topological space and $\leq$ be the specialisation preorder of θ. Let $U_1, \dots, U_n$ be upwards closed subsets of $\Sigma^\star$ for $\leq_{\mathsf{w}}$, and H be a downwards closed subset of $\Sigma^\star$ for $\leq_{\mathsf{w}}$. Then, $[U_1, \dots, U_n] \cap H = [U_1 \cap H, \dots, U_n \cap H] \cap H$.

Proof. Let $w \in [U_1, \dots, U_n] \cap H$. Then $w = \prod_{i=1}^{n} (\alpha_i u_i) \alpha_{n+1}$ where $\alpha_i \in \Sigma^\star$ and $u_i \in U_i$ for $1 \leq i \leq n$. For all $1 \leq i \leq n$, $u_i \leq_{\mathsf{w}} w$, hence $u_i \in H$ because the latter is downwards closed. As a consequence, $w \in [U_1 \cap H, \dots, U_n \cap H]$.

For the converse, notice that $[U_1 \cap H, \dots, U_n \cap H] \subseteq [U_1, \dots, U_n]$. ∎

This seemingly benign remark points out a particular dynamic of $\mathsf{E}_{\mathrm{words}}$ in general with respect to subset restriction. As we will see later on, this is the crucial property that $\mathsf{R}_{\mathsf{pref}}$ fails to exhibit.

Corollary 7.2.15. *Let θ be a topology over Σ, and τ be a topology over $\Sigma^\star$ such that $\tau \subseteq \mathsf{E}^{\theta}_{\mathrm{words}}(\tau)$. Let H be a closed subset of τ. Then $\mathsf{E}_{\mathrm{words}}(\tau){\downarrow}H \subseteq \mathsf{E}_{\mathrm{words}}(\tau{\downarrow}H){\downarrow}H$.*

Remark 7.2.16. Let $\tau \overset{\mathsf{def}}{=} \{\emptyset, a\Sigma^\star, b\Sigma^\star, \Sigma^\star\}$. Then $H \overset{\mathsf{def}}{=} a\Sigma^\star \cup \{\varepsilon\}$ is a closed subset of τ, and $\tau \subseteq \mathsf{R}_{\mathsf{pref}}(\tau)$. However, $\mathsf{R}_{\mathsf{pref}}(\tau){\downarrow}H$ evaluates to $\{\emptyset, aa\Sigma^\star, ab\Sigma^\star, a\Sigma^\star, \Sigma^\star\}$, which differs from $\mathsf{R}_{\mathsf{pref}}(\tau{\downarrow}H){\downarrow}H = \{\emptyset, aa\Sigma^\star, a\Sigma^\star, \Sigma^\star\}$.

As an intuition, we will say that Corollary 7.2.15 formalises the fact that the $\mathsf{E}_{\mathrm{words}}$ is “context-free” in the sense that two words u and v will be related (in the limit) if and only if they are when restricting our attention to a closed subset containing both u and v. We are now ready to formalise the definition of a topology expander, following the insights of Corollary 7.2.15.

Definition 7.2.17. *A* topology expander *is a refinement function* E *that satisfies the following extra property: for all Noetherian topology τ satisfying $\tau \subseteq \mathsf{E}(\tau)$, for all closed subset H in τ, $\mathsf{E}(\tau){\downarrow}H \subseteq \mathsf{E}(\tau{\downarrow}H){\downarrow}H$. We say that* E respects subsets *if that property is satisfied.*

One might notice that the definition of Definition 7.2.17 relies on the construction $\tau{\downarrow}H$ rather than on the other version $\tau{\uparrow}H$. Moreover, Corollary 7.2.15 continues to hold (for the same reasons) when substituting the former by the latter. Let us show the "alternative" definition of topology expander obtained by substituting $\tau{\uparrow}H$ to $\tau{\downarrow}H$ still allows least fixed points to be non-Noetherian.

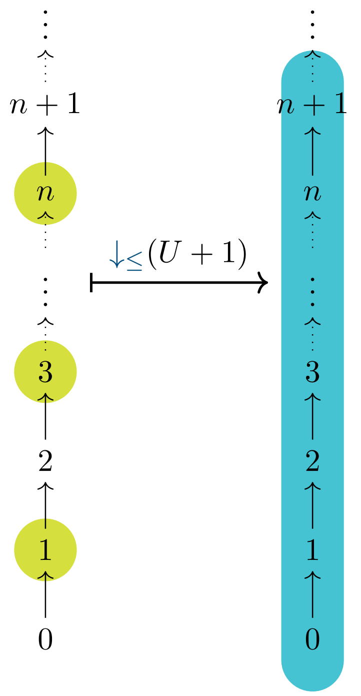


Figure 7.11.: Illustration of the computation of $\mathsf{E}_{\mathrm{nat}}$. On the left, an open subset U of a topology τ. On the right, the corresponding open subset in $\mathsf{E}_{\mathrm{nat}}(\tau)$.

Example 7.2.18. Let us fix $\mathsf{X} \overset{\mathsf{def}}{=} \mathbb{N}$ and write $\leq$ for the usual ordering on natural numbers. We define a refinement function over $\mathbb{N}$ as follows:

$$\mathsf{E}_{\mathrm{nat}}(\tau) \overset{\mathsf{def}}{=} \langle \{ 0 \} \cup \{ \downarrow_{\leq}(U+1) \colon U \in \tau \} \rangle_{\mathsf{topo}} \quad .$$

Then, $\mathsf{E}_{\mathrm{nat}}$ satisfies $\mathsf{E}_{\mathrm{nat}}(\tau){\uparrow}H \subseteq \mathsf{E}_{\mathrm{nat}}(\tau{\uparrow}H){\uparrow}H$ for every topology τ satisfying $\tau \subseteq \mathsf{E}_{\mathrm{nat}}(\tau)$ and every closed subset H in τ. However, $\mathsf{lfp}_{\tau}.\mathsf{E}_{\mathrm{nat}}(\tau)$ is not Noetherian, and $\mathsf{E}_{\mathrm{nat}}$ is not a topology expander.

Proof. We first have to check that $\mathsf{E}_{\mathrm{nat}}$ is a refinement function. While monotonicity is clear from the definition, let us briefly demonstrate that $\mathsf{E}_{\mathrm{nat}}(\tau)$ remains Noetherian whenever τ is. Assume for contradiction that $(V_i)_{i\in\mathbb{N}}$ is an infinite sequence of subbasic open subsets of $\mathsf{E}_{\mathrm{nat}}(\tau)$, which is enough thanks to Alexander's subbase lemma. Without loss of generality, one can assume that $V_i \neq \{ 0 \}$ for all $i \in \mathbb{N}$. Therefore, for all $i \in \mathbb{N}$, there exists $U_i \in \tau$ such that $V_i = \downarrow_{\leq}(U_i + 1)$. Because τ is Noetherian, there exists $n_0 \in \mathbb{N}$ such that $\bigcup_{i\in\mathbb{N}} U_i = \bigcup_{i\leq n_0} U_i$. Hence, $\bigcup_{i\in\mathbb{N}} V_i = \bigcup_{i\leq n_0} V_i$. We have proven that $\mathsf{E}_{\mathrm{nat}}(\tau)$ is Noetherian.

An easy induction demonstrates that $\mathsf{lfp}_{\tau}.\mathsf{E}_{\mathrm{nat}}$ is exactly $\{ \emptyset, \mathbb{N} \} \cup \{ \downarrow_{\leq}m \colon m \in \mathbb{N} \}$. The latter is not Noetherian, as witnessed by the sequence $(\downarrow_{\leq}m)_{m\in\mathbb{N}}$ which is infinite and strictly increasing.

To prove that $\mathsf{E}_{\mathrm{nat}}$ is not a topology expander, it suffices to consider $\tau_0 \overset{\mathsf{def}}{=} \{ \emptyset, \{0\}, \mathbb{N} \}$ and $H \overset{\mathsf{def}}{=} \mathbb{N} \setminus \{0\} = \uparrow_{<}0$. Let us now compute the different topologies appearing in the definition of a topology expander given τ_0 and H:

$$\begin{aligned}
\mathsf{E}_{\mathrm{nat}}(\tau_0) &= \{ \emptyset, \{0\}, \{0,1\}, \mathbb{N} \} && = \boxed{\mathrm{B}} \\
\mathsf{E}_{\mathrm{nat}}(\tau_0){\downarrow}H &= \{ \emptyset, \{1\}, H, \mathbb{N} \} && = \boxed{\mathrm{C}} \\
\tau_0{\downarrow}H &= \{ \emptyset, H, \mathbb{N} \} && = \boxed{\mathrm{A}} \\
\mathsf{E}_{\mathrm{nat}}(\tau_0{\downarrow}H) &= \{ \emptyset, \{0\}, \mathbb{N} \} && = \boxed{\mathrm{D}} \\
\mathsf{E}_{\mathrm{nat}}(\tau_0{\downarrow}H){\downarrow}H &= \{ \emptyset, H, \mathbb{N} \} && = \boxed{\mathrm{A}}
\end{aligned}$$

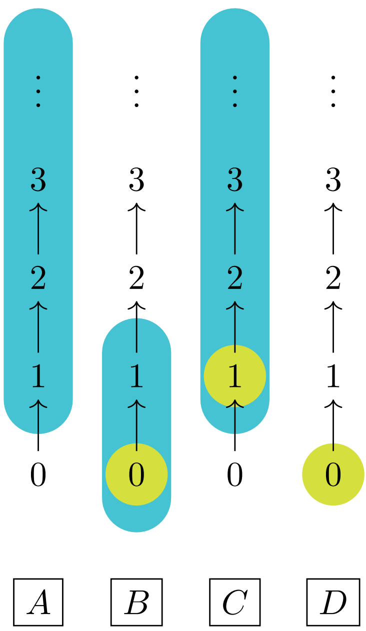


Figure 7.12.: Graphical representation of the different topologies obtained when combining the closed subset $\{ 0 \}$, the topology $\{ \emptyset, \{0\}, \mathbb{N} \}$, and the refinement function $\mathsf{E}_{\mathrm{nat}}$. We do not represent the sets $\mathbb{N}$ and $\emptyset$ that are always present.

Hence, we have proven that $\mathsf{E}_{\mathrm{nat}}(\tau_0){\downarrow}H$ is not a subset of $\mathsf{E}_{\mathrm{nat}}(\tau_0{\downarrow}H){\downarrow}H$ where H is a closed subset of a Noetherian topology τ_0 satisfying $\tau_0 \subseteq \mathsf{E}_{\mathrm{nat}}(\tau_0)$. ■

Because we will often consider spaces constructed via subbasic open subsets, let us formalise the fact that the property of being a topology expander can be checked over given subbases. If it is quite easy to see that one only needs to consider subbasic open subsets of the output topology, this we strengthen this remark by proving that one can also restrict the property to subbasic closed subsets of the input topology, thanks to the assumption that the latter is Noetherian.

Lemma 7.2.19. *Let* E *be a refinement function. Assume that for every Noetherian topology* τ *there exists a subbasis* B *of* τ *and a subbasis* B' *of* $\mathsf{E}(\tau)$ *such that for every subbasic closed subset* $H \in B$ *and every subbasic open subset* $U \in B'$, $H \cap U \in \mathsf{E}(\tau{\downarrow}H){\downarrow}H$.

Then, E *is a topology expander.*

Proof. Given a closed subset H of τ, the inclusion $\mathsf{E}(\tau){\downarrow}H \subseteq \mathsf{E}(\tau{\downarrow}H){\downarrow}H$ amounts to checking that for every $U \in B'$, $U \cap H \in \mathsf{E}(\tau{\downarrow}H){\downarrow}H$.

Let us fix U in the subbasis B'. The family of closed subsets H such that $U \cap H$ belongs to $\mathsf{E}(\tau{\downarrow}H){\downarrow}H$, contains B, is closed under finite intersections, and is closed under finite unions. Let us prove that this set is also closed under arbitrary intersections, which concludes the proof.

Let $\bigcap_{i\in I} H_i$ be an arbitrary intersection of closed subsets, such that $U \cap H_i$ belongs to $\mathsf{E}(\tau{\downarrow}H){\downarrow}H$. Because τ is Noetherian, there exists a finite subset $J \subseteq_{\mathsf{fin}} I$ such that $U \cap \bigcap_{i\in I} H_i = U \cap \bigcap_{i\in J} H_j$. Hence, we can leverage the closure under finite intersections. ∎

7.2.3. The (Generic) Topological Minimal Bad Sequence Argument

This section is entirely devoted to one of the main contribution of this thesis: a "master" fixed point theorem. Informally, the theorem states that topology expanders are refinement functions whose least fixed points are Noetherian topologies. This is the focus of [72] and the topic of a blog post [46].

[72]: Lopez (2023), 'Fixed Points and Noetherian Topologies'

[46]: Goubault-Larrecq (2022), *Aliaume Lopez' master theorem of Noetherian spaces*

The proof of the theorem relies on a transfinite induction, which necessitate to properly define what we mean by "iterating E α-times" when α is a general ordinal. This is the purpose of the next definition.

Definition 7.2.20. Let (X, τ) be a topological space, and E be a topology expander. The *limit topology* is defined by induction via:

$$\mathsf{E}^{\alpha}(\tau) = \begin{cases} \tau & \text{when: } \alpha = 0 \\ \mathsf{E}(\mathsf{E}^{\beta}(\tau)) & \text{when: } \alpha = \beta + 1 \\ \langle \bigcup_{\beta<\alpha} \mathsf{E}^{\beta}(\tau) \rangle_{\mathsf{topo}} & \text{when: } \alpha \text{ is a limit ordinal} \end{cases}$$

In the Definition 7.2.20, the sequence $\left(\mathsf{E}^{\beta}(\tau)\right)_{\beta<\alpha}$ is may increase. To obtain a non-decreasing sequence, it suffices to assume that $\tau \subseteq \mathsf{E}(\tau)$.

As mentioned in the introduction of Subsection 7.2.1, every refinement function has a least fixed point thanks to the Knaster-Tarski Fixed Point Theorem. It is folklore that this statement can be reinforced as follows: for every refinement function R, there exists an ordinal α such that $\mathsf{R}^{\alpha}(\tau_{\mathsf{triv}})$ is the least fixed point of R. In order to study the least fixed point of topology expanders, the only interesting case is $\mathsf{E}^{\alpha}(\tau_{\mathsf{triv}})$ when α is a limit ordinal, as the other topologies are then trivially Noetherian.

As a teaser, let α be a limit ordinal and assume that $\mathsf{E}^{\beta}(\tau_{\mathsf{triv}})$ is Noetherian for all $\beta < \alpha$. A typical infinite increasing sequence $(U_i)_{i \in \mathbb{N}}$ of subbasic open subsets in $\mathsf{E}^{\alpha}(\tau_{\mathsf{triv}})$ behaves as depicted in Figure 7.13. Indeed, the sequence must "step-up" infinitely often: there are finitely many indices $i \in \mathbb{N}$ such that $U_i \in \mathsf{E}^{\beta}(\tau_{\mathsf{triv}})$ given a $\beta < \alpha$, because the latter was supposed to be Noetherian. Intuitively, one can also assume without loss of generality that the sequence does not "step-down." The core of the proof is actually stating that one cannot "step-up" infinitely often when E is a topology expander. Incidentally, we will show that $\mathsf{E}^{\alpha}(\tau_{\mathsf{triv}})$ is simply the union $\bigcup_{\beta<\alpha} \mathsf{E}^{\beta}(\tau_{\mathsf{triv}})$.

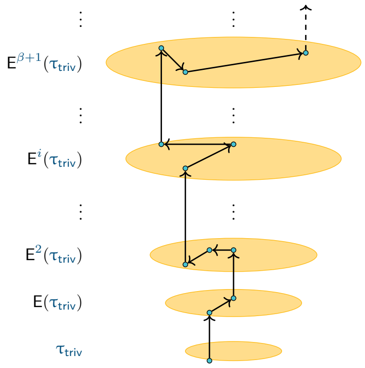


Figure 7.13.: Increasing sequence of subbasic open subsets in in $\mathsf{E}^{\alpha}(\tau_{\mathsf{triv}})$. The yellow ellipses represent "steps." The sequence stays in some steps for a finite number of iterations, and then jumps to a new step. In the drawing, arrows are always going up or staying in a given step, but they might be pointing down due to deformations induced by a two-dimensional representation of a three-dimensional object.

Noetherian Techniques. In order to study increasing sequences in limit topologies, we recall the efficient tools that were developed by Goubault-Larrecq in his proofs of the topological counterparts to Higman's Lemma and Kruskal's Tree Theorem. One proof technique that we used quite often was to rely on Alexander's subbase lemma and the characterisation of Noetherian spaces as those that do not have infinite increasing sequences of open subsets (see Lemma D.6.5). Let us wrap these two results into an easier to understand "topological" counterpart of the notion of bad sequence, a term that appears in the statements of [45, Lemma 9.7.15 and Lemma 9.7.31] and that we extract here.

To get some intuition, notice that in a quasi-ordered set $(\mathsf{X}, \leq)$, a sequence $(x_i)_{i\in\mathbb{N}}$ is a bad sequence if and only if $(\uparrow_{\leq} x_i)_{i\in\mathbb{N}}$ is a topological bad sequence.

Definition 7.2.21 [45, Lemma 9.7.15 and Lemma 9.7.31]**.** Let (X, τ) be a topological space. A sequence $(U_i)_{i\in\mathbb{N}}$ of open subsets is *good* if there exists $i \in \mathbb{N}$ such that $U_i \subseteq \bigcup_{j<i} U_j$.

A sequence that is not *good* is called *bad*.

Lemma 7.2.22 [45, Lemma 9.7.15]**.** A topological space (X, τ) is Noetherian if and only if it has no infinite topological bad sequence.

Proof Sketch. The union of the open subsets in an infinite topological bad sequence defines a non compact subset. Conversely, given a non-compact subset, one can extract an infinite topological bad sequence from a suitable open cover. In particular, this last step requires extracting a *countable* sequence from the open cover. Details of the proof are found in [45, Lemma 9.7.15]. ∎

One of the key ingredients in our statement of Theorems 6.3.24 and 6.3.25 was the "simpler" preorder (respectively the suffix ordering $\leq_{\mathsf{suf}}$ or the structural subtree ordering $\leq_{\mathsf{t\text{-}suf}}$) that allowed us to consider a *minimal* bad sequence. First, let us formally define the *lexicographic ordering* on sequences of (subbasic) open subsets

Definition 7.2.23. Let $(P, \leq)$ be a partial order. Then P^{ω} can be quasi-ordered via the following co-inductive definition: $(x_i)_{i\in\mathbb{N}} \leq_{\mathsf{lex}} (y_i)_{i\in\mathbb{N}}$ if and only if $x_0 < y_0$ or $x_0 = y_0$ and $(x_i)_{i\geq 1} \leq_{\mathsf{lex}} (y_i)_{i\geq 1}$.

Equivalently, $(x_i)_{i\in\mathbb{N}} \leq_{\mathsf{lex}} (y_i)_{i\in\mathbb{N}}$ if and only if the two sequences are equal, or if there exists $i \in \mathbb{N}$ such that $\forall n < i, x_n = y_n$, and $x_i < y_i$.

Beware that $(P^{\omega}, \leq_{\mathsf{lex}})$ has no reason to be well-founded whenever $(P, \leq)$ is: for instance, the sequence $(0^i 1^{\omega})_{i\in\mathbb{N}}$ is a strictly decreasing sequence of elements in $\mathbb{N}^\omega$ for $\leq_{\mathsf{lex}}$. Surprisingly, we can work around this issue and build minimal representatives for topological bad sequences.

Lemma 7.2.24 [45, Lemma 9.7.31]**.** Let (X, τ) be a topological space that is not Noetherian, $\mathcal{B}$ be a subbasis of τ and $\leq$ be a well-founded ordering on $\mathcal{B}$. The set of topological bad sequences in $(\mathcal{B}^\omega, \leq_{\mathsf{lex}})$ has a minimal element.

Proof Sketch. This element is constructed by considering a $\leq$ minimal open subset starting a bad sequence, a $\leq$-minimal open subset that continues this bad sequence, etc. The process yields a bad sequence that is minimal by construction. ∎

We call the sequence provided by Lemma 7.2.24 a *minimal topological bad sequence*.

Taming the Minimal Bad Sequences. Let us now explain how we can leverage the inductive definition of $\mathsf{E}^\alpha(\tau_{\mathsf{triv}})$ to actually define a well-founded order. The first step is to notice that given a topology expander E and a limit ordinal α, $\bigcup_{\beta<\alpha} \mathsf{E}^\beta(\tau_{\mathsf{triv}})$ is a subbasis of $\mathsf{E}^\alpha(\tau_{\mathsf{triv}})$. Let us refine this analysis by studying the "step" (see Figure 7.13) at which a given open subset appears in the union.[9]

9: If the sequence $\left(\mathsf{E}^\beta(\tau)\right)_{\beta<\alpha}$ is non-decreasing, and $U \in \mathsf{E}^{\beta_0}(\tau)$, for some $\beta_0 < \alpha$, then $U \in \mathsf{E}^\beta(\tau)$, for all $\beta_0 \leq \beta \leq \alpha$.

Definition 7.2.25. *Let E be a topology expander, τ be a topology, and α be an ordinal. We define for $U \in \mathsf{E}^\alpha(\tau)$ the ordinal $\mathrm{depth}(U)$ as the smallest $\beta \leq \alpha$ such that $U \in \mathsf{E}^\beta(\tau)$.*

We define a total partial order over open subsets of $(\mathsf{X}, \mathsf{E}^\alpha(\tau))$ by defining $U \lhd V$ if and only if $\mathrm{depth}(U) < \mathrm{depth}(V)$. Thanks to Lemma 7.2.24, we are guaranteed that it suffices to consider sequences of subbasic open subsets that are lexicographically minimal for $\trianglelefteq_{\mathsf{lex}}$ to prove/disprove that $\mathsf{E}^\alpha(\tau)$ is Noetherian for a given limit ordinal α. Let us continue our analysis and further restrict the shape of sequences we have to analyse. In particular, we are first going to prove that not every depth is necessary.[10]

10: The proof of Lemma 7.2.26 actually shows that all topological minimal bad sequences satisfy the property, and not only that there exists one. However, we only need the weaker statement: we are trying to build *one* topological minimal bad sequence.

Lemma 7.2.26. *Let E be a topology expander, τ be a topology, and α be an ordinal. Assume that for all $\beta < \alpha$, $\mathsf{E}^\beta(\tau)$ is Noetherian, and $\mathsf{E}^\alpha(\tau)$ is not Noetherian.*

There exists a topological minimal bad sequence $(U_i)_{i\in\mathbb{N}}$ for $\trianglelefteq_{\mathsf{lex}}$ such that, for all $i \in \mathbb{N}$, $\mathrm{depth}(U_i)$ is not *a limit ordinal.*

Proof. We know from Lemma 7.2.24 that there exists a topological minimal bad sequence $(U_i)_{i\in\mathbb{N}}$ for $\trianglelefteq_{\mathsf{lex}}$ such that $\mathrm{depth}(U_i) < \alpha$ for all $i \in \mathbb{N}$. Assume for contradiction that there exists $i \in \mathbb{N}$ such that $\mathrm{depth}(U_i)$ is a limit ordinal $d_i < \alpha$.

Let $S \stackrel{\text{def}}{=} \bigcup_{\beta<d_i} \mathsf{E}^\beta(\tau)$. Then, $U_i \in \mathsf{E}^{d_i}(\tau) \stackrel{\text{def}}{=} \langle S \rangle_{\mathsf{topo}}$. Because S is closed under finite intersections, U_i is a union of open subsets of S.

Since $\mathsf{E}^{d_i}(\tau)$ is Noetherian, one can define U_i as a finite union of open subsets of S, hence a finite union of open subsets of depth less than d_i. As a consequence, $\mathrm{depth}(U_i) < d_i$, which is absurd. ∎

We have proven that one can always assume that the depths in a minimal bad sequence are 0 or successor ordinals strictly below α. Let us get back to the picture in Figure 7.13: we shall now prove that the depth can be assumed to be non-decreasing, and even to increase infinity often.

Lemma 7.2.27. *Let E be a topology expander, τ be a topology, and α be an ordinal. Assume that for all $\beta < \alpha$, $\mathsf{E}^{\beta}(\tau)$ is Noetherian, and that $\mathsf{E}^{\alpha}(\tau)$ is not Noetherian.*

There exists a topological minimal bad sequence $(U_i)_{i\in\mathbb{N}}$ for $\trianglelefteq_{\mathsf{lex}}$ such that, for every $i \in \mathbb{N}$, $\mathrm{depth}(U_i)$ is not *a limit ordinal. Moreover, the sequence $(\mathrm{depth}(U_i))_{i\in\mathbb{N}}$ can be assumed non-decreasing and not stationary.*

Proof. Thanks to Lemma 7.2.26, we have a topological minimal bad sequence $(U_i)_{i\in\mathbb{N}}$ such that $\mathrm{depth}(U_i)$ is not a limit ordinal for every $i \in \mathbb{N}$.

Assume for contradiction that there exist $i < j$ such that $U_j \triangleleft U_i$. Notice that the sequence obtained by deleting $U_i, U_{i+1}, \ldots, U_{j-1}$ from the original sequence yields an infinite topological bad sequence. Moreover, it is strictly smaller than the original one for $\trianglelefteq_{\mathsf{lex}}$ by construction, which contradicts the latter's minimality.

Assume for contradiction that there exists an ordinal $\gamma < \alpha$ such that $T \stackrel{\text{def}}{=} \{ i \in \mathbb{N} \colon \mathrm{depth}(U_i) = \gamma \}$ is infinite. Then the subsequence $(U_i)_{i\in T}$ is an infinite topological bad sequence of open subsets of $\mathsf{E}^{\gamma}(\tau)$, and this contradicts the hypothesis that the latter is Noetherian. ■

The Structural Simplification We are now in a position where the shape of a minimal topological bad sequence is well understood. Notice that up to this point, we never actually used the property that singularise topology expanders among refinement functions, and this is exactly the part where subset restrictions and intersections will show up.[11] Before considering these, let us define the equivalent of the set S in the joint proof of Theorems 6.3.24 and 6.3.25, a construction that was gathering elements "structurally simpler" than those of the original sequence.

11: At the appropriate time, a note in the proof will highlight why $\tau{\downarrow}H$ crucially uses intersections and not unions, which was discussed before.

Definition 7.2.28. *Let E be a topology expander, τ be a topology, and α be an ordinal. Given $U \in \mathsf{E}^{\alpha}(\tau)$, the topology $\mathsf{Down}^{\alpha}(U)$ is defined as $\langle \{ V \in \mathsf{E}^{\alpha}(\tau) \colon V \triangleleft U \} \rangle_{\mathsf{topo}}$.*

Let $(U_i)_{i\in\mathbb{N}}$ be a topological minimal bad sequence in $(\mathsf{X}, \mathsf{E}^{\alpha}(\tau))$, where E is a topology expander and α is a limit ordinal. Because the sequence is bad, for every $i \in \mathbb{N}$, $U_i \not\subseteq \bigcup_{j<i} U_j \stackrel{\text{def}}{=} V_i$. Letting $H_i \stackrel{\text{def}}{=} \mathsf{X} \setminus V_i$ for every $i \in \mathbb{N}$, this equation can be rewritten as $U_i \cap H_i \neq \emptyset$. Notice that for every $i \in \mathbb{N}$, H_i is not only a closed subset of E^{α} but also a closed subset of $\mathsf{Down}^{d_i}(U_i)$, where $d_i \stackrel{\text{def}}{=} \mathrm{depth}(U_i)$. This means the following definition is sound.

Definition 7.2.29. *Let E be a topology expander, τ be a topology, and α be an ordinal. Let $(U_i)_{i\in\mathbb{N}}$ be a topological minimal bad sequence for $\trianglelefteq_{\mathsf{lex}}$. Let $d_i \stackrel{\text{def}}{=} \mathrm{depth}(U_i)$ for all $i \in \mathbb{N}$. We define the minimal topology $\mathsf{E}^{\alpha}_{\mathsf{min}}((U_i)_{i\in\mathbb{N}}, \tau)$ as the one generated by*

$$\bigcup_{i\in\mathbb{N}} \mathsf{Down}^{d_i}(U_i){\downarrow}H_i \quad ,$$

where $H_i \stackrel{\text{def}}{=} \mathsf{X} \setminus (\bigcup_{j<i} U_j)$.

In order to make the notation lighter, and because we only ever deal with one minimal bad sequence at once, we will simply write $\mathsf{E}^{\alpha}_{\mathsf{min}}$ to denote the full expression $\mathsf{E}^{\alpha}_{\mathsf{min}}((U_i)_{i\in\mathbb{N}}, \tau)$, when τ and $(U_i)_{i\in\mathbb{N}}$ are clear from the context. The main property of this minimal topology is that it is Noetherian, in a similar way that S was well-quasi-ordered in the proof of Theorem 6.3.24.

Lemma 7.2.30. *Let E be a topology expander, τ be a topology, and α be an ordinal. Let $(U_i)_{i\in\mathbb{N}}$ be a topological minimal bad sequence for $\trianglelefteq_{\mathsf{lex}}$. Then, the minimal topology $\mathsf{E}^{\alpha}_{\mathsf{min}}$ is Noetherian.*

Proof. Assume for contradiction that $\mathsf{E}^{\alpha}_{\mathsf{min}}$ is not Noetherian. Let us define $V_i \stackrel{\text{def}}{=} \bigcup_{j<i} U_j$, H_i as $\mathsf{X} \setminus V_i$, and $d_i \stackrel{\text{def}}{=} \mathrm{depth}(U_i)$, for all $i \in \mathbb{N}$.

Thanks to Lemma 7.2.22, there must be some topological bad sequence $(W_i)_{i\in\mathbb{N}}$ for $\mathsf{E}^{\alpha}_{\mathsf{min}}$ where, for every $i \in \mathbb{N}$, there exists $\rho(i) \in \mathbb{N}$ such that $W_i \in \mathsf{Down}^{d_{\rho(i)}}(U_{\rho(i)}){\downarrow}H_{\rho(i)}$. This amounts to the existence of an open subset $T_{\rho(i)}$ in $\mathsf{Down}^{d_{\rho(i)}}(U_{\rho(i)})$, such that $T_{\rho(i)} \triangleleft U_{\rho(i)}$, and $W_i = T_{\rho(i)} \setminus V_{\rho(i)}$.

