Finite Model Theory

Martin Otto

Winter Term 2008/09

Contents

Part I

Finite vs. Classical Model Theory and The Role of First-Order Logic

Chapter 0

Background and Examples from Classical Model Theory

Preliminary conventions. Vocabularies are typically denoted τ, σ and the like. In general τ may consist of families of relation symbols $(R)_{R\in\tau}$ (of specified arities), function symbols $(f)_{f\in\tau}$ (of specified arities), and constant symbols $(c)_{c\in\tau}$. A τ-structure $\mathfrak{A}$ with universe A is then typically denoted like $\mathfrak{A} = (A, (R^{\mathfrak{A}})_{R\in\tau}, (f^{\mathfrak{A}})_{f\in\tau}, (c^{\mathfrak{A}})_{c\in\tau})$, with $R^{\mathfrak{A}} \subseteq A^n$ if the arity of R is n, $f^{\mathfrak{A}}\colon A^n \to A$ if the arity of f is n, and $c^{\mathfrak{A}} \in A$.

Structures with finite tuples of parameters will often be denoted as in $\mathfrak{A}, \mathbf{a}$, for a tuple $\mathbf{a}$ from A. Notation like $\mathfrak{A} \models \varphi[\mathbf{a}]$ signifies that $\mathfrak{A}$ satisfies φ under the assignment of $\mathbf{a}$ to $\mathbf{x}$; $\mathfrak{A}, \mathbf{a} \models \varphi$ is used interchangeably, with the same meaning.

0.1 Compactness and Expressive Completeness

0.1.1 Embeddings and extensions

Definition 0.1.1 Let $\mathfrak{A}, \mathfrak{B}$ be τ-structures, $h\colon A \to B$ a map.

(i) $h\colon \mathfrak{A} \xrightarrow{\text{hom}} \mathfrak{B}$ is a *homomorphic embedding* if it preserves all atomic formulae $\varphi(\mathbf{x}) \in \mathrm{FO}(\tau)$: f.a. $\mathbf{a}$ in $\mathfrak{A}$: $\mathfrak{A} \models \varphi[\mathbf{a}] \;\Rightarrow\; \mathfrak{B} \models \varphi[h(\mathbf{a})]$.

(ii) $h\colon \mathfrak{A} \xrightarrow{\subseteq} \mathfrak{B}$ is an *isomorphic embedding* if it preserves all atomic and negated atomic (equivalently: all quantifier-free) formulae $\varphi(\mathbf{x}) \in \mathrm{FO}(\tau)$, i.e., if $\mathfrak{A} \models \varphi[\mathbf{a}] \;\Leftrightarrow\; \mathfrak{B} \models \varphi[h(\mathbf{a})]$ for all quantifier-free $\varphi(\mathbf{x}) \in \mathrm{FO}(\tau)$ and $\mathbf{a}$ from $\mathfrak{A}$.

(iii) $h\colon \mathfrak{A} \xrightarrow{\preccurlyeq} \mathfrak{B}$ is an *elementary embedding* if it preserves all formulae $\varphi(\mathbf{x}) \in \mathrm{FO}(\tau)$, i.e., if $\mathfrak{A} \models \varphi[\mathbf{a}] \;\Leftrightarrow\; \mathfrak{B} \models \varphi[h(\mathbf{a})]$ for all $\varphi(\mathbf{x}) \in \mathrm{FO}(\tau)$ and $\mathbf{a}$ from $\mathfrak{A}$.

In the special and important case in which $h = \mathrm{id}_A$ and $A \subseteq B$, we write $\mathfrak{A} \subseteq \mathfrak{B}$ for (ii) and call $\mathfrak{A}$ a substructure of $\mathfrak{B}$ ($\mathfrak{B}$ an extension of $\mathfrak{A}$); and $\mathfrak{A} \preccurlyeq \mathfrak{B}$ for (iii) and call $\mathfrak{A}$ an elementary substructure of $\mathfrak{B}$ ($\mathfrak{B}$ an elementary extension of $\mathfrak{A}$).

NB: the general case of (ii) says that h is an isomorphism between $\mathfrak{A}$ and the substructure $\mathfrak{B} \restriction h(A) \subseteq \mathfrak{B}$; similarly in (iii) we have an isomorphism between $\mathfrak{A}$ and the elementary substructure $\mathfrak{B} \restriction h(A) \preccurlyeq \mathfrak{B}$.

For a τ-structure $\mathfrak{A}$ and some subset $B \subseteq A$, we let τ_B be the extension of τ by new constant symbols for the elements of B (officially we use new constant $c_b \notin \tau$, but we may also identify b itself with its constant symbol where this causes no confusion). Then $\mathfrak{A}_B$ denotes the natural expansion of $\mathfrak{A}$ with these constant names for every $b \in B$. We shall be careless and allow notation like $\varphi(\mathbf{b}) \in \mathrm{FO}(\tau_B)$ for the result of substituting the constants corresponding to $\mathbf{b}$ into $\varphi(\mathbf{x}) \in \mathrm{FO}(\tau)$, such that $\mathfrak{A}_B \models \varphi(\mathbf{b})$ iff $\mathfrak{A} \models \varphi[\mathbf{b}]$.

Definition 0.1.2 (i) The *positive algebraic diagram* of $\mathfrak{A}$ is the set of atomic τ_A-sentences

$$\mathrm{D}^+(\mathfrak{A}) := \{\varphi \in \mathrm{FO}(\tau_A) \colon \varphi \text{ atomic }, \mathfrak{A}_A \models \varphi\}.$$

(ii) The *algebraic diagram* of $\mathfrak{A}$ is the set of quantifier-free τ_A-sentences

$$\mathrm{D}(\mathfrak{A}) := \{\varphi \in \mathrm{FO}(\tau_A) \colon \varphi \text{ quantifier-free }, \mathfrak{A}_A \models \varphi\}.$$

(iii) The *elementary diagram* of $\mathfrak{A}$ is the set of τ_A-sentences

$$\mathrm{E}(\mathfrak{A}) := \{\varphi \in \mathrm{FO}(\tau_A) \colon \mathfrak{A}_A \models \varphi\}.$$

Observation 0.1.3 For a τ_A-structure $\hat{\mathfrak{B}}$ with τ-reduct $\mathfrak{B}$ and for the map

$$\begin{array}{rcl} h \colon A & \longrightarrow & B \\ a & \longmapsto & c_a^{\hat{\mathfrak{B}}} \end{array} :$$

(i) If $\hat{\mathfrak{B}} \models \mathrm{D}^+(\mathfrak{A})$, then $h \colon \mathfrak{A} \xrightarrow{\text{hom}} \mathfrak{B}$ is a homomorphic embedding.

(ii) If $\hat{\mathfrak{B}} \models \mathrm{D}(\mathfrak{A})$, then $h \colon \mathfrak{A} \xrightarrow{\subseteq} \mathfrak{B}$ is an isomorphic embedding.

(iii) If $\hat{\mathfrak{B}} \models \mathrm{E}(\mathfrak{A})$, then $h \colon \mathfrak{A} \xrightarrow{\preccurlyeq} \mathfrak{B}$ is a elementary embedding.

Lemma 0.1.4 *For a τ-structure $\mathfrak{B}$ let $A \subseteq B$. Then A is the universe of an induced elementary substructure $\mathfrak{A} := \mathfrak{B}\restriction A \preccurlyeq \mathfrak{B}$ if, and only if, for all $\varphi(\mathbf{a}, x) \in \mathrm{FO}(\tau_A)$:*

$$\mathfrak{B}_A \models \exists x \varphi(\mathbf{a}, x) \quad \Rightarrow \quad \mathfrak{B}_A \models \varphi(\mathbf{a}, a) \text{ for some } a \in A.$$

This criterion says that witnesses for existential statements in $\mathfrak{B}$ can already be found within $\mathfrak{A}$. To prove sufficiency of the criterion, one establishes by syntactic induction on $\psi(\mathbf{x}) \in \mathrm{FO}(\tau)$ that for all parameters $\mathbf{a}$ from $\mathfrak{A}$: $\mathfrak{B}_A \models \psi(\mathbf{a})$ iff $\mathfrak{A}_A \models \psi(\mathbf{a})$.

0.1.2 Two classical preservation theorems

A formula $\varphi(\mathbf{x}) \in \mathrm{FO}(\tau)$ is *preserved under extensions* if, whenever $h \colon \mathfrak{A} \xrightarrow{\subseteq} \mathfrak{B}$, then $\mathfrak{A} \models \varphi[\mathbf{a}]$ implies $\mathfrak{B} \models \varphi[h(\mathbf{a})]$. Similarly, φ is *preserved under substructures* if in this situation always $\mathfrak{B} \models \varphi[h(\mathbf{a})]$ implies $\mathfrak{A} \models \varphi[\mathbf{a}]$. (Of course one could phrase these in terms of situations $\mathfrak{A} \subseteq \mathfrak{B}$, because any formula is preserved under isomorphism!)

One checks that φ is preserved under extensions if $\neg\varphi$ is preserved under substructures and vice versa.

A formula $\varphi(\mathbf{x}) \in \mathrm{FO}(\tau)$ is *preserved under homomorphisms* if $\mathfrak{A} \models \varphi[\mathbf{a}]$ implies $\mathfrak{B} \models \varphi[h(\mathbf{a})]$ whenever $h \colon \mathfrak{A} \xrightarrow{\text{hom}} \mathfrak{B}$.

An $\mathrm{FO}(\tau)$-formula is existential (resp. universal) if is is generated from atomic and negated atomic formualae with just $\wedge, \vee$ and $\exists$-quantification (resp. $\forall$-quantification). If w.l.o.g. we restrict attention to prenex normal form fromulae, then existential (resp. universal) formulae have the form $\exists \mathbf{x} \xi$ (resp. $\forall \mathbf{x} \xi$) with quantifier-free ξ.

An $\mathrm{FO}(\tau)$-formula is existential positive if is generated from atomic formualae with just $\wedge, \vee$ and $\exists$-quantification.

The following are two typical classical results on *preservation* and *expressive completeness*. The preservation assertions in these statements are straightforward: as is very easy to see by syntactic induction, existential formulae are preserved under extensions and existential positive formulae are preserved under homomorphisms. The corresponding, non-trivial expressive completeness assertions state that, conversely, up to logical equivalence these syntactic forms can always be achieved.

Theorem 0.1.5 (Łos–Tarski) *The following are equivalent for any $\varphi \in \mathrm{FO}(\tau)$:*

(i) *φ is preserved under extensions.*

(ii) *$\varphi \equiv \hat{\varphi}$ for some existential $\hat{\varphi} \in \mathrm{FO}(\tau)$.*

The obviously equivalent, dual assertion similarly links preservation under substructures and universal formalisability.

Theorem 0.1.6 (Lyndon–Tarski) *The following are equivalent for any $\varphi \in \mathrm{FO}(\tau)$:*

(i) *φ is preserved under homomorphisms.*

(ii) *$\varphi \equiv \hat{\varphi}$ for some positive existential $\hat{\varphi} \in \mathrm{FO}(\tau)$.*

Proof We indicate the proof of the crucial expressive completeness part for the first theorem, in the alternative version that preservation under substructures implies equivalence to some universal formula. For notational simplicity (and almost w.l.o.g.) we treat the case of a sentence. So, let φ be an $\mathrm{FO}(\tau)$-sentence preserved under substructures. Put

$$\Psi := \big\{ \psi \in \mathrm{FO}(\tau) \colon \varphi \models \psi, \psi \text{ universal } \big\}.$$

Clearly $\varphi \models \Psi$. We want to show that also

$$\Psi \models \varphi. \qquad (*)$$

This implies, by compactness, that $\Psi_0 \models \varphi$ for some finite subset Ψ_0 of universal formulae from Ψ. But then $\hat{\varphi} := \bigwedge \Psi_0 \equiv \varphi$ is as desired.

Towards $(*)$, let $\mathfrak{A} \models \Psi$. We need to show that $\mathfrak{A} \models \varphi$. We claim that

$$\mathrm{D}(\mathfrak{A}) \cup \{\varphi\}$$

is satisfiable. Otherwise (because of compactness and as $\mathrm{D}(\mathfrak{A})$ is closed under conjunctions), there would have to be some $\xi(\mathbf{a}) \in \mathrm{D}(\mathfrak{A})$ for which $\varphi \models \neg\xi(\mathbf{a})$. As (the constant symbols associated to) $\mathbf{a}$ do not occur in φ, this moreover implies that $\varphi \models \forall \mathbf{x} \neg \xi(\mathbf{x})$. But this latter formula is universal, and thus in Ψ. But this contradicts $\mathfrak{A} \models \Psi$ since clearly $\mathfrak{A} \models \exists \mathbf{x} \xi(\mathbf{x})$ if $\xi(\mathbf{a}) \in \mathrm{D}(\mathfrak{A})$.

So there is some model $\hat{\mathfrak{B}} \models \mathrm{D}(\mathfrak{A}) \cup \{\varphi\}$. We let $\mathfrak{B}$ be the τ-reduct of this τ_A-structure $\hat{\mathfrak{B}}$. By Observation 0.1.3, $\mathfrak{A}$ is isomorphic to a substructure of $\mathfrak{B}$. As $\mathfrak{B} \models \varphi$, $\mathfrak{A} \models \varphi$ follows by preservation of φ under substructures. $\Box$

Exercise 0.1.7 Provide a similar proof for the Lyndon–Tarski theorem.

Exercise 0.1.8 Let $\Delta \subseteq \mathrm{FO}(\tau)$ some (syntactic) fragment of $\mathrm{FO}(\tau)$ that is closed under $\wedge$ and $\vee$. Let us define the Δ *transfer relationship*

$$\mathfrak{A} \Longrightarrow_{\Delta} \mathfrak{B}$$

between τ-structures $\mathfrak{A}$ and $\mathfrak{B}$ by the condition that

$$\mathfrak{A} \models \psi \quad \Rightarrow \quad \mathfrak{B} \models \psi$$

for all $\mathrm{FO}(\tau)$-sentences $\psi \in \Delta$ (we could similarly treat formulae and structures with parameter tuples). Using compactness, show that the following are equivalent, for every $\mathrm{FO}(\tau)$-sentence φ:

(i) φ is preserved under $\Longrightarrow_\Delta$.

(ii) $\varphi \equiv \hat{\varphi}$ for some $\hat{\varphi} \in \Delta$.

Hint: For new unary predicates U, V, say, consider relativisations $[\xi]^U$ and $[\xi]^V$ of sentences $\xi \in \Delta$, which assert that ξ holds true in the substructures with universes U and V, respectively. The Δ transfer relationship between such τ-structures induced on the U- and V-parts of any suitable $\tau \cup \{U, V\}$-structure is then captured by the set of implications

$$\Phi^{\Rightarrow\Delta}_{U,V} := \{[\xi]^U \to [\xi]^V : \xi \in \Delta \cap \mathrm{FO}(\tau)\}.$$

Exercise 0.1.9 Use the previous exercise to give an alternative proof of the Łos–Tarski theorem, based on the following auxiliary assertion, where $\Longrightarrow_\exists$ stands for the transfer relationship w.r.t. existential sentences:

$$\varphi \text{ preserved under extensions} \quad \Rightarrow \quad \varphi \text{ preserved under } \Longrightarrow_\exists .$$

This can be established by showing that every relationship of the form $\mathfrak{A} \Longrightarrow_\exists \mathfrak{B}$ can be *upgraded* to an ismorphic embedding of $\mathfrak{A}$ into some elementary extension of $\hat{\mathfrak{B}} \succcurlyeq \mathfrak{B}$:

$$\begin{tikzcd}
\mathfrak{A} \arrow[r, dash, "\Longrightarrow_\exists"] \arrow[dr, "\subseteq"'] & \mathfrak{B} \arrow[d, "\preccurlyeq"] \\
 & \hat{\mathfrak{B}}
\end{tikzcd}$$

For that it suffices to establish the satisfiability of $\mathrm{D}(\mathfrak{A}) \cup \mathrm{E}(\mathfrak{B})$ where we may w.l.o.g. assume that $A \cap B = \emptyset$.

0.2 Compactness, Elementary Chains, and Interpolation

0.2.1 Elementary chains and the Tarski union property

Definition 0.2.1 A family of τ-structures $(\mathfrak{A}_i)_{i \in \mathbb{N}}$ forms a *chain* if $\mathfrak{A}_i \subseteq \mathfrak{A}_{i+1}$ for all $i \in \mathbb{N}$. It forms an *elementary chain* if $\mathfrak{A}_i \preccurlyeq \mathfrak{A}_{i+1}$ for all $i \in \mathbb{N}$.

The *limit* of a chain $(\mathfrak{A}_i)_{i \in \mathbb{N}}$ of τ-structures $\mathfrak{A}_i$ is the τ-structure

$$\mathfrak{A} := \bigcup_{i \in \mathbb{N}} \mathfrak{A}_i$$

with universe $A := \bigcup_i A_i$ und τ-interpretation according to:

$$\begin{aligned} c^{\mathfrak{A}} &:= c^{\mathfrak{A}_i} \ \text{(independent of } i \text{ by } \subseteq); \\ f^{\mathfrak{A}} &:= \textstyle\bigcup_i f^{\mathfrak{A}_i} \ \text{(well-defined due to } \subseteq); \\ \text{and } R^{\mathfrak{A}} &:= \textstyle\bigcup_i R^{\mathfrak{A}_i}. \end{aligned}$$

Theorem 0.2.2 (Tarski union property) *If $(\mathfrak{A}_i)_{i \in \mathbb{N}}$ is an elementary chain of τ-structures with limit $\mathfrak{A} = \bigcup_i \mathfrak{A}_i$, then $\mathfrak{A}_i \preccurlyeq \mathfrak{A}$ for all $i \in \mathbb{N}$.*

$\mathfrak{A}_i \preccurlyeq \mathfrak{A}$ implies in particular that $\mathfrak{A} \equiv \mathfrak{A}_i$ for all i. This is not generally true for chains of elementarily equivalent structures $\mathfrak{A}_i$.

Exercise 0.2.3 Find a counter example for an analogous statement for chains of elementarily equivalent structures, e.g., within a suitable elementary class of linear orderings like dense linear orderings with a last element.

Exercise 0.2.4 Prove the theorem by induction on $\varphi(\mathbf{x})$, showing that for all $\mathbf{a}$ from $\mathfrak{A}_i$: $\mathfrak{A}_i \models \varphi[\mathbf{a}] \;\Leftrightarrow\; \mathfrak{A} \models \varphi[\mathbf{a}]$.

0.2.2 Robinson consistency property

Recall that a set of sentences $\Phi \subseteq \mathrm{FO}(\tau)$ is called complete (a complete $\mathrm{FO}(\tau)$-theory) if for every $\mathrm{FO}(\tau)$-sentence φ: $\Phi \models \varphi$ or $\Phi \models \neg\varphi$.

Theorem 0.2.5 (Robinson consistency theorem) *Let $\Phi_i \subseteq \mathrm{FO}(\tau_i)$ be sets of sentences in different vocabularies τ_i for $i = 0, 1, 2$, where $\tau_0 = \tau_1 \cap \tau_2$, and such that*

- *Φ_0 is complete w.r.t.* $\mathrm{FO}(\tau_0)$, *and*
- *$\Phi_1 \supseteq \Phi_0$ and $\Phi_2 \supseteq \Phi_0$ are both satisfiable.*

Then $\Phi_1 \cup \Phi_2$ is also satisfiable.

Proof The proof is by a nice chain construction. The idea is to start from a model of Φ_0 (a τ_0-structure) and then alternatingly expand and extend to models of Φ_1 and Φ_2 in such a way that we get two interleaving elementary chains:

- one an elementary chain of τ_1-structures satisfying Φ_1,
- the other an elementary chain of τ_2-structures satisfying Φ_2,

whose limits, however, share the same τ_0-reduct. This limit will thus carry $\tau_1 \cup \tau_2$-structure and be a model of $\Phi_1 \cup \Phi_2$.

For the situation at the start of the chain, we first observe that for any τ_1-structure $\mathfrak{A}_1 \models \Phi_1$ we have $\mathfrak{A} := \mathfrak{A}_1 \restriction \tau_0 \models \Phi_0$ and that this τ_0-structure admits a τ_0-elementary extension/expansion to a τ_2-structure $\mathfrak{B} \models \Phi_2$. I.e., there is some $h \colon \mathfrak{A} \xrightarrow{\preccurlyeq_0} \mathfrak{B} \models \Phi_2$, where $\preccurlyeq_0$ stands for a τ_0-elementary embedding.

By a variant of Observation 0.1.3 it is easily seen that any model of $\mathrm{E}(\mathfrak{A}) \cup \Phi_2$ will give rise to such h and $\mathfrak{B}$. We claim that $\mathrm{E}(\mathfrak{A}) \cup \Phi_2$ is satisfiable. Otherwise, by compactness and since $E(\mathfrak{A})$ is closed under conjunction, there would have to be some $\varphi(\mathbf{a}) \in \mathrm{E}(\mathfrak{A})$ such that $\Phi_2 \models \neg\varphi(\mathbf{a})$. Since $\mathbf{a}$ does not occur in Φ_2, $\Phi_2 \models \forall\mathbf{x}\neg\varphi(\mathbf{x})$. The latter sentence is in $\mathrm{FO}(\tau_0)$. Hence, by completeness of Φ_0 and consistency with Φ_2, $\Phi_0 \models \forall\mathbf{x}\neg\varphi(\mathbf{x})$. But this contradicts $\mathfrak{A} \models \Phi_0$ ($\mathfrak{A} \models \varphi[\mathbf{a}]$ since $\varphi(\mathbf{a}) \in \mathrm{D}(\mathfrak{A})$).

For the typical extension/expansion step in the chain, consider, e.g., a situation in which $\mathfrak{A} \preccurlyeq_0 \mathfrak{B}$ for τ_1-structure $\mathfrak{A} \models \Phi_1$ and τ_2-structure $\mathfrak{B} \models \Phi_2$. We want to find another τ_1-structure $\mathfrak{A}'$ such that the following diagram of $\preccurlyeq_0$- and $\preccurlyeq_2$-embeddings commutes:

$$\begin{tikzcd} \mathfrak{A} \arrow[rr, "\preccurlyeq_1"] \arrow[dr, "\preccurlyeq_0"'] & & \mathfrak{A}' \\ & \mathfrak{B} \arrow[ur, "\preccurlyeq_0"'] & \end{tikzcd}$$

We assume that $\mathfrak{A} \preccurlyeq_0 \mathfrak{B}$, $A \subseteq B$, and hence $(\mathfrak{B} \restriction \tau_0)_A \models \mathrm{E}(\mathfrak{A} \restriction \tau_0)$. We then obtain a structure $\mathfrak{A}'$ as required from any model of

$$\mathrm{E}(\mathfrak{B} \restriction \tau_0) \cup \mathrm{E}(\mathfrak{A}).$$

Note that $\mathrm{E}(\mathfrak{A})$ is a diagram w.r.t. τ_1. We need to argue that this union is satisfiable. Otherwise, by compactness, there are

- $\varphi_1(\mathbf{a}) \in \mathrm{E}(\mathfrak{A})$ (with contants for elements $\mathbf{a}$ from A),
- $\varphi_0(\mathbf{a}, \mathbf{b}) \in \mathrm{E}(\mathfrak{B} \restriction \tau_0)$ (with the same $\mathbf{a}$ and in addition $\mathbf{b}$ from $B \setminus A$),

such that $\mathfrak{A}_A \models \varphi_1(\mathbf{a})$ and $\mathfrak{B}_B \models \varphi_0(\mathbf{a}, \mathbf{b})$, but

$$\varphi_1(\mathbf{a}) \models \neg\varphi_0(\mathbf{a}, \mathbf{b}).$$

As $\mathbf{b}$ does not occur in $\varphi_1(\mathbf{a})$, we find

$$\varphi_1(\mathbf{a}) \models \neg\exists\mathbf{y}\varphi_0(\mathbf{a},\mathbf{y}),$$

and observe that $\neg\exists\mathbf{y}\varphi_0(\mathbf{a},\mathbf{y}) \in \mathrm{FO}((\tau_0)_A)$. As $\mathfrak{A}_A \models \varphi_1(\mathbf{a})$, therefore, $\mathfrak{A}_A \models \neg\exists\mathbf{y}\varphi_0(\mathbf{a},\mathbf{y})$. As $\mathfrak{A} \preccurlyeq_0 \mathfrak{B}$, $\mathfrak{B}_A \models \neg\exists\mathbf{y}\varphi_0(\mathbf{a},\mathbf{y})$, too. But this contradicts $\mathfrak{B}_B \models \varphi_0(\mathbf{a},\mathbf{b})$.

An extension/expansion pattern is similarly established for the symmetric situation of $\mathfrak{B} \preccurlyeq_0 \mathfrak{A}$ for τ_2-structure $\mathfrak{B} \models \Phi_2$ and τ_1-structure $\mathfrak{A} \models \Phi_1$, in which we seek a τ_2-structure $\mathfrak{B}'$ such that following diagram commutes:

$$\begin{tikzcd}
& \mathfrak{A} \arrow[dr, "\preccurlyeq_0"] & \\
\mathfrak{B} \arrow[ur, "\preccurlyeq_0"] \arrow[rr, "\preccurlyeq_2"'] & & \mathfrak{B}'
\end{tikzcd}$$

We thus inductively obtain interleaving chains as follows:

$$\begin{tikzcd}
\mathfrak{A}_1 \arrow[rr, "\preccurlyeq_1"] \arrow[dr, "\preccurlyeq_0"'] & & \mathfrak{A}_2 \arrow[rr, "\preccurlyeq_1"] \arrow[dr, "\preccurlyeq_0"'] & & \cdots & \cdots \arrow[r] & \mathfrak{A} \models \Phi_1 & \tau_1\text{-limit} \\
& \mathfrak{B}_1 \arrow[ur, "\preccurlyeq_0"] \arrow[rr, "\preccurlyeq_2"'] & & \mathfrak{B}_2 \arrow[ur, "\preccurlyeq_0"] \arrow[rr, "\preccurlyeq_2"'] & & \cdots \arrow[r] & \mathfrak{B} \models \Phi_2 & \tau_2\text{-limit}
\end{tikzcd}$$

As $\mathfrak{A}{\restriction}\tau_0 = \mathfrak{B}{\restriction}\tau_0$, the superposition of the τ_1 and τ_2-structures on this common universe $A = B$ of the limits is compatible with the τ_0-parts, and provides the desired model of $\Phi_1 \cup \Phi_2$. $\square$

0.2.3 Interpolation and Beth's Theorem

An *interpolant* for an implication $\varphi \models \psi$ is an intermediate formula χ such that $\varphi \models \chi$ and $\chi \models \psi$. Craig's interpolation theorem asserts the existence of interpolants which only use symbols that occur in both φ and ψ: a bottle-neck w.r.t. vocabulary that has interesting consequences for definability issues like Beth's theorem, Theorem 0.2.8 below.

Theorem 0.2.6 (Craig interpolation theorem) *Let $\varphi \in \mathrm{FO}(\tau_1)$, $\psi \in \mathrm{FO}(\tau_2)$ be sentences such that $\varphi \models \psi$. Then this implication has an interpolant in the common vocabulary: there is some sentence $\chi \in \mathrm{FO}(\tau_1 \cap \tau_2)$ such that*

$$\varphi \models \chi \quad \textit{and} \quad \chi \models \psi.$$

Proof By reduction to the Robinson consistency property of FO. Suppose to the contrary that there were no such interpolant. W.l.o.g. τ_1 and τ_2 are finite so that the set of all $\mathrm{FO}(\tau_0)$-sentences can be enumerated as $(\varphi_n)_{n\in\mathbb{N}}$. We want to construct a complete $\mathrm{FO}(\tau_0)$-theory Φ s.t. $\Phi \cup \{\varphi\}$ and $\Phi \cup \{\neg\psi\}$ are both satisfiable. An application of Theorem 0.2.5 then yields a model of $\Phi \cup \{\varphi, \neg\psi\}$, contradicting $\varphi \models \psi$. We obtain Φ as a limit of finite approximations $\Phi := \bigcup_{n\in\mathbb{N}} \Phi_n$. Φ is chosen inductively such that each Φ_{n+1} comprises either φ_n or $\neg\varphi_n$ (this guarantees completeness for Φ), and such that the following is maintained for all n:

$$\text{there is no } \chi \in \mathrm{FO}(\tau_0) \text{ s.t.} \quad \Phi_n \models (\varphi \to \chi) \wedge (\chi \to \psi). \quad (*)$$

This invariant corresponds to failure of interpolation relative to every Φ_n, and hence, by compactness, eventually also relative to Φ. We start with $\Phi_0 := \emptyset$ so that $(*)$ is true by assumption. Suppose $(*)$ holds for Φ_n, but neither for $\Phi_{n+1} = \Phi_n \cup \{\varphi_n\}$ nor for $\Phi_{n+1} = \Phi_n \cup \{\neg\varphi_n\}$. This means that in both these cases we do find interpolants:

$$\begin{array}{rcl} \Phi_n \cup \{\varphi_n\} & \models & (\varphi \to \chi_1) \wedge (\chi_1 \to \psi), \text{ and} \\ \Phi_n \cup \{\neg\varphi_n\} & \models & (\varphi \to \chi_2) \wedge (\chi_2 \to \psi) \end{array}$$

for suitable $\chi_{1/2} \in \mathrm{FO}(\tau_0)$. But then

$$\chi := (\varphi_n \wedge \chi_1) \vee (\neg\varphi_n \wedge \chi_2)$$

violates $(*)$ for Φ_n.

Finally, $\Phi := \bigcup_n \Phi_n$ is as desired: Φ is a complete $\mathrm{FO}(\tau_0)$-theory by the inductive construction. Φ is satisfiable together with φ, since otherwise $\chi := \bot$ would violate $(*)$ for Φ, and hence, by compactness also for some Φ_n. Similarly Φ is satisfiable together with $\neg\psi$, as otherwise $\chi := \top$ would violate $(*)$ for Φ and hence for some Φ_n. □

Compare Proposition 1.2.8 for the failure of interpolation in finite model theory.

For the statement of Beth's theorem recall the following notions.

Definition 0.2.7 Let $R \notin \tau$. A sentence $\varphi = \varphi(R) \in \mathrm{FO}(\tau \cup \{R\})$ *defines relation R implicitly* if each τ-structure $\mathfrak{A}$ has at most one expansion to a $\tau \cup \{R\}$-structure $\hat{\mathfrak{A}} = (\mathfrak{A}, R^{\hat{\mathfrak{A}}})$ such that $\hat{\mathfrak{A}} \models \varphi$.

An explicit definition, on the other hand, is given by a formula $\xi(\mathbf{x}) \in \mathrm{FO}(\tau)$ that defines R (as a global relation across all τ-structures $\mathfrak{A}$) via $R^{\mathfrak{A}} := \{\mathbf{a} \in A^n : \mathfrak{A} \models \xi[\mathbf{a}]\}$. Coming from an implicit definition $\varphi(R)$, we say that ξ defines that same relation implicitly if

$$\varphi(R) \models \forall \mathbf{x}(R\mathbf{x} \leftrightarrow \xi(\mathbf{x})).$$

Clearly, explicit definability entails implicit definability. The converse, that implicit definitions can be eliminated in favour of explicit ones, is regarded as an indicator of a good balance between syntax and semantics.

Theorem 0.2.8 (Beth's definability theorem) *Every implicit* FO*-definition admits an explicit* FO*-definition.*

Proof Suppose $\varphi(R)$ defines R implicitly. Let $\varphi(R')$ be the result of exchanging R for a new relation symbol R' of the same arity. Then, by the nature of implicit definitions,

$$\varphi(R) \wedge \varphi(R') \models \forall \mathbf{x}(R\mathbf{x} \leftrightarrow R'\mathbf{x}),$$

or, for new constants $\mathbf{c}$:

$$\varphi(R) \wedge R\mathbf{c} \models \varphi(R') \to R'\mathbf{c}.$$

Let $\chi(\mathbf{c}) \in \mathrm{FO}(\tau \cup \{\mathbf{c}\})$ be an interpolant for this last implication, where we display the new constants explicitly as substituted into some underlying $\chi(\mathbf{x}) \in \mathrm{FO}(\tau)$. Note that χ, being an interpolant in the common language, does not mention R or R'. Now $\chi(\mathbf{x})$ is the desired explicit definition:

$$\begin{array}{rcl} \varphi(R) \wedge R\mathbf{c} \models \chi(\mathbf{c}) & \Rightarrow & \varphi(R) \models \forall \mathbf{x}(R\mathbf{x} \to \chi(\mathbf{x})), \text{ and} \\ \chi(\mathbf{c}) \models \varphi(R) \to R\mathbf{c} & \Rightarrow & \varphi(R) \models \forall \mathbf{x}(\chi(\mathbf{x}) \to R\mathbf{x}). \end{array}$$

□

0.2.4 Order-invariant definability

Let $<$ be a binary relation symbol (for a linear ordering) not in τ, $\mathcal{C}$ a class of τ-structures. With $\mathcal{C}$ associate the class of all linearly ordered versions of structures form $\mathcal{C}$, considered as a class of $\tau_<$-structures for $\tau_< := \tau \cup \{<\}$.

$$\mathcal{C}_< := \big\{ \hat{\mathfrak{A}} = (\mathfrak{A}, <^{\hat{\mathfrak{A}}}) \colon \mathfrak{A} \in \mathcal{C}, <^{\hat{\mathfrak{A}}} \text{ any linear ordering of } A \big\}.$$

We say that an $\mathrm{FO}(\tau \cup \{<\})$-sentence $\varphi(<)$ defines the class $\mathcal{C}$ of τ-structures in an *order-invariant* manner, if

$$\mathrm{Mod}(\varphi) = \mathcal{C}_<.$$

Note that this means that, while φ may refer to the linear ordering $<$, the truth or falsity of $(\mathfrak{A}, <^{\hat{\mathfrak{A}}}) \models \varphi$ must by definition be independent of the chosen ordering $<^{\hat{\mathfrak{A}}}$ of A and must only depend on the τ-reduct $\mathfrak{A}$.

If $<_1^{\hat{\mathfrak{A}}}$ and $<_2^{\hat{\mathfrak{A}}}$ are two linear orderings of the same τ-structure $\mathfrak{A}$, then $(\mathfrak{A}, <_1^{\hat{\mathfrak{A}}}) \models \varphi$ iff $(\mathfrak{A}, <_2^{\hat{\mathfrak{A}}}) \models \varphi$. We model the situation with two distinct binary relations symbols $<$ and $<'$, both not in τ. Let $\xi_{\text{lin ord}}(<) \in \mathrm{FO}(<)$ be a sentence saying that $<$ is a linear ordering of the universe; and similarly $\xi_{\text{lin ord}}(<') \in \mathrm{FO}(<')$ for $<'$. We then capture order-invariance by the validity of the implication

$$\xi_{\text{lin ord}}(<) \wedge \xi_{\text{lin ord}}(<') \models \varphi(<) \leftrightarrow \varphi(<').$$

From this we get

$$\xi_{\text{lin ord}}(<) \wedge \varphi(<) \models \xi_{\text{lin ord}}(<') \rightarrow \varphi(<').$$

It is now easy to see that any interpolant $\chi \in \mathrm{FO}(\tau)$ for this implication defines the class $\mathcal{C}$ without reference to either ordering.

Corollary 0.2.9 *Any class of τ-structures that is $\mathrm{FO}(\tau_<)$-definable in an order-invariant manner is also $\mathrm{FO}(\tau)$-definable without order.*

Exercise 0.2.10 Prove the above by filling in the details in the argument sketched.

We shall see that the same is not true in the sense of finite model theory: there are classes of finite structures that are order-invariantly $\mathrm{FO}(\tau_<)$-definable within the class of all finite structures, but not $\mathrm{FO}(\tau)$-definable within the class of all finite structures (cf. Example 1.2.11).

0.3 A Lindström Theorem

Lindström's first theorem characterises the expressive power of FO as maximal among competitors that satisfy some natural basic properties and two key constituents of first-order model theory: compactness and the Löwenheim–Skolem theorem. Lindström's second theorem similarly characterises FO as maximally expressive among effective logical systems that (instead of the compactness assumption) are recursively enumerable for validity – which serves as an abstract counterpart of the existence of an effective proof system.

We briefly review the key ideas in the proof of Lindsröm's first theorem, because it serves as an excellent example of a meta-level result on expressive completeness, and of a particularly appealing mix of techniques from classical model theory: compactness plus back-and-forth techniques related to Ehrenfeucht–Fraïssé games.

We recall some basic notions. We take a naive stance w.r.t. the concept of an *abstract logic*: a logic $\mathcal{L}$ consists of an association of a set $\mathcal{L}(\tau)$ of $\mathcal{L}$-sentences with every vocabulary τ; its semantics is specified by a satisfaction relation $\models_{\mathcal{L}}$ between τ-structures and $\mathcal{L}(\tau)$-sentences, which tells us when a τ-structure $\mathfrak{A}$ satisfies an $\mathcal{L}(\tau)$-sentence φ. In other words, $\models_{\mathcal{L}}$ just specifies in an abstract, purely extensional manner whether or not $\mathfrak{A} \models_{\mathcal{L}} \varphi$.

We summarise some additional but basic properties of abstract logics $\mathcal{L}$ that will always be assumed in the following discussion.

Definition 0.3.1 (i) $\models_{\mathcal{L}}$ is *invariant under isomorphism*: $\mathfrak{A} \simeq \mathfrak{A}'$ implies that $\mathfrak{A} \models_{\mathcal{L}} \varphi$ iff $\mathfrak{A}' \models_{\mathcal{L}} \varphi$.

(ii) $\mathcal{L}$ is *monotone* w.r.t. vocabularies: if $\tau \subseteq \tau'$, then $\mathcal{L}(\tau) \subseteq \mathcal{L}(\tau')$ and for any τ'-structure $\mathfrak{A}'$ with τ-reduct $\mathfrak{A}$ and for any $\varphi \in \mathcal{L}(\tau)$: $\mathfrak{A}' \models \varphi$ iff $\mathfrak{A} \models \varphi$.

(iii) $\mathcal{L}$ is *closed under renaming*: if τ' is obtained from τ by a renaming of symbols, then for every $\varphi \in \mathcal{L}(\tau)$ there is a corresponding $\varphi' \in \mathcal{L}(\tau')$ such that whenever $\mathfrak{A}'$ is the renaming of a τ-structures $\mathfrak{A}$, then $\mathfrak{A}' \models \varphi'$ iff $\mathfrak{A} \models \varphi$.

(iv) $\mathcal{L}$ is *closed under boolean connectives*: for all $\varphi_1, \varphi_2 \in \mathcal{L}(\tau)$ there are sentences of $\mathcal{L}(\tau)$ whose semantics is that of $\varphi_1 \wedge \varphi_2$, $\varphi_1 \vee \varphi_2$ and $\neg\varphi_1$, respectively.

(v) $\mathcal{L}$ admits *relativisation*: for every $\varphi \in \mathcal{L}(\tau)$ and every unary predicate $U \notin \tau$ there is some sentence $\psi \in \mathcal{L}(\tau \cup \{U\})$ such that for any $(\tau \cup \{U\})$-structure $\hat{\mathfrak{A}} = (\mathfrak{A}, U^{\hat{\mathfrak{A}}})$ for which $U^{\hat{\mathfrak{A}}}$ is the domain of a τ-substructure $\mathfrak{A}{\restriction}U^{\hat{\mathfrak{A}}} \subseteq \mathfrak{A}$, we have $\hat{\mathfrak{A}} \models_{\mathcal{L}} \psi$ iff $\mathfrak{A}{\restriction}U^{\hat{\mathfrak{A}}} \models_{\mathcal{L}} \varphi$.

W.l.o.g. for our concerns, we may adopt the usual first-order notations for derived fromulae as guaranteed by the above (we are just interested in the semantics of $\mathcal{L}$-sentences, not their concrete syntactic representations). In this manner we write, for instance, $\neg\varphi$ and $[\varphi]^U$ and treat these as $\mathcal{L}$-sentences obtained from the $\mathcal{L}$-sentence φ. For a logic $\mathcal{L}$ extending FO, we assume that $\mathrm{FO}(\tau) \subseteq \mathcal{L}(\tau)$ in the sense also of a syntactic inclusion. For such logics we can use the usual first-order definitions to eliminate function and constant symbols in favour of relations, and thus may restrict attention to purely relational vocabularies τ.

We turn to some specific properties which are familiar from FO.

Definition 0.3.2 (i) $\mathcal{L}$ has the *finite occurrence property*, if for every $\varphi \in \mathcal{L}(\tau)$ there is some finite $\tau_0 \subseteq \tau$ such that whether or not $\mathfrak{A} \models \varphi$ only depends on the τ_0-reduct $\mathfrak{A}{\restriction}\tau_0$ (and we may assume that $\varphi \in \mathcal{L}(\tau_0)$).

(ii) $\mathcal{L}$ has the *compactness property*, if a set of $\mathcal{L}$-sentences is satisfiable iff every finite subset is satisfiable.

(iii) $\mathcal{L}$ has the *Löwenheim–Skolem property*, if every countable set of $\mathcal{L}$-sentences that is satisfiable has a countable model.

Remark 0.3.3 *Compactness implies the finite occurrence property.*

Idea: Suppose τ consists of an infinite family of relation symbols $(R_i)_{i \in I}$ and let $\varphi \in \mathcal{L}(\tau)$. For a disjoint renaming of all relation symbols in τ, $\tau' := (R_i')_{i \in I}$, we obtain $\varphi' \in \mathcal{L}(\tau')$ according to the basic renaming property. We may then apply compactness to the implication

$$\big\{\forall \mathbf{x}(R_i\mathbf{x} \leftrightarrow R_i'\mathbf{x})\colon i \in I\big\} \models \varphi \leftrightarrow \varphi'$$

in order to establish that φ semantically only depends on finitely many of the symbols in τ after all.

