Contents

1 Erhenfeucht-Fraïssé Game **1**

2 FO logic, rank-q type, Ehrenfeucht-Fraïssé Theorem **2**

3 Inexpressibility in FO logic **4**

4 Proof of Ehrenfeucht-Fraïssé Theorem **6**

5 EF game for MSO logic **9**

1 Erhenfeucht-Fraïssé Game

Erhenfeucht-Fraïssé game (EF game in short) is a central notion for *finite* relational structures and logic. In this game, we have two players SPOILER and DUPLICATOR who play a game on two τ-structure $\mathbb{A}$ and $\mathbb{B}$. The goal of SPOILER is to reveal the difference between the two structures, where as $\mathbb{B}$ claims that they are the same. The SPOILER wins the game if it succeeds to reveal the difference and the DUPLICATOR wins if it manages to hide the difference during the game (or none exists in fact). Let us explain the rule of Erhenfeucht-Fraïssé game.

One round of EF game on $(\mathbb{A}, \mathbb{B})$ consists of the two moves:

1. the SPOILER selects one of $\mathbb{A}$ and $\mathbb{B}$ and chooses an element of the chosen structure at will,
2. the DUPLICATOR selects an element of the other structure; $a \in \mathbb{A}$ if the SPOILER chose $b \in \mathbb{B}$ and $b \in \mathbb{B}$ if the SPOILER chose $a \in \mathbb{A}$.

An *n-round Eherenfeucht-Fraïsse game* is expressed by two n-tuples $\vec{a} \in \mathbb{A}^n$ and $\vec{b} \in \mathbb{B}^n$, where a_i and b_i are the elements chosen by SPOILER and DUPLICATOR during the i-th round of the game; We do not distinguish whether the SPOILER chose a_i or b_i.

To present the winning condition of the EF game, we introduce the notion of *partial isomorphism.*

Definition 1 (Partial isomorphism)**.** *Let $\mathbb{A}$ and $\mathbb{B}$ be two τ-structures and let $\vec{a} \in A^n$ and $\vec{b} \in A^n$ be two n-tuples in $\mathbb{A}$ and $\mathbb{B}$ respectively. We say that $(\vec{a}, \vec{b})$ defines a* partial isomorphism between $\mathbb{A}$ and $\mathbb{B}$ *if the following holds.*

- *for every $i, j \leq \ell$, $a_i = a_j$ if and only if $b_i = b_j$*
- *for every constant symbol $c \in \tau$ and for every $i \leq \ell$, $a_i = c^{\mathbb{A}}$ if and only if $b_i = c^{\mathbb{B}}$.*

- *for every predicate $R \in \tau$ with $\mathsf{ar}(R) = k$ and for every k-tuple $(i_1, \ldots, i_k)$ of indices (not necessarily distinct, not necessarily in order) from $[n]$, $(a_{i_1}, \ldots, a_{i_k}) \in R^{\mathbb{A}}$ if and only if $(b_{i_1}, \ldots, b_{i_k}) \in R^{\mathbb{B}}$.*

After an n-round EF game on $\mathbb{A}$ and $\mathbb{B}$, let $\vec{a} \in A^n$ and $\vec{b} \in A^n$ be the game moves. The DUPLICATOR *is the winner of the n-round EF-game* (described by the game moves) $(\vec{a}, \vec{b})$ if $\left(\vec{c}^{\mathbb{A}} \circ \vec{a}, \vec{c}^{\mathbb{B}} \circ \vec{b}\right)$ defines a partial isomorphism between $\mathbb{A}$ and $\mathbb{B}$. The SPOILER wins the game otherwise. Here, $\vec{c}^{\mathbb{A}}$ is the tuple of constants in $\mathbb{A}$, each entry being $c^{\mathbb{A}} \in \mathbb{A}$ corresponding to the constant symbol $c \in \tau$. The concatenation of two tuples $\vec{c}^{\mathbb{A}}$ and $\vec{a}$ is written as $\vec{c}^{\mathbb{A}} \circ \vec{a}$.

Notice that the role of $\vec{c}^{\mathbb{A}}$ and $\vec{c}^{\mathbb{B}}$; it is as if EF game has been already played for $|\{c \mid c \text{ is a constant symbol in } \tau\}|$ rounds, and the SPOILER and DUPLICATOR continues the game n rounds.

We say that *the* DUPLICATOR *has a winning strategy in the n-round EF game on $\mathbb{A}$ and $\mathbb{B}$* if the DUPLICATOR has a matching sequence of moves to win the game regardless of the moves of the SPOILER, and write as $\mathbb{A} \equiv_n \mathbb{B}$. It is easy to see that $\mathbb{A} \equiv_n \mathbb{B}$ implies $\mathbb{A} \equiv_{n'} \mathbb{B}$ for every $n' \leq n$.

2 FO logic, rank-q type, Ehrenfeucht-Fraïssé Theorem

Let us momentarily turn the attention to FO logic. FO logic is a special case of MSO-logic and we import all the notations developed for MSO-logic to work with FO logic. Now that we do not allow a free set variable in an FO-formula, an atomic formula over τ is in the form

- $x = y$, where x (respectively y) is a free (individual) variable or a constant symbol $c \in \tau$,
- $R(x_1, \ldots, x_{\mathsf{ar}(R)})$ for some $R \in \tau$.

The set of all FO-formulas of quantifier rank at most q is denoted by FO$[q]$.

