Topics in Logic and Complexity
Handout 6

Anuj Dawar

MPhil Advanced Computer Science, Lent 2010

Expressive Power of First-Order Logic

We noted that there are computationally easy properties that are not definable in first-order logic.

- There is no sentence ϕ of first-order logic such that $\mathbb{A} \models \phi$ if, and only if, $|A|$ is even.
- There is no sentence ϕ that defines exactly the *connected* graphs.

How do we *prove* these facts?

Our next aim is to develop the tools that enable such proofs.

Quantifier Rank

The *quantifier rank* of a formula ϕ, written $\text{qr}(\phi)$ is defined inductively as follows:

1. if ϕ is atomic then $\text{qr}(\phi) = 0$,
2. if $\phi = \neg\psi$ then $\text{qr}(\phi) = \text{qr}(\psi)$,
3. if $\phi = \psi_1 \vee \psi_2$ or $\phi = \psi_1 \wedge \psi_2$ then $\text{qr}(\phi) = \max(\text{qr}(\psi_1), \text{qr}(\psi_2))$.
4. if $\phi = \exists x\psi$ or $\phi = \forall x\psi$ then $\text{qr}(\phi) = \text{qr}(\psi) + 1$

More informally, $\text{qr}(\phi)$ is the *maximum depth of nesting of quantifiers* inside ϕ.

Formulas of Bounded Quantifier Rank

Note: For the rest of this lecture, we assume that our signature consists only of relation and constant symbols. That is, there are *no function symbols of non-zero arity*.

With this proviso, it is easily proved that in a finite vocabulary, for each q, there are (up to logical equivalence) only finitely many sentences ϕ with $\text{qr}(\phi) \leq q$.

To be precise, we prove by induction on q that for all m, there are only finitely many formulas of quantifier rank q with at most m free variables.

Formulas of Bounded Quantifier Rank

If $\mathrm{qr}(\phi) = 0$ then ϕ is a Boolean combination of atomic formulas. If it is has m variables, it is equivalent to a formula using the variables $x_1, \ldots, x_m$. There are finitely many formulas, *up to logical equivalence*.

Suppose $\mathrm{qr}(\phi) = q + 1$ and the *free variables* of ϕ are among $x_1, \ldots, x_m$. Then ϕ is a Boolean combination of formulas of the form

$$\exists x_{m+1} \psi$$

where ψ is a formula with $\mathrm{qr}(\psi) = q$ and free variables $x_1, \ldots, x_m, x_{m+1}$.

By induction hypothesis, there are only finitely many such formulas, and therefore finitely many Boolean combinations.

Equivalence Relation

For two structures $\mathbb{A}$ and $\mathbb{B}$, we say $\mathbb{A} \equiv_q \mathbb{B}$ if for any sentence ϕ with $\mathrm{qr}(\phi) \leq q$,

$$\mathbb{A} \models \phi \text{ if, and only if, } \mathbb{B} \models \phi.$$

More generally, if $\mathbf{a}$ and $\mathbf{b}$ are m-tuples of elements from $\mathbb{A}$ and $\mathbb{B}$ respectively, then we write $(\mathbb{A}, \mathbf{a}) \equiv_q (\mathbb{B}, \mathbf{b})$ if for any formula ϕ with m free variables $\mathrm{qr}(\phi) \leq q$,

$$\mathbb{A} \models \phi[\mathbf{a}] \text{ if, and only if, } \mathbb{B} \models \phi[\mathbf{b}].$$

Partial Isomorphisms

A map f is a partial isomorphism between structures $\mathbb{A}$ and $\mathbb{B}$, if

- the domain of $f = \{a_1, \ldots, a_l\} \subseteq A$, including the interpretation of all constants;
- the range of $f = \{b_1, \ldots, b_l\} \subseteq B$, including the interpretation of all constants; and
- f is an isomorphism between its domain and range.

Note that if f is a partial isomorphism taking a tuple $\mathbf{a}$ to a tuple $\mathbf{b}$, then for any *quantifier-free* formula θ

$$\mathbb{A} \models \theta[\mathbf{a}] \text{ if, and only if, } \mathbb{B} \models \theta[\mathbf{b}].$$

Ehrenfeucht-Fraïssé Games

The q-round Ehrenfeucht game on structures $\mathbb{A}$ and $\mathbb{B}$ proceeds as follows:

- There are two players called Spoiler and Duplicator.
- At the ith round, Spoiler chooses one of the structures (say $\mathbb{B}$) and one of the elements of that structure (say b_i).
- Duplicator must respond with an element of the other structure (say a_i).
- If, after q rounds, the map $a_i \mapsto b_i$ is a partial isomorphism, then Duplicator has won the game, otherwise Spoiler has won.

Equivalence and Games

Write $\mathbb{A} \sim_q \mathbb{B}$ to denote the fact that *Duplicator* has a *winning strategy* in the q-round Ehrenfeucht game on $\mathbb{A}$ and $\mathbb{B}$.