Theorem 0.3.4 (First Lindström Theorem) *No logic $\mathcal{L}$ that is compact and satisfies the Löwenheim–Skolem property can properly extend* FO*. In other words, any extension of* FO *with these properties is just as expressive and no more expressive than* FO*: for any $\varphi \in \mathcal{L}(\tau)$ there is some $\hat{\varphi} \in \mathrm{FO}(\tau)$ such that $\hat{\varphi} \equiv \varphi$.*

And indeed, some common and natural extensions of FO, like second-order logic or infinitary first-order logic FO_∞ fail on both the crucial properties.

Proof We sketch the core argument that no compact logic $\mathcal{L}$ with Löwenheim–Skolem can properly extend FO. Any $\varphi \in \mathcal{L}(\tau)$ for such a logic must be equivalent to some $\hat{\varphi} \in \mathrm{FO}(\tau)$. We may assume that τ is finite (Remark 0.3.3) and purely relational.

Assuming that $\varphi \in \mathcal{L}(\tau)$ were not equivalent to any $\hat{\varphi} \in \mathrm{FO}(\tau)$ we first construct a complete set of $\mathrm{FO}(\tau)$-sentences Φ such that both $\Phi \cup \{\varphi\}$ and $\Phi \cup \{\neg\varphi\}$ are satisfiable. Such Φ is obtained inductively, from finite approximations along an enumeration of $\mathrm{FO}(\tau)$. One can proceed in complete analogy with the proof of Theorem 0.2.6: this time, the invariant to be inductively maintained for the Φ_n is that, even under Φ_n, φ is not equivalent to any $\hat{\varphi} \in \mathrm{FO}(\tau)$.

Given this complete $\mathrm{FO}(\tau)$-theory Φ, which is compatible with both φ and with $\neg\varphi$, we have τ-structures $\mathfrak{A} \models \Phi \cup \{\varphi\}$ and $\mathfrak{B} \models \Phi \cup \{\neg\varphi\}$. Since Φ is complete, $\mathfrak{A}$ and $\mathfrak{B}$ are elementarily equivalent: $\mathfrak{A} \equiv \mathfrak{B}$. We want to upgrade this to a situation in which corresponding models of φ and $\neg\varphi$ are not just elementarily equivalent, but

(a) in a first step: partially isomorphic: $\mathfrak{A} \simeq_{\mathrm{part}} \mathfrak{B}$. For this we use compactness of $\mathcal{L}$.

(b) in a second step: $\mathfrak{A} \simeq_{\mathrm{part}} \mathfrak{B}$ and both structures countable. For this we use Löwenheim–Skolem for $\mathcal{L}$.

(c) (as an automatic consequence in the situation of (b):) even $\mathfrak{A} \simeq \mathfrak{B}$.

As $\mathfrak{A} \models \varphi$ and $\mathfrak{B} \models \neg\varphi$, the situation in (c) contradicts isomorphism invariance.

Point (a) involves ideas centering on Ehrenfeucht–Fraïssé games and back-and-forth systems (cf. section 2.1).

For (a) recall that two τ-structures $\mathfrak{A}$ and $\mathfrak{B}$ in a relational vocabulary τ are *partially isomorphic*, $\mathfrak{A} \simeq_{\mathrm{part}} \mathfrak{B}$, if there is a system I of local isomorphisms (isomorphisms between induced substructures of $\mathfrak{A}$ and $\mathfrak{B}$) that has the back-and-forth properties:

(*forth*) for every $p \in I$ and every $a \in \mathfrak{A}$ there is some extension $p \subseteq q \in I$ for which $a \in \mathrm{dom}(q)$.

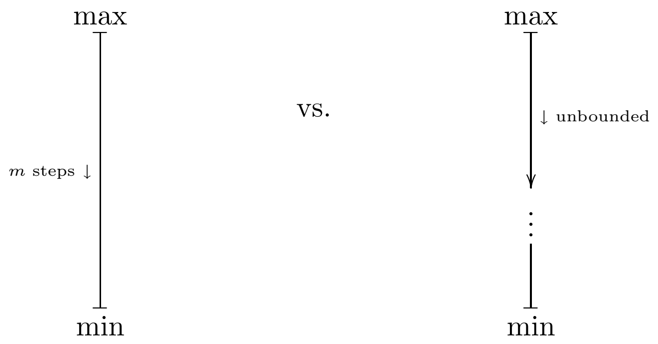
(*back*) for every $p \in I$ and for every $b \in \mathfrak{B}$ there is some extension $p \subseteq q \in I$ for which $b \in \mathrm{range}(q)$.

The Ehrenfeucht–Fraïssé theorem (see Theorem 2.1.7 below) for FO tells us that elementary equivalence between $\mathfrak{A}$ and $\mathfrak{B}$ is similarly captured by the existence of back-and-forth systems of every finite height: $(I_n)_{0 \leqslant n \leqslant m}$ such that, for all $n < m$,

(*forth*) for every $p \in I_{n+1}$ and every $a \in \mathfrak{A}$ there is some extension $p \subseteq q \in I_n$ for which $a \in \mathrm{dom}(q)$.

(*back*) for every $p \in I_{n+1}$ and for every $b \in \mathfrak{B}$ there is some extension $p \subseteq q \in I_n$ for which $b \in \mathrm{range}(q)$.

So the difference between what we have with $\mathfrak{A} \equiv \mathfrak{B}$ on one hand, and the desired situation for (a): $\mathfrak{A} \simeq_{\mathrm{part}} \mathfrak{B}$, on the other hand, is reflected by a just a difference in terms of the layering of the sets of local isomorphisms. For $\equiv$ we index sets of local isomorphisms (linked by back-and-forth requirements) by a linear ordering that allows us to descend finitely many (m many) steps; for $\simeq_{\text{part}}$ we can imagine to descend along an infinite downward chain of copies of the one given set I of local isomorphisms.



On the left, a finite discrete linear ordering allows m predecessor steps from its maximal element max, for some $m \in \mathbb{N}$; on the right an infinite discrete linear ordering allows an infinite sequence of predecessor steps from its maximal element max. In either secnario, we may encode systems of local isomorphisms between $\mathfrak{A}$ and $\mathfrak{B}$ in first-order terms within suitable super-structures that contain

– both $\mathfrak{A}$ and $\mathfrak{B}$ as τ-substructures
– some coding machinery for sets of local isomorphisms (ternary relations for labelled sets of graphs of functions),
– and the index structures to which these sets of local isomorphisms are attached, with predecessor links.

Then the back-and-forth conditions along predecessor links can be expressed as first-order requirements. The difference between arbitrarily long finite index structures (corresponding to the situation of $\mathfrak{A} \equiv \mathfrak{B}$) and an index structure with an infinite predecessor chain at max (corresponding to the situation of $\mathfrak{A} \simeq_{\text{part}} \mathfrak{B}$) is then easily bridged by a compactness argument. Think of forcing a non-standard variant in the class of all finite discrete linear orderings with first and last elements, in which the maximal element has an unbounded chain of predecessors.

To the indicated FO-specification of a super-structure for partially ismorphic $\mathfrak{A}$ and $\mathfrak{B}$, we add suitable relativisations of φ (in the $\mathfrak{A}$-part) and $\neg\varphi$ (in the $\mathfrak{B}$-part) in $\mathcal{L}$.

Applying compactness for $\mathcal{L}$, we get (a); and with Löwenheim–Skolem for $\mathcal{L}$, also (b). The following typical back-and-forth argument then yields (c) and thus clinches the argument.

Proposition 0.3.5 *If $\mathfrak{A} \simeq_{part} \mathfrak{B}$ and both structures are countable, then $\mathfrak{A} \simeq \mathfrak{B}$.*

Proof Let I be a set of local isomorphisms for $\mathfrak{A} \simeq_{\text{part}} \mathfrak{B}$. Let $A = \{a_n : n \in \mathbb{N}\}$ and $B = \{b_n : n \in \mathbb{N}\}$. Inductively obtain a monotone chain of local isomorphisms $(p_n)_{n\in\mathbb{N}}$ within I such that p_{n+1} extends p_n and such that $a_n \in \text{dom}(p_{2n+1})$ and $b_n \in \text{range}(p_{2n+2})$. This is achieved by application of the back-and-forth properties of I. Then the limit $h := \bigcup_n p_n$ is a well-defined local isomorphism between $\mathfrak{A}$ and $\mathfrak{B}$. Since it is also bijective by construction, h is in fact an isomorphism: $h\colon \mathfrak{A} \simeq \mathfrak{B}$. $\Box$

0.4 Classical Inexpressibility via Compactness

That some property of τ-structures is not definable in $\text{FO}(\tau)$ over the class of all (finite and infinite) τ-structures, can often be shown by compactness arguments. Another classical technique for inexpressibility proofs centers on Ehrenfeucht–Fraïsé games: that will concern us much more in this course, because this is one of the few methods from classical model theory that fully translate into finite model theory. See section 2.1 for that.

0.4.1 Some very typical examples

Example 0.4.1 Connectivity of graphs is not FO-definable.

Suppose there were an FO(E)-sentence φ such that, for all graphs $\mathfrak{A} = (A, E^{\mathfrak{A}})$, $\mathfrak{A} \models \varphi$ iff $\mathfrak{A}$ is a connected graph. Consider the following set of FO(E)-formulae:

$$\Phi := \{\varphi\} \cup \{\delta_n(x, y) \colon n \in \mathbb{N}\},$$

where $\delta_n(x, y) \in \mathrm{FO}(E)$ is a formula saying that the distance between x and y is greater than 2^n. Such formulae are easily obtained inductively: let $\delta_0(x, y) := \neg x = y \wedge \neg Exy$ and $\delta_{n+1}(x, y) := \forall z(\delta_n(x, z) \vee \delta_n(z, y))$.

It is easy to see that every finite $\Phi_0 \subseteq \Phi$ is satisfiable. By compactness Φ is satisfiable, but in every model $\mathfrak{A}, a, a' \models \Phi$, the instantiations a and a' of x and y are not at finite distance, whence $\mathfrak{A}$ must be disconnected. Therefore $\mathfrak{A} \models \varphi$ contradicts the assumptions.

Example 0.4.2 The class of fields of characteristic zero is definable by an infinite set of FO-sentences, but not by a single FO-sentence (in the vocabulary $\tau_{\mathrm{ar}} := \{+, \cdot, 0, 1\}$); the class of all fields of characteristic $\neq 0$ is not FO-definable.

Clearly the FO(τ_{ar})-sentence

$$\varphi_p := \neg \underbrace{1 + \cdots + 1}_{p\text{-times}} = 0$$

stipulates that the characteristic cannot be p. The set Φ consisting of the field axioms together with these sentences φ_p for all primes p, therefore, axiomatises the class of fields of characteristic zero.

Compactness shows that no single sentence can suffice for this. In fact, any sentence $\varphi \in \mathrm{FO}(\tau_{\mathrm{ar}})$ for which $\Phi \models \varphi$ is, by compactness, already a consequence of some finite subset $\Phi_0 \subseteq \Phi$. Hence φ is true also in all finite fields of sufficiently large characteristic.

With the statement from the next exercise we see that therefore the class of fields of characteristic $\neq 0$ is not FO-definable, not even by an infinite set of sentences.

Exercise 0.4.3 Show that, whenever a class $\mathcal{C}$ of τ-structures as well as its complement are both axiomatisable by sets of FO(τ)-sentences, then both are in fact axiomatisable by single FO(τ)-sentences. Hint: look at the unsatisfiable union of the two sets of sentences, and apply compactness.

0.4.2 A rather atypical example

Example 0.4.4 Evenness of the length of finite linear orderings in not FO-definable (not even within the class of just finite linear orderings, this is).

This is a typical first example for the use of Ehrenfeucht–Fraïssé games for inexpressibility results in finite as well as classical model theory, see section 2.1, Example 2.1.13.

Surprisingly, there is also a classical, compactness based proof for this.

Suppose there were a sentence $\varphi \in \mathrm{FO}(<)$ such that a *finite* linear ordering satisfies φ if, and only if, it is of even length. Let $\psi(x) := [\varphi]^{\leqslant x}$ be the relativisation of φ to the initial segment formed by x. Let $\psi_0 \in \mathrm{FO}(<)$ be the usual characterisation of discrete linear orderings with first and without last element; $\psi_1 \in \mathrm{FO}(<)$ the assertion that precisely every other element x satisfies $\psi(x) = [\varphi]^{\leqslant x}$.

Then $\psi_0 \wedge \psi_1$ would characterise the order type of $(\omega, <)$, which is impossible by compactness (non-standard models do exist!).

Consider any non-standard model $\mathfrak{A} = (A, <)$ of ψ_0 as in the sketch. Since the non-standard part of $\mathfrak{A}$ consists of an ordered sum of parts ordered like $(\mathbb{Z}, <)$, the successor operation induces an automorphism within each $\mathbb{Z}$-component and hence of $\mathfrak{A}$.

Therefore $\psi(x) = [\varphi]^{\leqslant x}$ cannot distinguish next neighbours within the non-standard part, and $\mathfrak{A} \not\models \psi_0 \wedge \psi_1$.

Exercise 0.4.5 Use the insights from the preceding example to show that the analogues of Beth's definability threoreem and of Craig's interpolation theorem must fail in the sense of finite model theory.

E.g., for $\tau = \{<\}$, let P be the unary predicate that is empty in finite $<$-structures unless $<$ is a linear ordering of the universe; for finite linear orderings, let P consist of the first and every other element w.r.t. $<$. Show that this P is implicitly definable in retsriction to the class of all finite τ-structures, but cannot be explicitly definable over this class. The classical version of Beth's theorem also implies that there is no implicit definition of any P that would extend the given specification to the class of all rather than just finite τ-structures.

Chapter 1

Introduction

1.1 Finite Model Theory: Topical and Methodological Differences

Model theory: analysis of syntax vs. semantics. Some main issues:

- definability of structural properties with logical means
- algebraic properties in axiomatic theories
- classification of models of theories

Some key theorems of classical first-order model theory:

- the compactness theorem (!)
- the Löwenheim-Skolem theorem
- the Craig interpolation theorem (and other interpolation theorems)
- the Łos–Tarski preservation theorem (and many other preservation theorems)
- classification of theories w.r.t. their spectrum in all infinite cardinalities (Morley's theorem)

Finite Model Theory: only consider finite structures.
Key notions correspondingly shift:

$$\begin{array}{rcl} \mathsf{STR}(\tau) & \rightsquigarrow & \mathsf{FIN}(\tau) \\ \mathsf{MOD}(\varphi) & \rightsquigarrow & \mathsf{FMOD}(\varphi) \\ \varphi \models \psi & \rightsquigarrow & \varphi \models_{\mathrm{fin}} \psi \\ \varphi \equiv \psi & \rightsquigarrow & \varphi \equiv_{\mathrm{fin}} \psi \\ \mathrm{SAT} & \rightsquigarrow & \mathrm{FINSAT} \\ \mathrm{VAL} & \rightsquigarrow & \mathrm{FINVAL} \end{array}$$

Finite structures are usually disregarded in classical model theory. But finiteness often matters (for adequate modelling), and restriction to just finite structures dramatically changes the picture.

Of the above central results, all are either meaningless (like Löwenheim-Skolem) or just no longer valid in the sense of finite model theory (like compactness, and in its wake, most other theorems of classical model theory).

Consequently, finite model theory (FMT) has developed into a very different discipline from classical model theory, with distinct methods, themes and applications of its own. Connections with computer science (theory and applications), algorithmic issues and complexity theory have strong influence on the development of FMT.

Emphasis on construction and analysis of finite models leads to stronger interplay with combinatorics, graph theory and related branches of discrete mathematics, including probabilistic methods. Algorithmic issues lead to new themes like model checking complexity (with finite structures and formulae as inputs) and the field of descriptive complexity (matching logical definability against computational complexity). Models of fixed finite sizes can be counted (up to isomorphism), leading to the study of asymptotic probabilities (0-1-laws).

Conventions: All vocabularies are finite (and mostly relational); FO stands for first-order logic, as well as for the set of formulae of FO (= FO(τ) when τ is fixed). Writing $\varphi(\mathbf{x})$ it is understood that the free variables of φ are among those listed.

$\mathsf{STR}(\tau)$ and $\mathsf{FIN}(\tau)$ are the classes of all and of all finite τ-structures, respectively. Structures are typically denoted as in $\mathfrak{A} = (A, R^{\mathfrak{A}}, U^{\mathfrak{A}}, c^{\mathfrak{A}})$; structures with parameters $\mathfrak{A}, \mathbf{a}$ for tuples $\mathbf{a}$ from A. Notation like $\mathfrak{A} \models \varphi[\mathbf{a}]$ signifies that $\mathfrak{A}$ satisfies φ under the assignment of $\mathbf{a}$ to $\mathbf{x}$. In this context, $\mathfrak{A}, \mathbf{a} \models \varphi$ is used interchangeably.

Example 1.1.1 For a binary relation symbol R, let $\varphi_0 \in \mathrm{FO}(\{R\})$ say that R is the graph of a total injective function; $\varphi_1 \in \mathrm{FO}(\{R\})$ say that R is the graph of a total surjective function. Then $\varphi_0 \equiv_{\mathrm{fin}} \varphi_1$, but classically neither $\varphi_0 \models \varphi_1$ nor $\varphi_1 \models \varphi_0$. A sentence $\varphi \in \mathrm{FO}(\{R\})$ with only infinite models $\varphi \in \mathrm{SAT(FO)} \setminus \mathrm{FINSAT(FO)})$ is easily obtained from these. So FO does not have the finite model property.

Definition 1.1.2 A logic $\mathcal{L}$ has the finite model property (FMP) if any satisfiable formula of $\mathcal{L}$ has a finite model, i.e., if $\mathrm{SAT}(\mathcal{L}) = \mathrm{FINSAT}(\mathcal{L})$.

While FO does not have the FMP, some interesting fragments (e.g., the 2-variable fragment or modal logics) do.

Exercise 1.1.3 If $\mathcal{L}$ is closed under boolean connectives and has the FMP, then the consequence relation for $\mathcal{L}$, $\models^{\mathcal{L}}$, coincides with its FMT variant, $\models^{\mathcal{L}}_{\mathrm{fin}}$.

1.2 Failure of Classical Methods and Results

Proposition 1.2.1 FO *does not have the compactness property for finite models.*

Proof Look, for $\tau = \emptyset$, at the set of sentences $\exists x_1 \ldots \exists x_n \bigwedge_{1 \leqslant i < j \leqslant n} \neg x_i = x_j$, for $n \geqslant 1$. □

Compare the following with the classical situation: there VAL(FO) is recursively enumerable (due to the existence of a complete proof calculus), but SAT(FO) is not.

Recall that Trakhtenbrot's theorem is proved via reduction of the halting problem (for Turing machines or register machine programs). There is a computable reduction that maps a machine/program $\mathcal{M}$ and input instance λ to an FO-sentence $\varphi_{\mathcal{M},\lambda}$ such that $\mathcal{M}$ terminates on λ iff $\varphi_{\mathcal{M},\lambda} \in \mathrm{FINSAT}$.

Theorem 1.2.2 (Trakhtenbrot)
FINVAL(FO) *is not recursively enumerable while* FINSAT(FO) *is recursively enumerable.*[1]

Corollary 1.2.3 FO *cannot have any (finitistic, effective) complete proof calculus for* $\models_{\text{fin}}$.

Exercise 1.2.4 Review and discuss the status of VAL and SAT on the one hand and of FINVAL and FINSAT on the other hand. What can be said about the status of fragments of FO that do have the finite model property?

Consider the example of a typical classical preservation theorem, the Łos–Traski theorem, see Theorem 0.1.5 above, which we re-state for convenience.

Theorem 1.2.5 (Łos–Tarski) *The following are equivalent for any* $\text{FO}(\tau)$*-sentence* φ*:*

(i) φ *is preserved under substructures, i.e., for all* $\mathfrak{B} \subseteq \mathfrak{A}$*, if* $\mathfrak{A} \models \varphi$ *then* $\mathfrak{B} \models \varphi$ *(equivalently:* $\mathsf{MOD}(\varphi)$ *closed under the substructure relation).*

(ii) $\varphi \equiv \tilde{\varphi}$ *for some universal* $\text{FO}(\tau)$*-sentence* $\hat{\varphi}$.

Proposition 1.2.6 (Tait, Gurevich) *The analogue of the Łos–Tarski theorem fails in FMP. There are* FO*-sentences* φ *for which* $\mathsf{FMOD}(\varphi)$ *is closed under the substructure relation, but* $\mathsf{FMOD}(\varphi) \neq \mathsf{FMOD}(\psi)$ *for any universal first-order sentence* ψ.

Proof Consider the vocabulary τ consisting of two constants min and max, two binary relation symbols $<$ and R, and a unary relation symbol P. Let $\varphi_0 \in \text{FO}(\tau)$ be a universal (!) sentence saying that "$<$ is a total linear order of the universe with first element min, last element max and R is a subset of the successor relation w.r.t. $<$."

Let $\varphi_1 = \forall x(x = \mathsf{max} \vee \exists y Rxy)$. Note that $\varphi_0 \wedge \varphi_1$ forces R to be the real successor relation w.r.t. $<$. Consider

$$\varphi := \varphi_0 \wedge (\varphi_1 \rightarrow \exists z Pz).$$

Claim 1: $\mathsf{FMOD}(\varphi)$ is closed under substructures.

Let $\mathfrak{A} \models \varphi$ and $\mathfrak{B} \subseteq \mathfrak{A}$. Clearly $\mathfrak{B} \models \varphi_0$ (φ_0 *is* universal). If $\mathfrak{B} \not\models \varphi_1$, then $\mathfrak{B} \models \varphi$. If $\mathfrak{B} \models \varphi_1$, then $\mathfrak{B} = \mathfrak{A}$ (!) and hence trivially $\mathfrak{B} \models \varphi$.

Claim 2: φ is not equivalent on $\mathsf{FIN}(\tau)$ to any sentence $\psi = \forall \mathbf{x}\chi(\mathbf{x})$ with qfr-free χ. Assume to the contrary that $\psi \equiv_{\text{fin}} \varphi$ for such ψ. Let $\mathbf{x} = (x_1, \dots, x_n)$ the variables of χ. Let $\mathfrak{A}_0$ be the standard model of $\varphi_0 \wedge \varphi_1$ on $A = \{0, \dots, n+2\}$. Then $(\mathfrak{A}, P^{\mathfrak{A}}) \models \varphi$ iff $P^{\mathfrak{A}} \neq \emptyset$. In particular, $(\mathfrak{A}, \emptyset) \not\models \varphi$ implies that $(\mathfrak{A}, \emptyset) \not\models \forall \mathbf{x}\chi(\mathbf{x})$. So $(\mathfrak{A}, \emptyset) \models \neg\chi[\mathbf{a}]$ for suitable $\mathbf{a} \in A^n$. Choosing $b \in A$ disjoint from $\mathbf{a}$, $\mathsf{min}^{\mathfrak{A}}$ and $\mathsf{max}^{\mathfrak{A}}$ (note that $|A| > n+2$), we still have $(\mathfrak{A}, \{b\}) \models \neg\chi[\mathbf{a}]$, because $\mathfrak{A} \restriction \mathbf{a}$ is unchanged. So $(\mathfrak{A}, \{b\}) \not\models \forall \mathbf{x}\chi(\mathbf{x})$ and therefore $(\mathfrak{A}, \{b\}) \not\models \psi$ while on the other hand $(\mathfrak{A}, \{b\}) \models \varphi$. Contradiction. $\square$

Many more classical results are known to fail in FMT. For instance, the following, which we discussed in the classical context above, cf. Theorem 0.2.6, is known to fail in FMT.

Theorem 1.2.7 (Craig interpolation) *For sentences* $\varphi \in \text{FO}(\tau_1)$ *and* $\psi \in \text{FO}(\tau_2)$ *such that* $\varphi \models \psi$*, there is an interpolant* $\chi \in \text{FO}(\tau_1 \cap \tau_2)$ *for which* $\varphi \models \chi$ *and* $\chi \models \psi$.

[1] More precisely: for any fixed finite τ, FINSAT(FO(τ)), the set of FO(τ)-sentences that are satisfiable in finite models, is r.e.; for any fixed finite finite vocabulary τ with at least one binary relation symbol, FINVAL(FO(τ)), the set of FO(τ)-sentences that are valid in all finite τ-structures, is not r.e.

Proposition 1.2.8 *Craig interpolation fails for* $\models_{\mathrm{fin}}$.

Exercise 1.2.9 Construct a counterexample to interpolation in FMT using sentences $\varphi \in \mathrm{FO}(\tau_1)$ and $\psi \in \mathrm{FO}(\tau_2)$, such that φ forces the universe of its finite models to have an odd number of elements, with ψ not in FINVAL but valid in all finite τ_2-structures with an odd number of elements.

That there is no $\chi \in \mathrm{FO}(\emptyset)$ whose finite models are just the odd size sets follows, e.g., with a very simple Ehrenfeucht-Fraïssé argument (cf. the next section, 2.1), as well as from Example 0.4.4 above.

One can similarly show that the FMT analogues of Beth's definability theorem, Theorem 0.2.8, fails – as we already concluded in Exercise 0.4.5 above.

Similarly, the analogue of Corollary 0.2.9 fails in the sense of FMT.

Proposition 1.2.10 *Order-invariant* FO *is strictly more expressive than plain* FO *in the sense of FMT: for suitable* τ *there are* $\mathrm{FO}(\tau \cup \{<\})$*-sentences that are order-invariant over* $\mathsf{FIN}(\tau)$ *but not equivalent to any plain* $\mathrm{FO}(\tau)$*-sentence over* $\mathsf{FIN}(\tau)$.

This claim is exemplified by the following.

Example 1.2.11 Let $\tau = \{U, P, E\}$ have unary predicates U, P and a binary relation E. The intended models for this example are structures $\mathfrak{A} = (A, U^{\mathfrak{A}}, P^{\mathfrak{A}}, E^{\mathfrak{A}})$ whose universe A is partitioned according to $A = U^{\mathfrak{A}} \dot\cup P^{\mathfrak{A}}$ (disjoint union). With $E^{\mathfrak{A}}$, we want the elements of $P^{\mathfrak{A}}$ to represent the power set $\mathcal{P}(U^{\mathfrak{A}})$, through the association of $p \in P^{\mathfrak{A}}$ with the subset

$$E_p^{\mathfrak{A}} := \{u \in U^{\mathfrak{A}} \colon (u, p) \in E^{\mathfrak{A}}\} \subseteq U^{\mathfrak{A}}.$$

Let

$$\psi_0 := \forall x\big((Ux \vee Px) \wedge \neg(Ux \wedge Px)\big) \wedge \forall x \forall y(Exy \rightarrow (Ux \wedge Py)) \wedge \exists x Ux \wedge \exists x Px.$$

This sentence ψ_0 says that we are dealing with one-sorted representations of two-sorted structures with sorts U and P, linked by a directed relation E with edges from the first to the second sort sort. In the following we use variables u and u_i for the U-parts and variables p and p_i for the P-parts to indicate these sorts. Formally one may, in models of ψ_0, restrict all FO-quantifications to relativised quantifications of the forms

$$\begin{array}{ll} \forall p(Pp \rightarrow \ldots) & \forall u(Uu \rightarrow \ldots) \\ \exists p(Pp \wedge \ldots) & \exists u(Uu \wedge \ldots) \end{array}$$

that explicitly prescribe sorts. We now add to ψ_0 the following requirements:

(1) the subset $E_p := \{u \in U \colon (u, p) \in E\} \subseteq U$ uniquely determines $p \in P$ (extensionality: $p = p'$ iff $E_p = E_{p'}$).

(2) the subset of $\mathcal{P}(U)$ represented by P is non-empty and closed under finite differences (it suffices to assert that for every p and u there are p', p'' such that $E_{p'} = E_p \cup \{u\}$ and $E_{p''} = E_p \setminus \{u\}$).

Let ψ be the conjunction of ψ_0 with natural formalisations of (1) and (2) in $\mathrm{FO}(\tau)$.

It follows that in finite models $\mathfrak{A} \models \psi$,

$$\{E_p^{\mathfrak{A}} \colon p \in P^{\mathfrak{A}}\} = \mathcal{P}(U^{\mathfrak{A}}).^2$$

[2]No FO-sentence can force the equality in infinite models, by the Löwenheim–Skolem theorem.

Up to isomorphism, therefore, $\mathsf{FMOD}(\psi)$ is represented by the following standard models:

$$\mathfrak{A}_n = \big([n] \cup \mathcal{P}([n]), [n], \mathcal{P}([n]), \in_n\big)$$

where $[n] := \{1, \dots, n\}$ and $\in_n \subseteq [n] \times \mathcal{P}([n])$ is the usual $\in$-relation, for $n \geqslant 1$.

We now consider the subclass

$$\mathcal{C} := \{\mathfrak{A} \in \mathsf{FMOD}(\psi) \colon |U^{\mathfrak{A}}| \equiv 1 \bmod 2\}$$

corresponding to the power set structures of finite sets of odd cardinality.

Ehrenfeucht–Fraïssé methods to be treated in sections 2.1 and 2.4 below will show that $\mathcal{C}$ is not definable in $\mathrm{FO}(\tau)$: there is no $\varphi \in \mathrm{FO}(\tau)$ such that $\mathcal{C} = \mathsf{FMOD}(\varphi)$; equivalently: no $\varphi \in \mathrm{FO}(\tau)$ can uniformly distinguish the $\mathfrak{A}_n$ with odd n from those with even n.

In fact this will follow from an Ehrenfeucht–Fraïssé analysis of monadic second-order logic MSO, which extends FO by quantification over subsets of the universe of the structures at hand (see section 2.4). The connection is made through the following translation of $\mathrm{FO}(\tau)$-formulae $\varphi(u_1, \dots, u_k; p_1, \dots, p_\ell)$ interpreted over the two-sorted powerset structures $\mathfrak{A}_n = ([n] \cup \mathcal{P}([n]), \dots)$ into $\mathrm{MSO}(\emptyset)$-formulae $\tilde{\varphi}(u_1, \dots, u_k; X_1, \dots, X_\ell)$ interpreted over the corresponding one-sorted naked set structures $\mathfrak{B}_n := [n]$:

$$\begin{array}{rcl} \varphi(u_1, \dots, u_k; p_1, \dots, p_\ell) & \leftrightsquigarrow & \tilde{\varphi}(u_1, \dots, u_k; X_1, \dots, X_\ell) \\ Eu_ip_j & \leftrightsquigarrow & X_ju_i, \\ \forall u_i(Uu \rightarrow \varphi), & \leftrightsquigarrow & \forall u_i(Uu \rightarrow \tilde{\varphi}), \\ \forall p_j(Pp_j \rightarrow \varphi), & \leftrightsquigarrow & \forall X_j\tilde{\varphi}, \end{array}$$

where we have suprressed some trivial clauses. One easily verifies by syntactic induction on φ that for all n and all assignments $u_i \in [n]$ and $p_j \in \mathcal{P}([n])$ in corresponding sorts of $\mathfrak{A}_n$ (such that the p_j are appropriate also as assignments to X_j over $\mathfrak{B}_n$):

$$\mathfrak{A}_n \models \varphi[\mathbf{u}, \mathbf{p}] \quad \Leftrightarrow \quad \mathfrak{B}_n \models \tilde{\varphi}[\mathbf{u}, \mathbf{p}].$$

Hence, the fact that no $\mathrm{MSO}(\emptyset)$-sentence distinguishes the $\mathfrak{B}_n$ with even n from those with odd n, see Example 2.4.8, implies that no $\mathrm{FO}(\tau)$-sentence can distinguish the $\mathfrak{A}_n$ with even n from those with odd n.

Looking at any linearly ordered version $(\mathfrak{A}_n, <)$, on the other hand, the following $\mathrm{FO}(\tau_<)$-sentence φ does draw exactly this distinction in an order-invariant manner. In the notation of section 0.2.4, $\mathcal{C}_< = \mathsf{FMOD}(\psi) \cap \mathsf{FMOD}(\varphi(<))$ for

$$\varphi(<) := \exists p\big(p \in P \wedge \psi(U, p, <)\big),$$

where $\psi(U, p, <)$ says that E_p contains the $<$-minimal/maximal elements of U, and exactly one of every pair of immediate successors w.r.t. the order induced by $<$ on U.

A notable exception of a classical Tarski style preservation theorem that does survive in FMT due to a recent result of Rossmann (LICS 2005 best paper award) is the Lyndon–Tarski preservation theorem, cf. Theorem 0.1.6 for the classical version.

Theorem 1.2.12 (Rossmann) *The following are equivalent for any sentence $\varphi \in \mathrm{FO}(\tau)$ for relational vocabulary τ, also in the sense of FMT:*

(i) *φ is preserved under homomorphisms.*

(ii) *$\varphi \equiv \hat{\varphi}$ for some positive existential $\mathrm{FO}(\tau)$-sentence $\hat{\varphi}$.*

1.3 Global Relations, Queries and Definability

Definition 1.3.1 Let $\mathsf{FIN}(\tau)$ be the class of all finite τ-structures.

For any $n \in \mathbb{N}$ and map $\pi \colon A \to B$, we lift π to all powers A^n through the maps $\pi \colon A^n \to B^n$ that send $(a_1, \ldots, a_n) \in A^n$ to $(\pi(a_1), \ldots, \pi(a_n)) \in B^n$.

For any set A, we identify A^n for $n = 0$ with the singleton set whose only element is the empty tuple. The only subsets of A^0 thus are $\emptyset$ and A^0; these we identify with the boolean values 0 (false, for the empty subset of A^0) and 1 (true, for A^0 itself). The corresponding lift of $\pi \colon A \to B$ to $\pi \colon A^0 \to B^0$ is the identity on $\mathbb{B} = \{0, 1\}$.

Definition 1.3.2 A *global relation* or *query* of arity n ($n \in \mathbb{N}$) over $\mathsf{FIN}(\tau)$ is a mapping

$$\mathfrak{A} \in \mathsf{FIN}(\tau) \longmapsto Q^{\mathfrak{A}} \subseteq A^n$$

that is compatible with $\simeq$ in the sense that for $\pi \colon \mathfrak{A} \simeq \mathfrak{B}$ always $Q^{\mathfrak{B}} = \pi\big(Q^{\mathfrak{A}}\big)$. Queries or global relations of arity 0 are called *boolean*.

Queries and global relations over suitable subclasses of $\mathsf{FIN}(\tau)$ (or of $\mathsf{STR}(\tau)$) can be similarly defined.

Remark 1.3.3 *A boolean query Q on $\mathsf{FIN}(\tau)$ is identified with the subclass (also called Q) $Q = \{\mathfrak{A} \in \mathsf{FIN}(\tau) \colon Q^{\mathfrak{A}} = 1\}$. The compatibility condition means that Q is closed under $\simeq$.*

Example 1.3.4 In the following $<, E$ are binary relation symbols, U is a unary relation symbol, $+, \cdot$ are binary function symbols, and $0, 1$ are constant symbols.

(i) for $\tau = \{E\}$: GRAPH $:= \{\mathfrak{A} = (A, E^{\mathfrak{A}}) \colon \mathfrak{A}$ a finite undirected graph $\}$.[3]

(ii) for $\tau = \{<\}$: ORD $:= \{\mathfrak{A} = (A, <^{\mathfrak{A}}) \in \mathsf{FIN}(\tau) \colon <^{\mathfrak{A}}$ a total linear order of A $\}$.

(iii) for $\tau = \emptyset$: EVEN $:= \{A \colon |A|$ even $\}$.

(iv) for $\tau = \{+, \cdot, 0, 1\}$: FIELD $:= \{\mathfrak{A} = (A, +^{\mathfrak{A}}, \cdot^{\mathfrak{A}}, 0^{\mathfrak{A}}, 1^{\mathfrak{A}}) \colon \mathfrak{A}$ a finite field $\}$.

(v) for $\tau = \{+, \cdot, 0, 1\}$: unary Q defined as

$$Q^{\mathfrak{A}} := \begin{cases} \{a \in A \colon a \text{ a unit in } \mathfrak{A}\ \} & \text{if } \mathfrak{A} \text{ is a ring} \\ \emptyset & \text{else} \end{cases}$$

(vi) for $\tau = \{E\}$, $\ell \in \mathbb{N}$: binary D_ℓ defined as

$$D^{\mathfrak{A}}_{\leqslant \ell} := \begin{cases} \{(a, b) \in A^2 \colon d(a, b) \leqslant \ell\} & \text{if } \mathfrak{A} \in \text{GRAPH}^4 \\ \emptyset & \text{else} \end{cases}$$

Derived from these, the following binary and boolean graph queries:

$$\begin{array}{rcll} D^{\mathfrak{A}}_{<\infty} & := & \bigcup_{\ell \in \mathbb{N}} D^{\mathfrak{A}}_{\ell} & \text{(reachability)} \\ \text{CONN} & := & \{\mathfrak{A} \colon \mathfrak{A} \text{ a finite connected undirected graph } \} & \text{(connectivity)} \\ & = & \{\mathfrak{A} \colon D^{\mathfrak{A}}_{<\infty} = A^2\} & \end{array}$$

(vii) for $\tau = \{E\}$: BIPART $:= \{\mathfrak{A} \colon \mathfrak{A} \in$ GRAPH bipartite $\}$.

(viii) for $\tau = \{E, U\}$: U-BIPART $:= \{\mathfrak{A} \colon \mathfrak{A}$ a finite graph, bipartite w.r.t. $U^{\mathfrak{A}}$ $\}$.

[3] $(A, E^{\mathfrak{A}})$ is an undirected graph if $E^{\mathfrak{A}}$ is symmetric and irreflexive.

[4] d is the usual graph distance.

(ix) for $\tau = \{E, U\}$: MATCH $:= \{\mathfrak{A} \in$ U-BIPART: $\mathfrak{A}$ has a perfect matching $\}$.

Definition 1.3.5 An n-ary query (global relation) is definable in the logic $\mathcal{L}$ (e.g., in FO) if for some $\varphi \in \mathcal{L}(\tau)$, we have that for all $\mathfrak{A} \in \mathsf{FIN}(\tau)$

$$Q^{\mathfrak{A}} = \{\mathbf{a} \in A^n : \mathfrak{A} \models \varphi[\mathbf{a}]\}.$$

In the boolean case ($n = 0$) this means that $Q = \mathsf{FMOD}(\varphi)$ for some $\mathcal{L}$-sentence φ.

Exercise 1.3.6 Provide FO-definitions for the following queries among the above examples: (i), (ii), (iv), (v), each $D_{\leqslant \ell}$ in (vi), as well as (viii). The others are in fact not FO-definable (see later).

With any class $\Delta = \Delta(\mathbf{x})$ of formulae (in a fixed tuple of free variables $\mathbf{x}$) we may associate the induced notion of Δ-equivalence of structures (with parameters as assignments to $\mathbf{x}$):

$$\mathfrak{A}, \mathbf{a} \equiv_\Delta \mathfrak{B}, \mathbf{b} \qquad \text{iff} \qquad \text{for all } \varphi(\mathbf{x}) \in \Delta \colon \mathfrak{A} \models \varphi[\mathbf{a}] \Leftrightarrow \mathfrak{B} \models \varphi[\mathbf{b}].$$

If $\Delta(\mathbf{x})$ is finite up to logical equivalence (over $\mathsf{FIN}(\tau)$), then $\equiv_\Delta$ has finite index (over $\mathsf{FIN}(\tau)$). A query Q is closed under $\equiv_\Delta$ if $\mathfrak{A}, \mathbf{a} \equiv_\Delta \mathfrak{B}, \mathbf{b}$ implies that $\mathbf{a} \in Q^{\mathfrak{A}}$ iff $\mathbf{b} \in Q^{\mathfrak{B}}$.

Proposition 1.3.7 *Let $\Delta(\mathbf{x})$ be a class of formulae that is closed under boolean connectives and finite up to logical equivalence over $\mathsf{FIN}(\tau)$, Q a query over $\mathsf{FIN}(\tau)$. Then the following are equivalent:*

(i) *Q is definable by a formula from Δ.*
(ii) *Q closed under $\equiv_\Delta$.*

Proof (of (ii) $\Rightarrow$ (i)) For $\mathfrak{A}, \mathbf{a}$ consider its Δ-type $\{\delta(\mathbf{x}) \in \Delta \colon \mathfrak{A} \models \delta[\mathbf{a}]\} \subseteq \Delta$. By assumption this set is finite up to logical equivalence over $\mathsf{FIN}(\tau)$ and hence logically equivalent to a single formula $\delta_{\mathfrak{A},\mathbf{a}} \in \Delta$: $\delta_{\mathfrak{A},\mathbf{a}} \equiv \bigwedge \{\delta(\mathbf{x}) \in \Delta \colon \mathfrak{A} \models \delta[\mathbf{a}]\}$ (and Δ is closed under conjunction). $\delta_{\mathfrak{A},\mathbf{a}}$ characterises $\mathfrak{A}, \mathbf{a}$ up to $\equiv_\Delta$-equivalence in the sense that $\mathfrak{B}, \mathbf{b} \equiv_\Delta \mathfrak{A}, \mathbf{a}$ iff $\mathfrak{B} \models \delta_{\mathfrak{A},\mathbf{a}}(\mathbf{b})$ (this uses that Δ is closed under negation). For similar reasons the following disjunction is (up to equivalence) a formula $\varphi \in \Delta$ and defines Q: $\varphi(\mathbf{x}) := \bigvee \{\delta_{\mathfrak{A},\mathbf{a}} \colon \mathfrak{A} \in \mathsf{FIN}(\tau), \mathbf{a} \in Q^{\mathfrak{A}}\}$. $\square$

As we shall see in more detail later, FO is too weak to capture even some very basic structural properties of finite τ-structures: for instance, EVEN or the natural boolean graph queries CONN, BIPART, MATCH and the binary reachability query $D_{<\infty}$ of (v) are not FO-definable.

