Decision Procedures — Lecture on Logic — Oct 31 2007

Pete Manolios

These lecture notes on logic are mostly based on material from the book by Ebbinghaus, Flum, and Thomas entitled "Mathematical Logic."

1 Intro to Logic

Logic forms the foundation of mathematics. Let's start with an example. A *group* is a triple $\langle G, \circ, e\rangle$ such that

- (G1) For all x, y, z: $(x \circ y) \circ z = x \circ (y \circ z)$.
- (G2) For all x: $x \circ e = x$.
- (G3) For all x there is a y such that: $x \circ y = e$.

The following are groups: $\langle \mathbb{Z}, +, 0\rangle$ and $\langle \mathbb{R}, +, 0\rangle$. The following are not: $\langle \mathbb{N}, +, 0\rangle$ and $\langle \mathbb{R}, \cdot, 1\rangle$.

Here is a theorem about groups.

Theorem 1 *For every x, there is a y such that: $y \circ x = e$.*

The axioms only directly mention a right inverse, but the above claims that left inverses also exist.

Proof By (G3), there is a y such that $x \circ y = e$ and a z such that $y \circ z = e$. Taking associativity (G1) into account, we have

$y \circ x = y \circ x \circ e = y \circ x \circ y \circ z = y \circ e \circ z = y \circ z = e$ $\square$

This example already highlights many of the features of modern mathematics. In mathematics, we study the properties of various objects, *e.g.*, groups. The properties that these objects enjoy are captured with "non-logical" axioms, *e.g.*, in the case of group theory, (G1)-(G3). The theory of groups consists of all theorems that are derivable from the "non-logical axioms" via logical reasoning alone. This separation is really fundamental. We cannot appeal to intuition or "obvious truths" about groups (or geometry, or the reals, etc). So, what exactly is a "proof", then? This question naturally leads to computer science and historically that is what happened, as a proof has to be machine-checkable.

Other questions naturally arise. When we prove theorems about groups, then the results apply to every instance of a group, *e.g.*, $\langle \mathbb{Z}, +, 0\rangle$ and $\langle \mathbb{R}, +, 0\rangle$, but if some formula φ holds in every group (denoted $\{(G1), (G2), (G3)\} \models \varphi$), then does there necessarily exist a proof (denoted $\{(G1), (G2), (G3)\} \vdash \varphi$)? Note that proofs are finite, machine checkable things, whereas there are many groups; how many? By a result we will prove, the number of groups is uncountable (and in fact there are so many groups, that they do not even form a set, so we have no simple way of measuring them). We will see how to make this question precise, *i.e.*, we will present a simple proof theory. Then, we will see that for any set of sentences Φ and any sentence φ, $\Phi \models \varphi$ iff $\Phi \vdash \varphi$, (where $\Phi \vdash \varphi$ denotes that there is a proof of φ from Φ). This is Gödel's completeness theorem, perhaps the most important result in logic, as it relates syntax with semantics.

2 Syntax of FOL

When one presents a mathematical language to a mature audience, *e.g.*, a programming language, one starts with the syntax and then the semantics. The syntax tells us what markings, what sequences of symbols, belong to the language. If we were to think about programming languages, this corresponds to the syntax checker. We will insist that the problem of checking whether a sequence of symbols is syntactically well-formed is decidable, that is there exists a program that can say "yes" or "no" when presented with a sequence of symbols. Syntax can be presented using BNF or any other precise method. When presenting the syntax, there is no need to mention the meaning or semantics of the strings; all that we do is that we determine what is and what is not a "statement" in the language. The semantics, or meaning, is given later.

We do not want to look at a specific language, instead, we want to describe the syntax of any first-order language (FOL). All FOLs have the following in common.

Definition 1 *$\mathcal{A}$ contains the following symbols:*

1. $v_0, v_1, v_2, \ldots$ *(variables);*
2. $\neg, \wedge, \vee, \Rightarrow, \leftrightarrow$ *(boolean connectives);*
3. $\forall, \exists$ *(quantifiers);*
4. $\equiv$ *(equality symbol);*
5. $), ($ *(parenthesis);*

Depending on the first-order theory (FOT) in question, there may be other symbols in a FOL, *e.g.*, in the theory of groups we had $\circ$, a 2-ary function symbol and e, a constant. In set theory we have $\in$, a 2-ary relation symbol, and so on.

Definition 2 *The symbol set S of a FOL contains*

1. *for every $n \geq 1$ a (possibly empty) set of n-ary relation symbols.*
2. *for every $n \geq 1$ a (possibly empty) set of n-ary function symbols.*
3. *a (possibly empty) set of constant symbols.*

S may be empty and the symbols mentioned in the definition of S must be distinct from each other and from the symbols in $\mathcal{A}$. S determines a FOL and $\mathcal{A}_S := \mathcal{A} \cup S$ is the alphabet of this language.

We shall use the letters $P, Q, R, \ldots$ for relation symbols, $f, g, h, \ldots$ for function symbols, $c, c_0, c_1, \ldots$ for constants, and $x, y, z, \ldots$ for variables.

2.1 Terms

To motivate the definition of terms and formulas, let me give you a preview of the semantics. FOL are interpreted over structures, *e.g.*, in the FOL of groups, $\circ$ corresponds to group multiplication say of group G. Terms are expressions that denote elements of G. Formulas are expressions that make statements about G, *e.g.*, that all elements of a certain type have a certain property.

Definition 3 *The set of S-*terms*, denoted T^S is the least set closed under the following rules.*

1. Every variable is an S-term.
2. Every constant in S is an S-term.
3. If $t_1, \ldots, t_n$ are S-terms and f is an n-ary function symbol in S, then $ft_1 \ldots t_n$ is an S-term.

Note that $T^S \subseteq \mathcal{A}_S^*$.

Here is an analogy with English. Bill, the father of John, etc. all denote elements in our universe. Similarly, x, c, fxy, etc. denote elements of the universe of a first-order theory.

Note that parentheses are not used in terms. They are not needed and do not result in any ambiguity.

2.2 Formulas

Terms name objects in our domain, whereas formulas correspond to statements about our domain.

Recall our analogy with English. Bill, the father of John, etc. all denote elements in our universe. Similarly, x, c, fxy, etc. denote elements of the universe of a first-order theory.

Similarly, statements such as "Bob has three siblings" are statements about the universe. They are either true or false. That is the role played by formulas.

Definition 4 *An* atomic formula *of S is either of the form $t_1 \equiv t_2$ or $Rt_1 \ldots t_n$, where $t_1, t_2, \ldots, t_n$ are S-terms and R is an n-ary relation symbol in S.*

Definition 5 *The set of S-*formulas *is the least set closed under the following rules.*

1. Every atomic formula is an S-formula.
2. If φ, ψ are S-formulas and x is a variable, then $\neg\varphi$, $(\varphi \vee \psi)$, and $\exists x\varphi$ are S-formulas.

We can define $\forall x\varphi$ to be $\neg\exists x\neg\varphi$. Also, all Boolean connectives can be defined in terms of $\neg$ and $\vee$.

L^S denotes the set of S-formulas.

Is there a string that is both a formula and a term? (No)

Can you think of a formula that can be parsed in more than one way? (No)

Lemma 1 *If $|S| \leq \omega$, then $|T^S| = |L^S| = \omega$.*

Proof?

$T^S \subseteq \mathcal{A}_S^*$; $L^S \subseteq \mathcal{A}_S^*$ and both are infinite.

