9.1 Skolemization: Getting rid of the $\exists$'s

Idea: Once φ is in RPF, we will convert φ to it's Skolemization: φ_S. To do this, we replace existentially quantified variables with a new function symbol applied to the variables that have been previously universally quantified.

We will see:

Theorem 9.1 (Trevor & Skolem's Theorem) *For any $\varphi \in \mathcal{L}(\Sigma)$, we can construct universal formula $\varphi_S \in \mathcal{L}(\Sigma')$ such that φ and φ_S are equi-satisifiable, i.e., $\varphi \in$ FO-SAT $\Leftrightarrow \varphi_S \in$ FO-SAT. (Σ' is the result of adding some new Skolem function symbols to Σ.)*

α_S will usually not be equivalent to α, but it will be the case that $(\alpha_S \to \alpha) \in$ FO-VALID.

Example 9.2 Let $\text{od2}(x) \equiv \exists y \exists z \forall w (E(x,y) \wedge E(x,z) \wedge y \neq z \wedge (\neg E(x,w) \vee w = y \vee w = z))$. The formula $\text{od2}(x)$ has x as a free variable and says that vertex x has out-degree 2. To Skolemize $\text{od2}(x)$, we will get rid of the "$\exists y$" and replace all occurrences of y by the term $f_y(x)$ where f_y is a new unary function symbol. Similarly for "$\exists z$"

$\text{od2}(x) \equiv \forall w (E(x, f_y(x)) \wedge E(x, f_z(x)) \wedge f_y(x) \neq f_z(x) \wedge (\neg E(x,w) \vee w = f_y(x) \vee w = f_z(x)))$. $\square$

Example 9.3 $\alpha \equiv \exists x \forall y (x + y = y)$. Note that α says that there is an identity element for addition.

$\alpha_S \equiv \forall y (c + y = y)$. This says that c is the identity element for addition. Note that α_S implies α and it is more specific: not ony is there an identity element for addition, but we pick a particular one and give it a name, the new Skolem constant symbol, c. $\square$

Example 9.4 $\beta \equiv \forall x \exists y\, E(x,y)$. β says that every vertex has an outgoing edge.

$\beta_S \equiv \forall x\, E(x, f(x))$

Here f is a new unary Skolem function symbol. Observe that $\beta_S \to \beta$. β_S is more specific than β. Not only does every vertex have at least one outgoing edge, but the function f picks one. $\square$

For any $\varphi \in \mathcal{L}(\Sigma)$, we produce $\varphi_S \in \mathcal{L}(\Sigma')$ which is universal and in RPF-CNF form and is equi-satisfiable with φ, i.e, $\varphi \in$ FO-SAT $\Leftrightarrow$ $\varphi_S \in$ FO-SAT.

$$\varphi_s \equiv \forall x_1 \cdots x_k \bigwedge_{i=1}^{a} \bigvee_{j=1}^{b} \ell_{ij}$$

9.2 Herbrand Theory:

We will assume that our vocabulary, Σ, always includes at least one constant symbol, c. For example, let $\Sigma = (R_1^{a_1}, \ldots, R_s^{a_s}; c, c_2, \ldots, c_k, f_1^{r_1}, \ldots, f_t^{r_t})$. recall that we defined $\text{term}(\Sigma)$ by induction. A term is *closed* if it has no variables. Since every vocabulary has the constant symbol, c, $c \in \text{closedTerm}(\Sigma)$, so $\text{closedTerm}(\Sigma) \neq \emptyset$.

Example 9.5 $\Sigma_{\text{st-graph}} = (E^2; s, t)$. $\text{closedTerm}(\Sigma_{\text{st-graph}}) = \{s, t\}$. □

Example 9.6 **Example:** $\Sigma_{\#\text{thy}} = (\le^2; 0, 1, +^2[\text{infix}], *^2[\text{infix}])$. $\text{closedTerm}(\Sigma_{\#\text{thy}}) = \{0, 1, 0+0, 0+1, 0*0, (1+1)*(1+1), (1+1)*(1+1)+1, \ldots\}$. □

Recall that each term $t \in \text{term}(\Sigma)$ is a sequence of symbols. Each structure $\mathcal{A} \in \text{STRUC}[\Sigma]$ interprets the term t as an element of its universe: $t^{\mathcal{A}} \in |\mathcal{A}|$.

Definition 9.7 The *Herbrand Universe* for $(\Sigma) \stackrel{\text{def}}{=} \text{closedTerm}(\Sigma)$. □

Definition 9.8 **Def.** A *Herbrand Structure* $\mathcal{H} \in \text{STRUC}[\Sigma]$ has the Herbrand Universe as its universe, i.e., $|\mathcal{H}| = \text{closedTerm}(\Sigma)$. Furthermore, for every function symbol $f^r \in \Sigma$, and every $t_1, \ldots, t_r \in |\mathcal{H}|$, we have,

Gizem's Condition: $\quad f^{\mathcal{H}}(t_1, \ldots, t_r) \;=\; f(t_1, \ldots, t_r).$ □

Proposition 9.9 *Let $\mathcal{H}$ be a Herbrand Structure of vocabulary Σ and let t be any closed term of Σ. Then $t^{\mathcal{H}} = t$.*

Proof: By induction on t. □

Definition 9.10 Let $\varphi \in \mathcal{L}(\Sigma)$. If $\mathcal{H}$ is a Herbrand structure and $\mathcal{H} \models \varphi$ then we say that $\mathcal{H}$ is a **Herbrand model** of φ. □