The reachability query $D_{<\infty}$, for instance, though not FO-definable, is computationally and algebraically very basic.

On graphs $\mathfrak{A}$, $D^{\mathfrak{A}}_{<\infty}$ is just the reflexive transitive closure of $E^{\mathfrak{A}}$. In terms of the adjacency matrix $\mathbb{A}$ of $E^{\mathfrak{A}}$, iterated powers of $\mathbb{A} + \mathbb{I}$ w.r.t. a boolean matrix product operation will yield the adjacency matrix for $D^{\mathfrak{A}}_{<\infty}$.

$D_{<\infty}$ is also generated by recursive iteration of a DATALOG program

$$\begin{array}{lcl} Xxx & \leftarrow & \\ Xxy & \leftarrow & Xxz, Ezy \end{array}$$

based on simple qfr-free FO-rules for the iteration. The n-th iteration of this program on a finite graph $\mathfrak{A}$ evaluates to $D^{\mathfrak{A}}_{\leqslant n}$, and $D^{\mathfrak{A}}_{<\infty}$ is reached as the limit (union) of this monotone chain of stages within $|A|$ many iterations. [Can you improve this to a logarithmic bound, with a modified program?]

Whether a given pair (a, b) of nodes of a finite graph $\mathfrak{A}$ is in $D^{\mathfrak{A}}_{<\infty}$ can also easily be checked by a breadth-first search algorithm (in a number of iterations that is linear in the number of edges).

Curious phenomenon: FO and in particular elementary equivalence $\equiv$ between finite structures are "too strong" to be of model theoretic interest, as any finite structure is characterised up to isomorphism by a single FO-sentence.

Observation 1.3.8 For finite vocabulary τ and any $\mathfrak{A} \in \mathsf{FIN}(\tau)$ there is a sentence $\varphi_{\mathfrak{A}} \in \mathrm{FO}(\tau)$ with $\mathsf{MOD}[\varphi_{\mathfrak{A}}] = \{\mathfrak{B} \colon \mathfrak{B} \simeq \mathfrak{A}\}$. Elementary equivalence and isomorphism agree on $\mathsf{FIN}(\tau)$.

Exercise 1.3.9 Let τ be finite. For $\mathfrak{A} \in \mathsf{FIN}(\tau)$ of size n provide a prenex sentence $\varphi_{\mathfrak{A}}$ that characterises $\mathfrak{A}$ up to $\simeq$, with $\mathrm{qr}(\varphi_{\mathfrak{A}}) = n + 1$.

Show that $\mathfrak{A} \equiv \mathfrak{B}$ implies $\mathfrak{A} \simeq \mathfrak{B}$ whenever at least one of $\mathfrak{A}$ of $\mathfrak{B}$ is finite. (For this, τ does not even have to be finite, why?)

Summary In FMT FO loses the unique status it holds in classical model theory. Consequently, FO will only be one logic among many others to be considered. Fragments (with better algorithmic behaviour, more suited to specific tasks, or inducing more interesting notions of equivalence over finite structures) and various extensions (with stronger expressive power for defining queries) feature importantly.

Some other logics Besides FO we here encounter, on the one hand, its restrictions to a fixed finite supply of k distinct variable symbols. FO^k, the k-variable fragment of FO. induces a logically non-trivial notion of elementary equivalence on finite structures, with useful game characterisations etc (this is where FO was "too strong").

On the other hand, for powerful extensions beyond FO, we look at the fragment of second-order logic which adds with quantification over subsets to FO (monadic second-order logic MSO) and at extensions of FO by several mechanisms for relational recursion (fixpoint logics LFP, IFP and PFP).

Applications and context; outlook FMT has strong links with computer science, both as an application area and as a source of motivation for model theoretic questions particular to finite structures. The following are some key connections:

- expressive power of various logics over finite structures (Part I).
  - database query languages (SQL essentially based on FO; extensions with various recursion operators like, e.g., transitive closures; DATALOG as a purely relational version of PROLOG, etc)
  - languages for formal specification and verification of systems and processes (model checking with various modal logics, temporal logics, process logics)
- algorithmic properties of logics over finite structures.
  - model checking algorithms and their complexity
  - SAT/FINSAT as central logic problems for many applications areas (process logics, description logics, logics for knowledge bases, etc)

- logic and complexity (Part II).
  - logics designed to match levels of computational complexity
  - transfer between model theory and theory of complexity

Chapter 2

Expressiveness and Definability via Games

2.1 The Ehrenfeucht-Fraïssé Method

We explicitly deal with relational vocabularies only. Unless otherwise mentioned all vocabularies τ are finite and consist of relation symbols only. Constants could easily be incorporated (in an obvious manner); the inclusion of function symbols would necessitate an analysis of the contribution that the quantifier rank and the complexity of terms make to the expressiveness of FO. Note, however, that functions can be eliminated in favour of relations that describe the graphs of functions.

Ehrenfeucht-Fraïssé games provide a key methodology for the analysis of the expressive power of various logics. The methodology itself is applicable in classical model theory as well as in FMT. In FMT it is well adapted to the often more combinatorial character of model construction and analysis.

We denote finite maps p from $\mathrm{def}(p) \subseteq A$ to $\mathrm{im}(p) \subseteq B$ as

$$p = (\mathbf{a} \mapsto \mathbf{b})$$

if $\mathbf{a} = (a_1, \dots, a_n) \in A^n$ is such that $\mathrm{def}(p) = \{a_1, \dots, a_n\}$ and $\mathbf{b} = (b_1, \dots, b_n)$ where $b_i = p(a_i)$.

$p \subseteq p'$ means that p' extends p in the sense that $\mathrm{def}(p) \subseteq \mathrm{def}(p')$ and $p'(a) = p(a)$ for all $a \in \mathrm{def}(p)$.

A map $p = (\mathbf{a} \mapsto \mathbf{b})$ is a *local isomorphism*[1] between τ-structures $\mathfrak{A}$ and $\mathfrak{B}$ if $p \colon \mathfrak{A} \restriction \mathrm{def}(p) \simeq \mathfrak{B} \restriction \mathrm{im}(p)$ is an isomorphism of induced substructures (useful shorthand: $p \colon \mathfrak{A} \restriction \mathbf{a} \simeq \mathfrak{B} \restriction \mathbf{b}$.) We admit the empty partial isomorphism $p = \emptyset$ as a special case of a partial isomorphism.

Definition 2.1.1 For $\mathfrak{A}, \mathfrak{B} \in \mathsf{STR}(\tau)$ let $\mathrm{Part}(\mathfrak{A}, \mathfrak{B})$ be the set of all finite local (classically: partial) isomorphisms between $\mathfrak{A}$ and $\mathfrak{B}$.

Example 2.1.2 For linear orderings $\mathfrak{A}$ and $\mathfrak{B}$, $\mathrm{Part}(\mathfrak{A}, \mathfrak{B})$ consists of all order-preserving maps $p = (\mathbf{a} \mapsto \mathbf{b})$. These are representable by $\mathbf{a}$ and $\mathbf{b}$ such that $\mathbf{a}$ is strictly increasing w.r.t. $<^{\mathfrak{A}}$ and $\mathbf{b}$ is strictly increasing w.r.t. $<^{\mathfrak{B}}$.

[1] We prefer the term "local isomorphism" to the standard terminology of "partial isomorphisms."

2.1.1 The basic Ehrenfeucht–Fraïssé game; FO pebble game

Review of basic idea: two players **I** (challenger, spoiler, male) and **II** (duplicator, female) play over two structures $\mathfrak{A}, \mathfrak{B} \in \mathsf{STR}(\tau)$. Roles: **I** tries to demonstrate differences, **II** similarity between $\mathfrak{A}$ and $\mathfrak{B}$.

Game positions: configurations $(\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b})$ where $\mathbf{a} = (a_1, \dots, a_n)$, $\mathbf{b} = (b_1, \dots, b_n)$, $n \in \mathbb{N}$. In pebble game terms: two sets of pebbles numbered $i = 1, \dots, n$, placed on elements a_i and b_i of $\mathfrak{A}$ and $\mathfrak{B}$, respectively.

Single round: challenge/response according to:

I	places next pebble on some element of either $\mathfrak{A}$ or $\mathfrak{B}$
II	responds by placing the opposite pebble in the opposite structure

This exchange of moves leads the play from some position $(\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b})$ to a new position $(\mathfrak{A}, \mathbf{a}a; \mathfrak{B}, \mathbf{b}b)$. The newly placed pebble pair extends the correspondence $\mathbf{a} \mapsto \mathbf{b}$ in the previous position to $\mathbf{a}a \mapsto \mathbf{b}b$.

Winning conditions/constraints: **II** loses (and **I** wins) the play as soon as the mapping $\mathbf{a} \mapsto \mathbf{b}$ induced by the current position is *not* a partial isomorphism. Otherwise, we speak of *isomorphic pebble configurations* if $(\mathbf{a} \mapsto \mathbf{b}) \in \mathrm{Part}(\mathfrak{A}, \mathfrak{B})$, and play may continue.

The m-round game

Definition 2.1.3 The m-round game $\mathsf{G}_m(\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b})$ continues for m rounds starting from position $(\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b})$. **II** wins any play in which she maintains isomorphic pebble configurations through all m rounds, and loses otherwise.

In any game, we say that **II** *wins the game* if she has a winning strategy (so that she wins every play that she plays according that strategy). It is obvious that in any m-round game over finite structures precisely one of the players has a winning strategy. Here this even follows directly from the finiteness of the game tree of all possible plays, which could also be analysed by exhaustive search to determine who can force a win.[2] The analysis below yields better insights, though.

Winning strategies and back-and-forth systems

Definition 2.1.4

(i) Let $I \subseteq \mathrm{Part}(\mathfrak{A}, \mathfrak{B})$, $p = (\mathbf{a} \mapsto \mathbf{b}) \in \mathrm{Part}(\mathfrak{A}, \mathfrak{B})$. *$p$ has back-and-forth extensions in I* if

$$\begin{array}{ll} \textit{forth} & \forall a \in A \exists b \in B\colon (\mathbf{a}a \mapsto \mathbf{b}b) \in I \\ \textit{back} & \forall b \in B \exists a \in A\colon (\mathbf{a}a \mapsto \mathbf{b}b) \in I \end{array}$$

(ii) Let $I_i \subseteq \mathrm{Part}(\mathfrak{A}, \mathfrak{B})$ for $0 \leqslant i \leqslant m$. Then $(I_i)_{0 \leqslant i \leqslant m}$ is a *back-and-forth system* for $\mathsf{G}_m(\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b})$ if
- $(\mathbf{a} \mapsto \mathbf{b}) \in I_m$
- for $1 \leqslant k \leqslant m$, every $p \in I_k$ has back-and-forth extensions in I_{k-1}.

[2] That games of this kind are *determined* in this sense follows in a much wider context, including infinite play with not necessarily finite branching on moves.

(iii) If $(I_i)_{0 \leqslant i \leqslant m}$ is a back-and-forth system for $\mathsf{G}_m(\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b})$, we write

$$(I_i)_{0 \leqslant i \leqslant m} \colon \mathfrak{A}, \mathbf{a} \simeq_m \mathfrak{B}, \mathbf{b}$$

and say that $\mathfrak{A}, \mathbf{a}$ and $\mathfrak{B}, \mathbf{b}$ are m-isomorphic, $\mathfrak{A}, \mathbf{a} \simeq_m \mathfrak{B}, \mathbf{b}$.

Observation 2.1.5 **II** wins $\mathsf{G}_m(\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b})$ (i.e., she has a winning strategy for this game) iff $\mathfrak{A}, \mathbf{a} \simeq_m \mathfrak{B}, \mathbf{b}$ (i.e., if there is a back-and-forth system for $\mathsf{G}_m(\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b})$).

Proof (sketch) For "$\Leftarrow$" extract winning strategy from back-and-forth conditions: with k more rounds to play, **II** can maintain positions in I_k.

For "$\Rightarrow$" show that the system $I_k := \big\{(\mathbf{a} \mapsto \mathbf{b}) \colon \textbf{ II} \text{ wins } \mathsf{G}_k(\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b})\big\}$ satisfies the back-and-forth conditions. $\Box$

Reminder: $\equiv_m$ stands for elementary equivalence up to qfr-rank m. $\mathfrak{A}, \mathbf{a} \equiv_m \mathfrak{B}, \mathbf{b}$ iff for all $\varphi(\mathbf{x}) \in \mathrm{FO}$ with $\mathrm{qr}(\varphi) \leqslant q$ we have $\mathfrak{A} \models \varphi[\mathbf{a}] \Leftrightarrow \mathfrak{B} \models \varphi[\mathbf{b}]$. In the sense of section 1.3, $\equiv_m$ is $\equiv_\Delta$ where Δ is the set of all FO-formulae of qfr-rank up to m.

Exercise 2.1.6 Show that for finite relational τ, $\equiv_m$ has finite index (over $\mathsf{FIN}(\tau)$ as well as over $\mathsf{STR}(\tau)$).

Theorem 2.1.7 (Ehrenfeucht-Fraïssé Theorem) *The following are equivalent for all $\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b}$ and m:*

(i) $\mathfrak{A}, \mathbf{a} \simeq_m \mathfrak{B}, \mathbf{b}$.

(ii) **II** *wins* $\mathsf{G}_m(\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b})$.

(iii) $\mathfrak{A}, \mathbf{a} \equiv_m \mathfrak{B}, \mathbf{b}$.

See logic course for detailed proof.

(ii) $\Rightarrow$ (iii) can be shown by induction on m; one outermost quantifier corresponds to the first round in the game.

For (iii) $\Rightarrow$ (i) one can show in an ad-hoc manner that the system $I_k := \big\{(\mathbf{a} \mapsto \mathbf{b}) \colon \mathfrak{A}, \mathbf{a} \equiv_k \mathfrak{B}, \mathbf{b}\big\}$ satisfies the back-and-forth conditions. Alternatively, one may use the following lemma with additional benefit.

Lemma 2.1.8 *For $\mathfrak{A}, \mathbf{a}$ and m there is a formula $\chi(\mathbf{x}) = \chi^m_{\mathfrak{A},\mathbf{a}}(\mathbf{x})$ of qfr-rank m that characterises the $\simeq_m$-class of $\mathfrak{A}, \mathbf{a}$ in the sense that for all $\mathfrak{B}, \mathbf{b}$:*

$$\mathfrak{B} \models \chi[\mathbf{b}] \qquad \textit{iff} \qquad \mathfrak{B}, \mathbf{b} \simeq_m \mathfrak{A}, \mathbf{a}.$$

The $\chi^m_{\mathfrak{A},\mathbf{a}}(\mathbf{x})$ are constructed by induction on m, for all $\mathfrak{A}, \mathbf{a}$ simultaneously:

$\chi^0_{\mathfrak{A},\mathbf{a}}$ consists just of conjunctions over all atomic and negated atomic formulae true of $\mathbf{a} \in \mathfrak{A}$.

Inductively, χ^{m+1} expresses the back-and-forth conditions relative to the given χ^m, in the following typical format:

$$\chi^{m+1}_{\mathfrak{A},\mathbf{a}}(\mathbf{x}) := \underbrace{\bigwedge\big\{\exists y\, \chi^m_{\mathfrak{A},\mathbf{a}a}(\mathbf{x}, y) \colon a \in A\big\}}_{\text{forth: responses for challenges in } \mathfrak{A}} \wedge \underbrace{\forall y \bigvee \big\{\chi^m_{\mathfrak{A},\mathbf{a}a}(\mathbf{x}, y) \colon a \in A\big\}}_{\text{back: responses for challenges in } \mathfrak{B}}.^3$$

Corollary 2.1.9 *A query (global relation) Q on $\mathsf{FIN}(\tau)$ is FO-definable at qfr-rank m iff Q is closed under $\simeq_m$ in the sense that for $\mathfrak{A}, \mathbf{a} \simeq_m \mathfrak{B}, \mathbf{b}$ we have $\mathbf{a} \in Q^{\mathfrak{A}} \Leftrightarrow \mathbf{b} \in Q^{\mathfrak{B}}$. It follows that Q is FO-definable iff Q is closed under $\simeq_m$ for some $m \in \mathbb{N}$.*

[3] Over infinite structures $\mathfrak{A}$ one uses the fact that there are only finitely many qfr-rank m formulae up to logical equivalence in order to see that these conjunctions and disjunctions can be made finite.

Proof For (i) $\Rightarrow$ (ii) let $\varphi(\mathbf{x}) \in \mathrm{FO}$ define Q, $\mathrm{qr}(\varphi) = m$, and let $\mathfrak{A}, \mathbf{a} \simeq_m \mathfrak{B}, \mathbf{b}$. By the theorem, $\mathfrak{A}, \mathbf{a} \equiv_m \mathfrak{B}, \mathbf{b}$, so $\mathfrak{B} \models \varphi[\mathbf{b}] \Leftrightarrow \mathfrak{A} \models \varphi[\mathbf{a}]$, and thus $\mathbf{a} \in Q^{\mathfrak{A}} \Leftrightarrow \mathbf{b} \in Q^{\mathfrak{B}}$.

For (ii) $\Rightarrow$ (i) let Q be closed under $\simeq_m$, hence under $\equiv_m$. The claim follows with Proposition 1.3.7. A defining formula for Q is

$$\varphi(\mathbf{x}) := \bigvee \{\chi^m_{\mathfrak{A},\mathbf{a}}(\mathbf{x}) \colon \mathbf{a} \in Q^{\mathfrak{A}}\}.$$

Note again how this disjunction is essentially finite. □

Exercise 2.1.10 Let $\mathfrak{A}_i, \mathbf{a}_i \simeq_m \mathfrak{B}_i, \mathbf{b}_i$ for $i = 1, 2$. Let $\mathfrak{A}$ be the disjoint union of $\mathfrak{A}_1$ and $\mathfrak{A}_2$, similarly $\mathfrak{B}$ that of the $\mathfrak{B}_i$. Show that $\mathfrak{A}, \mathbf{a}_1\mathbf{a}_2 \simeq_m \mathfrak{B}, \mathbf{b}_1\mathbf{b} - 2$. Argue for strategies in the game or with the corresponding back-and-forth systems. What does this imply about FO w.r.t. the operation of taking disjoint unions?

Exercise 2.1.11 Inductively define m-types of structures with parameters, $\mathrm{TP}^m(\mathfrak{A}, \mathbf{a})$, as follows:

$$\begin{aligned} \mathrm{TP}^0(\mathfrak{A}, \mathbf{a}) &:= \{\varphi(\mathbf{x}) \colon \varphi \text{ atomic and } \mathfrak{A} \models \varphi[\mathbf{a}]\}; \\ \mathrm{TP}^{m+1}(\mathfrak{A}, \mathbf{a}) &:= \{\mathrm{TP}^m(\mathfrak{A}, \mathbf{a}a) \colon a \in A\}. \end{aligned}$$

Show that TP^m characterises $\simeq_m$ classes (m-isomorphism types) in the sense that

$$\mathrm{TP}^m(\mathfrak{A}, \mathbf{a}) = \mathrm{TP}^m(\mathfrak{B}, \mathbf{b}) \qquad \text{iff} \qquad \mathfrak{A}, \mathbf{a} \simeq_m \mathfrak{B}, \mathbf{b}.$$

2.1.2 Inexpressibility via games

Based on Corollary 2.1.9 we can show that certain queries cannot be expressed in FO. For instance, for a boolean query Q, we establish that Q is not FO-definable over $\mathsf{FIN}(\tau)$ if we can exhibit structures $\mathfrak{A}_m \in Q$ and $\mathfrak{B}_m \notin Q$ for every $m \in \mathbb{N}$ such that $\mathfrak{A} \simeq_m \mathfrak{B}$.

Examples

Example 2.1.12 EVEN $\subseteq \mathsf{FIN}(\emptyset)$ is not FO-definable. Trivially any two naked sets of sizes $\geqslant m$ are m-isomorphic. Taking sets of sizes m and $m+1$ we see that EVEN is not closed under $\simeq_m$.

Example 2.1.13 The class Q of even length finite linear orderings is not FO-definable.

Proof Let $\mathfrak{A}_n$ be the standard ordering of $\mathbb{N}$ in restriction to $[n] := \{1, \dots, n\}$. On $(\mathbb{N}, <)$ consider the usual distance $d(i, j) = |j - i|$. We use truncated distances d_k (for $k \in \mathbb{N}$) with values in $\{0, \dots, 2^k - 1\} \cup \{\infty\}$ defined as

$$d_k(i, j) := \begin{cases} d(i, j) & \text{if } d(i, j) < 2^k \\ \infty & \text{else} \end{cases}$$

Consider strictly increasing $\mathbf{a} = (a_1, \dots, a_s)$ in $[n]$ and $\mathbf{b} = (b_1, \dots, b_s)$ in $[n']$. Put $(\mathbf{a} \mapsto \mathbf{b})$ into $I_k \subseteq \mathrm{Part}(\mathfrak{A}_n, \mathfrak{A}_{n'})$ if (for $s > 0$)

$$\begin{aligned} d_k(0, a_1) &= d_k(0, b_1) \\ d_k(a_i, a_{i+1}) &= d_k(b_i, b_{i+1}) \text{ for } 1 \leqslant i < s \\ d_k(a_s, n+1) &= d_k(b_s, n'+1), \end{aligned}$$

and $\emptyset \in I_k$ iff $d_k(0, n+1) = d_k(0, n'+1)$. One checks that $(I_k)_{0 \leqslant k \leqslant m}$ satisfies the back-and-forth conditions, and that $\emptyset \in I_m$ whenever $n = n'$ or $n, n' \geqslant 2^m - 1$. Hence $\mathfrak{A}_n \simeq_m \mathfrak{A}_{n'}$ for $n, n' \geqslant 2^m - 1$. Putting $n := 2^m$ and $n' := 2^m - 1$ we see that Q is not closed under $\simeq_m$. □

Exercise 2.1.14 Connectivity of finite linearly ordered graphs is not FO-definable. Modify the above example by choosing an edge relation E that is FO-definable in terms of the underlying linear orderings and such that $(A_n, E^{\mathfrak{A}_n})$ is connected precisely for even n. Describe a suitable choice of E and detail the argument establishing the non-definability claim.

Exercise 2.1.15 Connectivity of finite graphs, CONN, is not FO-definable. This follows from the previous, but one may also modify the truncated distance idea of Example 2.1.13 directly. We shall see a variant proof in Corollary 2.2.9 and a stronger result in Example 2.2.11.

Exercise 2.1.16 The class of finite bipartite graphs is not FO-definable. Hint: find an FO-definable edge relation E on the linear orderings $\mathfrak{A}_n$ from Example 2.1.13 such that the resulting graph is bipartite or not according to the parity of n.

2.2 Locality of FO: Hanf and Gaifman Theorems

All vocabularies finite and purely relational.

Definition 2.2.1

(i) With $\mathfrak{A} \in \mathsf{FIN}(\tau)$ associate its *Gaifman graph*, $G(\mathfrak{A}) := (A, E)$ where the edge relation is

$$E = E(\mathfrak{A}) := \bigcup_{R\in\tau} \bigcup_{\mathbf{a}\in R^{\mathfrak{A}}} \{(a_i, a_j) \colon a_i \neq a_j\}.$$

(ii) The *Gaifman distance* $d(a, b)$ between $a, b \in A$ is defined to be the usual graph distance in $G(\mathfrak{A})$ (with values in $\mathbb{N} \cup \{\infty\}$).

(iii) The *Gaifman neighbourhood* of radius ℓ of $a \in A$ is $N^\ell(a) := \{b \in A \colon d(a, b) \leqslant \ell\}$. For tuples $\mathbf{a} \in A^n$ define $N^\ell(\mathbf{a}) := \bigcup_{1\leqslant i\leqslant n} N^\ell(a_i)$.

(iv) A tuple $\mathbf{a}$ is called *ℓ-scattered* in $\mathfrak{A}$ if $d(a_i, a_j) > 2\ell$ for $i \neq j$. Equivalently, if $N^\ell(a_i) \cap N^\ell(a_j) = \emptyset$ for $i \neq j$.

Observation 2.2.2 The following global relations are FO-definable for all $\ell, n \in \mathbb{N}$:

(i) The edge relation E of the Gaifman graph of $\mathfrak{A}$.

(ii) $D_{\leqslant\ell}$ where $D^{\mathfrak{A}}_{\leqslant\ell} = \{(a, b) \in A^2 \colon d(a, b) \leqslant \ell\}$.

(iii) Similarly defined global relations $D_{*\ell}$ for $* =<, \geqslant, >, =$.

(iv) $\mathrm{SC}_{n,\ell}$ where $\mathrm{SC}^{\mathfrak{A}}_{n,\ell} = \{\mathbf{a} \in A^n \colon \mathbf{a}\ \ell\text{-scattered}\ \}$.

We use shorthand notation like "$d(x, y) \leqslant \ell$" for corresponding FO-formulae.
Also, "$d(\mathbf{x}, y) \leqslant \ell$" for $\mathbf{x} = (x_1, \ldots, x_n)$ is shorthand for $\bigvee_{1\leqslant i\leqslant n} d(x_i, y) \leqslant \ell$.

Exercise 2.2.3 Provide a formula defining the edge relation E of the Gaifman graph. By induction on ℓ, generate formulae "$d(x, y) \leqslant \ell$". [The qfr-rank of $d(x, y) \leqslant \ell$ can be bounded logarithmically in ℓ.]

Relativisation to Gaifman neighbourhoods Let $\varphi(\mathbf{x}) \in \mathrm{FO}(\tau)$, $\mathbf{y}$ a tuple of variables, w.l.o.g. not bound in φ. Let $\varphi^{N^\ell(\mathbf{y})}(\mathbf{x}, \mathbf{y})$ be the formula that relativises φ to the substructure that is induced on $N^\ell(\mathbf{y})$, the ℓ-neighbourhood of $\mathbf{y}$. One obtains this relativisation by induction on φ, as follows.

atomic φ:	$\varphi^{N^\ell(\mathbf{y})} := \varphi$
propositional connectives:	commute with relativisation
$\varphi = \exists z\psi$:	$\varphi^{N^\ell(\mathbf{y})} := \exists z\big(d(\mathbf{y},z) \leqslant \ell \wedge \psi^{N^\ell(\mathbf{y})}\big)$
$\varphi = \forall z\psi$:	$\varphi^{N^\ell(\mathbf{y})} := \forall z\big(d(\mathbf{y},z) \leqslant \ell \rightarrow \psi^{N^\ell(\mathbf{y})}\big)$

The crucial model theoretic property of these relativisations is that for all $\mathfrak{A}$, $\mathbf{a}, \mathbf{b}$ such that $\mathbf{a} \in N^\ell(\mathbf{b})$:

$$\mathfrak{A} \models \varphi^{N^\ell(\mathbf{y})}[\mathbf{a},\mathbf{b}] \qquad \text{iff} \qquad \mathfrak{A} \restriction N^\ell(\mathbf{b}) \models \varphi[\mathbf{a}].$$

Definition 2.2.4 (i) A formula $\varphi(\mathbf{x}) \in \mathrm{FO}(\tau)$ is *ℓ-local* iff $\varphi \equiv \varphi^{N^\ell(\mathbf{x})}$.

(ii) For any $\varphi(\mathbf{x})$ we write $\varphi^\ell(\mathbf{x})$ for the ℓ-local version $\varphi^{N^\ell(\mathbf{x})}$ of φ. If $q = \mathrm{qr}(\varphi)$, we refer to φ^ℓ as a local formula of Gaifman rank (ℓ, q).

(iii) A basic ℓ-local sentence is a sentence of the form

$$\exists x_1 \ldots \exists x_m \bigwedge_{i<j} d(x_i,x_j) > 2\ell \;\wedge\; \bigwedge_i \psi^\ell(x_i),$$

asserting the existing of an ℓ-scattered m-tuple whose components satisfy the ℓ-local formula $\psi^\ell(x)$. If $q = \mathrm{qr}(\psi)$, we regard the above basic local sentence as one of Gaifman rank (ℓ, q, m).

Example 2.2.5 The formula expressing $d(x,y) \leqslant \ell$ is $\lceil \ell/2 \rceil$-local (about x and y); $\exists y\big(d(x,y) \leqslant k \wedge \varphi^\ell(y)\big)$ is $(k+\ell)$-local (about x).

As $N^\ell(\mathbf{a}) \subseteq N^{\ell'}(\mathbf{a})$ for $\ell \leqslant \ell'$, any ℓ-local formula is also ℓ'-local for any $\ell' \geqslant \ell$.

Locality properties of FO Locality criteria can be used to establish $\simeq_m$, and hence $\equiv_m$ between relational structures $\mathfrak{A}, \mathfrak{B}$. Consider systems of sets of partial isomorphisms

$$\begin{array}{rrcl} \text{(Hanf)} & I_k & = & \big\{\mathbf{a} \mapsto \mathbf{b} \in \mathrm{Part}(\mathfrak{A},\mathfrak{B})\colon \mathfrak{A} \restriction N^{\ell_k}(\mathbf{a}),\mathbf{a} \simeq \mathfrak{B} \restriction N^{\ell_k}(\mathbf{b}),\mathbf{b}\big\} \\ \text{or, (Gaifman)} & I_k & = & \big\{\mathbf{a} \mapsto \mathbf{b} \in \mathrm{Part}(\mathfrak{A},\mathfrak{B})\colon \mathfrak{A} \restriction N^{\ell_k}(\mathbf{a}),\mathbf{a} \simeq_{r_k} \mathfrak{B} \restriction N^{\ell_k}(\mathbf{b}),\mathbf{b}\big\} \end{array}$$

for suitable choices of the parameters ℓ_k and r_k. The theorems of Hanf and Gaifman, respectively, give suitable overall conditions on local behaviours in $\mathfrak{A}$ and $\mathfrak{B}$ such that corresponding systems satisfy the back-and-forth conditions. Thus, local similarity conditions allow us to establish similarity in the sense of $\simeq_m$ and $\equiv_m$. Qualitatively, these techniques show that all of FO is of an essentially local nature over relational structures (classically as well as in FMT).

Hanf's theorem In Hanf's theorem, the overall condition specifies that each isomorphism type of ℓ-neighbourhoods is realised in $\mathfrak{A}$ and $\mathfrak{B}$ by the same number of elements. An N^ℓ isomorphism type ι is specified by a structure $\mathfrak{C}, c$ with distinguished element c (its centre) such that $\mathfrak{C} \restriction N^\ell(c) = \mathfrak{C}$; an element a of $\mathfrak{A}$ realises this isomorphism type if $\mathfrak{A} \restriction N^\ell(a), a \simeq \mathfrak{C}, c$.

Definition 2.2.6 $\mathfrak{A}$ and $\mathfrak{B}$ are *ℓ-Hanf-equivalent* if, for every N^ℓ isomorphism type ι, the number of elements in $\mathfrak{A}$ and $\mathfrak{B}$, respectively, that realise ι are equal.

Note that ℓ'-Hanf-equivalence implies ℓ-Hanf-equivalence if $\ell \leqslant \ell'$.

Lemma 2.2.7 *Let $\mathfrak{A}, \mathfrak{B} \in \mathsf{FIN}(\tau)$ be ℓ-Hanf-equivalent, $L = 3\ell + 1$. Then any $p = \mathbf{a} \mapsto \mathbf{b}$ such that*

$$\mathfrak{A} \restriction N^L(\mathbf{a}), \mathbf{a} \simeq \mathfrak{B} \restriction N^L(\mathbf{b}), \mathbf{b}$$

admits back-and-forth extensions $p' = \mathbf{a}a \mapsto \mathbf{b}b$ for which

$$\mathfrak{A} \restriction N^\ell(\mathbf{a}a), \mathbf{a}a \simeq \mathfrak{B} \restriction N^\ell(\mathbf{b}b), \mathbf{b}b.$$

Consequently, if $\mathfrak{A}$ and $\mathfrak{B}$ are $\frac{3^{m-1}-1}{2}$-Hanf-equivalent, then the following form a back-and-forth system for $\mathsf{G}_m(\mathfrak{A}, \mathfrak{B})$:

$$I_k = \big\{(\mathbf{a} \mapsto \mathbf{b}) \in \mathrm{Part}(\mathfrak{A}, \mathfrak{B}) \colon \mathfrak{A} \restriction N^{\ell_k}(\mathbf{a}), \mathbf{a} \simeq \mathfrak{B} \restriction N^{\ell_k}(\mathbf{b}), \mathbf{b}\big\} \quad \textit{for} \quad \ell_k = \frac{3^k - 1}{2}.$$

So any $\frac{3^{m-1}-1}{2}$-Hanf-equivalent $\mathfrak{A}$ and $\mathfrak{B}$ are m-isomorphic.

Proof Let $p \colon \mathbf{a} \mapsto \mathbf{b}$ such that $\rho \colon \mathfrak{A} \restriction N^L(\mathbf{a}), \mathbf{a} \simeq \mathfrak{B} \restriction N^L(\mathbf{b}), \mathbf{b}$, $L = 3\ell + 1$.

We show that p has extensions p' as required. Consider w.l.o.g. the forth-requirement for some $a \in A$. We need to provide $b \in B$ s.t. $\mathfrak{A} \restriction N^\ell(\mathbf{a}a), \mathbf{a}a \simeq \mathfrak{B} \restriction N^\ell(\mathbf{b}b), \mathbf{b}b$.

Case 1: $a \in N^{2\ell+1}(\mathbf{a})$. Then $N^\ell(a) \subseteq N^L$. Choosing $b := \rho(a)$ we get $\rho \colon \mathfrak{A} \restriction N^L(\mathbf{a}), \mathbf{a}a \simeq \mathfrak{B} \restriction N^L(\mathbf{b}), \mathbf{b}b$. Therefore $\rho' := \rho \restriction N^\ell(\mathbf{a}a)$ shows that $p' \colon \mathbf{a}a \mapsto \mathbf{b}b$ is as desired.

Case 2: $a \notin N^{2\ell+1}(\mathbf{a})$. Then $N^\ell(a) \cap N^\ell(\mathbf{a}) = \emptyset$ and moreover $\mathfrak{A} \restriction \big(N^\ell(a) \cup N^\ell(\mathbf{a})\big)$ is the disjoint union of $\mathfrak{A} \restriction N^\ell(a)$ and $\mathfrak{A} \restriction N^\ell(\mathbf{a})$. We therefore just need to find $b \in B$ of the same N^ℓ isomorphism type as a and such that also $b \notin N^{2\ell+1}(\mathbf{b})$. Then the restriction of ρ to $N^\ell(\mathbf{a})$ can be combined with an isomorphism between the ℓ-neighbourhoods of a and b, respectively.

Let ι be the isomorphism type of the ℓ-neighbourhood of a. The isomorphism type of $N^L(\mathbf{a})$ determines the number of realisations of ι within $N^{2\ell+1}(\mathbf{a})$, which (through ρ) must be the same as the number of realisations of ι within $N^{2\ell+1}(\mathbf{b})$ (note that $y \in N^{2\ell+1}(\mathbf{x})$ implies that $N^\ell(y) \subseteq N^L(\mathbf{x})$).

By ℓ-Hanf-equivalence, therefore, $\mathfrak{B}$ must also have the same number of realisations of ι outside $N^{2\ell+1}(\mathbf{b})$ as $\mathfrak{A}$ has outside $N^{2\ell+1}(\mathbf{a})$. Any such realisation will do.

For $\ell_k = \frac{3^k-1}{2}$ we have $\ell_{k+1} = 3\ell_k + 1$. At the bottom level, for $k = 0$, observe that even isomorphism of the 0-neighbourhoods of $\mathbf{a}$ and $\mathbf{b}$ implies that $(\mathbf{a} \mapsto \mathbf{b}) \in \mathrm{Part}(\mathfrak{A}, \mathfrak{B})$. So the $\simeq_m$ claim follows. □

Example 2.2.8 Consider finite undirected graphs built from connected components that are simple cycles $\mathfrak{C}_n$ of length n, for various n. Clearly any two elements have isomorphic ℓ-neighbourhoods in cycles of lengths $n > 2\ell + 1$. Hence, for $n > 2\ell + 1$, $\mathfrak{C}_{2n}$ and the disjoint union of two copies of $\mathfrak{C}_n$ are ℓ-Hanf-equivalent. We therefore get the following (compare Exercise 2.1.15).

Corollary 2.2.9 *Connectivity of finite undirected graphs,* CONN, *is not* FO*-definable.*

Exercise 2.2.10 Show that k-connectivity of finite graphs is not FO-definable. (A graph is k-connected if it remains connected after the removal of any k edges.)

As a further application, there is this stronger result about connectivity.

Proposition 2.2.11 CONN *is not even definable even by a monadic existential second-order sentence. I.e., there is no second-order sentence of the form* $\exists X_1 \ldots \exists X_s \psi(\mathbf{X})$ *with* $\psi \in \mathrm{FO}(\{E\} \cup \{X_1, \ldots, X_s\})$ *for unary relation variables* X_i *such that for all finite undirected graphs* $\mathfrak{A}$

$$\mathfrak{A} \textit{ connected} \quad \textit{iff} \quad (\mathfrak{A}, P_1, \ldots, P_s) \models \psi \textit{ for some } P_i \subseteq A.$$

Proof (by cutting and gluing) Suppose to the contrary that $\exists \mathbf{X}\psi(\mathbf{X})$ were as desired. Consider a cycle $\mathfrak{C}_N$ of length N. We let $C_N := \mathbb{Z}_N$ and put a symmetric edge between u and u' where $'\colon u \mapsto u' := u + 1 \bmod N$.

As $\mathfrak{C}_N$ is connected, there is an expansion $(\mathfrak{C}_N, \mathbf{P}) \models \psi$. Let ℓ be such that ℓ-Hanf-equivalence preserves ψ. For $N > 2\ell + 1$ any ℓ-neighbourhood in $\mathfrak{C}$ has precisely $2\ell + 1$ elements, and in $(\mathfrak{C}_N, \mathbf{P})$ there are only $s^{2\ell+1}$ many distinct $\mathbf{P}$-colourings of $(2\ell + 1)$-chains.

Hence, for sufficiently large N, there must be two nodes, u and v at distance $d(u, v) > 2\ell + 1$ in $(\mathfrak{C}_N, \mathbf{P})$ such that

$$(\mathfrak{C}_N, \mathbf{P}) \restriction N^\ell(u), u \simeq (\mathfrak{C}_N, \mathbf{P}) \restriction N^\ell(v), v$$

via an isomorphism of the form $x \mapsto x + m \bmod N$, for a suitable translation m. In particular this isomorphism also maps u' to v'. We now change just the edge relation near u and v by swapping u' and v':

$$E' := \big(E \setminus \{(u, u'), (u', u), (v, v'), (v', v)\}\big) \cup \{(u, v'), (v', u), (v, u'), (u', v)\}.$$

Then in $(\mathfrak{C}'_N, \mathbf{P})$ every node has exactly the same N^ℓ isomorphism type as in $(\mathfrak{C}_N, \mathbf{P})$, up to isomorphism. So $(\mathfrak{C}'_N, \mathbf{P})$ and $(\mathfrak{C}_N, \mathbf{P})$ are ℓ-Hanf-equivalent and therefore $(\mathfrak{C}'_N, \mathbf{P}) \models \psi$, too. But $\mathfrak{C}'_N$ is disconnected, consisting of two cycles rather than one. Contradiction. $\Box$

Exercise 2.2.12 Show that the binary reachability query $D_{<\infty}$ on the other hand is defined by the existential monadic second-order formula $\varphi(x, y) = \exists X \psi(x, y, X)$ where $\psi \in \mathrm{FO}(\{E, X\})$ says that either $x = y$ or

$x, y \in X$,
x and y each have precisely one immediate E-neighbour in X
all elements of X apart form x and y have precisely two direct E-neighbours in X.

Does this work in infinite graphs as well?

Gaifman's theorem This theorem shows that FO can only express structural properties of an essentially local nature. Compare definition 2.2.4 above.

Theorem 2.2.13 (Gaifman's theorem) *Any formula of* FO *is logically equivalent to a boolean combination of local formulae and basic local sentences.*

One can prove this "directly" by induction on φ (Gaifman's original proof). We make a detour through games.

Definition 2.2.14 (ℓ, q, m)-Gaifman-equivalence, $\mathfrak{A}, \mathbf{a} \equiv^\ell_{q,m} \mathfrak{B}, \mathbf{b}$, is defined by the following conditions:

(i) $\mathfrak{A}, \mathbf{a}$ and $\mathfrak{B}, \mathbf{b}$ satisfy the same ℓ-local formulae $\varphi^\ell(\mathbf{x})$ for $\mathrm{qr}(\varphi) \leqslant q$.

(ii) $\mathfrak{A}$ and $\mathfrak{B}$ satisfy the same basic local sentences of ranks (ℓ', q', m') for all $\ell' \leqslant \ell$, $q' \leqslant q$ and $m' \leqslant m$.

Note that $\equiv^\ell_{q,m}$ has finite index, and can be regarded as induced by the class $\Delta(\ell, q, m)$ of basic local formulae and sentences of rank up to (ℓ, q, m) in the sense of Proposition 1.3.7. To prove the theorem, it suffices to show that for any given m $(m = \mathrm{qr}(\varphi))$ there are (ℓ, q, n) such that for any $\mathfrak{A}, \mathfrak{B}$,

$$\mathfrak{A}, \mathbf{a} \equiv^\ell_{q,n} \mathfrak{B}, \mathbf{b} \qquad \Rightarrow \qquad \mathfrak{A}, \mathbf{a} \simeq_m \mathfrak{B}, \mathbf{b},$$

for then φ is equivalent to a boolean combination of basic local formulae and sentences of ranks up to (ℓ, q, n).

