Towards proof automation: Herbrand's Theorem and Skolemization

Benjamin Wack

Université Grenoble Alpes

March 2024

Reminder about expansion

Every man is mortal.	$\forall x(man(x) \Rightarrow mortal(x))$
Socrates is a man.	$\wedge man(Socrates)$
Hence Socrates is mortal.	$\Rightarrow mortal(Socrates)$

- Look for a counter-model using a 1-expansion then a 2-exp.
  - 1-expansion :
    $(man(0) \Rightarrow mortal(0)).man(Socrates) \Rightarrow mortal(Socrates)$
    We can only interpret *Socrates* as 0 : no counter-model.
  - 2-expansion : $(man(0) \Rightarrow mortal(0))$.
    $(man(1) \Rightarrow mortal(1)).man(Socrates) \Rightarrow mortal(Socrates)$
    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

Plan

Introduction

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

Semi-decision algorithm:

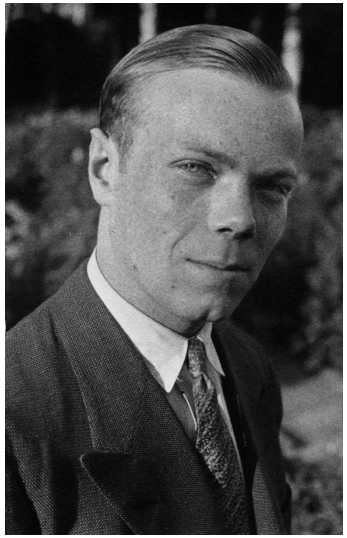
1. If it terminates then it correctly decides whether the formula is valid or not.
   When the formula is valid, the decision generally comes with a proof.
2. If the formula is valid, then the program terminates. However, the execution can be long!

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

Plan

Jacques Herbrand (1908-1931)

- Works in number fields
- 1930 : reduces the validity of a first-order formula to a set of propositional formulas
- Correspondence with Gödel about the consistency of arithmetic



Universal closure

Definition 5.1.1

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

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

Example 5.1.2

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

$$\boxed{\forall x \forall y (P(x) \wedge R(x,y)) \quad \text{or} \quad \forall y \forall x (P(x) \wedge R(x,y))}$$

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

Assumptions

We consider that

- the formulas 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 universe (domain) and Herbrand base

Definition 5.1.4

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

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

2. The Herbrand base $\boldsymbol{B_\Sigma}$ is the set of closed atomic formulas over Σ.

Example 5.1.5

1. Let $\Sigma = \{a^{f0}, b^{f0}, P^{r1}, Q^{r1}\}$: $D_\Sigma = \{a, b\}$ and

   $$\boxed{B_\Sigma = \{P(a), P(b), Q(a), Q(b)\}.}$$

2. Let $\Sigma = \{a^{f0}, f^{f1}, P^{r1}\}$: $D_\Sigma = \{f^n(a) \mid n \in \mathbb{N}\}$ and

   $$\boxed{B_\Sigma = \{P(f^n(a)) \mid n \in \mathbb{N}\}}$$

Plan

Herbrand Interpretation

Definition 5.1.6

A Herbrand interpretation H uses the domain D_Σ and:

- If t is a term, $[\![t]\!]_H = t$
  (terms are interpreted as themselves)
- In order to interpret formulas, it is sufficient to choose the set $E \subseteq B_\Sigma$ of true atomic closed formulas:
  $[s(t_1, \ldots, t_n)]_H = 1$ if and only if $s(t_1, \ldots, t_n) \in E$

(which amounts to setting the truth value of relations for any term in D_Σ)

Example 5.1.8

Let $\Sigma = \{a^{f0}, b^{f0}, P^{r1}, Q^{r1}\}$

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 Γ be a set of formulas with no quantifier over the signature Σ.

