Cours 11

Base de démonstration automatique : Méthode de Herbrand

11.1 Complétude et décidabilité

IL est important de préciser que la logique du premier ordre est indécidable, cela signifie qu'il n'existe pas d'algorithme pour déterminer si une formule est valide ou non valide. Ce résultat fut établi par Alonzo Church et Alan Türing en 1936 et 1937. Ils ont aussi montré que cette logique est semi-décidable, c'est-à-dire que nous pouvons écrire un programme, qui prend en entrée une formule et qui a le comportement suivant :

1. S'il termine alors il décide correctement si la formule est valide ou non. Lorsque la formule est valide, la décision est généralement accompagnée d'une preuve.
2. Si la formule est valide, alors il termine. Cependant, l'exécution peut être longue ! Notons que si la formule n'est pas valide, la terminaison de ce programme n'est pas garantie. En effet si le programme terminait pour toute formule, alors, d'après le premier point, ce serait un algorithme pour déterminer si une formule est valide ou non. Or, ainsi qu'il est dit ci-dessus, Church et Türing ont prouvé qu'un tel algorithme n'existait pas.

Théorème 1. ***(Théorème de complétude Godel)*** *Le calcul des prédicats est complet : on peut donner un nombre fini de principes (axiomes logiques, schémas d'axiomes logiques et règles de déduction) qui suffisent pour déduire de façon mécanique toutes les formules universellement valide de la logique du premier ordre.*
Il existe en fait différents systèmes de preuves complets et corrects pour la logique du premier ordre, citons par exemple : • Le système de Hilbert. • la déduction naturelle • la résolution • le calcul des séquents

Théorème 2. *La logique du premier ordre n'est pas décidable, mais elle est semi-décidable.*

Bien qu'il existe un système de preuve correct et complet, le calcul des prédicats n'est pas décidable. Si par exemple on considère le calcul des séquents, on ne va pas pouvoir déterminer qu'une formule n'est pas universellement valide car il y a un nombre non borné d'explorations possibles. Toutefois, puisqu'une preuve est un objet fini et que tous les objets considérés sont dénombrables, et donc énumérables, il est possible d'énumérer toutes les preuves, et donc tous les théorèmes.

Le calcul des prédicats est donc semi-décidable : si on veut déterminer si une formule F est universellement valide(tautologie), il suffit de la comparer avec chacun des théorèmes : le calcul terminera si F est effectivement un théorème (mais dans un temps non borné), le calcul peut ne pas terminer mais dans ce cas, on n'a pas la certitude que F n'est pas un théorème. En effet, on ne sait pas si le calcul des théorèmes n'a pas été mené assez loin pour identifier F comme théorème ou si ce calcul ne se terminera pas parce que F n'est pas un théorème.

11.2 Méthodes de Herbrand

Jacques Herbrand (1908-1931) est un mathématicien et logicien français. Il a établi en 1930 un lien entre calcul des prédicats et calcul des propositions.

11.3 Univers de Herbrand

L'univers de Herbrand

Définition 1. *L'univers de Herbrand est stratifié en niveaux.*

- *H_0 est constitué par l'ensemble des constantes apparaissant dans l'ensemble de clauses S . Si aucune constante n'apparaît $H_0 = \{a\}$ où a est une constante choisie arbitrairement.*
- *On obtient H_{i+1} à partir de H_i en formant l'union de H_i et de l'ensemble de tous les termes de la forme $f_n(t_1, t_2, \ldots, t_n)$ où les t_i prennent leurs valeurs dans H_i et f_n un symbole de fonction d'arité n qui apparaît dans S.*
- *H_i est appelé l'ensemble des constantes de niveau i.*
- *$H_\infty = \cup_{i\geq 0} H_i$ est appelé univers de Herbrand ou domaine de Herbrand*

En bref, on peut dire que le domaine de Herbrand c'est l'ensemble des termes fermés (sans variables) définis sur la signature associée à l'ensemble de clauses S. Pour ne pas avoir d'ensemble vide, on ajoute la constante de Herbrand a à la signature si l'ensemble S ne contient pas de constantes.

