Base de la démonstration automatique : Théorème de Herbrand et Skolémisation

Benjamin Wack

Université Grenoble Alpes

Mars 2025

Rappel sur l'expansion

Tous les hommes sont mortels.	$\forall x(homme(x) \Rightarrow mortel(x))$
Socrate est un homme.	$\wedge homme(Socrate)$
Donc Socrate est mortel.	$\Rightarrow mortel(Socrate)$

- Rechercher un contre-modèle par 1-expansion puis par 2-exp.
  - 1-expansion :
    $(homme(0) \Rightarrow mortel(0)).homme(Socrate) \Rightarrow mortel(Socrate)$
    On est forcé d'interpréter *Socrate* comme 0 : pas de c.-m.
  - 2-expansion : $(homme(0) \Rightarrow mortel(0))$.
    $(homme(1) \Rightarrow mortel(1)).homme(Socrate) \Rightarrow mortel(Socrate)$
    On peut interpréter *Socrate* comme 0 ou 1, mais aucun ne donne de contre-modèle.
- Que peut-on en conclure ?
  Rien ! Si ce n'est que cette formule est satisfaisable.

Plan

Plan

Introduction

En logique du premier ordre, il n'y a pas d'algorithme pour décider si une formule est valide ou non valide.

On devra se contenter d'un algorithme de semi-décision :

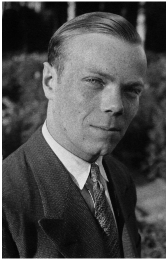
1. S'il termine alors il décide correctement si la formule est valide. Lorsque la formule est valide, la décision est accompagnée d'une preuve.
2. Si la formule est valide, alors il termine. Cependant, l'exécution peut être longue !

Si la formule n'est pas valide, la terminaison d'un tel algorithme n'est pas garantie.

Plan

Jacques Herbrand (1908-1931)

- Résultats en théorie des nombres
- 1930 : réduction de la validité d'une formule du premier ordre à un ensemble de formules propositionnelles
- Correspondance avec Gödel à propos de la cohérence de l'arithmétique



Fermeture universelle

Définition 5.1.1

Soit C une formule ayant pour variables libres $x_1, \ldots, x_n$.

La fermeture universelle de C, notée $\forall(C)$, est la formule $\forall x_1 \ldots \forall x_n C$.

Exemple 5.1.2

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

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

Soit Γ un ensemble de formules : $\forall(\Gamma) = \{\ \forall(A) \mid A \in \Gamma\ \}$.
Par exemple : $\forall(\ \{\ P(x), Q(x)\ \}\ ) = \{\forall x P(x), \forall x Q(x)\}$

Hypothèses

Nous ne considérons que :

- des formules sans $=$, $\top$ ni $\perp$ (car leur sens est fixé).
- des signatures qui comportent au moins une constante (quitte à ajouter **la constante *a***).

Domaine et base de Herbrand

Définition 5.1.4

1. Domaine de Herbrand $\boldsymbol{D_\Sigma}$ = ensemble des termes fermés de Σ (*i.e.* sans variable)
   Remarque : il n'est jamais vide, car $a \in D_\Sigma$.
2. Base de Herbrand $\boldsymbol{B_\Sigma}$ = ensemble des formules atomiques fermées de Σ

Exemple 5.1.5

1. Soit $\Sigma = \{a^{f0}, b^{f0}, P^{r1}, Q^{r1}\}$: $D_\Sigma = \{a, b\}$ et
   $\boxed{B_\Sigma = \{P(a), P(b), Q(a), Q(b)\}.}$
2. Soit $\Sigma = \{a^{f0}, f^{f1}, P^{r1}\}$: $D_\Sigma = \{f^n(a) \mid n \in \mathbb{N}\}$ et
   $\boxed{B_\Sigma = \{P(f^n(a)) \mid n \in \mathbb{N}\}}$

Plan

Interprétation de Herbrand

Définition 5.1.6

Une interprétation de Herbrand H a pour domaine D_Σ et :

- Si t est un terme, $[\![t]\!]_H = t$
  (les termes sont interprétés par eux-mêmes)
- Pour interpréter les formules, il suffit de choisir l'ensemble $E \subseteq B_\Sigma$ des formules atomiques fermées vraies :
  $[s(t_1, \ldots, t_n)]_H = 1$ si et seulement si $s(t_1, \ldots, t_n) \in E$

(Cela revient à fixer la valeur booléenne des relations pour chaque terme de D_Σ.)

Exemple 5.1.8

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

Le domaine de Herbrand est $D_\Sigma = \{a, b\}$.