Let us prove that we can assume ρ to be non-decreasing. For that, we will extract a subsequence of $(W_i)_{i\in\mathbb{N}}$ as follows: take $\iota(0) \stackrel{\text{def}}{=} 0$, and then $\iota(n+1) \stackrel{\text{def}}{=} \min\{ i \in \mathbb{N} \colon i > \iota(n) \wedge \rho(i) > \rho(\iota(n)) \}$. This construction is sound as soon as we can argue that for every $k \in \mathbb{N}$, $\rho^{-1}(k) \stackrel{\text{def}}{=} \{ i \in \mathbb{N} \colon \rho(i) = k \}$ is finite. Let $k \in \mathbb{N}$, then the sequence $(W_i)_{i\in\rho^{-1}(k)}$ is a sequence of open subsets of $\mathsf{Down}^{d_k}(U_k){\downarrow}H_k$. Because $d_k < \alpha$, $\mathsf{Down}^{d_k}(U_k) \subseteq \mathsf{E}^{d_k}(\tau)$ is Noetherian. As an easy consequence, $\mathsf{Down}^{d_k}(U_k){\downarrow}H_k$ is Noetherian too. Therefore, assuming for contradiction that $\rho^{-1}(k)$ is infinite provides us with an $i \in \rho^{-1}(k)$, such that $W_i \subseteq \bigcup_{\rho(j)=k\wedge j<i} W_j$, and *a fortiori* $W_i \subseteq \bigcup_{j<i} W_j$. Because $(W_i)_{i\in\mathbb{N}}$ is a topological bad sequence, this is absurd, and we have proven that $\rho^{-1}(k)$ is in fact finite.

Let us build the sequence $(Y_i)_{i\in\mathbb{N}}$ defined by $Y_i \stackrel{\text{def}}{=} U_i$ if $i < \rho(0)$ and $Y_i \stackrel{\text{def}}{=} T_{\rho(i)}$ otherwise. Because this is a sequence of open sets in $\mathsf{E}^{\alpha}(\tau)$ that is lexicographically smaller than $(U_i)_{i\in\mathbb{N}}$, it must be a good sequence. Therefore, there exists $p \in \mathbb{N}$ such that $Y_p \subseteq \bigcup_{j<p} Y_j$. Let us now analyse what it implies on the original sequences depending on the ordering between p and $\rho(0)$.

If $p < \rho(0)$**:** then $U_p \subseteq \bigcup_{j<p} U_j$ which is impossible because the sequence $(U_i)_{i\in\mathbb{N}}$ is a topological bad sequence.

If $p \geq \rho(0)$**:** let us write $Y_p = T_{\rho(p)} \subseteq \bigcup_{j<\rho(0)} U_j \cup \bigcup_{j<p} T_{\rho(j)}$. By taking the intersection with $H_{\rho(p)}$,[12] we obtain $W_p \subseteq \bigcup_{j<p} W_j$, hence the sequence $(W_i)_{i\in\mathbb{N}}$ is a topological good sequence which is absurd. ■

12: This is the crucial way intersections are used, and cannot be replaced by unions. See the other notes about this fact.

The Lifting Argument Now that we have extracted a Noetherian topology out of a suitable topological minimal bad sequence, let us prove that one can "lift" this construction to obtain a contradiction. This is the part where the key property of topology expanders plays a role by ensuring that this lifting is possible. In order to understand what the informal notion of lifting means, let us highlight the following fact.

Fact 7.2.31. Let α be an ordinal, E be a topology expander and $U \in \mathsf{E}^\alpha(\tau)$. If $d \stackrel{\text{def}}{=} \text{depth}(U)$ is a successor ordinal, then $U \in \mathsf{E}(\mathsf{Down}^d(U))$.

The main idea of the following proof is that a topological minimal bad sequence $(U_i)_{i\in\mathbb{N}}$ of $\mathsf{E}^\alpha(\tau)$ in fact belongs to $\mathsf{E}(\mathsf{E}^\alpha_{\mathsf{min}})$. This immediately yields a contradiction, as E preserves Noetherian topologies and $\mathsf{E}^\alpha_{\mathsf{min}}$ has just been proven to be Noetherian in Lemma 7.2.30. However, this rough idea must be refined, and we will not prove that the sequence belongs to $\mathsf{E}(\mathsf{E}^\alpha_{\mathsf{min}})$ (which is false in general).

Lemma 7.2.32. *Let α be an ordinal, X be a topology, and E be a topology expander. If $\mathsf{E}^\beta(\tau)$ is Noetherian for all $\beta < \alpha$, and $\tau \subseteq \mathsf{E}(\tau)$, then $\mathsf{E}^\alpha(\tau)$ is Noetherian.*

Proof. If α is a successor ordinal, then $\alpha = \beta+1$ and $\mathsf{E}^\alpha(\tau) = \mathsf{E}(\mathsf{E}^\beta(\tau))$. Because E respects Noetherian topologies, we immediately conclude that $\mathsf{E}^\alpha(\tau)$ is Noetherian. We are therefore only interested in the case where α is a limit ordinal.

Assume for contradiction that $\mathsf{E}^\alpha(\tau)$ is not Noetherian, we leverage Lemma 7.2.27 to obtain a topological minimal bad sequence $(U_i)_{i\in\mathbb{N}}$, such that $\text{depth}(U_i) < \alpha_i$ is either 0 or a successor ordinal written β_i+1. Moreover, we know that $\text{depth}(U_i)$ is non-decreasing and increases infinitely often.

We can therefore construct a strictly increasing map $\rho\colon \mathbb{N} \to \mathbb{N}$ such that $\text{depth}(U_{\rho(i)})$ is a strictly increasing map from $\mathbb{N}$ to non-zero successor ordinals below α.[13]

Let us consider some $p = \rho(i)$ for some $i \in \mathbb{N}$.[14] Let us write $V_p \stackrel{\text{def}}{=} \bigcup_{j<p} U_j$, and $H_p \stackrel{\text{def}}{=} X \setminus V_p$. As E is a topology expander, we derive the following inclusions:[15]

$$\begin{aligned}\mathsf{E}(\mathsf{Down}^{d_p}(U_p))\!\downarrow\! H_p &\subseteq \mathsf{E}(\mathsf{Down}^{d_p}(U_p)\!\downarrow\! H_p)\!\downarrow\! H_p \\ &\subseteq \mathsf{E}(\mathsf{E}^\alpha_{\mathsf{min}})\!\downarrow\! H_p \quad .\end{aligned}$$

13: In particular, the map ρ never selects open subsets U of depth zero. This is necessary for the following computations to make sense.

14: In particular, $d_p = \text{depth}(U_p) = \text{depth}(U_{\rho(i)}) = d_{\rho(i)}$ is a non-zero successor ordinal!

15: This is the key part where we require that E is a topology expander!

Recall that $U_p \in \mathsf{E}(\mathsf{Down}^{d_p}(U_p))$ thanks to Fact 7.2.31, and because $\text{depth}(U_p)$ is a non-zero successor ordinal. Hence, the intersection $U_p \cap$

H_p belongs to $\mathsf{E}(\mathsf{Down}^{d_p}(U_p)){\downarrow}H_p$, and therefore $U_p \cap H_p \in\subseteq \mathsf{E}(\mathsf{E}^{\alpha}_{\mathsf{min}}){\downarrow}H_p$. As a consequence of the above inclusion of topologies, there exists a $W_p \in \mathsf{E}(\mathsf{E}^{\alpha}_{\mathsf{min}})$ such that $U_p \cap H_p = U_p \setminus V_p = W_p \setminus V_p = W_p \cap H_p$. We have constructed a sequence $\left(W_{\rho(i)}\right)_{i\in\mathbb{N}}$ of open subsets of $\mathsf{E}(\mathsf{E}^{\alpha}_{\mathsf{min}})$ such that for all $i \in \mathbb{N}$, $U_{\rho(i)} \setminus V_{\rho(i)} = W_{\rho(i)} \setminus V_{\rho(i)}$.

Thanks to Lemma 7.2.30, and preservation of Noetherian topologies through topology expanders, $\mathsf{E}(\mathsf{E}^{\alpha}_{\mathsf{min}})$ is a Noetherian topology. This provides an $i \in \mathbb{N}$ such that $W_{\rho(i)} \subseteq \bigcup_{\rho(j)<\rho(i)} W_{\rho(j)}$. In particular,

$$\begin{aligned}
U_{\rho(i)} \setminus V_{\rho(i)} &= W_{\rho(i)} \setminus V_{\rho(i)} \\
&\subseteq \left(\bigcup_{\rho(j)<\rho(i)} W_{\rho(j)} \right) \setminus V_{\rho(i)} \\
&\subseteq \bigcup_{\rho(j)<\rho(i)} W_{\rho(j)} \setminus V_{\rho(j)} \\
&= \bigcup_{\rho(j)<\rho(i)} U_{\rho(j)} \setminus V_{\rho(j)} \\
&\subseteq \bigcup_{j<\rho(i)} U_j \\
&= V_{\rho(i)} \quad .
\end{aligned}$$

This proves that $U_{\rho(i)} \subseteq V_{\rho(i)}$, i.e., that $U_{\rho(i)} \subseteq \bigcup_{j<\rho(i)} U_j$. Finally, this contradicts the fact that $(U_i)_{i\in\mathbb{N}}$ is a topological bad sequence. ■

We have effectively proven that being well-behaved with respect to closed subspaces is enough to consider least fixed points of refinement functions. As an immediate consequence, we obtain our master theorem.

Theorem 7.2.33 (Main Result). *Let* X *be a set,* E *be a topology expander, and* τ *be a topology over* X *such that* $\tau \subseteq \mathsf{E}(\tau)$*. The least fixed point of* E *above* τ *is a Noetherian topology over* X*.*

7.3. Variations Around the Main Theorem

We discuss applications of Theorem 7.2.33 in a separate chapter dedicated to inductive constructions, namely Chapter 8 (Inductive Constructions). For the moment, we are interested in the limits, variations, and consequences of the theorem from a theoretical perspective.

7.3.1. When is a Refinement Function a Topology Expander?

Recall that the collection of topologies over a given set X is a set, which is moreover a complete lattice for inclusion. Let us decompose the three properties that define a topology expander.[16] This allows us to prove that the topology expanders “almost” form a complete lattice, which is enough to prove that any refinement function has a best approximation

16: Monotonicity, preserving Noetherian topologies, and the compatibility with subset restriction, as obtained by combining Definition 7.2.1 and Definition 7.2.17.

in terms of topology expander. Before that, let us provide some generic examples and non-examples of topology expanders.

Example 7.3.1. Let (X, θ) be a topological space. The constant map $\mathrm{const}_\theta \colon \tau \mapsto \theta$ is a topology expander if and only if θ is Noetherian.

Proof. The constant map is always monotone for the pointwise ordering. Let τ be a Noetherian topology satisfying $\tau \subseteq \theta$, and $H \in \mathcal{H}(\tau)$. Then, $\mathrm{const}_\theta(\tau){\downarrow}H = \theta{\downarrow}H = \mathrm{const}_\theta(\tau{\downarrow}H){\downarrow}H$. As a consequence, the only property left to check for const_θ to be a topology expander is that $\mathrm{const}_\theta(\tau)$ is Noetherian whenever τ is. This holds if and only if θ is itself Noetherian, because there exists at least one Noetherian topology over X (τ_{triv}). ■

In particular, this proves that topology expanders are enough to recover any Noetherian topology over a set X.

Remark 7.3.2. Let (X, θ) be a topological space. Then, θ is Noetherian if and only if it is the least fixed point of a topology expander.

Proof. It suffices to consider $\theta = \mathsf{lfp}_\tau.\,\mathrm{const}_\theta(\tau)$ which is a topology expander thanks to Example 7.3.1. ■

While Remark 7.3.2 is a kind of completeness result, it is not satisfactory in the sense that it does not help one to construct a Noetherian topology on a given set. Let us now turn our attention to a more practical question. As we saw in Corollary 7.2.7, refinement functions can have least fixed points that are not Noetherian topologies. The question is the following: given a refinement function R, does there exist a best approximation of R via topology expanders?

Let us write $[\![A \to B]\!]_{\mathsf{mon}}$ to denote the set of monotone functions from a quasi-order $(A, \leq_A)$ to another quasi-order $(B, \leq_B)$, and for two functions $f, g \in [\![A \to B]\!]_{\mathsf{mon}}$, we say that f is *pointwise below* g, written $f \leq_{\mathrm{pt}} g$, whenever for all $x \in A$, $f(x) \leq_B g(x)$.

Fact 7.3.3. Let $(A, \leq_A)$ be a quasi-order and $(B, \leq_B)$ be a complete lattice. Then $([\![A \to B]\!]_{\mathsf{mon}}, \leq_{\mathrm{pt}})$ is a complete lattice, where infima are computed pointwise, i.e., $(\inf_{i \in I} f)(x) \stackrel{\mathsf{def}}{=} \inf_{i \in I} f(x)$.

Recall that $\mathrm{Top}(\mathsf{X})$ is a complete lattice and $\mathrm{NTop}(\mathsf{X})$ is a downwards closed subset of $\mathrm{Top}(\mathsf{X})$ for inclusion. This is immediately lifted to functions.

Remark 7.3.4. The set of refinement functions is a downwards closed subset of $[\![\mathrm{Top}(\mathsf{X}) \to \mathrm{Top}(\mathsf{X})]\!]_{\mathsf{mon}}$.

Let us fix a set X. The collection of topologies over X is itself a set, and forms a complete lattice for inclusion.

Definition 7.3.5. *Let X be a set. Let us write $[\![\mathrm{Top}(\mathsf{X}) \to \mathrm{Top}(\mathsf{X})]\!]^{\mathsf{sub}}_{\mathsf{mon}}$ for the set of functions $f \in [\![\mathrm{Top}(\mathsf{X}) \to \mathrm{Top}(\mathsf{X})]\!]_{\mathsf{mon}}$ such that $f = \mathrm{const}_{\tau_{\mathsf{disc}}}$, or f is a topology expander.*

Lemma 7.3.6. *Let X be a set. Then, $[\![\mathrm{Top}(\mathsf{X}) \to \mathrm{Top}(\mathsf{X})]\!]^{\mathsf{sub}}_{\mathsf{mon}}$ is a complete lattice where infima are computed pointwise.*

Proof. Let $(f_i)_{i\in I}$ be a family of functions in $[\![\mathrm{Top}(\mathsf{X}) \to \mathrm{Top}(\mathsf{X})]\!]^{\mathsf{sub}}_{\mathsf{mon}}$. Let us construct an infimum to this family. Assume without loss of generality that for all $i \in I$, $f_i \neq \mathrm{const}_{\tau_{\mathsf{disc}}}$, which does not change the infimum. In particular, f_i is a topology expander for all $i \in I$.

If I is empty, then $\mathrm{const}_{\tau_{\mathsf{disc}}}$ is the infimum (and is computed pointwise). If I is non-empty, then let us write $f : \tau \mapsto \bigcap_{i\in I} f_i(\tau)$. Notice that f is monotone. Because I is non-empty, there exists $i \in I$ such that $f \leq_{\mathrm{pt}} f_i$, hence f is a refinement function thanks to Remark 7.3.4. Let τ be a Noetherian topology such that $\tau \subseteq f(\tau)$, and H be a closed subset of τ. For all $i \in I$, $\tau \subseteq f(\tau) \subseteq f_i(\tau)$; moreover, because f_i is a topology expander, $f_i(\tau){\downarrow}H \subseteq f_i(\tau{\downarrow}H){\downarrow}H$. We conclude that f is a topology expander as follows:

$$\begin{aligned} f(\tau){\downarrow}H &= \left(\bigcap_{i\in I} f_i(\tau)\right){\downarrow}H \\ &= \bigcap_{i\in I} (f_i(\tau){\downarrow}H) \\ &\subseteq \bigcap_{i\in I} (f_i(\tau{\downarrow}H){\downarrow}H) \\ &= f(\tau{\downarrow}H){\downarrow}H \end{aligned}$$ ■

Corollary 7.3.7. *Let X be a set, and R be a refinement function over X. There exists a largest topology expander E_{R} below R for the pointwise ordering.*

Proof. Let f be the supremum of $\{ g \in [\![\mathrm{Top}(\mathsf{X}) \to \mathrm{Top}(\mathsf{X})]\!]^{\mathsf{sub}}_{\mathsf{mon}} : g \leq_{\mathrm{pt}} \mathsf{R} \}$, which exists thanks to Lemma 7.3.6. It is an easy check that $f \leq_{\mathrm{pt}} \mathsf{R}$, hence f is a refinement function because of Remark 7.3.4. We have two cases, either $f = \mathrm{const}_{\tau_{\mathsf{disc}}}$, in which case $f(\tau_{\mathsf{triv}}) = \mathrm{const}_{\tau_{\mathsf{disc}}}(\tau_{\mathsf{triv}}) = \tau_{\mathsf{disc}}$ Noetherian, and f is a topology expander thanks to Example 7.3.1; or f is a topology expander by Definition 7.3.5. ■

Note that Corollary 7.3.7 is not constructive, and we have not managed to provide an explicit description of this "largest approximation." Recall that over finite words the "natural" refinement function $\mathsf{R}_{\mathsf{pref}}$ was not a topology expander, and an ad-hoc topology expander $\mathsf{E}_{\mathrm{words}}$ was designend to recover the subword topology (that is Noetherian). The following example shows that one cannot obtain $\mathsf{E}_{\mathrm{words}}$ as the largest approximation of $\mathsf{R}_{\mathsf{pref}}$.

Example 7.3.8. Let $\Sigma \stackrel{\text{def}}{=} \{a, b\}$ be a finite alphabet. Then, $\mathsf{E}^{\tau_{\mathsf{disc}}}_{\text{words}} \not\leq_{\text{pt}} \mathsf{R}_{\mathsf{pref}}$. In particular, $\mathsf{E}^{\tau_{\mathsf{disc}}}_{\text{words}} \not\leq_{\text{pt}} \mathsf{E}_{\mathsf{R}_{\mathsf{pref}}}$. Therefore, $\mathsf{E}^{\tau_{\mathsf{disc}}}_{\text{words}}$ is not the largest approximation of $\mathsf{R}_{\mathsf{pref}}$.

7.4. Discussion

What About Ordinal Invariants? Let us now briefly explain that Theorem 7.2.33 does not provide any information about the ordinal invariants associated with the space, mainly due to its non-constructive nature. For that, let us first recall that one can associate an ordinal invariant with each Noetherian spaces, similarly to what is done for well-quasi-orderings. This ordinal invariant is called the stature of a Noetherian space (X, τ), and is defined by $\|(\mathsf{X}, \tau)\| \stackrel{\text{def}}{=} \mathrm{h}(\mathcal{H}(\tau)) - 1$[17].

17: The subtraction is defined as the identity on limit ordinals, and the predecessor on successor ordinals. It is needed because the stature of $(\mathbb{N}, \mathsf{Alex}(\leq))$ is expected to be ω, but $\mathrm{h}(\mathcal{H}(\mathsf{Alex}(\leq))) = \omega + 1$ because of the presence of the entire space $\mathbb{N}$ as a closed subset.

The stature provides a fined grained control over the "complexity" of the Noetherian space at hand. It is tempting to refine the result of Theorem 7.2.33 under the assumption that the effect of the topology expander E on the stature of the topologies is controlled. Let us show that the naïve construction that first comes to mind cannot be applied.

Example 7.4.1. Let $\mathsf{X} = \Sigma^\star$ where Σ is a two letter alphabet with the discrete topology τ_{disc}. Then, $\|\mathsf{lfp}_\tau.\mathsf{E}_{\text{words}}(\tau)\| = \|{\mathsf{E}_{\text{words}}}^{\omega}(\tau_{\mathsf{triv}})\| = \omega^\omega$, but $\omega^\omega \not\leq \sup_{n<\omega} \|{\mathsf{E}_{\text{words}}}^{n}(\tau_{\mathsf{triv}})\| = \omega$.

[50]: Goubault-Larrecq and Laboureix (2023), 'Statures and Sobrification Ranks of Noetherian Spaces'

Proof. We have proven that the subword topology is exactly the least fixed point of $\mathsf{E}_{\text{words}}$ and is obtained after ω iterations. Moreover, it is known from [50, Theorem 12.22] that the stature of the subword topology over a two letter alphabet is exactly ω^ω. Notice that if τ is finite, then $\mathsf{E}_{\text{words}}(\tau)$ is finite too. As a consequence, the stature $\|{\mathsf{E}_{\text{words}}}^{n}(\tau_{\mathsf{triv}})\|$ is finite for all $n \in \mathbb{N}$. ■

The main problem identified by Example 7.4.1 is that even though every closed set H in the subword topology eventually appears in some ${\mathsf{E}_{\text{words}}}^{n}(\tau_{\mathsf{triv}})$ where $n \in \mathbb{N}$, decreasing sequences of such sets might go arbitrarily "up and down." This was already encountered when dealing with topological bad sequences of open subsets $(U_i)_{i\in\mathbb{N}}$, where the minimality allowed us to control the evolution of the parameter n (written $\mathrm{depth}(U_i)$).

One possible direction to overcome this difficulty would be to follow the usual techniques to prove upper bounds on the stature of Noetherian spaces, namely, the detour through the sobrification of the space [50]. Let us remark that a generic upper bound would implicitly reprove Theorem 7.2.33.

Towards Effective Representations. In order to use Noetherian spaces in practical verification algorithms, one needs to be able to "effectively" represent those spaces. As a consequence, it is a natural to ask whether Theorem 7.2.33 can be adapted to provide an effective description under reasonable assumptions. First, let us recall the definition of an effective representation for a Noetherian space.

Definition 7.4.2 S-representation [34, Definition 4.1]**.** Let (X, τ) be a topological space. An *S-representation* of (X, τ) is a tuple $(\mathsf{S}, [\![\cdot]\!], \trianglelefteq, \vec{s}, \sqcap)$ where

- S is a recursively enumerable set;
- $[\![\cdot]\!] \colon \mathsf{S} \twoheadrightarrow \mathcal{S}(\mathsf{X})$;
- $\trianglelefteq$ is a decidable relation such that for all $a, b \in \mathsf{S}$, $a \trianglelefteq b \iff [\![a]\!] \subseteq [\![b]\!]$;
- $\vec{s}$ is a finite subset of S such that $\mathsf{X} = \bigcup_{a \in \vec{s}} [\![a]\!]$;
- $\sqcap \colon \mathsf{S} \times \mathsf{S} \to \mathcal{P}_{\mathsf{fin}}(\mathsf{S})$ is a computable function such that for all $a, b \in \mathsf{S}$, $[\![a]\!] \cap [\![b]\!] = \bigcup_{c \in a \sqcap b} [\![c]\!]$.

Notice that the effective representations rely on the characterisation of the irreducible closed subsets of the topological space. As topology expanders are better suited do deal with open subsets, let us briefly demonstrate how Theorem 7.2.33 gives us a coarse control on the irreducible closed subsets of a limit topology.

Beware that in Lemma 7.4.3, it is not true that a subset is irreducible closed in the limit topology if and only if it is irreducible closed in some step. Consider for instance a Noetherian topology θ that is Noetherian and where X is not irreducible closed. Then, const_θ is a topology expander, X is an irreducible closed subset of τ_{triv}, but X is not an irreducible closed subset of $\mathsf{lfp}_\tau.\,\mathrm{const}_\theta(\tau) = \theta$.

Lemma 7.4.3. *Let E be a topology expander over a set X, and α be a limit ordinal. A subset H of X is irreducible closed in $\mathsf{E}^\alpha(\tau_{\mathsf{triv}})$ if and only if there exists $\beta < \alpha$ such that for all $\beta \leq \gamma < \alpha$, H is an irreducible closed subset of $\mathsf{E}^\gamma(\tau_{\mathsf{triv}})$.*

Proof. Because the topology $\mathsf{E}^\alpha(\tau_{\mathsf{triv}})$ is Noetherian and generated by $\bigcup_{\beta<\alpha} \mathsf{E}^\beta(\tau_{\mathsf{triv}})$ (the latter being closed under finite intersections), we know that $\mathsf{E}^\alpha(\tau_{\mathsf{triv}}) = \bigcup_{\beta<\alpha} \mathsf{E}^\beta(\tau_{\mathsf{triv}})$.

Let H be an irreducible closed subset of $\mathsf{E}^\alpha(\tau_{\mathsf{triv}})$. In particular, there exists $\beta < \alpha$ such that H is a closed subset of $\mathsf{E}^\beta(\tau_{\mathsf{triv}})$. It is immediate from the monotonicity of $(\mathsf{E}^\gamma(\tau_{\mathsf{triv}}))_{\gamma<\alpha}$ that H is a closed subset of $\mathsf{E}^\gamma(\tau_{\mathsf{triv}})$ for all $\beta \leq \gamma < \alpha$. Assume for contradiction that H is reducible in $\mathsf{E}^\gamma(\tau_{\mathsf{triv}})$ for some $\beta \leq \gamma < \alpha$. Because $\mathsf{E}^\gamma(\tau_{\mathsf{triv}}) \subseteq \mathsf{E}^\alpha(\tau_{\mathsf{triv}})$, we conclude that H is reducible in $\mathsf{E}^\alpha(\tau_{\mathsf{triv}})$ which is absurd.

Conversely, assume that there exists $\beta < \alpha$ such that for all $\beta \leq \gamma < \alpha$, H is a closed subset of $\mathsf{E}^\gamma(\tau_{\mathsf{triv}})$ that is irreducible. Let U_1, U_2 be two open subsets of $\mathsf{E}^\alpha(\tau_{\mathsf{triv}})$ that intersect H. There exists $\gamma_1, \gamma_2 < \alpha$ such that $U_1 \in \mathsf{E}^{\gamma_1}(\tau_{\mathsf{triv}})$, and $U_2 \in \mathsf{E}^{\gamma_2}(\tau_{\mathsf{triv}})$. Because the sequence $(\mathsf{E}^\gamma(\tau_{\mathsf{triv}}))_{\gamma<\alpha}$ is monotone, we conclude that there exists $\gamma < \alpha$ such that

- $\gamma_1 \leq \gamma$,
- $\gamma_2 \leq \gamma$,
- $\beta \leq \gamma$,
- $U_1, U_2 \in \mathsf{E}^\gamma(\tau_{\mathsf{triv}})$.

Because H is irreducible in $\mathsf{E}^{\gamma}(\tau_{\mathsf{triv}})$, $H \cap U_1 \cap U_2 \neq \emptyset$. We have proven that H is irreducible in $\mathsf{E}^{\alpha}(\tau_{\mathsf{triv}})$, ∎

In order to build an effective representation of a limit topology, we would like to add an "effectiveness" condition to the topology expander. It should work directly on effective representations, but should also allow us to convert closed sets between the input representation and the output representation.

Definition 7.4.4. *Let E be a topology expander over a set X. We say that E is an effective topology expander if there exists a computable map SE such that sends an S-effective representation $(\mathsf{S}, [\![\cdot]\!], \trianglelefteq, \vec{s}, \sqcap)$ of a Noetherian topology τ, to an S-effective representation $(\mathsf{S}', [\![\cdot]\!]', \trianglelefteq', \vec{s}', \sqcap')$ of the topology $\mathsf{E}(\tau)$, and a computable map $\bar{\mathsf{SE}} \colon \mathsf{S} \to \mathcal{P}_{\mathsf{fin}}(\mathsf{S}')$ such that for every code $c \in \mathsf{S}$, $\bigcup_{c' \in \bar{\mathsf{SE}}(c)} [\![c']\!]' = [\![c]\!]$.*

Lemma 7.4.5. *Let E be an effective topology expander over a set X. Assume moreover that $\mathcal{S}(\tau) \subseteq \mathcal{S}(\mathsf{E}(\tau))$ for all Noetherian topologies τ that satisfy $\tau \subseteq \mathsf{E}(\tau)$. Then, $\mathsf{E}^{\omega}(\tau_{\mathsf{triv}})$ has an effective representation.*

Proof. It is an easy check that the trivial topology has an effective representation, which we will call $(\mathsf{S}_0, [\![\cdot]\!]_0, \trianglelefteq_0, \vec{s}_0, \sqcap_0)$. Let us write, for every $n \in \mathbb{N}$, $(\mathsf{S}_{n+1}, [\![\cdot]\!]_{n+1}, \trianglelefteq_{n+1}, \vec{s}_{n+1}, \sqcap_{n+1})$ as the image of $(\mathsf{S}_n, [\![\cdot]\!]_n, \trianglelefteq_n, \vec{s}_n, \sqcap_n)$ through SE. Because of the extra assumption over E, the representation map $\bar{\mathsf{SE}} \colon \mathsf{S}_n \to \mathcal{P}_{\mathsf{fin}}(\mathsf{S}_{n+1})$ is actually a map from S to S_n, for every $n \in \mathbb{N}$. Let us now define the representation of the limit as follows:

- $\mathsf{S} \overset{\text{def}}{=} \sum_{i \in \mathbb{N}} \mathsf{S}_i$;
- $[\![\cdot]\!] \colon \mathsf{S} \twoheadrightarrow \mathcal{S}(\mathsf{X})$ that maps (i, a) where $a \in \mathsf{S}_i$ to $[\![a]\!]_i$;
- $(i, a) \trianglelefteq (j, b) \in \mathsf{S}$ if and only if $\bar{\mathsf{SE}}^{i'}(a) \trianglelefteq_{\max(i,j)} \bar{\mathsf{SE}}^{j'}(b)$, where $i' \overset{\text{def}}{=} \max(i, j) - i$ and $j' \overset{\text{def}}{=} \max(i, j) - j$;
- $\vec{s} = \{ (0, c) \colon c \in \vec{s}_0 \}$;
- $(i, a) \sqcap (j, b)$ is defined as $\bar{\mathsf{SE}}^{i'}(a) \sqcap_{\max(i,j)} \bar{\mathsf{SE}}^{j'}(b)$, where $i' \overset{\text{def}}{=} \max(i, j) - i$ and $j' \overset{\text{def}}{=} \max(i, j) - j$.