Definition 2 (Rank-k type)**.** *For a relational structure $\mathbb{A}$ over τ and an ℓ-tuple $\vec{v} = (v_1, \ldots, v_\ell) \in A^\ell$ of elements of A, we define the* FO rank-q ℓ-type *of $(\mathbb{A}, \vec{v})$ as the set of all* FO*-formulas of quantifier rank at most q with ℓ free variables satisfied by $(\mathbb{A}, \vec{v})$. That is,*

$$\mathsf{fo\text{-}type}_q(\mathbb{A}, \vec{v}) = \{\psi \in \mathrm{FO}[q] \mid \mathbb{A} \models \psi(\vec{v})\}.$$

We often omit to explicitly mention FO*, and say rank-q ℓ-type of $(\mathbb{A}, \vec{v})$ or even just the type of $(\mathbb{A}, \vec{v})$.*

When $\ell = 0$, the rank-q ℓ-type of a structure $\mathbb{A}$ is simply called the rank-q type *of $\mathbb{A}$. Notice that $\mathsf{fo\text{-}type}_q(\mathbb{A})$ is the set of all* FO*-sentences of quantifier rank at most k which holds on $\mathbb{A}$.*

An rank-q ℓ-type *(when a τ-structure $\mathbb{A}$ is not specified) is the set S of* FO*-formulas in* FO$[q]$ *with ℓ free individual variables which is* consistent and complete*; there exists $\mathbb{A}$ and $\vec{a} \in \mathbb{A}^\ell$ such that $\mathbb{A} \models \varphi(\vec{a})$ for every $\varphi \in S$ and for every* FO*-formula ψ with ℓ free variables, either $\psi \in S$ or $\neg\psi \in S$ holds.*

What we previously observed in MSO rank-k ℓ, m-types also hold for (FO) rank-q ℓ-types. We make those observations explicit below.

Lemma 3. *The following holds.*

1. *For any fixed q, l, the number of pairwise logically nonequivalent* FO*-formulas of quantifier rank at most q with ℓ free variables is finite, determined solely by q, ℓ and the vocabulary τ.*

2. *For any fixed q, l, the number of* FO *rank-q ℓ-types is finite, determined solely by k, ℓ and the vocabulary τ.*

3. *Let $T_1, \dots, T_r$ be all the rank-q ℓ-types. There exists* FO*-formulas in* $\mathrm{FO}[q]$ $\varphi_1(\vec{x}), \dots, \varphi_r(\vec{x})$ *such that*

   (a) *For every τ-structure $\mathbb{A}$ and $\vec{a} \in \mathbb{A}^\ell$, $\mathsf{fo\text{-}type}_q(\mathbb{A}, \vec{a}) = T_i$ if and only if $\mathbb{A} \models \varphi_i(\vec{a})$, and*

   (b) *for every formula $\varphi \in \mathrm{FO}[q]$ with ℓ free variables, φ is equivalent to a disjunction of some φ_j's.*

Proof: We skip the proof of (1). Let $R = \{\psi_1(\vec{x}), \dots, \psi_M(\vec{x})\}$ be a set of all pairwise nonequivalent FO-formulas with ℓ free variables. Due to (1), there is a finite such set R. To each T_i associate the set $J_i \subseteq R$ precisely consisting of $\psi_j(\vec{x})$'s satisfied by $(\mathbb{A}, \vec{a})$ of type T_i. That is,

$$J_i := R \cap T_i$$

Note that $\psi_j \in J_i$ if and only if $\psi_j \in T_i$ and all FO-formulas logically equivalent to ψ_j is also contained in T_i.

To see (2), it suffices to argue that $J_i = J_{i'}$ implies $i = i'$. Indeed, if $i \neq i'$ there exists an FO-formula $\alpha \in \mathrm{FO}[q]$ such that $\alpha \in T_i$ and $\neg\alpha \in T_{i'}$. Let $\psi \in R$ be the FO-formula logically equivalent to α. then $\psi \in J_i$ and $\psi \notin J_{i'}$, therefore we have $J_i \neq J_{i'}$. We conclude that each type T_i is uniquely associated with some subset of R, and (1) implies (2).

To see (3)-(a), let

$$\varphi_i := \bigwedge_{\psi \in J_i} \psi$$

and we claim that $\mathsf{fo\text{-}type}(\mathbb{A}, \vec{a}) = T_i$ if and only if $\mathbb{A} \models \varphi_i(\vec{a})$. The forward direction is straightforward by the construction of J_i. For the backward direction, suppose $\mathbb{A} \models \varphi_i(\vec{a})$ holds for some $(\mathbb{A}, \vec{a})$ and let $\mathsf{fo\text{-}type}_q(\mathbb{A}, \vec{a}) = T_{i'}$. As $(\mathbb{A}, \vec{a})$ satisfies ψ for all $\psi \in J_i$, it holds that $J_i \subseteq J_{i'}$. If $J_{i'} \subsetneq J_i$, then there is a formula $\gamma \in J_{i'} \setminus J_i \subseteq R \setminus J_i$ and thus we have $\mathbb{A} \models \neg\gamma(\vec{a})$ by the construction of J_i. On the other hand, γ is in the type $T_{i'}$ of $(\mathbb{A}, \vec{a})$ by construction of $J_{i'}$, thus $A \models \gamma(\mathbb{A}, \vec{a})$. We reach a contradiction as both γ and $\neg\gamma$ holds on $(A, \vec{a})$. Therefore, it holds that $i' = i$.