2.3 Definitions on terms and formulas

Define a function that given an S-term returns the set of variables occuring in it.

$$var(x) = \{x\}$$
$$var(c) = \{\}$$
$$var(ft_1 \ldots t_n) = var(t_1) \cup \cdots \cup var(t_n)$$

Is the above really a definition? Why? Because there is only one way of decomposing a term into its parts, so we do not inadvertently allow *var* to assign different values to the same argument.

Looked at another way, we can define functions on terms (and formulas) by using recursive definitions based on the rules defining terms (and formulas).

Define a function that given an S-formula returns the set of free variables occuring in it.

$$free(t_1 \equiv t_2) = var(t_1) \cup var(t_2)$$
$$free(Rt_1 \ldots t_n) = var(t_1) \cup \cdots \cup var(t_n)$$
$$free(\neg\varphi) = free(\varphi)$$
$$free((\varphi \star \psi)) = free(\varphi) \cup free(\psi), \text{for } \star \text{ a boolean connective}$$
$$free(Qx\varphi) = free(\varphi) \setminus \{x\}, \text{for } Q = \forall, \exists$$

Formulas without free variables are called *sentences*.

3 Semantics of FOL

We will now go beyond the grammatical, syntactic aspects of FOL to discuss what terms and formulas mean. Notions such as *free*, *term*, *formula* are purely syntactic.

Here is an example of something that isn't syntactic: what does $\forall v_0 R v_0 v_1$ mean? Well, it depends on what R means, *i.e.*, what relation is it and over what domain? and what v_1 means, *i.e.*, what element of the domain is it? Say that R is $<$ on $\mathbb{N}$ and v_1 is 0, then the statement is false. If R is $\geq$, then it is true.

3.1 Structures and Interpretations

Definition 6 *An* S-structure *is a pair* $\mathbf{U} = \langle A, \mathbf{a} \rangle$*, where* A *is a non-empty set, the* domain *or* universe*, and* $\mathbf{a}$ *is a function with domain* S *such that:*

1. If $c \in S$ is a constant symbol, then $\mathbf{a}.c \in A$.

2. If $f \in S$ is an n-ary function symbol, then $\mathbf{a}.f : A^n \to A$.

3. If $R \in S$ is an n-ary relation symbol, then $\mathbf{a}.R \subseteq A^n$.

Instead of $\mathbf{a}.R, \mathbf{a}.f$, and $\mathbf{a}.c$ we often write $R^{\mathbf{U}}, f^{\mathbf{U}}$, and $c^{\mathbf{U}}$ or even R^A, f^A, and c^A. In addition, instead of denoting a structure $\mathbf{U}$ as a pair $\langle A, \mathbf{a} \rangle$, we often replace $\mathbf{a}$ by a list of its values, *e.g.*, we would write an $\{f, R, c\}$-structure as $\langle A, f^{\mathbf{U}}, R^{\mathbf{U}}, c^{\mathbf{U}} \rangle$.

Here are some examples. The symbol sets

$$S_{ar} := \{+, \cdot, 0, 1\} \text{ and } S_{ar}^< := \{+, \cdot, 0, 1, <\}$$

play an important role, and we use $\mathcal{N}$ to denote the S_{ar}-structure $\langle \mathbb{N}, +^{\mathbb{N}}, \cdot^{\mathbb{N}}, 0^{\mathbb{N}}, 1^{\mathbb{N}} \rangle$ and $\mathcal{N}^<$ to denote the $S_{ar}^<$-structure $\langle \mathbb{N}, +^{\mathbb{N}}, \cdot^{\mathbb{N}}, 0^{\mathbb{N}}, 1^{\mathbb{N}}, <^{\mathbb{N}} \rangle$.

Similarly, we use $\mathcal{R}$ to denote the S_{ar}-structure $\langle \mathbb{R}, +^{\mathbb{R}}, \cdot^{\mathbb{R}}, 0^{\mathbb{R}}, 1^{\mathbb{R}} \rangle$ and $\mathcal{R}^<$ to denote the $S_{ar}^<$-structure $\langle \mathbb{R}, +^{\mathbb{R}}, \cdot^{\mathbb{R}}, 0^{\mathbb{R}}, 1^{\mathbb{R}}, <^{\mathbb{R}} \rangle$.

Notice that $+^{\mathbb{R}}$ and $+^{\mathbb{N}}$ are very different objects. Even so, we will drop the subscripts when (we think) no ambiguity will arise.

Are we done? Can we give a precise meaning to terms and formulas?

What about $\forall v_0 < v_0 v_0$? (not true in $\mathbb{R}$ nor in $\mathbb{N}$)

What about $\forall v_0 \exists v_1 < v_1 v_0$? (not true in $\mathbb{N}$, true in $\mathbb{R}$)

What about our initial example, $\forall v_0 < v_0 v_1$?

It depends on what v_1 means, so let's go on.

Definition 7 *An* S-interpretation $\mathcal{J}$ *is a pair* $\langle \mathbf{U}, \beta \rangle$*, where* $\mathbf{U} = \langle A, \mathbf{a} \rangle$ *is an* S*-structure and* $\beta : \mathit{Var} \to A$*, is an* assignment*, a function that assigns values to the variables.*

We define the meaning of any term t in interpretation $\mathcal{J}$, denoted $\mathcal{J}.t$, as follows.

1. If $v \in \mathit{Var}$, then $\mathcal{J}.v = \beta.v$.

2. If $c \in S$ is a constant symbol, then $\mathcal{J}.c = c^{\mathbf{U}}$.

3. If $ft_1 \dots t_n$ is a term, then $\mathcal{J}(ft_1 \dots t_n)$ is $(f^{\mathbf{U}})(\mathcal{J}.t_1, \dots, \mathcal{J}.t_n)$.

Let's look at an example. If $S = S_{gr}$ and $\mathcal{J} = \langle \mathbf{U}, \beta \rangle$, where $\mathbf{U} = \langle \mathbb{Z}, +, 0 \rangle$ and $\beta.v_0 = 2, \beta.v_1 = 4$, then what is $\mathcal{J}(\circ v_0 \circ e v_1)$?
$= +^{\mathbb{Z}}(\mathcal{J}.v_0, \mathcal{J}(\circ e v_1))$
$= \beta.v_0 + +^{\mathbb{Z}}(e^{\mathbb{Z}}, \mathcal{J}.v_1)$
$= 2 + (0 + \beta.v_1)$

$= 2 + (0 + 4)$
$= 6$

If β is an assignment, then $\beta\frac{a}{x}(y)$ is a if $y = x$ and $\beta.y$ otherwise. For $\mathcal{J} = \langle \mathbf{U}, \beta \rangle$, $\mathcal{J}\frac{a}{x}$ denotes $\langle \mathbf{U}, \beta\frac{a}{x} \rangle$.

We now define what it means for an interpretation to satisfy a formula.

1. $\mathcal{J} \models (t_1 \equiv t_2)$ iff $\mathcal{J}.t_1 \;=\; \mathcal{J}.t_2$.
2. $\mathcal{J} \models R(t_1 \ldots t_n)$ iff $\langle \mathcal{J}.t_1, \ldots, \mathcal{J}.t_n \rangle \in R^{\mathbf{U}}$.
3. $\mathcal{J} \models \neg\varphi$ iff not $\mathcal{J} \models \varphi$.
4. $\mathcal{J} \models (\varphi \vee \psi)$ iff $\mathcal{J} \models \varphi$ or $\mathcal{J} \models \psi$.
5. $\mathcal{J} \models \exists x \varphi$ iff for some $a \in A$, $\mathcal{J}\frac{a}{x} \models \varphi$.