It is a routine check that the functions defined are computable. The only real thing to check is that $[\![\cdot]\!]$ is a surjective map from S to $\mathcal{S}(\mathsf{E}^{\omega}(\tau_{\mathsf{triv}}))$. This follows from the extra assumption of E, together with the characterisation of irreducible closed subsets that was obtained in Lemma 7.4.3. ∎

It remains for us to show that the extra assumption that irreducible closed subsets remain irreducible after applying E is reasonable. However, this is not an easy task. For instance, if (X, τ) is a Noetherian space that is not irreducible, then const_{τ} is a topology expander, but $\mathcal{S}(\tau_{\mathsf{triv}}) \not\subseteq \mathcal{S}(\mathrm{const}_{\tau}(\tau_{\mathsf{triv}}))$ because $\mathsf{X} \notin \mathcal{S}(\tau)$.

We know for a fact that this does not happen in the case of finite words, where irreducible closed subsets are well-understood, but we were unable to prove the following conjecture about $\mathsf{E}^{\theta}_{\mathrm{words}}$:

Conjecture 7.4.6 (Effective word expander)**.** Let (X, θ) be a Noetherian space. Then, $\mathsf{E}^{\theta}_{\text{words}}$ is a topology expander over $\mathsf{X}^{\star}$ such that $\mathcal{S}(\tau) \subseteq \mathcal{S}(\mathsf{E}^{\theta}_{\text{words}}(\tau))$ for all $\tau \in \mathsf{Top}(\mathsf{X}^{\star})$ satisfying $\tau \subseteq \mathsf{E}^{\theta}_{\text{words}}(\tau)$.

Inductive Constructions? Because of the connection between lpps and Noetherian spaces,[18] one might think that the techniques introduced to prove Theorem 7.2.33 could be transported, by carefully handling the Boolean subalgebra of definable subsets. However, the very notion of topological minimal bad sequence ceases to make sense in the realm of lpps: one has to first chose an definable open subset, and then a sequence that is a covering of this subset. In particular, the analogue of Lemma 7.2.24 cannot be proven.

18: Which is discussed in Section 6.2 and particularly via Lemma 6.2.1 on page 146.

Furthermore, if one considers the topology and the Boolean subalgebra of definable subsets independently, analogues of Theorem 7.2.33 trivialise. Let us assume that a fixed point theorem exists akin to the one for Noetherian topologies. Let $\langle\!\langle X, \tau, \mathcal{B} \rangle\!\rangle$ be such a least fixed point, then one can build the space $\langle\!\langle X, \tau_{\mathsf{triv}}, \mathcal{P}(\mathcal{B}) \rangle\!\rangle$ that is a lpps, the least fixed point above this set should also be lpps by the same least fixed point argument, however it would also be Noetherian. As a consequence, the original space was also Noetherian, and one could already prove this fact using Theorem 7.2.33.

The Telescope Topology. There exists a largest topology expander under any refinement function (see Corollary 7.3.7). However, this non-constructive argument does not provide a description of the topology expander corresponding to $\mathsf{R}_{\mathsf{pref}}$. We conjecture that its least fixed point is connected to the "telescope topology" introduced in [45, Exercise 9.7.36].[19] This would demonstrate how Corollary 7.3.7 can be leveraged to construct *non-trivial* "best effort" Noetherian topologies for free.

19: Beware that the "telescope topology" is called the "prefix topology" in the exercise, which is not consistent with the definitions used in this document.

Definition 7.4.7. *Let (Σ, θ) be a Noetherian space. We define $\mathsf{E}^{\theta}_{\text{tel}}(\tau)$ over $\Sigma^{\star}$ as the topology generated by the closed subsets $\{ \varepsilon \} \cup HH'$ where H is a closed subset of θ and H' is a closed subset of τ satisfying $H' = \downarrow_{\leq_{\mathsf{suf}}} H'$.*

Conjecture 7.4.8 (Telescope Topology)**.** Let θ be a topology. Then, $\mathsf{E}^{\theta}_{\text{tel}} \leq_{\text{pt}} \mathsf{R}_{\mathsf{pref}}$, $\mathsf{E}^{\theta}_{\text{tel}}$ is a topology expander, and the least fixed point of $\mathsf{E}^{\theta}_{\text{tel}}$ is the "telescope topology" introduced in [45, Exercise 9.7.36].

Open Questions. Let us conclude with a conjecture that generalises Example 7.3.1 to non-constant functions. Notice that Theorem 7.2.33 proves that Item 2 implies Item 1 in Conjecture 7.4.9.

Conjecture 7.4.9 (Completeness of topology expanders)**.** Let R be a refinement function over a set X. The following are equivalent:

1. For all Noetherian topology τ, for all ordinal α, $\mathsf{R}^{\alpha}(\tau)$ is Noetherian;
2. There exists a smallest topology expander E above R.

References

[19] Jean Daligault, Michael Rao and Stéphan Thomassé. 'Well-Quasi-Order of Relabel Functions'. In: *Order* 27 (2010), pp. 301–315. DOI: 10.1007/s11083-010-9174-0. URL: https://doi.org/10.1007/s11083-010-9174-0 (cit. on pp. 37, 40, 44–46, 179, 235).

[23] Stéphane Demeri, Alain Finkel, Jean Goubault-Larrecq, Sylvain Schmitz and Philippe Schnoebelen. 'Algorithmic Aspects of WQO Theory (MPRI course)'. 2012. URL: https://cel.archives-ouvertes.fr/cel-00727025 (cit. on pp. 3, 139, 178, 181, 239).

[24] Nachum Dershowitz and Iddo Tzameret. 'Gap Embedding for Well-Quasi-Orderings'. In: *Electronic Notes in Theoretical Computer Science* 84 (2003), pp. 80–90. DOI: 10.1016/S1571-0661(04)80846-6. URL: http://www.sciencedirect.com/science/article/pii/S1571066104808466 (cit. on p. 179).

[34] Alain Finkel and Jean Goubault-Larrecq. 'Forward analysis for WSTS, part I: completions'. In: *Mathematical Structures in Computer Science* 30 (2020), pp. 752–832. DOI: 10.1017/S0960129520000195. (Visited on 13/10/2022) (cit. on pp. 185, 201).

[36] Anton Freund. 'From Kruskal's theorem to Friedman's gap condition'. In: *Mathematical Structures in Computer Science* 30 (2020), pp. 952–975. DOI: 10.1017/s0960129520000298 (cit. on pp. 179, 184, 229).

[40] Jean H. Gallier. 'Ann. Pure Appl. Logic: Erratum to "What's so special about Kruskal's Theorem and the ordinal Γ_0? A survey of some results in proof theory" [53 (1991) 199–260]'. In: *Annals of Pure and Applied Logic* 89 (1997), p. 275. DOI: 10.1016/S0168-0072(97)00043-2 (cit. on pp. 178, 179).

[44] Jean Goubault-Larrecq. 'Noetherian Spaces in Verification'. In: *Automata, Languages and Programming*. Ed. by Samson Abramsky, Cyril Gavoille, Claude Kirchner, Friedhelm Meyer auf der Heide and Paul G. Spirakis. Springer Berlin Heidelberg, 2010, pp. 2–21. DOI: 10.1007/978-3-642-14162-1_2 (cit. on pp. 53, 139, 178).

[45] Jean Goubault-Larrecq. *Non-Hausdorff Topology and Domain Theory*. Vol. 22. Cambridge University Press, 2013. DOI: 10.1017/cbo9781139524438 (cit. on pp. 3, 50, 52–54, 139, 159, 178, 179, 185, 186, 192, 193, 203, 208, 224, 239, 254, 255).

[46] Jean Goubault-Larrecq. *Aliaume Lopez' master theorem of Noetherian spaces*. 2022. URL: http://projects.lsv.ens-cachan.fr/topology/?p=6105 (visited on 08/02/2023) (cit. on p. 191).

[47] Jean Goubault-Larrecq. 'Infinitary Noetherian Constructions I. Infinite Words'. In: *Colloquium Mathematicum* 168 (2022), pp. 257–268. DOI: https://doi.org/10.4064/cm8077-4-2021 (cit. on pp. 139, 175, 178, 179, 209–211).

[48] Jean Goubault-Larrecq. 'Non-Hausdorff Topology and Domain Theory. Electronic supplements to the book – Errata'. 2022. URL: https://projects.lsv.ens-cachan.fr/topology/?page_id=12 (visited on 10/12/2022) (cit. on pp. 178, 179).

[49] Jean Goubault-Larrecq, Simon Halfon and Aliaume Lopez. *Infinitary Noetherian Constructions II. Transfinite Words and the Regular Subword Topology*. 2022 (cit. on pp. 175, 178, 179, 209, 211).

[50] Jean Goubault-Larrecq and Bastien Laboureix. 'Statures and Sobrification Ranks of Noetherian Spaces'. In: 2023, pp. 1–76. URL: https://arxiv.org/abs/2112.06828 (cit. on p. 200).

[51] Jean Goubault-Larrecq and Aliaume Lopez. 'On the word topology, and beyond'. 2020. URL: https://projects.lsv.ens-cachan.fr/topology/?page_id=2735 (visited on 08/02/2023) (cit. on p. 177).

[53] Jean Goubault-Larrecq, Monika Seisenberger, Victor L. Selivanov and Andreas Weiermann. 'Well Quasi-Orders in Computer Science (Dagstuhl Seminar 16031)'. In: *Dagstuhl Reports* 6 (2016), pp. 69–98. DOI: 10.4230/DagRep.6.1.69. URL: https://doi.org/10.4230/DagRep.6.1.69 (cit. on pp. 44, 178).

[58] Ryu Hasegawa. 'Two applications of analytic functors'. In: *Theoretical Computer Science* 272 (2002), pp. 113–175. DOI: 10.1016/S0304-3975(00)00349-2. URL: https://www.sciencedirect.com/science/article/pii/S0304397500003492 (cit. on pp. 179, 184, 207, 215, 220, 221, 224, 225, 227, 229).

[59] Graham Higman. 'Ordering by divisibility in abstract algebras'. In: *Proceedings of the London Mathematical Society* 3 (1952), pp. 326–336. DOI: 10.1112/plms/s3-2.1.326 (cit. on pp. 44, 158, 178, 181).

[65] Joseph B. Kruskal. 'The theory of well-quasi-ordering: A frequently discovered concept'. In: *Journal of Combinatorial Theory, Series A* 13 (1972), pp. 297–305. DOI: 10.1016/0097-3165(72)90063-5 (cit. on pp. 44, 158, 178, 179, 181).

[69] Aliaume Lopez. *On the word topology, and beyond*. 2020. URL: https://www.irif.fr/~alopez/posts/2020-09-23-word-topology-and-beyond.html (visited on 08/02/2023) (cit. on p. 177).

[72] Aliaume Lopez. 'Fixed Points and Noetherian Topologies'. In: *FOSSACS'23* (2023), pp. 1–14. DOI: 10.1145/3531130.3532498. URL: https://doi.org/10.1145/3531130.3532498 (cit. on pp. 63, 180, 191).

[76] Crispin St. John Alvah Nash-Williams. 'On well-quasi-ordering transfinite sequences'. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 61 (1965), pp. 33–39 (cit. on pp. 177–181).

[83] Benjamin Rossman. 'Homomorphism preservation theorems'. In: *Journal of the ACM* 55 (2008), 15:1–15:53. DOI: 10.1145/1379759.1379763 (cit. on pp. i, 24, 35, 37, 38, 40, 42, 55, 74, 175, 177, 235).

[88] Devendra Singh, Ali Maianguwa Shuaibu and Muhammad Ndayawo. 'Simplified Proof of Kruskal's Tree Theorem'. In: *Mathematical Theory and Modeling* 3 (2013), pp. 93–100. DOI: 10.13140/RG.2.2.12298.39363 (cit. on p. 179).

Inductive Constructions 8.

Outline of the chapter

In this chapter, we will leverage the master Theorem 7.2.33 from Chapter 7 (Topology expanders and Noetherian Topologies) to construct families of Noetherian spaces.

Goals of the chapter

At the end of the chapter, I hope you will be convinced that most of the previously known Noetherian topologies can be recovered through Theorem 7.2.33. Furthermore, we will have shown how the simplicity of the theorem allows us to use it in a categorical framework describing inductive datatypes, and obtain a generic Noetherian topology for those inductive datatypes, that correctly generalises what was known in the case of well-quasi-orderings.

Genesis. It is a thing to define a Noetherian topology inductively, but it is another to justify that "it is the right one." While a least-fixed-point definition is arguably simpler to defend, the next step was to prove that, when starting from an Alexandroff topology (i.e., starting from a well-quasi-order), the construction coincides with the "usual preorders." It turns out that inductive constructions in the case of well-quasi-orderings had already been studied since 2002 [58], through a completely different approach. This point of reference allowed us to state a correctness property in Theorem 8.2.33.

[58]: Hasegawa (2002), 'Two applications of analytic functors'

Contributions In Section 8.1, we demonstrate how one recovers previously known topologies as least-fixed-points, through a simple methodology. This pattern is then generalised to (suitably defined) inductive construction in Section 8.2 where the most important definition is the one of divisibility expander[1]. We then prove that the divisibility expander correctly generalises the picture in well-quasi-orderings via Theorem 8.2.33 of Subsection 8.2.2.

1: See Definition 8.2.17 p. 221

8.1. Applications

We now briefly explore topologies that can be proven to be Noetherian using Theorem 7.2.33. It should not be surprising that both the topological Higman lemma and the topological Kruskal theorem fit in the framework of topology expanders, as both were already proven using a minimal bad sequence argument. However, we will proceed to extend the use of topology expander to spaces for which the original proof did not use a minimal bad sequence argument, and illustrate how they can easily be used to define new Noetherian topologies.

8.1.1. Finite Words and Trees

Finite words. As a first example, we recover the topological Higman lemma [45, Theorem 9.7.33] because the subword topology is the least fixed point of $\mathsf{E}^{\theta}_{\text{words}}$, which is a topology expander (see Lemma 7.2.9 and Corollary 7.2.15).

Finite trees. It does not require much effort to generalise this proof scheme to the case of the topological Kruskal theorem [45, Theorem 9.7.46], the topological analogue of Kruskal's Tree Theorem. Before introducing this construction, let us use (as for the words) a shorthand notation for some specific set of trees.

[45]: Goubault-Larrecq (2013), *Non-Hausdorff Topology and Domain Theory*

Let U be a subset of Σ, and V be a subset of $\mathsf{T}(\Sigma)^{\star}$. Let us define $\diamond U\langle V\rangle$ to be the set of trees t such that there exists a subtree[2] t' of t with a root labelled by an element of U, and whose list of children belong to V.

2: That is, $t' \leq_{\mathsf{t\text{-}suf}} t$.

As for the subword topology, the definition is ad-hoc but correctly generalises the tree embedding relation because the tree topology of $\mathsf{Alex}(\leq)$ is the Alexandroff topology of $\leq_{\mathsf{t}}$, for every ordering well-quasi ordering $\leq$ over Σ [45, Exercise 9.7.48].

Recall that we write $u \leq_{\mathsf{w}} v$ when u is a scattered subword of v, and $t \leq_{\mathsf{t}} t'$ when t embeds in t' as a tree (see page 158).

Definition 8.1.1 [45, Definition 9.7.39]**.** Let (Σ, θ) be a topological space, and $\mathsf{T}(\Sigma)$ be the space of finite trees over Σ. The *tree topology* θ^{tree} is defined by induction as the coarsest topology such that $\diamond U\langle V\rangle$ is open whenever $U \in \theta$, and V is an open set of $((\mathsf{T}(\Sigma), \theta^{\mathsf{tree}})^{\star}, \theta^{\mathsf{tree}\star,\mathsf{t}})$.

Theorem 8.1.2 [45, Theorem 9.7.46]**.** Let (Σ, θ) be a topological space. Then, $(\mathsf{T}(\Sigma), \theta^{\mathsf{tree}})$ is Noetherian if and only if (Σ, θ) is Noetherian.

This is the *topological Kruskal theorem*.

The inductive nature of Definition 8.1.1 naturally leads to a corresponding topology expander.

Definition 8.1.3. *Let (Σ, θ) be a topological space. Let $\mathsf{E}^{\theta}_{\text{trees}}$ be the function that maps a topology τ to the topology generated by the sets $\diamond U\langle V\rangle$, for $U \in \theta$ and $V \in \tau^{\star,t}$.*

The proof of the following lemma is roughly the same as the corresponding ones for the subword topology, see Lemmas 7.2.9 and 7.2.10 and Corollary 7.2.15.

Lemma 8.1.4. *The tree topology is the least fixed point of $\mathsf{E}^{\theta}_{\text{trees}}$, which is a topology expander.*

Proof. The fact that $\mathsf{E}^{\theta}_{\text{trees}}$ is a refinement function whose least fixed point is the tree topology is left as an exercise. The main technical point is to check that it is a topology expander.

Notice that the sets $\diamond U\langle[V_1,\ldots,V_n]\rangle$, where U ranges over θ, n ranges over $\mathbb{N}$, and V_i ranges over τ for $1 \leq i \leq n$, form a subbasis of the topology $\mathsf{E}^{\theta}_{\text{trees}}(\tau)$. Indeed, the map $f : V \mapsto \diamond U\langle V\rangle$ satisfies $f(V_1 \cup V_2) = f(V_1) \cup f(V_2)$, and the open subsets $[V_1,\ldots,V_n]$ form a basis of the subword topology over $\mathsf{T}(\Sigma)^{\star}$.

Thanks to Lemma 7.2.19, proving that $\mathsf{E}^{\theta}_{\text{trees}}$ is a topology expander amounts to the following property: for every Noetherian topology τ satisfying $\tau \subseteq \mathsf{E}^{\theta}_{\text{trees}}(\tau)$, for every closed subset H of τ, for every open subset $\diamond U\langle[V_1,\ldots,V_n]\rangle$ in $\mathsf{E}^{\theta}_{\text{trees}}(\tau)$, $H \cap \diamond U\langle[V_1,\ldots,V_n]\rangle = H \cap \diamond U\langle[V_1 \cap H,\ldots,V_n \cap H]\rangle$.

Notice that in this last equality, one inclusion is always true: $H \cap \diamond U\langle[V_1,\ldots,V_n]\rangle \supseteq H \cap \diamond U\langle[V_1 \cap H,\ldots,V_n \cap H]\rangle$.

For the converse inclusion, remark that if H is a closed subset of $\mathsf{E}^{\theta}_{\text{trees}}(\tau)$, then it is downwards closed for $\leq_{\mathsf{t}}$. In particular, H is closed under taking subtrees. Let t be a tree in $H \cap \diamond U\langle[V_1,\ldots,V_n]\rangle$. There exists a subtree t' of t, whose root belongs to U an list of children belong to $[V_1,\ldots,V_n]$. Now, every child of t' is a subtree of t, hence belongs to H. As a consequence, t belongs to $H \cap \diamond U\langle[V_1 \cap H,\ldots,V_n \cap H]\rangle$. ∎

Corollary 8.1.5. *Let* (Σ, θ) *be a Noetherian space.*

Then, $(\mathsf{T}(\Sigma), \theta^{\mathsf{tree}})$ *is a Noetherian space.*

8.1.2. Ordinal Datatypes

Let us now demonstrate how Theorem 7.2.33 can be applied to spaces which are proved to be Noetherian without using a minimal bad sequence argument. This will also be the occasion to go beyond the world of well-quasi-orderings: we will define constructors that create non well-quasi-ordered spaces even when applied to some finite sets.

Ordinal words. For that, let us consider the set $\Sigma^{<\alpha}$ of words of ordinal length less than α, where α is a fixed ordinal. The mapping based definition of $\leq_{\mathsf{w}}$ (Definition 6.3.22) naturally generalises to ordinal length words, but is in general not a well quasi order. As a consequence, this space will also provide an example of a topological minimal bad sequence argument that has no counterpart in the realm of well-quasi-orders.

Let us start [3] with $\alpha = \omega + 1$, in which case $\Sigma^{<\alpha} = \Sigma^{\star} \uplus \Sigma^{\omega}$. Let us recover a Noetherian topology. In this case, we adapt Definition 6.3.26 to talk about infinite periodic behaviours. To that end, let us write $[U_1,\ldots,U_n|U_\infty]$ a shorthand notation for the set

$$\Sigma^* U_1 \Sigma^* \ldots \Sigma^* U_n (\Sigma^* U_\infty \Sigma^*)^\omega \quad .$$

3: We start with this “first infinite case” both for historical reasons – $\omega + 1$ was handled in [47] while the general construction only appeared in [49] – and because there are deep connection between this space and the sobrification of finite words.

Definition 8.1.6 [47, Definition 5.1]**.** Let (Σ, θ) be a topological space. The *asymptotic subword topology* $\theta^{\leq \omega}$ is defined as the coarsest topology containing the following open subsets: $[U_1, \dots, U_n | U_\infty]$ where $n \in \mathbb{N}$, $U_i \in \theta$ for $1 \leq i \leq n$, and $U_\infty \in \theta$.

Definition 8.1.7. *Let (Σ, θ) be a Noetherian space. We define the* asymptotic subword expander $\mathsf{E}^{\theta}_{\text{i-words}}$ *as mapping a topology τ to the topology generated by the following sets:*

- $\uparrow_{\leq_w} UV$ *for $U, V \in \tau$;*
- $\uparrow_{\leq_w} [|W]$*, for $W \in \theta$;*
- $\uparrow_{\leq_w} W$*, for $W \in \theta$.*

As per Lemma 8.1.4, the upcoming lemma follows the same proof scheme as Lemmas 7.2.9 and 7.2.10 and Corollary 7.2.15.

Lemma 8.1.8. *Let (Σ, θ) be a Noetherian space. The asymptotic subword topology is the least fixed point of $\mathsf{E}^{\theta}_{\text{i-words}}$ which is a topology expander.*

Proof. Let us first briefly check that $\mathsf{E}^{\theta}_{\text{i-words}}$ is a refinement function. It is clearly defined in a monotone way. Furthermore, if τ is Noetherian, then the following topologies are too, for similar reasons as for the case of $\mathsf{E}^{\theta}_{\text{words}}$:

- $\langle \uparrow_{\leq_w} UV : U, V \in \tau \rangle_{\mathsf{topo}}$;
- $\langle \uparrow_{\leq_w} [|W] : W \in \theta \rangle_{\mathsf{topo}}$;
- $\langle \uparrow_{\leq_w} W : W \in \theta \rangle_{\mathsf{topo}}$.

Leveraging the fact that the join of Noetherian topologies remains Noetherian (see Lemma D.6.6), we conclude that $\mathsf{E}^{\theta}_{\text{i-words}}$ is Noetherian.

Let us now prove that $\mathsf{E}^{\theta}_{\text{i-words}}$ is a topology expander. For that, assume that $\tau \subseteq \mathsf{E}^{\theta}_{\text{i-words}}(\tau)$ and is Noetherian. Then, closed subsets of τ are downwards closed for $\leq_w$ by construction. As a consequence, if U, V are open subsets of τ, and H is a closed subset of τ, we have

$$\uparrow_{\leq_w} UV \cap H = \uparrow_{\leq_w} (U \cap H)(V \cap H) \cap H \quad .$$

Therefore,[4] it is clear that for every closed subset H of τ, the following inclusion holds:

$$\mathsf{E}^{\theta}_{\text{i-words}}(\tau) {\downarrow} H) \subseteq \mathsf{E}^{\theta}_{\text{i-words}}(\tau {\downarrow} H) {\downarrow} H) \quad .$$

Furthermore, a simple induction shows that for all $n \in \mathbb{N}$, $U_1, \dots, U_n \in \tau$ and $W \in \theta$, the set $[U_1, \dots, U_n | W]$ belongs to ${\mathsf{E}^{\theta}_{\text{i-words}}}^{n}(\tau_{\mathsf{triv}})$.

Finally, the only claim left to address is the fact that the regular transfinite subword topology is a fixed point of $\mathsf{E}^{\theta}_{\text{i-words}}$. The only technical element is to notice that the sets $[U_1, \dots, U_n | W][V_1, \dots, V_n | W']$ and $[U_1, \dots, U_n | W][V_1, \dots, V_n |]$ are always empty, as the length of the words they contain is greater than ω. ∎

4: To be completely precise, one leverages Lemma 7.2.19, stating that expanders need only be checked on a subbasis of the topologies.

Corollary 8.1.9. *Let* (Σ, θ) *be a Noetherian space.*

Then, $(\Sigma^{<\omega+1}, \tau^{\leq\omega}\theta)$ *is a Noetherian space.*

Let us pause for a moment to appreciate the fact that the original proof did not use a topological minimal bad sequence argument, but rather a decomposition of the behaviours into a "finite part" and an "recurrent part." The main idea is that given a Noetherian topology over infinite words, and a word $w \in \Sigma^\omega$, one can build the "recurrent letters" of a word as a closed subset defined by $\overline{\bigcap_{i\in\mathbb{N}} \{ w_j : j < i \}}$. Because the topology θ is Noetherian, the decreasing sequence of closed subsets of the adherence of the letters of suffixes is constant after a given $n_0 \in \mathbb{N}$. This provides a way to split words that is compatible with the asymptotic topology and allows us to reduce it to a product of two Noetherian spaces (see [47, Section 5] for a full proof).

Let us now shift our attention to general ordinal length words. Again, the original proof [49] does not use a minimal bad sequence argument, but directly computes an upper bound on the stature of the topology to prove that it is Noetherian. As a consequence, the proof is quite technical and relies on the in-depth study of the possible inclusions between the subbasic closed sets. For that reason, the topology was originally given through its closed subsets, and a concrete description of the open subsets has yet to be given.

[49]: Goubault-Larrecq, Halfon and Lopez (2022), *Infinitary Noetherian Constructions II. Transfinite Words and the Regular Subword Topology*

Remark that if $H \subseteq \Sigma$ and $\beta \leq \alpha$, then $H^{<\beta} \subseteq \Sigma^{<\alpha}$ and consists of words of length less than β with all of their letters in H.

Definition 8.1.10 [49]**.** Let (Σ, θ) be a topological space. The *regular transfinite subword topology*, over $\Sigma^{<\alpha}$, written $\theta^{<\alpha}$, is the topology generated by the following closed subsets: $H_1^{<\beta_1} \cdots H_n^{<\beta_n}$, for $n \in \mathbb{N}$, H_i closed in θ, $\beta_i < \alpha$.

Before we define a suitable topology expander, given an ordinal β and a set $U \subseteq \Sigma^{<\alpha}$, let us write $[|\beta|U]$ as the set of words w such that $w_{>\gamma} \in U$ for every $0 \leq \gamma < \beta$.[5]

5: We have written $w_{>\gamma}$ for the word w where the prefix $w_{\leq\gamma}$ has been removed.

Definition 8.1.11. *Let* (Σ, θ) *be a topological space, and* α *be an ordinal. The function* $\mathsf{E}^{\theta}_{\text{o-words}}$ *maps a topology* τ *to the topology generated by the following sets:*

- $\uparrow_{\leq_w} UV$ *for* $U, V \in \tau$*;*
- $\uparrow_{\leq_w} [|\beta|U]$*, for* $U \in \tau$, $\beta \leq \alpha$*;*
- $\uparrow_{\leq_w} W$*, for* $W \in \theta$*.*

Let us slowly prove that $\mathsf{E}^{\theta}_{\text{o-words}}$ is a topology expander whose least fixed point *contains* the regular transfinite subword topology. Note that the equality is not known to hold, and that no inductive description of the open subsets of the regular transfinite subword topology is known. The inclusion is sufficient to conclude that the regular transfinite subword topology is Noetherian, this time using a (much shorter) fixed point argument.

Let us first notice that the new constructor of $\mathsf{E}^{\theta}_{\text{o-words}}$ preserves Noetherian spaces.

Lemma 8.1.12. *Let (Σ, θ) be a Noetherian space, and τ be a Noetherian topology. Then $\mathsf{E}^{\theta}_{\text{o-words}}(\tau)$ is Noetherian.*

Proof. As a consequence of Lemma 7.2.10, the topology generated by the sets $\uparrow_{\leq_{\mathsf{w}}} UV$, and $\uparrow_{\leq_{\mathsf{w}}} W$ is Noetherian. Thanks to Lemma D.6.6, it suffices to check that the topology generated by the sets $\uparrow_{\leq_{\mathsf{w}}} [|\beta|U]$ is Noetherian to conclude that $\mathsf{E}^{\theta}_{\text{o-words}}(\tau)$ is too.

For that, consider a bad sequence $[|\beta_i|U_i]$ of open sets, indexed by $\mathbb{N}$. Because for every $i \in \mathbb{N}$, $\beta_i < \alpha+1$, we can extract our sequence so that $\beta_i \leq \beta_j$ when $i \leq j$. The extracted sequence is still bad and infinite. Because τ is Noetherian, there exists $i \in \mathbb{N}$, such that $U_i \subseteq \bigcup_{j<i} U_j$. Let us now conclude that $[|\beta_i|U_i] \subseteq \bigcup_{j<i} [|\beta_j|U_j]$, which is absurd.

Let $w \in [|\beta_i|U_i]$, and assume by contradiction that for all $j < i$, there exists a $\gamma_j < \beta_j \leq \beta_i$ such that $w_{>\gamma_j} \notin U_j$. Let $\gamma \stackrel{\text{def}}{=} \max_{j<i} \gamma_j < \beta_i$. The word $w_{>\gamma}$ does not belong to U_j for $j < i$, because U_j is upwards closed for $\leq_{\mathsf{w}}$. As a consequence, $w_{>\gamma} \notin \bigcup_{j<i} U_j$. However, $w_{>\gamma} \in U_i$, which is absurd. ∎

It is now an easy check that $\mathsf{E}^{\theta}_{\text{o-words}}$ is a topology expander.