To see (3)-(b), let $j^* \in [M]$ be the index such that β is logically equivalent to ψ_{j^*}. Take $I^* \subseteq [r]$ as the set of all types among $T_1, \dots, T_r$ *compatible* with β, i.e.

$$I^* := \{i \in [r] \mid \psi_{j^*} \in J_i\}.$$

We claim that $\mathbb{A} \models \beta(\vec{a})$ if and only if $\mathbb{A} \models \varphi_i(\vec{a})$ for some $i \in I^*$, or equivalently $(\mathbb{A}, \vec{a})$ satisfies $\bigvee_{i \in I^*} \varphi_i$. The backward implication immediately follows from that whenever φ_i holds on $(\mathbb{A}, \vec{a})$ for some $i \in I^*$, ψ_{j^*} and thus β holds on $(\mathbb{A}, \vec{a})$. For the forward implication, notice that

$$\mathsf{fo\text{-}type}_q(\mathbb{A}, \vec{a}) \cap \{\psi_{j^*}, \neg\psi_{j^*}\} = \{\psi_{j^*}\}$$

holds and thus k belongs to I^*, where $T_k := \mathsf{fo\text{-}type}_q(\mathbb{A}, \vec{a})$. We complete the proof by applying (3)-(a). $\square$

The central theorem by Ehrenfeucht and Fraïssé [1, 2] connects the q-round EF game to the the equivalence of two structures for FO logic, i.e. rank-q types of the structures. It says that if the DUPLICATOR has a winning strategy for q-round EF game on $\mathbb{A}$ and $\mathbb{B}$, the two structures satisfy precisely the same set of

FO-sentences of quantifier rank up to q. Moreover, if the SPOILER can play the game to win (so that the DUPLICATOR fails to make a matching sequence of moves), then the corresponding game move reveals that $\mathbb{A}$ satisfies some specific FO-sentence ψ of quantifier rank at most q where as $\mathbb{B}$ satisfies $\neg\psi$.

Theorem 4 (Ehrenfeucht-Fraïssé Theorem). *[1, 2] Let $\mathbb{A}$ and $\mathbb{B}$ be two τ-structures. The followings are equivalent.*

- $\mathbb{A} \equiv_n \mathbb{B}$.
- $\mathsf{fo\text{-}type}_n(\mathbb{A}) = \mathsf{fo\text{-}type}_n(\mathbb{B})$.

3 Inexpressibility in FO logic

As one of its fundamental consequences, Eherenfeucht-Fraïssé Theorem provides a means to demonstrate the inexpressibility of some property (as a set of τ-structures) in FO logic. Suppose that a property $\mathcal{P}$ of τ-structures is FO-definable. Equivalently, there exists some finite $q \geq 0$ and an FO-sentence $\varphi \in \text{FO}[q]$ such that

$$\mathbb{A} \in \mathcal{P} \qquad \text{if and only if } \mathbb{A} \models \varphi.$$

With the implication $(I) \Rightarrow (II)$, this means that for any two τ-structures $\mathbb{A}, \mathbb{B}$ with $\mathbb{A} \equiv_q \mathbb{B}$, $\mathbb{A} \in \mathcal{P}$ implies $\mathbb{B} \in \mathcal{P}$. The contraposition is stated as in the following.

Corollary 5. *A property $\mathcal{P}$ of τ-structures is not expressible (definable) in* FO *if for every $q \geq 1$, there exists two τ-structures $\mathbb{A}$ and $\mathbb{B}$ such that $\mathbb{A} \equiv_q \mathbb{B}$, $\mathbb{A} \in \mathcal{P}$ and $\mathbb{B} \notin \mathcal{P}$.*

Proof: We prove the contraposition. Suppose that $\mathcal{P}$ is defined by an FO-sentence Φ and let $n = \mathsf{qr}(\Phi)$. For any two τ-structures $\mathbb{A}$ and $\mathbb{B}$, if $\mathbb{A} \equiv_n \mathbb{B}$ then by Theorem 4 $\mathsf{fo\text{-}type}_n(\mathbb{A}) = \mathsf{fo\text{-}type}_n(\mathbb{B})$. In particular, $\mathbb{A} \models \Phi$ if and only if $\mathbb{B} \models \Phi$, completing the proof. □

Using Corollary 5, one can show that some property is not FO-definable.

Example 6. *Consider the vocabulary $\tau = \emptyset$. Then any τ-structure is simply a set.*

1. *Consider two sets $\mathbb{A}$ and $\mathbb{B}$ with whose ground set have size at least n respectively. Then $\mathbb{A} \equiv_n \mathbb{B}$. (What is the* DUPLICATOR*'s winning strategy?*
2. *Let* EVEN *be the property of all (finite) sets of even size. For every $n \geq 1$, one can choose $\mathbb{A}$ to be a set of size n and $\mathbb{B}$ to be a set of size $n+1$. We have $\mathbb{A} \equiv_n \mathbb{B}$ but only one of them belongs to* EVEN*. So the property* EVEN *is not* FO*-definable.*

Lemma 7. *Consider the vocabulary $\tau = \{<\}$ and two τ-structures $\mathbb{A}$ and $\mathbb{B}$ in each of which $<$ is interpreted as a linear order. If both $\mathbb{A}$ and $\mathbb{B}$ have universe size at least $2^n + 1$, then $\mathbb{A} \equiv_n \mathbb{B}$.*

Proof: We use the induction on n. When $n = 0$, the equivalence is trivial. For the inductive step, we need the next claim.

Claim 1. *Let $a \in \mathbb{A}$ and $b \in \mathbb{B}$ such that $\mathbb{A}^{\leq a} \equiv_q \mathbb{B}^{\leq b}$ and $\mathbb{A}^{\geq a} \equiv_q \mathbb{B}^{\geq b}$. Then $\mathbb{A} \equiv_q \mathbb{B}$. Moreover, in any $(q+1)$-round game which begins with $a \in \mathbb{A}$ and $b \in \mathbb{B}$, the* DUPLICATOR *has a winning strategy.*

PROOF OF THE CLAIM: Let $a_{i_0} := a$ and $b_{i_0} = b$ for some i. The DUPLICATOR's strategy for continuing this game is to replicate its move in the prefix $\mathbb{B}^{\leq b}$ when the SPOILER chooses an element from $\mathbb{A}^{\leq a}$ and vice versa, and do the same if the SPOILER choose an element from the prefix. Note that the resulting game is in the form

$$(a_0, a_1, \ldots, a_q, b_0, b_1, \ldots, b_q).$$

It remains to observe that

- for every $0 \leq i < j \leq q$; if $a_i = a_j$ then any winning game for the DUPLICATOR on the prefix / suffix game will ensure $b_i = b_j$ (the converse also holds).
- for every $0 \leq i < j \leq q$; suppose $a_i < a_j$
  1. if both a_i and a_j are from the prefix, any winning game for the DUPLICATOR on the prefix / suffix game will ensure $b_i < b_j$. The same argument remains valid when both elements are from the suffix.
  2. if a_i is from the prefix and a_j is from the suffix, the DUPLICATOR's strategy of restricting the same in the prefix/suffix as the SPOILER has chosen ensures that $b_i < b_j$.