The relation $\sim_q$ is, in fact, an *equivalence relation*.

Theorem (Fraïssé 1954; Ehrenfeucht 1961)
$\mathbb{A} \sim_q \mathbb{B}$ if, and only if, $\mathbb{A} \equiv_q \mathbb{B}$

While one direction $\mathbb{A} \sim_q \mathbb{B} \Rightarrow \mathbb{A} \equiv_q \mathbb{B}$ is true for an arbitrary vocabulary, the other direction assumes that the vocabulary is *finite* and has *no function symbols*.

Proof

To prove $\mathbb{A} \sim_q \mathbb{B} \Rightarrow \mathbb{A} \equiv_q \mathbb{B}$, it suffices to show that if there is a sentence ϕ with $\mathrm{qr}(\phi) \leq q$ such that

$$\mathbb{A} \models \phi \quad \text{and} \quad \mathbb{B} \not\models \phi$$

then *Spoiler* has a winning strategy in the q-round Ehrenfeucht game on $\mathbb{A}$ and $\mathbb{B}$.

Assume that ϕ is in *negation normal form*, i.e. all negations are in front of atomic formulas.

Proof

We prove by induction on q the stronger statement that if ϕ is a formula with $\mathrm{qr}(\phi) \leq q$ and $\mathbf{a} = (a_1, \dots, a_m)$ and $\mathbf{b} = (b_1, \dots, b_m)$ are tuples of elements from $\mathbb{A}$ and $\mathbb{B}$ respectively such that

$$\mathbb{A} \models \phi[\mathbf{a}] \quad \text{and} \quad \mathbb{B} \not\models \phi[\mathbf{b}]$$

then *Spoiler* has a winning strategy in the q-round Ehrenfeucht game which starts from a position in which $a_1, \dots, a_m$ and $b_1, \dots, b_m$ have *already been selected*.

Proof

When $q = 0$, ϕ is a quantifier-free formula. Thus, if

$$\mathbb{A} \models \phi[\mathbf{a}] \quad \text{and} \quad \mathbb{B} \not\models \phi[\mathbf{b}]$$

there is an *atomic* formula θ that distinguishes the two tuples and therefore the map taking $\mathbf{a}$ to $\mathbf{b}$ is not a *partial isomorphism*.

When $q = p + 1$, there is a subformula θ of ϕ of the form $\exists x \psi$ or $\forall x \psi$ such that $\mathrm{qr}(\psi) \leq p$ and

$$\mathbb{A} \models \theta[\mathbf{a}] \quad \text{and} \quad \mathbb{B} \not\models \theta[\mathbf{b}]$$

If $\theta = \exists x \psi$, *Spoiler* chooses a witness for x in $\mathbb{A}$.

If $\theta = \forall x \psi$, $\mathbb{B} \models \exists x \neg \psi$ and *Spoiler* chooses a witness for x in $\mathbb{B}$.

Using Games

To show that a class of structures S is not definable in FO, we find, for every q, a pair of structures $\mathbb{A}_q$ and $\mathbb{B}_q$ such that

- $\mathbb{A}_q \in S$, $\mathbb{B}_q \in \overline{S}$; and
- *Duplicator* wins a q-round game on $\mathbb{A}_q$ and $\mathbb{B}_q$.

This shows that S is not closed under the relation $\equiv_q$ for *any* q.

Fact:

> S is definable by a first order sentence if, and only if, S is closed under the relation $\equiv_q$ for some q.

The direction from right to left requires a *finite, function-free* vocabulary.

Evenness

Let $\mathbb{A}$ be a structure in the *empty vocabulary* with q elements and $\mathbb{B}$ be a structure with $q + 1$ elements.

Then, it is easy to see that $\mathbb{A} \sim_q \mathbb{B}$.

It follows that there is no first-order sentence that defines the structures with an even number of elements.

If $S \subseteq \mathbb{N}$ is a set such that

$$\{\mathbb{A} \mid |\mathbb{A}| \in S\}$$

is definable by a first-order sentence then S is finite or co-finite.

Linear Orders

Let L_n denote the structure in one binary relation $\leq$ which is a linear order of n elements. Then $L_6 \not\equiv_3 L_7$ but $L_7 \equiv_3 L_8$.

In general, for $m, n \geq 2^p - 1$,

$$L_m \equiv_p L_n$$

Duplicator's strategy is to maintain the following condition after r rounds of the game:

for $1 \leq i < j \leq r$,

- *either* $\text{length}(a_i, a_j) = \text{length}(b_i, b_j)$
- *or* $\text{length}(a_i, a_j), \text{length}(b_i, b_j) \geq 2^{p-r} - 1$

Evenness is not first order definable, even on linear orders.

Reading List for this Handout

1. Ebbinghaus and Flum. Chapter 2.
2. Libkin. Chapter 3.
3. Grädel et al. Section 2.3.