Lemma 8.1.13. *Let (Σ, θ) be a topological space, and α be an ordinal. The map $\mathsf{E}^{\theta}_{\text{o-words}}$ is a topology expander.*

Proof. It is clear that $\mathsf{E}^{\theta}_{\text{o-words}}$ is monotone. Moreover, the closed sets H in $\mathsf{E}^{\theta}_{\text{o-words}}(\tau)$ are downwards closed with respect to $\leq_{\mathsf{w}}$. As a consequence, for every closed subset H, $(\uparrow_{\leq_{\mathsf{w}}} UV) \cap H = (\uparrow_{\leq_{\mathsf{w}}} (U \cap H)(V \cap H)) \cap H$, and $(\uparrow_{\leq_{\mathsf{w}}} [|\beta|U]) \cap H = (\uparrow_{\leq_{\mathsf{w}}} [|\beta|(U \cap H)]) \cap H$. To conclude that $\mathsf{E}^{\theta}_{\text{o-words}}$ is a topology expander, it remains to prove that it preserves Noetherian topologies, but this is exactly Lemma 8.1.12. ∎

Lemma 8.1.14. *Let (Σ, θ) be a topological space, and α be an ordinal. The least fixed point of $\mathsf{E}^{\theta}_{\text{o-words}}$ contains the regular transfinite subword topology.*

Proof. Let us check that every open subset in the regular transfinite subword topology is open in the least fixed point of $\mathsf{E}^{\theta}_{\text{o-words}}$. We prove by induction over n that a product $F_1^{<\beta_1} \dots F_n^{<\beta_n}$ has a complement that is open. The empty product is the set $\{ \varepsilon \}$, and is the complement of $\uparrow_{\leq_{\mathsf{w}}} \mathsf{X}$, which belongs to the least fixed point of $\mathsf{E}^{\theta}_{\text{o-words}}$.

Assume that $P \stackrel{\text{def}}{=} F^{<\beta} P'$. By induction hypothesis, P'^c is an open U in the least fixed point topology. Let us prove that $P^c = A \cup B$, where $A \stackrel{\text{def}}{=} \uparrow_{\leq_{\mathsf{w}}} \{ av \colon a \notin F \wedge av \in U \}$, and $B \stackrel{\text{def}}{=} \uparrow_{\leq_{\mathsf{w}}} [|\beta|U]$. To that end, let us first show that A and B are open in the least fixed point of $\mathsf{E}^{\theta}_{\text{o-words}}$. The set B is open because U is open. Let us prove by induction that whenever U is open and F is closed in θ, the set $F \rtimes U$ defined as $\uparrow_{\leq_{\mathsf{w}}} \{ av \colon a \notin F, av \in U \}$ is open. It is easy to check that $F \rtimes (\uparrow_{\leq_{\mathsf{w}}} W) = \uparrow_{\leq_{\mathsf{w}}} (W \cap F^c) \cup \uparrow_{\leq_{\mathsf{w}}} F^c W$. Moreover, $F \rtimes (\uparrow_{\leq_{\mathsf{w}}} UV) =$

$\uparrow_{\leq_w}(F \rtimes U)V$. Finally, for $\beta \geq 1$, $F \rtimes (\uparrow_{\leq_w}[|\beta|U]) = \uparrow_{\leq_w} F^c([|\beta'|U])$ with $\beta' = \beta$ if β is limit, and $\beta' = \gamma$ if $\beta = \gamma + 1$.

Let us now prove that $P^c \subseteq A \cup B$. Let $w \notin P$ and distinguish two cases.

- Either there exists a $\gamma < \beta$ such that $w_\gamma \notin F$. In which case we can assume that γ is the smallest ordinal with this property. Then, $w = w_{<\gamma} w_\gamma w_{>\gamma}$. Since $\gamma < \beta$, $w_\gamma \in F^{<\beta}$, hence $w_{\geq\gamma} \in U$ because $w \notin P$. As a consequence, $w \in A$, since $w_\gamma w_{>\gamma} \in A$, and $w_\gamma w_{>\gamma} \leq_w w$.
- Or $w_\gamma \in F$ for every $\gamma < \beta$. However, this proves that $w_{>\gamma} \in U$ for every $\gamma < \beta$, which means that $w \in B$.

Now, let us prove that $A \subseteq P^c$. Because P is downwards closed for $\leq_w$, it suffices to check that every word av with $a \notin F$ and $av \in U$ lies in P^c. Assume by contradiction that $av \in P$, then $av = u_1 u_2$ with $u_1 \in F^{<\beta}$ and $u_2 \in P'$. Because $a \notin F$, this proves that u_1 is the empty word, and that $u_2 = w \in P'$. This is absurd because $w \in U = (P')^c$.

Finally, let us show that $B \subseteq P^c$. Because P is downwards closed for $\leq_w$ it suffices to check that every word $w \in [|\beta|U]$ belongs to P^c. Assume by contradiction that such a word w is in P. One can write $w = uv$ with $u \in F^{<\beta}$ and $v \in P'$. However, $|u| = \gamma < \beta$, hence, $v = w_{>\gamma} \in U = (P')^c$ which is absurd.

We have proven that P^c is open. ■

Corollary 8.1.15. *Let* (Σ, θ) *be a Noetherian space.*

Then, the space $(\Sigma^{<\alpha}, \theta^{<\alpha})$ *is a Noetherian space.*

Remark that Definitions 7.2.8, 8.1.3 and 8.1.11 all follow the same blueprint: new open sets are built as upward closures for the corresponding quasi-order of the natural constructors associated with the space. We argue that this blueprint mitigates the canonicity issue and the complexity of Definitions 6.3.26, 8.1.1 and 8.1.10.

Ordinal branching trees. As an example of a new Noetherian topology derived using Theorem 7.2.33, we will consider *α-branching trees* $\mathsf{T}^{<\alpha}(\Sigma)$, i.e., the least fixed point of the constructor $X \mapsto \Sigma \times X^{<\alpha}$ where α is a given ordinal. This example was not known to be Noetherian, fails to be a well-quasi-order, and illustrates how Theorem 7.2.33 easily applies on inductively defined spaces, based on the "blueprint" of Definitions 7.2.8, 8.1.3 and 8.1.11.

As per ordinal length words, notice that the tree embedding relation defined on page 158 naturally generalises to ordinal branching trees. Similarly, the notation $\diamond U\langle V\rangle$ defined on page 208 also makes sense for ordinal branching trees.

Definition 8.1.16. *Let* (Σ, θ) *be a Noetherian space, and* α *be an ordinal. The* transfinite subtree expander $\mathsf{E}^{\theta}_{\text{o-trees}}$ *maps a topology* τ *to the topology generated by the sets* $\diamond U\langle V\rangle$ *where* $U \in \theta$, $V \in \tau^{<\alpha}$.

Definition 8.1.17. *Let* (Σ, θ) *be a Noetherian space, and* α *be an ordinal. The* transfinite subtree topology *over* $\mathsf{T}^{<\alpha}(\Sigma)$ *is the least* *fixed point of* $\mathsf{E}^{\theta}_{\text{o-trees}}$.

Lemma 8.1.18. *Let* (Σ, θ) *be a Noetherian space, and* α *be an ordinal. The transfinite subtree expander* $\mathsf{E}^{\theta}_{\text{o-trees}}$ *is a topology expander.*

Proof. It is clear that $\mathsf{E}^{\theta}_{\text{o-trees}}$ is monotone, and a closed set of $\mathsf{E}^{\theta}_{\text{o-trees}}(\tau)$ is always closed under subtrees. Therefore, $\diamond U \langle V \rangle \cap H = \diamond U \langle V \cap H^{<\alpha} \rangle \cap H$. Notice that $H^{<\alpha} \cap V \in (\tau{\downarrow}H)^{<\alpha}$. As a consequence, $V \cap H \in \mathsf{E}^{\theta}_{\text{o-trees}}(\tau{\downarrow}H){\downarrow}H$.

Let us now check that $\mathsf{E}^{\theta}_{\text{o-trees}}$ preserves Noetherian topologies. Let $W_i \stackrel{\text{def}}{=} \diamond U_i \langle V_i \rangle$ be a $\mathbb{N}$-indexed sequence of open sets in $\mathsf{E}^{\theta}_{\text{o-trees}}(\tau)$ where τ is Noetherian. The product of the topology θ and the regular transfinite subword topology over τ is Noetherian thanks to Lemma D.6.4 and Corollary 8.1.15. Hence, there exists an $i \in \mathbb{N}$ such that $U_i \times V_i \subseteq \bigcup_{j<i} U_j \times V_j$. As a consequence, $W_i \subseteq \bigcup_{j<i} W_j$. We have proven that $\mathsf{E}^{\theta}_{\text{o-trees}}(\tau)$ is Noetherian. ■

Theorem 8.1.19. *The* α*-branching trees endowed with the transfinite subtree topology forms a Noetherian space.*

At this point, we have proven that the framework of topology expanders allows us to build non-trivial Noetherian spaces. We argue that this bears several advantages over ad-hoc proofs:

1. the ad-hoc proofs are often tedious and error-prone
2. the proof that E is a topology expander is rather short
3. this proof scheme reduces the canonicity issue of Noetherian topologies to the choice of a suitable topology expander.

8.2. Divisibility Topologies

So far, the process of constructing Noetherian spaces has been the following: first build a set of points, then compute a topology that is Noetherian as a least fixed point. In the case where the set of points itself is inductively defined (such as finite words or finite trees), the second step might seem redundant. By automating the second step, we provide a satisfactory answer to the canonicity concerns about Noetherian topologies.

Before studying inductive definition of topological spaces, the notion of least fixed-point in this setting has to be made precise. To that purpose, let us now introduce some basic notions of category theory. In this paper only three categories will appear, the category Set of sets and functions, the category Top of topological spaces and continuous maps, and the category Ord of quasi-ordered spaces and monotone maps. Using this language, a unary constructor G in the algebra of wqos defines an endofunctor from objects of the category Ord to objects of the category Ord preserving well-quasi-orderings.

In our study of Noetherian spaces (resp. well-quasi-orderings), we will often see constructors G' as first building a new set of structures, and then adapting the topology (resp. ordering) to this new set. In categorical terms, we are interested in endofunctors G' that are U-lifts of endofunctors on Set, where U is the forgetful functor from Top (resp. Ord) to Set.

8.2.1. Handling Inductively Defined Datatypes

We are going to work in a framework where the space of interest (words, trees, etc.) is the solution to a least fixed point equation. Recall that both finite trees and finite words can be obtained as least fixed points (see Fact 7.0.1 and Fact 7.0.2).

We are not only interested in the set of points of the least fixed point, but in the topology/order that one can define over this least fixed point so that it is Noetherian/wqo. In the combined proof of Kruskal's tree theorem and Higman's lemma (Theorem 7.1.5), we leveraged a notion of "substructure" ordering (suffixes, structural subtrees). Similarly, in the examples developed in Section 8.1, we defined our refinement functions using upward closures for well-chosen orderings, that allowed us to prove that they were actually topology expanders. Most of this section is devoted to a uniform definition of a "substructure" ordering over inductively defined datatype.

One possible way to design a notion of "substructure" over an inductively defined datatype is to consider type constructors that are analytic functors. These are endofunctors of Set with additional properties that where introduced by Joyal to study "combinatorial classes of structures" in [62]. This notion of analytic functor was then used by Hasegawa to define a divisibility preorder over inductively defined datatypes, and prove that the latter is a well-quasi-order under mild assumptions [58, Theorem 2.10]. We formally introduce the constructions of Hasegawa in the upcoming Subsection 8.2.2. For now, we are interested in leveraging analytic functors to design Noetherian topologies over spaces defined as least fixed points. To that end, we will first define what an analytic functor is.

[62]: Joyal (1986), 'Foncteurs analytiques et espèces de structures'

[58]: Hasegawa (2002), 'Two applications of analytic functors'

Instead of using the classical definition of analytic functors given by Joyal, we will take as definition an equivalent characterisation of analytic functors that was given by Hasegawa in [58].

> **Theorem 8.2.1** [58, Theorem 1.6]**.** An endofunctor $\mathbf{G}$ of Set is an *analytic functor* whenever its category of elements $\mathsf{Elt}(\mathbf{G})$ has the weak normal form property, and for every weak normal form $f \in \mathrm{Hom}((X,x),(Y,y))$ in $\mathsf{Elt}(\mathbf{G})/(Y,y)$, X is a finite set.

Let us now explain the terms employed in Theorem 8.2.1 in the specific setting of the inductive definition of finite words using $\mathrm{CWords}_\Sigma \colon \mathsf{X} \mapsto \mathbf{1} + \Sigma \times \mathsf{X}$ (see Fact 7.0.1).[6] An element w of $\mathrm{CWords}_\Sigma(\mathsf{X})$ is either the unique element $\star$ of $\mathbf{1}$, or a pair (a,u) with $a \in \Sigma$ and $u \in \mathsf{X}$. As a consequence, the same word w is also an element of $\mathrm{CWords}_\Sigma(\{ u \})$, and

6: The map CWords_Σ defines a functor by stating that $\mathrm{CWords}_\Sigma(f)(\star) \overset{\text{def}}{=} \star$ and $\mathrm{CWords}_\Sigma(f)((a,u)) = (a, f(u))$ whenever $f \colon X \to Y$ is a function.

even an element of $\mathrm{CWords}_\Sigma(\emptyset)$ if $w = \star$. The idea of the category of elements and the notion of weak normal form is to extract, for an element $w \in \mathrm{CWords}_\Sigma(\mathsf{X})$, a "minimal" set X' such that $w \in \mathrm{CWords}_\Sigma(\mathsf{X}')$. The objects of interest for us are therefore pairs (X, w) such that $w \in \mathrm{CWords}_\Sigma(\mathsf{X})$, which forms what is known as the category of elements.[7]

7: The definition below only works on Set: we use the notion of membership, and explicitly require that morphisms are maps.

Definition 8.2.2. Given $\mathbf{G}$ an endofunctor of Set, the *category of elements* $\mathsf{Elt}(\mathbf{G})$ has as objects pairs (X, x) with $x \in \mathbf{G}(X)$, and as morphisms between (X, x) and (Y, y) maps $c \colon X \to Y$ such that $\mathbf{G}(c)(x) = y$.

Intuitively, an element (X, x) in $\mathsf{Elt}(\mathbf{G})$ is a witness that x can be produced through $\mathbf{G}$ by using elements of X. Morphisms in the category of elements are witnessing how relations between elements of $\mathbf{G}(X)$ and $\mathbf{G}(Y)$ arise from relations between X and Y.

The "weak normal forms" that are yet to be defined are morphisms f in the *slice category* $\mathsf{Elt}(\mathbf{G})/(Y, y)$, which we define hereafter.

Definition 8.2.3. Given an object A in a category $\mathcal{C}$, the *slice category* $\mathcal{C}/A$ is the category whose objects are elements of $\mathrm{Hom}(B, A)$ when B ranges over objects of $\mathcal{C}$ and morphisms between $c_1 \in \mathrm{Hom}(B_1, A)$ and $c_2 \in \mathrm{Hom}(B_2, A)$ are morphisms $f \in \mathrm{Hom}(B_1, B_2)$ such that $c_2 \circ f = c_1$.

Let us unfold the definition in the specific case of $\mathsf{Elt}(\mathbf{G})$ hereafter to gain intuition about the objects and morphisms in this category.

Fact 8.2.4. Let $\mathbf{G}$ be an endofunctor of Set, and (Y, y) be an object of $\mathsf{Elt}(\mathbf{G})$, that is, $y \in \mathbf{G}(Y)$.

The objects of the slice category $\mathsf{Elt}(\mathbf{G})/(Y, y)$ are morphisms $c \in \mathrm{Hom}((X, x), (Y, y))$, that is, maps $c \colon X \to Y$ such that $\mathbf{G}(c)(x) = y$.

The morphisms f between two objects $c_1 \in \mathrm{Hom}((X, x), (Y, y))$ and $c_2 \in \mathrm{Hom}((X', x'), (Y, y))$ of the slice category $\mathsf{Elt}(\mathbf{G})/(Y, y)$ are maps $f \colon X \to X'$ that satisfy the following equations: $c_1 = c_2 \circ f$, and $\mathbf{G}(f)(x) = x'$.

The slice category is used to abstract away the concrete representation of the elements in $\mathsf{Elt}(\mathbf{G})$. Concretely, for CWords_Σ, an element of the slice category $\mathsf{Elt}(\mathrm{CWords}_\Sigma)/(Y, y)$, where $y \in \mathrm{CWords}_\Sigma(Y)$ is taken to be different from $\star$ (i.e., $y = (a, y')$ for some $a \in \Sigma$ and $y' \in Y$) is given by the following data

- A set X and an element $x \in \mathrm{CWords}_\Sigma(X)$,
- A function $c \colon X \to Y$ such that $\mathrm{CWords}_\Sigma(c)(x) = y = (a, y')$. In particular, we obtain that x must be different from $\star$, and therefore $x = (b, x')$ for some $b \in \Sigma$ and $x' \in X$. From this, we conclude that $\mathrm{CWords}_\Sigma(c)(x) = (b, c(x'))$, thus that $a = b$ and $c(x) = y'$.

From the above reasoning, we conclude that an element of the slice category (in the case of $y \neq \star$) is "essentially" a set X, a point $x' \in X$, and a map $c\colon X \to Y$ such that $c(x') = y'$. Note that there are absolutely no constraints on what c should be doing for elements in $X \setminus \{ x' \}$.

Let us now argue how this abstracts the pair $y = (a, y')$. We can represent it as an element of the slice category as follows: $X \stackrel{\text{def}}{=} \{ y' \}$ and $c\colon y' \mapsto y'$, but also as the following simple set: $X \stackrel{\text{def}}{=} 1$, and $c\colon \star \mapsto y'$. The actual "content" of y' is hidden in the definition of the map c.

The example of CTrees_Σ (associated with finite trees) is better suited to explain the pertinence of this notion than CWords_Σ. An element $y \in \mathrm{CTrees}_\Sigma(Y)$ is a pair $(a, y_1 \cdots y_n)$ with $a \in \Sigma$ and $y_i \in \mathsf{X}$ for every $1 \leq i \leq n$. One can construct the set $Y \stackrel{\text{def}}{=} \{ y_1, \dots, y_n \}$ and notice that $y \in \mathrm{CTrees}_\Sigma(Y)$. To reason about the way that t is constructed, the actual values of $y_1, \dots, y_n$ are not relevant, and we can abstract Y as $X \stackrel{\text{def}}{=} \{ 1, \dots, |Y| \}$. Now, the tree $y = (a, y_1 \cdots y_n)$ can be represented in $\mathrm{CTrees}_\Sigma(X)$ by choosing a bijective map $f\colon X \to Y$, and defining $x \stackrel{\text{def}}{=} (a, f^{-1}(y_1) \cdots f^{-1}(y_n))$. Remark that any choice of bijective function $f\colon X \to Y$ leads to a map from $\mathrm{CTrees}_\Sigma(X)$ to $\mathrm{CTrees}_\Sigma(Y)$, but only one actually maps x_i to y_i. What we are noticing, is that the slice category allows us to retain from the construction of words and trees the "combinatorial information:" what element were actually used, where those elements were used, but without considering the actual values of the elements.

In order to build the "minimal" object in the slice category, we see that in the case of trees we have to reason "up to permutations," which is precisely what the notion of transitive object is made for.

We write $\mathrm{Aut}(X)$ for the collection of automorphisms of X, that is, of maps $f \in \mathrm{Hom}(X, X)$ such that there exists $g \in \mathrm{Hom}(X, X)$ that satisfies $g \circ f = f \circ g = \mathsf{Id}_X$. In the case of Set, these are real bijections.

Definition 8.2.5. A *transitive object* in a category $\mathcal{C}$ is an object X satisfying the following two conditions for every object A of $\mathcal{C}$:

1. the collection $\mathrm{Hom}(X, A)$ in $\mathcal{C}$ is non-empty;
2. the right action of $\mathrm{Aut}(X)$ on $\mathrm{Hom}(X, A)$ by composition is transitive, that is, for all $f, g \in \mathrm{Hom}(X, A)$, there exists a $h \in \mathrm{Aut}(X)$, such that $f \circ h = g$.

The intuition behind Definition 8.2.5 is that the transitive object X represents something that appears inside every object of the category *up to permutations*. Concretely, for every object $A \in \mathcal{C}$, there is a map $f \in \mathrm{Hom}(X, A)$, and for any other map $g \in \mathrm{Hom}(X, A)$, there exists a "permutation" $\sigma \in \mathrm{Aut}(X)$, such that $g \circ \sigma = f$.

To witness the way permutations may appear, the example of CTrees_Σ is not ideal because of the uniqueness property of the morphism $f \in \mathrm{Hom}(X, A)$ (see the paragraph before the definition of a transitive object): the automorphism group is trivial.

Instead, one can work with the finite multiset functor, that maps a set X to the collection of its finite multisets (with the natural definition on morphisms). For this functor, a transitive object in the slice category of an object $(\mathbb{N}, \{\!|\, 1,1,1,3,2 \,|\!\})$ is the morphism

$$c \in \mathrm{Hom}((\{\, a,b,c,d,e \,\}, \{\!|\, a,b,c,d,e \,|\!\}), (\mathbb{N}, \{\!|\, 1,1,1,3,2 \,|\!\}))\,,$$

such that $a \mapsto 1$, $b \mapsto 1$, $c \mapsto 1$, $d \mapsto 3$, $e \mapsto 2$. Indeed, given

$$d \in \mathrm{Hom}((Z, M_Z), (\mathbb{N}, \{\!|\, 1,1,1,3,2 \,|\!\}))\,,$$

we can define $f\colon \{\, a,b,c,d,e \,\} \to Z$ by selecting, inside M_Z the elements that are mapped to 1 (there are exactly 3 of them) and choosing a bijection between those and $\{\, a,b,c \,\}$, selecting the elements that map to 3 (there is exactly one element in M_Z that maps to 3) and map d to this element, and proceeding similarly for 2. While there was a single possibility to define f with respect to d and e (we must find an element that maps to 3 or 2 and belongs to M_Z), we can choose any bijection between $\{\, a,b,c \,\}$ and the three elements of M_Z that map to 1 through d. More precisely, *any* morphism $f' \in \mathrm{Hom}((\{\, a,b,c,d,e \,\}, \{\!|\, a,b,c,d,e \,|\!\}), (Z, M_Z))$ such that $d \circ f = c$ must coincide with f on d and e, and only permutes the image of f over a, b, c, which is precisely the definition of a transitive object.

Definition 8.2.6. A *weak normal form* of an object A in a category $\mathcal{C}$ is a transitive object in $\mathcal{C}/A$.

We now have all the necessary definitions to introduce the last undefined term of Theorem 8.2.1.

Definition 8.2.7. A category $\mathcal{C}$ has the *weak normal form property* whenever every object A has a weak normal form.

Let us now formally prove that CWords_Σ is an analytic functor, which should help the reader to digest the above definitions.

Example 8.2.8. Let Σ be a set. The Set-endofunctor CWords_Σ is an analytic functor.

Proof. Let $(X, w) \in \mathsf{Elt}(\mathrm{CWords}_\Sigma)$. By definition of CWords_Σ, either $w = \star$ (where $\star$ is the unique element of 1) or $w = (a, u)$ with $a \in \Sigma$ and $u \in X$.

In the first case, let us define $\iota \in \mathrm{Hom}((\emptyset, \star), (X, w))$ via $\iota(\star) = \star$. It is a morphism in $\mathsf{Elt}(\mathrm{CWords}_\Sigma)$ because $\mathrm{CWords}_\Sigma(\iota)(\star) = \star$. As a consequence, it is an element of $\mathsf{Elt}(\mathrm{CWords}_\Sigma)/(X, w)$. Notice that $\emptyset$ is finite, and that it suffices for us to prove that ι is a transitive object to conclude. Let $f \in \mathrm{Hom}((Y, v), (X, w))$ be a morphism in $\mathsf{Elt}(\mathrm{CWords}_\Sigma)$. Then $f\colon Y \to X$, and $\mathrm{CWords}_\Sigma(f)(v) = w$. In particular, $\mathrm{CWords}_\Sigma(f)(v) = \star$, hence $v = \star$. Let us define g as the unique map from $\emptyset$ to Y. Then, g satisfies that $\mathrm{CWords}_\Sigma(g)(\star) = \star = v$, and $f \circ g = \iota$. As a consequence, g is a morphism between ι and f in the slice category. Furthermore, any other map g' from ι to f in the slice category corresponds to a map $g'\colon \emptyset \to Y$, hence must be equal to g.

In the second case, let us define $\iota \in \mathrm{Hom}((\{ u \}, (a, u)), (X, w))$ via $\iota(u) \stackrel{\text{def}}{=} u$, which is a morphism because $\mathrm{CWords}_\Sigma(\iota)((a,u)) = (a, \iota(u)) = (a,u) = w$. Notice that $\{ u \}$ is a finite set, and that it suffices to prove that ι is a transitive object in $\mathsf{Elt}(\mathrm{CWords}_\Sigma)/(X, w)$ to conclude. Let $f \in \mathrm{Hom}((Y, v), (X, w))$ be an element of $\mathsf{Elt}(\mathrm{CWords}_\Sigma)/(X, w)$. Then, $\mathrm{CWords}_\Sigma(f)(v) = w = (a, u)$. In particular, $v \neq \star$, and $v = (a, u')$ for some $u' \in Y$. Let us define $g \colon \{ u \} \to Y$ via $g(u) \stackrel{\text{def}}{=} u'$. Then, $\mathrm{CWords}_\Sigma(g)((a,u)) = (a, g(u)) = (a, u') = v$, and g is a morphism in $\mathrm{Hom}((\{ u \}, (a, u)), (Y, v))$. Furthermore, $f \circ g = \iota$, and therefore g is a morphism from ι to f in the slice category. Finally, let g' be another morphism in $\mathrm{Hom}((\{ u \}, (a, u)), (Y, v))$. Then, $\mathrm{CWords}_\Sigma(g')((a,u)) = v = (a, u')$, and this implies that $g'(u) = u'$. In particular, we have proven that $g' = g$. ■

In the above example, we saw that the notion of weak normal form (see Definition 8.2.6) captures the "minimal information needed to build an object" in the specific case of CWords_Σ. Let us illustrate this fact again for a more involved Set-endofunctor. For that, recall that there exists a canonical map $\mathrm{letters} : \Sigma^\star \to \mathcal{P}_{\mathsf{fin}}(\Sigma)$ that maps a word $w \in \Sigma^\star$ to the finite set of its letters.

Example 8.2.9. The Set-endofunctor $\square^\star \colon X \mapsto X^\star$ is an analytic functor. Moreover, a weak normal form of an element (X, w) in the category of elements $\mathsf{Elt}(\square^\star)$ is the morphism

$$\iota \in \mathrm{Hom}((\mathrm{letters}\,(w), w), (X, w)) \quad ,$$

where $\iota(a) \stackrel{\text{def}}{=} a$ for all $a \in \mathrm{letters}\,(w)$.

Example 8.2.10. The Set-endofunctor $\square^{<\omega+1} \colon X \mapsto X^{<\omega+1}$ is not an analytic functor.

Proof. Let $e \stackrel{\text{def}}{=} (\mathbb{N}, (i)_{i \in \mathbb{N}})$ be an element in $\mathsf{Elt}(\square^{<\omega+1})$). Let $f \in \mathrm{Hom}((S, s), e)$ be a weak normal form of e, By definition, $s \in S^{<\omega+1}$, $f(S) \subseteq \mathbb{N}$, and $f^{<\omega+1}(s) = (f(s_i))_{i \in \mathbb{N}} = (i)_{i \in \mathbb{N}}$. This implies that S is infinite, hence $\square^{<\omega+1}$ is not an analytic functor. ■

For finite words and finite trees, the notions of suffix ordering and structural subtree ordering where crucial parts of the minimal bad sequence arguments in Theorems 6.3.24 and 6.3.25. Moreover, they appear crucially when checking that $\mathsf{E}^{\theta}_{\text{words}}$ and $\mathsf{E}^{\theta}_{\text{trees}}$ are topology expanders (see. Corollary 7.2.15 and Lemma 8.1.4).

In order to define this notion of structural embedding in general, let us first notice that analytic functors come with a definition of "support": an element $a \in F(\mathsf{X})$ can be seen as being built using finitely many elements in X. This is the intuitive meaning behind the following formal proposition/definition.

> **Definition 8.2.11.** *Let F be an analytic functor, (X, x) be an element in $\mathsf{Elt}(F)$ and $f \in \mathrm{Hom}((Y, y), (X, x))$ be a weak normal form of (X, x). We define $f(Y)$ as the support of x in X, written $\mathrm{support}_X(x)$.*

$$\begin{tikzcd}
(Y, y) \arrow[d, "h_1"] \arrow[dr, "f"] & \\
(Z, z) \arrow[r, "g"] \arrow[d, "h_2"] & (X, x) \\
(Y, y) \arrow[ur, "f"] &
\end{tikzcd}$$

Figure 8.1.: Since f and g are transitive objects of $\mathsf{Elt}(F)/(X, x)$, there exists functions h_1, h_2 such that the above diagram commutes in $\mathsf{Elt}(F)$.

Proof that the above definition does not depend on Y. Let F be an analytic functor, (X, x) be an element in $\mathsf{Elt}(F)$. Let f be an element of $\mathrm{Hom}((Y, y), (X, x))$, g be an element of $\mathrm{Hom}((Z, z), (X, x))$, and assume that both f and g are weak normal forms of (X, x) in $\mathsf{Elt}(F)$. Let us prove that $f(Z) = g(Y)$.

Since f and g are transitive objects of $\mathsf{Elt}(F)/(X, x)$, there exists functions h_1, h_2 such that the diagram of Figure 8.1 commutes in $\mathsf{Elt}(F)$.

Because f is a transitive object, $\mathrm{Hom}(f, f) = \mathrm{Aut}(f)$ in $\mathsf{Elt}(F)/(X, x)$. As a consequence, the morphism $h_1; h_2$ is an automorphism, and in particular $h_2 \circ h_1$ is a bijection of Y into itself. Finally,

$$f(Y) = f(h_2(h_1(Y))) = g(h_1(Y)) = g(Z) \quad . \qquad \blacksquare$$

We provide the support functions associated with some analytic functors in Figure 8.2, in order to justify the name and intuition about their definition.

$F(X) \overset{\text{def}}{=} \Sigma \times X + 1$

$\mathrm{support}_X((a, b)) = \{ b \}$
$\mathrm{support}_X(1) = \emptyset$

$F(X) \overset{\text{def}}{=} \Sigma \times X^\star$

$\mathrm{support}_X((a, w)) = \mathrm{letters}(w)$

$F(X) \overset{\text{def}}{=} \mathsf{T}(X)$

$\mathrm{support}_X(t) = \mathrm{nodes}(t)$

Figure 8.2.: Examples of analytic functors and their corresponding support functions.

