Towards proof automation: Herbrand’s Theorem and Skolemization

Benjamin Wack
Paolo Torrini

Université Grenoble Alpes

March 2021
Reminder about expansion

<table>
<thead>
<tr>
<th>Every man is mortal.</th>
<th>$\forall x (\text{man}(x) \Rightarrow \text{mortal}(x))$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socrates is a man.</td>
<td>$\land \text{man}(\text{Socrates})$</td>
</tr>
<tr>
<td>Hence Socrates is mortal.</td>
<td>$\Rightarrow \text{mortal}(\text{Socrates})$</td>
</tr>
</tbody>
</table>

- Look for a counter-model using a 1-expansion then a 2-exp.
  - 1-expansion :
    $$(\text{man}(0) \Rightarrow \text{mortal}(0)).\text{man}(\text{Socrates}) \Rightarrow \text{mortal}(\text{Socrates})$$
    We can only interpret Socrates as 0 : no counter-model.
  - 2-expansion :
    $$(\text{man}(0) \Rightarrow \text{mortal}(0)).(\text{man}(1) \Rightarrow \text{mortal}(1)).\text{man}(\text{Socrates}) \Rightarrow \text{mortal}(\text{Socrates})$$
    We may interpret Socrates as 0 or 1, but neither yields a counter-model.

- What can you conclude ?
  Nothing! Except that this formula is satisfiable.
Overview

Introduction

Herbrand Universe (domain) and Herbrand Base

Herbrand Interpretation

Herbrand’s Theorem

Skolemization
  Motivation, properties and examples
  Definitions and procedure

Conclusion
Plan

Introduction

Herbrand Universe (domain) and Herbrand Base

Herbrand Interpretation

Herbrand’s Theorem

Skolemization
  Motivation, properties and examples
  Definitions and procedure

Conclusion
Introduction

In first-order logic, there is no algorithm for deciding whether a formula is valid or not.

Semi-decision algorithm:

1. If it terminates then it correctly decides whether the formula is valid or not. When the formula is valid, the decision generally comes with a proof.

2. If the formula is valid, then the program terminates. However, the execution can be long!

Note that if the formula is not valid, termination is not guaranteed.
Plan

Introduction

Herbrand Universe (domain) and Herbrand Base

Herbrand Interpretation

Herbrand’s Theorem

Skolemization
  Motivation, properties and examples
  Definitions and procedure

Conclusion
Universal closure

Definition 5.1.1

Let $C$ be a formula with free variables $x_1, \ldots, x_n$.

The **universal closure** of $C$, denoted by $\forall(C)$, is the formula
$\forall x_1 \ldots \forall x_n C$.

Example 5.1.2

$\forall(P(x) \land R(x, y)) =$

$\forall x \forall y (P(x) \land R(x, y))$ or $\forall y \forall x (P(x) \land R(x, y))$

Let $\Gamma$ be a set of formulae: $\forall(\Gamma) = \{ \forall(A) \mid A \in \Gamma \}$.
For example: $\forall(\{P(x), Q(x)\}) = \{ \forall x P(x), \forall x Q(x) \}$
Assumptions

We consider that

- the formulae do not contain neither $=,$ nor $\top$ or $\bot$ (since their truth value is fixed)
- every signature contains at least one constant (add an arbitrary constant $a$ if need be.)
Herbrand’s Theorem
Herbrand Universe (domain) and Herbrand Base

Herbrand universe (domain) and Herbrand base

**Definition 5.1.4**

1. The Herbrand universe $D_\Sigma$ is the set of closed terms (i.e., without variable) over $\Sigma$.

   **Remark:** this set is never empty, since $a \in D_\Sigma$.

2. The Herbrand base $B_\Sigma$ is the set of closed atomic formulae over $\Sigma$.

**Example 5.1.5**

1. Let $\Sigma = \{a^0, b^0, P^1, Q^1\}$: $D_\Sigma = \{a, b\}$ and

   \[ B_\Sigma = \{P(a), P(b), Q(a), Q(b)\}. \]

2. Let $\Sigma = \{a^0, f^1, P^1\}$: $D_\Sigma = \{f^n(a) \mid n \in \mathbb{N}\}$ and

   \[ B_\Sigma = \{P(f^n(a)) \mid n \in \mathbb{N}\}. \]
Plan

Introduction

Herbrand Universe (domain) and Herbrand Base

Herbrand Interpretation

Herbrand’s Theorem

Skolemization
  Motivation, properties and examples
  Definitions and procedure

Conclusion
## Herbrand Interpretation

### Definition 5.1.6

Let $E \subseteq B_\Sigma$.