$E = \{P(b), Q(a)\}$ définit l'interprétation de Herbrand H où :

- les constantes a et b ont pour valeur elles-mêmes
- $P_H = \{b\}$ et $Q_H = \{a\}$

Formule universelle et modèle de Herbrand

Théorème 5.1.16

Soit Γ un ensemble de formules sans quantificateur sur la signature Σ.

$\forall(\Gamma)$ a un modèle

si et seulement si

$\forall(\Gamma)$ a un modèle qui est une interprétation de Herbrand.

- Preuve : cf poly (il « suffit » de bien choisir E)
- Conséquence : pas la peine d'en chercher un autre !

Exemple

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

Soit I l'interprétation de domaine $\{0, 1\}$ où :

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

Le domaine de Herbrand est $D_\Sigma = \{a, b\}$.

$E = \{P(b), Q(a)\}$ définit l'interprétation de Herbrand H où :

- les constantes a et b ont pour valeur elles-mêmes
- $P_H = \{b\}$ et $Q_H = \{a\}$

I est modèle d'un ensemble $\forall(\Gamma)$ ssi H est un modèle de $\forall(\Gamma)$.

Plan

Théorème de Herbrand

Théorème 5.1.17

Soit Γ un ensemble de formules sans quantificateur de signature Σ.

$\forall(\Gamma)$ a un modèle

si et seulement si

Tout ensemble fini de formules de Γ
instanciées par des termes de D_Σ
admet un modèle propositionnel $B_\Sigma \to \{0, 1\}$.

Rappels :

- Σ comporte au moins une constante *a* et pas de signe $=$
- Instancier : substituer un terme à chaque variable

Variante du théorème de Herbrand

Corollaire 5.1.18

Soit Γ un ensemble de formules sans quantificateur de signature Σ.

$\forall(\Gamma)$ est insatisfaisable

si et seulement si

Il existe un ensemble fini insatisfaisable d'instances fermées des formules de Γ.

Preuve.

C'est la « contraposée » du théorème de Herbrand. $\square$

Procédure de semi-décision : insatisfaisabilité de $\forall(\Gamma)$

Soit Γ un ensemble fini de formules sans quantificateur.
Énumérer l'ensemble des instances fermées des formules de Γ et :

1. si on trouve un ensemble insatisfaisable, $\forall(\Gamma)$ est insatisfaisable.
2. si on termine sans contradiction (pour un Σ *sans fonction*), $\forall(\Gamma)$ est satisfaisable.
3. en attendant, on ne peut pas conclure :
   - soit $\forall(\Gamma)$ est satisfaisable (et on ne s'arrêtera jamais) ;
   - soit $\forall(\Gamma)$ est insatisfaisable, mais on n'a pas encore énuméré assez d'instances pour obtenir une contradiction.

Exemple 5.1.19 (1/5)

Soit $\Gamma = \{P(x), Q(x), \neg P(a) \vee \neg Q(b)\}$ et $\Sigma = \{a^{f0}, b^{f0}, P^{r1}, Q^{r1}\}$.

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

L'ensemble d'instances $\{P(a), Q(b), \neg P(a) \vee \neg Q(b)\}$
est insatisfaisable, donc $\forall(\Gamma)$ aussi.

Exemple 5.1.19 (2/5)

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

L'ensemble de toutes les instances sur D_Σ est :
$\{P(a) \vee Q(a),\ \ P(b) \vee Q(b),\ \ \neg P(a),\ \ \neg Q(b)\}$
Il a un modèle : $E = \{P(b), Q(a)\}$.

Donc $\forall(\Gamma)$ a un modèle (l'interprétation de Herbrand associée à E).

Exemple 5.1.19 (3/5)

Soit $\Gamma = \{P(x), \neg P(f(x))\}$ et $\Sigma = \{a^{f0}, f^{f1}, P^{r1}\}$.

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

$$\begin{array}{rl} P(x) & < x := f(a) > \text{donne } P(f(a)) \\ \text{et } \neg P(f(x)) & < x := a > \text{donne } \neg P(f(a)) \end{array}$$

L'ensemble $\{P(f(a)), \neg P(f(a))\}$ est insatisfaisable,
donc $\forall(\Gamma)$ est insatisfaisable.

Exemple 5.1.19 (4/5)

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

$\left\{ \begin{array}{l} \neg P(a), \\ P(a) \vee \neg P(f(a)), \\ P(f(a)) \vee \neg P(f(f(a))), \\ P(f(f(a))) \end{array} \right\}$ est insatisfaisable, donc $\forall(\Gamma)$ aussi.