If $\mathcal{J} \models \varphi$ we say that φ holds in $\mathcal{J}$; we also say that $\mathcal{J}$ is a model of φ; we also say that $\mathcal{J}$ satisfies φ .

Given, Φ, a set of formulas, $\mathcal{J} \models \Phi$ ($\mathcal{J}$ is a model of Φ) iff for every $\varphi \in \Phi$, $\mathcal{J} \models \varphi$.

You should convince yourself that $\mathcal{J} \models \varphi$ iff φ is *true* under interpretation $\mathcal{J}$.

Let's look at an example. If $S = S_{gr}$ and $\mathcal{J} = \langle \mathbf{U}, \beta \rangle$, where $\mathbf{U} = \langle \mathbb{Z}, +, 0 \rangle$ and $\beta.v_0 = 2, \beta.v_1 = 4$, as before, then what is the value of $\mathcal{J} \models \forall v_0 \exists v_1 \circ v_0 e \equiv v_1$?

$\mathcal{J} \models \forall v_0 \exists v_1 \circ v_0 e \equiv v_1$
iff for all $i \in \mathbb{Z}, \mathcal{J}\frac{i}{v_0} \models \exists v_1 \circ v_0 e \equiv v_1$
iff for all $i \in \mathbb{Z}$, there is a $j \in \mathbb{Z}$ such that $(\mathcal{J}\frac{i}{v_0})\frac{j}{v_1} \models \circ v_0 e \equiv v_1$
iff for all $i \in \mathbb{Z}$, there is a $j \in \mathbb{Z}$ such that $(\mathcal{J}\frac{i}{v_0})\frac{j}{v_1}(\circ v_0 e) = (\mathcal{J}\frac{i}{v_0})\frac{j}{v_1}(v_1)$
iff for all $i \in \mathbb{Z}$, there is a $j \in \mathbb{Z}$ such that $\circ^{\mathbf{U}}((\mathcal{J}\frac{i}{v_0})\frac{j}{v_1}(v_0), (\mathcal{J}\frac{i}{v_0})\frac{j}{v_1}(e)) = j$
iff for all $i \in \mathbb{Z}$, there is a $j \in \mathbb{Z}$ such that $i + e^{\mathbf{U}} = j$
iff for all $i \in \mathbb{Z}$, there is a $j \in \mathbb{Z}$ such that $i + 0 = j$
true, set j to i

Note that the meaning of a sentence does not depend on the assignment. In general, we are interested in sentences, but to evaluate them, we have to evaluate subformulas, which may not be sentences, therefore, the need for assignments. This kind of thing comes up in programming a lot.

Using the notion of satisfaction, we define the notion of consequence.

Definition 8 *Let Φ be a set of formulas and φ a formula. Then $\Phi \models \varphi$ (φ is a consequence of Φ) iff for every interpretation, $\mathcal{J}$, which is a model of Φ, we have that $\mathcal{J} \models \varphi$.*

We have developed enough mathematical machinery to reconsider, in a more rigourous way, one of our initial goals. Recall, that we were interested in whether $\Phi \models \varphi$ iff $\Phi \vdash \varphi$. For example, we saw a proof that groups have a left inverse, *i.e.*, $\Phi_{gr} \vdash \forall v_o \exists v_1 (v_1 \circ v_0) \equiv e$, and you should be convinced that such a proof implies $\Phi_{gr} \models \forall v_o \exists v_1 (v_1 \circ v_0) \equiv e$, where $\Phi_{gr} = \{\forall v_0 \forall v_1 \forall v_2 (v_0 \circ v_1) \circ v_2 \equiv v_0 \circ (v_1 \circ v_2), \forall v_0 v_0 \circ e \equiv v_0, \forall v_0 \exists v_1 v_0 \circ v_1 = e\}$. Once we develop the notion of proof more carefully, this will be an easy theorem to prove.

What is not as clear is whether the opposite direction holds. The completeness theorem will establish this. That comes after we define what a proof is and will be the first main theorem we prove.

We now continue to build our vocabulary.

Definition 9 *A formula φ is* valid *iff $\emptyset \models \varphi$, which we abbreviate by $\models \varphi$.*

Definition 10 *A formula φ is* satisfiable*, written Sat φ iff there is an interpretation which is a model of φ; similarly, a set of formulas Φ is satisfiable, Sat Φ iff there is an interpretation which is a model of all the formulas in Φ.*

Lemma 2 *For all Φ and all φ, $\Phi \models \varphi$ iff not Sat $\Phi \cup \{\neg\varphi\}$.*

Proof $\Phi \models \varphi$
iff for all $\mathcal{J}$, $\mathcal{J} \models \Phi$ implies $\mathcal{J} \models \varphi$
iff there is no $\mathcal{J}$ such that $\mathcal{J} \models \Phi$ but not $\mathcal{J} \models \varphi$
iff there is no $\mathcal{J}$ such that $\mathcal{J} \models \Phi \cup \{\neg\varphi\}$
iff not Sat $\Phi \cup \{\neg\varphi\}$. $\square$

As a consequence, φ is valid iff $\neg\varphi$ is not satisfiable.

We now prove some straight-forward lemmas that clarify the situation and suggest new notations.

The first lemma, the "coincidence lemma" isolates what parts of an interpretation can affect the meaning of terms and formulas.

Lemma 3 *(Coincidence Lemma). Let $\mathcal{J}_1 = \langle \mathbf{U}_1, \beta_1 \rangle$ be an S_1-interpretation and and $\mathcal{J}_2 = \langle \mathbf{U}_2, \beta_2 \rangle$ be an S_2-interpretation, both with the same domain. Let $S = S_1 \cap S_2$.*

1. *Let t be an S-term. If $\mathcal{J}_1$ and $\mathcal{J}_2$ agree on the S-symbols occurring in t and on the variables occuring in t, then $\mathcal{J}_1(t) = \mathcal{J}_2(t)$.*

2. *Let φ be an S-formula. If $\mathcal{J}_1$ and $\mathcal{J}_2$ agree on the S-symbols and on the variables occurring free in φ, then $\mathcal{J}_1 \models \varphi$ iff $\mathcal{J}_2 \models \varphi$.*

Proof By induction on S-terms and then on S-formulas. $\square$

Note that the coincidence lemma tells us that the meaning of a formula φ under an interpretation $\mathcal{J}$ depends only on the free variables in φ, which form a finite part of an assignment.

If the variables are among $v_0, v_1, \ldots, v_{n-1}$ (denoted $\varphi \in L_n^S$, so $\varphi \in L_0^S$ is the set of S-sentences), and if $\beta.v_i = a_i$, instead of $\langle \mathbf{U}, \beta \rangle \models \varphi$, we often write the more suggestive

$$\mathbf{U} \models \varphi[a_0, \ldots, a_{n-1}]$$

Similary, if t is an S-term such that $var(t) \subseteq \{v_0, \ldots, v_{n-1}\}$, instead of $\mathcal{J}(t)$, we may write $t^{\mathbf{U}}[a_0, \ldots, a_{n-1}]$.

If φ is a sentence ($\varphi \in L_0^S$) then we write $\mathbf{U} \models \varphi$.

If Φ is a set of sentence, then, as expected, $\mathbf{U} \models \Phi$ means that for each $\varphi \in \Phi$, $\mathbf{U} \models \varphi$.