$\diamondsuit$

Without loss of generality, let $a \in \mathbb{A}$ is the first move of the SPOILER. We want to determine the DUPLICATOR's choice $b \in \mathbb{B}$ so that it satisfies the precondition of Claim 1 for $q = n - 1$. If this is achievable for any choice of $a \in \mathbb{A}$, then we are done by Claim 1.

So we describe how the DUPLICATOR chooses $b \in \mathbb{B}$ depending on the choice $a \in \mathbb{A}$ of the SPOILER. Let $\min(\mathbb{A})$ be the minimum element of $\mathbb{A}$, $\max(\mathbb{A})$ is defined as the maximum element. The distance between $x, y \in \mathbb{A}$ is the number of elements in-between x and y, plus 1. Note that the assumption on $\mathbb{A}$ and $\mathbb{B}$ means that $\mathsf{dist}_{\mathbb{A}}(\min(\mathbb{A}), \max(\mathbb{A})) \geq 2^n$ and $\mathsf{dist}_{\mathbb{B}}(\min(\mathbb{B}), \max(\mathbb{B})) \geq 2^n$.

- If $\mathsf{dist}(\min(\mathbb{A}), a) < 2^{n-1}$, then choose $b \in \mathbb{B}$ so that $\mathsf{dist}(\min(B), b) = \mathsf{dist}(\min(A), a)$.
- If $\mathsf{dist}(\max(\mathbb{A}), a) < 2^{n-1}$, then choose $b \in \mathbb{B}$ so that $\mathsf{dist}(\max(B), b) = \mathsf{dist}(\min(A), a)$.
- Otherwise, choose any $b \in \mathbb{B}$ so that $\mathsf{dist}(\min(\mathbb{B}), b) \geq 2^{n-1}$ and $\mathsf{dist}(\max(\mathbb{B}), b) \geq 2^{n-1}$

By the assumption on the size of the universe, $\mathsf{dist}(\min(\mathbb{A}), \max(\mathbb{A})) \geq 2^n$ and exactly one of the above cases hold. For each choice of $b \in \mathbb{B}$, the precondition of Claim 1 holds: in the last case, by induction hypothesis and in the first two cases, due to the isomorphism and the induction hypothesis on the larger side. $\square$

One can use Lemma 7 and Corollary 5 to prove that the property EVEN of linear orders of even cardinality is not FO-definable.

Eherenfeucht-Fraïssé Theorem furthermore indicates that the converse of Corollary 5 holds as well. Therefore, if a property $\mathcal{P}$ is not FO-definable, then one can always[1] prove the inexpressibility in FO by demonstrating such pair of structures as depicted in Corollary 5.

Corollary 8. *Let $\mathcal{P}$ be a property of τ-structures. Then $\mathcal{P}$ is not FO-definable if and only if the following*

[1]The phenomenon is referred to as the *completeness of Ehrenfeucht-Fraïssé game for* FO *logic*.

holds for every $n \geq 0$:

there exist τ-structures $\mathbb{A}, \mathbb{B}$ such that $\mathbb{A} \equiv_n \mathbb{B}$, $\mathbb{A} \in \mathcal{P}$ and $\mathbb{B} \notin \mathcal{P}$.

Proof: The backward implication is in Corollary 5. To prove the forward implication, suppose that there exists $q \geq 0$ such that for every $\mathbb{A}, \mathbb{B}$ with $\mathbb{A} \equiv_q \mathbb{B}$, $\mathbb{A} \in \mathcal{P}$ holds if and only if $\mathbb{B} \in \mathcal{P}$ holds. This means that $\mathcal{P}$, as a set of τ-structures, is a disjoint union of equivalence classes by $\equiv_q$; any equivalence class under $\equiv_q$ is either fully contained in $\mathcal{P}$ or disjoint from $\mathcal{P}$. By Theorem 4 and (3)-(a) of Lemma 3, there exist a finite set of FO-sentences $\varphi_1, \ldots, \varphi_t$ in FO$[q]$ which defines the equivalence classes contained in $\mathcal{P}$. Note that the disjunction of these sentences define the property $\mathcal{P}$. $\square$

Another consequence of Theorem 10 is given below.