> **Remark 8.2.12.** Let F be an analytic functor, (X, x) be an element of F. Then, $f \colon \mathrm{support}_X(x) \to X$ defined as the identity map is a morphism in $\mathrm{Hom}((\mathrm{support}_X(x), x), (X, x))$ that is a weak normal form of (X, x).

> **Remark 8.2.13.** Let F be an analytic functor, $Y \subseteq X$ be sets, and $x \in F(X)$. Then, $x \in F(Y)$ if and only if $\mathrm{support}_X(x) \subseteq Y$.

Let us now fix F to be an analytic functor from Set to Set, and let us consider an initial algebra $(\mu G, \delta)$ where μF is a set, and $\delta \colon F(\mu F) \to F$ is a function. Such an initial algebra always exists for analytic functors thanks to [58, Lemma 1.26]. We refer the reader to Definition F.2.3 for a formal definition of the notion of initial algebra, and for the rest of the section, we will simply use Lambek's lemma [67] to conclude that δ is an isomorphism in Set, i.e., that δ is a bijection between μF and $F(\mu F)$. Leveraging our notion of support, we define a structural ordering on an initial algebra as follows.

[58]: Hasegawa (2002), 'Two applications of analytic functors'

[67]: Lambek (1968), 'A Fixpoint Theorem for complete Categories.'

> **Definition 8.2.14.** *Let G be an analytic functor and let $(\mu G, \delta)$ be an initial algebra of G. We say that $a \in \mu G$ is a child of $b \in \mu G$ whenever $a = b$ or $a \in \mathrm{support}_{\mu G}(\delta^{-1}(b))$.*
>
> *The transitive closure of the child relation is called the substructure ordering of μG and written $\sqsubseteq$.*

Let us check that this abstract definition actually specialises into the usual suffix ordering and structural subtree ordering when applied to CWords_Σ and CTrees_Σ.

Example 8.2.15. The substructure ordering associated with the initial algebra $(\Sigma^\star, \delta)$ of CWords_Σ is the suffix ordering. Where $\delta(\star)$ is defined as ε, and $\delta((a, u))$ is defined as au.

Proof. Referring to Figure 8.2, we know that $\mathrm{support}_{\Sigma^\star}(\delta^{-1}(\varepsilon)) = \emptyset$, and $\mathrm{support}_{\Sigma^\star}(\delta^{-1}(au)) = \{ u \}$. As a consequence, an easy induction proves that $u \sqsubseteq w$ if and only if $u \leq_{\mathsf{suf}} w$. ■

Example 8.2.16. The substructure ordering associated with the initial algebra $(\mathsf{T}(\Sigma), \delta)$ of CTrees_Σ is the structural subtree ordering. Where $\delta((a, u)) = a(u)$.

Proof. Referring to Figure 8.2, we know that if $t = a(t_1, \ldots, t_n) \in \mathsf{T}(\Sigma)$, then $\mathrm{support}_{\mathsf{T}(\Sigma)}(\delta^{-1}(t)) = \{ t_1, \ldots, t_n \}$. As a consequence, an easy induction proves that $t \sqsubseteq t'$ if and only if $t \leq_{\mathsf{t\text{-}suf}} t'$. ■

We leverage the notion of substructure ordering to define a suitable topology expander over initial algebras of analytic functors.[8] Note that this ordering appears implicitly in the construction of [58, Definition 2.7].

8: Recall that these initial algebras always exist for analytic functors, see [58, Lemma 1.26].

[58]: Hasegawa (2002), 'Two applications of analytic functors'

Definition 8.2.17. *Let $\bar{G} \colon \mathsf{Top} \to \mathsf{Top}$ be a U-lift of an analytic functor G, and $(\mu F, \delta)$ an initial algebra of G.*

Let us define the divisibility expander $\mathsf{E}[\bar{G}]_\diamond(\tau)$ as the topology generated by the sets $\uparrow_{\sqsubseteq}\delta(U)$, where U ranges over the open subsets of $\bar{G}(\mu F, \tau)$ (recall that $\sqsubseteq$ was introduced in Definition 8.2.14).

Using this definition of $\mathsf{E}[\bar{G}]_\diamond$, we can uniformly equip the initial algebras of F with a topology as follows.

Definition 8.2.18. *Let $\bar{G}$ be a U-lift of an analytic functor G. Let $(\mu F, \delta)$ be an initial algebra of G, the divisibility topology over μG with respect to $\bar{G}$ is defined as $\mathsf{lfp}_\tau.\mathsf{E}[\bar{G}]_\diamond(\tau)$.*

Recall that a topological embedding is a way of generalising subspaces with the induced topology by allowing to consider bijections between the carrier sets. Formally, a map h is an embedding when it is a homeomorphism onto its image (see Definition D.3.3). This extra hypothesis easily entails that the divisibility expander is a refinement function.

Lemma 8.2.19. *Let $\bar{G} \colon \mathsf{Top} \to \mathsf{Top}$ be an U-lift of an analytic functor G, and let $(\mu F, \delta)$ be an initial algebra of G. Moreover, we suppose that $\bar{G}$ preserves topological embeddings. Then, the map*

$\mathsf{E}[\bar{G}]_\diamond$ is a refinement function.

Proof. Let us prove that $\mathsf{E}[\bar{G}]_\diamond$ sends Noetherian topologies to Noetherian topologies. Let τ be a Noetherian topology. Let $(\uparrow_\sqsubseteq \delta(U_i))_{i \in \mathbb{N}}$ be a sequence of subbasic open subsets of $\mathsf{E}[\bar{G}]_\diamond(\tau)$. Because $\bar{G}(\mu F, \tau)$ is a Noetherian space, there exists $n_0 \in \mathbb{N}$ such that $\bigcup_{i \leq n_0} U_i \supseteq \bigcup_{i \in \mathbb{N}} U_i$. Now, we conclude that $\bigcup_{i \leq n_0} \delta(U_i) \supseteq \bigcup_{i \in \mathbb{N}} \delta(U_i)$, which implies that the desired inclusion holds: $\bigcup_{i \leq n_0} \uparrow_\sqsubseteq \delta(U_i) \supseteq \bigcup_{i \in \mathbb{N}} \uparrow_\sqsubseteq \delta(U_i)$.

Let us show that $\mathsf{E}[\bar{G}]_\diamond$ is monotone. Consider $\tau \subseteq \tau'$ two topologies on μF. Let us write $X \stackrel{\text{def}}{=} (\mu F, \tau)$ and $Y \stackrel{\text{def}}{=} (\mu F, \tau')$. By definition of the inclusion of topologies, the identity map over μF can be lifted to a continuous function $\iota \colon X \to Y$ in Top. Because $\bar{G}$ is a U-lift, $\mathsf{U}\bar{G} = G\mathsf{U}$, we conclude that $\mathsf{U}(X) = \mathsf{U}(Y) = G(\mu F)$. Moreover, $\mathsf{U}\bar{G}(\iota) = G\mathsf{U}(\iota) = G(\mathsf{Id}_{\mu F}) = \mathsf{Id}_{\mu F}$. As a consequence, $\mathsf{E}[\bar{G}]_\diamond(\tau) \subseteq \mathsf{E}[\bar{G}]_\diamond(\tau')$. ∎

In order to prove that $\mathsf{E}[\bar{G}]_\diamond$ is a topology expander, we will first have to understand how the property of $\bar{G}$ being a functor translates in terms of sub-spaces, and then in terms of subset restrictions. The following fact helps us do this translation.

Fact 8.2.20. Let $\bar{G} \colon \mathsf{Top} \to \mathsf{Top}$ be a U-lift of an analytic functor G, and $(\mu F, \delta)$ an initial algebra of G. Let us consider a Noetherian topology τ such that $\tau \subseteq \mathsf{E}[\bar{G}]_\diamond(\tau)$, and a closed subset H of τ. Then, $\delta^{-1}(H) \subseteq G(H)$.

Proof. Let $t \in H$, because H is downwards closed for $\sqsubseteq$, for every $u \in \mathrm{support}_{\mu F}(\delta^{-1}(t))$, $u \in H$. As a consequence, $\mathrm{support}_{\mu F}(\delta^{-1}(t)) \subseteq H$, and this means that $\delta^{-1}(t) \in G(H)$ because of Remark 8.2.13. ∎

We are now ready to state the main theorem, that the divisibility expander is indeed a topology expander.

Theorem 8.2.21. *Let $\bar{G} \colon \mathsf{Top} \to \mathsf{Top}$ be a U-lift of an analytic functor G, and $(\mu F, \delta)$ an initial algebra of G. Moreover, we suppose that $\bar{G}$ preserves topological embeddings. Then, the map $\mathsf{E}[\bar{G}]_\diamond$ is a topology expander.*

Proof. Let us consider a Noetherian topology τ such that $\tau \subseteq \mathsf{E}[\bar{G}]_\diamond(\tau)$, a closed subset H of τ, and let us prove that

$$\mathsf{E}[\bar{G}]_\diamond(\tau){\downarrow}H \subseteq \mathsf{E}[\bar{G}]_\diamond(\tau{\downarrow}H){\downarrow}H \quad .$$

Let $U = \uparrow_\sqsubseteq \delta(V)$ be a subbasic open subset of $\mathsf{E}[\bar{G}]_\diamond(\tau)$. Recall that H is a closed subset of $\mathsf{E}[\bar{G}]_\diamond(\tau)$ because $\tau \subseteq \mathsf{E}[\bar{G}]_\diamond(\tau)$. Therefore, we can leverage Fact 8.2.20 to derive the following equality

$$\begin{aligned} U \cap H &= (\uparrow_{\sqsubseteq}(\delta(V))) \cap H \\ &= (\uparrow_{\sqsubseteq}(\delta(V) \cap H)) \cap H && \text{because } H = \downarrow_{\sqsubseteq} H \\ &= (\uparrow_{\sqsubseteq}(\delta(V) \cap \delta(G(H)))) \cap H && \text{because } H \subseteq \delta(G(H)) \\ &= (\uparrow_{\sqsubseteq}(\delta(V \cap G(H)))) \cap H && \text{because } \delta \text{ is bijective.} \end{aligned}$$

To conclude that $U \cap H$ is open in $\mathsf{E}[\bar{G}]_{\diamond}(\tau{\downarrow}H){\downarrow}H$ it suffices to show that $V \cap G(H)$ can be rewritten as $W \cap G(H)$ where W is open in $\bar{G}(\mu F, \tau{\downarrow}H)$. To that end, let us consider the two inclusion maps $e_1 \colon (H, \tau_H) \to (\mu F, \tau)$, and $e_2 \colon (H, \tau_H) \to (\mu F, \tau{\downarrow}H)$. These two maps are topological embeddings by definition, hence preserved by $\bar{G}$. As a consequence, $V \cap G(H) = G(e_1)^{-1}(V)$, which is open because $\bar{G}(e_1)$ is a topological embedding and V is an open subset. Because $\bar{G}(e_2)$ is an embedding, there exists an open subset W of $\bar{G}(\mu F, \tau{\downarrow}H)$ such that $G(e_2)^{-1}(W) = V \cap G(H)$. This can be rewritten as $W \cap G(H) = V \cap G(H)$. ■

Because we have built a topology expander, we are now able to associate a Noetherian topology to a wide range of analytic functors.

Corollary 8.2.22. *Let $\bar{G} \colon \mathsf{Top} \to \mathsf{Top}$ be a U-lift of an analytic functor that preserves topological embeddings. Then, for every initial algebra $(\mu G, \delta)$ of G, the divisibility topology over μG with respect to the lift $\bar{G}$ is Noetherian.*

As a sanity check, we apply Theorem 8.2.21 to the sets of finite words and finite trees, and we recover the subword topology and the tree topology that were obtained in an ad-hoc fashion before. In addition to validating the usefulness of Corollary 8.2.22, we believe that these are strong indicators that the topologies introduced prior to this work were the right generalisations of Higman's word embedding and Kruskal's tree embedding in a topological setting, and addresses the canonicity issue of the aforementioned topologies.

Lemma 8.2.23. *The subword topology over $\Sigma^\star$, is the divisibility topology over $\Sigma^\star$ seen as an initial algebra of CWords_Σ.*

Proof. It suffices to remark that the functions $\mathsf{E}[\mathrm{CWords}_\Sigma]_\diamond$ and $\mathsf{E}^{\theta}_{\mathrm{words}}$ have the same least fixed point, and conclude using Lemma 7.2.9. For that, let us compute $\mathsf{E}[\mathrm{CWords}_\Sigma]_\diamond(\tau)$.

We know from Example 8.2.15 that the substructure ordering for words is the suffix ordering. As a consequence, $\mathsf{E}[\mathrm{CWords}_\Sigma]_\diamond(\tau)$ is generated by the sets of the form $\uparrow_{\leq_{\mathsf{suf}}} WV$, where W is an open subset of Σ and $V \in \tau$.

It then follows from an easy (transfinite) induction that open subsets in ${\mathsf{E}[\mathrm{CWords}_\Sigma]_\diamond}^{\alpha}(\tau_{\mathsf{triv}})$ are all upwards closed for the subword ordering.

As a consequence, ${\mathsf{E}[\mathrm{CWords}_\Sigma]_\diamond}^{\alpha}(\tau_{\mathsf{triv}}) = {\mathsf{E}^{\theta}_{\mathrm{words}}}^{\alpha}(\tau_{\mathsf{triv}})$ for all α. ■

Lemma 8.2.24. *The tree topology over $\mathsf{T}(\Sigma)$, is the divisibility topology over $\mathsf{T}(\Sigma)$ seen as an initial algebra of $CTrees_\Sigma$.*

Proof. The proof follows the exact same pattern as Lemma 8.2.23, except that Lemma 8.1.4 already considers the structural subtree ordering hence expanding the definition shows the equality of the two expanders immediately. ∎

8.2.2. A Correctness Result

We have provided a general definition of a divisibility topology on inductively defined datatypes. However, we only checked on two examples (finite words and finite trees) that this definition was sound. Let us strengthen this result by proving in full generality that the divisibility topology correctly generalises the corresponding notions on quasi-orderings. In the case of finite words, this translates to the equation $\mathsf{Alex}(\leq)^{\star,\mathsf{t}} = \mathsf{Alex}((\leq)_{\mathsf{w}})$ [45, Exercise 9.7.30]. We do this by relating the divisibility topology to the divisibility preorder introduced by [58, Definition 2.7].

[45]: Goubault-Larrecq (2013), *Non-Hausdorff Topology and Domain Theory*

[58]: Hasegawa (2002), 'Two applications of analytic functors'

Let us quickly go through the definition of the divisibility preorder given by Hasegawa in [58, Section 2.2]. Let $\bar{G}\colon \mathsf{Ord} \to \mathsf{Ord}$ be a U-lift of an analytic functor G that respects embeddings and well-quasi-orderings, let us build a family A_i of quasi-orders and $e_i\colon A_i \to A_{i+1}$ of embeddings as follows:

- $A_0 = \emptyset$, $A_1 = \bar{G}(A_0)$ and e_0 is the empty map.
- $e_{n+1} = \bar{G}(e_n)$ and A_{n+1} has as carrier set $G(A_n)$ and as preordering $\preceq_{n+1}$ the transitive closure of the union of the two following relations:
  - the quasi-order $\leq_{\bar{G}(A_n)}$,
  - the collection of $b \lhd_{n+1} a$ for each weak normal form $f \in \mathrm{Hom}((X,z),(A_n,a))$ in $\mathsf{Elt}(G)$ and each $b \in \mathrm{Im}(e_n \circ f)$.

The *divisibility ordering* $\preceq$ is the ω-inductive limit in the category Ord of the diagram $A_0 \to^{e_0} A_1 \to^{e_1} \cdots$. As remarked by Hasegawa, the maps e_n are injective order embeddings, and so are the morphisms $c_n\colon A_n \to \mu F$ of the colimiting cone [58, Lemma 2.8]. Without loss of generality, we can assume that $A_0 \subseteq A_1 \ldots$ and that the colimit $(\mu F, \delta)$ is the union of the sets A_i for $0 \leq i < \omega$. In particular, δ is the identity map in this setting.

In order to illustrate this definition, let us take as a concrete example the case of finite words over an (ordered) alphabet $(\Sigma, \leq)$, and see that we indeed recover the subword ordering. In this case, the functor $\bar{G}$ is CWords_Σ equipped with the natural product and sum orderings. We have $A_0 = \emptyset$ by definition. Then, $A_1 = \mathrm{CWords}_\Sigma(A_0) = \Sigma \times \emptyset + 1$, i.e., $A_1 = 1$. Moreover, $e_0\colon A_0 \to A_1$ is the empty map. Because $A_{n+1} = \mathrm{CWords}_\Sigma(A_n)$, an easy induction proves that A_{n+1} is the set of finite words over Σ length at most n. Moreover, a similar induction proves that the function e_n is the identity sending words of length at most n to words of length at most $n+1$. Let us now describe the orderings of the sets $(A_n)_{n\in\mathbb{N}}$, and in particular prove that A_{n+1} is equipped with the subword ordering restricted to words of length at most n. Let us decompile the two relations that make the ordering over A_{n+1}:

- $u, v \in \mathrm{CWords}_\Sigma(A_n)$ are related for $\mathrm{CWords}_\Sigma(A_n)$ if and only if $u = v = 1$ or $u = (a, u')$, $v = (b, v')$, $a \leq b$ and $u' \leq v'$ in A_n. In terms of finite words and using the induction hypothesis that A_n is equipped with the subword ordering, this translates to $u = v = \varepsilon$ or $u = au'$, $v = bv'$, $a \leq b$ and $u' \leq_{\mathsf{w}} v'$.
- $u, v \in \mathrm{CWords}_\Sigma(A_n)$ satisfy $u \lhd_{n+1} v$ if and only if there exists a weak normal form $f \in \mathrm{Hom}((X, x), (A_n, v))$ such that $u \in \mathrm{Im}(e_n \circ f)$. In particular, $u = e_n(f(c))$ for some $c \in X$, and by definition of the support, $u = e_n(v')$ for some $v' \in \mathrm{support}_{A_n}(v)$. Thanks to the computations of Figure 8.2, this last equation is rewritten as $v = bv'$. As a consequence, $u = e_n(v') = v'$, since e_n is the identity function.

As a consequence of this case analysis, the preorder $\preceq_{n+1}$ over A_{n+1} is the transitive closure of the following two relations: the first relation compares the first letters pointwise, and the suffixes using subword ordering, and the second relation is the suffix ordering. We easily recognise the (recursive) description of the subword ordering given in Figure 7.2.

Hopefully, one has gained a some intuition about Hasegawa's construction, and in particular the following lemma relating the construction of the ordering on A_{n+1} to the notion of support of the functor G should not be surprising. Because the proof is mainly "symbol pushing", it can safely be skipped as all the intuitions already have been introduced prior to the lemma.

Lemma 8.2.25. *For all $n \in \mathbb{N}$, for all $a, b \in A_{n+1}$, $b \lhd_{n+1} a$ if and only if $b \in \mathrm{support}_{\mu F}(a)$.*

Proof $\Rightarrow$. Let $n \in \mathbb{N}$ and $a, b \in A_{n+1}$ such that $b \lhd_{n+1} a$. There exists a weak normal form $f \in \mathrm{Hom}((X, x), (A_n, a))$ such that $b \in \mathrm{Im}(e_n \circ f)$. Notice that $c_n \circ f \in \mathrm{Hom}((X, x), (\mu F, a))$ remains a weak normal form thanks to [58, Lemma 1.5]. As a consequence, $c_{n+1}(b) \in \mathrm{Im}(c_{n+1} \circ e_n \circ f)$. However, $c_{n+1} \circ e_n = c_n$ is the inclusion map from A_n to μF. Therefore, $b \in \mathrm{Im}(c_n \circ f) = \mathrm{support}_{\mu F}(a)$. ■

[58]: Hasegawa (2002), 'Two applications of analytic functors'

Proof $\Leftarrow$. Assume that $b \in \mathrm{support}_{\mu F}(a)$. By definition of the support, there exists a weak normal form $f \in \mathrm{Hom}((X, x), (\mu F, a))$ such that $b \in \mathrm{Im}(f)$. Remark that the inclusion $c_n \colon A_n \to \mu F$ is in fact an object $c_n \in \mathrm{Hom}((A_n, a), (\mu F, a))$ in the slice category $\mathsf{Elt}(G)/(\mu F, a)$. Because f is a transitive object, there exists a morphism g from f to c_n in this category, which in particular gives a $g \colon X \to A_n$ such that $c_n \circ g = f$. However, this proves that $b \in (c_n \circ g)(X)$, hence $b \in c_n(A_n) = A_n = e_n(A_n)$.

To conclude, it suffices to notice that $g \in \mathrm{Hom}((X, x), (A_n, a))$ is a weak normal form, but this is a direct consequence of the fact that f was one. ■

Now that we have a better understanding of the relation $\triangleleft$ in terms of support, it is not difficult to mimic the analysis done in the case of finite words, that relates $\preceq$ to the structural ordering $\sqsubseteq$.

Corollary 8.2.26. *For every $n \in \mathbb{N}$, A_n is a downwards closed subset of μF for $\sqsubseteq$.*

Corollary 8.2.27. *For every $n \in \mathbb{N}$, $\preceq_{n+1} = (\sqsubseteq \leq_{\bar{G}(A_n)})^*$.*

Corollary 8.2.28. *The preorder $\preceq$ equals $(\sqsubseteq\preceq)^*$, i.e., $\sqsubseteq \subseteq \preceq$.*

Analysing Corollaries 8.2.27 and 8.2.28, one finds that they are looking similar to the definition of the the divisibility topology expander of a functor. One main difference is that for the divisibility expander, one considers the upward closure of sets, where the above corollaries only talk about transitive closures of relations: i.e, the order of composition is important for the topology expander. It turns out that getting this stratification is not particualry difficult for the divisibility ordering too.

Lemma 8.2.29. *For all $n \in \mathbb{N}$, the following equalities hold over A_{n+1}*

$$\preceq_{n+1}\sqsubset \quad = \quad (\sqsubseteq \leq_{\bar{G}(A_n)})^* \sqsubset \quad = \quad \leq_{\bar{G}(A_n)}\sqsubset$$

Proof. Let $n \in \mathbb{N}$. Thanks to Corollary 8.2.27, it is clear that the first equality holds:

$$\preceq_{n+1}\sqsubset \quad = \quad (\sqsubseteq \leq_{\bar{G}(A_n)})^* \sqsubset \quad .$$

For the second equality, the following inclusion is trivial:

$$(\sqsubseteq \leq_{\bar{G}(A_n)})^* \sqsubset \quad \supseteq \quad \leq_{\bar{G}(A_n)}\sqsubset \quad .$$

Let us now prove by induction over $k \in \mathbb{N}$ that the following holds:

$$\forall k \in \mathbb{N}, (\leq_{\bar{G}(A_n)}\sqsubseteq)^k \sqsubset \quad \subseteq \quad \leq_{\bar{G}(A_n)}\sqsubset \tag{IH}$$

For $k = 0$, Equation IH follows from the reflexivity of $\leq_{\bar{G}(A_n)}$. Let us now consider the case $k+1$. For that, let us unpack the definitions and leverage the induction hypothesis to extract the core argument of the proof.

$$\begin{aligned}(\leq_{\bar{G}(A_n)}\sqsubseteq)^{k+1} \sqsubset &\subseteq (\leq_{\bar{G}(A_n)}\sqsubseteq) \leq_{\bar{G}(A_n)}\sqsubset \\ &= (\leq_{\bar{G}(A_n)}\sqsubset\leq_{\bar{G}(A_n)}\sqsubset) \cup (\leq_{\bar{G}(A_n)}\sqsubset)\end{aligned}$$

Now, let $a, b \in A_{n+1}$ be such that $a \leq_{\bar{G}(A_n)}\sqsubset\leq_{\bar{G}(A_n)}\sqsubset b$. There exists $c, d, e \in A_{n+1}$ such that $a \leq_{\bar{G}(A_n)} c \sqsubset d \leq_{\bar{G}(A_n)} e \sqsubset b$. As a consequence of $c \sqsubset d$ and $e \sqsubset b$, $c \in A_n$ and $e \in A_n$. Because $e_n \colon A_n \to A_{n+1}$ is an embedding, $c \preceq_{n+1} e$ implies $c \preceq_n e$. Moreover, e_n is a monotone map from $(A_n, \preceq_n)$ to $(A_{n+1}, \leq_{\bar{G}(A_n)})$ by construction. As a consequence, $c \leq_{\bar{G}(A_n)} e$. We have proven that $a \leq_{\bar{G}(A_n)} \leq_{\bar{G}(A_n)} \sqsubset b$, i.e, that $a \leq_{\bar{G}(A_n)} \sqsubset b$. ■

Corollary 8.2.30. *For every $n \in \mathbb{N}$, $\preceq_{n+1} = \leq_{\bar{G}(A_n)} \sqsubseteq$.*

Having gained a finer understanding of Hasegawa's construction of the divisibility preorder, we can now provide an inductive definition of its upwards closed subsets. This is the final step in relating the divisibility topology to the divisibility preorder, as the upwards closed subsets are exactly the open subsets in an Alexandroff topology.

Lemma 8.2.31. *For every $E \subseteq \mu G$, $\uparrow_{\preceq} E = \uparrow_{\bar{G}(\preceq) \sqsubseteq} E$.*

Proof. It suffices to notice that $\preceq = \bar{G}(\preceq) \sqsubseteq$. For that, recall that $x \preceq y$ if and only if there exists $n \in \mathbb{N}$ such that $x, y \in A_{n+1}$ and $x \preceq_{n+1} y$. Leveraging Corollary 8.2.30, we conclude that $x \preceq y$ if and only if there exists $n \in \mathbb{N}$ such that $x, y \in A_{n+1}$ and $x \leq_{\bar{G}(A_n)} \sqsubseteq y$. The latter is equivalent over A_{n+1} to $x \leq_{\bar{G}(\preceq)} \sqsubseteq y$ and we have concluded. ■

Lemma 8.2.32. *For all $E \subseteq \mu G$, $\uparrow_{\sqsubseteq} \uparrow_{\bar{G}(\preceq\downarrow_{A_n})} E = \uparrow_{\preceq\downarrow_{A_{n+1}}} E$.*

Proof. It suffices to notice that $\bar{G}(\preceq\downarrow_{A_n}) \sqsubseteq = \preceq\downarrow_{A_{n+1}}$.

Because $\bar{G}$ preserves order embeddings, $\bar{G}(\preceq) = \bar{G}(\preceq\downarrow_{A_n})$ over A_{n+1}.

Recall that $x \preceq y$ if and only if there exists $n \in \mathbb{N}$ such that $x, y \in A_{n+1}$ and $x \preceq_{n+1} y$. Leveraging Corollary 8.2.30, we conclude that $x \preceq y$ if and only if there exists $n \in \mathbb{N}$ such that $x, y \in A_{n+1}$ and $x \leq_{\bar{G}(A_n)} \sqsubseteq y$. Which thanks to the above remark is equivalent to and $x \leq_{\bar{G}(\preceq\downarrow_{A_n})} \sqsubseteq y$. ■

We now have all the tools to prove the main theorem of this section, namely that the divisibility preorder is the trace of the divisibility topology in the realm of wqos. We are now ready to state our correctness theorem, i.e., that the divisibility topology is a correct generalisation to the topological setting of the divisibility preorder from Hasegawa.

Theorem 8.2.33. *Let $\bar{G}$ be the lift of an analytic functor respecting Alexandroff topologies, Noetherian spaces, and embeddings. Then, the divisibility topology of μG is the Alexandroff topology of the divisibility preorder of μG, which is a well-quasi-ordering.*

Proof of the well-quasi-ordering. Assuming that the two topologies coincide, it suffices to notice that the specialisation preorder of a Noetherian Alexandroff topology is a well-quasi-ordering. ■

Proof of the first inclusion $\subseteq$. It suffices to prove that $\mathsf{Alex}(\preceq)$ is a post-fixed point of $\mathsf{E}[\bar{G}]_\diamond$.

Let us consider a subbasic open subset $U \in \mathsf{E}[\bar{G}]_\diamond(\mathsf{Alex}(\preceq))$. It is of the form $U = \uparrow_\sqsubseteq \delta(V)$ for some V open subset of $\bar{G}(\mu G, \mathsf{Alex}(\preceq))$. Recall that δ is the identity function. Moreover, V is an upwards closed subset of $\bar{G}(\mu G, \mathsf{Alex}(\preceq))$ in its specialisation preorder $\bar{G}(\preceq)$. We can leverage Lemma 8.2.31 to conclude:

$$\begin{aligned}\uparrow_\sqsubseteq V &= \uparrow_\sqsubseteq \uparrow_{\bar{G}(\preceq)} V \\ &= \uparrow_{\bar{G}(\preceq)\sqsubseteq} V \\ &= \uparrow_\preceq V \in \mathsf{Alex}(\preceq) \quad .\end{aligned}$$

We have proven that $U \in \mathsf{Alex}(\preceq)$. ■

Proof of the second inclusion $\supseteq$. It suffices to prove that, for all $n \in \mathbb{N}$, $\mathsf{E}[\bar{G}]_\diamond(\mathsf{Alex}(\preceq)\downarrow A_n) = \mathsf{Alex}(\preceq)\downarrow A_{n+1}$.