3.2 Substitution

We want to define a notion of substitution so that if we substitute term t for variable x in formula φ, obtaining φ', then φ' says about t what φ says about x. Substitution is know to be error-prone. Here is an example of how we have to be careful.

Consider $\varphi = \exists z z + z \equiv x$.

Note that $\langle \mathcal{N}, \beta \rangle \models \varphi$ iff $\beta.x$ is even.

Replacing x by y gives, $\varphi' = \exists z z + z \equiv y$, where $\langle \mathcal{N}, \beta \rangle \models \varphi$ iff $\beta.y$ is even. Good.

What about replacing x by z? This gives $\varphi' = \exists z z + z \equiv z$, but $\mathcal{N} \models \varphi$, so here we have a problem. In order to get a φ' which expresses about z what φ expresses about x, we can first replace bound occurences of z by a new variable u in φ, and then proceed as before.

We will define how to perform simultaneous substitution for terms, where the x_i are distinct.

1. $x\frac{t_0 \ldots t_r}{x_0 \ldots x_r} = \begin{cases} x & \text{if } x \neq x_0, \ldots, x \neq x_r, \\ t_i & \text{if } x = x_i \end{cases}$
2. $c\frac{t_0 \ldots t_r}{x_0 \ldots x_r} = c$
3. $[ft'_1 \ldots t'_n]\frac{t_0 \ldots t_r}{x_0 \ldots x_r} = ft'_1\frac{t_0 \ldots t_r}{x_0 \ldots x_r} \ldots t'_n\frac{t_0 \ldots t_r}{x_0 \ldots x_r}$

The square brackets are for easier reading. Now, we define substitution for formulas

1. $[t'_1 \equiv t'_2]\frac{t_0 \ldots t_r}{x_0 \ldots x_r} = t'_1\frac{t_0 \ldots t_r}{x_0 \ldots x_r} \equiv t'_2\frac{t_0 \ldots t_r}{x_0 \ldots x_r}$
2. $[Rt'_1 \ldots t'_n]\frac{t_0 \ldots t_r}{x_0 \ldots x_r} = Rt'_1\frac{t_0 \ldots t_r}{x_0 \ldots x_r} \ldots t'_n\frac{t_0 \ldots t_r}{x_0 \ldots x_r}$
3. $[\neg\varphi]\frac{t_0 \ldots t_r}{x_0 \ldots x_r} = \neg[\varphi\frac{t_0 \ldots t_r}{x_0 \ldots x_r}]$
4. $(\varphi \vee \psi)\frac{t_0 \ldots t_r}{x_0 \ldots x_r} = (\varphi\frac{t_0 \ldots t_r}{x_0 \ldots x_r} \vee \psi\frac{t_0 \ldots t_r}{x_0 \ldots x_r})$
5. Suppose $x_{i_1}, \ldots, x_{i_s} (i_1 < \cdots < i_s)$ are exactly the variables x_i among the $x_0, \ldots, x_r$ such that

   $x_i \in \mathit{free}(\exists x\varphi)$ and $x_i \neq t_i$

   Then, set

   $$[\exists x\varphi]\frac{t_0 \ldots t_r}{x_0 \ldots x_r} = \exists u[\varphi\frac{t_{i_1} \ldots t_{i_s} u}{x_{i_1} \ldots x_{i_s} x}],$$

where u is x if x does not occur in $t_{i_1} \dots t_{i_s}$; otherwise u is the first variable in the list $v_0, v_1, v_2, \dots$ which does not occur in $\varphi, t_{i_1} \dots t_{i_s}$.

Notice that this definition is very much like a program and in fact, similar definitions need to be given in actual languages.

Let's look at some examples.

1. $[Pv_0 f v_1 v_2]\frac{v_2 v_0 v_1}{v_1 v_2 v_3} = Pv_0 f v_2 v_0$
2. $[\exists v_0 P v_0 f v_1 v_2]\frac{v_4 f v_1 v_1}{v_0 v_2} = \exists v_0 [P v_0 f v_1 v_2 \frac{f v_1 v_1 v_0}{v_2 v_0}] = \exists v_0 P v_0 f v_1 f v_1 v_1$
3. $[\exists v_0 P v_0 f v_1 v_2]\frac{v_0 v_2 v_4}{v_1 v_2 v_0} = \exists v_3 [P v_0 f v_1 v_2 \frac{v_0 v_3}{v_1 v_0}] = \exists v_3 P v_3 f v_0 v_2$

There are some lemmas about substitution that will be important later on, and that is what we will get to after some definitions.

First, some definitions that extend existing notations. Let $\mathcal{J} = \langle \mathbf{U}, \beta \rangle$ with $a_0, \dots, a_r \in A$. Then:

$\beta\frac{a_0 \dots a_r}{x_0 \dots x_r}(y) = \begin{cases} \beta.y & \text{if } y \neq x_0, \dots, y \neq x_r \\ a_i & \text{if } y = x_i \end{cases}$

and

$\mathcal{J}\frac{a_0 \dots a_r}{x_0 \dots x_r} = \langle \mathbf{U}, \beta\frac{a_0 \dots a_r}{x_0 \dots x_r} \rangle$

Here then is the main result about substitution.

Lemma 4 *1. For every term t, $\mathcal{J}(t\frac{t_0 \dots t_r}{x_0 \dots x_r}) = \mathcal{J}\frac{\mathcal{J}(t_0) \dots \mathcal{J}(t_r)}{x_0 \dots x_r}(t)$*

2. For every formula φ, $\mathcal{J} \models \varphi\frac{t_0 \dots t_r}{x_0 \dots x_r}$ iff $\mathcal{J}\frac{\mathcal{J}(t_0) \dots \mathcal{J}(t_r)}{x_0 \dots x_r} \models \varphi$

Proof By induction on terms and formulas. $\square$

4 Proof Theory

4.1 Introduction

Remember that we are on our way to proving $\Phi \models \varphi$ iff $\Phi \vdash \varphi$. We defined what $\Phi \models \varphi$ means, that is when φ is a consequence of Φ. Now we will define $\Phi \vdash \varphi$, that is when φ is provable from Φ. There are many ways of defining the notion of proof and at first glance it may seem a hopeless task to nail down exactly what it is that is allowed in a proof. Don't mathmaticians expand their set of techniques every so often? It will turn out that we will give a fairly simple set of obvious proof rules that will be enough to prove the completeness theorem. What we are doing is defining a calculus and the formulas derivable in the calculus are exactly the provable formulas.

4.2 Sequent Rules

We will use the notion of a *sequent*: a nonempty list (sequence) of formulas. For example, $\varphi_1 \ldots \varphi_n \varphi$ is a sequent. $\varphi_1 \ldots \varphi_n$ is called the *antecedent* and φ is the *succedent*. From the unique decomposition of formulas, we know that we can uniquely determine the antecedent and succedent of a sequent. The antecedent can be empty, but the succedent is not.

We will use $\Gamma, \Delta, \ldots$ to denote (possibly empty) sequences of formulas. We will now define a sequent calculus. Here is an example.

$$\frac{\begin{array}{lll}\Gamma & \neg\varphi & \psi \\ \Gamma & \neg\varphi & \neg\psi\end{array}}{\begin{array}{ll}\Gamma & \varphi\end{array}}$$

Think of this as saying that if you have a proof of both ψ and $\neg\psi$ from $\Gamma \cup \{\neg\varphi\}$ then that constitutes a proof of φ from Γ.