Corollary 9 (Hintikka formula)**.** *For each $q \geq 0$, τ-structure $\mathbb{A}$ and a tuple $\vec{a}_0 \in \mathbb{A}^\ell$, there exists for a formula $\psi := \psi^q_{(\mathbb{A},\vec{a})}$ of quantifier rank q and ℓ free variables which defines the set of pairs $(\mathbb{B}, \vec{b}_0)$, where $\mathbb{B}$ is a τ-structure and $\vec{b}_0 \in \mathbb{B}^\ell$, that are equivalent with $(\mathbb{A}, \vec{a}_0)$ under q-round EF game. That is, $(\mathbb{B}, \vec{b}_0) \equiv_q (\mathbb{A}, \vec{a}_0)$ if and only if $\mathbb{B} \models \psi(\vec{b}_0)$. (Such a formula is known as a hintikka formula in the literature.)*

Proof: Let $\mathsf{fo\text{-}type}_q(\mathbb{A}, \vec{a}_0) = T_i$ and let $\varphi_i(\vec{x})$ be the FO-formula as given in (3)-(a) of Lemma 3. By Theorem 10, the claim follows. $\square$

4 Proof of Ehrenfeucht-Fraïssé Theorem

To prove Theorem 4 by induction on n, we need a more general version of Ehrenfeucht-Fraïssé game in which we a k-configuration. For two τ-structures $\mathbb{A}$ and $\mathbb{B}$, a *s-configuration* is a pair $(\vec{a}_0, \vec{b}_0)$ consisting of two k-tuples $\vec{a}_0 \in \mathbb{A}^s$ and $\vec{b}_0 \in \mathbb{B}^s$.

An n-round Ehrenfeucht-Fraïssé game on $(\mathbb{A}, \vec{a}_0, \mathbb{B}, \vec{b}_0)$, two structures $\mathbb{A}$ and $\mathbb{B}$ together with an s-configuration $\vec{a}_0 \in \mathbb{A}^s$ and $\vec{b}_0 \in \mathbb{B}^s$, works exactly like the usual EF game. The difference is the winning condition for the DUPLICATOR, namely the SPOILER and the DUPLICATOR alternates to select an element of $\mathbb{A}$ and an element of $\mathbb{B}$, described with a pair $(\vec{a}, \vec{b})$ of n-tuples of $\mathbb{A}$ and $\mathbb{B}$. The DUPLICATOR wins the n-round EF game $(\vec{a}, \vec{b})$ on $(\mathbb{A}, \vec{a}_0, \mathbb{B}, \vec{b}_0)$ if

the pair $(\vec{c}^{\mathbb{A}} \circ \vec{a}_0 \circ \vec{a}, \vec{c}^{\mathbb{B}} \circ \vec{b}_0 \circ \vec{b})$ defines a partial isomorphism between $\mathbb{A}$ and $\mathbb{B}$.

Intuitively, the s-configuration describes the moves of the SPOILER and the DUPLICATOR played beforehand by the time a sequence of n moves begin. So, the winning condition takes the configuration into account by defining the partial isomorphism in this way. Clearly, EF game with 0-configuration (empty configuration) is the basic EF game we learned earlier.

We write $(\mathbb{A}, \vec{a}_0) \equiv_n (\mathbb{B}, \vec{b}_0)$ if the DUPLICATOR has a winning strategy in the n-round EF game on $(\mathbb{A}, \vec{a}_0, \mathbb{B}, \vec{b}_0)$.

Theorem 10 (Ehrenfeucht-Fraïssé Theorem)**.** *[1, 2] Let $\mathbb{A}$ and $\mathbb{B}$ be two τ-structures, and let $\vec{a}_0 \in \mathbb{A}^s$ and $\vec{b}_0 \in \mathbb{B}^s$ for some $s \geq 0$. The followings are equivalent.*

1. $(\mathbb{A}, \vec{a}_0) \equiv_n (\mathbb{B}, \vec{b}_0)$.

II. $\mathsf{fo\text{-}type}_n(\mathbb{A}, \vec{a}_0) = \mathsf{fo\text{-}type}_n(\mathbb{B}, \vec{b}_0)$.

What does $(\mathbb{A}, \vec{a}_0) \equiv_0 (\mathbb{B}, \vec{b}_0)$ mean? With s-configuration (possibly empty) and constants (possibly empty as well), there is a partial isomorphism between $\mathbb{A}$ and $\mathbb{B}$. How about $\mathsf{fo\text{-}type}_n(\mathbb{A}, \vec{a}_0) = \mathsf{fo\text{-}type}_n(\mathbb{B}, \vec{b}_0)$ for $n = 0$? Notice that a formula of quantifier rank 0 is a quantifier-free formula (and vice versa), and a quantifier-free formula is a formula which is obtained as a boolean combination of atomic formulas, i.e. $x = y$ or $R(x_1, \dots, x_r)$ for some $R \in \tau$.

The validity of Theorem 10 for the base case $n = 0$ is tedious to check.

Lemma 11. *Let $\mathbb{A}$ and $\mathbb{B}$ be two τ-structures, and let $\vec{a}_0 \in \mathbb{A}^s$ and $\vec{b}_0 \in \mathbb{B}^s$ for some $s \geq 0$. The followings are equivalent.*

(i) $(\mathbb{A}, \vec{a}_0) \equiv_0 (\mathbb{B}, \vec{b}_0)$.

(ii) $(\vec{c}^A \circ \vec{a}_0, \vec{c}^{\mathbb{B}} \circ \vec{b}_0)$ *defines a partial isomorphism between $\mathbb{A}$ and $\mathbb{B}$.*

(iii) for every quantifier-free formula $\varphi(\vec{x})$ with at most s free variables, $\mathbb{A} \models \varphi(\vec{a}_0)$ if and only if $\mathbb{B} \models \varphi(\vec{b}_0)$.

(iv) for every atomic formula $\varphi(\vec{x})$ with at most s free variables, $\mathbb{A} \models \varphi(\vec{a}_0)$ if and only if $\mathbb{B} \models \varphi(\vec{b}_0)$.

Proof: (i) and (ii) are equivalent by definition of $\equiv_0$ and the winning condition of the DUPLICATOR. To see (iv) implies (ii), recall that every atomic formula formula $\varphi(\vec{x})$ is of the form $t_1 = t_2$ or $R(t_1, \dots, t_r)$ for some predicate $R \in \tau$, where t_i is a term, i.e. either a constant $c^{\mathbb{A}}$ for some constant symbol $c \in \tau$ or a free variable. It is trivial to verify the following.