For the first inclusion, let U be a subbasic open subset of the topology $\mathsf{E}[\bar{G}]_\diamond(\mathsf{Alex}(\preceq)\downarrow A_n)$. It is of the form $U = \uparrow_\sqsubseteq \delta(V)$ for some V open subset of the topology $\bar{G}(\mu G, \mathsf{Alex}(\preceq)\downarrow A_n)$. Because δ is assumed to be the identity map, and $\bar{G}(\preceq\downarrow_{A_n})$ is the specialisation preorder of the resulting space, $V = \delta(V) = \uparrow_{\bar{G}(\preceq\downarrow_{A_n})} V$. We can leverage Lemma 8.2.32 to conclude:

$$\begin{aligned}\uparrow_\sqsubseteq V &= \uparrow_\sqsubseteq \uparrow_{\bar{G}(\preceq\downarrow_{A_n})} V \\ &= \uparrow_{\bar{G}(\preceq\downarrow_{A_n})\sqsubseteq} V \\ &= \uparrow_{\preceq\downarrow A_{n+1}} V \quad .\end{aligned}$$

Conversely, let $U = \uparrow_{\preceq\downarrow A_{n+1}} E$, then $U = \uparrow_\sqsubseteq \uparrow_{\bar{G}(\preceq\downarrow_{A_n})} E$, and $V \stackrel{\text{def}}{=} \uparrow_{\bar{G}(\preceq\downarrow_{A_n})} E$ is an open subset of $\bar{G}(\mu G, \mathsf{Alex}(\preceq)\downarrow A_n)$ because the latter is an Alexandroff topology with specialisation preorder $\preceq\downarrow_{A_n}$. ■

8.3. Discussion

Variations on Trees. When considering trees, we have implicitly restricted our attention to unranked, ordered trees of finite height. This was natural because both the Kruskal tree theorem and topological Kruskal theorem are formulated on this class of trees. However, it might be interesting to play with other definitions of trees, and their corresponding notions of "tree-embedding" as it was done in [52]. In particular, we hope to obtain the topological analogues of their results stated in terms of well-quasi-orders.

[52]: Goubault-Larrecq and Schmitz (2016), 'Deciding Piecewise Testable Separability for Regular Tree Languages'

The Divisibility Expander. There is an interesting notion behind the construction of a divisibility expander, namely the notion of a substructure ordering that accompanies inductive definitions made using analytic functors. The intervention of the substructure ordering somehow explains why the topologies placed on finite words and finite trees had the "shape" of finite words and finite trees.

Another Correctness Theorem. As mentioned in the introduction of Chapter 7 (Topology expanders and Noetherian Topologies), there is another way to inductively construct well-quasi-orderings. Based on the ideas of total order dilators introduced by [42], Freund [36] introduced a categorical construction (not unlike the one of Hasegawa) to study the gap embedding relation between finite trees. Computations are done inside the category PO of partial orders and quasi-embeddings. Instead of considering an endofunctor of PO that is a U-lift of an analytic functor, Freund consider partial order dilators defined as follows.

[42]: Girard (1981), 'Π12-logic, Part 1: Dilators'

[36]: Freund (2020), 'From Kruskal's theorem to Friedman's gap condition'

Definition 8.3.1 [36, Definition 2.1]**.** A *partial order dilator* (or *PO-dilator*) consists of

- a functor $F\colon \mathsf{PO} \to \mathsf{PO}$ that preserves embeddings and
- a natural transformation $\mathrm{supp}^F : F \Rightarrow \mathcal{P}_{\mathsf{fin}}$ satisfying the following condition: given any embedding $f\colon X \to Y$ of partial orders, the embedding $F(f)\colon F(X) \to F(Y)$ satisfies $\mathrm{Im}(F(f)) = \{ b \in F(Y)\colon \mathrm{supp}^F_Y(b) \subseteq \mathrm{Im}(f) \}$.

Whenever F is a PO-dilator that preserves well-quasi-orderings, we call it a *WPO-dilator*. A PO-dilator is called *normal* when for all $(X, \leq)$, and $a, b \in F(X)$, $a\ F(\leq)\ b$ implies $\mathrm{supp}^F_X(a) \leq_{\mathcal{H}} \mathrm{supp}^F_X(b)$.

We believe that an analogue of Theorem 8.2.33 holds for WPO-dilators, because the key notion of support is already baked in the definition. Proving such a result however would involve defining a suitable notion of "topology dilator", and extract the correct notion of "divisibility preorder" from [36, Definition 2.4]. Under these assumptions, we conjecture the following (second) correctness result.

Conjecture 8.3.2. For every topology dilator that respects Alexandroff topologies, Noetherian spaces, and embeddings. The divisibility topology over its initial algebra is the Alexandroff topology of Freund's divisibility preorder, which is a well-quasi-ordering.

References

[36] Anton Freund. 'From Kruskal's theorem to Friedman's gap condition'. In: *Mathematical Structures in Computer Science* 30 (2020), pp. 952–975. DOI: 10.1017/s0960129520000298 (cit. on pp. 179, 184, 229).

[42] Jean-Yves Girard. 'Π12-logic, Part 1: Dilators'. In: *Annals of Mathematical Logic* 21 (1981), pp. 75–219. DOI: 10.1016/0003-4843(81)90016-4 (cit. on p. 229).

[45] Jean Goubault-Larrecq. *Non-Hausdorff Topology and Domain Theory*. Vol. 22. Cambridge University Press, 2013. DOI: 10.1017/cbo9781139524438 (cit. on pp. 3, 50, 52–54, 139, 159, 178, 179, 185, 186, 192, 193, 203, 208, 224, 239, 254, 255).

[47] Jean Goubault-Larrecq. 'Infinitary Noetherian Constructions I. Infinite Words'. In: *Colloquium Mathematicum* 168 (2022), pp. 257–268. DOI: https://doi.org/10.4064/cm8077-4-2021 (cit. on pp. 139, 175, 178, 179, 209–211).

[49] Jean Goubault-Larrecq, Simon Halfon and Aliaume Lopez. *Infinitary Noetherian Constructions II. Transfinite Words and the Regular Subword Topology*. 2022 (cit. on pp. 175, 178, 179, 209, 211).

[52] Jean Goubault-Larrecq and Sylvain Schmitz. 'Deciding Piecewise Testable Separability for Regular Tree Languages'. In: *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*. Ed. by Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani and Davide Sangiorgi. Vol. 55. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016, 97:1–97:15. DOI: 10.4230/LIPIcs.ICALP.2016.97. URL: http://drops.dagstuhl.de/opus/volltexte/2016/6232 (visited on 02/02/2023) (cit. on pp. 3, 228).

[58] Ryu Hasegawa. 'Two applications of analytic functors'. In: *Theoretical Computer Science* 272 (2002), pp. 113–175. DOI: 10.1016/S0304-3975(00)00349-2. URL: https://www.sciencedirect.com/science/article/pii/S0304397500003492 (cit. on pp. 179, 184, 207, 215, 220, 221, 224, 225, 227, 229).

[62] André Joyal. 'Foncteurs analytiques et espèces de structures'. In: *Combinatoire énumérative*. Ed. by Gilbert Labelle and Pierre Leroux. Springer Berlin Heidelberg, 1986, pp. 126–159 (cit. on pp. 215, 239).

[67] Joachim Lambek. 'A Fixpoint Theorem for complete Categories.' In: *Mathematische Zeitschrift* 103 (1968), pp. 151–161. URL: http://eudml.org/doc/170906 (cit. on p. 220).

Part III.

Outlook

Concluding Remarks 9.

Outline of the chapter

This chapter is the last one of this thesis.

Meta Reading Notes. It is quite hard to write a self-contained document that maintains a coherence in the notations, motivations, and still lists widely different results. Even at this stage, it can be argued that Chapter 7 (Topology expanders and Noetherian Topologies) and Chapter 8 (Inductive Constructions) are not entirely fitting the big picture of studying preservation theorems. This explains the need for a conjunction in the title of this manuscript.

Reflecting on Locality. There is much to talk about with respect to Chapter 4 (Locality and Preservation) and Chapter 5 (A Local-to-Global Preservation Theorem). The main technical tool that we used in these two sections is the Gaifman locality theorem of first order logic. In view of the proof of Theorem 4.2.2, and the problems arising when using Theorem 5.1.2 on classes that are not sparse, it seems that this notion of locality is too coarse to faithfully represent the semantic properties of FO.

This is not a novel remark, and the notion of "differential games" [38], "twin-width" [9], and the newly introduced "flipper games" [30, 39], all try to account for the intrinsic combinatorial complexity of neighbourhoods. It is possible that the (recurring) proof technique that uses the notion of covering points and generic points can be adapted to this setting of "neighbourhood similarity", leading to better local normal forms and/or a strengthening of the local-to-global technique of Theorem 5.1.2.

[38]: Gajarsky, Gorsky and Kreutzer (2022), 'Differential Games, Locality, and Model Checking for FO Logic of Graphs'

[9]: Bonnet, Kim, Thomassé and Watrigant (2020), 'Twin-width I: tractable FO model checking'

[30]: Dreier, Mählmann, Siebertz and Toruńczyk (2022), *Indiscernibles and Wideness in Monadically Stable and Monadically NIP Classes*

[39]: Gajarský, Mählmann, McCarty, Ohlmann, Pilipczuk, Przybyszewski, Siebertz, Sokołowski and Toruńczyk (2023), *Flipper games for monadically stable graph classes*

Reflecting on Noetherian Constructions. There are plenty of open questions and conjectures in Chapter 7 (Topology expanders and Noetherian Topologies) and in Chapter 8 (Inductive Constructions). However, there is a deeper issue that can only be seen when considering this document as whole: the space of finite words (resp finite trees, infinite words, infinite trees) is presented in two different ways. The first one is as coloured structures,[1] and the second one is as inductively defined datatypes in Chapter 8 (Inductive Constructions).

1: That is, as wreath products in Chapter 6 (Logically Presented Spaces)

This opens an interesting question regarding the precise relationship between the wreath topology and the divisibility topology, that generalises a conjecture due to Pouzet regarding the labelling of well-quasi-orders.

Conjecture 9.0.1 (Labelling of Noetherian Spaces)**.** Let σ be a finite relational signature without unary predicates, $\mathcal{C}$ be a class of finite structures over σ, F be a fragment of $\mathsf{FO}[\sigma]$, such that $(\mathcal{C}, \leq_{\mathsf{F}})$

is ∞-wqo.

Then, for every Noetherian space (X, τ), $\mathcal{C} \rtimes \mathsf{X}$ with the wreath topology associated to F is Noetherian.

In the study logically presented pre-spectral spaces that was performed in Chapter 6, the notion of definable subset allowed us to prove a variation of the above conjecture in Theorem 6.3.44.

Reflecting on Backward Algorithms. We have spent a lot of time proving or disproving the relativisation of preservation theorems to classes of structures. While this is a sound question regarding the completeness of some logical fragment, a more pragmatic approach would ask: how can we leverage the compactness to build algorithms? Recall that lpps were introduced as a generalisation of wqos and Noetherian spaces, both of which have interesting applications in the verification of transition systems. One hope would be to leverage lpps to verify "data-centric transitions systems", that is, dynamical systems obtained using first order interpretations (see for instance the presentation of [8]).

[8]: Bojańczyk, Segoufin and Toruńczyk (2013), 'Verification of Database-Driven Systems via Amalgamation'

Unfortunately, it is not possible to use the theory of lpps to obtain an analogue of the "backward algorithm" that powers the decision procedures over well-structured transition systems and well-structured topological transition systems. The main issue comes from the fact that a generic backwards algorithm would build an increasing sequence of definable open subsets, and the termination of such algorithms (in the case of Noetherian spaces) comes from the compactness of the union[2] of the sets in this sequence. In the case of lpps, an arbitrary union of definable subsets might not be definable, and therefore one cannot conclude that the union is compact. That is, the generic backward algorithm may not terminate.

2: The union of an arbitrary sequence of open subsets is an open subset

As a concrete example, let us consider the set $\mathcal{C} \overset{\text{def}}{=} \Delta^2_{\text{deg}}$ of *directed* graphs of degree bounded by 2. We know that $\langle\!\langle \mathcal{C}, \mathsf{Alex}(\subseteq_i), \mathsf{FO} \rangle\!\rangle$ is a lpps. Now, let us produce the following FO-interpretation I from $\mathcal{C}$ to $\mathcal{C}$: $\varphi_{\mathsf{dom}}(x) \overset{\text{def}}{=} \top$, and $\varphi_E(x,y) \overset{\text{def}}{=} \exists z.E(x,z) \wedge E(z,y)$. It is quite clear that $\mathsf{I}\colon \mathcal{C} \to \mathcal{C}$ is a logical map that defines a deterministic transition system. Let us now consider a target set S of all graphs that contain a cycle $E(x,y) \wedge E(y,x)$. It is clear that S is an definable open subset of $\mathcal{C}$. The backward algorithm constructs the increasing sequence $\mathsf{I}^{-i}(S)$ for $i \geq 0$, but here, the union is precisely the set of all directed graphs that have a strongly connected component, which is neither definable (due to locality) nor compact.

Reflecting on the Chase. Instead of importing the backward algorithm from well-structured transition systems to database driven systems, it may be interesting to recast the well-known Chase algorithm in a topological setting. Using a slight reformulation of the analysis performed by [25, Definition 6], a Chase algorithm answers the following problem: given as input a set Σ of first-order constraints[3] together with an input database $\mathfrak{A} \in \mathcal{C}$, output a sentence $\psi \in \exists\mathsf{PQF}$ such that $[\![\psi]\!]_{\mathcal{C}} = \uparrow_{\preceq_h}([\![\Sigma]\!]_{\mathcal{C}} \cap \uparrow_{\preceq_h}\mathfrak{A})$. Usually, the algorithm will incrementally

[25]: Deutsch, Nash and Remmel (2008), 'The chase revisited'

3: usually in $\forall\exists\mathsf{QF}$

build ψ by "chasing" the constraints in Σ, but this is not relevant to this high-level description.

As noticed by [25, Theorem 17], the Homomorphism Preservation Theorem implies that one can characterise when a specific version of the Chase algorithm, the "core chase," terminates: the algorithm terminates with input $(\Sigma, \mathfrak{A})$ if and only if the output $\uparrow_{\preceq_h}(\llbracket\Sigma\rrbracket \cap \uparrow_{\preceq_h}\mathfrak{A})$ is first-order definable. We do not have a complete explanation for this result, but it seems that the "core chase step" given a set Σ of constraints is a logical map $\mathsf{I}_\Sigma : \mathcal{C} \to \mathcal{C}$ [25, Definition 4].

Reflecting on Regularity Conditions. In order to get closer to algorithmic applications of lpps, it might be necessary to not only ask for "effective representations", but also to strengthen the grasp that is given over $\mathcal{K}^\circ(\mathsf{X})$. One way to do so is to strengthen the condition $\tau \cap \mathcal{B} \subseteq \mathcal{K}^\circ(\mathsf{X})$ into the following one: for all definable subset $D \in \mathcal{B}$, its interior $\mathring{D}$ is compact.

This leads to the following definition: A lps $\langle\langle \mathsf{X}, \tau, \mathcal{B}\rangle\rangle$ is *interior compact* if and only if for every $D \in \mathcal{B}$, $\mathring{D} \in \mathcal{K}^\circ(\mathsf{X})$. It is an easy check that every interior compact lps is an lpps.

One reason to study these spaces is that the main result from [83, Theorem 5.15] can be rephrased as the fact that $\langle\langle \mathsf{Fin}(\sigma), \langle \mathsf{EPFO}\rangle_{\mathsf{topo}}, \mathsf{FO}\rangle\rangle$ is an interior compact lps. However, it is unclear whether the ability to consider interiors of definable subsets allows us to build proper algorithms.

Reflecting on Free Variables. In the introduction of this thesis, we introduced first order logic as a way to build queries for databases. However, most of the technical developments were considering first order sentences, because those define actual *subsets* which is meaningful from a topological point of view.

The problem of lifting results over sentences to formulas with free variables is not new and we know at least two way this can be done. The first one works mostly for homomorphisms and is the notion of "plebeian companion" that one can find in [19, 83]. Overall, this method encodes the query parameters as partially evaluated relations by extending the vocabulary.

[19]: Daligault, Rao and Thomassé (2010), 'Well-Quasi-Order of Relabel Functions'
[83]: Rossman (2008), 'Homomorphism preservation theorems'

Another, maybe more categorical way to add free variables to lpps would be to consider pointed structures, which are represented by pairing finite domain valuations to structures. Over this extended notion of structures, one can interpret *formulas* as subsets, and therefore apply the topological tools that were developed in this manuscript. However, once the space accounts for free variables, one must be careful with renaming and the domains of the valuations. In particular, the notion of F-embedding or even the semantics of a formula are not completely straightforward to define.

One application of these (involved) definition could be to recast the Chang-Łoś-Suszko Theorem in a simpler framework. Recall that this theorem investigates the semantic properties of $\exists\forall\mathsf{QF}$ sentences. By "removing one layer of quantification," we obtain $\forall\mathsf{QF}$ *formulas*, that we have to evaluate over pointed structures. This seems to be another formulation of the combinatorial approach introduced in [85, Chapter 11]. In particular, we would be able to reformulate sufficient conditions based on "coloured structures being wqos" [85, Theorem 11.2.2] together with structural stability properties [85, Chapter 10 Section 10.2], by leveraging the technical developments done in Chapter 6 (Logically Presented Spaces).

Part IV.

Cheat-Sheets

You are now entering the appendix containing various cheatsheets. Most of the results and definitions are folklore, and in the worst case scenario, can be found in one of the following list of "go-to" documents [3, 18, 23, 26, 45, 62, 68, 73].

[3]: Adámek, Herrlich and Strecker (1990), *Abstract and Concrete Categories : The Joy of Cats*
[18]: Community (2008), *The nLab wiki*
[23]: Demeri, Finkel, Goubault-Larrecq, Schmitz and Schnoebelen (2012), 'Algorithmic Aspects of WQO Theory (MPRI course)'
[26]: Dickmann, Schwartz and Tressl (2019), *Spectral Spaces*
[45]: Goubault-Larrecq (2013), *Non-Hausdorff Topology and Domain Theory*
[62]: Joyal (1986), 'Foncteurs analytiques et espèces de structures'
[68]: Libkin (2012), *Elements of finite model theory*
[73]: Mac Lane (1978), *Categories for the Working Mathematician*

The goal is not to provide a Wikipedia-like list of definitions, but rather to quickly answer to questions regarding some details in definitions, and/or formally state theorems that are used in the main part of the manuscript. As a consequence, these cheatsheets have little to no pedagogical virtue, are not exhaustive, and are for the most part unordered.

Set Theory Cheat Sheet A.

A.1. Sets

Definition A.1.1. *We write* $\{a, b, c, \dots\}$ *for the set with elements* a, b, *etc.*

Definition A.1.2. *We write* $|S|$ *for the number of elements in* S *when the set is finite.*

Definition A.1.3. *A set* S *is* countable *if there exists a map* $f : S \to \mathbb{N}$ *that is injective.*

Definition A.1.4. *We write* $A \subseteq B$ *to denote the inclusion of the set* A *inside the set* B*, that is, whenever* $\forall x. x \in A \Rightarrow x \in B$*.*

We write $A \subseteq_{\mathsf{fin}} B$ *to denote the fact that* A *is a finite set and* $A \subseteq B$*.*

Definition A.1.5. *We write* $\mathcal{P}(A)$ *for the set of subsets of* A*.*

Definition A.1.6. *We write* $\mathcal{P}_{\mathsf{fin}}(A)$ *for the set of finite subsets of* A*.*

Definition A.1.7. *We write* $\{x \in A : P(x)\}$ *for the usual comprehension axiom of set theory.*

Definition A.1.8. *A set* A *is a* co-finite *subset of* B*, whenever* $A \subseteq B$ *and* $B \setminus A$ *is finite.*

Definition A.1.9. *A* multiset M *over a set* X *is a function from* X *to* $\mathbb{N}$*.*

We write $\{\!|\, a, a, b \,|\!\}$ *for the multiset containing two occurrences of* a *and one occurrence of* b*.*

A multiset is finite *whenever* $\sum_{x \in \mathsf{X}} M(x)$ *is finite.*

A.2. Relations

Definition A.2.1. *A relation R between S and S' is a subset of $S \times S'$.*

Definition A.2.2. *A relation R over a set X is reflexive if xRx for all$x \in \mathsf{X}$.*

Definition A.2.3. *A relation R over a set X is transitive if xRy and yRz implies xRz for all $x, y, z \in \mathsf{X}$.*

Definition A.2.4. *The transitive closure of a relation R over a set X is the coarsest relation that is transitive and contains R.*

Definition A.2.5. *A relation R over a set X is symmetric if xRy implies yRx for all $x, y \in \mathsf{X}$.*

Definition A.2.6. *A relation R over a set X is antisymmetric if xRy and yRx implies $x = y$ for all $x, y \in \mathsf{X}$.*

Definition A.2.7. *A relation R over a set X is an equivalence relation if it is reflexive, transitive, and symmetric.*

Definition A.2.8. *Let R be an equivalence relation over a set X. The equivalence class of a point $x \in \mathsf{X}$ is the set $\{ y \in \mathsf{X} \colon xRy \}$.*

The quotient of X by R is the collection of its equivalence classes.

Definition A.2.9. *Let R be an equivalence relation over a set X. A subset S of X is saturated if it is a union of equivalence classes.*

Fact A.2.10. The equivalence classes of an equivalence relation R over X partition the space X.

Definition A.2.11. *An equivalence relation R over X is of finite index whenever it has finitely many distinct equivalence classes.*

A.3. Functions

Definition A.3.1. *We write* dom(f) *for the domain of a function* f.

Definition A.3.2. *We write* Im(f) *for the image of the function* f, *that is, the collection* $\{ f(x) : x \in \mathrm{dom}(f) \}$.

We often use the notation $f(S) \stackrel{\text{def}}{=} \{ f(x) : x \in S \}$, *in which case* $\mathrm{Im}(f) = f(\mathrm{dom}(f))$.

Definition A.3.3. *A function from* X *to* Y *is* injective *whenever* $\forall x, y \in X, f(x) = f(y) \Rightarrow x = y$.

Definition A.3.4. *A function from* X *to* Y *is* surjective *whenever* $f(X) = Y$.

Definition A.3.5. *A function from* X *to* Y *is* bijective *if it is injective and surjective.*

Definition A.3.6. *Let* $(\mathsf{X}, \leq_{\mathsf{X}})$ *and* $(\mathsf{Y}, \leq_{\mathsf{Y}})$ *be two partial orders. A map* $f : \mathsf{X} \to \mathsf{Y}$ *is a* quasi-embedding of partial orders *if for all* $x, x' \in \mathsf{X}$, $f(x) \leq_{\mathsf{Y}} f(x')$ *implies* $x \leq_{\mathsf{X}} x'$.

Definition A.3.7. *Let* $(\mathsf{X}, \leq_{\mathsf{X}})$ *and* $(\mathsf{Y}, \leq_{\mathsf{Y}})$ *be two quasi-orders. A map* $f : \mathsf{X} \to \mathsf{Y}$ *is an* embedding of quasi-orders *if for all* $x, x' \in \mathsf{X}$, $f(x) \leq_{\mathsf{Y}} f(x')$ *if and only if* $x \leq_{\mathsf{X}} x'$.

Definition A.3.8. *Let* $(\mathsf{X}, \leq_{\mathsf{X}})$ *and* $(\mathsf{Y}, \leq_{\mathsf{Y}})$ *be two quasi-orders. A map* $f : \mathsf{X} \to \mathsf{Y}$ *is an* isomorphism of quasi-orders *if* f *is an embedding of quasi-orders that as an inverse* g *that is also an embedding of quasi-orders.*

Definition A.3.9. *Let* $(\mathsf{X}, \leq_{\mathsf{X}})$ *and* $(\mathsf{Y}, \leq_{\mathsf{Y}})$ *be two partial orders. A map* $f : \mathsf{X} \to \mathsf{Y}$ *is* monotone *or* non-decreasing *if for all* $x, x' \in \mathsf{X}$, $x \leq_{\mathsf{X}} x'$ *implies* $f(x) \leq_{\mathsf{Y}} f(x')$.

Definition A.3.10. *Let* $f : \mathsf{X} \to \mathsf{X}$. *An element* $a \in \mathsf{X}$ *is a* fixed point *of* f *whenever* $f(a) = a$.

Order, Algebras, and Rings Cheat-Sheet

B.

B.1. Orders

Definition B.1.1. *A relation R over a set X is a* partial order *if it is reflexive, transitive, and antisymmetric.*

Definition B.1.2. *A partial order $\leq$ over a set X is a* total order *if for all $x, y \in \mathsf{X}$, either $x \leq y$ or $y \leq x$.*

Definition B.1.3. *A relation $\leq$ over a set X is a* quasi-order *if it reflexive, and transitive.*

We use the term preorder *as a synonym of* quasi-order.

Definition B.1.4. *Let $(\mathsf{X}, \leq)$ be a quasi-order. A sequence $(a_i)_{i\in\mathbb{N}}$ is a* decreasing sequence *whenever $a_i < a_j$ for all $i < j \in \mathbb{N}$.*

Definition B.1.5. *A quasi-order $(\mathsf{X}, \leq)$ is* well-founded *whenever every decreasing sequence $(x_i)_{i\in\mathbb{N}}$ of elements in X is finite.*

Definition B.1.6. *Let $(\mathsf{X}, \leq)$ be a quasi-order, and $S \subseteq \mathsf{X}$. A point $a \in \mathsf{X}$ is an* upper bound *of S if for all $b \in S$, $a \leq b$.*

Conversely, $a \in \mathsf{X}$ is an lower bound *of S if for all $b \in S$, $b \leq a$.*

Definition B.1.7. *Let $(\mathsf{X}, \leq)$ be a quasi-order. An element $a \in \mathsf{X}$ is a* maximum *if a is an upper bound of X. Conversely, it is a* minimum *if it is a lower bound of X.*

We write min.

Definition B.1.8. *Let $(\mathsf{X}, \leq)$ be a quasi-order. An element $a \in \mathsf{X}$ is a* maximal element *if for all $b \in \mathsf{X}$, $a \leq b \Rightarrow a = b$. Conversely, it is a* minimal element *if for all $b \in \mathsf{X}$, $b \leq a \Rightarrow a = b$.*

We write $\max S$ (resp. $\min S$) for the maximal element *(resp.* minimal element*) of S when it exists.*

Definition B.1.9. *Let* $(\mathsf{X}, \leq)$ *be a partial order. It is* inductive *when every totally ordered subset of* X *has an upper bound in* X.

Lemma B.1.10. *Let* $(\mathsf{X}, \leq)$ *be an inductive partial order. Then* X *contains at least one maximal element.*

This is known as Zorn's Lemma.

Definition B.1.11. *Let* $(\mathsf{X}, \leq)$ *be a partial order and* $S \subseteq \mathsf{X}$ *be a subset of* X.

A point $a \in \mathsf{X}$ *is the* supremum *of* S*, written* $a \stackrel{\text{def}}{=} \sup S$ *whenever* a *is the minimal upper bound of* S.

Similarly, a *is the* infimum *of* S*, written* $a \stackrel{\text{def}}{=} \inf S$ *whenever* a *is the maximal lower bound of* S.

Definition B.1.12. *Let* $(\mathsf{X}, \leq)$ *be a well-founded partial order. The* ordinal height *of* X*, written* $\mathrm{h}(\mathsf{X})$*, is defined by well-founded induction as follows:*

- *The* ordinal height *of* $\emptyset$ *is zero,*
- *The* ordinal height *of* X *is defined as the supremum of the* ordinal heights *of* $\{ y \in \mathsf{X} \colon y < x \}$ *where* x *ranges over elements in* X.

B.2. Lattices

Definition B.2.1. *A* lattice *is a partial order* $(\mathsf{X}, \leq)$ *such that every set* $\{ a, b \} \subseteq \mathsf{X}$ *has both a supremum and an infimum.*

Definition B.2.2. *Let* $(\mathsf{X}, \leq)$ *be a lattice. A set* $S \subseteq \mathsf{X}$ *is a* sublattice *of* $(\mathsf{X}, \leq)$ *when* S *is a lattice and the infimums and supremums are computed as in* X.

Definition B.2.3. *A lattice* $(\mathsf{X}, \leq)$ *is a* complete lattice *if every subset* $S \subseteq \mathsf{X}$ *has both a supremum and an infimum.*

In particular, a complete lattice *is non-empty.*

Theorem B.2.4 (Knaster-Tarski)**.** *Let* $(\mathsf{X}, \leq)$ *be a complete lattice and* $f \colon \mathsf{X} \to \mathsf{X}$ *be a monotone function. Then, the set of fixed points of* f *is a complete lattice.*

This is the Knaster-Tarski fixed point theorem.

B.3. Boolean Algebras

Definition B.3.1. *Let X be a set. A* Boolean subalgebra *of $\mathcal{P}(\mathsf{X})$ is a subset of $\mathcal{P}(\mathsf{X})$, stable under finite intersections, finite unions, complement, containing X, and containing $\emptyset$.*

In this document, we will use the term Boolean algebra *interchangeably with Boolean subalgebras of $\mathcal{P}(\mathsf{X})$ for some set X.*

Definition B.3.2. *Let $\mathsf{X}' \subseteq \mathsf{X}$, and $\mathcal{B}$ be a Boolean subalgebra of $\mathcal{P}(\mathsf{X})$. The* induced Boolean algebra *over X', written $\mathcal{B}_{|\mathsf{X}}$ is the collection of the sets $S \cap \mathsf{X}'$ for $S \in \mathcal{B}$.*

Definition B.3.3. *Let X be a set and $S \subseteq \mathcal{P}(\mathsf{X})$. The Boolean algebra* generated by *S, written $\langle S \rangle_{\mathsf{bool}}$, is the smallest Boolean subalgebra of $\mathcal{P}(\mathsf{X})$ containing S.*

Equivalently, it is the Boolean subalgebra of $\mathcal{P}(\mathsf{X})$ composed of finite unions of finite intersections of subsets of the form $D \cap \mathsf{X} \setminus D'$ where $D, D' \in S$.