Exercise 2.2.15 Fill in the details for the above arguments.

Lemma 2.2.16 *Let $\mathfrak{A}$ and $\mathfrak{B}$ be (L, Q, m)-Gaifman-equivalent for sufficiently large L, Q. Then any $p = \mathbf{a} \mapsto \mathbf{b}$ with $|\mathbf{a}| = |\mathbf{b}| < m$ and such that*

$$\mathfrak{A} \restriction N^L(\mathbf{a}), \mathbf{a} \equiv_Q \mathfrak{B} \restriction N^L(\mathbf{b}), \mathbf{b}$$

admits back-and-forth extensions $p' = \mathbf{a}a \mapsto \mathbf{b}b$ for which

$$\mathfrak{A} \restriction N^\ell(\mathbf{a}a), \mathbf{a}a \equiv_q \mathfrak{B} \restriction N^\ell(\mathbf{b}b), \mathbf{b}b.$$

Consequently, for some suitably fast growing sequence (ℓ_k, q_k) we get the following. If $\mathfrak{A} \equiv^\ell_{q,m} \mathfrak{B}$ for $(\ell, q) = (\ell_m, q_m)$, then the following form a back-and-forth system for $\mathsf{G}_m(\mathfrak{A}, \mathfrak{B})$:

$$I_k = \big\{(\mathbf{a} \mapsto \mathbf{b}) \in \mathrm{Part}(\mathfrak{A}, \mathfrak{B})\colon |\mathbf{a}| = |\mathbf{b}| \leqslant m - k\,,\ \mathfrak{A} \restriction N^{\ell_k}(\mathbf{a}), \mathbf{a} \simeq_{q_k} \mathfrak{B} \restriction N^{\ell_k}(\mathbf{b}), \mathbf{b}\big\}.$$

Similarly, for $\mathbf{a} \in A^n$ and $\mathbf{b} \in B^n$: if $\mathfrak{A}, \mathbf{a}$ and $\mathfrak{B}, \mathbf{b}$ are $(\ell_m, q_m, m+n)$-Gaifman-equivalent, then they are m-isomorphic.

Proof Let $\mathfrak{A}, \mathbf{a}$ and $\mathfrak{B}, \mathbf{b}$ be (L, Q, m)-Gaifman-equivalent (bounds on L, Q will be collected during the proof), $p = \mathbf{a} \mapsto \mathbf{b}$ with $|\mathbf{a}| = |\mathbf{b}| < m$ such that

$$\mathfrak{A} \restriction N^L(\mathbf{a}), \mathbf{a} \ \equiv_Q \ \mathfrak{B} \restriction N^L(\mathbf{b}), \mathbf{b},$$

and (for the forth property), $a \in A$ be given. We need to find $b \in B$ such that

$$\mathfrak{A} \restriction N^\ell(\mathbf{a}a), \mathbf{a}a \ \equiv_q \ \mathfrak{B} \restriction N^\ell(\mathbf{b}b), \mathbf{b}b.$$

Case 1 (a close to $\mathbf{a}$): $a \in N^{2\ell+1}(\mathbf{a})$. Then $N^\ell(a) \subseteq N^{3\ell+1}(\mathbf{a})$. We assume that $L \geqslant 3\ell + 1$. Then, by the forth property for $\mathfrak{A} \restriction N^L(\mathbf{a}), \mathbf{a} \simeq_Q \mathfrak{B} \restriction N^L(\mathbf{b}), \mathbf{b}$ we find $b \in N^L(\mathbf{b})$ such that $\mathfrak{A} \restriction N^L(\mathbf{a}), \mathbf{a}a \simeq_{Q-1} \mathfrak{B} \restriction N^L(\mathbf{b}), \mathbf{b}b$. Provided Q is sufficiently large this implies that also $N^\ell(b) \subseteq N^{3\ell+1}(\mathbf{b})$ and $\mathfrak{A} \restriction N^\ell(\mathbf{a}a), \mathbf{a}a \equiv_q \mathfrak{B} \restriction N^\ell(\mathbf{b}b), \mathbf{b}b$.

Case 2 (a far from $\mathbf{a}$): $a \not\in N^{2\ell+1}(\mathbf{a})$. $N^\ell(a)$ and $N^\ell(\mathbf{a})$ are disjoint and $\mathfrak{A} \restriction N^\ell(\mathbf{a}a)$ is the disjoint union of $\mathfrak{A} \restriction N^\ell(\mathbf{a})$ and $\mathfrak{A} \restriction N^\ell(a)$. It suffices to find $b \in B$ that is also far from $\mathbf{b}$ and such that $\mathfrak{B} \restriction N^\ell(b), b \simeq_q \mathfrak{A} \restriction N^\ell(a), a$. (Strategies for the q-round games are compatible with disjoint unions, cf. Exercise 2.1.10).

In this case we need to rely on basic local sentences to guarantee that in $\mathfrak{B}$ we can find a matching $b \notin N^{2\ell+1}(\mathbf{b})$. Let $\psi(x)$ be the qfr-rank q formula that characterises the $\simeq_q$-type of a in $\mathfrak{A} \restriction N^\ell(a)$, $\psi^\ell(x)$ its ℓ-local version.

Case 2.1 $\mathfrak{A}$ has a $(2\ell+1)$-scattered m-tuple of elements realising ψ^ℓ (i.e., with ℓ-neighbourhoods q-isomorphic to that of a).

For $L \geqslant 2\ell+1$ and $Q \geqslant q$ this fact is preserved in (L, Q, m)-Gaifman-equivalence. Hence $\mathfrak{B}$ also has such an m-tuple. As $m > |\mathbf{b}|$, at least one member of this tuple must lie outside $N^{2\ell+1}(\mathbf{b})$ (each $N^{2\ell+1}(b_i)$ can hold at most one member as its diameter is at most $2(2\ell+1)$).

Case 2.2 For some $n < m$, the maximal $(2\ell+1)$-scattered tuple of elements realising ψ^ℓ in $\mathfrak{A}$ has size n. Provided $L \geqslant 2\ell+1$ and $Q \geqslant q$, the same n works in $\mathfrak{B}$.

We now compare this n with n_0, the maximal size of any $(2\ell+1)$-scattered tuple of elements within $N^{2\ell+1}(\mathbf{a})$ realising ψ^ℓ. Clearly $n_0 \leqslant n$.

The same number n_0 works for $\mathfrak{B}$, provided Q is sufficiently large to express the existence of a $(2\ell+1)$-scattered n_0-tuple and non-existence of an (n_0+1)-tuple for ψ^ℓ, and if $L \geqslant 3\ell+1$.

If $n_0 < n$, then also in $\mathfrak{B}$ we find an element outside $N^{2\ell+1}(\mathbf{b})$ that satisfies ψ^ℓ and thus has an ℓ-neighbourhood q-isomorphic to that of a.

If $n = n_0$, then all realisations of ψ^ℓ in $\mathfrak{A}, \mathbf{a}$, together with their ℓ-neighbourhoods lie inside $N^L(\mathbf{a})$ if $L \geqslant 7\ell+3$. It follows that for $L \geqslant 7\ell+3$ and Q sufficiently large to express the existence of a witness for $\psi^\ell(x)$ at distance greater than $2\ell+1$ from $\mathbf{x}$, the existence of b as desired is guaranteed by $\mathfrak{A} \restriction N^L(\mathbf{a}), \mathbf{a} \simeq_Q \mathfrak{B} \restriction N^L(\mathbf{b}), \mathbf{b}$. $\Box$

2.3 Two Preservation Theorems on Restricted Classes

Minimal models For the two most fundamental classical preservation theorems – Łos–Tarski: Theorem 0.1.5 and Lyndon–Tarski: Theorem 0.1.6 – corresponding notions of minimal models play an important role. In connection with preservation under extensions, the minimal models of interest are minimal w.r.t. the substructure relation.

Definition 2.3.1 A structure $\mathfrak{B} \in \mathcal{C}$ is substructure-minimal ($\subseteq$-minimal) within the class $\mathcal{C}$ if $\mathcal{C}$ contains no proper substructure of $\mathfrak{B}$.

If $\mathcal{C} \subseteq \mathsf{FIN}(\tau)$ is closed under extensions within $\mathsf{FIN}(\tau)$, then it is generated by its $\subseteq$-minimal elements in the sense that

$$\mathcal{C} = \{\mathfrak{A} \in \mathsf{FIN}(\tau) \colon h \colon \mathfrak{B} \xrightarrow{\subseteq} \mathfrak{A} \text{ for some } \subseteq\text{-minimal } \mathfrak{B} \in \mathcal{C}\}.$$

Observation 2.3.2 Let $\mathcal{C} \subseteq \mathsf{FIN}(\tau)$ (τ finite and relational) be closed under extensions within $\mathsf{FIN}(\tau)$. Then the following are equivalent:

(i) $\mathcal{C} = \mathsf{FMOD}(\varphi)$ for some existential FO(τ)-sentence φ.

(ii) $\mathcal{C}$ has only finitely many $\subseteq$-minimal elements up to $\simeq$.

Proof Note that (ii) says that there are $\mathfrak{B}_1, \ldots, \mathfrak{B}_N \models \varphi$ such that

$$\mathcal{C} = \big\{\mathfrak{A} \in \mathsf{FIN}(\tau) \colon \text{ there exists } h \colon \mathfrak{B}_i \xrightarrow{\subseteq} \mathfrak{A} \text{ for some } 1 \leqslant i \leqslant N\big\}.$$

For (ii) ⇒ (i) it then suffices to let φ be the disjunction over the existential sentences $\chi[\mathfrak{B}]$ consisting of the existentially quantified algebraic diagram of $\mathfrak{B}$ for $\mathfrak{B} = \mathfrak{B}_1, \ldots, \mathfrak{B}_N$. Formally:

$$\varphi := \bigvee_{i=1,\ldots,N} \chi[\mathfrak{B}_i] \; , \qquad \text{where} \qquad \chi[\mathfrak{B}] := \exists (x_b)_{b \in B} \bigwedge D(\mathfrak{B})((x_b)_{b\in B})$$

is the result of formalising the algebraic diagram of $\mathfrak{B}$ (cf. Definition 0.1.2) with variables for the (finitely many) elements of B, rather than constants, and existentially quantifying over all of these.

For (i) ⇒ (ii), let $\varphi = \exists \mathbf{x} \xi(\mathbf{x})$ be an existential sentence as in (i), w.l.o.g. prenex with quantifier-free kernel formula $\xi(\mathbf{x})$ in a variable tuple $\mathbf{x} = (x_1, \ldots, x_k)$. It is clear that minimal models of φ cannot have more than k elements; hence there are only finitely many (τ is finite!). □

An analogous criterion applies in connection with preservation under homomorphisms. We are here seeking to generate the class $\mathcal{C}$ from finitely many $\mathfrak{B}_1, \ldots, \mathfrak{B}_N$ by homomorphic embeddings:

$$\mathcal{C} = \big\{ \mathfrak{A} \in \mathsf{FIN}(\tau) \colon \text{ there exists } h \colon \mathfrak{B}_i \xrightarrow{\text{hom}} \mathfrak{A} \text{ for some } 1 \leqslant i \leqslant N \big\}.$$

What kind of minimality criterion characterises the right generators $\mathfrak{B} \in \mathcal{C}$ in a class that is closed under homomorphisms?

For the following recall that the *weak* substructure relation $\mathfrak{A} \subseteq_w \mathfrak{B}$ between relational structures requires $A \subseteq B$ and $R^{\mathfrak{A}} \subseteq R^{\mathfrak{B}}$ for all $R \in \tau$ (rather than $R^{\mathfrak{A}} = R^{\mathfrak{B}} \restriction A$). Note that any weak substructure $\mathfrak{A} \subseteq_w \mathfrak{B}$ is homomorphically embedded into $\mathfrak{B}$ by the identity map. A proper weak susbstructure $\mathfrak{A} \subseteq_w \mathfrak{B}$ is one with $\mathfrak{A} \neq \mathfrak{B}$, which may be because $A \subsetneq B$ or because, even when $A = B$, $R^{\mathfrak{A}} \subsetneq R^{\mathfrak{B}}$ for some relation $R \in \tau$.

Definition 2.3.3 A structure $\mathfrak{B} \in \mathcal{C}$ is $\subseteq_w$-minimal within the class $\mathcal{C}$ if no proper weak substructure of $\mathfrak{B}$ is in $\mathcal{C}$.

The proof of the following is entirely analogous to the reasoning behind Observation 2.3.2 above: the role of the algebraic diagrams is now played by the positive algebraic diagrams (cf. Definition 0.1.2).

Observation 2.3.4 Let $\mathcal{C} \subseteq \mathsf{FIN}(\tau)$ (τ finite and relational) be closed under homomorphisms within $\mathsf{FIN}(\tau)$. Then the following are equivalent:

(i) $\mathcal{C} = \mathsf{FMOD}(\varphi)$ for some existential positive FO(τ)-sentence φ.

(ii) $\mathcal{C}$ has finitely many $\subseteq_w$-minimal members, up to $\simeq$.

Exercise 2.3.5 Give a detailed proof of Observation 2.3.4.

Exercise 2.3.6 Let τ be finite and relational. Show that, for any $\mathcal{C} \subseteq \mathsf{FIN}(\tau)$ closed under homomorphisms within $\mathsf{FIN}(\tau)$ the following are equivalent:

(i) $\mathcal{C}$ has finitely many $\subseteq_w$-minimal members, up to isomorphism.

(ii) $\mathcal{C}$ has finitely many $\subseteq$-minimal members, up to isomorphism.

Aside: homomorphism equivalence and cores We call two structures *homomorphically equivalent* if there are homomorphic embeddings in both directions: $h_1 \colon \mathfrak{A} \xrightarrow{\text{hom}} \mathfrak{B}$ and $h_2 \colon \mathfrak{B} \xrightarrow{\text{hom}} \mathfrak{A}$. Let $\mathfrak{A} \xleftrightarrow{\text{hom}} \mathfrak{B}$ be shorthand for this. Minimal generators of the equivalence class w.r.t. homomorphic equivalence can always be found among the weak substructures of a given finite structure as follows.

Definition 2.3.7 A τ-structure is called a *core* if it is not homomorphically equivalent to any proper weak substructure. A structure $\mathfrak{A}_0 \subseteq_w \mathfrak{A}$ such that $\mathfrak{A}_0 \xleftrightarrow{\text{hom}} \mathfrak{A}$ that is a core is called a *core of* $\mathfrak{A}$.

It follows immediately from the definitions that every $\subseteq_w$-minimal member of a class $\mathcal{C} \subseteq \mathsf{FIN}(\tau)$ that is closed under homomorphisms within $\mathsf{FIN}(\tau)$ is a core.

Exercise 2.3.8 Give some examples of graphs that are cores; in particular: which cycles are cores?

Exercise 2.3.9 Show the following, for any finite relational τ:

(i) every $\mathfrak{A} \in \mathsf{FIN}(\tau)$ has a core: $\mathfrak{A}$ is homomorphically equivalent to some core $\mathfrak{A}_0 \subseteq_w \mathfrak{A}$.

(ii) any two cores of any two homomorphically equivalent finite structures are isomorphic. In particular, the core of any given $\mathfrak{A} \in \mathsf{FIN}(\tau)$ is unique up to isomorphism.

(iii) any core $\mathfrak{A}_0 \subseteq_w \mathfrak{A}$, is related to $\mathfrak{A}$ by a *retract*, i.e., by a homomorphism whose restriction to $A_0 \subseteq A$ is the identity: $h \colon \mathfrak{A} \xrightarrow{\text{hom}} \mathfrak{A}_0$ with $h \restriction A_0 = \mathrm{id}_{A_0}$.

2.3.1 Łos–Tarski over finite successor strings

The following proposition re-captures some of the Łos–Tarski theorem, which fails in FMT, for some very restricted class of finite relational structures. Compare Theorem 0.1.5 for the classical theorem and Proposition 1.2.6 for its failure in FMT.

The result discussed here is a special case of a more general result of Atserias, Dawar and Grohe, which among others applies to classes of finite structures with acyclic Gaifman graphs, to so-called *wide* classes of structures (see section 2.3.2 below) and to the class of all finite structures of treewidth $\leqslant k$, for fixed $k \in \mathbb{N}$.

We deal with the class $\mathcal{C}$ of simple string structures. Let $\tau = \{S\} \cup \tau_0$ be a vocabulary with one binary relation S and a finite set τ_0 of unary predicates. Let $\mathcal{S} \subseteq \mathsf{FIN}(\tau)$ be the class of finite coloured successor strings, i.e., of finite τ-structures whose S-reduct is isomorphic to the disjoint union of S-structures of the form

$$([n], S) = \big(\{1, \ldots, n\}, \{(i, i+1) \colon 1 \leqslant i < n\}\big).$$

We want to show the analogue of the Łos–Tarski theorem over $\mathcal{S}$, on the basis of Gaifman's theorem.

Proposition 2.3.10 (Atserias–Dawar–Grohe) *Over the class $\mathcal{S} \subseteq \mathsf{FIN}(\tau)$ of finite string structures t.f.a.e. for any* $\mathrm{FO}(\tau)$*-sentence φ:*

(i) *φ is preserved under extensions within $\mathcal{S}$.*

(ii) *over $\mathcal{S}$, $\varphi \equiv \hat{\varphi}$ for some existential $\hat{\varphi} \in \mathrm{FO}(\tau)$.*

The technical core is the following. This will allow us to derive a bound on the size of $\subseteq$-minimal models.

Lemma 2.3.11 *There is a function $N(\ell, m)$ such that every connected $\mathfrak{A} \in \mathcal{S}$ of length greater than $N(\ell, m)$ possesses a proper substructure $\mathfrak{A}' \subsetneq \mathfrak{A}$ and a disjoint extension $\mathfrak{A} + \mathfrak{B}$, both within $\mathcal{S}$, such that $\mathfrak{A}'$ and $\mathfrak{A} + \mathfrak{B}$ are (ℓ, q, m)-Gaifman-equivalent (for all q).*

Note that, if $\mathfrak{A} \models \varphi$ and φ is preserved under extensions within $\mathcal{S}$ and under (ℓ, q, m)-Gaifman equivalence, then $\mathfrak{A} + \mathfrak{B} \models \varphi$ (by extension preservation) and $\mathfrak{A}' \models \varphi$ (by (ℓ, q, m)-Gaifman equivalence). Therefore no connected $\mathfrak{A}$ of length greater than $N(\ell, m)$ can be $\subseteq$-minimal in $\mathsf{FMOD}(\varphi)$.

Exercise 2.3.12 Use the fact that (ℓ, q, m)-Gaifman equivalence is preserved in disjoint unions to derive the following from the lemma, for any φ that is preserved under extensions within $\mathcal{S}$ as well as under (ℓ, q, m)-Gaifman equivalence:

(i) $\subseteq$-minimal models of φ have connected components of lengths $\leqslant N(\ell, m)$.

(ii) there is a bound on the size of $\subseteq$-minimal models of φ.

As the size – and therefore the number – of $\subseteq$-minimal models of φ is finitely bounded, a disjunction over the existentially quantified algebraic diagrams of all such models can serve as $\hat{\varphi}$ to prove the expressive completeness assertion in the proposition.

Proof [of the lemma] A crucial feature of $\mathcal{S}$ for our purposes is the following. For given $\alpha, s \in \mathbb{N}$ there is an N such that any connected $\mathfrak{A} \subseteq \mathcal{S}$ of length greater than N satisfies the following: for any subset A_0 of size up to s there is an interval of length α of $\mathfrak{A}$ that is disjoint from A_0. We shall use this in the following.

Let ℓ and m be given. Put $L = 2\ell + 1$ and let K be a bound on the number of isomorphism types of L-neighbourhoods that can occur in structures in $\mathcal{S}$. (The neighbourhoods in question are connected structures of length up to $2L + 1$ from $\mathcal{S}$.) Let us say that such an isomorphism type ι is *frequent* in a given connected $\mathfrak{A} \in \mathcal{S}$, if it is realised at least $4mL$ times in $\mathfrak{A}$. Note that we have an upper bound (of $4mLK$) for the number of elements that can realise non-frequent L-neighbourhoods in any $\mathfrak{A}$.

The multiplicity threshold $4mL$ is chosen such that any frequent type is necessarily realised by an L-scattered configuration of m elements in $\mathfrak{A}$: an L-neighbourhood can overlap with at most $4L$ others.

Now for $\alpha \in \mathbb{N}$, there is an $N(\alpha) \in \mathbb{N}$ such that any $\mathfrak{A}$ of length greater than $N(\alpha)$ must contain an interval I of length α such that

(i) only frequent ι are realised by elements of I,

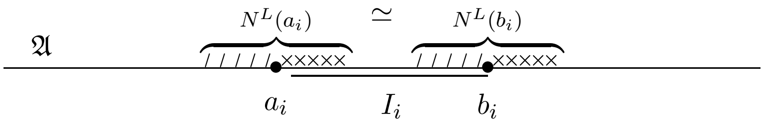
(ii) all frequent ι are still frequent in $\mathfrak{A} \restriction (A \setminus I)$.

By choosing the length α sufficiently large we are guaranteed an interval I of this kind such that moreover

(iii) I contains a sequence of $2m$ many L-scattered points whose (disjoint) L-neighbourhoods are fully contained in I and all of the same type ι.

We now choose the lower bound $N(\ell, m)$ such that an interval I satisfying (i), (ii) and (iii) must exist in any connected $\mathfrak{A} \in \mathcal{S}$ of length greater than $N(\ell, m)$.

W.l.o.g. we may assume that the given $\mathfrak{A}$ of length greater than $N(\ell, m)$ is realised on a set $A = [M] = \{1, \ldots, M\} \subseteq \mathbb{N}$ with the natural successor relation S from $\mathbb{N}$. Let $a_1 < b_1 < \cdots < a_m < b_m$ be a sequence of $2m$ elements of $\mathfrak{A}$ from an interval I according to (i), (ii) and (iii) above.



Let then I_i be the interval $I_i = (a_i, b_i] \subseteq A$. We let $\mathfrak{A}' \subseteq \mathfrak{A}$ be the substructure induced on $A \setminus \bigcup I_i$. Let $\mathfrak{A}_i = \mathfrak{A} \restriction I_i$, $\mathfrak{B}_i \simeq \mathfrak{A}_i$ a disjoint copy of $\mathfrak{A}_i$, and $\mathfrak{B}$ the disjoint union of these. We claim that

$$\mathfrak{A}' \equiv^{\ell}_{q,m} \mathfrak{A} + \mathfrak{B}.$$

For this, let us analyse the behaviour of ℓ-neighbourhoods $N^\ell(a)$ in $\mathfrak{A}'$, $\mathfrak{A}$ and $\mathfrak{B}$. Our aim is to find for any scattered configuration of up to m many ℓ-neighbourhoods in $\mathfrak{A}'$ a matching scattered configuration of even isomorphic ℓ-neighbourhoods in $\mathfrak{A} + \mathfrak{B}$, and vice versa. Clearly this implies (ℓ, q, m)-equivalence for every q.

Consider first $a \in \mathfrak{A}'$ such that $\mathfrak{A}' \restriction N^\ell(a) \not\simeq \mathfrak{A} \restriction N^\ell(a)$. This implies that, in $\mathfrak{A}$, $N^\ell(a) \subseteq N^L(a_i)$ or $N^\ell(a) \subseteq N^L(b_i)$ for some i. But then, since all these are disjoint and isomorphic, there are at least m many disjoint isomorphic copies of $\mathfrak{A}' \restriction N^\ell(a)$ in both $\mathfrak{A}'$ and in $\mathfrak{B}$.

The same reasoning applies to $a \in \mathfrak{A}_i \simeq \mathfrak{B}_i$ for which $\mathfrak{A}_i \restriction N^\ell(a) \not\simeq \mathfrak{A} \restriction N^\ell(a)$.

If $a \in A_0$ is such that $\mathfrak{A}' \restriction N^\ell(a) = \mathfrak{A} \restriction N^\ell(a)$, then $\mathfrak{A}' \restriction N^\ell(a) \subseteq \mathfrak{A} \subseteq \mathfrak{A} + \mathfrak{B}$.

It remains to argue that $\mathfrak{A} + \mathfrak{B}$ cannot have extra realisations of some isomorphism type of $\mathfrak{A} \restriction N^\ell(a)$ for $a \in I$, unless $\mathfrak{A}'$ already has m many disjoint realisations of this type. But any $\iota \simeq N^L(a)$ for $a \in I$ is frequent in $\mathfrak{A}$, and still frequent in $\mathfrak{A}' = \mathfrak{A} \restriction A \setminus \bigcup I_i$ by the choice of I. $\Box$

2.3.2 Lyndon–Tarski over wide classes of finite structures

Definition 2.3.13 A structure is (ℓ, m)-*wide* if its Gaifman graph contains an ℓ-scattered subset of size m. A class $\mathcal{C}$ of τ-structures is called *wide* if there is a function $N \colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$, such that, for all ℓ and m and $\mathfrak{A} \in \mathcal{C}$, if $|A| \geqslant N(\ell, m)$, then $\mathfrak{A}$ is (ℓ, m)-wide.

A typical example of a wide class is the class of graphs of fixed bounded degree.

Exercise 2.3.14 Show that the class of all finite graphs of degree up to d is wide, for any d. Hint: this involves a case distinction as to many connected components vs. components of large diameter.

The following is a restricted version of a more general result by Atserias, Dawar and Kolaitis.

Proposition 2.3.15 (Atserias–Dawar–Kolaitis) *Over any class $\mathcal{C}$ of finite τ-structures that is wide and closed under substructures and disjoint unions, t.f.a.e. for any* $\mathrm{FO}(\tau)$*-sentence φ:*

(i) *φ is preserved under homomorphisms within $\mathcal{C}$.*

(ii) *over $\mathcal{C}$, $\varphi \equiv \hat{\varphi}$ for some existential positive $\hat{\varphi} \in \mathrm{FO}(\tau)$.*

For the expressive completeness claim, it suffices to establish a bound on the size of $\subseteq_w$- or $\subseteq$-minimal models of φ in $\mathcal{C}$. Then Observation 2.3.4 and Exercise 2.3.6 prove the claim.

Since $\mathcal{C}$ is wide, a size bound may be inferred from any bound on the wideness of $\subseteq$-minimal models.

We sketch the argument that, for a first-order sentence φ that is preserved under $\equiv^{\ell}_{q,m}$ and under homomorphisms (within $\mathcal{C}$), there are $L, M \in \mathbb{N}$ such that no (L, M)-wide model of φ can be $\subseteq$-minimal. More precisely, there are

- M, large enough w.r.t. L, Q, such that such that within any L-scattered subset of size M in $\mathfrak{A} \models \varphi$ we find some pair of elements $a \neq b$ for which $\mathfrak{A}, a \equiv^{L}_{Q,0} \mathfrak{A}, b$;
- L and Q, large enough w.r.t. ℓ, q, such that $\mathfrak{A}, a \equiv^{L}_{Q,0} \mathfrak{A}, b$ implies the following transfer property for Gaifman rank $(\ell, q, 1)$-assertions:

$$\mathfrak{A} \Rightarrow^{\ell}_{q,1} \mathfrak{B} := \mathfrak{A}{\restriction}(A \setminus \{b\}),$$

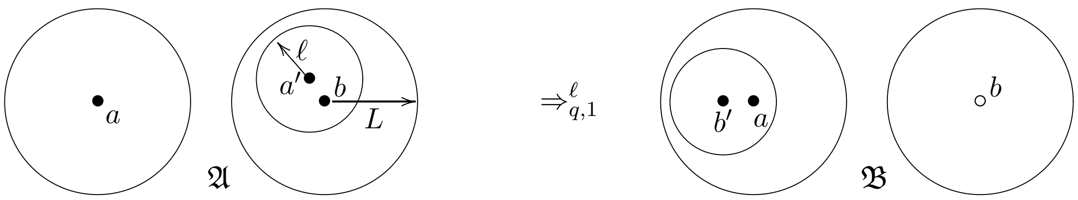
  meaning that every sentence of the form $\exists x \chi^{\ell}(x)$ where $\mathrm{qr}(\chi) \leqslant q$ that is true in $\mathfrak{A}$ remains true in $\mathfrak{B}$ (which is $\mathfrak{A}$ with b removed).

M simply needs to be chosen large w.r.t. the number of quantifier-rank Q types of single elements (in their L-neighbourhood) in order to guarantee the existence of distinct but $\equiv^{L}_{Q,0}$ equivalent nodes by the pigeon hole principle.

For such a and b, the desired transfer of $\exists x \chi^{\ell}(x)$-assertions follows from $\equiv^{L}_{Q,0}$ equivalence provided $L \geqslant 2\ell$ and Q large enough so that for all $\mathrm{qr}(\chi) \leqslant q$, the assertion

$$\exists x' \big(d(x, x') \leqslant \ell \wedge \chi^{\ell}(x') \big) \qquad (*)$$

is L-local and of quantifier rank $\leqslant Q$. Compare the diagram below for this proof sketch. In the non-trivial case $\mathfrak{A} \models \chi^{\ell}[a']$ for some $a' \in N^{\ell}(b)$, so that after the removal of b, there is no guarantee that still $\mathfrak{B} \models \chi^{\ell}[a']$. Using $\equiv^{L}_{Q,0}$ equivalence between a and b, though, $(*)$ is true of a if it is true at b. Hence there is a corresponding $b' \in N^{\ell}(a)$ such that $\mathfrak{A} \models \chi^{\ell}[b']$. So $\mathfrak{B} \models \chi^{\ell}[b']$ follows, since the L-neighbourhood of a is unaffected by the removal of b.



It follows that

$$\mathfrak{A} + m \cdot \mathfrak{B} \equiv^{\ell}_{q,m} m \cdot \mathfrak{B},$$

with disjoint sums of m isomorphic copies of $\mathfrak{B}$, plus one copy of $\mathfrak{A}$ on the left-hand side. Therefore, $\mathfrak{B} \models \varphi$ is a smaller model of φ, which shows that $\mathfrak{A}$ cannot have been $\subseteq$- or $\subseteq_w$-minimal:

$$\mathfrak{A} \xrightarrow{\text{hom}} \mathfrak{A} + m \cdot \mathfrak{B} \equiv^{\ell}_{q,m} m \cdot \mathfrak{B} \xrightarrow{\text{hom}} \mathfrak{B}.$$

2.4 Variation: Monadic Second-Order Logic and Its Game

Reminder: monadic second-order logic MSO extends FO by the possibility to quantify over subsets of the universe (unary relation variables) as well as over elements. We use letters like X, Y, Z for second-order relation variables, ranging over subsets of the universe of the structure at hand. So, e.g., $\mathfrak{A} \models \exists X \varphi(X)$ iff there is some $P \subseteq A$ for which $\mathfrak{A}, P \models \varphi$ (also written $\mathfrak{A} \models \varphi[P]$).

Example 2.4.1 Connectivity of undirected graphs is definable in MSO($\{E\}$, by the sentence

$$\forall X \big[\big(\exists x Xx \wedge \forall x \forall y (Xx \wedge Exy \rightarrow Xy) \big) \rightarrow \forall x Xx \big].$$

Many natural queries, in particular graph queries, are definable in MSO. The expressive power of MSO over finite linearly ordered coloured strings (word models) is analysed precisely in Part II.

Exercise 2.4.2 Give MSO-definitions for 3-colourability (as a boolean graph query) and of the binary reachability query over undirected graphs.

Formulae of MSO can have free first- and second-order variables, which we indicate as in $\varphi(\mathbf{X}, \mathbf{x})$: the free second-order variables are among those listed as $\mathbf{X}$ just as the free first-order variables are among those listed as $\mathbf{x}$. MSO quantifier rank is defined to represent the nesting depth of first- and second-order quantifications (counted as one). The qfr-rank of the above sample sentence is 3.

Definition 2.4.3 MSO(τ) stands for monadic second order logic over vocabulary τ. MSO-equivalence, $\equiv^{\mathrm{MSO}}$ and MSO-equivalence up to qfr-rank m, $\equiv^{\mathrm{MSO}}_m$, are defined in the obvious manner. E.g., $\mathfrak{A}, \mathbf{P}, \mathbf{a} \equiv^{\mathrm{MSO}}_m \mathfrak{B}, \mathbf{Q}, \mathbf{b}$ if for all $\varphi(\mathbf{X}, \mathbf{x}) \in \mathrm{MSO}(\tau)$ with $\mathrm{qr}(\varphi) \leqslant m$ we have $\mathfrak{A} \models \varphi[\mathbf{P}, \mathbf{a}] \Leftrightarrow \mathfrak{B} \models \varphi[\mathbf{Q}, \mathbf{b}]$.

The MSO game The FO-game is modified to allow a new kind of move that takes care of second-order quantifications. A position in the MSO-game over structures $\mathfrak{A}$ and $\mathfrak{B}$ consists of tuples of pebbled elements $\mathbf{a}$ and $\mathbf{b}$ (which establish a correspondence $\mathbf{a} \mapsto \mathbf{b}$ as before) and tuples of designated subsets, $\mathbf{P}$ and $\mathbf{Q}$, of A and B, respectively. We denote such a configuration as $(\mathfrak{A}, \mathbf{P}, \mathbf{a}; \mathfrak{B}, \mathbf{Q}, \mathbf{b})$.

A single round is governed by a challenge/response exchange, where **I** decides whether to play an element or a subset, and (as before) in which structure to play his challenge. **II** needs to respond by choosing a corresponding object (element or subset) in the opposite structure. So a set-move of **I** takes the game from position $(\mathfrak{A}, \mathbf{P}, \mathbf{a}; \mathfrak{B}, \mathbf{Q}, \mathbf{b})$ to some position $(\mathfrak{A}, \mathbf{P}P, \mathbf{a}; \mathfrak{B}, \mathbf{Q}Q, \mathbf{b})$; an element move, as before, takes the game from position $(\mathfrak{A}, \mathbf{P}, \mathbf{a}; \mathfrak{B}, \mathbf{Q}, \mathbf{b})$ to some position $(\mathfrak{A}, \mathbf{P}, \mathbf{a}a; \mathfrak{B}, \mathbf{Q}, \mathbf{b}b)$.

The winning conditions stipulate that **II** needs to maintain positions $(\mathfrak{A}, \mathbf{P}, \mathbf{a}; \mathfrak{B}, \mathbf{Q}, \mathbf{b})$ in which $(\mathbf{a} \mapsto \mathbf{b}) \in \mathrm{Part}((\mathfrak{A}, \mathbf{P}), (\mathfrak{B}, \mathbf{Q}))$.

Definition 2.4.4 The m-round MSO-game starting from position $(\mathfrak{A}, \mathbf{P}, \mathbf{a}; \mathfrak{B}, \mathbf{Q}, \mathbf{b})$, $\mathsf{G}^{\mathrm{MSO}}_m(\mathfrak{A}, \mathbf{P}, \mathbf{a}; \mathfrak{B}, \mathbf{Q}, \mathbf{b})$, consists of m rounds.

We say that **II** wins the game if she has a winning strategy to maintain locally isomorphic pebble configurations that also respect the subsets selected in set-moves.

Back-and-forth systems that correspond to winning strategies in the m-round MSO-game can be defined in analogy with those for the FO game, only that back-and-forth matches must be provided both for extensions by one further element and for extensions by one further subset.

$\mathfrak{A}, \mathbf{P}, \mathbf{a}$ and $\mathfrak{B}, \mathbf{Q}, \mathbf{b}$ are MSO-m-isomorphic, $\mathfrak{A}, \mathbf{P}, \mathbf{a} \simeq_m^{\mathrm{MSO}} \mathfrak{B}, \mathbf{Q}, \mathbf{b}$ if there is a back-and-forth system for $\mathsf{G}_m^{\mathrm{MSO}}(\mathfrak{A}, \mathbf{P}, \mathbf{a}; \mathfrak{B}, \mathbf{Q}, \mathbf{b})$, which is equivalent to the existence of a winning strategy of **II**. Characteristic formulae $\chi^m_{\mathfrak{A},\mathbf{P},\mathbf{a}}(\mathbf{X}, \mathbf{x})$ are defined inductively in the canonical way, to characterise $\mathfrak{A}, \mathbf{P}, \mathbf{a}$ up to $\simeq_m^{\mathrm{MSO}}$. One obtains the following MSO variant of the Ehrenfeucht-Fraïssé theorem.

Theorem 2.4.5 (MSO Ehrenfeucht-Fraïssé Theorem) *The following are equivalent for all $\mathfrak{A}, \mathbf{P}, \mathbf{a}; \mathfrak{B}, \mathbf{Q}, \mathbf{b}$ and m (in finite relational vocabulary τ):*

(i) $\mathfrak{A}, \mathbf{P}, \mathbf{a} \simeq_m^{\mathrm{MSO}} \mathfrak{B}, \mathbf{Q}, \mathbf{b}$.

(ii) **II** *wins* $\mathsf{G}_m^{\mathrm{MSO}}(\mathfrak{A}, \mathbf{P}, \mathbf{a}; \mathfrak{B}, \mathbf{Q}, \mathbf{b})$.

(iii) $\mathfrak{A}, \mathbf{P}, \mathbf{a} \equiv_m^{\mathrm{MSO}} \mathfrak{B}, \mathbf{Q}, \mathbf{b}$.

A composition lemma A useful feature of MSO is its compositionality w.r.t. to some simple composition operations on structures. We discuss (for later use) compositionality w.r.t. *ordered sums.* (FO inherits the same as a fragment of MSO.)

Let τ have a binary relation symbol $<$ (for a linear ordering). If $\mathfrak{A}_1 = (A_1, <^{\mathfrak{A}_1}, \dots)$ and $\mathfrak{A}_2 = (A_2, <^{\mathfrak{A}_2}, \dots)$ are τ-structures with disjoint universes, $A_1 \cap A_2 = \emptyset$, that are linearly ordered by $<$, then the ordered sum of $\mathfrak{A}_1$ and $\mathfrak{A}_2$ is the τ-structure

$$\begin{aligned} &\mathfrak{A}_1 \oplus \mathfrak{A}_2 := \big(A_1 \cup A_2, <^{\mathfrak{A}_1 \oplus \mathfrak{A}_2}, (R^{\mathfrak{A}_1} \cup R^{\mathfrak{A}_2})_{R \in \tau \setminus \{<\}}\big) \\ &\text{where } <^{\mathfrak{A}_1 \oplus \mathfrak{A}_2} = <^{\mathfrak{A}_1} \cup <^{\mathfrak{A}_2} \cup A_1 \times A_2. \end{aligned}$$

So $\mathfrak{A}_1 \oplus \mathfrak{A}_2$ is obtained from the $\mathfrak{A}_i$ by appending the order $(A_2, <^{\mathfrak{A}_2})$ to the order $(A_1, <^{\mathfrak{A}_1})$ and taking the disjoint union of the interpretations for all other relations. In case $\mathfrak{A}_1$ and $\mathfrak{A}_2$ are not disjoint, we pass to isomorphic copies that are.

Lemma 2.4.6 *For linearly ordered $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{B}_1, \mathfrak{B}_2$ with first- and second-order parameter tuples as indicated: if $\mathfrak{A}_1, \mathbf{P}_1, \mathbf{a}_1 \simeq_m^{\mathrm{MSO}} \mathfrak{B}_1, \mathbf{Q}_1, \mathbf{b}_1$ and $\mathfrak{A}_2, \mathbf{P}_2, \mathbf{a}_2 \simeq_m^{\mathrm{MSO}} \mathfrak{B}_2, \mathbf{Q}_2, \mathbf{b}_2$, then also*

$$(\mathfrak{A}_1, \mathbf{P}_1) \oplus (\mathfrak{A}_2, \mathbf{P}_2), \mathbf{a}_1\mathbf{a}_2 \ \simeq_m^{\mathrm{MSO}} \ (\mathfrak{B}_1, \mathbf{Q}_1) \oplus (\mathfrak{B}_2, \mathbf{Q}_2), \mathbf{b}_1\mathbf{b}_2.$$

Proof We argue with a composition of winning strategies and look at a single round in $\mathsf{G}_m^{\mathrm{MSO}}(\mathfrak{A}_1 \oplus \mathfrak{A}_2; \mathfrak{B}_1 \oplus \mathfrak{B}_2)$ (suppressing parameters for clarity). For instance, let **I** play a set move $P \subseteq A_1 \cup A_2$. **II** needs to come up with a matching $Q \subseteq B_1 \cup B_2$. Let $P_i := P \cap A_i$; then her strategies for $\mathsf{G}_m^{\mathrm{MSO}}(\mathfrak{A}_i; \mathfrak{B}_i)$ give **II** matching responses $Q_i \subseteq B_i$. Then $Q := Q_1 \cup Q_2$ is a good response for her in the combined game. For element moves, **II** refers to her strategy in that game $\mathsf{G}_m^{\mathrm{MSO}}(\mathfrak{A}_i; \mathfrak{B}_i)$ in which **I**'s challenge falls. □

Exercise 2.4.7 Show that similarly $\simeq_m^{\mathrm{MSO}}$ is compatible with disjoint unions of relational structures.

A particularly transparent application of this compositionality idea is given in the following, which was mentioned in connection with Proposition 1.2.10 and Example 1.2.11 above.