Exemple 1. $S = \{P(a), \neg P(x) \vee P(f(x))\}$. $H_0 = \{a\}$. $H_1 = \{a, f(a)\}$.
$H_\infty = \{a, f(a), f(f(a)), f(f(f(a))), \ldots\} = \{f^n(a) | n \geq 0\}$.

Exemple 2. $S = \{P(x), R(x) \vee Q(y, x), \neg Q(y, y)\}$. *Pas de constante, on choisit arbitrairement a.*
$H_0 = \{a\}$. $H_1 = \{a\}$ *car il n'y a pas de fonction.* $H_\infty = \{a\}$.

Exemple 3. $S = \{P(f(x)) \vee Q(a), Q(g(b)) \vee \neg P(y)\}$.
$H_0 = \{a, b\}$. $H_1 = \{a, b, f(a), f(b), g(a), g(b)\}$.
$H_2 = \{a, b, f(a), f(b), g(a), g(b), f(f(a)), f(g(a)), g(f(a)), g(g(a)), f(f(b)), f(g(b)), g(f(b)), g(g(b))\}$.

11.4 Base de Herbrand

Base de Herbrand

Définition 2. *Soit S un ensemble de clauses, l'ensemble des formules atomiques de base (atomes de base) de la forme $P_n(t_1, t_2, \ldots, t_n)$ pour tout prédicat d'arité n apparaissant dans S où $t_1, t_2, \ldots, t_n$ sont des éléments de l'univers de Herbrand est appelé la base de Herbrand.*

Autrement dit, la base de Herbrand pour S est l'ensemble des `formules atomiques fermées`(sans variables) de la signature associée à S.

Exemple 4.
— $A_1 = \{P(a), P(f(a)), P(f(f(a))), \ldots\}$ *est la base de Herbrand de l'exemple 1.*
— $A_2 = \{P(a), R(a), Q(a,a)\}$ *est la base de Herbrand de l'exemple 2.*
— $A_3 = \{P(a), Q(a), P(b), Q(b), P(f(a)), Q(f(a)), \ldots\}$ *est la base de Herbrand de l'exemple 3.*

11.5 Instances de Base

Définition 3. *Une instance de base d'une clause C (clause fondamentale) d'un ensemble S est une clause obtenue en remplaçant les variables libres de C par des éléments de l'univers de Herbrand H_∞.*

Exemple 5. $C = R(x) \vee Q(y,x)$. $H_\infty = \{a\}$. *La clause* $C_1 = R(a) \vee Q(a,a)$ *est une instance de C*

On appelle déploiement de Herbrand de S l'ensemble de toutes les clauses issues de chaque $A \in S$. Cet ensemble est noté $H(S)$.

Exemple 6. $S = \{\neg P(x,a), Q(x) \vee P(b,y)\}$.
- *La signature associée est* $P^{r2}, Q^{r1}, a^{f0}, b^{f0}$.
- *L'univers de Herbrand* $H_\infty = \{a,b\}$.
- *La base de Herbrand de S est* $\{P(a,a), P(a,b), P(b,a), P(b,b), Q(a), Q(b)\}$.
- *Le déploiement de Herbrand*

$$H(S) = \{\neg P(a,a), \neg P(b,a), Q(a) \vee P(b,a), Q(a) \vee P(b,b), Q(b) \vee P(b,a), Q(b) \vee P(b,b)\}.$$

11.6 Interprétation de Herbrand

Définition 4. *Soit S un ensemble de clauses, soit H l'univers de Herbrand, I une interprétation dans H est dite H-interprétation ou interprétation de Herbrand si elle satisfait aux conditions suivantes :*

1. *Elle applique toutes les constantes individuelles sur elle-même*
2. *Elle applique chaque symbole de fonction f d'arité n une fonction h qui fait correspondre à n termes $h_1, h_2, \ldots, h_n$, éléments de l'univers de Herbrand H, le terme $f(h_1, h_2, \ldots, h_n)$.*

Remarque : Les H interprétations ne diffèrent que par l'affectation des symboles de prédicats.