The **Herbrand interpretation** $H_{\Sigma,E}$ consists of the domain $D_\Sigma$ and:

1. Constants symbols $s$ are mapped to themselves.
2. If $s$ is a function symbol and if $t_1, \ldots, t_n \in D_\Sigma$ then
   
   $s_{H_{\Sigma,E}}(t_1, \ldots, t_n) = s(t_1, \ldots, t_n)$.
3. If $s$ is a propositional variable, $s_{H_{\Sigma,E}} = 1$ (true) iff $s \in E$.
4. If $s$ is a relation symbol then
   
   $s_{H_{\Sigma,E}} = \{(t_1, \ldots, t_n) \mid s(t_1, \ldots, t_n) \in E\}$.

Another way to put it:

- Terms are interpreted as themselves.
- $E$ is the set of true atomic formulae.
Example 5.1.8

Let \( \Sigma = \{ f^0, b^0, Pr^1, Qr^1 \} \)

The Herbrand universe is \( D_\Sigma = \{ a, b \} \).

The set \( E = \{ P(b), Q(a) \} \) defines the Herbrand interpretation \( H \) where:

- constants \( a \) and \( b \) are mapped to themselves and
- \( P_H = \{ b \} \) and \( Q_H = \{ a \} \).
Universal closure and Herbrand model

**Theorem 5.1.16**

Let $\Gamma$ be a set of formulae with no quantifier over the signature $\Sigma$.

$$\forall(\Gamma) \text{ has a model if and only if } \forall(\Gamma) \text{ has a model which is a Herbrand interpretation.}$$

- Proof: Cf. handout course notes (just choose the “right” $E$)
- Consequence: no need to look for another model!
Example

Let $\Sigma = \{a^0, b^0, P^1, Q^1\}$

Let $I$ be the interpretation of domain $\{0, 1\}$ where:

- $a_I = 0, b_I = 1$,
- $P_I = \{1\}$ and $Q_I = \{0\}$.

The Herbrand universe is still $D_S = \{a, b\}$.

The set $E = \{P(b), Q(a)\}$ defines the Herbrand interpretation $H$ where:

- Constants $a$ and $b$ are mapped to themselves
- $P_H = \{b\}$ and $Q_H = \{a\}$.

$I$ is a model of a set $\forall(\Gamma)$ of formulae iff $H$ is a Herbrand model of $\forall(\Gamma)$. 
Plan

Introduction

Herbrand Universe (domain) and Herbrand Base

Herbrand Interpretation

Herbrand’s Theorem

Skolemization
  - Motivation, properties and examples
  - Definitions and procedure

Conclusion
Herbrand’s Theorem

Theorem 5.1.17

Let $\Gamma$ be a set of formulae with no quantifiers over signature $\Sigma$. 

\[ \forall (\Gamma) \text{ has a model if and only if} \]

Every finite set of closed instances of formulae of $\Gamma$ has a propositional model $B_\Sigma \rightarrow \{0, 1\}$.

Reminders:

- $\Sigma$ contains at least one constant $a$ and no $=$ sign
- Instantiate = substitute each variable by a term
Other version of Herbrand’s Theorem

Corollary 5.1.18

Let $\Gamma$ be a set of formulae without quantifier over signature $\Sigma$.

$\forall (\Gamma)$ is unsatisfiable

if and only if

There is a finite unsatisfiable set of closed instances of formulae taken from $\Gamma$

Proof.

Negate each side of the equivalence of the previous statement of Herbrand’s theorem.
Semi-decision procedure: unsatisfiability of $\forall(\Gamma)$

Let $\Gamma$ be a finite set of formulae with no quantifier.
We enumerate the set of closed instances of the formulae of $\Gamma$ and:

1. if we find an unsatisfiable set, then $\forall(\Gamma)$ is unsatisfiable.
2. if we have enumerated all of them without contradiction (for a $\Sigma$ without functions), then $\forall(\Gamma)$ is satisfiable.
3. in the meantime, we cannot conclude:
   - either $\forall(\Gamma)$ is satisfiable (and we will never stop);
   - or $\forall(\Gamma)$ is unsatisfiable but we haven’t enumerated enough instances to reach a contradiction.
Example 5.1.19 (1/5)

Let $\Gamma = \{ P(x), Q(x), \neg P(a) \lor \neg Q(b) \}$ and $\Sigma = \{ a^{f_0}, b^{f_0}, P^{r_1}, Q^{r_1} \}$.