Example 9.11 $\Sigma_{\text{st-graph}}$ has exactly 16 Herbrand structures: $\mathcal{H}_0, \ldots, \mathcal{H}_{15}$ all with universe $\{s, t\}$, and with the 16 possible interpretations of E on a two-element universe. For example, let $E^{\mathcal{H}_2} = \{(t, s)\}$. Note that $\mathcal{H}_2 \models \forall x\, (\neg E(x, x))$. Thus $\mathcal{H}_2$ is a Herbrand model of $\forall x\, (\neg E(x, x))$. □

Let $\beta = \exists x \exists y\, (R(x) \land \neg R(y))$. Let $\mathcal{A}_2 = (\{0, 1\}, R^{\mathcal{A}_2} = \{1\})$. Note that $\mathcal{A}_2 \models \beta$. However, $\mathcal{H}_0 \models \neg\beta$ and $\mathcal{H}_1 \models \neg\beta$. Thus, all the Herbrand structures of vocabulary Σ_0 satisfy $\neg\beta$. Thus, β is satisfiable, but has no Herbrand model with vocabulary Σ_0

Theorem 9.12 (Herbrand's Theorem) *Let Σ be a vocabulary s.t. $\text{closedTerm}(\Sigma) \neq \emptyset$. Let $\varphi \in \mathcal{L}(\Sigma)$ be a universal sentence, $\varphi = \forall x_1 \cdots x_k(\alpha)$ where α is quantifier free. Then*

$$(\varphi \in \text{FO-SAT}) \quad \Leftrightarrow \quad (\varphi \text{ has a Herbrand model})\ .$$

Proof: $\Leftarrow$: If φ has a Herbrand model, $\mathcal{H} \models \varphi$, then φ has a model, so it is satisfiable.

$\Rightarrow$: Assume that $\varphi \in \text{FO-SAT}$ and let $\mathcal{A} \models \varphi$.

Goal: define a Herbrand structure, $\mathcal{H}$, s.t. $\mathcal{H} \models \varphi$.

Two parts of building $\mathcal{H}$ are trivial, i.e., we must have that $|\mathcal{H}| = \text{closedTerm}(\Sigma)$ and for all $f^r \in \Sigma$ and $t_1, \ldots, t_r, \in \text{closedTerm}(\Sigma)$, $f^{\mathcal{H}}(t_1, \ldots, t_r) = f(t_1, \ldots, t_r)$.

What remains to be defined is $R^{\mathcal{H}}$ for each $R^a \in \Sigma$. For these definitions, we just ask $\mathcal{A}$:

$$R^{\mathcal{H}} \stackrel{\text{def}}{=} \left\{(t_1, \ldots, t_a) \in |\mathcal{H}|^a \;\middle|\; \mathcal{A} \models R(t_1, \ldots, t_a)\right\}\ . \qquad (*)$$

Claim. For all closed quantifier-free $\beta \in \mathcal{L}(\Sigma)$, $\quad \mathcal{H} \models \beta \Leftrightarrow \mathcal{A} \models \beta$.

Proof: By induction on β.

base case: $\beta = R(t_1, \ldots, t_a)$. By (*), $\mathcal{H} \models \beta \Leftrightarrow \mathcal{A} \models \beta$.

inductive case: Assume true for β_1 and β_2.

$$\mathcal{H} \models \neg\beta_1 \Leftrightarrow \mathcal{H} \not\models \beta_1 \Leftrightarrow \mathcal{A} \not\models \beta_1 \Leftrightarrow \mathcal{A} \models \neg\beta_1$$

$$\mathcal{H} \models \beta_1 \vee \beta_a \Leftrightarrow (\mathcal{H} \models \beta_1) \text{ or } (\mathcal{H} \models \beta_2) \Leftrightarrow (\mathcal{A} \models \beta_1) \text{ or } (\mathcal{A} \models \beta_2) \Leftrightarrow \mathcal{A} \models \beta_1 \vee \beta_2 \ .$$

□

Finally we show that $\mathcal{H} \models \forall x_1 \cdots x_k \ (\alpha)$.

Since $\mathcal{A} \models \forall x_1 \cdots x_k \ (\alpha)$, it follows that for all closed terms, $t_1, \ldots, t_k$, $\quad \mathcal{A}[t_1^{\mathcal{A}}/x_1, \ldots, t_k^{\mathcal{A}}/x_k] \models \alpha$.

Thus, by the Translation Lemma, for all closed terms, $t_1, \ldots, t_k$, $\quad \mathcal{A} \models \alpha[t_1^{\mathcal{A}}/x_1, \ldots, t_k^{\mathcal{A}}/x_k]$.

Thus, by our Claim, for all closed terms, $t_1, \ldots, t_k$, $\quad \mathcal{H} \models \alpha[t_1^{\mathcal{A}}/x_1, \ldots, t_k^{\mathcal{A}}/x_k]$.

Thus, by the Translation Lemma, for all closed terms, $t_1, \ldots, t_k$, $\quad \mathcal{H}[t_1/x_1, \ldots, t_k/x_k] \models \alpha$.

Thus, for all $t_1, \ldots, t_k \in |\mathcal{H}|$, $\quad \mathcal{H}[t_1/x_1, \ldots, t_k/x_k] \models \alpha$.

Thus, by Tarski's Definition of Truth, $\mathcal{H} \models \forall x_1 \cdots x_k \ (\alpha)$. □