Definition B.3.4. *Let $\mathcal{B}$ and $\mathcal{B}'$ be two Boolean subalgebras respectively of $\mathcal{P}(\mathsf{X})$ and $\mathcal{P}(\mathsf{X}')$. The* product algebra *written $\mathcal{B} \times_b \mathcal{B}'$ is the Boolean subalgebra of $\mathsf{X} \times \mathsf{X}'$ generated by the sets $D \times D'$ where D ranges over $\mathcal{B}$, and D' ranges over $\mathcal{B}'$.*

For a family $(\mathsf{X}_i, \mathcal{B}_i)_{i \in I}$ of sets with associated Boolean subalgebras, the product algebra *is generated by the products $\prod_{i \in I} D_i$, where $D_i \in \mathcal{B}_i$ and $D_i = \mathsf{X}_i$ for all but finitely many i's. The* product algebra *is then written $\prod_{i \in I} \mathcal{B}_i$.*

Definition B.3.5. *Let $\mathcal{B}$ and $\mathcal{B}'$ be two Boolean subalgebras respectively of $\mathcal{P}(\mathsf{X})$ and $\mathcal{P}(\mathsf{X}')$. The* sum algebra *$\mathcal{B} +_b \mathcal{B}'$ is the Boolean subalgebra of $\mathsf{X} \uplus \mathsf{X}'$ generated by the sets $D \cup D'$, where D ranges over $\mathcal{B}$ and D' ranges over $\mathcal{B}'$.*

For a family $(\mathcal{B}_i)_{i \in I}$ of Boolean subalgebras, the sum algebra *is generated by the unions $\bigcup_{i \in I} D_i$, where $D_i \in \mathcal{B}_i$ The* sum algebra *is then written $\sum_{i \in I} \mathcal{B}_i$.*

Definition B.3.6. *A Boolean algebra is a* complete Boolean algebra *whenever it is stable under arbitrary intersections.*

Equivalently, a complete Boolean algebra *is a Boolean algebra that is stable under arbitrary unions.*

Definition B.3.7. *Let $S \subseteq \mathsf{X}$. The complete Boolean subalgebra of $\mathcal{P}(\mathsf{X})$ generated by S, $\langle S \rangle_{\mathsf{bool}}^{\mathsf{comp}}$ is the smallest complete Boolean subalgebra of $\mathcal{P}(\mathsf{X})$ that contains S.*

B.4. Rings

Definition B.4.1. *We write $\mathbb{N}$ for the set of* natural numbers*, $\mathbb{Z}$ for the set of* integers*, $\mathbb{Q}$ for the set of* fractions*, $\mathbb{R}$ for the set of* real numbers*, and $\mathbb{C}$ for the set of* complex numbers*.*

Definition B.4.2. *We write $K[X_1, \ldots, X_n]$ for the ring of polynomials with indeterminates $X_1, \ldots, X_n$ and coefficients in the field K.*

Definition B.4.3. *A (bilateral)* ideal *of a ring is a subset S that is stable under addition, and such that $a \times S \times b \subseteq S$ for all a, b elements of the ring.*

Definition B.4.4. *A ring is a* Noetherian ring *whenever every increasing sequence $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ of ideals has a maximal element.*

C. Well-Quasi-Orderings Cheat Sheet

Definition C.0.1. *Let* $(\mathsf{X}, \leq)$ *and* $(\mathsf{X}', \leq')$ *be two quasi-orders. The* co-product ordering *over* $\mathsf{X} \uplus \mathsf{X}'$ *is defined by*

$$x \leq_{\mathsf{X} \uplus \mathsf{X}'} y \iff \vee \begin{cases} x \leq y & \text{and } x, y \in \mathsf{X} \\ x \leq' y & \text{and } x, y \in \mathsf{X}' \end{cases} .$$

Definition C.0.2. *Let* $(\mathsf{X}, \leq)$ *and* $(\mathsf{X}', \leq')$ *be two quasi-orders. The* product ordering *over* $\mathsf{X} \times \mathsf{X}'$ *is defined by*

$$(x, y) \leq_{\mathsf{X} \times \mathsf{X}'} (x', y') \iff \wedge \left\{ x \leq x' y \leq' y' \right. .$$

Definition C.0.3. *Let* $(\mathsf{X}, \leq)$ *be a quasi-order. The* multiset ordering *relation* $\leq_{\circledast}$ *over* $\mathsf{M}^{\diamond}(\mathsf{X})$ *is defined inductively as follows:*

- $\emptyset \leq_{\circledast} m$ *for all* $m \in \mathsf{M}^{\diamond}(\mathsf{X})$,
- *If* $m \leq_{\circledast} m'$ *and* $a \in \mathsf{X}$, *then* $m \leq_{\circledast} m' \uplus \{\!| a |\!\}$,
- *If* $a \leq b$ *and* $m \leq_{\circledast} m'$ *then* $m \uplus \{\!| a |\!\} \leq_{\circledast} m' \uplus \{\!| b |\!\}$.

Definition C.0.4. *Let* $(\mathsf{X}, \leq)$ *be a quasi-order. The* Hoare quasi-ordering *relation* $\leq_{\mathcal{H}}$ *over* $\mathsf{P_f}(\mathsf{X})$ *is defined as follows:*

$$S \leq_{\mathcal{H}} S' \iff \forall x \in S, \exists y \in S', x \leq y \quad .$$

Lemma C.0.5. *Let* $(W, \leq)$ *be a quasi-ordered set. Let* $(x_i)_{i \in \mathbb{N}}$ *be a bad sequence in* W, *and* $\alpha \colon \mathbb{N} \to \mathbb{N}$ *be a strictly increasing map. Then, the sequence* $\left(x_{\alpha(i)}\right)_{i \in \mathbb{N}}$ *is a bad sequence.*

Topology Cheat Sheet D.

D.1. Basic Topology

Definition D.1.1. *Let (X, τ) be a topological space. We write $\mathcal{H}(\mathsf{X})$ for the collection of closed subsets of X.*

Definition D.1.2. *Let (X, τ) and (X, θ) be two topological spaces. The topology τ is* coarser *than θ if $\tau \subseteq \theta$, in this case, we also say that θ is* finer *than τ.*

Definition D.1.3. *Let (X, τ) be a topological space. A subset $S \subseteq \mathsf{X}$ is* dense *when for all non-empty subset $U \in \tau$, $S \cap U \neq \emptyset$.*

Definition D.1.4. *Let (X, τ) be a topological space, $x \in \mathsf{X}$ be a point, and $U \in \tau$ be an open subset. We say that U is an* open neighbourhood *of x whenever $x \in U$.*

Definition D.1.5. *Let (X, τ) be a topological space, and $S \subseteq \mathsf{X}$ be a subset. We write $\mathring{S}$ for the* interior *of S, i.e., the maximal open subset contained in S.*

Similarly, we define the closure *of S, written $\overline{S}$, as the minimal closed subset containing S.*

Definition D.1.6. *Let (X, τ) be a topological space. The space X is* irreducible *if all non-empty open subsets of X are dense, or equivalently if any two non-empty open subsets have a non-empty intersection.*

A subset $S \subseteq \mathsf{X}$ is irreducible *if it is* irreducible *in its induced topology.*

D.2. Building New Topologies

Definition D.2.1. *The co-finite topology over a set* X *is the topology composed of the co-finite subsets of* X.

Definition D.2.2. *Let* (X, τ) *and* (Y, θ) *be two topological spaces. The product topology* $\tau \times_t \theta$ *over* $\mathsf{X} \times \mathsf{Y}$ *is the topology generated by the sets* $U \times V$*, where* U *ranges over* τ *and* V *ranges over* θ.

For a family $(\mathsf{X}_i, \tau_i)_{i \in I}$ *of topological spaces, the product topology is generated by the products* $\prod_{i \in I} U_i$*, where* $U_i \in \tau_i$ *and* $U_i = \mathsf{X}_i$ *for all but finitely many* i*'s. The product topology is then written* $\prod_{i \in I} \tau_i$.

Definition D.2.3. *Let* (X, τ) *and* (Y, θ) *be two topological spaces. The sum topology over* $\mathsf{X} \uplus \mathsf{Y}$ *is the topology generated by the sets* $U \cup V$*, where* U *ranges over* τ *and* V *ranges over* θ.

For a family $(\mathsf{X}_i, \tau_i)_{i \in I}$ *of topological spaces, the sum topology is generated by the unions* $\bigcup_{i \in I} U_i$*, where* $U_i \in \tau_i$ *The sum topology is then written* $\sum_{i \in I} \tau_i$.

Definition D.2.4. *Let* (X, τ) *and* (X, θ) *be two topological spaces. The join topology over* X *is the topology generated by* $\tau \cup \theta$.

For a family $(\mathsf{X}, \tau_i)_{i \in I}$ *of topological spaces, the join topology is generated by the union* $\bigcup_{i \in I} \tau_i$*. The join topology is then written* $\bigvee_{i \in I} \tau_i$.

Definition D.2.5. *Let* (X, τ) *be a topological space and* $S \subseteq \mathsf{X}$ *be a subset of* X*. The topology induced by* τ *over* S *is the topology* $\tau_{|S} \stackrel{\text{def}}{=} \{\, U \cap S \colon U \in \tau \,\}$.

Definition D.2.6. *Let* (X, τ) *be a topological space and* $\equiv$ *be an equivalence relation over* X*. The quotient topology over the equivalence classes of* $\equiv$ *is the collection of sets*

$$\{\, E \colon E \cap U \neq \emptyset \,\} \quad ,$$

where U *ranges in* τ.

D.3. Topological Maps

Definition D.3.1. *Let X and Y be two topological spaces. A map $f: \mathsf{X} \to \mathsf{Y}$ is* continuous *when for all open subset U of Y, $f^{-1}(U)$ is an open subset of X.*

Definition D.3.2. *Let X and Y be two topological spaces. A map $f: \mathsf{X} \to \mathsf{Y}$ is an* homeomorphism *when f is a bijection, f is continuous, and f^{-1} is continuous.*

Definition D.3.3. *Let X and Y be two topological spaces. A map $f: \mathsf{X} \to \mathsf{Y}$ is a* topological embedding *if it is injective, continuous and $f: \mathsf{X} \to f(\mathsf{X})$ is an homeomorphism when $f(\mathsf{X})$ is equipped with the induced topology of Y.*

Definition D.3.4 ([26, Definition 1.2.2])**.** *Let X and Y be two topological spaces. A map $f: \mathsf{X} \to \mathsf{Y}$ is a* spectral map *if it is continuous and for all $U \in \mathcal{K}^\circ(\mathsf{Y})$, $f^{-1}(U) \in \mathcal{K}^\circ(\mathsf{X})$.*

D.4. Compactness

Theorem D.4.1. *The* Tychonoff theorem *states that the product of compact spaces is compact.*

Definition D.4.2. *Let (X, τ) be a topological space. We write $\overline{\mathcal{K}}(\mathsf{X})$ for the closed subsets of X that are also compact subsets of X.*

Lemma D.4.3. *The union of two compact subsets is compact.*

D.5. Separation

Definition D.5.1. *A topological space (X, τ) is* T_0 *if and only if for every pair $x, y \in \mathsf{X}$, there exists an open subset U such that $x \in U$ and $y \notin U$, or $y \in U$ and $x \notin U$. Equivalently, if $\leq_\tau$ is a partial order.*

Definition D.5.2. *The* T_0 quotient *of a topological space (X, τ) is the space obtained by considering equivalence classes for $\leq_\tau$.*

Definition D.5.3. *A topological space* (X, τ) *is a* sober space *whenever every (non-empty) irreducible closed subset of* X *is the topological closure of a single point.*

Definition D.5.4. *Let* (X, τ) *be a topological space. The* sobrification $\mathcal{S}(\mathsf{X})$ *is the collection of all irreducible closed subsets of* X *equipped with the lower Vietoris topology.*

The map $\mathsf{X} \mapsto \mathcal{S}(\mathsf{X})$ *is an endofunctor of* Top *that maps* X *to the "free sober space" generated by* X.[1]

1: See [45, Section 8.2.6].

D.5.1. Ordinal Invariants

Definition D.5.5. *Let* (X, τ) *be a Noetherian space. Then the* stature *of* τ*, written* $\|\tau\|$*, is the ordinal height of* $(\mathcal{H}(\mathsf{X}), \subseteq)$ *minus one.*

D.6. Noetherian Spaces

Definition D.6.1. *Let* $n \in \mathbb{N}$*. The* Zariski topology *over* $\mathbb{C}^n$ *is generated by the following closed subsets:*

$$\{\, \vec{x} \in \mathbb{C}^n : P(x_1, \ldots, x_n) = 0 \,\} \quad ,$$

where $P \in \mathbb{C}[X_1, \ldots, X_n]$.

Definition D.6.2. *Let* (X, τ) *be a topological space. Remark that* $\mathsf{M}^{\diamond}(\mathsf{X})$ *is exactly the quotient of* $\mathsf{X}^{\star}$ *by the equivalence relation* $\equiv$ *that permutes letters. The* multiset topology *over* $\mathsf{M}^{\diamond}(\mathsf{X})$ *is defined as the quotient topology of the subword topology over* $\mathsf{X}^{\star}$ *by the equivalence relation* $\equiv$.

Definition D.6.3. *Let* (X, τ) *be a topological space. The* lower Vietoris *over* $\mathcal{P}(\mathsf{X})$ *is the topology generated by the following open subsets:*

$$\{\, E \in \mathcal{P}(\mathsf{X}) : E \cap U \neq \emptyset \,\} \quad ,$$

where U *ranges in* τ.

Lemma D.6.4. *The product of two Noetherian spaces is Noetherian.*

Lemma D.6.5 ([45, Proposition 9.7.6])**.** *A space* (X, τ) *is Noetherian if and only if for every increasing sequence of open subsets* $(U_i)_{i \in \mathbb{N}}$*, there exists* $j \in \mathbb{N}$ *such that* $\bigcup_{i \in \mathbb{N}} U_i = \bigcup_{i \leq j} U_i$.

Lemma D.6.6 ([45, Proposition 9.7.18])**.** *If* (X, τ) *and* (X, τ') *are Noetherian spaces, then* $(\mathsf{X}, \tau \vee_t \tau')$ *is a Noetherian space.*

D.7. Strange Topologies

Definition D.7.1. *Let* $(P, \leq)$ *be a quasi-order. The* d-topology *is the coarsest topology having as closed subsets the subsets* $S \subseteq P$ *such that for all directed families* $(x_i)_{i \in I}$ *of elements in* S *that has a supremum* $x = \sup_{i \in I} x_i$ *in* P*,* $x \in S$*.*

D.8. Topology of the Real Line

Definition D.8.1. *Let* $a \leq b$ *be two real numbers. We define the* open interval *between* a *and* b $]a,b[$ *as the set* $\{ x \in \mathbb{R} : a < x < b \}$*. Similarly, we define the* closed interval *between* a *and* b $[a,b]$ *as the set* $\{ x \in \mathbb{R} : a \leq x \leq b \}$*.*

For completeness, let us also introduce the notations for the natural closed/open variants $]a,b]$ *and* $[a,b[$*.*

Lemma D.8.2. *Let* $x \in \mathbb{R}$*. Then, there exists a natural number* $n \in \mathbb{N}$ *such that* $n > x$*.*

This is called the Archimedean property*.*

Lemma D.8.3. *Let* $([a_i,b_i])_{i \in \mathbb{N}}$ *be a decreasing sequence of non-empty closed intervals. Formally,* $a_i \leq b_j$ *for all* $i, j \in \mathbb{N}$*,* $(a_i)_{i \in \mathbb{N}}$ *is a non-decreasing sequence, and* $(b_i)_{i \in \mathbb{N}}$ *is a non-increasing sequence. There exists* $x \in \bigcap_{i \in \mathbb{N}} [a_i,b_i]$*.*

This is called the Nested Intervals Theorem*.*

Logic Cheat Sheet

E.

E.1. Master Theorems of Model Theory

Theorem E.1.1. *Let σ be a countable relational signature, and T be a first order theory. If T has a model $\mathfrak{A}$, then there exists a countable model $\mathfrak{B}$ that satisfies T.*

This is the consequence of the Löwenheim-Skolem Theorem *for countable relational signatures.*

Theorem E.1.2. Lyndon's positivity theorem *states that a first order formula is preserved under X-positive homomorphisms on all structures if and only if it is logically equivalent to an X-positive formula.*

E.2. Monadic Second Order Logic

Definition E.2.1. *Let σ be a relational signature, $\mathbb{V}$ be a countable set of first order free variables, and $\mathbb{MV}$ be a countable set of* second order free variables*. The* Monadic Second Order formulas *with first order free variables $\vec{x} \subseteq_{\mathsf{fin}} \mathbb{V}$, and* second order free variables *$\vec{X} \subseteq_{\mathsf{fin}} \mathbb{MV}$ are defined inductively as follows:*

$$\begin{aligned}
\varphi(\vec{x}, \vec{X}) ::= {} & \varphi(\vec{x'}, \vec{X'}) \wedge \varphi(\vec{x''}, \vec{X''}) && \textit{when: } \vec{x} = \vec{x'} \cup \vec{x''} \wedge \vec{X} = \vec{X'} \cup \vec{X''} \\
\mid {} & R(x_1, \dots, x_n) && \textit{when: } R \in \sigma \wedge \mathsf{arity}(R) = n \\
\mid {} & x \in X \\
\mid {} & \neg\varphi(\vec{x}, \vec{X}) \\
\mid {} & \exists y.\varphi(\vec{x}y, \vec{X}) \\
\mid {} & \exists Y.\varphi(\vec{x}, \vec{X}Y) \\
\mid {} & \top
\end{aligned}$$

We call $\mathsf{MSO}[\sigma]$ *the collection of* Monadic Second Order formulas*, and omit σ when the signature is irrelevant or clear from the context.*

Definition E.2.2. *The* satisfaction relation $\models$ *is defined by induction on the formulas follows: given $\mathfrak{A} \in \mathsf{Struct}(\sigma)$, $\vec{x} \subseteq_{\mathsf{fin}} \mathbb{V}$,*

$\vec{X} \subseteq_{\mathsf{fin}} \mathbb{MV}$, $\upsilon\colon \vec{x} \to \mathfrak{A}$, $\Upsilon\colon \vec{X} \to \mathcal{P}(\mathfrak{A})$, *and* $\varphi(\vec{x}, \vec{X}) \in \mathsf{MSO}[\sigma]$

$$\begin{aligned}
\mathfrak{A}, \upsilon, \Upsilon \models R(\vec{y}) &\overset{\mathit{def}}{\iff} (\upsilon(y_1), \dots, \upsilon(y_n)) \in R^{\mathfrak{A}} \\
\mathfrak{A}, \upsilon, \Upsilon \models y \in Y &\overset{\mathit{def}}{\iff} \upsilon(y) \in \Upsilon(Y) \\
\mathfrak{A}, \upsilon, \Upsilon \models \psi_1 \wedge \psi_2 &\overset{\mathit{def}}{\iff} \mathfrak{A}, \upsilon, \Upsilon \models \psi_1 \text{ and } \mathfrak{A}, \upsilon, \Upsilon \models \psi_2 \\
\mathfrak{A}, \upsilon, \Upsilon \models \neg\psi &\overset{\mathit{def}}{\iff} \mathfrak{A}, \upsilon \not\models \psi_1 \\
\mathfrak{A}, \upsilon, \Upsilon \models \exists z.\psi &\overset{\mathit{def}}{\iff} \text{ there exists } a \in \mathfrak{A}, \mathfrak{A}, \upsilon[z \mapsto a], \Upsilon \models \psi \\
\mathfrak{A}, \upsilon, \Upsilon \models \exists Z.\psi &\overset{\mathit{def}}{\iff} \text{ there exists } S \subseteq \mathrm{dom}(\mathfrak{A}), \\
&\qquad \mathfrak{A}, \upsilon, \Upsilon[Z \mapsto S] \models \psi \\
\mathfrak{A}, \upsilon, \Upsilon \models \top &\overset{\mathit{def}}{\iff} \text{ always}
\end{aligned}$$

Definition E.2.3. *Let σ be a relational signature, $\mathcal{C}$ be a class of structures over σ, and $\varphi(\vec{x}) \in \mathsf{MSO}[\sigma]$ be a Monadic Second Order formula that only has first order free variables. The formula $\varphi(\vec{x})$ is an r-local MSO formula over $\mathcal{C}$ if and only if the following holds:*

$$\forall \mathfrak{A} \in \mathcal{C}, \forall \upsilon\colon \vec{x} \to \mathfrak{A}, \mathfrak{A}, \upsilon \models \varphi \text{ if and only if } \mathcal{N}_{\mathfrak{A}}(\upsilon(\vec{x}), r), \upsilon \models \varphi \quad .$$

E.3. Structural Properties

Definition E.3.1. *Let σ be a relational signature. A class $\mathcal{C}$ of structures has bounded degree whenever there exists $M \in \mathbb{N}$ such that for all $\mathfrak{A} \in \mathcal{C}$, for all $a \in \mathrm{dom}(\mathfrak{A})$, the number of edges in $\mathrm{Gaif}(\mathfrak{A})$ that contain a is bounded by M.*

Definition E.3.2. *A tree decomposition of a graph G is a tree T with nodes $X_1, \dots, X_n$, where each X_i is a subset of the vertices of G, satisfying the following properties:*

- *Every vertex of G belongs to some X_i,*
- *For every vertex v of G, the nodes of T that contain v form a (connected) subtree of T,*
- *For every edge (u, v) in the graph G, there exists $1 \leq i \leq n$ such that $\{ u, v \} \subseteq X_i$.*

The width of a tree decomposition is the maximum of $|X_i| - 1$ when $1 \leq i \leq n$.

The tree width of a graph G is defined as the minimum width of a tree decomposition of G.

A class has bounded tree width whenever there exists $M \in \mathbb{N}$ such that for all graph G in the class, the tree width of G is at most M.

Definition E.3.3. *The clique width of a graph G is defined as the minimum number of labels needed to construct G using the following* *operations:*

- *Creation of a labelled vertex*
- *Disjoint union of two labelled graphs*
- *Selecting a pair of labels (l, l') and connecting every vertex labelled by l to every vertex labelled by l'.*
- *Relabelling every vertex with label l into an l' labelled vertex.*

A class of graphs has bounded clique width whenever there exists $M \in \mathbb{N}$ such that for all graph G in the class, the clique width of G is at most M.

One can interpret m-partite cographs as graphs of clique width m without relabelling functions.

Definition E.3.4. *An m-partite cograph is a graph that can be build using the following operations:*

- *Creating a node with a label $1 \leq i \leq m$,*
- *Taking the disjoint union of two graphs,*
- *Selecting two labels $1 \leq i, j \leq m$ and connecting every vertex labelled by i to every vertex labelled by j.*

Definition E.3.5. *A graph is planar whenever it can be drawn on the two-dimensional plane in such a way that edges intersect only at their endpoints.*

Definition E.3.6. *The definition of Shrub Depth is quite involved, and we unfortunately have to redirect the reader to [41, Definition 3.3].*

[41]: Ganian, Hlinený, Nešetřil, Obdrzálek and Mendez (2017), 'Shrub-depth: Capturing Height of Dense Graphs'

Definition E.3.7. *A class $\mathcal{C}$ of finite structures is quasi-wide if there exists a function $f : \mathbb{N} \to \mathbb{N}$, $N : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$, such that for all $r, m \in \mathbb{N}$, for all structure $\mathfrak{A} \in \mathcal{C}$ with at least $N(r, m)$ elements, there exists a subset $C \subseteq \mathrm{dom}(\mathfrak{A})$ of size at most $f(r)$, such that $\mathrm{Gaif}(\mathfrak{A}) \setminus C \models \exists_r^{\geq m} x.\top$.*

Definition E.3.8. *A class $\mathcal{C}$ of finite graphs is nowhere dense if there exists a function $f : \mathbb{N} \to \mathbb{N}$, such that for all $r \in \mathbb{N}$ and every $G \in \mathcal{C}$, $K_{f(r)}$ is not an* depth-r minor *of G. We redirect the reader to [77, Chapter 4, Section 2, Shallow Minors] for the definition of depth-r minors.*

[77]: Nešetřil and Ossona de Mendez (2011), 'On nowhere dense graphs'

E.4. Positive Sentences

Definition E.4.1. *Let σ be a relational signature, and X be a unary relational symbol inside σ. A sentence $\varphi \in \mathsf{FO}[\sigma]$ is X-positive if and only if the symbol X occurs below an even number of negations* *in the syntax tree of φ.*

Definition E.4.2. *Let σ be a relational signature, $\mathfrak{A}, \mathfrak{B}$ be two relational structures over σ, and X be a unary relational symbol inside σ. A map $f : \mathfrak{A} \to \mathfrak{B}$ is a* X-positive homomorphism *if and only if f is an embedding of $\mathfrak{A}$ into $\mathfrak{B}$ over the signature $\sigma \setminus \{ (X, 1) \}$, and f is a homomorphism from $\mathfrak{A}$ into $\mathfrak{B}$ over the whole signature σ.*

Definition E.4.3. *Let σ be a relational signature containing a unary predicate X. A sentence $\varphi \in \mathsf{FO}[\sigma]$ is in* X-positive Gaifman normal form *whenever it is a Boolean combination of basic local sentences $(\varphi_i)_{1 \leq i \leq n}$, where φ_i is of the form $\exists^{\geq k_i}_{r_i} x.\psi_i(x)$, $r_i, k_i \in \mathbb{N}$, and ψ_i is an X-positive formula for all $1 \leq i \leq n$.*

F. Category Theory Cheat Sheet

F.1. Basic Category

As per usual, we have to be extremely careful with the size issues, and for the sake of simplicity, we will avoid mentioning any of these issues here.

Definition F.1.1. *A category $\mathcal{C}$ is a collection of object together with collections $\mathrm{Hom}(A, B)$ of morphisms between A and B, where A, B ranges over the objects of $\mathcal{C}$, and an associative composition operator $\circ$: $\mathrm{Hom}(A, B) \times \mathrm{Hom}(B, C) \to \mathrm{Hom}(A, C)$, such that:*

- *For all objects A, there exists a morphism $\mathsf{Id}_A \in \mathrm{Hom}(A, A)$,*
- *For all $f \in \mathrm{Hom}(A, B)$, $f \circ \mathsf{Id}_A = f$ and $\mathsf{Id}_B \circ f = f$.*

Definition F.1.2. *Let $\mathcal{C}$ be a category. Two objects A, B of $\mathcal{C}$ are isomorphic whenever there exists $f \in \mathrm{Hom}(A, B)$ and $g \in \mathrm{Hom}(B, A)$ such that $f \circ g = \mathsf{Id}_B$ and $g \circ f = \mathsf{Id}_A$.*

Definition F.1.3. *Let $\mathcal{C}$ be a category. A category $\mathcal{C}'$ is a full subcategory of $\mathcal{C}$ if its objects belong to $\mathcal{C}$ and for all A, B in $\mathcal{C}'$, $\mathrm{Hom}_{\mathcal{C}'}(A, B) = \mathrm{Hom}_{\mathcal{C}}(A, B)$.*

Definition F.1.4. *Let $\mathcal{C}$ be a category and A be an object of $\mathcal{C}$. We write $\mathrm{Aut}(A)$ denotes the set of automorphisms of A, i.e., invertible elements of the monoid $(\mathrm{Hom}(A, A), \circ)$.*

Definition F.1.5. *The category Set is defined as having sets as objects, and functions as morphisms.*

Definition F.1.6. *The category Top is defined as having topological spaces as objects, and continuous functions as morphisms.*

Definition F.1.7. *The category Ord is defined as having quasi-orders as objects, and monotone maps as morphisms.*

Definition F.1.8. *Let $\mathcal{C}$ be a category. An object A of $\mathcal{C}$ is an initial object if for every B, $\mathrm{Hom}(A, B)$ contains exactly one morphism.*

Definition F.1.9. *Let $\mathcal{C}$ be a category, and A, B be two objects of $\mathcal{C}$. The product of A and B is an object C with two morphisms*

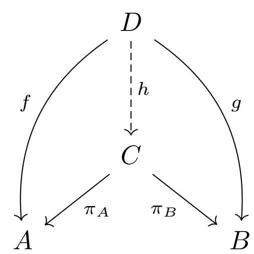


Figure F.1.: The commutative diagram of the product construction.

$\pi_A \in \mathrm{Hom}(C, A)$ and $\pi_B \in \mathrm{Hom}(C, B)$, such that for all D with morphisms $f \in \mathrm{Hom}(D, A)$, and $g \in \mathrm{Hom}(D, B)$, there exists a unique morphism $h \in \mathrm{Hom}(D, C)$ such that $f = \pi_A \circ h$ and $g = \pi_B \circ h$.

We refer the reader to Figure F.1 for a diagrammatic presentation of the above equations.

$$\begin{tikzcd}
& D & \\
& C \arrow[u, dashed, "h"] & \\
A \arrow[ur, "\iota_A"] \arrow[uur, bend left, "g"] & & B \arrow[ul, "\iota_B"'] \arrow[uul, bend right, "f"']
\end{tikzcd}$$

Figure F.2.: The commutative diagram of the co-product construction.

Definition F.1.10. *Let $\mathcal{C}$ be a category, and A, B be two objects of $\mathcal{C}$. The* co-product *of A and B is an object C with two morphisms $\iota_A \in \mathrm{Hom}(A, C)$ and $\iota_B \in \mathrm{Hom}(B, C)$, such that for all D with morphisms $f \in \mathrm{Hom}(A, D)$, and $g \in \mathrm{Hom}(B, D)$, there exists a unique morphism $h \in \mathrm{Hom}(C, D)$ such that $f = h \circ \iota_A$ and $g = h \circ \iota_B$.*

We refer the reader to Figure F.2 for a diagrammatic presentation of the above equations.