$D_\Sigma = \{ a, b \}$.

The set $\{ P(a), Q(b), \neg P(a) \lor \neg Q(b) \}$ of instances over the $D_\Sigma$ is unsatisfiable, hence $\forall(\Gamma)$ is unsatisfiable.
Example 5.1.19 (2/5)

Let \( \Gamma = \{ P(x) \lor Q(x), \neg P(a), \neg Q(b) \} \)

The set of all the instances over \( D_\Sigma \) is:
\[
\{ P(a) \lor Q(a), P(b) \lor Q(b), \neg P(a), \neg Q(b) \}
\]
It has a propositional model characterised by \( E = \{ P(b), Q(a) \} \).

Hence the Herbrand interpretation associated to \( E \) is a model of \( \forall(\Gamma) \).
Example 5.1.19 (3/5)

Let $\Gamma = \{ P(x), \neg P(f(x)) \}$ and $\Sigma = \{ a^{f^0}, f^{f^1}, P^{r^1} \}$.

The set $\{ P(f(a)), \neg P(f(a)) \}$ is unsatisfiable, hence $\forall(\Gamma)$ is unsatisfiable.
Example 5.1.19 (4/5)

Let $\Gamma = \{ \neg P(a), \ P(x) \lor \neg P(f(x)), \ P(f(f(a))) \}$

\[
\begin{align*}
\neg P(a), \\
P(a) \lor \neg P(f(a)), \\
P(f(a)) \lor \neg P(f(f(a))), \\
P(f(f(a)))
\end{align*}
\]

is unsatisfiable, hence $\forall(\Gamma)$ too.

**Remark:** note that we had to consider 2 instances ($x := a$ then $x := f(a)$) of the second formula of $\Gamma$ to obtain a contradiction.
Example 5.1.19 (5/5)

Let \( \Gamma = \{ R(x, s(x)), R(x, y) \land R(y, z) \Rightarrow R(x, z), \neg R(x, x) \} \)
and \( \Sigma = \{ a^f_0, s^f_1, R^r_2 \} \).

\[
D_\Sigma = \{ s^n(a) \mid n \in \mathbb{N} \}. \text{ This is an infinite domain.}
\]

Every finite set of instances of formulae of \( \Gamma \) has a model: the enumeration will never stop.

Indeed, \( \forall (\Gamma) \) has an infinite model: the interpretation \( I \) of domain \( \mathbb{N} \) with \( a_I = 0, \quad s_I(n) = n + 1 \) and \( R(x, y) = x < y \).

Remark: \( \forall (\Gamma) \) has no finite model, i.e., it is useless to look for one by \( n \)-expansions.
Plan

Introduction

Herbrand Universe (domain) and Herbrand Base

Herbrand Interpretation

Herbrand’s Theorem

Skolemization
  Motivation, properties and examples
  Definitions and procedure

Conclusion
Introduction

Herbrand’s theorem applies to the universal closure of a set of formulae with no quantifier.

For formulae with existential quantification, we use skolemization (Thoralf Albert Skolem).

**Skolemization**

- transforms a set of closed formulae to the universal closure of a set of formulae with no quantifier.
- preserves the existence of a model (satisfiability).
Example 5.2.1

The formula $\exists x P(x)$ is skolemized as $P(a)$.

We note the following relations between the two formulae:

1. $\exists x P(x)$ is a consequence of $P(a)$
2. $P(a)$ is not a consequence of $\exists x P(x)$, but a model of $\exists x P(x)$
   “provides” a model of $P(a)$.

   (Just choose to interpret $a$ as an element of $P_i$.)
Definitions

A first-order formula is in **normal form** if it does not contain $\leftrightarrow$ nor $\Rightarrow$ and if its negations only apply to **atomic formulae**.

**Definition 5.2.3**

A closed formula is said to be **proper**, if no variable is bound by two distinct quantifiers.

**Example 5.2.4**

- The formula $\forall x P(x) \lor \forall x Q(x)$ is **not proper**.
- The formula $\forall x P(x) \lor \forall y Q(y)$ is **proper**.
- The formula $\forall x (P(x) \Rightarrow \exists x Q(x) \land \exists y R(x, y))$ is **not proper**.
- The formula $\forall x (P(x) \Rightarrow \exists y R(x, y))$ is **proper**.
How to skolemize a closed formula $A$?