If there is a derivation of the sequent $\Gamma\ \varphi$, then we write $\vdash \Gamma\ \varphi$ and we say that $\Gamma\ \varphi$ is *derivable*.

Definition 11 *A formula φ is* formally provable *or* derivable *from a set Φ of formulas (written $\Phi \vdash \varphi$) iff there are finetely many formulas $\varphi_1, \ldots, \varphi_n$ in Φ such that $\vdash \varphi_1\ \ldots\ \varphi_n\ \varphi$.*

A sequent $\Gamma\ \varphi$ is *correct* if $\Gamma \models \varphi$ (more carefully $\{\psi : \psi$ is a member of $\Gamma\} \models \varphi$).

We will now introduce the rules of the sequent calculus and will show that they are *correct*: when applied to correct sequents, they return correct sequents.

Antecedent Rule (Ant)

$$\frac{\Gamma \quad \varphi}{\Gamma' \quad \varphi} \text{ if every member of } \Gamma \text{ is also a member of } \Gamma'.$$

Assumption Rule (Assm)

$$\frac{}{\Gamma \quad \varphi} \text{ if } \varphi \text{ is a member of } \Gamma.$$

Proof of correctness of the above rules is obvious, but let's look at a proof to make sure we know what is required. Remember showing that a rule is correct requires showing that if the rule is applied to correct sequents, it returns a correct sequent.

Correctness of Ant: If $\Gamma\ \varphi$ is correct, then by definition $\Gamma \models \varphi$, (here we are thinking of Γ as the set $\{\psi : \psi$ is a formula in $\Gamma\}$) but since $\Gamma \subseteq \Gamma'$, (again, we are thinking of Γ, Γ' as sets, when they are really sequences) $\Gamma' \models \varphi$ as well. Why? Note that $\Gamma \models \varphi$ means that any interpretation that satisfies Γ satisfies φ. Any interpretation that satisfies Γ' also satisfies Γ, this is sometimes called the monotonicity of FOL. By increasing a set of formulas, you either decrease or do not affect the class of models satisfying the formulas.

Proof by Cases Rule (PC)

$$\begin{array}{lll} \Gamma & \psi & \varphi \\ \Gamma & \neg\psi & \varphi \\ \hline \Gamma & & \varphi \end{array}$$

Proof of correctness?

Contradiction Rule (Ctr)

$$\begin{array}{lll} \Gamma & \neg\varphi & \psi \\ \Gamma & \neg\varphi & \neg\psi \\ \hline \Gamma & & \varphi \end{array}$$

$\vee$-Rule for the Antecedent ($\vee$ A)

$$\begin{array}{lll} \Gamma & \varphi & \xi \\ \Gamma & \psi & \xi \\ \hline \Gamma & (\varphi \vee \psi) & \xi \end{array}$$

$\vee$-Rule for the Succedent ($\vee$ S)

$$\text{(a)}\frac{\Gamma \quad \varphi}{\Gamma \quad (\varphi \vee \psi)} \qquad\qquad \text{(b)}\frac{\Gamma \quad \varphi}{\Gamma \quad (\psi \vee \varphi)}$$

Using the existing rules, we can derive various sequents. We can also show that rules themselves are derivable. These so called derived rules of inference are derived, instead of made base rules, for the same reasons that the connectives $\wedge, \rightarrow$, etc. are thought of as abbreviations. We want to keep things simple. By showing that they are derivable, we can use them as if they were built in, but do not have to reason about them, *i.e.*, they do not add to proof obligations. At the other extreme, where we are interested not in the simplicity of the logic (because we are exploring its inherent power), but where we are interested in the usability of the logic, as is the case with ACL2, we can think of the ACL2 system as one big derived rule of inference.

Tertium non datur (Ctr)

$$\frac{}{(\varphi \vee \neg\varphi)}$$

Proof? We can prove it by assuming φ, getting $\varphi \vee \neg\varphi$ and similarly with $\neg\varphi$.

$$\begin{array}{llll} 1. & \varphi & \varphi & \text{(Ant)} \\ 2. & \varphi & (\varphi \vee \neg\varphi) & (\vee\ \text{S}) \\ 3. & \neg\varphi & \neg\varphi & \text{(Ant)} \\ 4. & \neg\varphi & (\varphi \vee \neg\varphi) & (\vee\ \text{S}) \\ 5. & & (\varphi \vee \neg\varphi) & \text{(PC)} \end{array}$$

There are other rules. Here are some of them.

Second Contradiction Rule (Ctr')

$$\begin{array}{ll} \Gamma & \psi \\ \Gamma & \neg\psi \\ \hline \Gamma & \varphi \end{array}$$

Chain Rule (Ch)

$$\begin{array}{lll} \Gamma & & \varphi \\ \Gamma & \varphi & \psi \\ \hline \Gamma & & \psi \end{array}$$

Contraposition Rules (Cp)

(a) $\dfrac{\Gamma \quad \varphi \quad \psi}{\Gamma \quad \neg\psi \quad \neg\varphi}$ (c) $\dfrac{\Gamma \quad \neg\varphi \quad \psi}{\Gamma \quad \neg\psi \quad \varphi}$

(b) $\dfrac{\Gamma \quad \neg\varphi \quad \neg\psi}{\Gamma \quad \psi \quad \varphi}$ (d) $\dfrac{\Gamma \quad \varphi \quad \neg\psi}{\Gamma \quad \psi \quad \neg\varphi}$

Modus ponens

$$\begin{array}{ll} \Gamma & (\varphi \to \psi) \\ \Gamma & \varphi \\ \hline \Gamma & \psi \end{array}$$

4.3 Quantifier and Equality Rules

Now we will look at rules for quantifiers and equality.

$\exists$-Introduction in the Succedent ($\exists$ S)

$$\begin{array}{ll} \Gamma & \varphi\frac{t}{x} \\ \hline \Gamma & \exists x\varphi \end{array}$$

Proof Suppose $\Gamma \models \varphi\frac{t}{x}$. If $\mathcal{J} \models \Gamma$, we have $\mathcal{J} \models \varphi\frac{t}{x}$. By the substitution lemma, $\mathcal{J}\frac{\mathcal{J}.t}{x} \models \varphi$ and thus $\mathcal{J} \models \exists x\varphi$. $\square$

The next rule corresponds to an often used argument used to prove that ψ follows from $\exists x\varphi$. One assumes that for some new y, $\varphi\frac{y}{x}$. The intuition is that this is a valid thing to do because nothing is known about y.

$\exists$-Introduction in the Antecedent ($\exists$ A)

$$\frac{\Gamma \quad \varphi\frac{y}{x} \quad \psi}{\Gamma \quad \exists x\varphi \quad \psi} \text{ if } y \text{ is not free in } \Gamma\ \exists x\varphi\ \psi.$$

Proof So, $\Gamma\varphi\frac{y}{x} \models \psi$. Suppose $\mathcal{J} \models \Gamma$ and $\mathcal{J} \models \exists x\varphi$. Then there is an a such that $\mathcal{J}\frac{a}{x} \models \varphi$, but by the coincidence lemma, $(\mathcal{J}\frac{a}{y})\frac{a}{x} \models \varphi$. Since $\mathcal{J}\frac{a}{y}(y) = a$, we have $(\mathcal{J}\frac{a}{y})\frac{\mathcal{J}\frac{a}{y}(y)}{x} \models \varphi$ and by substitution lemma $\mathcal{J}\frac{a}{y} \models \varphi\frac{y}{x}$. Since $\mathcal{J} \models \Gamma$ and $y \notin \text{free}.\Gamma$, we get $\mathcal{J}\frac{a}{y} \models \Gamma$. Now, we get $\mathcal{J}\frac{a}{y} \models \psi$ and therefore $\mathcal{J} \models \psi$ because $y \notin \text{free}.\psi$. $\square$

Finally, two rules about equality.