Exemple 7. $A = \{P(a), Q(a), R(a), P(f(a)), Q(f(a)), R(f(a)), \ldots\}$
Une H interprétation I_1 est $I1 = \{V, V, V, F, F, F, \ldots\}$
ou $I1 = \{P(a), Q(a), R(a), \neg P(f(a)), \neg Q(f(a)), \neg R(f(a)), \ldots\}$

Exemple 8. $S = \{P(x) \vee \neg Q(f(x))\}$. $H_\infty = \{a, f(a), f((f(a)), \ldots\}$
$A = \{P(a), Q(a), P(f(a)), Q(f(a)), \ldots\}$
$I1 = \{\neg P(a), \neg Q(a), P(f(a)), \neg Q(f(a)), P(f(f(a)), \ldots\}$
$I2 = \{P(a), \neg Q(a), \neg P(f(a)), Q(f(a)), \ldots\}$

Exemple 9. $S = \{P(a) \vee \neg Q(f(x), R(x,y)\}$. *Soit l'interprétation I telle que :* $D_I = \{1,2\}$. *On définit a_I et f_I comme :*

a	f(1)	f(2)	P(1)	P(2)	Q(1)	Q(2)	R(1,1)	R(1,2)	R(2,1)	R(2,2)
2	2	2	V	F	V	V	F	V	F	F

Une H interprétation correspond à I avec $H = \{a, f(a), f(f(a)), \ldots\}$. *(domaine de Herbrand)*

$A = \{P(a), Q(a), R(a,a), P(f(a)), Q(f(a)), R(f(a),a), R(a,f(a)), \ldots\}$
$(P(a))_I = P(2) = F$; $(Q(a))_I = Q(2) = V$; $(R(a,a))_I = R(2,2) = F$; $(P(f(a)))_I = P(f(2)) = P(2) = F$;

$$IH = \{\neg P(a), Q(a), \neg R(a,a), \neg P(f(a)), \ldots\}.$$

Définition 5. *Étant donné une interprétation I sur un domaine D, une H interprétation I^* correspondant à I est une H interprétation qui satisfait les conditions suivantes :*

1. *Soient $h_1, h_2, \ldots, h_n$ des éléments de H à chaque h_i on affecte un $d_i \in D$,*
2. *Si $P(d_1, d_2, \ldots, d_n)$ est vrai (resp faux) dans I alors $P(h_1, h_2, \ldots, h_n)$ est vrai (resp faux) dans I^*.*

Lemme 1. *: Si une interprétation I sur un domaine D satisfait un ensemble S de clauses alors n'importe quelle interprétation de Herbrand I^* correspondant à I satisfait aussi S.*

Exemple 10. *Soit $S = \{P(x, f(x))\}$. Considérons l'interprétation $I : D = \{1, 2\}$. $f(1) = 1; f(2) = 2$. $P(1, 1) = V, P(1, 2) = F, P(2, 1) = F, P(2, 2) = V$. I est un modèle de S. $[\forall x P(x, f(x))]_I = P(1, f(1)) \wedge P(2, f(2)) = 1$.*

Soit I^* la H interprétation correspondante à I avec $a = 1$.
$H = \{a, f(a), f(f(a)), \ldots\}$
$A = \{P(a, a), P(a, f(a)), P(f(a), a), P(f(a), f(a)), \ldots\}$
$P(a, a) = P(1, 1) = V$. $P(a, f(a)) = P(1, 1) = V$. $P(f(a), a) = P(1, 1) = 1$. $P(f(a), f(a)) = P(1, 1) = 1, \ldots$
Donc I^* est modèle de S..
On peut considérer une autre H interprétation avec $a = 2$. Mais on vérifie que c'est aussi un modèle de S.

Théorème 3. ***Un ensemble de clauses S est insatisfiable ssi il est faux sous toutes les H interprétations.***

Eléments de preuve.

$\Rightarrow$ Si S est insatisfaisable, donc elle est fausse pour toutes les interprétations et donc en particulier pour les H interprétations.