$\forall(\Gamma)$ has a model
if and only if
$\forall(\Gamma)$ 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 $\Sigma = \{a^{f0}, b^{f0}, P^{r1}, Q^{r1}\}$

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

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 formulas iff H is a Herbrand model of $\forall(\Gamma)$.

Plan

Herbrand's Theorem

Theorem 5.1.17

Let Γ be a set of formulas with no quantifiers over signature Σ.

$\forall(\Gamma)$ has a model

if and only if

Every finite set of formulas of Γ
instantiated by terms of D_Σ
has a propositional model $B_\Sigma \to \{0, 1\}$.

Reminders:

- Σ contains at least one constant a and no $=$ sign
- Instantiate = substitute each variable with a term

Other version of Herbrand's Theorem

Corollary 5.1.18

Let Γ be a set of formulas without quantifier over signature Σ.

$\forall(\Gamma)$ is unsatisfiable

if and only if

There is a finite unsatisfiable set
of closed instances of formulas taken from Γ

Proof.

Negate each side of the equivalence of the previous statement of Herbrand's theorem. $\square$

Semi-decision procedure: unsatisfiability of $\forall(\Gamma)$

Let Γ be a finite set of formulas with no quantifier.
We enumerate the set of closed instances of the formulas of Γ 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 Σ *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 $\Gamma = \{P(x), Q(x), \neg P(a) \vee \neg Q(b)\}$ and $\Sigma = \{a^{f0}, b^{f0}, P^{r1}, Q^{r1}\}$.

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

The set $\{P(a), Q(b), \neg P(a) \vee \neg Q(b)\}$ of instances over the D_Σ is unsatisfiable, hence $\forall(\Gamma)$ is unsatisfiable.

Example 5.1.19 (2/5)

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

The set of all the instances over D_Σ is :
$\{P(a) \vee Q(a), P(b) \vee Q(b), \neg P(a), \neg Q(b)\}$
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 $\Sigma = \{a^{f0}, f^{f1}, P^{r1}\}$.

$D_\Sigma = \{f^n(a) | n \in \mathbb{N}\}$.

$P(x) \ < x := f(a) >$ gives $P(f(a))$
and $\neg P(f(x)) \ < x := a >$ gives $\neg P(f(a))$

The set $\{P(f(a)), \neg P(f(a))\}$ is unsatisfiable,
hence $\forall(\Gamma)$ is unsatisfiable.

Example 5.1.19 (4/5)

Let $\Gamma = \left\{ \begin{array}{l} \neg P(a), \\ P(x) \vee \neg P(f(x)), \\ P(f(f(a))) \end{array} \right\}$

$\left\{ \begin{array}{l} \neg P(a), \\ P(a) \vee \neg P(f(a)), \\ P(f(a)) \vee \neg P(f(f(a))), \\ P(f(f(a))) \end{array} \right\}$ is unsatisfiable, hence $\forall(\Gamma)$ too.

Remark: notice that we had to consider 2 instances ($x := a$ then $x := f(a)$) of the second formula of Γ to obtain a contradiction.

Example 5.1.19 (5/5)

Let $\Gamma = \left\{ \begin{array}{l} R(x, s(x)), \\ R(x,y) \wedge R(y,z) \Rightarrow R(x,z), \\ \neg R(x,x) \end{array} \right\}$ $\begin{array}{l} n < n+1 \\ x < y < z \Rightarrow x < z \\ \neg(x < x) \end{array}$

and $\Sigma = \{a^{f0}, s^{f1}, R^{r2}\}$.

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

Every finite set of instances of formulas of Γ 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 $R_I(x,y) = x < y$.

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

Plan

Introduction

Herbrand's theorem applies to the universal closure of a set of formulas with no quantifier.

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

Skolemization

- removes the $\exists$ from a closed formula
- transforms the $\forall$ into a universal closure
- 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 formulas:

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 formulas**.