Definition 5.2.5 (skolemization)

Let $A$ be a closed formula:

1. $B = \text{Normalize } A$
2. $C = \text{Make } B \text{ proper}$
3. $D = \text{Eliminate existential quantifiers from } C.$
   
   \textbf{This transformation only preserves the existence of a model.}

4. $E = \text{Remove the universal quantifiers from } D.$

$E$ is the \textbf{Skolem form of } $A$.

($E$ is a normal formula with no quantifier.)
1. Normalization

1. Eliminate the equivalences
2. Eliminate the implications
3. Move the negations towards the atomic formulae

Rules

1. et 2. As in propositional logic: \[
\begin{align*}
A \iff B & \equiv (A \implies B) \land (B \implies A) \\
A \implies B & \equiv \neg A \lor B
\end{align*}
\]

3. As in propositional logic: \[
\begin{align*}
\neg \neg A & \equiv A \\
\neg (A \land B) & \equiv \neg A \lor \neg B \\
\neg (A \lor B) & \equiv \neg A \land \neg B
\end{align*}
\]
Furthermore \[
\begin{align*}
\neg \forall x A & \equiv \exists x \neg A \\
\neg \exists x A & \equiv \forall x \neg A
\end{align*}
\]
Example 5.2.7

The normal form of $\forall y (\forall x P(x, y) \iff Q(y))$ is:

First, elimination of $\iff$:

$$\forall y ((\neg \forall x P(x, y) \lor Q(y)) \land (\neg Q(y) \lor \forall x P(x, y)))$$

then, move $\neg$:

$$\forall y ((\exists x \neg P(x, y) \lor Q(y)) \land (\neg Q(y) \lor \forall x P(x, y)))$$
Another example

Turn into normal form: \((\forall x (\exists y P(x, y) \iff Q(x))) \Rightarrow A:\)

Eliminate \iff: \((\forall x ((\exists y P(x, y) \Rightarrow Q(x)) \lor (Q(x) \Rightarrow \exists y P(x, y)))) \Rightarrow A\)

Eliminate \Rightarrow: \neg((\forall x (\neg \exists y P(x, y) \lor Q(x)) \land (\neg Q(x) \lor \exists y P(x, y))) \lor A\)

Pushing \neg inwards:
At \forall: \exists x ((\neg \exists y P(x, y) \lor Q(x)) \land (\neg Q(x) \lor \exists y P(x, y))) \lor A\)
At \land: \exists x ((\neg (\exists y P(x, y) \lor Q(x))) \lor (\neg Q(x) \lor \exists y P(x, y))) \lor A\)
At \lor: \exists x ((\neg \exists y P(x, y) \land \neg Q(x)) \lor (\neg Q(x) \land \neg \exists y P(x, y))) \lor A\)
At \exists, DNE: \exists x ((\exists y P(x, y) \land \neg Q(x)) \lor (Q(x) \land \forall y (\neg P(x, y)))) \lor A\)
2. Transformation to a proper formula

**Rename** bound variables, e.g., by choosing new names.

**Example 5.2.8**

- The formula $\forall x P(x) \lor \forall x Q(x)$ is changed to
  $$\forall x P(x) \lor \forall y Q(y)$$

- The formula $\forall x (P(x) \Rightarrow \exists x Q(x) \land \exists y R(x, y))$ is changed to
  $$\forall x (P(x) \Rightarrow \exists z Q(z) \land \exists y R(x, y))$$
3. Elimination of existential quantifiers

Let $\exists y B$ be a sub-formula of a closed normal and proper formula $A$. Let $x_1, \ldots, x_n$ be the free variables of $\exists y B$.

Let $f$ be a new symbol (if $n = 0$, then $f$ is a constant) and replace $\exists y B$ by $B < y := f(x_1, \ldots, x_n) >$ in $A$.

**Theorem 5.2.9**

The resulting formula $A'$ is a closed, normal and proper formula such that:

1. $A$ is a consequence of $A'$
2. If $A$ has a model then $A'$ has an identical model (up to the truth value of $f$).
Remark 5.2.10

The resulting formula $A'$ remains closed, normal and proper.