Reflexivity Rule for Equality ($\equiv$)

$$\frac{}{t \equiv t}$$

Substitution Rule for Equality (Sub)

$$\frac{\Gamma \qquad\qquad \varphi\frac{t}{x}}{\Gamma \quad t \equiv t' \quad \varphi\frac{t'}{x}}$$

Let's review. A formula φ is derivable from Φ, written $\Phi \vdash \varphi$, iff there are formulas $\varphi_1, \ldots, \varphi_n$ in Φ such that $\vdash \varphi_1 \ldots \varphi_n \varphi$. From this definition, the following lemma follows easily.

Lemma 5 *For all Φ and φ, $\Phi \vdash \varphi$ iff there is a finite subset Φ_0 of Φ such that $\Phi_0 \vdash \varphi$.*

We will prove a similar theorem, the compactness theorem, for $\models$. As a preview, once we prove the completeness theorem, namely that the notions $\models$ and $\vdash$ are "equivalent" then we will be able to transfer results such as this one from one realm to the other. The beauty is that sometimes results are trivial to prove in one realm, but seem very deep in the other.

Theorem 2 *For all Φ and φ, if $\Phi \vdash \varphi$ then $\Phi \models \varphi$.*

Proof The proof is by induction on the structure of a derivation. Suppose $\Phi \vdash \varphi$. Then, we have $\vdash \Gamma\varphi$, for $\Gamma \subseteq \Phi$. Since every rule is correct, every derivable sequent is correct, hence $\Gamma\varphi$ is correct, so $\Gamma \models \varphi$ and $\Phi \models \varphi$. $\square$

This is one direction of the completeness theorem. Note that we now know what $\Phi \models \varphi$ means and what $\Phi \vdash \varphi$ means. It is surprising that mathematical reasoning, the essence of mathematics, can be reduced to these simple proof rules.

5 Consistency

After we introduced $\models$, consequence, we introduced satisfiability. The syntactic counterpart is consistency.

Definition 12 *Φ is consistent, written Con Φ, iff there is no formula φ such that $\Phi \vdash \varphi$ and $\Phi \vdash \neg\varphi$.*

*Φ is inconsistent, written Inc Φ iff Φ is not consistent (*i.e.*, there is a formula φ such that $\Phi \vdash \varphi$ and $\Phi \vdash \neg\varphi$).*

Lemma 6 *Inc Φ iff for all φ: $\Phi \vdash \varphi$.*

Proof Only ($\Rightarrow$) is not obvious, but it follows from (Ctr'). $\square$

Lemma 7 *Con Φ iff there is a φ such that not $\Phi \vdash \varphi$.*

Proof Negate both sides of the previous lemma. $\square$

Lemma 8 *For all Φ, Con Φ iff Con Φ_0 for all finite subsets Φ_0 of Φ.*

Proof $\Phi \vdash \varphi$ iff $\Phi_0 \vdash \varphi$ for some finite subset Φ_0 of Φ. $\square$

Lemma 9 *Sat Φ implies Con Φ.*

Proof

Inc Φ

$\Rightarrow$ { Definition of Inc }
$\Phi \vdash \varphi$ and $\Phi \vdash \neg\varphi$

$\Rightarrow$ { Correctness of the sequent calculus }
$\Phi \models \varphi$ and $\Phi \models \neg\varphi$

$\Rightarrow$ { A formula is either true or false in a model }
not Sat Φ $\square$

Lemma 10 *For all Φ and φ the following holds:*

1. *$\Phi \vdash \varphi$ iff Inc $\Phi \cup \{\neg\varphi\}$.*
2. *$\Phi \vdash \neg\varphi$ iff Inc $\Phi \cup \{\varphi\}$.*
3. *If Con Φ, then Con $\Phi \cup \{\varphi\}$ or Con $\Phi \cup \{\neg\varphi\}$.*

Proof

$\Phi \vdash \varphi$

$\Rightarrow$ { }
$\Phi \cup \{\neg\varphi\} \vdash \varphi$ and $\Phi \cup \{\neg\varphi\} \vdash \neg\varphi$

$\Rightarrow$ { Definition of Inc }
Inc $\Phi \cup \{\neg\varphi\}$

$\Rightarrow$ { By definition of Inc, there is $\Gamma \subseteq \Phi$ }
$\vdash \Gamma\ \neg\varphi\quad \varphi$

$\Rightarrow$ { $\Gamma\ \neg\varphi\quad \varphi$
$\Gamma\quad \varphi\quad \varphi$ (Assm)
$\Gamma\qquad \varphi$ (PC) }

$\Phi \vdash \varphi$

The second part is similar.

$\text{Inc}\Phi \cup \{\varphi\} \quad \text{and} \quad \text{Inc}\Phi \cup \{\neg\varphi\}$

$\Rightarrow \quad \{$ Parts 1, 2, above $\}$

$\Phi \vdash \neg\varphi \quad \text{and} \quad \Phi \vdash \varphi$

$\Rightarrow \quad \{$ Definition of Inc $\}$

Inc Φ $\Box$

We have assumed a fixed symbol set S. When we need to consider several symbol sets simultaneously, we will use $\Phi \vdash_S \varphi$ to indicate that that there is a derivation with underlying symbol set S. Similarly $\text{Con}_S\ \Phi$ denotes Con Φ with underlying symbol set S.

Lemma 11 *For all $i \in \omega$, S_i is a symbol set and $S_i \subseteq S_{i+1}$. Similarly for all $i \in \omega$, Φ_i is a set of S_i-formulas such that $Con_{S_i}\ \Phi_i$ and $\Phi_i \subseteq \Phi_{i+1}$.*

Let $S = \cup_{i\in\omega} S_i$ and $\Phi = \cup_{i\in\omega}\Phi_i$. Then $Con_S\ \Phi$.

Proof

$\text{Inc}_S\Phi$

$\Rightarrow \quad \{$ $\text{Inc}_S\Psi$ for finite Ψ s.t. $\Psi \subseteq \Phi$, thus $\Psi \subseteq \Phi_k$ for some k $\}$

$\text{Inc}_S\Phi_k$

$\Rightarrow \quad \{$ Any derivation of φ, $\neg\varphi$ is finite so all symbols are in S_m for $m \geq k$ $\}$

$\text{Inc}_{S_m}\Phi_m$

6 Completeness Theorem

To show: For all Φ and φ: If $\Phi \models \varphi$ then $\Phi \vdash \varphi$. We will instead show: Every consistent set of formulas is satisfiable.