Example 2.4.8 EVEN $\subseteq$ $\mathsf{FIN}(\emptyset)$ is not MSO-definable. In fact, no quantifier rank n sentence in MSO($\emptyset$) distinguishes between naked sets of sizes $m, m' \geqslant 2^{n-1}$:

$$[m] \simeq_n^{\mathrm{MSO}} [m'] \quad \text{for } m, m' \geqslant 2^{n-1}. \qquad (\dagger)$$

Recall that $[m] = \{1, \ldots, m\}$ stands for the standard $\emptyset$-structure with m elements.

The argument is by induction on n and uses compositionality of strategies in the MSO game w.r.t. disjoint unions of structures with a single monadic predicate P, for $\tau = \{P\}$. As a special case of Exercise 2.4.7 above, we use that $\simeq_n^{\mathrm{MSO}}$ is compatible with disjoint unions of τ-structures of the form

$$\mathfrak{A}_1 = (A_1, A_1)\,,\ \mathfrak{A}_2 = (A_2, \emptyset) \quad \longmapsto \quad \mathfrak{A}_1 + \mathfrak{A}_2 = (A_1 \dot\cup A_2, A_1). \qquad (*)$$

The base case of ($\dagger$) for $n = 1$ is trivial. Let us consider the induction step from n to $n + 1$. By induction hypothesis we assume ($\dagger$) for n. Clearly then also the same equivalence holds for these (trivial) expansions of $[m]$ and $[m']$ to τ-structures:

$$([m], [m]) \simeq_n^{\mathrm{MSO}} ([m'], [m']) \quad \text{and} \quad ([m], \emptyset) \simeq_n^{\mathrm{MSO}} ([m'], \emptyset).$$

We now use compositionality w.r.t. the operation indicated in $(*)$, for structures $\mathfrak{A}_1 = ([m_1], [m_1])$, $\mathfrak{A}_2 = ([m_2], \emptyset)$, $\mathfrak{B}_1 = ([m_1'], [m_1'])$, $\mathfrak{B}_2 = ([m_2'], \emptyset)$. We know that $\mathfrak{A}_i \simeq_n^{\mathrm{MSO}} \mathfrak{B}_i$ for $i = 1, 2$ provided m_1, m_2, m_1', m_2' are such that $m_i = m_i'$ or $m_i, m_i' \geqslant 2^{n-1}$ for $i = 1, 2$. Hence

$$\begin{array}{rcl} \mathfrak{A}_1 + \mathfrak{A}_2 & \simeq_n^{\mathrm{MSO}} & \mathfrak{B}_1 + \mathfrak{B}_2 \\ ([m_1 + m_2], [m_1]) & \simeq_n^{\mathrm{MSO}} & ([m_1' + m_2'], [m_1']) \end{array}$$

for all m_1, m_2, m_1', m_2' such that for $i = 1, 2$: $m_i = m_i'$ or $m_i, m_i' \geqslant 2^{n-1}$.

Consider then a move by player **I** in the first round of the game $\mathsf{G}_{n+1}^{\mathrm{MSO}}([m]; [m'])$ for $m, m' \geqslant 2^n$. W.l.o.g., let **I** play a set move $P \subseteq [m]$ such that $([m], P) \simeq ([m_1 + m_2], [m_1])$ for suitable m_1, m_2 with $0 < m_1 < m$. Since $m = m_1 + m_2 \geqslant 2^n$, and since also $m' \geqslant 2^n$, there are m_1', m_2' such that $m' = m_1' + m_2'$ and $m_i = m_i'$ or $m_i, m_i' \geqslant 2^{n-1}$ for $i = 1, 2$. If **II** responds by playing $P' := [m_1'] \subseteq [m']$ for such a choice of m_1', m_2', then $([m_1 + m_2], [m_1]) \simeq_n^{\mathrm{MSO}} ([m_1' + m_2'], [m_1'])$ guarantees that she can force a win in the remaining game.

2.5 Variation: k Variables, k Pebbles

2.5.1 The k-variable fragment and k-pebble game

Reminder: all vocabularies finite and purely relational.

We consider the number of distinct variable symbols required as a logical resource. This also corresponds to the maximal arity of the 'auxiliary' queries defined by subformulae. The finite-variable fragments of FO play an important role in FMT. They induce non-trivial notions of elementary equivalence, and are helpful in the analysis also of the much more expressive extensions by FO by fixpoint constructors that we encounter in Part II. In the following $k \geqslant 2$ is arbitrary but fixed.

Definition 2.5.1 (i) The k-variable fragment of first-order logic, $\mathrm{FO}^k \subseteq \mathrm{FO}$, consists of those FO-formulae in which only the variable symbols $x_1, \ldots, x_k$ are used (free or bound).

(ii) $\mathfrak{A}, \mathbf{a}$ and $\mathfrak{B}, \mathbf{b}$ (with $|\mathbf{a}| = |\mathbf{b}| \leqslant k$) are k-variable equivalent, $\mathfrak{A}, \mathbf{a} \equiv^k \mathfrak{B}, \mathbf{b}$, if for all $\varphi(\mathbf{x}) \in \mathrm{FO}^k$ we have $\mathfrak{A} \models \varphi[\mathbf{a}] \Leftrightarrow \mathfrak{B} \models \varphi[\mathbf{b}]$.
Similarly, for $m \in \mathbb{N}$, $\mathfrak{A}, \mathbf{a} \equiv^k_m \mathfrak{B}, \mathbf{b}$ if they agree on all $\varphi(\mathbf{x}) \in \mathrm{FO}^k$ with $\mathrm{qr}(\varphi) \leqslant m$.

We often use variable symbols $x, y, \ldots$ (but only k distinct ones) instead of the official variables $x_1, \ldots, x_k$.

As atomic formulae in FO^k also cannot use more than k distinct variables, we only want to consider $\mathrm{FO}^k(\tau)$ in connection with relational vocabularies τ whose relation symbols have arities up to k at most. For similar reasons, we only consider parameter tuples (assignments to potentially free variables) of length k.

Example 2.5.2 (i) For $k = 3$, the class ORD is definable in $\mathrm{FO}^k(\{<\})$.

(ii) Over $\mathfrak{A} = (A, <^{\mathfrak{A}}) \in \mathrm{ORD}$, there are $\mathrm{FO}^2(\{<\})$-formulae φ_n defining the subset consisting of the first n elements w.r.t. $<$, for $n \geqslant 1$. Inductively, $\varphi_1(x_1) := \forall x_2 \neg x_2 < x_1$; $\varphi_{n+1}(x_1) := \forall x_2 \big(x_2 < x_1 \rightarrow \varphi_n(x_2)\big)$.[4]

It follows that $\equiv^k$ is not of finite index. Unlike $\equiv$, $\equiv^k$ does in general not trivialise to $\simeq$ over finite structures. It will follow from game considerations below, for instance, that $\mathrm{FO}^k[\emptyset]$ cannot distinguish between naked sets of different sizes $n \geqslant k$.

Exercise 2.5.3 Show that for every $k \geqslant 2$, $\equiv^k$ coincides with $\simeq$ on linearly ordered finite graphs.

The k-pebble game The k-pebble game is obtained as a simple variation of the FO pebble game. There are k pairs of pebbles numbered $1, \ldots, k$.

Positions in the k-pebble game over $\mathfrak{A}$ and $\mathfrak{B}$ are positions $(\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b})$ with $\mathbf{a} \in A^k$, $\mathbf{b} \in B^k$.[5]

A single round consists of a challenge/response exchange which now takes the following form. In position $(\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b})$, **I** chooses one pebble in one of the structures and relocates it on any element of that structure (e.g., pebble i in $\mathfrak{A}$ is moved to element $a \in A$); **II** has to respond by moving the corresponding pebble in the opposite structure (in the example, moving pebble i in $\mathfrak{B}$ to some element $b \in B$).

Writing $\mathbf{a}\frac{a}{i}$ for the result of replacing the i-th component of $\mathbf{a}$ by a, a round played with pebble i thus leads from a position $(\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b})$ to a position $(\mathfrak{A}, \mathbf{a}\frac{a}{i}; \mathfrak{B}, \mathbf{b}\frac{b}{i})$.

The constraints and winning conditions for the m-round k-pebble game are strictly analogous to those for the FO game, cf. Definition 2.1.3 and discussion there.

Definition 2.5.4 The m-round k-pebble game $\mathsf{G}^k_m(\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b})$ continues for m rounds starting from position $(\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b})$. **II** wins any play in which she maintains isomorphic pebble configurations through m rounds, and loses otherwise.

Definition 2.5.5 A *back-and-forth system* for G^k_m over $\mathfrak{A}$ and $\mathfrak{B}$ is a system $(I_i)_{0 \leqslant i \leqslant m}$ such that $\emptyset \neq I_i \subseteq \{(\mathbf{a} \mapsto \mathbf{b}) \in \mathrm{Part}(\mathfrak{A}, \mathfrak{B}) \colon \mathbf{a} \in A^k, \mathbf{b} \in B^k\}$, and, for $1 \leqslant i \leqslant m$, every $(\mathbf{a} \mapsto \mathbf{b}) \in I_n$ has back-and-forth extensions in I_{i-1}:

$$\begin{array}{ll} \textit{forth} & \forall j \in \{1, \ldots, k\}\ \forall a \in A\ \exists b \in B \colon (\mathbf{a}\frac{a}{j} \mapsto \mathbf{b}\frac{b}{j}) \in I_{i-1} \\ \textit{back} & \forall j \in \{1, \ldots, k\}\ \forall b \in B\ \exists a \in A \colon (\mathbf{a}\frac{a}{j} \mapsto \mathbf{b}\frac{b}{j}) \in I_{i-1}. \end{array}$$

[4] Here $\varphi_n(x_2)$ is obtained from $\varphi_n(x_1)$ be swapping variables x_1 and x_2 throughout.

[5] For simplicity we do not explicitly consider initial phases of the game during which not all pebbles would have to be placed; but this can be adapted where necessary.

$(I_i)_{0\leqslant i\leqslant m}$ is a back-and-forth system for $\mathsf{G}^k_m(\mathfrak{A},\mathbf{a};\mathfrak{B},\mathbf{b})$ if $(\mathbf{a}\mapsto\mathbf{b})\in I_m$. We write $(I_n)_{0\leqslant i\leqslant m}\colon \mathfrak{A},\mathbf{a}\simeq^k_m\mathfrak{B},\mathbf{b}$ in this situation, and say that $\mathfrak{A},\mathbf{a}$ and $\mathfrak{B},\mathbf{b}$ are k-pebble m-equivalent.

The following variation of the Ehrenfeucht-Fraïssé theorem is strictly analogous to the case of the m-round FO game, Theorem 2.1.7.

Theorem 2.5.6 *The following are equivalent for all* $\mathfrak{A},\mathbf{a};\mathfrak{B},\mathbf{b}$ *and* m*:*

(i) $\mathfrak{A},\mathbf{a}\simeq^k_m\mathfrak{B},\mathbf{b}$.

(ii) **II** *wins* $\mathsf{G}^k_m(\mathfrak{A},\mathbf{a};\mathfrak{B},\mathbf{b})$.

(iii) $\mathfrak{A},\mathbf{a}\equiv^k_m\mathfrak{B},\mathbf{b}$.

Exercise 2.5.7 Prove the above in analogy to the remarks in connection with Theorem 2.1.7.

Just as for the FO-game, one uses characteristic formulae $\chi^m_{\mathfrak{A},\mathbf{a}}\in\mathrm{FO}^k$, $\mathrm{qr}(\chi^m_{\mathfrak{A},\mathbf{a}})=m$, that characterise the $\simeq^k_m$-class of $\mathfrak{A},\mathbf{a}$. (Their shape is analogous to the $\chi^m_{\mathfrak{A},\mathbf{a}}$ for the FO game, with the obvious adaptation that the back-and-forth conditions now are for the moves in the k-pebble game.)

For $m=0$, $\chi^m_{\mathfrak{A},\mathbf{a}}$ is the conjunction over all atomic and negated atomic FO^k-formulae that are true of $\mathbf{a}$ in $\mathfrak{A}$. Inductively,

$$\chi^{m+1}_{\mathfrak{A},\mathbf{a}}(\mathbf{x}) := \chi^m_{\mathfrak{A},\mathbf{a}}\wedge\underbrace{\textstyle\bigwedge_{1\leqslant j\leqslant k}\bigwedge\big\{\exists x_j\chi^m_{\mathfrak{A},\mathbf{a}\frac{a}{j}}(\mathbf{x})\colon a\in A\big\}}_{\text{forth}}\wedge\underbrace{\textstyle\bigwedge_{1\leqslant j\leqslant k}\forall x_j\bigvee\big\{\chi^m_{\mathfrak{A},\mathbf{a}\frac{a}{j}}(\mathbf{x})\colon a\in A\big\}}_{\text{back}}.$$

Exercise 2.5.8 Show that **II** wins $\mathsf{G}^k_m(\mathfrak{A},\mathbf{a};\mathfrak{B},\mathbf{b})$ iff $\mathfrak{B}\models\chi^m_{\mathfrak{A},\mathbf{a}}[\mathbf{b}]$.

2.5.2 The unbounded k-pebble game and k-variable types

Definition 2.5.9 The infinite or unbounded k-pebble game $\mathsf{G}^k_\infty(\mathfrak{A},\mathbf{a};\mathfrak{B},\mathbf{b})$ starts from $(\mathfrak{A},\mathbf{a};\mathfrak{B},\mathbf{b})$ and consists of an unending succession of rounds in which isomorphic pebble configurations are maintained or ends with a loss for **II** when local isomorphism is violated. Correspondingly, **II** wins the game $\mathsf{G}^k_\infty(\mathfrak{A},\mathbf{a};\mathfrak{B},\mathbf{b})$ if she has a strategy to maintain isomorphic pebble configurations indefinitely in any play starting from $(\mathfrak{A},\mathbf{a};\mathfrak{B},\mathbf{b})$.

Note that the game graph for G^k_∞ over any two fixed finite $\mathfrak{A}$ and $\mathfrak{B}$ is finite (there are only $|A|^k\cdot|B|^k$ distinct positions. The analysis of the game is therefore essentially finite, since any sufficiently long play must eventually repeat some configuration.

Exercise 2.5.10 Try to argue game theoretically that if **I** has a winning strategy for $\mathsf{G}^k_\infty(\mathfrak{A},\mathbf{a};\mathfrak{B},\mathbf{b})$, then he has a strategy to force a win within $|A|^k\cdot|B|^k$ many rounds.

Back-and-forth systems for the infinite game

Definition 2.5.11 A *back-and-forth system* for G^k_∞ over $\mathfrak{A}$ and $\mathfrak{B}$ is a single set $I\subseteq\{(\mathbf{a}\mapsto\mathbf{b})\in\mathrm{Part}(\mathfrak{A},\mathfrak{B})\colon\mathbf{a}\in A^k,\mathbf{b}\in B^k\}$, such that every $(\mathbf{a}\mapsto\mathbf{b})\in I$ has back-and-forth extensions in I. I is a back-and-forth system for $\mathsf{G}^k_\infty(\mathfrak{A},\mathbf{a};\mathfrak{B},\mathbf{b})$ if $(\mathbf{a}\mapsto\mathbf{b})\in I$. We write $I\colon\mathfrak{A},\mathbf{a}\simeq^k_\infty\mathfrak{B},\mathbf{b}$ in this situation, and say that $\mathfrak{A},\mathbf{a}$ and $\mathfrak{B},\mathbf{b}$ are k-pebble equivalent.

Remark (for those familiar with these classical notions): $\simeq^k_\infty$ is the k-variable variant of the classical notion of partial isomorphy $\simeq_{\text{part}}$. [However, over finite structures, the distinction between notions of finite isomorphy and partial isomorphy is blurred.]

Analysis of k-variable types

Definition 2.5.12 For $\mathfrak{A}$ and $\mathbf{a} \in A^k$ define

(i) The *k-variable type* $\mathrm{tp}^k(\mathfrak{A}, \mathbf{a}) := \{\varphi(\mathbf{x}) \in \mathrm{FO}^k : \mathfrak{A} \models \varphi[\mathbf{a}]\}$.

(ii) The *rank m k-variable type* $\mathrm{tp}^k_m(\mathfrak{A}, \mathbf{a}) := \mathrm{tp}^k(\mathfrak{A}, \mathbf{a}) \cap \{\varphi \in \mathrm{FO}^k : \mathrm{qr}(\varphi) \leqslant m\}$.

By the Ehrenfeucht-Fraïssé theorem above, the rank m k-variable types $\mathrm{tp}^k_m(\mathfrak{A}, \mathbf{a})$ exactly specify the $\simeq^k_m$-equivalence class of $\mathfrak{A}, \mathbf{a}$. It is therefore also determined by the characteristic formula $\chi^m_{\mathfrak{A},\mathbf{a}}$, which is itself a member of this type.

Inductive refinement Consider an individual $\mathfrak{A} \in \mathsf{FIN}(\tau)$. An inductive refinement generates $\simeq^k_i$ and, as their limit, $\simeq^k_\infty$, as equivalence relations on A^k. Let $\sim_0$ be the equivalence relation corresponding to qfr-rank 0 equivalence:

$$\mathbf{a} \sim_0 \mathbf{a}' \quad \text{iff} \quad \mathfrak{A}, \mathbf{a} \simeq^k_0 \mathfrak{A}, \mathbf{a}' \quad \text{iff} \quad \mathrm{tp}^k_0(\mathfrak{A}, \mathbf{a}) = \mathrm{tp}^k_0(\mathfrak{A}, \mathbf{a}') \quad \text{iff} \quad \mathfrak{A} \restriction \mathbf{a}, \mathbf{a} \simeq \mathfrak{A} \restriction \mathbf{a}', \mathbf{a}'.$$

Suppose the equivalence relation $\sim_i$ on A^k is given such that

$$\mathbf{a} \sim_i \mathbf{a}' \quad \text{iff} \quad \mathfrak{A}, \mathbf{a} \simeq^k_i \mathfrak{A}, \mathbf{a}' \quad \text{iff} \quad \mathrm{tp}^k_i(\mathfrak{A}, \mathbf{a}) = \mathrm{tp}^k_i(\mathfrak{A}, \mathbf{a}').$$

For any $1 \leqslant j \leqslant k$, any $\sim_i$ equivalence class $\alpha \in A^k/{\sim_i}$, and $\mathbf{a} \in A^k$ define

$$\iota_{j,\alpha}(\mathbf{a}) := \begin{cases} 1 & \text{if } \exists a \in A(\mathbf{a}\tfrac{a}{j} \in \alpha) \\ 0 & \text{else.} \end{cases}$$

Now put

$$\mathbf{a} \sim_{i+1} \mathbf{a}' \quad \text{iff} \quad \mathbf{a} \sim_i \mathbf{a}' \text{ and } \forall j \forall \alpha \colon \iota_{j,\alpha}(\mathbf{a}) = \iota_{j,\alpha}(\mathbf{a}').$$

I.e., for $\mathbf{a} \simeq^k_i \mathbf{a}'$ we put $\mathbf{a} \sim_{i+1} \mathbf{a}'$ if, and only if, $\mathbf{a} \mapsto \mathbf{a}'$ has back-and-forth extensions which maintain $\simeq^k_i$ equivalence. It follows that, as desired, $\sim_{i+1}$ coincides with $\simeq^k_{i+1}$ (and $\equiv^k_{i+1}$) on $\mathfrak{A}$.

Clearly the sequence $(\sim_i)_{i \geqslant 0}$ is a monotone sequence of successively refined equivalence relation on the finite set A^k. Hence for some $r \leqslant |A|^k$ we must have $\sim_r = \sim_{r+1} = \sim_{r+s}$ for all $s \in \mathbb{N}$. The minimal such r is called the k-rank of $\mathfrak{A}$, k-rank$(\mathfrak{A})$.

Lemma 2.5.13 (i) *For all $i \in \mathbb{N}$: $\mathbf{a} \sim_i \mathbf{a}'$ iff $\mathfrak{A}, \mathbf{a} \simeq^k_i \mathfrak{A}, \mathbf{a}'$.*

(ii) *For $r = k$-rank$(\mathfrak{A})$: $\mathbf{a} \sim_r \mathbf{a}'$ iff $\mathfrak{A}, \mathbf{a} \simeq^k_\infty \mathfrak{A}, \mathbf{a}'$.*

Proof (i) follows directly from the definition of the $\sim_i$, by induction on i.

For (ii) consider $I := \{\mathbf{a} \mapsto \mathbf{a}' : \mathbf{a} \sim_r \mathbf{a}'\}$. We claim that I has back-and-forth extensions, so that $I \colon \mathfrak{A}, \mathbf{a} \simeq^k_\infty \mathfrak{A}, \mathbf{a}'$ for any $(\mathbf{a} \mapsto \mathbf{a}') \in I$. Let $(\mathbf{a} \mapsto \mathbf{a}') \in I$, $0 \leqslant j \leqslant k$ and $a \in A$. As $\sim_r = \sim_{r+1}$, $\mathbf{a} \sim_{r+1} \mathbf{a}'$. Hence $\mathfrak{A}, \mathbf{a} \simeq^k_{r+1} \mathfrak{A}, \mathbf{a}'$, which guarantees the existence of some $a' \in A$ for which $\mathfrak{A}, \mathbf{a}\frac{a}{j} \simeq^k_r \mathfrak{A}, \mathbf{a}'\frac{a'}{j}$. It follows that $(\mathbf{a}\frac{a}{j} \mapsto \mathbf{a}'\frac{a'}{j}) \in I$ is as desired. $\Box$

We conclude that for every $\mathfrak{A}$ and $\mathbf{a} \in A^k$ there is also a single FO^k-formula that characterises the $\simeq^k_\infty$ class of $\mathfrak{A}, \mathbf{a}$. Let $r = k$-rank$(\mathfrak{A})$ and put

$$\begin{aligned} \chi_{\mathfrak{A}} \;:=\; & \textstyle\bigwedge_{\mathbf{a} \in A^k} \exists \mathbf{x} \chi^r_{\mathfrak{A},\mathbf{a}} \;\wedge\; \forall \mathbf{x} \bigvee_{\mathbf{a} \in A^k} \chi^r_{\mathfrak{A},\mathbf{a}} \\ & \wedge \textstyle\bigwedge_{1 \leqslant j \leqslant k} \bigwedge_{\mathbf{a} \in A^k} \forall \mathbf{x} \Big[\chi^r_{\mathfrak{A},\mathbf{a}} \rightarrow \big(\bigwedge_{a \in A} \exists x_j \chi^r_{\mathfrak{A},\mathbf{a}\frac{a}{j}} \;\wedge\; \forall x_j \bigvee_{a \in A} \chi^r_{\mathfrak{A},\mathbf{a}\frac{a}{j}} \big) \Big]. \end{aligned}$$

The first two conjuncts say that precisely the rank r k-types of $\mathfrak{A}$ are realised; the third conjunct implies that $\sim_{r+1}=\sim_r$, whence the rank r types determine the $\simeq^k_\infty$ types.

For $\mathbf{a} \in A^k$ put

$$\chi_{\mathfrak{A},\mathbf{a}}(\mathbf{x}) := \chi_{\mathfrak{A}} \wedge \chi^r_{\mathfrak{A},\mathbf{a}}(\mathbf{x}).$$

Note that $\mathrm{qr}(\chi_{\mathfrak{A},\mathbf{a}}) = r + k + 1$.

Suppose $\mathfrak{B} \models \chi_{\mathfrak{A}}$. Then $\mathfrak{B}$ has exactly the same rank r types as $\mathfrak{A}$ and k-rank$(\mathfrak{B}) = r$. Moreover, the following system has back-and-forth extensions:

$$I := \big\{\mathbf{a} \mapsto \mathbf{b} \colon \mathrm{tp}^k_r(\mathfrak{B}, \mathbf{b}) = \mathrm{tp}^k_r(\mathfrak{A}, \mathbf{a})\big\}.$$

If $\mathfrak{B} \models \chi_{\mathfrak{A},\mathbf{a}}[\mathbf{b}]$, then also $\mathrm{tp}^k_r(\mathfrak{B}, \mathbf{b}) = \mathrm{tp}^k_r(\mathfrak{A}, \mathbf{a})$ and $\mathfrak{A}, \mathbf{a} \simeq^k_\infty \mathfrak{B}, \mathbf{b}$ via I.

Theorem 2.5.14 *Let $\mathfrak{A} \in \mathsf{FIN}(\tau)$, $r := k\text{-rank}(\mathfrak{A})$, $\mathbf{a} \in A^k$. Then the following are equivalent for any $\mathfrak{B} \in \mathsf{FIN}(\tau)$ and $\mathbf{b} \in B^k$:*

(i) $\mathfrak{A}, \mathbf{a} \simeq^k_\infty \mathfrak{B}, \mathbf{b}$.

(ii) $\mathfrak{A}, \mathbf{a} \simeq^k_{r+k+1} \mathfrak{B}, \mathbf{b}$.

(iii) $\mathfrak{A}, \mathbf{a} \equiv^k_{r+k+1} \mathfrak{B}, \mathbf{b}$.

(iv) $\mathfrak{A}, \mathbf{a} \equiv^k \mathfrak{B}, \mathbf{b}$.

Proof The equivalence between (ii) and (iii) is from Theorem 2.5.6. As $\simeq^k_\infty$ equivalence implies $\simeq^k_i$ equivalence for all i, we clearly have (i) $\Rightarrow$ (ii), and Theorem 2.5.6 also gives (i) $\Rightarrow$ (iv). (iv) $\Rightarrow$ (iii) is trivial. As (iii) implies $\mathfrak{B} \models \chi_{\mathfrak{A},\mathbf{a}}[\mathbf{b}]$, (iii) $\Rightarrow$ (i) follows from the consideration above. □

A global pre-ordering w.r.t. k-variable types We upgrade the inductive refinement that separated out the different $\simeq^k_\infty$-types over an individual finite structure $\mathfrak{A}$ so that it provides a linear ordering of these types, i.e., a linear ordering of $A^k/{\simeq^k_\infty}$. This is achieved in stages that go along with the inductive refinement of the equivalence relations $(\sim_i)_{i\geqslant 0}$ on A^k we saw above. We now instead generate transitive, reflexive and total pre-ordering relations $\preccurlyeq_i$ on A^k such that

$$\mathbf{a} \sim_i \mathbf{a}' \quad\Leftrightarrow\quad \big(\mathbf{a} \preccurlyeq_i \mathbf{a}' \text{ and } \mathbf{a}' \preccurlyeq_i \mathbf{a}\big). \qquad (*)$$

It follows that $A^k/{\sim_i}$ is linearly ordered (in the sense of $\leqslant$) by $\preccurlyeq_i$.

At level $i = 0$ we use some arbitrary but fixed linear ordering of the (finitely many) qfr-free k-variable types. Inductively assume that $\preccurlyeq_i$ satisfies $(*)$, and hence induces a linear ordering on $A^k/{\sim_i}$. Consider the values of the boolean functions $\iota_{j,\alpha}$ on A^k, and recall that for $\mathbf{a} \sim_i \mathbf{a}'$ we have $\mathbf{a} \sim_{i+1} \mathbf{a}'$ iff $\iota_{j,\alpha}(\mathbf{a}) = \iota_{j,\alpha}(\mathbf{a}')$ for all j, α.

Consider the boolean tuple listing the $\iota_{j,\alpha}$-values in order of increasing j, and within the same j, increasing w.r.t. α in the sense of $\preccurlyeq_i$. The set of all such tuples carries a natural lexicographic ordering, based on the first position where the two tuples differ (if not equal). We now put

$$\mathbf{a} \preccurlyeq_{i+1} \mathbf{a}' \quad\text{if}\quad \big(\mathbf{a} \preccurlyeq_i \mathbf{a}' \text{ and } (\iota_{j,\alpha}(\mathbf{a}))_{j,\alpha} \leqslant_{\mathrm{lex}} (\iota_{j,\alpha}(\mathbf{a}'))_{j,\alpha}\big).$$

Then $\preccurlyeq_{i+1}$ is again transitive, reflexive and total and provides a linear ordering of the $\sim_{i+1}$-classes according to $(*)$. Note that $\preccurlyeq_{i+1}$ is uniformly FO-definable in terms of $\preccurlyeq_i$ over $\mathfrak{A}$. Let $\preccurlyeq$ be the limit $\preccurlyeq^{\mathfrak{A}}=\preccurlyeq^{\mathfrak{A}}_r$ for $r = k$-rank$(\mathfrak{A})$. Then $\preccurlyeq^{\mathfrak{A}}$ is a linear ordering (in the sense of $\leqslant$) on $A^k/{\simeq^k_\infty}$.

Lemma 2.5.15 *There is a global relation $\preccurlyeq$ of arity $2k$, uniformly definable by an inductive iteration of a first-order definable operation, such that $\preccurlyeq^{\mathfrak{A}}$ is a pre-ordering on A^k which linearly orders w.r.t. k-variable types.*

The linear ordering induced by $\preccurlyeq$ on $A^k/{\simeq^k_\infty}$ is completely determined by the k-variable types in $\mathfrak{A}$, and hence only depends on the $\equiv^k$-class of $\mathfrak{A}$. The same is true of the information about the qfr-free formulae in each $\simeq^k_\infty$-type,

$$P_\theta := \{\alpha \in A^k/{\simeq^k_\infty} \colon \mathfrak{A} \models \theta[\mathbf{a}] \text{ for } \mathbf{a} \in \alpha\} \quad \text{for each qfr-free } \theta \in \mathrm{FO}^k(\tau)$$

and the incidence between $\simeq^k_\infty$-types w.r.t. the relations describing moves of the j-th pebble in the game

$$E_j := \big\{(\alpha, \alpha') \colon \exists a (\mathbf{a}\frac{a}{j} \in \alpha') \text{ for } \mathbf{a} \in \alpha\big\} \quad \text{for } j = 1, \dots, k.$$

For a distinguished tuple $\mathbf{a} \in A^k$, we may also identify its $\simeq^k_\infty$-class $[\mathbf{a}]_{\simeq^k_\infty}$ as a distinguished element in the quotient $A^k/{\simeq^k_\infty}$.

Definition 2.5.16 The k-variable invariant of $\mathfrak{A}, \mathbf{a}$ is defined to be the linearly ordered quotient structure $\mathfrak{I}^k(\mathfrak{A}, \mathbf{a}) := \big(A^k/{\simeq^k_\infty}, \leqslant, (P_\theta), (E_j), [\mathbf{a}]_{\simeq^k_\infty}\big)$.

Proposition 2.5.17 *For $\mathfrak{A}, \mathbf{a}$ and $\mathfrak{B}, \mathbf{b}$: $\mathfrak{A}, \mathbf{a} \simeq^k_\infty \mathfrak{B}, \mathbf{b}$ iff $\mathfrak{I}^k(\mathfrak{A}, \mathbf{a}) \simeq \mathfrak{I}^k(\mathfrak{B}, \mathbf{b})$.*

Proof "$\Leftarrow$" is clear from the definition of $\mathfrak{I}^k$. "$\Rightarrow$" follows from the fact that the actual k-variable types realised in $\mathfrak{A}$, and in particular the k-variable type of $\mathbf{a}$ in $\mathfrak{A}$ can be identified from just $\mathfrak{I}^k(\mathfrak{A}, \mathbf{a})$. For this one determines $\mathrm{tp}^k_m(\mathfrak{A}, \mathbf{a}')$ for $\mathbf{a}' \in \alpha \in \mathfrak{I}^k$, by induction on m. $\square$

As the invariants are polynomial time computable, and isomorphism between finite linearly ordered structures is trivially decidable in polynomial time (also cf.~Part II), we get the following.

Corollary 2.5.18 *k-variable equivalence can be decided by a polynomial time algorithm.*

Chapter 3

Zero-One Laws

3.1 Asymptotic Probabilities

For fixed finite relational vocabulary τ and $n \geqslant 1$, the set of all finite τ-structures of size n is finite up to $\simeq$. If we let $\mathsf{FIN}_n(\tau)$ stand for the set of τ-structures over the standard size n universe $[n] := \{1, \dots, n\}$, then $\mathsf{FIN}_n(\tau)$ is a finite set for each n. Given a sentence $\varphi \in \mathcal{L}(\tau)$, we regard

$$\mu_n(\varphi) := \frac{|\{\mathfrak{A} \in \mathsf{FIN}_n(\tau) \colon \mathfrak{A} \models \varphi\}|}{|\mathsf{FIN}_n(\tau)|} = \frac{|\mathsf{MOD}(\varphi) \cap \mathsf{FIN}_n(\tau)|}{|\mathsf{FIN}_n(\tau)|}$$

as the probability that a randomly chosen $\mathfrak{A} \in \mathsf{FIN}_n(\tau)$ satisfies φ (the boolean query Q defined by φ). If the limit exists,

$$\mu(\varphi) := \lim_{n \to \infty} \mu_n(\varphi)$$

is called the *asymptotic probability* of φ (the query defined by φ). In particular, we say that

$$\begin{array}{ll} \varphi \textit{ is almost surely true} & \text{if } \mu(\varphi) = 1, \\ \varphi \textit{ is almost surely false} & \text{if } \mu(\varphi) = 0. \end{array}$$

An equivalent and intuitive characterisation of this probability space based on $\mathsf{FIN}_n(\tau)$ with the uniform distribution is as follows. For every relational τ-atom $R\mathbf{x}$ and every assignment $\mathbf{a}$ for $\mathbf{x}$ over $[n]$, toss a fair coin to determine whether $\mathbf{a} \in R^{\mathfrak{A}}$ or not, with probability 1/2. Treating all instantiated atoms in this way, independent of each other, we arrive at a τ-structure $\mathfrak{A} \in \mathsf{FIN}_n(\tau)$ as the outcome of a random experiment. For any query Q, $\mu_n(Q)$ then is the probability that this random experiment yields $\mathfrak{A} \in Q$.

For technical reasons we also consider probabilities for formulae with free variables. For $\varphi(\mathbf{x})$ and fixed assignment $\mathbf{a}$ to its free variables $\mathbf{x}$ over $[n]$, we let

$$\mu_n(\varphi[\mathbf{a}]) := \frac{\{\mathfrak{A} \in \mathsf{FIN}_n(\tau) \colon \mathfrak{A} \models \varphi[\mathbf{a}]\}}{|\mathsf{FIN}_n(\tau)|}$$

be the probability that $\mathfrak{A} \in \mathsf{FIN}_n$ makes $\varphi[\mathbf{a}]$ true.

Definition 3.1.1 (i) A logic $\mathcal{L}$ *has asymptotic probabilities* if all $\mathcal{L}$-sentences have asymptotic probabilities.

(ii) A logic satisfies a *zero-one law* if it has asymptotic probabilities, and these take values in $\{0, 1\}$ only. I.e., if every sentence of $\mathcal{L}(\tau)$ is either almost surely true or almost surely false on $\mathsf{FIN}(\tau)$.

Variation Often the reference class of structures is not $\mathsf{FIN}(\tau)$ but a proper subclass. We only look at the special case of $\tau = \{E\}$ and the class GRAPH of all finite undirected graphs. Generally, for $\mathrm{K} \subseteq \mathsf{FIN}(\tau)$, the above notions are adapted accordingly, by letting

$$\mu_n^{\mathrm{K}}(\varphi) = \frac{|\mathsf{MOD}(\varphi) \cap \mathrm{K}_n|}{|\mathrm{K}_n|},$$

where $\mathrm{K}_n = \mathrm{K} \cap \mathsf{FIN}_n(\tau)$.

Example 3.1.2 (i) The boolean query EVEN does not have an asymptotic probability, as $\mu_n(\mathrm{EVEN}) = n \bmod 2$.

(ii) We shall see below that finite undirected graphs almost surely have diameter 2, whence in particular they are almost surely connected.

Asymptotic probabilities can also be useful in connection with FINSAT issues. Note that if $\mu_n(\varphi) > 0$ the φ must have models of size n; if $\mu(\varphi) > 0$, then φ must have arbitrarily large finite models. Probabilistic arguments can thus be used to prove the existence of finite models.

3.2 Extension Axioms and The Almost Sure Theory

Fix finite relational τ. A *basic k-type* is a maximal consistent set of atomic and negated τ-formulae in variables $x_1, \ldots, x_k$ comprising in particular the formulae $\neg x_i = x_j$ for $1 \leqslant i < j \leqslant k$.

In the following we write $\theta(\mathbf{x})$ where $\mathbf{x} = (x_1, \ldots, x_k)$ for such basic types. As $\theta(\mathbf{x})$ is finite, we may identify it with the single qfr-free formula $\bigwedge \theta$. A non-degenerate tuple $\mathbf{a}$ in $\mathfrak{A}$ realises the type θ if $\theta = \{\varphi(\mathbf{x})\colon \varphi \text{ (negated) atomic}, \mathfrak{A} \models \varphi[\mathbf{a}]\}$. i.e. if $\mathfrak{A} \models \theta[\mathbf{a}]$.

Exercise 3.2.1 Let $\theta(\mathbf{x}), \theta'(\mathbf{x})$ be basic k-types, $\mathbf{a}, \mathbf{a}'$ disjoint non-degenerate assignments over $[n]$ to $\mathbf{x}$. Show that

(i) $\mu_n(\theta[\mathbf{a}]) = \mu_n(\theta[\mathbf{a}'])$.

(ii) $\mu_n(\theta[\mathbf{a}] \wedge \theta[\mathbf{a}']) = \mu_n(\theta[\mathbf{a}])\mu_n(\theta[\mathbf{a}'])$ (independence).

(iii) $\mu_n(\theta[\mathbf{a}]) = \mu_n(\theta'[\mathbf{a}])$, i.e., any two basic types have the same probability to be realised by a particular tuple.

For basic k-type $\theta(\mathbf{x})$ and $(k+1)$-type $\theta'(\mathbf{x}, x_{k+1})$, we say that θ' is an extension of θ if $\theta \subseteq \theta'$ (if we regard them as sets of formulae), or equivalently, if $\theta' \models \theta$ (if we regard them as formulae). We call such (θ, θ') an *extension pair.*

Definition 3.2.2 The *extension axiom* for the extension pair (θ, θ') is the $\mathrm{FO}(\tau)$-sentence

$$\mathrm{Ext}_{\theta,\theta'} := \forall \mathbf{x}\big(\theta(\mathbf{x}) \rightarrow \exists x_{k+1} \theta'(\mathbf{x}, x_{k+1})\big).$$

We explicitly include the case of extension pairs $(\emptyset, \theta(x_1))$, for which $\mathrm{Ext}_{\emptyset,\theta} \equiv \exists x_1 \theta(x_1)$. Let $\mathrm{EXT}(\tau) \subseteq \mathrm{FO}(\tau)$ be the set of all extension axioms.

Definition 3.2.3 The *almost sure theory* for τ is the set of $\mathrm{FO}(\tau)$-sentences

$$\mathrm{AST}(\tau) := \{\varphi \in \mathrm{FO}(\tau) : \mu(\varphi) = 1\}.$$

In the remainder of this section we show the following result.

Theorem 3.2.4 (Fagin) *For any finite relational vocabulary τ:*

(i) $\mathrm{EXT}(\tau) \subseteq \mathrm{AST}(\tau)$, *i.e., every extension axiom is almost surely true. It follows that every finite collection of extension axioms has a finite model, and that (by compactness)* $\mathrm{EXT}(\tau)$ *is satisfiable (in an infinite model).*

(ii) $\mathrm{EXT}(\tau)$ *has, up to isomorphism, precisely one countably infinite model, the so-called random τ-structure $\mathfrak{R}_\tau$.*

(iii) $\mathrm{AST}(\tau)$ *is the* FO*-theory of the random τ-structure:*

$$\mathrm{AST}(\tau) = \{\varphi \in \mathrm{FO}(\tau) \colon \mathfrak{R}_\tau \models \varphi\}.$$

(iv) $\mathrm{FO}(\tau)$ *satisfies a zero-one law.*

We first establish (i).

Lemma 3.2.5 *For every extension pair (θ, θ'), $\mathrm{Ext}_{\theta,\theta'}$ is almost surely true.*

Proof We show that $\mu_n(\mathrm{Ext}_{\theta,\theta'}) \rightarrow 1$. Let $\theta = \theta(\mathbf{x})$ where $\mathbf{x} = (x_1, \ldots, x_k)$, and, writing y instead of x_{k+1} for clarity, $\theta' = \theta'(\mathbf{x}, y)$. Note that, as a basic type, $\theta'(\mathbf{x}, y)$ stipulates that $y \neq x$ for x in $\mathbf{x}$.

Consider a fixed non-degenerate assignment $\mathbf{a}$ in $[n]$ to the variables $\mathbf{x}$. For every $b \in [n] \backslash \{\mathbf{a}\}$, every basic type for $(\mathbf{a}, b)$ is equally likely in $\mathfrak{A} \in \mathsf{FIN}_n(\tau)$ (cf. Exercise 3.2.1 (iii)). Let $\delta > 0$ be this probability that the basic type of $(\mathbf{a}, b)$ is θ' (or any other specified basic type).

$$\delta := \mu_n(\theta'[\mathbf{a}, b]) > 0.$$

Then

$$\mu_n\big((\theta \wedge \neg\theta')[\mathbf{a}, b]\big) < 1 - \delta,$$

and as any $b \in [n] \setminus \{\mathbf{a}\}$ has the same probability to realise θ' (by an argument similar to Exercise 3.2.1 (ii)), for fixed $\mathbf{a}$

$$\mu_n\big(\theta \wedge \neg\exists y\theta')[\mathbf{a}]\big) < (1 - \delta)^{n-k}.$$

Clearly, each non-degenerate $\mathbf{a}$ fails $\theta \rightarrow \exists y\theta'$ with the same probability (e.g., by an isomorphism argument). Therefore

$$\mu_n(\neg\mathrm{Ext}_{\theta,\theta'}) = \mu_n\big(\exists\mathbf{x}(\theta \wedge \neg\exists y\theta')\big) < n^k(1 - \delta)^{n-k} \quad \stackrel{n\to\infty}{\longrightarrow} \; 0.$$

It follows that $\mu(\neg\mathrm{Ext}_{\theta,\theta'}) = 0$ and $\mu(\mathrm{Ext}_{\theta,\theta'}) = 1$. $\square$

Note that this implies in particular that any finite collection of extension axioms is satisfiable in finite models. By compactness, $\mathrm{EXT}(\tau)$ is satisfiable, but has no finite models (why?). By Löwenheim-Skolem, it must have countably infinite models.