Remarque : observez qu'il a fallu prendre 2 instances ($x := a$ puis $x := f(a)$) de la deuxième formule de Γ pour obtenir une contradiction.

Exemple 5.1.19 (5/5)

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

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

$D_\Sigma = \{s^n(a) \mid n \in \mathbb{N}\}$ (domaine infini)

Tout ensemble fini d'instances des formules de Γ a un modèle : leur énumération ne s'arrêtera jamais.

En effet, $\forall(\Gamma)$ a un modèle (infini) : l'interprétation I de domaine $\mathbb{N}$ avec $a_I = 0$, $s_I(n) = n+1$ et $R_I(x,y) = x < y$.

Remarque : $\forall(\Gamma)$ n'a aucun modèle fini (inutile d'en chercher un par n-expansion).

Plan

Pourquoi la Skolémisation ?

Le théorème de Herbrand s'applique à la fermeture universelle d'un ensemble de formules sans quantificateur.

Pour des formules avec quantificateur existentiel on utilise la skolémisation (Thoralf Albert Skolem).

La skolémisation

- élimine les $\exists$ d'une formule fermée
- transforme les $\forall$ en fermeture universelle
- préserve l'existence d'un modèle (la satisfaisabilité)

Exemple 5.2.1

La formule $\exists x P(x)$ est skolémisée en $P(a)$.

On observe les relations suivantes entre ces deux formules :

1. $P(a)$ a pour conséquence $\exists x P(x)$
2. $\exists x P(x)$ n'a pas pour conséquence $P(a)$
   mais un modèle de $\exists x\, P(x)$ « donne » un modèle de $P(a)$.
   (Il suffit d'interpréter a comme un élément satisfaisant P.)

Définitions

Définition

Une formule est en forme normale si elle n'a ni $\Leftrightarrow$ ni $\Rightarrow$ et les négations portent uniquement sur les **formules atomiques**.

Définition 5.2.3

Une formule fermée est propre si aucune variable n'est liée par deux quantificateurs distincts.

Exemple 5.2.4

- La formule $\forall x P(x) \vee \forall x Q(x)$ n'est **pas propre**.
- La formule $\forall x P(x) \vee \forall y Q(y)$ est **propre**.
- La formule $\forall x (P(x) \Rightarrow \exists x Q(x) \wedge \exists y R(x, y))$ n'est **pas propre**.
- La formule $\forall x (P(x) \Rightarrow \exists y R(x, y))$ est **propre**.

Comment skolémiser une formule fermée A ?

Définition 5.2.5 (skolémisation)

Soit A une formule fermée :

1. B = Mettre A en forme normale
2. C = Mettre B en forme propre
3. D = **Éliminer les quantificateurs existentiels** de C
   (Cette transformation préserve seulement l'existence de modèle.)
4. E = Enlever les quantificateurs universels de D

E est la forme de Skolem de A.

E est une formule normale sans quantificateur.

1. Normalisation

1. Remplacer les équivalences
2. Remplacer les implications
3. Déplacer les négations vers les formules atomiques

Règles

1. et 2. Comme dans le cas propositionnel : $\begin{cases} A \Leftrightarrow B \equiv (A \Rightarrow B) \wedge (B \Rightarrow A) \\ A \Rightarrow B \equiv \neg A \vee B \end{cases}$

3. Comme dans le cas propositionnel : $\begin{cases} \neg\neg A \equiv A \\ \neg(A \wedge B) \equiv \neg A \vee \neg B \\ \neg(A \vee B) \equiv \neg A \wedge \neg B \end{cases}$

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

Exemple 5.2.7

La forme normale de $\forall y(\forall x P(x,y) \Leftrightarrow Q(y))$ est :

On remplace $\Leftrightarrow$:

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

puis on déplace $\neg$:

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

2. Transformation en formule propre

Renommer les variables liées, par exemple en choisissant de nouveaux noms.

Exemple 5.2.8

- La formule $\forall x P(x) \vee \forall x Q(x)$ est changée en
  $\boxed{\forall x P(x) \vee \forall y Q(y)}$
- La formule $\forall x (P(x) \Rightarrow \exists x Q(x) \wedge \exists y R(x,y))$ est changée en
  $\boxed{\forall x (P(x) \Rightarrow \exists z Q(z) \wedge \exists y R(x,y))}$

3. Élimination des quantificateurs existentiels

Soit $\exists y B$ une sous-formule d'une formule A fermée normale et propre.
Soient $x_1, \ldots x_n$ les variables libres de $\exists y B$.