Proof

not $\Phi \vdash \varphi$ implies not $\Phi \models \varphi$

$\equiv \quad \{$ Lemma 10 $\}$

Con $\Phi \cup \{\neg\varphi\}$ implies Sat $\Phi \cup \{\neg\varphi\}$

$\Leftarrow \quad \{$ Instance of $\}$

Con Ψ implies Sat Ψ $\Box$

6.1 Henkin's Theorem

If Φ is consistent, then all we have is the syntactical info that this provides. Let's use it to find a model $\mathcal{J} = \langle \mathbf{U}, \beta \rangle$ of Φ. If A is T^S and $\beta(v_i) = v_i$, $f^{\mathbf{U}}(t) = ft$, ..., then for variable x we have $\mathcal{J}(fx) = f^{\mathbf{U}}(\beta.x) = fx$, so $\mathcal{J}(fv_0) \neq \mathcal{J}(fv_1)$, but what if $fv_0 \equiv fv_1 \in \Phi$? To overcome this, we define an equivalence relation on terms.

First, we define an equivalence relation on T^S: $t_1 \sim t_2$ iff $\Phi \vdash t_1 \equiv t_2$.

Lemma 12

1. *$\sim$ is an equivalence relation.*
2. *If $t_1 \sim t'_1, \ldots, t_n \sim t'_n$ then for n-ary $f \in S$: $ft_1 \ldots t_n \sim ft'_1 \ldots t'_n$ and for n-ary $R \in S$: $\Phi \vdash Rt_1 \ldots t_n$ iff $\Phi \vdash Rt'_1 \ldots t'_n$.*

Proof Follows from previous chapter, *e.g.*, there it is shown that $\equiv$ is an equivalence relation.

$$t_1 \sim t'_1, \ldots, t_n \sim t'_n$$

$\equiv$ { Definition of $\sim$ }

$$\Phi \vdash t_1 \equiv t'_1, \ldots, \Phi \vdash t_n \equiv t'_n$$

$\Rightarrow$ { Results of last chapter }

$$\Phi \vdash ft_1 \ldots t_n \equiv ft'_1 \ldots t'_n$$

$\equiv$ { Definition of $\sim$ }

$$ft_1 \ldots t_n \sim ft'_1 \ldots t'_n$$

Let $\bar{t} = \{t' \in T^S : t \sim t'\}$, *i.e.*, $\bar{t}$ is the equivalence class of t.

Let T^Φ be the set of equivalence classes: $T^\Phi = \{\bar{t} : t \in T^S\}$. Note that T^Φ is not empty. We now define the term structure over T^Φ, $\mathcal{T}^\Phi$ as follows.

1. $c^{\mathcal{T}^\Phi} = \bar{c}$
2. $f^{\mathcal{T}^\Phi}(\overline{t_1}, \ldots, \overline{t_n}) = \overline{ft_1 \ldots t_n}$
3. $R^{\mathcal{T}^\Phi}\overline{t_1} \ldots \overline{t_n}$ iff $\Phi \vdash Rt_1 \ldots t_n$

Note that by Lemma 12, the definitions of $f^{\mathcal{T}^\Phi}$ and $R^{\mathcal{T}^\Phi}$ make sense.

We define the *term interpretation* associated with Φ to be $\mathcal{J}^\Phi = \langle \mathcal{T}^\Phi, \beta^\Phi \rangle$, where $\beta^\Phi(x) = \bar{x}$.

Lemma 13

1. *For all t, $\mathcal{J}^\Phi(t) = \bar{t}$.*
2. *For every atomic formula φ, $\mathcal{J}^\Phi \models \varphi$ iff $\Phi \vdash \varphi$.*

3. *For every formula φ and pairwise disjoint variables $x_1, \ldots, x_n$*

   (a) $\mathcal{J}^{\varphi} \models \exists x_1 \ldots \exists x_n \varphi$ *iff there are* $t_1, \ldots, t_n \in T^S$ *s.t.* $\mathcal{J}^{\Phi} \models \varphi\frac{t_1 \ldots t_n}{x_1 \ldots x_n}$.

   (b) $\mathcal{J}^{\varphi} \models \forall x_1 \ldots \forall x_n \varphi$ *iff for all* $t_1, \ldots, t_n \in T^S$ *we have* $\mathcal{J}^{\Phi} \models \varphi\frac{t_1 \ldots t_n}{x_1 \ldots x_n}$.

Proof (1) By induction on terms. By definition it holds for variables and constants. If $t = ft_1 \ldots t_n$ then

$$
\begin{aligned}
& \mathcal{J}^{\Phi}(ft_1 \ldots t_n) \\
\equiv & \quad \{ \text{ Definitions } \} \\
& f^{\mathcal{T}^{\Phi}}(\mathcal{J}^{\Phi}(t_1), \ldots, \mathcal{J}^{\Phi}(t_n)) \\
\equiv & \quad \{ \text{ Induction hypothesis } \} \\
& f^{\mathcal{T}^{\Phi}}(\overline{t_1}, \ldots, \overline{t_n}) \\
\equiv & \quad \{ \text{ Definition of } f^{\mathcal{T}^{\Phi}} \} \\
& \overline{ft_1 \ldots t_n}
\end{aligned}
$$

(2)

$$
\begin{aligned}
& \mathcal{J}^{\Phi} \models t_1 \equiv t_2 \\
\equiv & \quad \{ \text{ Definitions } \} \\
& \mathcal{J}^{\Phi}(t_1) = \mathcal{J}^{\Phi}(t_2) \\
\equiv & \quad \{ \text{ by part (1) } \} \\
& \overline{t_1} = \overline{t_2} \\
\equiv & \quad \{ \text{ definition of } \bar{t} \} \\
& t_1 \sim t_2 \\
\equiv & \quad \{ \text{ Definition of } \sim \} \\
& \Phi \vdash t_1 \equiv t_2
\end{aligned}
$$

$$
\begin{aligned}
& \mathcal{J}^{\Phi} \models Rt_1 \ldots t_n \\
\equiv & \quad \{ \text{ Definitions } \} \\
& R^{\mathcal{T}^{\Phi}}(\mathcal{J}^{\Phi}(t_1)) \ldots (\mathcal{J}^{\Phi}(t_n)) \\
\equiv & \quad \{ \text{ by part (1) } \} \\
& R^{\mathcal{T}^{\Phi}}\overline{t_1} \ldots \overline{t_n} \\
\equiv & \quad \{ \text{ Definition of } R^{\mathcal{T}^{\Phi}} \}
\end{aligned}
$$

$\Phi \vdash Rt_1 \ldots t_n$

(c)

$\mathcal{J}^\Phi \models \exists x_1 \ldots \exists x_n \varphi$

$\equiv$ { Definitions }

there are $a_1, \ldots, a_n \in T^\Phi$ s.t. $\mathcal{J}^\Phi \dfrac{a_1 \ldots a_n}{x_1 \ldots x_n} \models \varphi$

$\equiv$ { $T^\Phi = \{\bar{t} : t \in T^S\}$ }

there are $t_1, \ldots, t_n \in T^S$ s.t. $\mathcal{J}^\Phi \dfrac{\overline{t_1} \ldots \overline{t_n}}{x_1 \ldots x_n} \models \varphi$

$\equiv$ { by part (1) }

there are $t_1, \ldots, t_n \in T^S$ s.t. $\mathcal{J}^\Phi \dfrac{\mathcal{J}^\Phi(t_1) \ldots \mathcal{J}^\Phi(t_n)}{x_1 \ldots x_n} \models \varphi$

$\equiv$ { Substitution lemma }

there are $t_1, \ldots, t_n \in T^S$ s.t. $\mathcal{J}^\Phi \models \varphi \dfrac{t_1 \ldots t_n}{x_1 \ldots x_n}$

Part 2 of c is similar. $\square$

Where are we? Well, by the previous lemma $\mathcal{J}^\Phi$ is a model of the atomic formulas in Φ, but we do not know that it is a model of all formulas in Φ. In fact, it isn't. Consider $\Phi = \{\exists x Rx\}$. Then, by (3) of the previous lemma, $\mathcal{J}^\Phi \models \Phi$ iff there is a term (in our case a variable) y such that $\exists x Rx \vdash Ry$, but this does not hold, as one of the exercises requires you to show. Consider $\Phi \cup \{\neg Ry : y$ is a variable $\}$. This set is satisfiable, thus consistent, but for no term $t \in T^S$ do we have $\Phi \vdash Rt$.