Corollary 3.2.6 $\mathrm{EXT}(\tau)$ *is satisfiable in a countably infinite model.*

The following lemma yields (ii) of the theorem. The proof is a familiar back-and-forth argument from classical model theory (partially isomorphic countable structures are isomorphic).

Lemma 3.2.7 *Let $\mathfrak{A}, \mathfrak{B} \models \mathrm{EXT}(\tau)$, $\mathfrak{A}$ and $\mathfrak{B}$ both countable. Then $\mathfrak{A} \simeq \mathfrak{B}$.*

Proof (back-and-forth argument) Suppose Let $\mathfrak{A}, \mathfrak{B} \models \mathrm{EXT}(\tau)$. Then the following system of all finite partial isomorphisms between $\mathfrak{A}$ and $\mathfrak{B}$ has back-and-forth extensions:

$$I := \{ p = (\mathbf{a} \mapsto \mathbf{b}) \colon p \in \mathrm{Part}(\mathfrak{A}, \mathfrak{B}), \mathrm{def}(p) \text{ finite } \}.$$

Consider $(\mathbf{a} \mapsto \mathbf{b}) \in I$, and, for instance, $a \in A$. W.l.o.g. $\mathbf{a}$ is non-degenerate and a disjoint from $\mathbf{a}$. Let $\theta(\mathbf{x})$ and $\theta(\mathbf{x}, y)$ be the basic types of $\mathbf{a}$ and $\mathbf{a}a$ in $\mathfrak{A}$, respectively. Then (θ, θ') is an extension pair. As $\mathbf{a} \mapsto \mathbf{b}$ is an isomorphism between $\mathfrak{A} \restriction \mathbf{a}$ and $\mathfrak{B} \restriction \mathbf{b}$, $\mathbf{b}$ satisfies $\theta(\mathbf{x})$ in $\mathfrak{B}$. As $\mathfrak{B} \models \mathrm{Ext}_{\theta,\theta'}$, we find $b \in B$ such that $\mathfrak{B} \models \theta'[\mathbf{b}, b]$. It follows that $(\mathbf{a}a \mapsto \mathbf{b}b) \in I$.

If $\mathfrak{A}$ and $\mathfrak{B}$ are both countable, let A and B be enumerated as $(a_i)_{i \in \mathbb{N}}$ and $(b_i)_{i \in \mathbb{N}}$. Define a sequence of finite partial isomorphisms $(p_n)_{n \in \mathbb{N}}$, where $p_n \in I$ for all n, $p_0 \subseteq p_1 \subseteq p_2 \subseteq \cdots$ an increasing chain and such that $\bigcup_n \mathrm{def}(p_n) = A$ and $\bigcup_n \mathrm{im}(p_n) = B$. The limit p of the sequence $(p_n)_{n \in \mathbb{N}}$ is an isomorphism, $p = \bigcup_n p_n \colon \mathfrak{A} \simeq \mathfrak{B}$ (check!).

The p_n are chosen inductively, starting from $p_0 = \emptyset$, where

(i) for even $n = 2m$, $p_{n+1} \supseteq p_n$ is a *forth*-extension of p_n with $a_m \in \mathrm{def}(p_{n+1})$.

(ii) for odd $n = 2m + 1$, $p_{n+1} \supseteq p_n$ is a *back*-extension of p_n with $b_m \in \mathrm{im}(p_{n+1})$.

$\Box$

Together with Corollary 3.2.6, we see that $\mathrm{EXT}(\tau)$ has a unique (up to $\simeq$) countably infinite model $\mathfrak{R}_\tau \models \mathrm{EXT}(\tau)$, called *the random τ-structure*. Moreover:

$$\mathrm{EXT}(\tau) \models \varphi \quad \Leftrightarrow \quad \mathfrak{R}_\tau \models \varphi. \qquad (*)$$

For "$\Leftarrow$" note that consistency of $\neg\varphi$ with $\mathrm{EXT}(\tau)$ would imply the existence of a (countable) model of $\mathrm{EXT}(\tau) \cup \{\neg\varphi\}$, whence, by the last lemma, $R_\tau \models \neg\varphi$.

$$\mathrm{EXT}(\tau) \models \varphi \quad \Rightarrow \quad \mu(\varphi) = 1 \quad (\varphi \in \mathrm{AST}) \qquad (**)$$

For this note that, by compactness, $\mathrm{EXT}(\tau) \models \varphi$ implies that φ is a consequence of finitely many extension axioms. As those are almost surely true, so is φ.

Proof (of Theorem 3.2.4) Parts (i) and (ii) have been dealt with in the preceding lemmas. For (iii) we argue that, by $(*)$ and $(**)$, also

$$\mathfrak{R}_\tau \models \varphi \quad \Leftrightarrow \quad \mu(\varphi) = 1 \quad (\varphi \in \mathrm{AST}).$$

"$\Rightarrow$" is immediate from $(*)$ and $(**)$. For "$\Leftarrow$" it suffices to consider "$\Rightarrow$" for $\neg\varphi$: if $\mathfrak{R}_\tau \not\models \varphi$, then $\mathfrak{R}_\tau \models \neg\varphi$, so $\mu(\neg\varphi) = 1$ and $\mu(\varphi) = 0$.

(iv) follows, since for all sentences φ either $\varphi \in \mathrm{AST}$ or $\neg\varphi \in \mathrm{AST}$. In the first case, $\mu(\varphi) = 1$, in the second case $\mu(\varphi) = 0$. $\Box$

3.3 The Random Graph

The whole approach outlined above carries through for so-called parametric classes of finite τ-structures instead of the whole of $\mathsf{FIN}(\tau)$ as the basic reference set. This is important, as one cannot directly apply the above results to classes like G := GRAPH. The naive attempt to identify those FO-sentences φ that are almost surely true in finite graphs by looking at $\mu(\varphi_0 \to \varphi)$, where $\varphi_0 \in \mathrm{FO}(\{E\})$ defines GRAPH, fails. In fact, $\mu(\varphi_0) = 0$ (why?) and therefore $\mu(\varphi_0 \to \varphi) = 1$ for any φ.

One needs to look at conditional probabilities instead. In effect this necessitates a repetition of the above arguments involving extension axioms for those basic types that are compatible with GRAPH (consistent with φ_0).

For the probabilities we now want the following, where in our case G := GRAPH and $\mathrm{G}_n := \mathrm{GRAPH} \cap \mathsf{FIN}_n(\{E\})$, and φ_0 says that E is irreflexive and symmetric.

$$\mu_n^{\mathrm{G}}(\varphi) = \frac{|\mathsf{MOD}(\varphi) \cap \mathrm{G}_n|}{|\mathrm{G}_n|} = \frac{|\mathsf{MOD}(\varphi_0 \wedge \varphi) \cap \mathsf{FIN}_n(\{E\})|}{|\mathsf{MOD}(\varphi_0) \cap \mathsf{FIN}_n(\{E\})|}.$$

Extension axioms for graphs

A basic type is compatible with φ_0 if it can be realised in graphs, which is the case if and only if it describes the isomorphism type of a finite (sub-)graph. A one-point extension of a finite (sub-)graph $\mathfrak{G}_0$ by one extra new vertex b is fully determined by stipulating to which vertices a of $\mathfrak{G}_0$ the new vertex b is linked by an edge. If $\mathbf{a}$ is a non-degenerate tuple listing all the vertices of $\mathfrak{G}_0$, we just need to partition $\mathbf{a}$ into disjoint tuples $\mathbf{a}^+, \mathbf{a}^-$, where $\mathbf{a}^+$ lists those $a \in \mathbf{a}$ for which $(a,b),(b,a) \in E$ and $\mathbf{a}^-$ lists those $a \in \mathbf{a}$ for which $(a,b),(b,a) \notin E$ in the extension. Consistency with internal edges within $\mathfrak{G}_0$ is trivial, as there is no dependency. Therefore, the following formalisation of graph extension axioms is sufficient.

$$\mathrm{Ext}_{n,m} := \forall x_1 \ldots \forall x_{n+m}\Big(\bigwedge_{1 \leqslant i < j \leqslant n+m} \neg x_i = x_j \quad \rightarrow \exists y\big(\bigwedge_{1 \leqslant i \leqslant n} Ex_iy \quad \wedge \bigwedge_{n+1 \leqslant i \leqslant n+m} \neg Ex_iy\big)\Big).$$

The theory corresponding to $\mathrm{EXT}(\tau)$ in the graph setting then is

$$\mathrm{EXT}^{\mathrm{G}} := \{\varphi_0\} \cup \big\{\mathrm{Ext}_{n,m} \colon n, m \in \mathbb{N}\big\},$$

while the analogue of the almost sure τ-theory now is the *almost sure theory of undirected graphs*

$$\mathrm{AST}^{\mathrm{G}} := \big\{\varphi \in \mathrm{FO}(\{E\}) \colon \mu^{\mathrm{G}}(\varphi) = 1\big\}.$$

Theorem 3.3.1

(i) $\mu^{\mathrm{G}}(\mathrm{Ext}_{n,m}) = 1$ *for all* n, m.

(ii) $\mathrm{EXT}^{\mathrm{G}}$ *has, up to isomorphism, precisely one countably infinite model, the so-called random graph, or Rado graph,* $\mathfrak{R}$.

(iii) ASTG *is the* FO*-theory of the random graph. For all* $\mathrm{FO}(\{E\})$*-sentences* φ*:*

$$\mu^{\mathrm{G}}(\varphi) = 1 \quad \textit{iff} \quad \mathfrak{R} \models \varphi.$$

Exercise 3.3.2 Outline the proof of the above theorem in analogy with the proof of Theorem 3.2.4.

Exercise 3.3.3 One explicit representation of (the isomorphism type of) the random graph $\mathfrak{R}$ is given by the following structure. $\mathfrak{A} := (\mathbb{N}, E)$, where E is defined as follows. Let $p_0 = 2, p_1 = 3, p_3 = 5, \ldots$ be the enumeration of all primes. For $0 \leqslant n < m$ put

$$(n,m),(m,n) \in E \quad \text{iff} \quad p_n | m.$$

Check that $\mathfrak{A}$ satisfies the graph extension axioms.
It follows that $\mathfrak{A} \simeq \mathfrak{R}$ *is* (a presentation of) the random graph.

Example 3.3.4 An undirected finite graph almost surely has diameter 2. I.e., the following is almost surely true in undirected graphs (true in the random graph $\mathfrak{R}$):

$$\exists x \exists y \neg Exy \ \wedge \ \forall x \forall y (x = y \vee Exy \vee \exists z (Exz \wedge Ezy)).$$

In fact even $\exists x \exists y \neg Exy \ \wedge \ \forall x \forall y \exists z (Exz \wedge Ezy)$ is almost surely true. For the first conjunct, one may use the extension axiom $\mathrm{Ext}_{0,1}$ and $\mathrm{Ext}_{2,0}$ for the second part.

Example 3.3.5 Every finite graph $\mathfrak{G}_0$ is isomorphically embedded in the random graph $\mathfrak{R}$. Moreover, if $\mathfrak{G}_0 \subseteq \mathfrak{G}_1$ are finite graphs, and if $\rho_0 \colon \mathfrak{G}_0 \simeq \mathfrak{G}_0' \subseteq \mathfrak{R}$ is an isomorphic embedding of $\mathfrak{G}_0$ into $\mathfrak{R}$, then ρ_0 can be extended to an isomorphic embedding of $\mathfrak{G}_1$ into $\mathfrak{R}$: there exists $\rho_1 \supseteq \rho_0$ such that $\rho_1 \colon \mathfrak{G}_1 \simeq \mathfrak{G}_1' \subseteq \mathfrak{R}$.

For this one establishes first the second claim in the special case where $\mathfrak{G}_1$ is an extension of $\mathfrak{G}_0$ by a single vertex. The appropriate graph extension axiom takes care of this. Then, we proceed by induction on the number of vertices in $\mathfrak{G}_1 \setminus \mathfrak{G}_0$.

Exercise 3.3.6 For arbitrary finite relational τ consider the diameter of the Gaifman graph of finite τ-structures $\mathfrak{A}$, diameter$(G(\mathfrak{A}))$. Show that almost surely

$$\mathrm{diameter}(G(\mathfrak{A})) = \begin{cases} 2 & \text{if } \tau \text{ has at least one binary but no ternary relation,} \\ 1 & \text{if } \tau \text{ has at least one ternary relation.} \end{cases}$$

Part II

Logic and Complexity: Descriptive Complexity

Chapter 4

Monadic Second-Order Logic and Büchi's Theorem

4.1 Word Models

Fix finite alphabet Σ; recall:

- Σ^* the set of all Σ-words;
- $w = a_1 \dots a_n \in \Sigma^*$ ($a_i \in \Sigma$) has length $|w| = n$;
- $\varepsilon \in \Sigma^*$ the empty word, $\Sigma^+ = \Sigma^* \setminus \{\varepsilon\}$;
- $(\Sigma^*, \cdot, \varepsilon)$ with concatenation operation $\cdot$ the monoid of Σ-words;
- $(\Sigma^+, \cdot)$ the semigroup of non-empty Σ-words.

Definition 4.1.1 Let $\tau_\Sigma = \{<\} \cup \{P_a : a \in \Sigma\}$.

(i) The canonical *word model* associated with $w = a_1 \dots a_n \in \Sigma^+$ is the linearly ordered τ_Σ-structure $\mathfrak{A}_w = \big([n], <, (P_a^{\mathfrak{A}_w})_{a \in \Sigma}\big)$, where $[n] = \{1, \dots, n\}$ with the usual ordering $<$ and

$$P_a^{\mathfrak{A}_w} = \{i : a_i = a\}.$$

(ii) A (Σ-)word model is any τ_Σ-structure isomorphic to some $\mathfrak{A}_w$ for $w \in \Sigma^+$; $\underline{\Sigma}^+ \subseteq \mathsf{FIN}(\tau_\Sigma)$ the class of all Σ word models (the $\simeq$-closure of $\{\mathfrak{A}_w : w \in \Sigma^+\}$).

(iii) For Σ-languages $L \subseteq \Sigma^+$ put $\underline{L} := \{\mathfrak{B} : \mathfrak{B} \simeq \mathfrak{A}_w$ for some $w \in L\} \subseteq \underline{\Sigma}^+$ (the $\simeq$-closure of $\{\mathfrak{A}_w : w \in L\}$).

The connection between words and word models induces exact correspondences:

Σ-words $w \in \Sigma^+$	–	isomorphism classes of word models $\mathfrak{A}_w$
Σ-languages $L \subseteq \Sigma^+$	–	isomorphism classes of word models
concatenation of words	–	ordered sums of word models

Büchi's Theorem will enrich this picture by the correspondence between regularity and MSO-definability.

4.2 Regular Languages

Definition 4.2.1 For $L \subseteq \Sigma^*$ define *syntactic congruence* $\approx_L$ as an equivalence relation on Σ^* by

$$w \approx_L w' \quad \text{if} \quad \text{for all } x, y \in \Sigma^*\colon\ xwy \in L \ \Leftrightarrow\ xw'y \in L.$$

Observe that $\approx_L$ is a congruence w.r.t. concatenation: $v \approx_L v'$ and $w \approx_L w'$ implies $vw \approx_L v'w'$. The quotient of $(\Sigma^*, \cdot, \varepsilon)$ w.r.t. $\approx_L$ (well-defined) is called the syntactic monoid of L.

A right-handed version $\sim_L$ may be defined by

$$w \sim_L w' \quad \text{if} \quad \text{for all } x \in \Sigma^*\colon\ wx \in L \ \Leftrightarrow\ w'x \in L.$$

$\sim_L$ is right-invariant: $w \sim_L w'$ implies $wa \sim_L w'a$, but in general not left-invariant (and not a congruence).

$\approx_L$ is a refinement of $\sim_L$ and in particular also right-invariant.

Clearly L is a union of classes of $\sim_L$ as well as a union of classes of $\approx_L$.

Exercise 4.2.2 Show that if $L = L(\mathcal{A})$ for a deterministic automaton $\mathcal{A}$ with m states, then $\approx_L$ has at most m^m equivalence classes. Hint: associate unary operations on the sate set with $\approx_L$ classes.

Definition 4.2.3 A language $L \subseteq \Sigma^*$ is *regular* if it is recognised by a (deterministic) finite automaton, $L = L(\mathcal{A})$ for a DFA $\mathcal{A}$.

Note that this means that the word problem for L,

on input $w \in \Sigma^*$ decide whether $w \in L$

is solved by a DFA (with constant memory). The following is proved in automata and formal languages.

Theorem 4.2.4 (Myhill–Nerode) *T.f.a.e. for any $L \subseteq \Sigma^*$:*

(i) *L regular (DFA/NFA recognisable).*

(ii) *$\approx_L$ has finite index.*

(iii) *$\sim_L$ has finite index.*

(iv) *there is some right-invariant equivalence relation of finite index on Σ^* such that L is a union of equivalence classes.*

4.3 Büchi's Theorem

We concentrate on languages $L \subseteq \Sigma^*$, as we have no word model for the empty word. Note that $L \subseteq \Sigma^*$ is regular iff $L \setminus \{\varepsilon\} \subseteq \Sigma^+$ is regular.

Recall MSO, now over the signature τ_Σ for word models. We use first-order variables x,y,z, ... (ranging over elements) and second-order variables X,Y,Z, ... (ranging over subsets). MSO(τ_Σ) is the closure of the atomic formulae (of types $x = y$, $x < y$, $P_a x$ and Xx) under $\wedge$, $\vee$, $\neg$ and first- and second-order existential and universal quantification. MSO quantifier rank counts first- and second-order quantification without distinction.

Recall $\equiv_m^{\mathrm{MSO}}$ (MSO-equivalence up to quantifier rank m between structures with first- and second-order parameters) and its game characterisation in $\mathsf{G}_m^{\mathrm{MSO}}(\mathfrak{A}, \mathbf{P}, \mathbf{a}; \mathfrak{B}, \mathbf{Q}, \mathbf{b})$. Recall from Lemma 2.4.6 in Part I that $\equiv_m^{\mathrm{MSO}}$ is compatible with ordered sums.

Exercise 4.3.1 There exists $\varphi_0 \in \mathrm{FO}(\tau_\Sigma)$ for which $\underline{\Sigma}^+ = \mathsf{FMOD}(\varphi_0)$.

Call a class $K \subseteq \underline{\Sigma}^+$ MSO-definable if $K = \mathsf{FMOD}(\varphi) \cap \underline{\Sigma}^+$ for some sentence $\varphi \in \mathrm{MSO}(\tau_\Sigma)$ (if $K = \mathsf{FMOD}(\varphi \wedge \varphi_0)$ for φ_0 as in the exercise).

Similarly, K is said to be $\exists$-MSO-definable if $K = \mathsf{FMOD}(\varphi) \cap \underline{\Sigma}^+$ for some sentence $\varphi = \exists \mathbf{X} \psi(\mathbf{X})$ with $\psi(\mathbf{X}) \in \mathrm{FO}(\tau_\Sigma \cup \{\mathbf{X}\})$.

Theorem 4.3.2 (Büchi) *T.f.a.e. for any $K \subseteq \underline{\Sigma}^+$:*

(i) K MSO-*definable.*

(ii) $K = \underline{L}$ *for some regular* $L \subseteq \Sigma^+$.

(iii) K $\exists$-MSO-*definable.*

Corollary 4.3.3 MSO *and* $\exists$-MSO *have the same expressive power over word models (unlike, e.g., over the class of finite graphs).*

The proof of the theorem is immediate from the following two lemmas.

Lemma 4.3.4 *For any NFA $\mathcal{A}$ there is an* $\exists$-MSO *sentence $\varphi_{\mathcal{A}}$ such that for all $w \in \Sigma^+$:*

$$w \in L(\mathcal{A}) \quad \textit{iff} \quad \mathfrak{A}_w \models \varphi_{\mathcal{A}}.$$

Idea: use second-order variables X_q for the states q of $\mathcal{A}$ and an FO *formula $\psi(\mathbf{X})$ describing accepting runs of $\mathcal{A}$ on w in terms of a state-assignment $(\mathfrak{A}_w, (P_q))$ over $\mathfrak{A}_w$.*

The following lemma states that MSO model checking over word models can be done by DFA.

Lemma 4.3.5 *For any sentence $\varphi \in$* $\mathrm{MSO}(\tau_\Sigma)$ *there is a DFA $\mathcal{A}_\varphi$ such that for all $w \in \Sigma^+$:*

$$w \in L(\mathcal{A}_\varphi) \quad \textit{iff} \quad \mathfrak{A}_w \models \varphi.$$

Proof By Theorem 4.2.4, it suffices to provide some right-invariant equivalence relation on Σ^* such that $L_\varphi := \{w \in \Sigma^+ : \mathfrak{A}_w \models \varphi\}$ is a union of equivalence classes.

If $\mathrm{qr}(\varphi) = m$, then $\equiv_m^{\mathrm{MSO}}$ induces such an equivalence relation. Put

$$w \sim w' \quad \text{if} \quad \mathfrak{A}_w \equiv_m^{\mathrm{MSO}} \mathfrak{A}_{w'}.$$

Then clearly $\sim$ has finite index (just like $\equiv_m^{\mathrm{MSO}}$), and L_φ is a union of $\sim$-classes.

Right-invariance follows from compatibility of $\equiv_m^{\mathrm{MSO}}$ with ordered sums, Lemma 2.4.6 in Part I. In fact, $\sim$ is even a congruence w.r.t. concatenation. $\Box$

Variations (on the proofs) Instead of automata one may rely on the characterisation of the class of all regular Σ-languages as the smallest class of Σ-languages comprising the empty language $\emptyset$ and the singleton languages $\{a\}$ for all $a \in \Sigma$ that is closed under the language operations of union, concatenation and star.

Exercise 4.3.6 Show MSO-definability of all (classes of word models of) regular languages by direct induction on the generation of regular languages (or on the syntax of regular expressions). More precisely, provide, by induction, for every regular language L an MSO-definition of $\underline{L \setminus \{\varepsilon\}}$.

An inductive approach to the opposite direction, from MSO to automata or to regular languages, is seemingly hampered by the problem of free variables in subformulae.

Free second-order variables can be treated through an appropriate extension of the alphabet Σ. In order to describe a structure $(\mathfrak{A}_w, \mathbf{P})$ where $\mathbf{P} = (P_1, \ldots, P_k)$ as a word, we may use the alphabet $\Sigma \times \mathbb{B}^k$ and associate with position $i \in [n] = A$ the letter $(a, b_1, \ldots, b_k)$ if $i \in P_a$ and $i \in P_j$ for precisely those j with $b_j = 1$.

First-order variables (and parameters) may be eliminated in favour of second-order variables as follows. Let $\mathrm{MSO}^\circ(\tau_\Sigma)$ be the variant MSO logic with atomic formulae

$$\begin{array}{ll} X \subseteq P_a & \text{(with the obvious semantics)} \\ X \subseteq Y & \text{(with the obvious semantics)} \\ X < Y & \text{(with semantics: ``}\emptyset \neq X \times Y \subseteq <\text{'')} \end{array}$$

closed under $\wedge$, $\vee$, $\neg$ and existential and universal second-order quantification.
There is an effective translation from ordinary $\mathrm{MSO}(\tau_\Sigma)$ into $\mathrm{MSO}^\circ(\tau_\Sigma)$.

Exercise 4.3.7 (a) Provide MSO°-formalisations for "$X = \emptyset$", "$X \cap Y = Z$", "$X \cup Y = Z$", "X is a singleton set" and "$<$ is a linear ordering of the domain".

(b) Sketch a construction of model checking automata $\mathcal{A}_\varphi$ for $\varphi \in \mathrm{MSO}^\circ(\tau_\sigma)$, by induction on φ. For $\varphi = \varphi(X_1, \ldots, X_k)$ we want that

$$\mathfrak{A}_w \models \varphi[\mathbf{P}] \quad \text{iff} \quad \mathcal{A}_\varphi \text{ accepts } w_{\mathfrak{A},\mathbf{P}},$$

where $w_{\mathbf{P}}$ is the canonical word representation of $(\mathfrak{A}_w, \mathbf{P})$ over the alphabet $\Sigma \times \mathbb{B}^k$ as indicated above.

Chapter 5

Excursion: Computational Complexity

5.1 Turing Machines

We here look at decision problems exclusively. A decision problem is a problem which requires "yes"/"no" answers on any admissible input; it can be specified by two sets

$$\begin{array}{rl} I & \text{the set of } \textit{instances} \\ D \subseteq I & \text{the subset of } \textit{positive instances}, \end{array}$$

where $D \subseteq I$ is just the set of those instances for which the answer is "yes", and $I \setminus D$ the set of those for which the answer is "no". At a syntactic level, I and D are languages over some alphabet Σ used for the encoding of the instances. $D \subseteq I \subseteq \Sigma^*$ is thus a (sub-)language recognition problem.

We use Turing machines of a particular format suitable for the input/output requirements of decision problems to formalise the notion of an algorithmic solution of a decision problem. For complexity consideration one treats both deterministic and non-deterministic Turing machines.

Deterministic Turing machines The following definition of a Turing machine is suited to language recognition problems. For some language $L \subseteq \Sigma^*$, we want to deal with inputs $w \in \Sigma^*$, which the machine is to accept or reject according to whether $w \in L$ or not.

Definition 5.1.1 [DTM]
A deterministic Turing machine (DTM) with work tape alphabet $\Gamma \supseteq \Sigma \cup \{\Box\}$ ($\Box$ the *blank*) is a tuple

$$\mathcal{M} = \big(\Gamma, Q, q_0, q^+, q^-, q^\perp, \delta\big)$$

Q the finite *set of states* with distinct special states $q_0, q^+, q^-, q^\perp \in Q$:

$$\begin{array}{ll} q_0 \in Q & \text{the } \textit{initial state} \\ q^+ \in Q & \text{the } \textit{accepting final state} \\ q^- \in Q & \text{the } \textit{rejecting final state} \\ q^\perp \in Q & \text{the } \textit{garbage state} \end{array}$$

δ the *transition function.*

The transition function has the format

$$\delta\colon Q \times \Gamma \rightarrow \Gamma \times \{-1, 0, +1\} \times Q$$

and, depending on internal state and symbol currently read, specifies symbol to be printed, head movement to be carried out, and successor state.

We assume that $\delta(q, b) = (b, 0, q)$ for $q \in \{q^+, q^-, q^\perp\}$ (which will guarantee stationary configurations).

For a computation of $\mathcal{M}$ on some input word $w = a_0 \dots a_{n-1} \in \Sigma^*$ we assume an initialisation that has a_i written into tape cell i for $i = 0, \dots, n-1$, all other tape cells blank; the head is located at tape cell 0; $\mathcal{M}$ in state q_0.

A *configuration* of $\mathcal{M}$ is a complete description of its overall state, comprising of

internal (control) state:	$q \in Q$
head position over tape:	tape cell index $\ell \in \mathbb{N}$
full tape content:	a function $\rho\colon \mathbb{N} \rightarrow \Gamma$.

We use notation $C = (q, \ell, \rho)$ for configurations, and write $C_t[w]$ for the configuration in step t of the computation on input w. The function ρ is used for specifying the tape content in tape cell i as $\rho(i)$. In any configuration arising in a computation of $\mathcal{M}$, $\rho(i)$ will differ from $\Box$ only in a finite region around $i = 0 \in \mathbb{N}$, because all tape cells not yet visited by the head must still be blank.

The *initial configuration* on input w, according to our initialisation convention, is given by $C_0[w] = (q_0, 0, \rho_0)$ where

$$\rho_0(i) = \begin{cases} a_i & \text{for } i < n = |w| \\ \Box & \text{else.} \end{cases}$$

A *run* of $\mathcal{M}$ on input w is the sequence $C_0[w], C_1[w], \dots$ of configurations, starting with the initial configuration $C_0[w]$, and $C_{t+1}[w]$ always determined as the *successor configuration* $C_t[w]'$ of $C_t[w]$. Successor configurations are determined according to the transition function δ:

$$C = (q, \ell, \rho) \longmapsto C' = (q', \ell + d, \rho') \qquad \text{with } \rho'(i) = \begin{cases} b' & \text{for } i = \ell \\ \rho(i) & \text{for } i \neq \ell \end{cases}$$

if $\delta(q, \rho(\ell)) = (b', d, q')$ and $\ell + d \geqslant 0$.

$$C = (q, \ell, \rho) \longmapsto C' = (q^\perp, 0, \rho)$$

else, i.e., if $\ell = 0$ and $\delta(q, \rho(\ell)) = (b', -1, q')$ [head trying to fall off the tape].

In words: $\delta(q, \rho(\ell)) = (b', d, q')$ tells $\mathcal{M}$ to print b over the currently read letter (which is $\rho(\ell)$), move the head by d to the right, and to enter state q'.

Termination the computation of $\mathcal{M}$ on w terminates (within k steps) if a configuration $C_t[w] = (q, \ell, \rho)$ with $q \in \{q^+, q^-\}$ is reached (for some $t < k$).

Acceptance A run of $\mathcal{M}$ on input w is accepting if it terminates with $\mathcal{M}$ in the accepting final state q^+, rejecting if it terminates in the rejecting final state q^-. Note that there can also be non-terminating runs; these are also called divergent and are neither accepting nor accepting (no decision reached).

We use the following symbolic notation for acceptance, rejection, termination and divergence:

$w \xrightarrow{\mathcal{M}} q^+$	"the run of $\mathcal{M}$ on input w is accepting"
$w \xrightarrow{\mathcal{M}} q^-$	"the run of $\mathcal{M}$ on input w is rejecting"
$w \xrightarrow{\mathcal{M}} < k$	"the run of $\mathcal{M}$ on input w terminates within k steps"
$w \xrightarrow{\mathcal{M}} \infty$	"the run of $\mathcal{M}$ on input w diverges"

Definition 5.1.2 A DTM $\mathcal{M}$ solves the decision problem associated with $D \subseteq \Sigma^*$ (or decides D) if for all $w \in \Sigma^*$:

$$\begin{aligned} w &\xrightarrow{\mathcal{M}} q^+ \quad \text{for } w \in D, \\ w &\xrightarrow{\mathcal{M}} q^- \quad \text{for } w \notin D. \end{aligned}$$

(For $D \subseteq I \subseteq \Sigma^*$, we only require correct termination for admissible inputs $w \in I$.)

Non-deterministic Turing machines A non-deterministic Turing machine (NTM) has the format $\mathcal{M} = \big(\Gamma, Q, q_0, q^+, q^-, \Delta\big)$ with a *transition relation*

$$\Delta \subseteq Q \times \Gamma \times \Gamma \times \{-1, 0, +1\} \times Q.$$

Tuples (q, b, b', d, q') give rise to *possible* transitions to successor configurations in the obvious manner: when in state q and reading letter b, the run may proceed to a successor configuration obtained by overwriting b with b', moving the head by d and entering control state q' (provided the head was not already in position 0 and $d = -1$).

A *run* of $\mathcal{M}$ on input w is a (finite or infinite) sequence $C_0[w], C_1[w], \ldots$ of configurations, starting with the initial configuration $C_0[w]$, and with $C_t[w]$ one of the allowed successor configurations of $C_{t-1}[w]$ according to Δ.

As in general a configuration may have one or several possible successor configurations or none, the set of runs on a given input forms a tree.

An *accepting run* is one in which state q^+ is reached. A run terminates if q^+, q^- or some configuration without successor is reached.

Definition 5.1.3 An NTM $\mathcal{M}$ *solves the decision problem* associated with $D \subseteq \Sigma^*$ (or decides D) if for all $w \in \Sigma^*$, all runs of $\mathcal{M}$ on w terminate and $\mathcal{M}$ has an accepting run on w if, and only if, $w \in D$. (Modification for $D \subseteq I \subseteq \Sigma^*$ as above.)

5.2 Resource Bounds and Complexity Classes

Definition 5.2.1 For a function $f : \mathbb{N} \to \mathbb{N}$ we say that an NTM or DTM $\mathcal{M}$ is

(i) f *time bounded* if for all inputs w, all computations of $\mathcal{M}$ on input w terminate within $f(|w|)$ many steps.

(ii) f *space bounded* if for all inputs w, all computations of $\mathcal{M}$ on input w have head positions $\ell \leqslant f(|w|)$ throughout.

In particular, an NTM or DTM is polynomially time or space bounded if it is f time or space bounded for some polynomial $f(n)$.

Definition 5.2.2 A decision problem is in

(i) P (or Ptime, deterministic polynomial time), if it is solvable by some polynomially time bounded DTM.

(ii) NP (non-deterministic polynomial time), if it is solvable by some polynomially time bounded NTM.
(iii) Pspace (polynomial space) if it is solvable by some polynomially space bounded DTM.

Straightforward arguments show that

$$\text{Ptime} \subseteq \text{NP} \subseteq \text{Pspace} \subseteq \text{Exptime},$$

where Exptime is the class defined via exponentially time bounded DTM (with time bounds $f(n) = 2^{p(n)}$, p a polynomial). It is not currently known which of these inclusions are strict (e.g., the famous P/NP problem), apart from the known

$$\text{Ptime} \subsetneq \text{Exptime}.$$

The complexity class P is generally regarded as an appropriate idealisation of *feasibly* solvable decision problems. Many important decision problems are known to be in NP (through polynomially bounded "guessing and checking"), with no indication how they could be solved deterministically with polynomial time bounds. Moreover, many natural NP problems can be shown to be NP complete in the sense that a deterministic polynomial time bounded solution would imply that all of NP collapses to P.

Note that the complexity measures used here are essentially *asymptotic.* They distinguish different growth rates of required resources, in their dependence on input size, as input size tends to infinity. In particular, one may always ignore finitely many inputs (all inputs of some bounded size), as they could be treated trivially by "table look-up".

Major complexity classes like the above are very robust in the sense that natural variations in the specific details of the machine model do not affect the classes.

All the usual algorithms based on some natural intuition of polynomially bounded iterations can in fact be "implemented" in polynomially bounded DTM.

5.3 Finite Structures As Inputs: Unavoidable Coding

We want to consider structures $\mathfrak{A} \in \mathsf{FIN}(\tau)$ (for finite relational τ) as inputs for decision problems $Q \subseteq \mathsf{FIN}(\tau)$ where Q is a boolean query on finite τ-structures (i.e., an isomorphism closed subclass. Think of Q as the subclass of those finite τ-structures that have some structural property we are interested in; then we want to decide, given any finite τ-structure, whether it has this property.) However, $\mathfrak{A} \in \mathsf{FIN}(\tau)$ cannot be fed to a Turing machine directly, but via some encoding as an input word. For this purpose we may for instance associate binary strings $\langle\mathfrak{A}\rangle \in \mathbb{B}^*$ with $\mathfrak{A} \in \mathsf{FIN}(\tau)$.

Example 5.3.1 Consider graphs $\mathfrak{A} = (V, E^{\mathfrak{A}})$ with universes $V = \{0, \dots, n-1\}$ of size n. We regard $n = |\mathfrak{A}|$ as the input size.

$\mathfrak{A}$ is faithfully representable by the boolean *adjacency matrix* $A_{\mathfrak{A}} \in \mathbb{B}^{n,n}$ with entries $a_{ij} = 0$ or $a_{ij} = 1$ depending on whether $(i, j) \in E^{\mathfrak{A}}$ or not. We may then encode $\mathfrak{A}$ as a single boolean string $\langle\mathfrak{A}\rangle$ by concatenating the rows of this matrix to form one word of length n^2 over the alphabet $\mathbb{B}$.

Since we are here only interested in sizes and complexity bounds up to polynomial re-scalings, the discrepancy between input size $|\mathfrak{A}|$ (taken to be the universe size) and the length of the actual encoding $\langle\mathfrak{A}\rangle$ (which is n^2 in the example) does not matter. The actual encoding details do not matter too much either; the robustness of complexity classes such as P, NP, Pspace also covers natural variations in coding.[1]

If τ has several relations (of various arities) we may chose an encoding that similarly consists of some systematic concatenation of the entries in (higher-dimensional) boolean matrices, one for each relation.

Definition 5.3.2 For finite relational τ let $\mathsf{STAN}(\tau) \subseteq \mathsf{FIN}(\tau)$ be the class of finite τ-structures $\mathfrak{A}$ with a standard universe of the form $\{0, \dots, n-1\}$, $n = |\mathfrak{A}| \geqslant 1$.
For $\mathfrak{A} \in \mathsf{STAN}(\tau)$, we let $\mathfrak{A}_<$ be its expansion by the natural ordering $<$ on its domain $\{0, \dots, |\mathfrak{A}| - 1\}$.

For notational convenience we often write just n for the standard set $\{0, \dots, n-1\}$ of n elements, and similarly for instance n^k for the standard set $\{0, \dots, n^k - 1\}$ of n^k elements, etc.

We assume a fixed natural encoding scheme for structures $\mathfrak{A} \in \mathsf{STAN}(\tau)$ by $\langle\mathfrak{A}\rangle \in \mathbb{B}^*$, such that

- for $\mathfrak{A} \in \mathsf{STAN}(\tau)$, $\langle\mathfrak{A}\rangle$ uniquely determines $\mathfrak{A}$.
- $\langle\mathfrak{A}\rangle$ is of polynomial length in n for $|\mathfrak{A}| = n$, $|\langle\mathfrak{A}\rangle| < n^k$;
  coding numbers in $n^k = \{0, \dots, n^k - 1\}$ as k-digit numbers to base n, the boolean entries of the word $\langle\mathfrak{A}\rangle$ are FO definable in $\mathfrak{A}_<$.
- the set $I = \{\langle\mathfrak{A}\rangle \colon \mathfrak{A} \in \mathsf{STAN}(\tau)\}$ is decidable at low complexity;
  say, for our purposes, at least in P.

These are easily checked for instance for the adjacency list encoding of binary relations (and its generalisation) indicated above.

Definition 5.3.3 Let $Q \subseteq \mathsf{FIN}(\tau)$ be a boolean query (an isomorphism closed class of finite τ-structures). We say that Q is in P (or in NP, or in Pspace) if the associated decision problem $D \subseteq I$,

$$\begin{aligned} I &= \{\langle\mathfrak{A}\rangle \colon \mathfrak{A} \in \mathsf{STAN}(\tau)\} \\ D &= \{\langle\mathfrak{A}\rangle \in I \colon Q^{\mathfrak{A}} = 1\} \;=\; \{\langle\mathfrak{A}\rangle \in I \colon \mathfrak{A} \in Q\} \end{aligned}$$

is in P (or in NP, or in Pspace, respectively).

Remark 5.3.4 *It is not hard to see (but tedious to detail) that for instance all* FO *definable classes are in P; that algorithms based on polynomially bounded iteration, like transitive closures, depth or breadth first search in graphs, etc., can all be implemented in polynomially time bounded DTM, and are thus available for algorithmic solutions establishing membership of structural decision problems $Q \subseteq \mathsf{FIN}(\tau)$ in P.*

Since everything that is of interest for us is already reflected in the notationally much simpler cases of (ordered or unordered) graphs, we shall often deal explicitly with $\tau = \{E\}$ (one binary relation for, e.g., graphs) and $\tau_< = \{E, <\}$ (two binary relations for, e.g., linearly ordered graphs). All considerations do generalise to arbitrary fixed finite relational τ.

[1] But beware of exponential gaps, as can occur for instance in the encoding of numerical values, between unary and binary encodings.

Note on the role of order When deciding a structural property for finite τ-structures (a boolean query, which is explicitly required to be isomorphism invariant!), we implicitly require that all standard realisations of the same structure (or its isomorphic copies) produce the same result. If $\mathfrak{A} \simeq \mathfrak{B}$, $\mathfrak{A}, \mathfrak{B} \in \mathsf{STAN}(\tau)$, are two (standard) realisations of the same isomorphism type, then $\langle \mathfrak{A} \rangle \in D \Leftrightarrow \langle \mathfrak{B} \rangle \in D$.

This is a crucial semantic consistency constraint on algorithms (Turing machines) that decide queries.

If we are dealing with a class of linearly ordered structures, over $\tau_< = \tau \cup \{<\}$, and restrict the inputs to be (encodings of) structures that interpret $<$ as a linear ordering of their universe, this problem can be avoided. The isomorphism type of any such structure has a unique representative determined as *the* member of its isomorphism class in STAN that interprets the ordering $<$ as the standard ordering on $\{0, \dots, |\mathfrak{A}|-1\}$. As a result we can have, for linearly ordered structures, a one-to-one correspondence between encodings $\langle \mathfrak{A} \rangle$ and isomorphism classes $\{\mathfrak{B} \colon \mathfrak{B} \simeq \mathfrak{A}\}$.

Chapter 6

Existential Second-Order Logic and Fagin's Theorem

6.1 Existential Second-Order Logic

Definition 6.1.1 Second-order logic SO is the extension of FO by quantification over relation variables of any arity. We write $\mathrm{SO}(\tau)$ for the set of SO formulae over signature τ. Existential second-order logic $\exists$-SO is the fragment of SO consisting of formulae of the form

$$\exists X_1 \ldots \exists X_s \;\; \psi(X_1, \ldots, X_s),$$

where $\psi \in \mathrm{FO}(\tau \cup \{X_1, \ldots, X_s\})$, X_i a second-order variable of arity r_i.