On prend f un nouveau symbole (si $n = 0$, f est une constante)
et on remplace $\exists y B$ par $B < y := f(x_1, \ldots x_n) >$.

Théorème 5.2.9

La formule A' obtenue est fermée, normale, propre et vérifie :

1. A' a pour conséquence A
2. Si A a un modèle alors A' a un modèle identique (sauf pour f).

Remarque 5.2.10

La formule obtenue reste fermée, normale et propre.

Par application répétée en choisissant un **nouveau** symbole à chaque quantificateur éliminé, on obtient :

- une formule B fermée, normale, propre et sans $\exists$
- telle que A a un modèle si et seulement si B en a un.

Exemple 5.2.11

En éliminant les quantificateurs existentiels de la formule $\exists x \forall y P(x, y) \wedge \exists z \forall u \neg P(z, u)$ on obtient

$\forall y P(a, y) \wedge \forall u \neg P(b, u)$
Il est facile d'en trouver un modèle.

Mais si on fait l'erreur d'éliminer les deux $\exists$ avec la même constante a, on obtient $\forall y P(a, y) \wedge \forall u \neg P(a, u)$

qui est insatisfaisable.

Exemple 5.2.12

En éliminant les quantificateurs existentiels de la formule $\exists x \forall y \exists z P(x, y, z)$ nous obtenons deux solutions possibles :

- si on élimine $\exists x$ d'abord :
  $\forall y \exists z P(a, y, z) \quad \rightarrow \quad \forall y P(a, y, f(y))$
- si on élimine $\exists z$ d'abord :
  $\exists x \forall y P(x, y, g(x, y)) \quad \rightarrow \quad \forall y P(b, y, g(b, y))$

L'existence d'un modèle est préservée dans les deux cas.

4. Transformation en formule universelle

Théorème 5.2.13

Soit A une formule fermée, normale, propre et sans quantificateur existentiel.
Soit B la formule obtenue en enlevant tous les $\forall$.

A est équivalente à $\forall(B)$.

Démonstration.

Cela revient à effectuer tous les remplacements de la forme

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

où x est non libre dans D $\square$

Propriété de la skolémisation

Propriété 5.2.14

Soit A une formule fermée et E sa forme de Skolem :
A a un modèle si et seulement si $\forall(E)$ a un modèle.

Démonstration.

A une formule fermée		
$\downarrow$	Normaliser	*(équivalente)*
B		
$\downarrow$	Rendre propre	*(équivalente)*
C		
$\downarrow$	Élim. $\exists$	*("préserve" les modèles)*
D		
$\downarrow$	Enlever $\forall$	*(équivalente à $\forall(E)$)*
E forme de Skolem		

□

Exemple 5.2.15 : prouvons que A est valide.

Soit $A = \forall x(P(x) \Rightarrow Q(x)) \Rightarrow (\forall x P(x) \Rightarrow \forall x Q(x))$. On skolémise $\neg A$.

1. $\neg A$ est transformée en la formule normale :
   $\forall x(\neg P(x) \vee Q(x)) \wedge \forall x P(x) \wedge \exists x \neg Q(x)$
2. La formule normale est transformée en la formule propre :
   $\forall x(\neg P(x) \vee Q(x)) \wedge \forall y P(y) \wedge \exists z \neg Q(z)$
3. Le quantificateur existentiel est « remplacé » par une constante :
   $\forall x(\neg P(x) \vee Q(x)) \wedge \forall y P(y) \wedge \neg Q(a)$
4. Les quantificateurs universels sont enlevés :
   $(\neg P(x) \vee Q(x)) \wedge P(y) \wedge \neg Q(a)$.

L'instanciation $x := a, y := a$ donne $(\neg P(a) \vee Q(a)) \wedge P(a) \wedge \neg Q(a)$.

Donc (théorème de Herbrand) la forme de Skolem de $\neg A$ est insatisfaisable.

Puisque la skolémisation préserve la satisfaisabilité, $\neg A$ est insatisfaisable.

Plan

Aujourd'hui

Rappel : pour montrer que A est satisfaisable :

- Recherche d'un modèle (fini) par n-**expansions**

- Pour montrer que A est insatisfaisable :
  - **Skolemisation**
  - Recherche d'un **ensemble (fini) insatisfaisable d'instances** sur D_{Σ}
    `https://www.cs.unm.edu/~mccune/mace4/examples/2009-11A/misc/index.html`
- Méthodes non terminantes et limitées aux interprétations finies
- Pour prouver la validité de A, on étudie plutôt $\neg A$.

La prochaine fois

Méthode de déduction au premier ordre :

- Forme clausale
- Unification
- Résolution au premier ordre