Definition F.1.11. *Let $\mathcal{C}$ be a category. A* projective system *is a family of objects $(E_i)_{i \in I}$, together with a family of morphisms $(f_{i,j})_{i \geq j \in I}$, such that $f_{i,j} \in \mathrm{Hom}(E_i, E_j)$, and such that $f_{i,i} = \mathsf{Id}_{E_i}$, and $f_{i,j} \circ f_{j,k} = f_{i,k}$ for all $i \geq j \geq k \in I$,*

Definition F.1.12. *Let $\mathcal{C}$ be a category. A* limit *of a projective system $(f_{i,j} \in \mathrm{Hom}(E_i, E_j))_{i \geq j \in I}$, is an object A of $\mathcal{C}$ together with morphisms $f_i \in \mathrm{Hom}(A, E_i)$, and such that $f_i = f_{i,j} \circ f_j$ for all $i \geq j \in I$.*

The limit *of a projective system is unique up to isomorphism.*

F.2. Functors

Definition F.2.1. *Let $\mathcal{C}$ and $\mathcal{C}'$ be two categories. A* functor *F from $\mathcal{C}$ to $\mathcal{C}'$ is a mapping from objects A of $\mathcal{C}$ to objects $F(A)$ of $\mathcal{C}'$, that associates with morphisms $f \in \mathrm{Hom}(A, B)$ of $\mathcal{C}$ a morphism $F(f) \in \mathrm{Hom}(F(A), F(B))$ in $\mathcal{C}'$. Furthermore, a* functor *should satisfy that*

- *$F(\mathsf{Id}_A) = \mathsf{Id}_{F(A)}$ for all objects A of $\mathcal{C}$,*
- *$F(f \circ g) = F(f) \circ F(g)$ for all morphisms f, g of $\mathcal{C}$.*

Definition F.2.2. *Let $\mathcal{C}$ be a category. An* endofunctor *of $\mathcal{C}$ is a functor from $\mathcal{C}$ to itself.*

Definition F.2.3. *Let $\mathcal{C}$ be a category, and F be an endofunctor of $\mathcal{C}$. An* initial algebra *of F is a pair $(\mu F, \delta)$ where μF is an object of $\mathcal{C}$ and $\delta \in \mathrm{Hom}(F(\mu F), \mu F)$, such that for all (A, ρ) where $\rho \in$*

Hom$(F(A), A)$, there exists a unique $h \in$ Hom$(\mu F, A)$ such that $h \circ \delta = \rho \circ F(h)$.

F.3. Liftings

Definition F.3.1. *The forgetful functor from* Top *(resp.* Ord*) to* Set *is the mapping from* (X, τ) *(resp.* $(P, \leq)$*) to the underlying set* X *(resp.* P*), that furthermore associates the map* f *to the map* f *(forgetting that it was continuous or monotone).*

Definition F.3.2. *An endofunctor* G' *of* Top *is a lift of an endofunctor* G *of* Set *if the following diagram commutes, where* U *is the forgetful functor*

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

Index

E

F

G

H

I

J

K

Q

R

S

T

U

W

X

Z

Math Commands

Bibliography

[1] Parosh Aziz Abdulla, Karlis Čerāns, Bengt Jonsson and Yih-Kuen Tsay. 'General decidability theorems for infinite-state systems'. In: *Proceedings of LICS'96*. IEEE, 1996, pp. 313–321. DOI: 10.1109/LICS.1996.561359 (cit. on p. 44).

[2] Serge Abiteboul, Richard Hull and Victor Vianu. *Foundations of Databases*. Addison-Wesley, 1995 (cit. on p. 24).

[3] Jiří Adámek, Horst Herrlich and George E. Strecker. *Abstract and Concrete Categories : The Joy of Cats*. Wiley, 1990 (cit. on p. 239).

[4] Miklós Ajtai and Yuri Gurevich. 'Monotone versus positive'. In: *Journal of the ACM* 34 (1987), pp. 1004–1015. DOI: 10.1145/31846.31852 (cit. on pp. 35, 37, 113).

[5] Miklós Ajtai and Yuri Gurevich. 'Datalog vs first-order logic'. In: *Journal of Computer and System Sciences* 49 (1994), pp. 562–588. DOI: 10.1016/S0022-0000(05)80071-6 (cit. on pp. 35, 37, 113).

[6] Albert Atserias, Anuj Dawar and Martin Grohe. 'Preservation under extensions on well-behaved finite structures'. In: *SIAM Journal on Computing* 38 (2008), pp. 1364–1381. DOI: 10.1137/060658709 (cit. on pp. 37, 40, 47, 55, 64, 67, 68, 74, 97, 98, 117, 121, 127).

[7] Albert Atserias, Anuj Dawar and Phokion G. Kolaitis. 'On preservation under homomorphisms and unions of conjunctive queries'. In: *Journal of the ACM* 53 (2006), pp. 208–237. DOI: 10.1145/1131342.1131344 (cit. on pp. 37, 40, 47, 55, 74).

[8] Mikołaj Bojańczyk, Luc Segoufin and Szymon Toruńczyk. 'Verification of Database-Driven Systems via Amalgamation'. In: *Proceedings of the 32nd ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*. Association for Computing Machinery, 2013, pp. 63–74. DOI: 10.1145/2463664.2465228. URL: https://doi.org/10.1145/2463664.2465228 (cit. on p. 234).

[9] Édouard Bonnet, Eun Jung Kim, Stéphan Thomassé and Rémi Watrigant. 'Twin-width I: tractable FO model checking'. In: *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. 2020, pp. 601–612 (cit. on p. 233).

[10] Simone Bova and Hubie Chen. 'How Many Variables are Needed to Express an Existential Positive Query?' In: vol. 63. 2019, pp. 1573–1594. DOI: 10.1007/s00224-018-9884-z. URL: https://doi.org/10.1007/s00224-018-9884-z (visited on 06/06/2023) (cit. on p. 56).

[11] Robert Brignall, Michael Engen and Vincent Vatter. 'A Counterexample Regarding Labelled Well-Quasi-Ordering'. In: *Graphs and Combinatorics* 34 (2018), pp. 1395–1409. DOI: 10.1007/s00373-018-1962-0 (cit. on p. 46).

[12] Florent Capelli, Arnaud Durand, Amélie Gheerbrant and Cristina Sirangelo. *A simple counter example of Lyndon's Theorem in the finite*. 2020 (cit. on p. 35).

[13] Donald D. Chamberlin and Raymond F. Boyce. 'SEQUEL: A Structured English Query Language'. In: (1974), pp. 249–264. DOI: 10.1145/800296.811515. URL: https://doi.org/10.1145/800296.811515 (cit. on p. 12).

[14] Ashok K. Chandra and Philip M. Merlin. 'Optimal implementation of conjunctive queries in relational databases'. In: *Proceedings of STOC'77*. 1977, pp. 77–90. DOI: 10.1145/800105.803397 (cit. on p. 23).

[15] Chen Chung Chang and H. Jerome Keisler. *Model Theory*. Vol. 73. Elsevier, 1990 (cit. on pp. i, 29, 32–34, 68).

[16] Yijia Chen and Jörg Flum. 'Forbidden Induced Subgraphs and the Łoś-Tarski Theorem'. In: *2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*. 2021, pp. 1–13. DOI: 10.1109/LICS52264.2021.9470742 (cit. on pp. 37, 39, 98, 113).

[17] Edgar F. Codd. 'A relational model of data for large shared data banks'. In: *FIXME Conference of the ACM* 13 (1970), pp. 377–387. DOI: 10.1145/362384.362685 (cit. on pp. 11, 17).

[18] Community. *The nLab wiki*. 2008. URL: https://ncatlab.org/nlab/show/HomePage (visited on 2023) (cit. on p. 239).

[19] Jean Daligault, Michael Rao and Stéphan Thomassé. 'Well-Quasi-Order of Relabel Functions'. In: *Order* 27 (2010), pp. 301–315. DOI: 10.1007/s11083-010-9174-0. URL: https://doi.org/10.1007/s11083-010-9174-0 (cit. on pp. 37, 40, 44–46, 179, 235).

[20] Anuj Dawar. 'Homomorphism preservation on quasi-wide classes'. In: *Journal of Computer and System Sciences* 76 (2010), pp. 324–332. DOI: 10.1016/j.jcss.2009.10.005 (cit. on pp. 47, 55, 74).

[21] Anuj Dawar, Martin Grohe, Stephan Kreutzer and Nicole Schweikardt. 'Approximation schemes for first-order definable optimisation problems'. In: *21st Annual IEEE Symposium on Logic in Computer Science (LICS'06)*. 2006, pp. 411–420. DOI: 10.1109/lics.2006.13 (cit. on pp. 68, 130, 131).

[22] Anuj Dawar and Abhisekh Sankaran. 'Extension Preservation in the Finite and Prefix Classes of First Order Logic'. In: *29th EACSL Annual Conference on Computer Science Logic (CSL 2021)*. Ed. by Christel Baier and Jean Goubault-Larrecq. Vol. 183. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2021, 18:1–18:13. DOI: 10.4230/LIPIcs.CSL.2021.18. URL: https://drops.dagstuhl.de/opus/volltexte/2021/13452 (cit. on pp. 35, 37).

[23] Stéphane Demeri, Alain Finkel, Jean Goubault-Larrecq, Sylvain Schmitz and Philippe Schnoebelen. 'Algorithmic Aspects of WQO Theory (MPRI course)'. 2012. URL: https://cel.archives-ouvertes.fr/cel-00727025 (cit. on pp. 3, 139, 178, 181, 239).

[24] Nachum Dershowitz and Iddo Tzameret. 'Gap Embedding for Well-Quasi-Orderings'. In: *Electronic Notes in Theoretical Computer Science* 84 (2003), pp. 80–90. DOI: 10.1016/S1571-0661(04)80846-6. URL: http://www.sciencedirect.com/science/article/pii/S1571066104808466 (cit. on p. 179).

[25] Alin Deutsch, Alan Nash and Jeffrey B. Remmel. 'The chase revisited'. In: *Proceedings of PODS'08*. 2008, pp. 149–158. DOI: 10.1145/1376916.1376938 (cit. on pp. 27, 33, 35, 234, 235).

[26] Max Dickmann, Niels Schwartz and Marcus Tressl. *Spectral Spaces*. Vol. 35. Cambridge University Press, 2019. DOI: 10.1017/9781316543870.003 (cit. on pp. 56, 146–148, 174, 239, 253).

[27] Guoli Ding. 'Subgraphs and well-quasi-ordering'. In: *Journal of Graph Theory* 16 (1992), pp. 489–502. DOI: 10.1002/jgt.3190160509 (cit. on pp. 37, 40, 44, 45, 122).

[28] Gaëtan Douéneau-Tabot and Aliaume Lopez. 'Rejoindre l'administration publique après un doctorat'. 2022. URL: https://www.irif.fr/~alopez/posts/2022-11-29-presentation-dptinfo.html (cit. on p. 7).

[29] Gaëtan Douéneau-Tabot, Aliaume Lopez and Laurent Prosperi. 'Recrutement et emploi des docteurs dans les administrations publiques'. 2021. URL: https://www.irif.fr/~alopez/ressources/docteurs-admin (cit. on p. 7).

[30] Jan Dreier, Nikolas Mählmann, Sebastian Siebertz and Szymon Toruńczyk. *Indiscernibles and Wideness in Monadically Stable and Monadically NIP Classes*. 2022. DOI: 10.48550/arXiv.2206.13765. URL: https://arxiv.org/abs/2206.13765 (cit. on p. 233).

[31] Heinz-Dieter Ebbinghaus and Jörg Flum. *Finite Model Theory*. Springer, 1995 (cit. on p. 11).

[32] William Ewald. 'The Emergence of First-Order Logic'. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Metaphysics Research Lab, Stanford University, 2019. URL: https://plato.stanford.edu/archives/spr2019/entries/logic-firstorder-emergence/ (visited on 07/11/2022) (cit. on p. 13).

[33] Solomon Feferman and Robert Vaught. 'The first order properties of products of algebraic systems'. In: *Fundamenta Mathematicae* 47 (1959), pp. 57–103. DOI: 10.4064/fm-47-1-57-103 (cit. on pp. 78, 155).

[34] Alain Finkel and Jean Goubault-Larrecq. 'Forward analysis for WSTS, part I: completions'. In: *Mathematical Structures in Computer Science* 30 (2020), pp. 752–832. DOI: 10.1017/S0960129520000195. (Visited on 13/10/2022) (cit. on pp. 185, 201).

[35] Alain Finkel and Philippe Schnoebelen. 'Well-structured transition systems everywhere!' In: *Theoretical Computer Science* 256 (2001), pp. 63–92. DOI: 10.1016/S0304-3975(00)00102-X. URL: https://linkinghub.elsevier.com/retrieve/pii/S030439750000102X (visited on 19/06/2023) (cit. on p. 44).

[36] Anton Freund. 'From Kruskal's theorem to Friedman's gap condition'. In: *Mathematical Structures in Computer Science* 30 (2020), pp. 952–975. DOI: 10.1017/s0960129520000298 (cit. on pp. 179, 184, 229).

[37] Haim Gaifman. 'On local and non-local properties'. In: *Proceedings of the Herbrand Symposium*. Vol. 107. Elsevier, 1982, pp. 105–135. DOI: 10.1016/S0049-237X(08)71879-2 (cit. on pp. 72, 73).

[38] Jakub Gajarsky, Maximilian Gorsky and Stephan Kreutzer. 'Differential Games, Locality, and Model Checking for FO Logic of Graphs'. In: 216 (2022). Ed. by Florin Manea and Alex Simpson, 22:1–22:18. DOI: 10.4230/LIPIcs.CSL.2022.22 (cit. on pp. 71, 233).

[39] Jakub Gajarský, Nikolas Mählmann, Rose McCarty, Pierre Ohlmann, Michał Pilipczuk, Wojciech Przybyszewski, Sebastian Siebertz, Marek Sokołowski and Szymon Toruńczyk. *Flipper games for monadically stable graph classes*. 2023. DOI: 10.48550/arXiv.2301.13735. URL: https://arxiv.org/abs/2301.13735 (cit. on p. 233).

[40] Jean H. Gallier. 'Ann. Pure Appl. Logic: Erratum to "What's so special about Kruskal's Theorem and the ordinal Γ_0? A survey of some results in proof theory" [53 (1991) 199–260]'. In: *Annals of Pure and Applied Logic* 89 (1997), p. 275. DOI: 10.1016/S0168-0072(97)00043-2 (cit. on pp. 178, 179).

[41] Robert Ganian, Petr Hlinený, Jaroslav Nešetřil, Jan Obdrzálek and Patrice Ossona de Mendez. 'Shrub-depth: Capturing Height of Dense Graphs'. In: *CoRR* abs/1707.00359 (2017). URL: http://arxiv.org/abs/1707.00359 (cit. on p. 259).

[42] Jean-Yves Girard. 'Π12-logic, Part 1: Dilators'. In: *Annals of Mathematical Logic* 21 (1981), pp. 75–219. DOI: 10.1016/0003-4843(81)90016-4 (cit. on p. 229).

[43] Kurt Gödel. 'Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme'. In: *Monatshefte für Mathematik und Physik* 38 (1931), pp. 173–198. DOI: 10.1007/BF01700692 (cit. on p. 18).

[44] Jean Goubault-Larrecq. 'Noetherian Spaces in Verification'. In: *Automata, Languages and Programming*. Ed. by Samson Abramsky, Cyril Gavoille, Claude Kirchner, Friedhelm Meyer auf der Heide and Paul G. Spirakis. Springer Berlin Heidelberg, 2010, pp. 2–21. DOI: 10.1007/978-3-642-14162-1_2 (cit. on pp. 53, 139, 178).

[45] Jean Goubault-Larrecq. *Non-Hausdorff Topology and Domain Theory*. Vol. 22. Cambridge University Press, 2013. DOI: 10.1017/cbo9781139524438 (cit. on pp. 3, 50, 52–54, 139, 159, 178, 179, 185, 186, 192, 193, 203, 208, 224, 239, 254, 255).

[46] Jean Goubault-Larrecq. *Aliaume Lopez' master theorem of Noetherian spaces*. 2022. URL: http://projects.lsv.ens-cachan.fr/topology/?p=6105 (visited on 08/02/2023) (cit. on p. 191).

[47] Jean Goubault-Larrecq. 'Infinitary Noetherian Constructions I. Infinite Words'. In: *Colloquium Mathematicum* 168 (2022), pp. 257–268. DOI: https://doi.org/10.4064/cm8077-4-2021 (cit. on pp. 139, 175, 178, 179, 209–211).

[48] Jean Goubault-Larrecq. 'Non-Hausdorff Topology and Domain Theory. Electronic supplements to the book – Errata'. 2022. URL: https://projects.lsv.ens-cachan.fr/topology/?page_id=12 (visited on 10/12/2022) (cit. on pp. 178, 179).

[49] Jean Goubault-Larrecq, Simon Halfon and Aliaume Lopez. *Infinitary Noetherian Constructions II. Transfinite Words and the Regular Subword Topology*. 2022 (cit. on pp. 175, 178, 179, 209, 211).

[50] Jean Goubault-Larrecq and Bastien Laboureix. 'Statures and Sobrification Ranks of Noetherian Spaces'. In: 2023, pp. 1–76. URL: https://arxiv.org/abs/2112.06828 (cit. on p. 200).

[51] Jean Goubault-Larrecq and Aliaume Lopez. 'On the word topology, and beyond'. 2020. URL: https://projects.lsv.ens-cachan.fr/topology/?page_id=2735 (visited on 08/02/2023) (cit. on p. 177).

[52] Jean Goubault-Larrecq and Sylvain Schmitz. 'Deciding Piecewise Testable Separability for Regular Tree Languages'. In: *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*. Ed. by Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani and Davide Sangiorgi. Vol. 55. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016, 97:1–97:15. DOI: 10.4230/LIPIcs.ICALP.2016.97. URL: http://drops.dagstuhl.de/opus/volltexte/2016/6232 (visited on 02/02/2023) (cit. on pp. 3, 228).

[53] Jean Goubault-Larrecq, Monika Seisenberger, Victor L. Selivanov and Andreas Weiermann. 'Well Quasi-Orders in Computer Science (Dagstuhl Seminar 16031)'. In: *Dagstuhl Reports* 6 (2016), pp. 69–98. DOI: 10.4230/DagRep.6.1.69. URL: https://doi.org/10.4230/DagRep.6.1.69 (cit. on pp. 44, 178).

[54] Erich Grädel, Phokion G Kolaitis, Leonid Libkin, Maarten Marx, Joel Spencer, Moshe Y Vardi, Yde Venema and Scott Weinstein. *Finite Model Theory and its Applications*. Springer, 2007 (cit. on p. 11).

[55] Martin Grohe and Stefan Wöhrle. 'An existential locality theorem'. In: *Annals of Pure and Applied Logic* 129 (2004), pp. 131–148. DOI: 10.1016/j.apal.2004.01.005 (cit. on pp. 67, 68, 75, 76, 85, 93, 113).

[56] Yuri Gurevich. 'Toward logic tailored for computational complexity'. In: *Computation and Proof Theory, Proceedings of LC'84*. Vol. 1104. Springer, 1984, pp. 175–216. DOI: 10.1007/BFb0099486 (cit. on p. 35).

[57] Frederik Harwath, Lucas Heimberg and Nicole Schweikardt. 'Preservation and decomposition theorems for bounded degree structures'. In: *Logical Methods in Computer Science* Volume 11, Issue 4 (2015). DOI: 10.2168/LMCS-11(4:17)2015. URL: https://lmcs.episciences.org/1618 (cit. on p. 74).

[58] Ryu Hasegawa. 'Two applications of analytic functors'. In: *Theoretical Computer Science* 272 (2002), pp. 113–175. DOI: 10.1016/S0304-3975(00)00349-2. URL: https://www.sciencedirect.com/science/article/pii/S0304397500003492 (cit. on pp. 179, 184, 207, 215, 220, 221, 224, 225, 227, 229).

[59] Graham Higman. 'Ordering by divisibility in abstract algebras'. In: *Proceedings of the London Mathematical Society* 3 (1952), pp. 326–336. DOI: 10.1112/plms/s3-2.1.326 (cit. on pp. 44, 158, 178, 181).

[60] Ehud Hrushovski, Joël Ouaknine, Amaury Pouly and James Worrell. 'Polynomial Invariants for Affine Programs'. In: (2018), pp. 530–539. DOI: 10.1145/3209108.3209142. URL: https://dl.acm.org/doi/10.1145/3209108.3209142 (cit. on p. 53).

[61] Peter T Johnstone. *Stone Spaces*. Vol. 3. Cambridge university press, 1982 (cit. on p. 56).

[62] André Joyal. 'Foncteurs analytiques et espèces de structures'. In: *Combinatoire énumérative*. Ed. by Gilbert Labelle and Pierre Leroux. Springer Berlin Heidelberg, 1986, pp. 126–159 (cit. on pp. 215, 239).

[63] J. C. Kelly. 'Bitopological Spaces'. In: *Proceedings of the London Mathematical Society* s3-13 (1963), pp. 71–89. DOI: 10.1112/plms/s3-13.1.71. URL: https://onlinelibrary.wiley.com/doi/abs/10.1112/plms/s3-13.1.71 (visited on 15/11/2022) (cit. on p. 146).

[64] Igor Kříž and Robin Thomas. 'On well-quasi-ordering finite structures with labels'. In: *Graphs and Combinatorics* 6 (1990), pp. 41–49 (cit. on p. 44).

[65] Joseph B. Kruskal. 'The theory of well-quasi-ordering: A frequently discovered concept'. In: *Journal of Combinatorial Theory, Series A* 13 (1972), pp. 297–305. DOI: 10.1016/0097-3165(72)90063-5 (cit. on pp. 44, 158, 178, 179, 181).

[66] Denis Kuperberg. '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, 2021, pp. 1–13. DOI: 10.1109/LICS52264.2021.9470602 (cit. on pp. 37, 39, 68, 98, 113, 137).

[67] Joachim Lambek. 'A Fixpoint Theorem for complete Categories.' In: *Mathematische Zeitschrift* 103 (1968), pp. 151–161. URL: http://eudml.org/doc/170906 (cit. on p. 220).

[68] Leonid Libkin. *Elements of finite model theory*. Springer, 2012. DOI: 10.1007/978-3-662-07003-1 (cit. on pp. 11, 68, 239).

[69] Aliaume Lopez. *On the word topology, and beyond*. 2020. URL: https://www.irif.fr/~alopez/posts/2020-09-23-word-topology-and-beyond.html (visited on 08/02/2023) (cit. on p. 177).

[70] Aliaume Lopez. 'Preservation Theorems Through the Lens of Topology'. In: *Proceedings of CSL'21*. Vol. 183. 2021, 32:1–32:17. DOI: 10.4230/LIPIcs.CSL.2021.32 (cit. on pp. 63, 138, 148, 174, 175).

[71] Aliaume Lopez. 'When Locality Meets Preservation'. In: *LICS'22* (2022), pp. 1–14. DOI: 10.1145/3531130.3532498. URL: https://doi.org/10.1145/3531130.3532498 (cit. on pp. 63, 67, 117).

[72] Aliaume Lopez. 'Fixed Points and Noetherian Topologies'. In: *FOSSACS'23* (2023), pp. 1–14. DOI: 10.1145/3531130.3532498. URL: https://doi.org/10.1145/3531130.3532498 (cit. on pp. 63, 180, 191).

[73] Saunders Mac Lane. *Categories for the Working Mathematician*. Vol. 5. Springer, 1978. DOI: 10.1007/978-1-4757-4721-8. URL: http://link.springer.com/10.1007/978-1-4757-4721-8 (visited on 13/04/2023) (cit. on p. 239).

[74] Johann A. Makowsky. 'Algorithmic uses of the Feferman–Vaught Theorem'. In: *Annals of Pure and Applied Logic* 126 (2004), pp. 159–213. DOI: 10.1016/j.apal.2003.11.002 (cit. on pp. 78, 155).

[75] Eric C Milner. 'Basic wqo-and bqo-theory'. In: *Graphs and order*. Springer, 1985, pp. 487–502 (cit. on p. 53).

[76] Crispin St. John Alvah Nash-Williams. 'On well-quasi-ordering transfinite sequences'. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 61 (1965), pp. 33–39 (cit. on pp. 177–181).

[77] Jaroslav Nešetřil and Patrice Ossona de Mendez. 'On nowhere dense graphs'. In: *European Journal of Combinatorics* 32 (2011), pp. 600–617. DOI: 10.1016/j.ejc.2011.01.006. URL: https://www.sciencedirect.com/science/article/pii/S0195669811000151 (cit. on pp. 46, 121, 259).

[78] Jaroslav Nešetřil and Patrice Ossona de Mendez. *Sparsity: Graphs, Structures, and Algorithms*. Springer, 2012. DOI: 10.1007/978-3-642-27875-4 (cit. on pp. 25, 120).

[79] Oleg Pikhurko and Oleg Verbitsky. 'Descriptive complexity of finite structures: Saving the quantifier rank'. In: *The Journal of Symbolic Logic* 70 (2005), pp. 419–450 (cit. on p. 24).

[80] R. Rado. 'Partial well-ordering of sets of vectors'. In: *Mathematika* 1 (1954), pp. 89–95. DOI: 10.1112/S0025579300000565 (cit. on p. 53).

[81] Eric Rosen. 'Finite model theory and finite variable logics'. PhD thesis. 1995 (cit. on pp. 35, 55).

[82] Eric Rosen. 'Some aspects of model theory and finite structures'. In: *Bulletin of Symbolic Logic* 8 (2002), pp. 380–403. DOI: 10.2178/bsl/1182353894 (cit. on p. 35).

[83] Benjamin Rossman. 'Homomorphism preservation theorems'. In: *Journal of the ACM* 55 (2008), 15:1–15:53. DOI: 10.1145/1379759.1379763 (cit. on pp. i, 24, 35, 37, 38, 40, 42, 55, 74, 175, 177, 235).

[84] Abhisekh Sankaran, Bharat Adsul and Supratik Chakraborty. 'A generalization of the Łoś-Tarski preservation theorem over classes of finite structures'. In: *International Symposium on Mathematical Foundations of Computer Science*. 2014, pp. 474–485. DOI: 10.1007/978-3-662-44522-8_40 (cit. on p. 37).

[85] Abhisekh Sankaran, Bharat Adsul and Supratik Chakraborty. 'A generalization of the Łoś–Tarski preservation theorem'. In: *Annals of Pure and Applied Logic* 167 (2016), pp. 189–210 (cit. on pp. 137, 236).

[86] Abhisekh Sankaran, Bharat Adsul, Vivek Madan, Pritish Kamath and Supratik Chakraborty. 'Preservation under substructures modulo bounded cores'. In: *International Workshop on Logic, Language, Information, and Computation*. 2012, pp. 291–305. DOI: 10.1007/978-3-642-32621-9_22 (cit. on pp. 34, 35).

[87] Thomas Schwentick and Klaus Barthelmann. 'Local normal forms for first-order logic with applications to games and automata'. In: *Discrete Mathematics and Theoretical Computer Science* 3 (1999), pp. 109–124. DOI: 10.46298/dmtcs.254. URL: https://www.dmtcs.org/dmtcs-ojs/index.php/dmtcs/article/view/101/234.html (cit. on pp. 67, 75).

[88] Devendra Singh, Ali Maianguwa Shuaibu and Muhammad Ndayawo. 'Simplified Proof of Kruskal's Tree Theorem'. In: *Mathematical Theory and Modeling* 3 (2013), pp. 93–100. DOI: 10.13140/RG.2.2.12298.39363 (cit. on p. 179).

[89] Alexei P. Stolboushkin. 'Finitely monotone properties'. In: (1995), pp. 324–330. DOI: 10.1109/LICS.1995.523267 (cit. on pp. 35, 113, 137).

[90] William W. Tait. 'A counterexample to a conjecture of Scott and Suppes'. In: *Journal of Symbolic Logic* 24 (1959), pp. 15–16. DOI: 10.2307/2964569 (cit. on pp. i, 35, 37, 39, 97, 100, 113, 137).

[91] Boris Avraamovitch Trakhtenbrot. 'The impossibility of an algorithm for the decision problem for finite classes'. In: *Doklady Akademii Nauk SSSR. Novaya Seriya* 70 (1950), pp. 569–572. DOI: 10.1090/trans2/023 (cit. on p. 18).

[92] Victor Vianu. 'Databases and finite-model theory'. In: *Proceedings of a DIMACS Workshop on Descriptive Complexity and Finite Models*. Ed. by Neil Immerman and Phokion G. Kolaitis. Vol. 31. AMS, 1997, pp. 97–148 (cit. on p. 12).

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

université PARIS-SACLAY

Titre : Théorèmes de préservation pour la logique au premier ordre : localité, topologie et constructions limites.

Mots clés : Théorème de Préservation, Topologie, Localité, Logique au Premier Ordre, Espaces Noethériens.

Résumé :
Les théorèmes de préservation en logique du premier ordre constituent une collection de résultats provenant de la théorie classique des modèles. Ceux-ci établissent une correspondance directe entre les propriétés sémantiques des formules et les contraintes syntaxiques imposées par le langage utilisé pour les exprimer. Cependant, l'étude de ces théorèmes devient particulièrement difficile lorsqu'on se restreint à des modèles finis, ce qui est d'autant plus regrettable que le domaine de la théorie des modèles finis est mieux adaptée pour décrire les phénomènes observés en informatique. Cette thèse propose une approche systématique pour étudier les théorèmes de préservation dans le cadre de la théorie des modèles finis. Les preuves ad-hoc déjà existantes sont comprises dans le cadre d'une théorie plus générale qui englobe les techniques basées sur la localité, reposant sur une présentation topologique des théorèmes de préservation appelée espaces préspectraux logiquement présentés. L'introduction de ces espaces topologiques permet le développement d'une théorie compositionnelle des théorèmes de préservation. De plus, cette thèse constitue une première étape vers l'examen systématique des théorèmes de préservation dans des classes de structures finies définies par induction. Cela est réalisé en démontrant un théorème de point fixe générique pour une classe restreinte d'espaces préspectraux logiquement présentés, plus précisément les espaces Noethériens.

Title : First Order Preservation Theorems in Finite Model Theory : Locality, Topology, and Limit Constructions

Keywords : Preservation Theorem, Topology, Locality, First Order Logic, Noetherian Space.

Abstract : Preservation Theorems in first-order logic are a collection of results derived from classical Model Theory. These results establish a direct correspondence between the semantic properties of formulas and the syntactic constraints imposed on the language used to express them. However, studying these theorems becomes notably challenging when focusing on finite models, which is unfortunate given that the field of Finite Model Theory is better equipped to describe phenomena occurring in Computer Science. This thesis presents a systematic approach to investigating Preservation Theorems within the realm of Finite Model Theory. The traditional ad-hoc proofs are replaced with a theoretical framework that generalizes techniques based on locality, and introduces a topological presentation of preservation theorems called logically presented pre-spectral spaces. Introducing these topological spaces enables the development of a compositional theory for preservation theorems. Additionally, this thesis takes an initial stride towards systematically examining preservation theorems across inductively defined classes of finite structures. It accomplishes this by proving a generic fixed point theorem for a topological restriction of logically presented pre-spectral spaces, specifically Noetherian spaces.