Exercise 6.1.2 Show that the following graph queries are definable in $\exists$-SO over GRAPH $\subseteq$ $\mathsf{FIN}(\tau)$ for $\tau = \{E\}$:

(i) BIPART $:= \{\mathfrak{A} \in \mathrm{GRAPH}\colon\; \mathfrak{A}$ bipartite $\}$.
(ii) MATCH $:= \{\mathfrak{A} \in \mathrm{BIPART}\colon\; \mathfrak{A}$ has a perfect matching $\}$.
(iii) 3-COL $:= \{\mathfrak{A} \in \mathrm{GRAPH}\colon\; \mathfrak{A}$ 3-colourable $\}$.
(iv) HAM $:= \{\mathfrak{A} \in \mathrm{GRAPH}\colon\; \mathfrak{A}$ has a Hamilton cycle $\}$.

Remark: (i) and (ii) are in P, (iii) and (iv) NP-complete.

Lemma 6.1.3 *Any $\exists$-SO definable query is in NP.*

Proof Let $Q = \mathsf{FMOD}(\varphi)$, $\varphi = \exists X_1 \;\ldots\; \exists X_s \, \psi(X_1, \ldots, X_s)$ with $\psi \in \mathrm{FO}$. Consider a polynomially time bounded DTM for the FO query defined by ψ,

$$Q_\psi := \big\{(\mathfrak{A}, R_1, \ldots, R_s)\colon (\mathfrak{A}, R_1, \ldots, R_s) \models \psi\big\} \subseteq \mathsf{FIN}(\tau \cup \{R_1, \ldots, R_s\})$$

with relations R_i of arities r_i matching the second-order variables X_i of φ. (Compare Remark 5.3.4 above.) We obtain an NTM deciding Q which operates in two phases:

(1) extend the input $\langle\mathfrak{A}\rangle$ to an encoding $\langle\mathfrak{A}, R_1, \ldots, R_s\rangle$ of an expansion $(\mathfrak{A}, R_1, \ldots, R_s)$. In an adjacency matrix encoding this is achieved by non-deterministically appending an arbitrary $\{0,1\}$-word of the appropriate length.
(2) simulate the DTM that checks for satisfaction of ψ on the result of phase 1.

As the encoding in phase 1 is polynomial length, the overall computation remains polynomially time bounded. It clearly decides Q. $\square$

6.2 Coding Polynomially Bounded Computations

Recall how existential MSO (the existential fragment of the monadic fragment of SO) is used in Büchi's theorem to encode acceptance of a word by an automaton over the corresponding word model. We want to do the same for acceptance of (the encoding of) a structure $\mathfrak{A}$ by a polynomially bounded NTM.

Assume that $Q \subseteq \mathsf{FIN}(\tau)$ is in NP: there is a polynomially time bounded NTM $\mathcal{M}$ which on input $\langle \mathfrak{A} \rangle$ determines whether $\mathfrak{A} \in Q$. We want to translate this into the definability of the class Q in a suitable logic $\mathcal{L}$. I.e., we look for a logical description of "acceptance of $\langle \mathfrak{A} \rangle$ by $\mathcal{M}$" as a property of $\mathfrak{A}$.

To this end we firstly encode possible (accepting) runs of $\mathcal{M}$ on $\langle \mathfrak{A} \rangle$ within $\mathfrak{A}$ itself, through suitable relations over the universe $\{0, \ldots, n-1\}$ of $\mathfrak{A} \in \mathsf{STAN}(\tau)$ $(n = |\mathfrak{A}|)$.

For the given NTM $\mathcal{M}$, enumerate the set of tape symbols as $\{b_0 = \Box, b_1, \ldots, b_{r-1}\}$ and the state set as $\{q_0, q_1 = q^+, \ldots, q_{s-1}\}$. We may identify Γ with the set $\{0, \ldots, r-1\}$ and the state set with the set $\{0, \ldots, s-1\}$. Chose $k > 0$ and $m \in \mathbb{N}$ such that $m \geqslant r, s$ and such that for inputs $\langle \mathfrak{A} \rangle$ of size $|\mathfrak{A}| = n \geqslant m$ all computations of $\mathcal{M}$ terminate within n^k many steps.

Recall that we write just n for the standard set $\{0, \ldots, n-1\}$ of n elements, similarly for instance n^k for the standard set $\{0, \ldots, n^k - 1\}$ of n^k elements, etc. We now also associate numbers in the range $\{0, \ldots, n^k - 1\}$ (e.g., for time indices and tape cell indices in configurations of $\mathcal{M}$) with k-tuples over $\{0, \ldots, n-1\}$, simply by use of number representations to base n. In other words, we treat the elements of the universe of $\mathfrak{A}$, $n = \{0, \ldots, n-1\}$, as digits. Tuples $\mathbf{t} = (t_{k-1}, \ldots, t_0) \in \{0, \ldots, n-1\}^k$ encode numbers $0 \leqslant t < n^k$ as $t = \sum_i t_i n^i$ in n^k.

Exercise 6.2.1 Over the universes $\{0, \ldots, n-1\}$ with the natural linear ordering $<$ regard k-tuples $\mathbf{a}, \mathbf{b}, \mathbf{c}$ as representations to base n of numbers $a, b, c < n^k$.

(a) provide FO definitions $\varphi(\mathbf{x}, \mathbf{y})$ of "$a < b$" and $\psi(\mathbf{x}, \mathbf{y})$ of "$b = a + 1$".

(b) show that (for $k = 1$) the graph of addition is not uniformly FO definable in restriction to $\{0, \ldots, n-1\}$ with the natural order: there is no $\varphi(x, y, z) \in \mathrm{FO}(<)$ such that for all n, $a, b, c < n$: $a + b = c$ iff $(n, <) \models \varphi[a, b, c]$.
(Hint: EVEN is not FO definable.)

(c) show that there is $\varphi \in \mathrm{FO}(<, R)$ for $3k$-ary relation symbol R such that for all n and $R \subseteq \{0, \ldots, n-1\}^{3k}$:
$(n, <) \models \varphi$ iff $R = \{(\mathbf{a}, \mathbf{b}, \mathbf{c}) \in \{0, \ldots, n-1\}^{3k} \colon a + b = c\}$.
(Hint: use the inductive definition of addition to force the correct R.)

Consider any configuration $C = (q, \ell, \rho)$ in a run of $\mathcal{M}$ on size n input (for $n \geqslant m$). We may code this configuration numerically by associating with the state $q = q_i$ its index $i < s \leqslant n$, with the head position $\ell < n^k$, a k-tuple $\boldsymbol{\ell}$ representing this number to base n, and with ρ (the graph of) a function from tape cell indices $i < n^k$ to indices $j < r \leqslant n$ of the letters $\rho(i) = b_j$ in those positions.

A run of $\mathcal{M}$ is a sequence of length $T \leqslant n^k$ of such configurations, $(C_t)_{t<T} = (q_t, \ell_t,$ where also the dependency on time indices $t < n^k$ can be coded through k-tuples $\mathbf{t}$. The corresponding functions are:

$t \mapsto q_t$	with graph $S = \{(\mathbf{t}, q_t) \colon t < T\} \subseteq \{0, \ldots, n-1\}^{k+1}$
$t \mapsto \boldsymbol{\ell}_t$	with graph $H = \{(\mathbf{t}, \boldsymbol{\ell}_t) \colon t < T\} \subseteq \{0, \ldots, n-1\}^{2k}$
$t \mapsto \rho_t, \quad \rho_t : i \mapsto \rho_t(i)$	with graph $R = \{(\mathbf{t}, \mathbf{i}, \rho_t(i)) \colon t < T, i < n^k\} \subseteq \{0, \ldots, n-1\}^{2k+1}$

In order to identify elements as digits in $n = \{0, \ldots, n-1\}$ (the universe of $\mathfrak{A}$), we use a linear ordering $<$ on the universe. Note that $\mathfrak{A}$ need not be linearly ordered of its own, i.e., as a τ-structure, even if we refer to standard realisations on universes $\{0, \ldots, n-1\}$ that do have a natural ordering.)

The goal is now to capture acceptance of $\langle\mathfrak{A}\rangle$ by $\mathcal{M}$ in the form

$$\exists < \exists S \, \exists H \, \exists R \;\; \psi(<, S, H, R),$$

where ψ needs to express that $<$ is a linear ordering of the domain, and that (w.r.t. this ordering) S, H and R encode an accepting run in the intended manner, over input $\langle\mathfrak{A}\rangle$. The proof of the lemma indicates how this can be done with suitable $\psi \in$ FO, whence $\varphi \in \exists$-SO.

Lemma 6.2.2 *Let $\mathcal{M}$ be polynomially bounded NTM deciding the boolean query $Q \subseteq \mathsf{FIN}(\tau)$. Then Q is $\exists$-SO definable as a subclass of $\mathsf{FIN}(\tau)$. For a suitable sentence $\varphi \in \exists\text{-SO}(\tau)$:*

$$Q = \mathsf{FMOD}(\varphi).$$

Proof Let the given NTM be as above, n sufficiently large for the encoding of states and tape symbols and such that $\mathcal{M}$ is n^k time bounded on all inputs of size n. Using new relations $<, S, H, R$ or arities $2, k+1, 2k, 2k+1$, respectively, as indicated above, we essentially need to express the following in FO

(i) $<$ is a linear ordering of the domain.

The following conditions refer to numbers in the range $\{0, \ldots, n^k - 1\}$ in representations to base n as k-tuples, as outlined above.

(ii) There is some $T \leqslant n^k$ such that
   - (a) S is the graph of a function from T to $s = \{0, \ldots, s-1\}$ with values 0 (for q_0) at 0 and 1 (for $q_1 = q^+$) at $T - 1$.
   - (b) H is the graph of a function from T to n^k with value 0 at 0.
   - (c) for $t < T$, encoded as $\mathbf{t}$: $R\mathbf{t}\_ := \{(\mathbf{i}, j) \colon (\mathbf{t}, \mathbf{i}, j) \in R\}$ is the graph of a total function from n^k to $r = \{0, \ldots, r-1\}$.

(iii) $R\mathbf{0}\_$ represents the function that encodes tape content $\langle\mathfrak{A}\rangle$ w.r.t. to the ordering $<$ of $\mathfrak{A}$.

(iv) The encoded functions are updated according to admissible transitions in Δ as $t \mapsto t' = t + 1$ for $t + 1 < T$.

All this can indeed be expressed in FO, and $\varphi = \exists < \exists S \, \exists H \, \exists R \;\; \psi(<, S, H, R)$ is as desired (for all sufficiently large $\mathfrak{A}$, the rest can be taken care of by FO sentences that characterise them up to isomorphisms). $\square$

Together Lemmas 6.1.3 and 6.2.2 give Fagin's correspondence between NP and $\exists$-SO, a key result of descriptive complexity.

Theorem 6.2.3 (Fagin)
For any class $Q \subseteq \mathsf{FIN}(\tau)$, that is closed under isomorphism, t.f.a.e.:

(i) *Q is in NP.*

(ii) *Q is $\exists$-SO definable within $\mathsf{FIN}(\tau)$: $Q = \mathsf{FMOD}(\varphi)$ for some $\varphi \in \exists\text{-SO}(\tau)$.*

Note that it also offers a machine-independent, natural characterisation of the complexity class NP. It also reflects (via the proof of Lemma 6.1.3) a known and useful normal form for NP algorithms, consisting of two phases:

(1) non-deterministic "guessing" of a polynomially size bounded "certificate", (viz., the predicates S, H, R in our formalisation), followed by
(2) deterministic polynomial time validation of this certificate, (viz., checking the FO-query defined by ψ in our formalisation).

Chapter 7

Fixpoint Logics

7.1 Recursion On First-Order Operators

As noted in Part I, FO has very limited expressive power e.g., for non-local properties, and also for properties that intuitively involve some iterative or dynamic concepts. The transitive closure of a binary relation E, for instance, is very easily generated by a recursion based on the relational operation $R \mapsto R \cup R \circ R$ where $R \circ R := \{(x, y) \colon \exists z(Rxz \wedge Rzy)\}$. Over a structure $\mathfrak{A} = (A, E^{\mathfrak{A}})$ of size n, the iterative application of this operator, starting with the initialisation $R_0 := E^{\mathfrak{A}}$ produces a sequence of stages $R_{i+1} = R_i \cup R_i \circ R_i$. This iteration terminates in the sense of reaching a stage R_i with $R_i \circ R_i \subseteq R_i$ and hence a fixpoint $R_{i+1} = R_i$. (How many steps can this take?) This final value for R_i is the transitive closure of $E^{\mathfrak{A}}$.

In this section we examine several extensions of FO by recursive mechanisms that allow us to define and use the results of well-defined recursion mechanisms based on definable relational operations.

Consider any formula $\varphi(X, \mathbf{x}) \in \mathcal{L}(\tau \cup \{X\})$ (any suitable logic $\mathcal{L}$) with free second-order variable X of arity r and matching tuple of first-order variables $\mathbf{x} = (x_1, \dots, x_r)$. On τ-structures $\mathfrak{A}$, one needs to supply assignments $P \subseteq A^r$ for X and $\mathbf{a} \in A^r$ for $\mathbf{x}$ in order to evaluate $\varphi[P, \mathbf{a}]$ (to a boolean value). Alternatively we may think of φ as mapping an assignment $P \subseteq A^r$ for X to the r-ary relation $P' := \{\mathbf{a} \in A^r \colon \mathfrak{A} \models \varphi[P, \mathbf{a}]\} \subseteq A^r$. In this sense φ is a *global predicate transformer*, operating on the set of r-ary relations over each τ-structure.

Definition 7.1.1 With a formula $\varphi(X, \mathbf{x})$ with free variables as indicated and of matching arity $r > 0$, associate the operator

$$\begin{array}{rcl} F_\varphi^{\mathfrak{A}} \colon \mathcal{P}(A^r) & \longrightarrow & \mathcal{P}(A^r) \\ P & \longmapsto & F_\varphi^{\mathfrak{A}}(P) := \{\mathbf{a} \in A^r \colon \mathfrak{A} \models \varphi[P, \mathbf{a}]\}. \end{array}$$

over all structures $\mathfrak{A}$ that interpret φ (up to assignments for X and $\mathbf{x}$ that is).
When $\mathfrak{A}$ is clear from context, we often just write F_φ. Formulae with additional free first- and second-order variables, beside X and $\mathbf{x}$ can be similarly treated, with assignments to these extra free variables as parameters.

Definition 7.1.2 An operation $F \colon \mathcal{P}(D) \to \mathcal{P}(D)$, over a finite domain D (e.g., $D = A^r$ for some $r > 0$, A the universe of a structure $\mathfrak{A}$) is called

(i) *monotone* if for all $P_1, P_2 \subseteq D$: $P_1 \subseteq P_2 \Rightarrow F(P_1) \subseteq F(P_2)$.

(ii) *inductive* (meaning inductive on $\emptyset$) if the sequence $(F^n(\emptyset))_{n\in\mathbb{N}}$ is increasing in the sense that $\emptyset \subseteq F(\emptyset) \subseteq F(F(\emptyset)) \subseteq \cdots$

(iii) *eventually constant* (meaning eventually constant on $\emptyset$) if the sequence $(F^n(\emptyset))_{n\in\mathbb{N}}$ is eventually constant in the sense that $F^{i+1}(\emptyset) = F^i(\emptyset)$ for some $i \in \mathbb{N}$.

$P \subseteq D$ is called *a fixpoint* of the operation F if $F(P) = P$.

NB: For iterates F^i of an operator F, we always set F^0 to be the identity operation, and inductively put $F^{i+1} := F \circ F^i$.

Note that over finite D the sequence $(F^i(\emptyset))_{i\in\mathbb{N}}$ must be eventually periodic. It can fail to be eventually constant only if it becomes periodic of a non-trivial period.

Exercise 7.1.3 For operations $F\colon \mathcal{P}(D) \to \mathcal{P}(D)$ as in the definition

(i) show that (F monotone) $\Rightarrow$ (F inductive) $\Rightarrow$ (F eventually constant).

(ii) give examples (of FO definable operations over suitable structures $\mathfrak{A}$, $D = A^r$) that such operations can be eventually constant without being inductive, or inductive without being monotone.

(iii) show that for monotone F, the sequence $F^i(D)_{i\in\mathbb{N}}$ is monotone decreasing: $A^r \supseteq F(D) \supseteq F(F(D)) \supseteq \cdots$

(iv) show that the induced operation $F^+\colon P \mapsto P \cup F(P)$ is always inductive (but not necessarily monotone).

Definition 7.1.4 For $\varphi(X, \mathbf{x}) \in FO$, we say that φ is *positive in* X if X only appears in the scope of an even number of negations within φ.[1] This syntactic notion extends to many other logics, in particular the fixpoint extensions of FO to be introduced below.

Examples: $Xxx \vee \neg\exists y \forall z(\neg Xxy \vee Yxy)$ is positive in X but not in Y.

It is a fact from classical FO logic that the operation F_φ associated with an FO formula $\varphi(X, \mathbf{x})$ that is positive in X is monotone (over all structures that interpret φ up to the free variables). This is easily shown by syntactic induction, if the claim is generalised to natural maps $F^{\mathfrak{A}}_\varphi\colon \mathcal{P}(A^r) \to \mathcal{P}(A^s)$ induced by formulae with not necessarily matching arities for their free first- and second-order variables.

Exercise 7.1.5 Prove that positivity implies monotonicity for FO definable maps.

The following is a very special case of a more general statement which is of great value in much more general settings.[2]

Lemma 7.1.6 *(Knaster–Tarski) Any monotone operator $F\colon \mathcal{P}(D) \to \mathcal{P}(D)$ has unique $\subseteq$-minimal and $\subseteq$-maximal fixpoints, called the least and greatest fixpoints of F, denoted $\mu(F)$ and $\nu(F)$, respectively. Moreover (over finite domains D of size $|D| = n$), these fixpoint are reached within n steps as the limits of the monotone (increasing, respectively decreasing) sequences*

$$\begin{array}{rl} (F^i(\emptyset))_{i\in\mathbb{N}} \longrightarrow \quad F^n(\emptyset) = F^{n+1}(\emptyset) = \bigcup_{i\in\mathbb{N}} F^i(\emptyset) = \mu(F) & \textit{the least fixpoint of } F. \\ (F^i(D))_{i\in\mathbb{N}} \longrightarrow F^n(D) = F^{n+1}(D) = \bigcap_{i\in\mathbb{N}} F^i(D) = \nu(F) & \textit{the greatest fixpoint of } F. \end{array}$$

[1] Here we appeal to official FO syntax comprising the boolean connectives $\neg, \wedge, \vee$, but not $\rightarrow$ or $\leftrightarrow$. The latter are regarded as abbreviations, whose elimination does introduce extra negations, as in $\varphi_1 \rightarrow \varphi_2 \equiv \neg\varphi_1 \vee \varphi_2$.

[2] The concepts of monotonicity and least and greatest fixpoints make sense over (not necessarily finite) complete partial orderings; we here only use it for finite partial $\subseteq$ orderings. Also the inductive generation of least/greatest fixpoints generalises, but on infinite domains the fixpoint is typically reached only in a transfinite ordinal sequence of stages.

Proof We treat the case of the least fixpoint. By monotonicity of F, the sequence $(F^i(\emptyset))_{i\in\mathbb{N}}$ is monotone increasing: $F^0(\emptyset) = \emptyset \subseteq F(\emptyset) \subseteq F^2(\emptyset) \subseteq \cdots \subseteq D$.

As D is finite, this sequence must be eventually constant, i.e., reaches a fixpoint of F. If $|D| = n$, there can be at most n strict increases, and clearly $F^{i+1}(\emptyset) = F^i(\emptyset)$ implies $F^{i+m}(\emptyset) = F^i(\emptyset)$ for all m, whence certainly $F^n(\emptyset) = F^{n+1}(\emptyset)$ is a fixpoint.

Let $F(P) = P$ be any fixpoint of F. Then $\emptyset \subseteq P$ and monotonicity of F imply (by induction on i) that $F^i(\emptyset) \subseteq F^i(P) = P$ for all $i \in \mathbb{N}$. Therefore the fixpoint $F^n(\emptyset)$ is contained in P, which implies that it is *the* $\subseteq$-minimal fixpoint of F. □

Exercise 7.1.7 Give examples of monotone operators that have just one, exactly two, more than two fixpoints, respectively.

Exercise 7.1.8 Let $F\colon \mathcal{P}(D) \to \mathcal{P}(D)$ be monotone. Show that the *dual operator*

$$\hat{F}\colon P \longmapsto \overline{F(\overline{P})} \quad \text{(complementation before and after } F\text{)}$$

is also monotone, and that the greatest fixpoint of F is the complement of the least fixpoint of $\hat{F}$ and vice versa: $\overline{\mu(F)} = \nu(\hat{F})$ and $\overline{\nu(F)} = \mu(\hat{F})$.

7.2 Least and Inductive Fixpoint Logics

Fixpoint logics enrich the syntax and semantics of FO by constructs that provide closure under certain fixpoint constructs. The most important one is *least fixpoint logic* which is a smallest well-behaved logic extending FO in which the least and greatest fixpoints of positively definable operators (which are monotone) are definable.

7.2.1 Least fixpoint logic LFP

Definition 7.2.1 The syntax of least fixpoint logic LFP(τ) is the extension of FO syntax with second-order variables (of any arity) by closure under μ and ν:

If $\varphi(X, \mathbf{Z}, \mathbf{x}, \mathbf{z}) \in \mathrm{LFP}(\tau)$ with free variables as indicated is positive in X, X of arity r and $\mathbf{x} = (x_1, \ldots, x_r)$ (pairwise distinct), then the following are also formulae of LFP(τ):

$$\psi_1(\mathbf{Z}, \mathbf{z}, \mathbf{x}) = \mu_{X,\mathbf{x}}\varphi \quad \text{and} \quad \psi_2(\mathbf{Z}, \mathbf{z}, \mathbf{x}) = \mu_{X,\mathbf{x}}\varphi,$$

with free variables as indicated. Here ψ_i is positive in Z if φ is positive in Z.

The semantics of ψ_i in τ-structures $\mathfrak{A}$ with assignments $\mathbf{R}, \mathbf{c}$ to the parameters $\mathbf{Z}, \mathbf{z}$, and for assignment $\mathbf{a}$ to $\mathbf{x}$, is given by

$$\begin{array}{lll} \mathfrak{A} \models (\mu_{X,\mathbf{x}}\varphi)[\mathbf{R}, \mathbf{c}, \mathbf{a}] & \text{iff} & \mathbf{a} \in \mu\big(F_\varphi^{\mathfrak{A},\mathbf{R},\mathbf{c}}\big) \\ \mathfrak{A} \models (\nu_{X,\mathbf{x}}\varphi)[\mathbf{R}, \mathbf{c}, \mathbf{a}] & \text{iff} & \mathbf{a} \in \nu\big(F_\varphi^{\mathfrak{A},\mathbf{R},\mathbf{c}}\big) \end{array}$$

where $F_\varphi^{\mathfrak{A},\mathbf{R},\mathbf{c}}$ is the monotone operator $F_\varphi^{\mathfrak{A},\mathbf{R},\mathbf{c}}\colon P \longmapsto \{\mathbf{a} \in A^r \colon \mathfrak{A} \models \varphi[P, \mathbf{R}, \mathbf{c}, \mathbf{a}]\}$ on $\mathcal{P}(A^r)$. In other words, the formulae ψ_i define the least and greatest fixpoints of the monotone operator defined by φ in terms of X and $\mathbf{x}$ (relative to fixed assignments to the other parameters).

Example 7.2.2 Consider, for a binary relation E, the FO-formulae $\varphi_1(X, x_1, x_2) = Ex_1x_2 \vee \exists y(Ex_1y \wedge Xyx_2)$ and $\varphi_2(Y, y) = \forall z(Eyz \to Yz)$, which are positive in X and Y, respectively. The inductive generation of the least fixpoint defined by $(\mu\varphi_1)(x_1, x_2)$

has stages $X^0 = \emptyset$, $X^1 = E$, $X^2 = E \cup E \circ E$, etc. The least fixpoint reached is the transitive closure of E.

The least fixpoint defined by $\mu\varphi_2$ is the set of all those elements from which there is no infinite E-path ($\mu\varphi_2$ defines the well-foundedness query).

Exercise 7.2.3 Consider a pair of two FO-formulae, $\varphi_1(X_1, X_2, \mathbf{x}_1)$ and $\varphi_2(X_1, X_2, \mathbf{x}_2)$, both positive in both X_i, with matching arities r_i between X_i and $\mathbf{x}_i$.

(a) Show that the simultaneous iteration based on simultaneous updates for (X_1, X_2) according to (φ_1, φ_2) converges to the least fixpoint of the monotone operator

$$\begin{array}{rcl} F\colon \mathcal{P}(A^{r_1}) \times \mathcal{P}(A^{r_2}) & \longrightarrow & \mathcal{P}(A^{r_1}) \times \mathcal{P}(A^{r_2}) \\ (P_1, P_2) & \longmapsto & (P_1', P_2') \\ & & \text{where } P_i' = \{\mathbf{a} \in A^{r_i} \colon \mathfrak{A} \models \varphi_i[P_1, P_2, \mathbf{a}]\}. \end{array}$$

(Monotonicity is w.r.t. the natural partial order of componentwise $\subseteq$).

(b) Show that the simultaneous least fixpoint from (a) is definable in LFP using a suitable simulation that encodes X_1 and X_2 in one relation (Hint: one could try to use $X_1 \times X_2$ but that causes some problem if one of these stays empty for some stages; instead, one may use auxiliary first-order parameters as tags, at least in structures with at least two distinct elements to instantiate these).

Remark: simultaneous (least or greatest) fixpoints of systems are generally reducible to ordinary fixpoints, albeit at the expense of increased arities; one can use this, e.g., in showing that DATALOG queries, whose semantics is precisely given through least fixpoints of systems over negation-free existential FO, are LFP definable.

Exercise 7.2.4 (compare Exercise 6.2.1) For fixed $k \geqslant 0$ provide an LFP($<$) formula $\varphi(\mathbf{x}, \mathbf{y}, \mathbf{z})$ that defines the graph of addition in restriction to standard structures $(\{0, \ldots, n-1\}, <)$ w.r.t. to number representations to base n. I.e., for all n and $a, b, c < n^k$ we want $(n, <) \models \varphi[\mathbf{a}, \mathbf{b}, \mathbf{c}]$ iff $a + b = c$.

Lemma 7.2.5 LFP *formulae can be evaluated in polynomial time over finite structures (polynomial time model checking for* LFP*).*

Proof Arguing by syntactic induction on LFP formulae $\varphi(\mathbf{Z}, \mathbf{z})$ with free variables as indicated, we want to show that the following boolean query Q_φ on $\tau \cup \{\mathbf{Z}\}$ structures with parameters $\mathbf{c}$ (as assignments for $\mathbf{z}$) is in P:

$$Q_\varphi = \big\{(\mathfrak{A}, \mathbf{R}, \mathbf{c}) \colon \mathfrak{A} \models \varphi[\mathbf{R}, \mathbf{c}]\big\}.$$

We treat the case of μ-application. Let $\psi = \psi(\mathbf{Z}, \mathbf{z}, \mathbf{x}) = \mu_{X,\mathbf{x}}\varphi(X, \mathbf{Z}, \mathbf{z}, \mathbf{x})$. By the inductive hypothesis, Q_φ is in P. Let r be the arity of X and $\mathbf{x}$. In order to evaluate Q_ψ on an input $(\mathfrak{A}, \mathbf{R}, \mathbf{c}, \mathbf{a})$ with $|\mathfrak{A}| = n$, successively compute the stages (X_i) of the inductive generation of the least fixpoint. $X_0 = \emptyset$ is trivial; inductively,

$$X_{i+1} = \big\{\mathbf{a}' \in A^r \colon (\mathfrak{A}, X_i, \mathbf{Z}, \mathbf{c}, \mathbf{a}') \in Q_\varphi\big\}$$

can be computed by evaluating Q_φ on each one of the n^r candidate tuples $\mathbf{a}'$ successively. The iteration can be terminated as soon as we find $\mathbf{a} \in X_i$ (then $(\mathfrak{A}, \mathbf{R}, \mathbf{c}, \mathbf{a}) \in Q_\psi$) or $X_{i+1} = X_i$ but $\mathbf{a} \notin X_i$ (then $(\mathfrak{A}, \mathbf{R}, \mathbf{c}, \mathbf{a}) \notin Q_\psi$). As the fixpoint is guaranteed to complete within n^r iterations, each of which takes polynomial time, the overall procedure is again polynomially time bounded. $\square$

Exercise 7.2.6 Write polynomially bounded For-loop or While-loop relational program for the evaluation of the least or greatest fixpoints of an operator F, which is itself computed by a black box program.

Exercise 7.2.7 On the relationship between LFP and SO:

(a) Give a translation from LFP(τ) formulae into logically equivalent SO(τ) formulae, based on the second-order definition of least and greatest fixpoints.

(b) For an LFP(τ) formula $\psi = \mu_{X,\mathbf{x}}\varphi(X, \mathbf{x})$ with $\varphi \in$ FO, find an $\exists$-SO(τ) formula $\hat{\psi}$ such that $\psi \equiv_{\mathsf{FIN}} \hat{\psi}$.

Hint: for (b) look for an FO formula that describes an encoding of the stages of the inductive generation of the fixpoint by means of auxiliary relations (which can then be quantified existentially in $\exists$-SO).

7.2.2 Capturing Ptime on ordered structures

The following central result in descriptive complexity is due to Immerman and Vardi, independently. It provides a logical characterisation of the complexity class P (deterministic polynomial time) in terms of the expressive power of LFP (FO plus least fixpoint recursion) *over linearly ordered finite structures.* The restriction to linearly ordered input structures seems

(i) technically necessary in order to enable sufficient coding machinery on the structural side.

(ii) unproblematic from a more conventional computational point of view, as actual computation always works with ordered input (representations).

(iii) unsatisfactory from a model theoretic (and database query language) point of view, as we do *not* obtain a logical language for all polynomial time boolean queries also over unordered structures that would guarantee semantic independence of input representation.

Point (iii) concerns a semantic safety requirement for queries put to not intrinsically ordered data: the answer needs to be independent of the (incidental) ordering that underlies the input presentation. In fact, the problem of whether there is a logic that captures P also over not necessarily ordered structures remains a major challenge in descriptive complexity. Note that in Fagin's characterisation of NP (which is good also over structures without order) we could just quantify over all possible orderings. This trick is not available at the level of P as there are $n!$ many orderings to consider.

To make the ordering explicit, consider finite relational vocabularies $\tau_< = \tau \cup \{<\}$, and the class

$$\mathsf{FORD}(\tau_<) := \big\{\mathfrak{A} \in \mathsf{FIN}(\tau_<)\colon\ <^{\mathfrak{A}} \text{ a linear ordering of the domain } \big\}.$$

The input representation $\langle\mathfrak{A}\rangle$ is then in one-to-one correspondence with *the* unique realisation of the isomorphism type of $\mathfrak{A}$ in $\mathsf{STAN}(\tau)$ for which the inner ordering (corresponding to $<^{\mathfrak{A}}$) is the natural ordering on the domain $\{0, \ldots, n\}$.

Theorem 7.2.8 (Immerman, Vardi)
For any class $Q \subseteq \mathsf{FORD}(\tau_<)$, that is closed under isomorphism, t.f.a.e.:

(i) *Q is in P.*

(ii) *Q is* LFP *definable: $Q = \mathsf{FMOD}(\psi)$ for some $\psi \in$ LFP($\tau_<$).*

Proof One direction is settled by Lemma 7.2.5. For the other direction, suppose Q is decided by the polynomially time bounded DTM $\mathcal{M}$ which terminates within n^k steps on input $\langle\mathfrak{A}\rangle$ with $|\mathfrak{A}| = n$ (for all sufficiently large n). We only consider $\mathfrak{A} \in \mathsf{FORD}(\tau_<)$ (which is FO definable within $\mathsf{FIN}(\tau_<)$).

We want to generate a relational description of the first $n^k - 1$ steps of the computation of $\mathcal{M}$ on input $\langle\mathfrak{A}\rangle$ as a least fixed point over $\mathfrak{A}$. Since $\mathfrak{A}$ is linearly ordered, we may use a numerical encoding of configurations similar to the one used for Fagin's theorem. The key difference: here we do not guess one possible computation path but rather describe the unique computation path.

Let the sate set of $\mathcal{M}$ be $s = \{0, \ldots, s-1\}$, and the set of tape symbols identified with $r = \{0, \ldots, r-1\}$ and assume $n \geqslant s, r$. In order to make do with just one single relation for the description of the computation we recombine the graphs of the functions $t \mapsto q_t$ (control state), $t \mapsto \ell_t$ (head position) and $t \mapsto \rho_t \colon n^k \to r$ (tape content) into the graph of one combined function (for $t \mapsto C_t$)

$$C = \big\{(\mathbf{t}, q_t, \boldsymbol{\ell}_t, \mathbf{i}, \rho_t(i)) \colon t, i < n^k\big\} \subseteq A^k \times A \times A^k \times A^k \times A = A^{3k+2}.$$

We want to generate the relation C over $\mathfrak{A}$ as a least fixpoint of a positive FO definable operation $F_\varphi \colon \mathcal{P}(A^{3k+2}) \to \mathcal{P}(A^{3k+2})$ whose stages $X_i = F_\varphi^i(\emptyset)$ precisely correspond to the initial segments of the computation:

$$X_i := C \cap \big(\{\mathbf{t} \in A^k \colon t < i\} \times A^{2k+2}\big).$$

It remains to provide an FO formulae $\varphi(X, \mathbf{x})$ which induces the desired operation F_φ. Here X is a second-order variable of arity $3k+2$ and $\mathbf{x}$ a matching tuple of distinct first-order variables. For better readability we write these first-order variables to suggest their intended instantiations as $\varphi(X, \mathbf{t}, q, \boldsymbol{\ell}, \mathbf{i}, b)$.

Assuming that X_i is as desired, we want $X_{i+1} = \{(\mathbf{t}, q, \boldsymbol{\ell}, \mathbf{i}, b) \colon \varphi(X_i, \mathbf{t}, q, \boldsymbol{\ell}, \mathbf{i}, b)\}$ to consist of X_i together with all those tuples $(\mathbf{t}, q, \boldsymbol{\ell}, \mathbf{i}, b)$ where $\mathbf{t}$ represents the immediate successor $t = t_{\rm prev} + 1$ of some $t_{\rm prev} < i$ represented by some $\mathbf{t}_{\rm prev}$, that make up the correct description of the successor configuration of a configuration described by $X_i \cap ((\{\mathbf{t}_{\rm prev}\} \times (A^2 k + 2))$.[3]

For this we use a formula $\varphi_0(\mathbf{t}, q, \boldsymbol{\ell}, \mathbf{i}, b)$ that defines the correct description of the initial configuration and then put

$$\begin{aligned} \varphi(X, \mathbf{t}, q, \boldsymbol{\ell}, \mathbf{i}, b) = \;& \varphi_0(\mathbf{t}, q, \boldsymbol{\ell}, \mathbf{i}, b) \vee \\ & \exists \mathbf{t}_{\rm prev} \exists q_{\rm prev} \exists \boldsymbol{\ell}_{\rm prev} \exists b_{\rm prev} \left[\begin{array}{l} \quad X \mathbf{t}_{\rm prev} q_{\rm prev} \boldsymbol{\ell}_{\rm prev} \mathbf{i}\, b_{\rm prev} \\ \wedge \text{“}t = t_{\rm prev} + 1\text{”} \\ \wedge\, \xi(X, q, \boldsymbol{\ell}, \mathbf{i}, b; \mathbf{t}_{\rm prev}, q_{\rm prev}, \boldsymbol{\ell}_{\rm prev}) \end{array} \right], \end{aligned}$$

where the formula ξ serves to select just those $(q, \boldsymbol{\ell}, \mathbf{i}, b)$ that belong in the correct description of the successor configuration of the configuration described by

$$X \cap \big(\{\mathbf{t}_{\rm prev}\} \times A^{2k+2}\big) = X \mathbf{t}_{\rm prev}\_.$$

More specifically, ξ is a disjunction of clauses of the following form, one for each possible transition $\delta(m, j) = (m', d, j')$ of $\mathcal{M}$:

$$\exists z \left[\begin{array}{l} X \mathbf{t}_{\rm prev} q_{\rm prev} \boldsymbol{\ell}_{\rm prev} \boldsymbol{\ell}_{\rm prev} z \wedge \text{“}q_{\rm prev} = m\text{”} \wedge \text{“}z = j\text{”} \\ \wedge \text{“}q = m'\text{”} \wedge \text{“}\ell = \ell + d\text{”} \\ \wedge \big[\big(\mathbf{i} \neq \boldsymbol{\ell}_{\rm prev} \wedge X \mathbf{t}_{\rm prev} q_{\rm prev} \boldsymbol{\ell}_{\rm prev} \boldsymbol{\ell} b\big) \vee \big(\mathbf{i} = \boldsymbol{\ell}_{\rm prev} \wedge \text{“}b = j'\text{”}\big)\big] \end{array} \right].$$

[3] It would seem to suffice to add to X_i just those tuples that describe the i-th configuration, but one cannot extract the maximal $\mathbf{t}'$-value represented in X_i by means of a positive formula (why?).

Here the first line identifies the state and tape symbol read by the head through inspection of suitable entries in $X\mathbf{t}_{\text{prev}\_}$; the second line sets the values for state and head position according to δ; the third line forces the tape content to be transcribed and modified accordingly. Numerical equalities like "$z = j$" are FO definable conditions for fixed values j, e.g., expressible as "z has precisely j many predecessors w.r.t. $<$".

The fixpoint formula

$$\psi_0(\mathbf{t}, q, \boldsymbol{\ell}, \mathbf{i}, b) := \mu_{X,\mathbf{x}}\varphi(X, \mathbf{t}, q, \boldsymbol{\ell}, \mathbf{i}, b)$$

defines the query that, over $\mathfrak{A} \in \mathsf{FORD}(\tau_<)$ returns the relational representation $C \subseteq A^{3k+2}$ of the computation of $\mathcal{M}$ on input $\langle\mathfrak{A}\rangle$. The desired LFP sentence ψ that defines the boolean query "acceptance by $\mathcal{M}$" over $\mathsf{FORD}(\tau_<)$ is then easily obtained in the form

$$\exists\mathbf{t}\exists q\exists\boldsymbol{\ell}\exists\mathbf{i}\exists b\big(\psi_0(\mathbf{t}, q, \boldsymbol{\ell}, \mathbf{i}, b) \wedge \text{“}q = q^+\text{”}\big).$$

$\square$

The proof also provides a normal form for LFP over finite linearly ordered structures.

Corollary 7.2.9 *Over* $\mathsf{FORD}(\tau_<)$, *every* LFP *sentence is equivalent to one that has only one μ-application, or: every polynomial time decidable boolean query on linearly ordered finite structures is* FO *definable in terms of the least fixed point of some* FO*-definable operation. (There are stronger normal forms for* LFP, *over all finite structures.)*

7.2.3 Inductive fixpoint logic IFP

Inductive fixpoint logic extends FO by *inductive fixpoints* rather than least or greatest fixpoints. The inductive fixpoint of an operation $F\colon \mathcal{P}(D) \to \mathcal{P}(D)$ is the limit of the increasing sequence of stages of the induced *inductive* operator $F^+\colon P \mapsto P \cup F(P)$:

$$\emptyset \subseteq F^+(\emptyset) \subseteq (F^+)^2(\emptyset) \subseteq \quad \cdots \quad \subseteq (F^+)^i(\emptyset) = (F^+)^{i+1}(\emptyset) \; = \mathrm{IFP}(F).$$

The limit is denoted $\mathrm{IFP}(F)$ and, by abuse of terminology, called the *inductive fixpoint* of F, even though it need not be a fixpoint of F (but only of F^+). If F itself is inductive, i.e., if the sequence of the $F^i(\emptyset)$ is increasing by itself, then $(F^+)^i(\emptyset) = F(\emptyset)$ for all i, and the limit is indeed a fixpoint of F. (This, again, is in particular the case if F is monotone, in which case $\mathrm{IFP}(F) = \mu(F)$.)

Clearly, for any F over a finite domain D, $\mathrm{IFP}(F)$ is reached within $|D|$ many iterations of F^+, for cardinality reasons.

Definition 7.2.10 The syntax of inductive fixpoint logic $\mathrm{IFP}(\tau)$ is the extension of FO syntax with second-order variables (of any arity) by closure under an IFP operation:
For $\varphi(X, \mathbf{Z}, \mathbf{x}, \mathbf{z}) \in \mathrm{IFP}(\tau)$ with free variables as indicated, X of arity r and $\mathbf{x} = (x_1, \ldots, x_r)$ (pairwise distinct), $\psi(\mathbf{Z}, \mathbf{z}, \mathbf{x}) = \mathrm{IFP}_{X,\mathbf{x}}\varphi$ is also a formula of $\mathrm{IFP}(\tau)$, with free variables as indicated.
The semantics of ψ in τ-structures $\mathfrak{A}$ with assignments $\mathbf{R}, \mathbf{c}$ to the parameters $\mathbf{Z}, \mathbf{z}$, and for assignment $\mathbf{a}$ to $\mathbf{x}$, is given by

$$\mathfrak{A} \models (\mathrm{IFP}_{X,\mathbf{x}}\varphi)[\mathbf{R}, \mathbf{c}, \mathbf{a}] \quad \text{iff} \quad \mathbf{a} \in \mathrm{IFP}(F_\varphi^{\mathfrak{A},\mathbf{R},\mathbf{c}}),$$

where $F_\varphi^{\mathfrak{A},\mathbf{R},\mathbf{c}}$ is the operator $F_\varphi^{\mathfrak{A},\mathbf{R},\mathbf{c}}\colon P \longmapsto \{\mathbf{a} \in A^r : \mathfrak{A} \models \varphi[P, \mathbf{R}, \mathbf{c}, \mathbf{a}]\}$ on $\mathcal{P}(A^r)$.