- $\mathbb{A} \models (t_1 = t_2)(a_i, a_j)$ if and only if $\mathbb{B} \models (t_1 = t_2)(b_i, b_j)$, which implies that the first two conditions of partial isomorphism between $(\vec{c}^A, \vec{a}_0)$ to $(\vec{c}^{\mathbb{B}}, \vec{b}_0)$ are satisfied,
- $\mathbb{A} \models R(t_1, \dots, t_r)(a_{i_1}, \dots, a_{i_r})$ if and only if $\mathbb{B} \models R(t_1, \dots, t_r)(b_{i_1}, \dots, b_{i_r})$, which implies that the third condition of partial isomorphism is satisfied.

That (ii) implies (iv) can be similarly verified. The equivalence between (iii) and (iv) is immediate from that any quantifier-free formula is a boolean combination of atomic formulas. $\square$

Note that Lemma 11 is stated for any s. Using this as the base case we prove Theorem 10 by induction on n.

Proof of Theorem 10: The base case is proved in Lemma 11 and we assume that I and II are equivalent for $n \geq 0$ as an induction hypothesis.

Observe that $(\mathbb{A}, \vec{a}_0) \equiv_{n+1} (\mathbb{B}, \vec{b}_0)$ is equivalent to that the following two conditions hold:

- **Forth:** for every SPOILER's move $a \in \mathbb{A}$, there exists a DUPLICATOR's move $b \in \mathbb{B}$ such that the DUPLICATOR has a winning strategy for the n-round EF game on $(\mathbb{A}, \vec{a}_0 \circ a, \mathbb{B}, \vec{b}_0 \circ b)$ (i.e. with $|\vec{a}_0| + 1$-configuration expanded by a and b respectively).
- **Back:** for every SPOILER's move $b \in \mathbb{B}$, there exists a DUPLICATOR's move $a \in \mathbb{A}$ such that the DUPLICATOR has a winning strategy for the n-round EF game on $(\mathbb{A}, \vec{a}_0 \circ a, \mathbb{B}, \vec{b}_0 \circ b)$.

We state this observation more succinctly below.

Claim 2. *The followings are equivalent for every $n \geq 0$ and $\vec{a}_0 \in \mathbb{A}^s, \vec{b}_0 \in \mathbb{B}^s$.*

1. $(\mathbb{A}, \vec{a}_0) \equiv_{n+1} (\mathbb{B}, \vec{b}_0)$.
2. *For every $a \in \mathbb{A}$, there exists $b \in B$ such that $(\mathbb{A}, \vec{a}_0 \circ a) \equiv_n (\mathbb{B}, \vec{b}_0 \circ b)$. (The* **forth** *condition holds). Conversely, every $b \in \mathbb{B}$, there exists $a \in A$ such that $(\mathbb{A}, \vec{a}_0 \circ a) \equiv_n (\mathbb{B}, \vec{b}_0 \circ b)$. (The* **back** *condition holds).*

($\Leftarrow$) Let us show that (II) implies (I). Suppose $\mathsf{fo\text{-}type}_{n+1}(\mathbb{A}, \vec{a}_0) = \mathsf{fo\text{-}type}_{n+1}(\mathbb{B}, \vec{b}_0)$ and consider an arbitrary move of the SPOILER in an $n+1$-round EF game on $(\mathbb{A}, \vec{a}_0, \mathbb{B}, \vec{b}_0)$. Without loss of generality, we assume that the SPOILER picked $a \in \mathbb{A}$. We claim that there exists a DUPLICATOR's response $b \in B$ such that $(\mathbb{A}, \vec{a}_0 \circ a) \equiv_n (\mathbb{B}, \vec{b}_0 \circ b)$

Consider the $\mathsf{fo\text{-}type}_n(\mathbb{A}, \vec{a}_0 \circ a)$. By (3)-(a) of Lemma 3 there exists an FO-formula $\varphi(\vec{x})$ of quantifier rank n (with $s+1$ free variables) which defines the type $\mathsf{fo\text{-}type}_n(\mathbb{A}, \vec{a}_0 \circ a)$. Moreover, note that $\mathbb{A} \models \psi(\vec{a}_0)$, where

$$\psi(\vec{x}_0) := \exists x \varphi(\vec{x}_0, x).$$

As ψ is a formula in FO$[n+1]$, the precondition (I) on quantifier rank $n+1$, implies that $\mathbb{B} \models \psi(\vec{b}_0)$. Therefore, there exists $b \in B$ such that $\mathbb{B} \models \varphi(\vec{b}_0, b)$. By the construction of φ, we derive that

$$\mathsf{fo\text{-}type}_n(\mathbb{A}, \vec{a}_0 \circ a) = \mathsf{fo\text{-}type}_n(\mathbb{B}, \vec{b}_0 \circ b).$$

Now we apply the induction hypothesis and observe $(\mathbb{A}, \vec{a} \circ a) \equiv_n (\mathbb{B}, \vec{b} \circ b)$, as claimed earlier. That is, the forth condition of Claim 2 is met. As we can symmetrically demonstrate the back condition, this establishes that (I) holds by Claim 2.

($\Rightarrow$) Let us show that (I) implies (II). Suppose $(\mathbb{A}, \vec{a}_0) \equiv_{n+1} (\mathbb{B}, \vec{b}_0)$. By induction hypothesis, it holds that $\mathsf{fo\text{-}type}_n(\mathbb{A}, \vec{a}_0) = \mathsf{fo\text{-}type}_n(\mathbb{B}, \vec{b}_0)$. Therefore it suffices to show that for every $\varphi(\vec{x})$ of quantifier rank $n+1$ with s variables, $\mathbb{A} \models \varphi(\vec{a})$ if and only if $\mathbb{B} \models \varphi(\vec{b})$. We prove the forward implication; the backward implication can be shown analogously. The key technical step is the following.