$\Leftarrow$ Supposons S est fausse pour toutes les H interprétations mais S est satisfaisable. Alors il existe une interprétation modèle qui satisfait S. Par le lemme précédent il existe une H interprétation satisfait S. Ce qui contredit l'hypothèse de départ.

Définition 6. *Un ensemble de clauses est contradictoire ssi la conjonction de ces clauses est fausse dans tous les modèles c'est-à-dire pour toute interprétation et tout domaine.*

Grâce aux théorème précédent et son corollaire, on peut se limiter à l'examen du modèle de Herbrand composé de l'univers de Herbrand et de la H interprétation.

Ou dit autrement : Un ensemble de clauses est sémantiquement inconsistant (contradictoire) ssi la conjonction de ces clauses est fausse dans toutes les interprétations I^* des S.

Propriétés

1. Une clause C est satisfaite par une H interprétation ssi si toute instance de base est satisfaite par I
2. Une clause C est falsifiée pour une H interprétation I ssi il y a au moins une instance de base qui soit fausse dans cette interprétation
3. Un ensemble de clauses est insatisfiable si et seulement si pour toute H interprétation il y a au moins une instance de base ou une clause qui soit fausse

Exemple 11. *Soit $S = \{\neg P(x), P(a)\}$. S est-il satisfaisable ?*

11.7 Théorème de Herbrand

Théorème 4. *L'ensemble fini de clauses S admet un modèle équivaut à $H(S)$ est satisfaisable.*

Corollaire 5. *L'ensemble fini de clauses S est contradictoire si et seulement si l'ensemble H(S) est contradictoire.*

Notons que $H(S)$ est un ensemble de formules du calcul propositionnels. Par le théorème de compacité du calcul propositionnel, $H(S)$ est contradictoire ssi il existe un sous ensemble fini de $H(S)$ (un ensemble d'instances de base fini) qui est contradictoire. Autrement dit :

Théorème de Herbrand première formulation

Théorème 6.
Un ensemble S de clauses est insatisfiable ssi il existe un ensemble fini S' d'instances de base de clauses de S contradictoire.

Herbrand a aussi montré qu'il est possible de générer de manière automatique cet ensemble fini. Son théorème a donc un caractère constructif. Le théorème de Herbrand et son corollaire suggèrent une procédure pour démontrer ***insatisfaisabilité*** des formules sous formes clausales. Pour une formule F sous forme clausale, nous engendrons successivement les instances de clauses G_i des clauses de F et tester si leur conjonction est contradictoire.

Cette procédure permet de détecter si F est insatisfaisable après un nombre fini d'étapes. Dans le cas contraire si F est satisfaisable la procédure peut ne pas se terminer. C'est une procédure de décision partielle.

Donnons un exemple. Soit la formule :

$$\exists y \forall z (P(z, y) \Leftrightarrow \neg \exists x (P(z, x) \wedge P(x, z)))$$

Après transformations de on obtient la forme clausale :

$\boxed{C_1}$ $\neg P(z, a) \vee \neg P(z, x) \vee \neg P(x, z)$ $\boxed{C_2}$ $P(z, a) \vee P(z, f(z))$ $\boxed{C_3}$ $P(z, a) \vee P(f(z), z)$

L'ensemble $S(F)$ contient

1. $\neg P(a,a)$ On remplace dans C_1 x, z par a et on simplifie
2. $P(a,a) \vee P(f(a),a)$ On remplace dans C_2 z par a.
3. $P(a,a) \vee P(a,f(a))$. On remplace dans C_3 z par a.
4. $\neg P(f(a),a) \vee \neg P(a,f(a))$. On remplace z par $f(a)$ et x par a.

En notant $A = P(a,a), B = P(a,f(a))$ et $C = P(f(a),a)$, le sous ensemble ci-dessus n'est autre que : $\{\neg A, A \vee B, A \vee C, \neg B \vee \neg C\}$. On peut vérifier aisément que cet ensemble est contradictoire. La formule F est donc insatisfaisable.