NB: One could, in the defining clause for the semantics of $\mathrm{IFP}_{X,\mathbf{x}}\varphi$, directly refer to the inductive operator $(F_\varphi)^+$ which is the same as $F_{\varphi'}$ for $\varphi'(X,\mathbf{x}) = X\mathbf{x} \vee \varphi(X,\mathbf{x})$.

Since for $\varphi(X,\mathbf{Z},\mathbf{z},\mathbf{x})$ that is positive in X we have $\mathrm{IFP}_{X,\mathbf{x}}\varphi \equiv \mu_{X,\mathbf{x}}\varphi$, we may regard $\mathrm{LFP}(\tau)$ as a sublogic of $\mathrm{IFP}(\tau)$.

Since IFP also has polynomial time model checking over finite structures, it follows that IFP captures P over $\mathsf{FORD}(\tau_<)$ just as LFP does.

Corollary 7.2.11 *Over linearly ordered finite structures,* IFP *and* LFP *have exactly the same expressive power.*

In fact, by a result of Gurevich and Shelah, which we state without proof, IFP and LFP are equally expressive over all (not necessarily ordered) finite structures; this result moreover even extends to all (not necessarily finite) structures, by a more recent result of Kreutzer.[4] These results are useful, because it is often much easier to formalise some inductive process in IFP an in LFP – without the necessity of making the process monotone and formalising it in a positive formula. We shall appeal to the Gurevich–Shelah result for this reason later.

Theorem 7.2.12 (Gurevich–Shelah; Kreutzer)
IFP *and* LFP *have the same expressive power.*

The proofs are based on the LFP-definability of relations that encode the stages of the inductive iteration sequence $(F^+_\varphi)^i(\emptyset)$.

7.3 Partial Fixpoint Logic

7.3.1 Partial fixpoints

Looking at arbitrary operations $F\colon \mathcal{P}(D) \rightarrow \mathcal{P}(D)$, one can "enforce" inductive behaviour by passage to $F^+\colon P \mapsto P \cup F(P)$ (as in IFP). Alternatively, we may iterate F itself on $\emptyset$ and associate with this iteration either its natural limit, if F is eventually constant on $\emptyset$, or a default value $\emptyset$ otherwise, i.e., if the sequence $(F^i(\emptyset))_{i\in\mathbb{N}}$ becomes non-trivially periodic. The *partial fixpoint* of F is defined in this way, "partial" because it may return $\emptyset$ as the default value when $\emptyset$ is not a fixpoint of F.

$$\mathrm{PFP}(F) := \begin{cases} F^{i+1}(\emptyset) = F^i(\emptyset) & \text{if such } i \text{ exists} \\ \emptyset & \text{otherwise.} \end{cases}$$

Definition 7.3.1 The syntax of partial fixpoint logic $\mathrm{PFP}(\tau)$ is the extension of FO syntax with second-order variables (of any arity) by closure under a PFP operation:
For $\varphi(X,\mathbf{Z},\mathbf{x},\mathbf{z}) \in \mathrm{PFP}(\tau)$ with free variables as indicated, X of arity r and $\mathbf{x} = (x_1,\ldots,x_r)$ (pairwise distinct), $\psi(\mathbf{Z},\mathbf{z},\mathbf{x}) = \mathrm{PFP}_{X,\mathbf{x}}\varphi$ is also a formula of $\mathrm{PFP}(\tau)$, with free variables as indicated.
The semantics of ψ in τ-structures $\mathfrak{A}$ with assignments $\mathbf{R},\mathbf{c}$ to the parameters $\mathbf{Z},\mathbf{z}$, and for assignment $\mathbf{a}$ to $\mathbf{x}$, is given by

$$\mathfrak{A} \models (\mathrm{PFP}_{X,\mathbf{x}}\varphi)[\mathbf{R},\mathbf{c},\mathbf{a}] \quad \text{iff} \quad \mathbf{a} \in \mathrm{PFP}(F^{\mathfrak{A},\mathbf{R},\mathbf{c}}_\varphi),$$

where $F^{\mathfrak{A},\mathbf{R},\mathbf{c}}_\varphi$ is the operator $F^{\mathfrak{A},\mathbf{R},\mathbf{c}}_\varphi\colon P \longmapsto \{\mathbf{a} \in A^r : \mathfrak{A} \models \varphi[P,\mathbf{R},\mathbf{c},\mathbf{a}]\}$ on $\mathcal{P}(A^r)$.

[4]The semantics of IFP over infinite structures is based on the transfinite inductive iteration of F^+_φ.

PFP is at least as expressive as LFP, or generalises LFP, since $\mathrm{PFP}(F) = \mu(F)$ for monotone F. It is also at least as expressive as IFP, since $\mathrm{IFP}(F) = \mathrm{PFP}(F^+)$.

Lemma 7.3.2 *Model checking for* PFP *over finite structures is in Pspace.*

Proof The evaluation of a partial fixpoint for a Pspace computable operation $F\colon \mathcal{P}(D) \to \mathcal{P}(D)$ over finite domain $|D| = n$ is again in Pspace. One merely needs to note that

$$\mathrm{PFP}(F) = \begin{cases} F^{2^n}(\emptyset) & \text{if } F^{2^n}(\emptyset) = F^{2^n+1}(\emptyset) \\ \emptyset & \text{otherwise.} \end{cases}$$

A (binary) counter for 2^n iterations can be implemented in space n. Note, however, that this procedure is only exponentially time bounded in general. $\Box$

7.3.2 Capturing Pspace on ordered structures

In analogy with the capturing result for P over ordered structures through LFP, we obtain a capturing result for Pspace over ordered structures through PFP.

Theorem 7.3.3 (Abiteboul–Vianu)
For any class $Q \subseteq \mathsf{FORD}(\tau_<)$, that is closed under isomorphism, t.f.a.e.:

(i) *Q is in Pspace.*
(ii) *Q is* PFP *definable within $\mathsf{FIN}(\tau_<)$: $Q = \mathsf{FMOD}(\psi)$ for some $\psi \in \mathrm{PFP}(\tau_<)$.*

Proof The proof of the crucial direction is similar to those given for the previous capturing results. Given an n^k space bounded DTM $\mathcal{M}$, otherwise of the same format as in the proof of Theorem 7.2.8 say, we now want to define a relational representation of its final configuration as a partial fixpoint of some FO formula.

For this we set up the underlying FO-formula $\varphi(X, \mathbf{x})$ such that the iterates $F_\varphi^{i+1}(\emptyset)$ over $\mathfrak{A}$ represent the i-th configuration C_i of $\mathcal{M}$ on input $\langle\mathfrak{A}\rangle$, for every i until termination. Note that we do not have to keep track of the time index explicitly.

Using a tuple of first-order variables $\mathbf{x} = (q, \boldsymbol{\ell}, \mathbf{i}, b)$ (to suggest the intended roles as representatives for state, head position, tape cell index and its contents) of arity $2k+2$ and matching X, we put

$$\varphi(X, q, \boldsymbol{\ell}, \mathbf{i}, b) := \big(\neg\exists\mathbf{x}X\mathbf{x} \wedge \varphi_0(q, \boldsymbol{\ell}, \mathbf{i}, b)\big) \vee \big(\exists\mathbf{x}X\mathbf{x} \wedge \xi(X, q, \boldsymbol{\ell}, \mathbf{i}, b)\big).$$

Here φ_0 defines the relational description of the initial configuration on input $\langle\mathfrak{A}\rangle$ over $\mathfrak{A}$; note that φ_0 is invoked precisely in the first iteration (when X is still empty), and thus provides the correct initialisation with $X_1 := C_0$).

The formula ξ collects the tuples $(q, \boldsymbol{\ell}, \mathbf{i}, b)$ that provide the description of the successor configuration of the configuration described by X, according to the transition function δ of $\mathcal{M}$, similar to the corresponding formula in the proof of Theorem 7.2.8. None of the further iterates will thus be empty.

$$\psi_0(\mathbf{x}) := \mathrm{PFP}_{X,\mathbf{x}}\varphi$$

is guaranteed to define over $\mathfrak{A}$ the non-empty relational description of the final configuration of $\mathcal{M}$ on input $\langle\mathfrak{A}\rangle$, because $\mathcal{M}$ does terminate. It follows that

$$\psi := \exists q \exists \boldsymbol{\ell} \exists \mathbf{i} \exists b \big(\psi_0(q, \boldsymbol{\ell}, \mathbf{i}, b) \wedge \text{“}q = q^+\text{”}\big)$$

defines acceptance by $\mathcal{M}$. $\Box$

Exercise 7.3.4 Fill in the details for the formula ξ in the proof above, in analogy with corresponding formalisation of the successor configuration in the proof of Theorem 7.2.8.

7.4 The Abiteboul–Vianu Theorem

It is not known whether P $\subsetneq$ Pspace: the inclusion is obvious, but strictness is one of the major open problems of computational complexity theory.

Since we can equate each side of this relationship with definability in a suitable fixpoint logic over FORD, we get the following equivalence:

$$\boxed{\mathrm{P} = \mathrm{Pspace}} \quad\Leftrightarrow\quad \boxed{\mathrm{LFP} \equiv \mathrm{PFP} \text{ over } \mathsf{FORD}}$$

In this section we outline the proof of a famous result by Abiteboul and Vianu, which allows us to remove the restriction to ordered structures in this equivalence.

$$\boxed{\mathrm{P} = \mathrm{Pspace}} \quad\Leftrightarrow\quad \boxed{\mathrm{LFP} \equiv \mathrm{PFP} \text{ over } \mathsf{FIN}}$$

Equal expressiveness between LFP and PFP over all finite structures is equivalent to the collapse of Pspace to P. In other words, Pspace = P if, and only if, the result of every PFP recursion can be equivalently obtained as the result of an LFP recursion, or if relational WHILE recursion is not more powerful than positive, monotone least fixpoint recursion in determining any property of finite structures.

Technically, this result involves a uniform reduction from fixpoint evaluations over a given not necessarily ordered finite structure $\mathfrak{A}$ to the evaluation of a variant of that fixpoint in some linearly ordered structure definable from $\mathfrak{A}$. This is achieved via a detour through infinitary finite variable logics and a simulation of the fixpoint evaluation over $\mathfrak{I}^k(\mathfrak{A})$, the k-variable invariant associated with $\mathfrak{A}$ from section 2.5.2 in Part I, for suitable k. See in particular Definition 2.5.16 and Proposition 2.5.17.

Our first step, therefore, is to embed the fixpoint logics into infinitary k-variable logics.

7.4.1 Fixpoint logics and finite variable logics

Definition 7.4.1 Infinitary k-variable logic FO^k_∞ is defined as an extension of k-variable first-order logic FO^k, augmenting the rules for formula formation in FO^k by allowing disjunctions and conjunctions over arbitrary (rather than just finite) sets of formulae. If $\Phi \subseteq \mathrm{FO}^k_\infty(\tau)$ is any set of formulae of the logic, the so are $\bigwedge \Phi$ and $\bigvee \Phi$. The semantics of these is the natural one:

$$\begin{array}{lll} \mathfrak{A}, \mathbf{a} \models \bigvee \Phi & \text{if} & \mathfrak{A}, \mathbf{a} \models \varphi \text{ for at least one } \varphi \in \Phi, \\ \mathfrak{A}, \mathbf{a} \models \bigwedge \Phi & \text{if} & \mathfrak{A}, \mathbf{a} \models \varphi \text{ for all } \varphi \in \Phi. \end{array}$$

Recall the analysis of the unbounded k-pebble game in section 2.5.2, which showed k-variable equivalence $\equiv^k$ to coincide with the equivalence defined through the unbounded k-pebble game, $\simeq^k_\infty$, over finite relational structures. The same analysis extends to show that $\simeq^k_\infty$ implies equivalence even at the level of FO^k_∞. Hence the unbounded k-pebble game may also be regarded as the Ehrenfeucht-Fraïssé game for infinitary k-variable logic (and *this* correspondence in actually good even over infinite structures).

In the following we always assume parameter tuples $\mathbf{a} \in A^k$, $\mathbf{b} \in B^k$ for the instantiation of the k variables that may be free in formulae of k-variable logic.

Lemma 7.4.2 *For finite structures* $\mathfrak{A}, \mathbf{a}$ *and* $\mathfrak{B}, \mathbf{b}$ *of the same finite relational type* τ, *t.f.a.e.:*

(i) $\mathfrak{A}, \mathbf{a} \equiv^k \mathfrak{B}, \mathbf{b}$*: for all* $\varphi(\mathbf{x}) \in \mathrm{FO}^k(\tau)$, $\mathfrak{A} \models \varphi[\mathbf{a}] \;\Leftrightarrow\; \mathfrak{B} \models \varphi[\mathbf{b}]$
(equivalence in FO^k*).*

(ii) $\mathfrak{A}, \mathbf{a} \simeq^k_\infty \mathfrak{B}, \mathbf{b}$
(equivalence w.r.t. G^k_∞*:* **II** *wins* $\mathsf{G}^k_\infty(\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b})$*).*

(iii) *for all* $\varphi(\mathbf{x}) \in \mathrm{FO}^k_\infty(\tau)$, $\mathfrak{A} \models \varphi[\mathbf{a}] \;\Leftrightarrow\; \mathfrak{B} \models \varphi[\mathbf{b}]$
(equivalence in FO^k_∞*).*

Proof We know (i) $\Rightarrow$ (ii) from section 2.5.2.
(iii) $\Rightarrow$ (i) is obvious, as $\mathrm{FO}^k \subseteq \mathrm{FO}^k_\infty$.

It suffices therefore to prove that $\neg$(iii) (inequivalence in FO^k_∞) gives **I** a winning strategy in $\mathsf{G}^k_\infty(\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b})$, and hence implies $\neg$(ii).

The claim about winning strategies for **I** is proved by induction on the (infinitary!) syntax of a formula $\varphi(\mathbf{x}) \in \mathrm{FO}^k_\infty$ that distinguishes between $\mathfrak{A}, \mathbf{a}$ ad $\mathfrak{B}, \mathbf{b}$. Assume, for instance, that $\mathfrak{A} \models \varphi[\mathbf{a}]$ while $\mathfrak{B} \not\models \varphi[\mathbf{b}]$, and, as our inductive hypothesis, that for all proper subformulae ψ of φ the claim is true (that inequivalence w.r.t. ψ gives **I** a winning strategy).

If $\varphi(\mathbf{x})$ is of the form $\neg\psi(\mathbf{x})$, then the inductive hypothesis for ψ works directly for φ.

If $\varphi(\mathbf{x}) = \exists x_j \psi(\mathbf{x})$, then there is an $a \in A$ such that $\mathfrak{A} \models \psi[\mathbf{a}\frac{a}{j}]$ while $\mathfrak{B} \models \neg\psi[\mathbf{b}\frac{b}{j}]$ for all $b \in B$. We advise **I** to play in $\mathfrak{A}$, move pebble j to a, and rely on the inductive hypothesis for ψ. The universal quantifier case is strictly analogous.

If $\varphi = \bigvee \Phi$, then there is some $\psi \in \Phi$ such that $\mathfrak{A} \models \psi[\mathbf{a}]$ while $\mathfrak{B} \models \neg\psi[\mathbf{b}]$. So **I** can use the strategy guaranteed by the distinguishing subformula ψ according to the inductive hypothesis. The case of an (infinite) conjunction is strictly analogous. $\Box$

What is the point of considering FO^k_∞ over finite structures then?
There are two answers:

- FO^k_∞ defines many classes of finite structures (queries) that are not FO definable. (See Exercise 7.4.3 below.)
- FO^k_∞ provides a natural way to define FO^k-types and arbitrary collections of FO^k-types over $\mathsf{FIN}(\tau)$ (unions of $\simeq^k_\infty$ classes).

Exercise 7.4.3 Show that the following classes of finite structures are definable in FO^k_∞ over FIN for suitable k and try to find the minimal k.

(a) $\tau = \{E\}$. The class of finite undirected graphs that are connected.

(b) $\tau = \{<\}$. For an arbitrary fixed subset $S \subseteq \mathbb{N}$: the class of finite linear orderings of length n for $n \in S$

Lemma 7.4.4 *For* $\varphi(X, \mathbf{x}) = \varphi(X, x_1, \dots, x_k) \in \mathrm{FO}^k_\infty(\tau \cup \{X\})$ *with* k*-ary* X*, let* F_φ *be the operation that is globally defined by* φ *as an operation on* k*-ary relations over structures* $\mathfrak{A} \in \mathsf{FIN}(\tau)$

$$\begin{array}{rcl} F^{\mathfrak{A}}_\varphi \colon \mathcal{P}(A^k) & \longrightarrow & \mathcal{P}(A^k) \\ P & \longmapsto & F^{\mathfrak{A}}_\varphi(P) = \{\mathbf{a} \in A^k \colon \mathfrak{A} \models \varphi[P, \mathbf{a}]\}. \end{array}$$

Then $\mu(F_\varphi)$, $\mathrm{IFP}(F_\varphi)$ *and* $\mathrm{PFP}(F_\varphi)$ *are globally definable in* $\mathrm{FO}^k_\infty(\tau)$.

Proof We first show by induction that the finite stages in the iteration of F_φ on $\emptyset$, $F^i_\varphi(\emptyset)$ are uniformly definable by suitable formulae $\varphi^i(\mathbf{x}) \in \mathrm{FO}^k_\infty$.

For $i = 0$, $\varphi^0(\mathbf{x}) = \neg\, x_1 = x_1$ is as desired.

Suppose $\varphi^i(\mathbf{x})$ is given. We want to obtain $\varphi^{i+1}(\mathbf{x})$ by a process of substituting $\varphi^i(\mathbf{y})$ for every atom $X\mathbf{y}$ inside φ. However, $\mathbf{y}$ can be any tuple of not necessarily distinct variables from $\{x_1, \ldots, x_k\}$. This can be dealt with as follows. If $\mathbf{y} = (x_{\pi(1)}, \ldots, x_{\pi(k)})$ for some permutation π of $\{1, \ldots, k\}$, we just apply π to (the indices of) all variables (free or bound) in $\varphi^i(\mathbf{x})$, in oder to obtain a formula $\varphi^i(\mathbf{y})$ that is as desired. The case where $\mathbf{y}$ has multiple occurrences of the same variable symbol, e.g., $\mathbf{y} = (x_1, x_1, x_3, \ldots, x_k)$, reduces to the former case through quantification and equality binding of those variable symbols that do not appear as components of $\mathbf{y}$, e.g., $Xx_1x_1x_3 \ldots x_k \equiv \exists x_2(x_2 = x_1 \wedge Xx_1x_2x_3 \ldots x_k)$.

Then $\mathrm{PFP}(F_\varphi)$ is globally defined by

$$\bigvee_{i\in\mathbb{N}}\big(\varphi^i(\mathbf{x}) \wedge \forall\mathbf{x}(\varphi^i(\mathbf{x}) \leftrightarrow \varphi^{i+1}(\mathbf{x}))\big).$$

Similarly $\mu(F_\varphi)$, for monotone φ, is globally defined by $\bigvee_{i\in\mathbb{N}} \varphi^i(\mathbf{x})$. For IFP we may similarly first obtain global definitions of the stages w.r.t. F^+_φ, which are the same as the stages of F_{φ^+} for $\varphi^+(X, \mathbf{x}) = X\mathbf{x} \vee \varphi(X, \mathbf{x})$. ☐

The same argument for definability of the stages $F^i_\varphi(\emptyset)$ goes through for $\varphi(X, \mathbf{z}, \mathbf{x}) \in \mathrm{FO}^k_\infty$ (with X of arity $r \leqslant k$, $\mathbf{x} = (x_{i_1}, \ldots, x_{i_r})$ distinct and disjoint from the parameters $\mathbf{z}$), if the variables $\mathbf{z}$ do not have bound occurrences in φ. (See Exercise 7.4.9 below for the necessity of some such restriction). We may then work with permutations of the variable tuple $\mathbf{x}$ (fixing $\mathbf{z}$) and equality bindings (possibly involving parameters $\mathbf{z}$).

Corollary 7.4.5 *Let $\varphi(X, \mathbf{z}, \mathbf{x}) \in \mathrm{FO}^k_\infty$ such that no variable in $\mathbf{z}$ occours bound in φ, X and $\mathbf{x}$ of matching arities suitable for corresponding fixpoints. Then these fixpoints (with parameters) are globally definable in FO^k_∞.*

Definition 7.4.6 Let PFP^k consist of the closure of FO^k under the formula formation rules of FO^k and PFP applications to formulae of the form $\mathrm{PFP}_{X,\mathbf{x}}\varphi(X, \mathbf{Z}, \mathbf{z}, \mathbf{x})$ such that the variables in $\mathbf{z}$ do not have bound occurrences in φ. Fragments LFP^k and IFP^k are similarly defined.

Every LFP, IFP or PFP formula can be transformed into an equivalent formula in LFP^k, IFP^k or PFP^k for some k, by a renaming of bound variables where necessary.

Corollary 7.4.7 *Every formula φ of $\mathrm{LFP}^k(\tau)$, $\mathrm{IFP}^k(\tau)$ or $\mathrm{PFP}^k(\tau)$ can be translated into a formula of $\mathrm{FO}^k_\infty(\tau)$ that is equivalent to φ over $\mathsf{FIN}(\tau)$. It follows that every formula of $\mathrm{LFP}(\tau)$, $\mathrm{IFP}(\tau)$ or $\mathrm{PFP}(\tau)$ is equivalent over $\mathsf{FIN}(\tau)$ to a formula of $\mathrm{FO}^k_\infty(\tau)$ for suitable $k \in \mathbb{N}$.*

Exercise 7.4.8 Show that for least and inductive fixpoints, the fixpoint w.r.t. X and $\mathbf{x}$ for the operator defined by $\varphi(X, \mathbf{z}, \mathbf{x})$ with parameters $\mathbf{z}$ is first-order inter-definable with a parameter-free fixpoint. Consider the fixpoint w.r.t. $Y$and $\mathbf{y}$ for the operator defined by $\hat{\varphi}(Y, \mathbf{y})$ where $\mathbf{y} = \mathbf{z}\mathbf{x}$, Y of matching arity, and $\hat{\varphi}(Y, \mathbf{z}\mathbf{x}) = \varphi(Y\mathbf{z}\_, \mathbf{z}, \mathbf{x})$.
Why does this not work directly for PFP?

Exercise 7.4.9 A fixpoint application to a formula $\varphi \in \mathrm{FO}^k$ which has first-order parameters can lead outside FO^k. Here is a simple example.

$$\varphi(X, x_1, x_2) := x_2 = x_1 \vee \exists x_1(Ex_1x_2 \wedge Xx_1) \in \mathrm{FO}^2$$

uses the free occurrence of x_1 as a parameter if we consider $\psi(x_1, x_2) := \mu_{X,x_2}\varphi \in \mathrm{LFP}(\{E\})$.

Check that $\psi(x_1, x_2)$ defines the global relation of reachability, which is not definable in FO^2_∞. Give a definition of the same query in FO^3_∞.

7.4.2 Simulating fixpoints over the invariants

The last corollary implies in particular that for every formula $\varphi(\mathbf{x}) \in \mathrm{PFP}$ there is some k such that $\varphi(\mathbf{x})$ is preserved under $\simeq^k_\infty$ in the sense that for all $\mathfrak{A}, \mathbf{a} \simeq^k_\infty \mathfrak{B}, \mathbf{b}$ we must have $\mathfrak{A} \models \varphi[\mathbf{a}] \;\Rightarrow\; \mathfrak{B} \models \varphi[\mathbf{b}]$. As the equivalence class of $(\mathfrak{A}, \mathbf{a})$ w.r.t. $\simeq^k_\infty$ is encoded in the k-variable invariant $\mathfrak{I}^k(\mathfrak{A}, \mathbf{a})$ from section 2.5.2, whether or not $\mathfrak{A} \models \varphi[\mathbf{a}]$ can be determined in terms of $\mathfrak{I}^k(\mathfrak{A}, \mathbf{a})$. In particular, for every PFP definable boolean query $Q \subseteq \mathsf{FIN}(\tau)$ there is a value of k and a corresponding query

$$\hat{Q} := \{\mathfrak{I}^k(\mathfrak{A}) \colon \mathfrak{A} \in Q\}$$

such that the function $\mathfrak{I}^k \colon \mathsf{FIN}(\tau) \to \mathsf{FIN}(\tau^k)$, which maps $\mathfrak{A}$ to its k-variable invariant $\mathfrak{I}^k(\mathfrak{A})$, is a polynomial time reduction from Q to $\hat{Q}$. The same applies to LFP or IFP definable queries and a similar statement also covers non-boolean queries. (Compare Lemma 7.4.10 below.)

Here τ^k is the relational vocabulary used for the k-variable invariant of τ-structures, $\tau^k = \{\leqslant\} \cup \{P_\theta \colon \theta \in \mathrm{FO}^k(\tau), \mathrm{qr}(\theta) = 0\} \cup \{E_j \colon 1 \leqslant j \leqslant k\}$.

We now want to see that, moreover, a $\mathrm{PFP}(\tau)$/$\mathrm{IFP}(\tau)$/$\mathrm{LFP}(\tau)$ formula can be evaluated in terms of the k-variable invariants, for a suitable value of k, using a matching $\mathrm{PFP}(\tau^k)$/$\mathrm{IFP}(\tau^k)$/$\mathrm{LFP}(\tau^k)$ formula there.

Recall the format of the k-variable invariants for $\mathfrak{A} \in \mathsf{FIN}(\tau), \mathbf{a} \in A^k$:

$$\mathfrak{I}^k(\mathfrak{A}, \mathbf{a}) = \big(A^k/{\simeq^k_\infty}, \leqslant, (P_\theta), (E_j), [\mathbf{a}]_{\simeq^k_\infty}\big).$$

The P_θ for quantifier-free $\theta \in \mathrm{FO}^k(\tau)$ indicate which quantifier-free formulae are satisfied in each $\simeq^k_\infty$ class of k-tuples represented in $\mathfrak{A}$; the E_j are equivalence relations linking $\simeq^k_\infty$ classes between which one can switch by changing the j-th components (moves with pebble j). We now write just $[\mathbf{a}]$ for the $\simeq^k_\infty$ equivalence class of $\mathbf{a}$ in A^k. If the distinguished parameter tuple $\mathbf{a}$ is irrelevant, we write $\mathfrak{I}^k(\mathfrak{A})$.

We want to give a translation

$$\begin{array}{rcl} \hat{\ } : \mathrm{PFP}^k(\tau) & \longrightarrow & \mathrm{PFP}(\tau^k) \\ \varphi(\mathbf{Z}, x_1, \ldots, x_k) & \longmapsto & \hat{\varphi}(\hat{\mathbf{Z}}, x), \end{array}$$

where $\mathbf{Z} = (Z_1, \ldots, Z_m)$, Z_s an r_s-ary second-order variable with $r_s \leqslant k$, and $\hat{\mathbf{Z}} = (\hat{Z}_1, \ldots, \hat{Z}_m)$ consists of monadic second-order variables $\hat{Z}_s$. The semantic condition is that for all $\mathfrak{A} \in \mathsf{FIN}(\tau)$, and for all assignments to the Z_s that are closed w.r.t. $\simeq^k_\infty$ (unions of $\simeq^k_\infty$ classes over A^{r_s}) and matching $\hat{Z}_s := \{[\mathbf{a}] \colon \mathbf{a} \in Z_s \times A^{k-r_s}\}$ we have:

$$\mathfrak{A} \models \varphi[\mathbf{Z}, \mathbf{a}] \quad \text{iff} \quad \mathfrak{I}^k(\mathfrak{A}) \models \hat{\varphi}(\hat{\mathbf{Z}}, [\mathbf{a}]).$$

The translation $\hat{\varphi}$ is given by induction on $\varphi(\mathbf{Z}, \mathbf{x}) = \varphi(\mathbf{Z}, x_1, \ldots, x_k)$. In order to describe substitutions of variable tuples $\mathbf{y} \in \{x_1, \ldots, x_k\}^r$, we first provide auxiliary formulae $\eta_\sigma \in \mathrm{IFP}(\tau^k)$ for any (partial) function $\sigma \colon \{1, \ldots, k\} \to \{1, \ldots, k\}$. $\eta_\sigma(x, y)$ is to define the binary relation

$$\{([\mathbf{a}], [\mathbf{b}]) \colon b_i = a_{\sigma(i)}\}$$

over $\mathfrak{I}^k(\mathfrak{A})$. Such formulae η_σ may be extracted from (the simulation of) the inductive process that successively refines $A^k/\simeq^k_m$ to $A^k/\simeq^k_{m+1}$ with limit $A^k/\simeq^k_\infty$. This simulation passes through representations of the rougher levels of $\simeq^k_m$ as equivalence relations on the universe $A^k/\simeq^k_\infty$ of $\mathfrak{I}^k(\mathfrak{A})$. The stages of corresponding approximations $\{([\mathbf{a}],[\mathbf{b}]) \in (A^k/\simeq^k_m)^2 \colon b_i = a_{\sigma(i)}\}$ reach the desired limit $\{([\mathbf{a}],[\mathbf{b}]) \in (A^k/\simeq^k_\infty)^2 \colon b_i = a_{\sigma i}\}$, which is therefore uniformly IFP definable in $\mathfrak{I}^k(\mathfrak{A})$.

The translation from $\mathrm{PFP}^k(\tau)$ *to* $\mathrm{PFP}(\tau^k)$*:*

Atomic formulae: Any quantifier-free $\varphi(\mathbf{x}) \in \mathrm{FO}^k(\tau)$ is logically equivalent to one of the formulae θ that give rise to the predicates P_θ in τ^k. Then $\hat{\varphi} := P_\theta x$ is as required. For atoms $\varphi = Z\mathbf{y}$ where $\mathbf{y} = (x_{\sigma(1)}, \ldots, x_{\sigma(r)})$, put $\hat{\varphi} := \exists y\big(\eta_\sigma(x,y) \wedge \hat{Z}y\big)$.

Boolean connectives: trivially commute with the translation. E.g., if $\varphi = \varphi_1 \wedge \varphi_2$ then $\hat{\varphi} := \hat{\varphi}_1 \wedge \hat{\varphi}_2$ works.

Quantification: For $\varphi = \exists x_j \psi$, put $\hat{\varphi} := \exists y(E_j xy \wedge \hat{\psi}(y))$; similarly for universal quantification.

Partial fixpoints: Consider $\varphi = \mathrm{PFP}_{X,\mathbf{x}_1}\psi(X,\mathbf{x})$, where $\mathbf{x} = \mathbf{x}_0\mathbf{x}_1$ and $\mathbf{x}_0$ acts as a parameter tuple (X and $\mathbf{x}_1$ of arity $r \leqslant k$). Let $\hat{\psi}(\hat{X},x)$ be the translation of ψ. We use the formula $\eta_\sigma(x,y)$, where σ is the identity on the components in $\mathbf{x}_0$. The formula $\mathrm{PFP}_{\hat{X},y}\big(\hat{\psi}(\hat{X},y) \wedge \eta_\sigma(x,y)\big)$ refers to a fixpoint in $\hat{X}$ and y whose y is coupled to the parameter x by η_σ in such a way as to admit only equivalence classes y that are compatible with equivalence class of the fixed $\mathbf{x}_0$. Since convergence of the fixpoint iteration in $\mathfrak{A}, \mathbf{a}$ only depends on the equivalence class of $\mathbf{a}_0$ (which is determined by the equivalence class of $\mathbf{a}$), convergence of the fixpoint process based on $\hat{\psi}(\hat{X},y) \wedge \eta_\sigma(x,y)$ in $\mathfrak{I}^k(\mathfrak{A})$ is dependent only on parameter α for x. Therefore the following is as desired:

$$\hat{\varphi}(x) := \exists y\big(x = y \wedge \mathrm{PFP}_{\hat{X},y}\big(\hat{\psi}(\hat{X},y) \wedge \eta_\sigma(x,y)\big)\big).$$

We have shown the following.

Lemma 7.4.10 *Every* PFP^k*-definable query* Q *over* $\mathsf{FIN}(\tau)$ *translates into a* PFP*-definable query* $\hat{Q}$ *over* $\mathsf{FIN}(\tau^k)$ *such that for all* $\mathfrak{A} \in \mathsf{FIN}(\tau)$ *and* $\mathbf{a} \in A^k$*:*

$$\mathbf{a} \in Q^{\mathfrak{A}} \quad \text{iff} \quad [\mathbf{a}] \in \hat{Q}^{\mathfrak{I}^k(\mathfrak{A})}.$$

Corollary 7.4.11 *Let* $k\text{-size}(\mathfrak{A}) := |A^k/\simeq^k_\infty|$ *stand for the number of equivalence classes w.r.t.* $\simeq^k_\infty$ *over* A^k*. Then the partial fixpoint of* $\varphi(X,\mathbf{x}) \in \mathrm{PFP}^k(\tau)$ *is either empty or reached within* $2^{k\text{-size}(\mathfrak{A})}$ *many steps over* $\mathfrak{A}$*.*

Exercise 7.4.12 Use the preceding fact to analyse the expressive power of $\mathrm{PFP}^k(\emptyset)$, i.e., of PFP over naked sets. An alternative analysis proceeds by direct induction on the syntax of $\mathrm{PFP}^k(\emptyset)$, to show that next to nothing (but what exactly?) is definable in PFP over sets without structure.

Exercise 7.4.13 Describe the inductive iteration for the fixpoint process underlying the formulae η_σ in terms of the $\mathfrak{I}^k(\mathfrak{A})$.

Exercise 7.4.14 Outline variant translations that work for IFP and LFP. Note that for LFP positivity in second-order variables needs to be preserved in the translation.

7.4.3 From the invariants back to the real structures

We now indicate a translation in the opposite direction, for expressing definable properties of $\mathfrak{I}^k(\mathfrak{A})$ as properties of the underlying $\mathfrak{A}$. We thus pull back PFP/IFP/LFP definabilty in terms of $\mathfrak{I}^k(\mathfrak{A})$ to PFP/IFP/LFP definability in terms of $\mathfrak{A}$. We explicitly treat LFP, for our purposes below, but translations for the other fixpoint logics could be obtained in the same fashion. Our main goal is a translation

$$\begin{array}{rcl} \check{} : \mathrm{LFP}(\tau^k) & \longrightarrow & \mathrm{LFP}(\tau) \\ \varphi(\mathbf{Z}, \mathbf{z}) & \longmapsto & \check{\varphi}(\check{\mathbf{Z}}, \check{\mathbf{z}}). \end{array}$$

In this translation we replace each first-order variable z_i (ranging over elements of $\mathfrak{I}^k(\mathfrak{A})$, corresponding to equivalence classes of k-tuples of elements of $\mathfrak{A}$) by a k-tuple of variables $\check{z}_i = \mathbf{x}_i = (x_{i1}, \dots, x_{ik})$. Similarly, we replace each second-order variable Z of arity r (ranging over sets of r-tuples over $\mathfrak{I}^k(\mathfrak{A})$) by a kr-ary relation $\check{Z}$ (ranging over sets of r-tuples of k-tuples over $\mathfrak{A}$).

For an r-ary relation R over $\mathfrak{I}^k(\mathfrak{A})$ let

$$\check{R} := \big\{(\mathbf{a}_1, \dots, \mathbf{a}_r) \colon ([\mathbf{a}_1], \dots, [\mathbf{a}_r]) \in R\big\} \subseteq A^{kr}.$$

Then the semantic condition on the translation $\varphi \mapsto \check{\varphi}$ is that for all $\mathfrak{A} \in \mathsf{FIN}(\tau)$, and for all assignments $\mathbf{R}$ to the second-order variables $\mathbf{Z}$ of φ,

$$\mathfrak{A} \models \check{\varphi}[\check{\mathbf{R}}, \mathbf{a}_1, \dots, \mathbf{a}_m] \quad \text{iff} \quad \mathfrak{I}^k(\mathfrak{A}) \models \varphi[\mathbf{R}, [\mathbf{a}_1], \dots, [\mathbf{a}_m]].$$

We firstly need an auxiliary formula that defines $\leqslant$, the linear ordering w.r.t. $\simeq^k_\infty$-types in $\mathfrak{I}^k(\mathfrak{A})$, in terms of $\mathfrak{A}$. This is precisely the result of the inductive refinement w.r.t. k-variable types from our analysis of the unbounded k-pebble game. Compare section 2.5.2 and in particular the paragraph on pre-ordering types on page 50.

We there obtained $\preccurlyeq$ as the (inductive) fixpoint of an FO-definable operation. By the Gurevich–Shelah theorem, Theorem 7.2.12, there is also an LFP-formula

$$\eta_\preccurlyeq(x_{11}, \dots, x_{1k}, x_{21}, \dots, x_{2k}) \in \mathrm{LFP}(\tau)$$

such that $\mathfrak{A} \models \eta_\preccurlyeq[\mathbf{a}_1, \mathbf{a}_2]$ iff $[\mathbf{a}_1] \leqslant [\mathbf{a}_2]$ in $\mathfrak{I}^k(\mathfrak{A})$. This means that $\eta_\preccurlyeq$ is the desired translation of $z_1 \leqslant z_2$. It then follows that

$$\eta_\approx(\check{z}_1, \check{z}_2) := \eta_\preccurlyeq(\check{z}_1, \check{z}_2) \wedge \eta_\preccurlyeq(\check{z}_2, \check{z}_1)$$

is the correct translation of $z_1 = z_2$ (check this against the semantic requirements!).

Again $\check{\varphi}$ is obtained by induction on φ.

Atomic formulae: $P_\theta z$ translates into $\theta(\check{z})$;
$Zz_1 \dots z_r$ into $\check{Z}\check{z}_1 \dots \check{z}_r$;
$z_1 = z_2$ into $\eta_\approx(\check{z}_1, \check{z}_2)$;
$z_1 \leqslant z_2$ into $\eta_\preccurlyeq(\check{z}_1, \check{z}_2)$;
$E_j z_1 z_2$ into $\exists x_{1j} \eta_\approx(\check{z}_1, \check{z}_2)$.

Boolean connectives: trivially commute with the translation.

Quantification: E.g., if $\varphi = \exists z_j \psi$, put $\check{\varphi} := \exists x_{j1} \dots \exists x_{jk} \check{\psi}$.

Least fixpoints: Note that the other steps preserve positivity, in the sense that $\check{\varphi}(\check{Z})$ is positive in $\check{Z}$ if φ is positive in Z. One may therefore just pull back least fixpoints. For instance, $\varphi = \mu_{X,\mathbf{x}}\psi(X,\mathbf{x})$ translates into $\check{\varphi}(\check{\mathbf{x}}) := \mu_{\check{X},\check{\mathbf{x}}}\check{\psi}(\check{X},\check{\mathbf{x}})$. Correctness is shown by induction on the stages of these fixpoints; one establishes that the i-th stage of the fixpoint w.r.t. $\check{\psi}$ over $\mathfrak{A}$ is the translation of the i-th stage of the fixpoint w.r.t. ψ over $\mathfrak{I}^k(\mathfrak{A})$.

This gives the following.

Lemma 7.4.15 *Every* LFP*-definable r-ary query Q over* $\{\mathfrak{I}^k(\mathfrak{A}) \colon \mathfrak{A} \in \mathsf{FIN}(\tau)\} \subseteq \mathsf{FIN}(\tau^k)$ *translates into an* LFP*-definable rk-ary query $\check{Q}$ over* $\mathsf{FIN}(\tau)$ *such that for all* $\mathfrak{A} \in \mathsf{FIN}(\tau)$ *and* $(\mathbf{a}_1, \ldots, \mathbf{a}_r) \in A^{rk}$*:*

$$(\mathbf{a}_1, \ldots, \mathbf{a}_r) \in \check{Q}^{\mathfrak{A}} \quad \textit{iff} \quad ([\mathbf{a}_1], \ldots, [\mathbf{a}_r]) \in Q^{\mathfrak{I}^k(\mathfrak{A})}.$$

Exercise 7.4.16 Check that for a (boolean) query Q that is PFP^k-definable over $\mathsf{FIN}(\tau)$, the passage of translations from Q to $Q_1 := \hat{Q}$ (according to Lemma 7.4.10) to $Q_2 := \check{Q}_1$ (according to Lemma 7.4.15) gives $Q_2 = Q$.

7.4.4 P versus Pspace

Lemmas 7.4.10 and 7.4.15, together with the capturing results for P and Pspace in the presence of order now prove the following.

Theorem 7.4.17 (Abiteboul–Vianu)
The following are equivalent:

(i) LFP *and* PFP *have the same expressive power – define exactly the same boolean queries – over finite relational structures.*

(ii) Pspace *collapses to* P*.*

Proof (i) $\Rightarrow$ (ii). Using (i) just for classes of linearly ordered finite structures, we obtain (ii) from Theorem 7.3.3 and Lemma 7.2.5.

(ii) $\Rightarrow$ (i). Assume Pspace = P and let Q be definable in PFP. For suitable k, Q is definable by a PFP^k-sentence φ. By Lemma 7.4.10, the associated $\hat{Q}$ is PFP-definable, and hence by Lemma 7.3.2 in particular in Pspace. From the assumption that Pspace = P we get that $\hat{Q}$ is in P. As the invariants are linearly ordered structures, $\hat{Q}$ is definable in LFP by Theorem 7.2.8. Then Lemma 7.4.15, together with Exercise 7.4.16, shows that Q is LFP-definable. $\square$