Claim 3. *Let $\varphi(\vec{x})$ be an* FO*-formula of quantifier rank $n+1$ with s free variables in the form $\varphi(\vec{x}) := \exists x\ \psi(\vec{x}, x)$. If $(\mathbb{A}, \vec{a}_0) \equiv_{n+1} (\mathbb{B}, \vec{b}_0)$ and $\mathbb{A} \models \varphi(\vec{a}_0)$, then it holds that $\mathbb{B} \models \varphi(\vec{b}_0)$.*

PROOF OF THE CLAIM: Recall how we evaluate $\varphi(\vec{x}) := \exists x\ \psi(\vec{x}, x)$ on $(\mathbb{A}, \vec{a}_0)$; it holds on $(\mathbb{A}, \vec{a}_0)$ if and only if there is an assignment of x to some element $a \in \mathbb{A}$ such that $\mathbb{A} \models \psi(\vec{a}_0), a)$. Consider the $n+1$-round EF game on $(\mathbb{A}, \vec{a}_0, \mathbb{B}, \vec{b}_b)$ in which the first move if the SPOILER to pick the very element $a \in \mathbb{A}$. By the precondition $(\mathbb{A}, \vec{a}_0) \equiv_{n+1} (\mathbb{B}, \vec{b}_0)$ and the forth condition in Claim 2, there exists a matching move $b \in \mathbb{B}$ by the DUPLICATOR such that $(\mathbb{A}, \vec{a}_0 \circ a) \equiv_n (\mathbb{B}, \vec{b}_0 \circ b)$. As ψ has quantifier rank (at most) n, the induction hypothesis holds and $\mathbb{B} \models \psi(\vec{b}_0, b)$. Equivalently, φ holds on $(\mathbb{B}, \vec{b}_0)$ as desired. $\Diamond$

By Claim 3, an FO-formula in FO$[n+1]$ which begins with an existential quantifier holds on $(\mathbb{A}, \vec{a}_0)$ if and only if it holds on $(\mathbb{B}, \vec{b}_0)$. It remains to observe that an FO-formula of quantifier rank $n+1$ can be recursively constructed as a boolean combination of formulas such that each basic formula either introduces an existential quantifier followed by a formula in FO$[n]$ or it is a formula in FO$[n]$ itself. Therefore (II) holds. $\square$

5 EF game for MSO logic

Can we design a similar game such that the DUPLICATOR having a winning strategy is equivalent to two structures having the same MSO rank types? It is possible with a slight generalization of the EF game (for FO).

In the n-round EF game forMSO logic, we distinguish between the *point move* and the *set move.* At each round, when the SPOILER chooses a structure between $\mathbb{A}$ and $\mathbb{B}$, it also chooses whether it will make a point move or set move. A point move is to choose a single element, and a set move is to choose a subset of elements in the chosen structure. Notice that EF game (for FO) is nothing but an EF game for MSO in which all moves are point moves.

To present the winning condition of the DUPLICATOR, one needs to define the notion of partial isomorphism which take the set moves into consideration. Note that the only difference is the last item, in which we check if the element picked at i-th point move belongs to a set chosen at j-th set move.

Definition 12 (Partial isomorphism for MSO)**.** *Let $\mathbb{A}$ and $\mathbb{B}$ be two τ-structures and let $\vec{a} \in A^n$ and $\vec{b} \in A^n$ be two n-tuples of elements in $\mathbb{A}$ and $\mathbb{B}$, let $\vec{U} \in (2^{\mathbb{A}})^m$ and $\vec{V} \in (2^V)^m$ be two m-tuples of sets in $\mathbb{A}$ and $\mathbb{B}$ respectively respectively. We say that $(\vec{a}, \vec{U}, \vec{b}, \vec{V})$ defines a* partial isomorphism between $\mathbb{A}$ and $\mathbb{B}$ *if the following holds.*

- *for every $i, j \leq \ell$, $a_i = a_j$ if and only if $b_i = b_j$*
- *for every constant symbol $c \in \tau$ and for every $i \leq \ell$, $a_i = c^{\mathbb{A}}$ if and only if $b_i = c^{\mathbb{B}}$.*
- *for every predicate $R \in \tau$ with $\mathsf{ar}(R) = k$ and for every k-tuple $(i_1, \ldots, i_k)$ of indices (not necessarily distinct, not necessarily in order) from $[n]$, $(a_{i_1}, \ldots, a_{i_k}) \in R^{\mathbb{A}}$ if and only if $(b_{i_1}, \ldots, b_{i_k}) \in R^{\mathbb{B}}$.*
- *for every $i \leq n$ and $j \leq m$, $a_i \in U_j$ if and only if $b_i \in V_j$.*

An n-round EF game allowing a set move is represented as $(\vec{a}, \vec{U}, \vec{b}, \vec{V})$, where $\vec{a}$ and $\vec{b}$ are tuples of the same length consisting of elements of $\mathbb{A}$ and $\mathbb{B}$ respectively, and $\vec{U}$ and $\vec{V}$ are tuples of the same length consisting of set of $\mathbb{A}$ and B, such that $|\vec{a}| + |\vec{U}| = n$. The DUPLICATOR wins in an n-round EF game $(\vec{a}, \vec{U}, \vec{b}, \vec{V})$ if $(\vec{a}, \vec{U}, \vec{b}, \vec{V})$ defines a *partial isomorphism between* $\mathbb{A}$ *and* $\mathbb{B}$. When the DUPLICATOR has a winning strategy for an n-round EF game for MSO logic, i.e. for every SPOILER's move there is a DUPLICATOR move so that after n-round the DUPLICATOR wins, we write