Hence, by repeatedly “applying” the theorem, choosing a new symbol for each eliminated quantifier, one can get:

- a closed, normal, proper formula $B$ without $\exists$
- such that $A$ has a model if and only if $B$ has one.
Example 5.2.11

By eliminating existential quantifiers in the formula
\[ \exists x \forall y P(x, y) \land \exists z \forall u \neg P(z, u) \]
we obtain
\[ \forall y P(a, y) \land \forall u \neg P(b, u). \]
It is easy to observe that this formula has a model.

But if we mistakenly eliminate both \( \exists \) using the same constant \( a \), we obtain
\[ \forall y P(a, y) \land \forall u \neg P(a, u) \]
which is unsatisfiable (it entails \( P(a, a) \) and \( \neg P(a, a) \)).
Exemple 5.2.12

By eliminating the existential quantifiers in the formula $\exists x \forall y \exists z P(x, y, z)$ we obtain two possible solutions:

- **is we eliminate first $\exists x$:**
  $$\forall y \exists z P(a, y, z) \rightarrow \forall y P(a, y, f(y))$$

- **if we eliminate first $\exists z$:**
  $$\exists x \forall y P(x, y, g(x, y)) \rightarrow \forall y P(b, y, g(b, y))$$

The existence of a model is preserved in both cases.
4. Transformation into a universal closure

Theorem 5.2.13

Let $A$ be a closed, normal, proper formula without existential quantifier. Let $B$ be the formula obtained by removing all the $\forall$ from $A$.

$A$ is equivalent to $\forall(B)$.

Proof.

What we are doing is actually applying repeatedly replacements such as

$\begin{align*}
\forall x (C) \land D & \equiv \forall x (C \land D) \\
\forall x (C) \lor D & \equiv \forall x (C \lor D)
\end{align*}$

where $x$ is not free in $D$
Property of skolemization

Property 5.2.14

Let $A$ be a closed formula and $E$ the Skolem form of $A$. $A$ has a model if and only if $\forall(E)$ has a model.

Proof.

\[
\begin{align*}
A \text{ a closed formula} & \Downarrow \quad \text{Normalize} \quad \text{(equivalent)} \\
B & \Downarrow \quad \text{Make proper} \quad \text{(equivalent)} \\
C & \Downarrow \quad \text{Eliminate } \exists \quad \text{("preserves" the models)} \\
D & \Downarrow \quad \text{Remove } \forall \quad \text{(equivalent to } \forall(E)) \\
E \text{ Skolem form} &
\end{align*}
\]
Example 5.2.15

Let $A = \forall x(P(x) \Rightarrow Q(x)) \Rightarrow (\forall x P(x) \Rightarrow \forall x Q(x))$. We skolemize $\neg A$.

1. $\neg A$ is transformed into the normal formula:
   $\forall x(\neg P(x) \vee Q(x)) \land \forall x P(x) \land \exists x \neg Q(x)$

2. The normal formula is made proper:
   $\forall x(\neg P(x) \vee Q(x)) \land \forall y P(y) \land \exists z \neg Q(z)$

3. The existential quantifier is “replaced” by a constant:
   $\forall x(\neg P(x) \vee Q(x)) \land \forall y P(y) \land \neg Q(a)$

4. The universal quantifiers are removed:
   $(\neg P(x) \vee Q(x)) \land P(y) \land \neg Q(a)$.

The instantiation $x := a, y := a$ yields $(\neg P(a) \vee Q(a)) \land P(a) \land \neg Q(a)$.

Hence (Herbrand’s theorem) the Skolem form of $\neg A$ is unsatisfiable.

Since skolemization preserves satisfiability, $\neg A$ is unsatisfiable.
Plan

Introduction

Herbrand Universe (domain) and Herbrand Base

Herbrand Interpretation

Herbrand’s Theorem

Skolemization
   Motivation, properties and examples
   Definitions and procedure

Conclusion
Today

- To prove that \( A \) is **satisfiable**:
  - Look for a (finite) model by \( n \)-expansions

- To prove that \( A \) est **unsatisfiable**:
  - **Skolemisation**
  - Look for a **(finite) unsatisfiable set of instances** over \( D_\Sigma \)
  - Herbrand’s theorem: then \( A \) is unsatisfiable too

- These methods are **non terminating and limited to finite interpretations**

- To find a counter-model or to prove the validity of \( A \), we proceed as before with \( \neg A \)
Next course

First-order deductive method:

- Clausal form
- Unification
- First-order resolution
- Consistency
- Completeness