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 $\forall x P(x) \vee \forall x Q(x)$ is **not proper.**
- The formula $\forall x P(x) \vee \forall y Q(y)$ is **proper.**
- The formula $\forall x (P(x) \Rightarrow \exists x Q(x) \wedge \exists y R(x, y))$ 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 = Normalize A
2. C = Make B proper
3. D= **Eliminate existential quantifiers from** C
   This transformation only preserves the existence of a model.
4. E = Remove the universal quantifiers from D

E is the 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 formulas

Rules

1. et 2. As in propositional logic: $\begin{cases} A \Leftrightarrow B \equiv (A \Rightarrow B) \wedge (B \Rightarrow A) \\ A \Rightarrow B \equiv \neg A \vee B \end{cases}$

3. As in propositional logic: $\begin{cases} \neg\neg A \equiv A \\ \neg(A \wedge B) \equiv \neg A \vee \neg B \\ \neg(A \vee B) \equiv \neg A \wedge \neg B \end{cases}$

Furthermore $\begin{cases} \neg\forall x A \equiv \exists x \neg A \\ \neg\exists x A \equiv \forall x \neg A \end{cases}$

Example 5.2.7

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

First, elimination of $\Leftrightarrow$:

$$\forall y((\neg\forall x P(x,y) \vee Q(y)) \wedge (\neg Q(y) \vee \forall x P(x,y)))$$

then, move $\neg$:

$$\forall y((\exists x \neg P(x,y) \vee Q(y)) \wedge (\neg Q(y) \vee \forall x P(x,y)))$$

2. Transformation to a proper formula

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

Example 5.2.8

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

$$\boxed{\forall x P(x) \vee \forall y Q(y)}$$

- The formula $\forall x (P(x) \Rightarrow \exists x Q(x) \wedge \exists y R(x, y))$ is changed to

$$\boxed{\forall x (P(x) \Rightarrow \exists z Q(z) \wedge \exists y R(x, y))}$$

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 $B < y := f(x_1, \ldots x_n) >$.

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 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 $\exists x \forall y P(x, y) \wedge \exists z \forall u \neg P(z, u)$ we obtain

$\forall y P(a, y) \wedge \forall u \neg P(b, u)$.
It is easy to find a model.

But if we mistakenly eliminate both $\exists$ using the same constant a, we obtain $\forall y P(a, y) \wedge \forall u \neg P(a, u)$

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

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

- $(\forall x C) \wedge D \equiv \forall x(C \wedge D)$
- $(\forall x C) \vee D \equiv \forall x(C \vee D)$

where x is not free in D $\square$

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.

A a closed formula □

$\downarrow$ Normalize *(equivalent)*

B

$\downarrow$ Make proper *(equivalent)*

C

$\downarrow$ Eliminate $\exists$ *("preserves" the models)*

D

$\downarrow$ Remove $\forall$ *(equivalent to $\forall(E)$)*

E Skolem form

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:
   $\forall x(\neg P(x) \vee Q(x)) \wedge \forall x P(x) \wedge \exists x \neg Q(x)$
2. The normal formula is made proper:
   $\forall x(\neg P(x) \vee Q(x)) \wedge \forall y P(y) \wedge \exists z \neg Q(z)$
3. The existential quantifier is "replaced" by a constant:
   $\forall x(\neg P(x) \vee Q(x)) \wedge \forall y P(y) \wedge \neg Q(a)$
4. The universal quantifiers are removed:
   $(\neg P(x) \vee Q(x)) \wedge P(y) \wedge \neg Q(a)$.

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

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

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

Plan

Today

- To prove that A is satisfiable :
  - Search for a (finite) model by n-**expansions**
- To prove that A est unsatisfiable :
  - **Skolemisation**
  - Search for a **(finite) unsatisfiable set of instances** over D_Σ
  - Herbrand's theorem: then A is unsatisfiable too
    `https://www.cs.unm.edu/~mccune/mace4/examples/2009-11A/misc/index.html`
- 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