11.8 Méthode de Herbrand

Le théorème de Herbrand donne une méthode (semi décidable) pour tester la validité d'une formule F
$\boxed{1.}$ prendre la négation de F $\boxed{2.}$ la mettre sous forme clausale
$\boxed{3.}$ générer un ensemble (fini) d'instances des clauses $\boxed{4.}$ tester que l'instance est non-satisfaisable.
$\boxed{5.}$ si elle est insatisfaisable, alors on s'arrête, sinon on recommence à l'étape 3.

11.9 Arbre sémantique

Les arbres sémantiques permettent l'examen de toutes les H interprétations Soit S un ensemble de clause et $A = \{A_1, A_2, \ldots, A_n, \ldots, \}$ la base de Herbrand de S. Un arbre sémantique T pour S est un arbre binaire tel que :

1. Les deux arcs sortant de chaque nœud de T sont étiquetés, l'un par une formule atomique de base, et l'autre par sa négation.

1. L'ensemble des littéraux étiquetant les arcs d'une même branche ne contient pas de littéraux complémentaire.

- Un arbre sémantique pour S est dit complet ssi en plus des conditions exigées de l'arbre sémantique la condition : Chaque branche de l'arbre sémantique contient pour tout atome de la base de Herbrand soit A_i soit $\neg A_i$.
- L'arbre sémantique complet de S est infini ssi S contient un symbole de fonction.
- Une branche correspond à la définition d'une H interprétation, un arbre sémantique complet correspond à l'examen de toutes les H interprétations

11.9.1 Noeud d'échec et noued d'inférence

Pour tout noeud N, Soit $I(N)$ l'union des littéraux étiquetant le chemin liant la racine au noeud N.

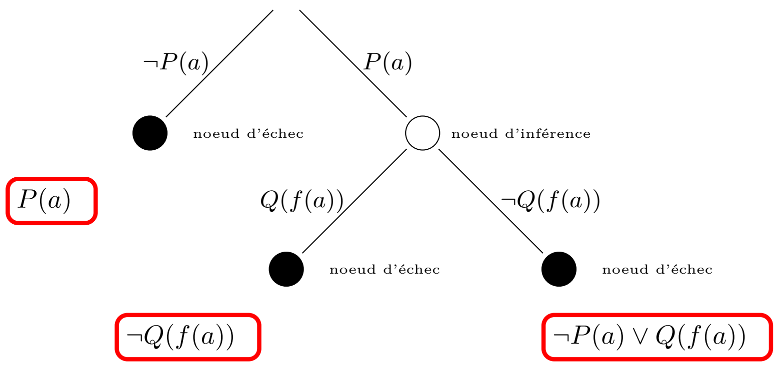
Définition 7 (nœud d'échec)**.** *Un nœud N est dit noeud d'échec si $I(N)$ falsifie une instance de base d'une clause de S et $I(N')$ pour tout N ancêtre de N ne falsifie aucune instance de base d'une clause de S*

Définition 8 (Arbre sémantique clos)**.**
Un arbre sémantique est clos ou fermé si et seulement si chacune de ses branches se termine par un noeud d'échec.

Définition 9. *Un noeud d'un arbre sémantique clos est dit noeud d'inférence si et seulement si tous les noeuds descendants immédiats de N sont des noeuds d'échec .*

Théorème 7 (Théorème de Herbrand)**.**
Un ensemble S de clauses fondamentales est insatisfiable si et seulement si à tout arbre sémantique complet correspond un arbre sémantique clos.

Exemple 12. $S = \{P(X), \neg P(X) \vee Q(f(X)), \neg Q(f(X))\}$; $\quad H_\infty = \{a, f(a), f((f(a)), \ldots\}$; $\quad A = \{P(a), Q(a), P(f(a)), Q(f(a)), \ldots\}$.



Remarque 1. *Les clauses sont des fermetures universelles. Rappelons que $\models A$ ssi $\models \forall x A$. Donc si on cherche à étudier la validité de F avec des variables libres, c'est équivalent à étudier la validité de sa fermeture universelle.*