$$\mathbb{A} \equiv_n^{\text{MSO}} \mathbb{B}.$$

Recall that we expressed via s-configuration the EF game in which the history of the game played already is given in the form of two s-tuples of elements in $\mathbb{A}$ and $\mathbb{B}$. Similarly, the game history played so far can be presented in the form of (ℓ, m)-configuration $(\vec{a}_0, \vec{U}_0, \vec{b}_0, \vec{V}_0)$ for two ℓ-tuples $\vec{a}_0$ and $\vec{b}_0$ of elements from $\mathbb{A}$ and $\mathbb{B}$ respectively, and two m-tuples $\vec{U}_0$ and $\vec{V}_0$ of sets from $\mathbb{A}$ and $\mathbb{B}$ respectively. If the DUPLICATOR has a winning strategy for an EF game on (ℓ, m)-configuration $(\vec{a}_0, \vec{U}_0, \vec{b}_0, \vec{V}_0)$, or equivalently on $(\mathbb{A}, \vec{a}_0, \vec{U}_0, \mathbb{B}, \vec{b}_0, \vec{V}_0)$, we write

$$(\mathbb{A}, \vec{a}_0, \vec{U}_0) \equiv_n^{\text{MSO}} (\mathbb{B}, \vec{b}_0, \vec{V}_0).$$

It is straightforward to verify that the proofs in the previous lecture notes generalize when you replace the EF game for FO by EF game for MSO and so on. We state Ehrenfeucht-Fraïssé Theorem for MSO.

Theorem 13 (Ehrenfeucht-Fraïssé Theorem for MSO)**.** *[1, 2] Let $\mathbb{A}$ and $\mathbb{B}$ be two τ-structures, and let $\vec{a}_0 \in \mathbb{A}^s$ and $\vec{b}_0 \in \mathbb{B}^s$ for some $s \geq 0$. The followings are equivalent.*

I. $(\mathbb{A}, \vec{a}_0, \vec{U}_0) \equiv_n (\mathbb{B}, \vec{b}_0, \vec{V}_0)$.

II. $\mathsf{mso\text{-}type}_n(\mathbb{A}, \vec{a}_0, \vec{U}_0) = \mathsf{mso\text{-}type}_n(\mathbb{B}, \vec{b}_0, \vec{V}_0)$.

The deferred proof of compositionality of MSO logic can be proved easily with Theorem 13.

Lemma 14. *Let $\mathbb{A}_1, \mathbb{A}_2, \mathbb{B}_1, \mathbb{B}_2$ be τ-structures such that* $\mathsf{mso\text{-}type}_q(\mathbb{A}_1) = \mathsf{mso\text{-}type}_q(\mathbb{B}_1)$ *and* $\mathsf{mso\text{-}type}_q(\mathbb{A}_2) = \mathsf{mso\text{-}type}_q(\mathbb{B}_2)$*. Then it holds that* $\mathsf{mso\text{-}type}_q(\mathbb{A}_1 \dot{\cup} \mathbb{A}_2) = \mathsf{mso\text{-}type}_q(\mathbb{B}_1 \dot{\cup} \mathbb{B}_2)$.

Proof: By Theorem 4, it holds that $\mathbb{A}_1 \equiv_q \mathbb{B}_1$ and $\mathbb{A}_2 \equiv_q \mathbb{B}_2$. It suffices to prove that $\mathbb{A}_1 \dot{\cup} \mathbb{A}_2 \equiv_q \mathbb{B}_1 \dot{\cup} \mathbb{B}_2$. Consider a q-round EF game $\vec{U} \cdot \vec{a} := (U_1, \ldots, U_\ell, a_{\ell+1}, \ldots, a_q)$ and $\vec{V} \cdot \vec{b} := (V_1, \ldots, V_\ell, b_{\ell+1}, \ldots, b_q)$ on $(\mathbb{A}, \mathbb{B}) := (\mathbb{A}_1 \dot{\cup} \mathbb{A}_2, \mathbb{B}_1 \dot{\cup} \mathbb{B}_2)$ in which the DUPLICATOR chooses $b \in \mathbb{B}_i$ (resp. $a \in \mathbb{A}_i$) whenever the SPOILER chooses $a \in \mathbb{A}_i$ (resp. $b \in \mathbb{B}_i$) for $i = 1, 2$ and the same for the set moves. Moreover, the DUPLICATOR chooses an element $b \in \mathbb{B}_i$ in a way the *subgame* restricted to $(\mathbb{A}_i, \mathbb{B}_i)$, the element b is a response to the choice $a \in \mathbb{A}_i$ of the SPOILER(the same with the set moves). In other words, whenever the SPOILERone of $i = 1, 2$, and some element, the DUPLICATOR simulates the EF game on $(\mathbb{A}_i, \mathbb{B}_i)$. It is tedious to verify that this is the winning strategy for the DUPLICATOR. □

Lemma 15 (MSO is compositional under concatenation)**.** *Let s_i, s'_i for $i = 1, 2$ be two strings over the alphabet Σ. If*

$$\mathsf{mso\text{-}type}_q(s_1) = \mathsf{mso\text{-}type}_k(s_2) \quad \textit{and} \quad \mathsf{mso\text{-}type}_q(s'_1) = \mathsf{mso\text{-}type}_q(s'_2),$$

then it holds that $\mathsf{mso\text{-}type}_q(s_1 \cdot s'_1) = \mathsf{mso\text{-}type}_q(s_2 \cdot s'_2)$

References

[1] Roland Fraïssé. *Sur quelques classifications des systèmes de relations*. Thèses présentées à la Faculté des Sciences de l'Université de Paris. impr. Durand, 1955.

[2] Andrzej Ehrenfeucht. An application of games to the completeness problem for formalized theories. *Fundamenta Mathematicae*, 49:129–141, 1961.