What is missing are some closure conditions that we now specify.

Definition 13

Φ is *negation complete* iff for every formula φ, $\Phi \vdash \varphi$ or $\Phi \vdash \neg\varphi$.

Φ *contains witnesses* iff for every formula of the form $\exists x \varphi$, there is a term t such that $\Phi \vdash (\exists x \varphi \rightarrow \varphi \frac{t}{x})$.

Lemma 14 *If Φ is consistent, negation complete, and contains witnesses, then for all φ and ψ.*

1. *$\Phi \vdash \neg\varphi$ iff not $\Phi \vdash \varphi$*
2. *$\Phi \vdash (\varphi \vee \psi)$ iff $\Phi \vdash \varphi$ or $\Phi \vdash \psi$*
3. *$\Phi \vdash \exists x \varphi$ iff there is a term t s.t. $\Phi \vdash \varphi \frac{t}{x}$*

Proof (a) Since Φ is negation complete, $\Phi \vdash \varphi$ or $\Phi \vdash \neg\varphi$. Since it is consistent, not both.

(b) ($\Leftarrow$): Use ($\vee$ S). ($\Rightarrow$): If not $\Phi \vdash \varphi$, then $\Phi \vdash \neg\varphi$ by negation completeness, but then $\Phi \vdash \psi$ by sequent calculus.

(c)

$$\Phi \vdash \exists x\varphi$$

$\Rightarrow$ { Φ contains witnesses, so $\exists\ t$ s.t. $\Phi \vdash (\exists x\varphi \rightarrow \varphi\frac{t}{x})$, modus ponens }

$$\Phi \vdash \varphi\frac{t}{x}$$

$\Rightarrow$ { ($\exists$S) sequent calculus }

$$\Phi \vdash \exists x\varphi \ \Box$$

Theorem 3 *(Henkin's Theorem)*

If Φ is consistent, negation complete, and contains witnesses, then for all φ, $\mathcal{J}^\Phi \models \varphi$ iff $\Phi \vdash \varphi$.

Proof By induction on the structure of formulas (number of connectives and quantifiers). We already proved it for atomic formulas.

(1) $\varphi = \neg\psi$

$$\mathcal{J}^\Phi \models \neg\psi$$

$\equiv$ { Defs }

not $\mathcal{J}^\Phi \models \psi$

$\equiv$ { Induction hypothesis }

not $\Phi \vdash \psi$

$\equiv$ { Lemma 14 }

$$\Phi \vdash \neg\psi$$

(2) $\varphi = (\psi \vee \xi)$

$$\mathcal{J}^\Phi \models (\psi \vee \xi)$$

$\equiv$ { Defs }

$\mathcal{J}^\Phi \models \psi$ or $\mathcal{J}^\Phi \models \xi$

$\equiv$ { Induction hypothesis }

$\Phi \vdash \psi$ or $\Phi \vdash \xi$

$\equiv$ { Lemma 14 }

$$\Phi \vdash (\psi \vee \xi)$$

(3) $\varphi = \exists x \psi$

$\mathcal{J}^\Phi \models \exists x \psi$

$\equiv$ { Defs, lemma 13 }

there is a t s.t. $\mathcal{J}^\Phi \models \psi \dfrac{t}{x}$

$\equiv$ { Induction hypothesis, rank $\psi \frac{t}{x}$ = rank $\psi <$ rank φ }

$\Phi \vdash \psi \dfrac{t}{x}$

$\equiv$ { Lemma 14 }

$\Phi \vdash \exists x \psi$

7 Satisfiability of Countable Consistent Sets

What we can do now is to show that and consistent set of formulas can be extended to one that is consistent, negation complete, and contains witnesses. Then, from Henkin's theorem we get the completeness theorem.

Once we show the equivalence between $\models$ and $\vdash$, we can transfer properties of one to the other, *e.g.*, we can prove the compactness theorem for $\models$ by tranfering it from the analogous theorem about $\vdash$.

Theorem 4 *(a) $\Phi \models \varphi$ iff there is a finite $\Phi_0 \subseteq \Phi$ such that $\Phi_0 \models \varphi$.*
(b) Sat Φ iff for all finite $\Phi_0 \subseteq \Phi$, Sat Φ_0.

In addition, given that the term interpretation is a model of a set of formulas and that the size of the term interpretation is bound by the size of T^S, we have the Löwenheim-Skolem theorem.

Theorem 5 *Every satisfiable and at most countable set of formulas is satisfiable over a domain which is at most countable.*

8 Gödel's Incompleteness Theorems

8.1 Gödel's First Incompleteness Theorem

Here is an overview of Gödel's incompleteness theorem applied to set theory. A set S is *recursive* iff there is a Turing machine that for any input returs yes or no, depending on whether the input is an element or not. Assuming Con(ZF) (that ZF is consistent), the set $\{\varphi : \text{ZF} \vdash \varphi\}$ is not recursive. (Why do we assume Con(ZF)? Otherwise, all formulas follow from ZF.) More generally, for any consistent extension C of ZF, we have $\{\varphi : C \vdash \varphi\}$ is not recursive. We will not prove this, but it should be intuitively clear: we can embed Turing machines in set theory and we can write a formula that folds iff some Turing machine terminates.

Theorem 6 *(Gödel's first incompleteness theorem.) If C is a recursive consistent extension of ZF, then it is* incomplete*,* i.e.*, there is a formula φ such that $C \not\vdash \varphi$ and $C \not\vdash \neg\varphi$.*

Proof Outine: If not, then for every φ, either $C \vdash \varphi$ or $C \vdash \neg\varphi$. We can now decide $C \vdash \varphi$: enumerate all proofs of C. Stop when a proof for φ or $\neg\varphi$ is found. □

In ZF, the axiom of choice is neither provable nor refutable. In ZFC, the continuum hypothesis is neither provable nor refutable. By Gödel's first incompleteness theorem, no matter how we extend ZFC, there will always be sentences which are neither provable nor refutable.

8.2 G'odel's Second Incompleteness Theorem

This material is from a post to FOM by Harvey Friedman that addresses both of Gödel's incompleteness theorems.

To make things as familiar as possible, we treat PA. We assume familiarity with Turing machines and their formalization in PA.

In particular, we will assume that every $n \geq 0$ is the Gödel number of a Turing machine. We write TM[n] for the n-th Turing machine.

We begin with the description of a particularly simple, fascinating(!) and diabolical(!) Turing machine TM.

At input n, TM searches for a proof in PA that "TM[n] does not halt at n". When it finds one, it immediately halts (and returns 0). Otherwise, TM will not halt.

Let TM be TM[k]. What if we run TM[k] at k?

Case 1. There is a proof in PA that "TM[k] does not halt at k". Then TM[k] halts at k (by the action of TM = TM[k]). But then PA proves "TM[k] halts at k". Since PA is CONSISTENT, this case is impossible.

Case 2. There is no proof in PA that "TM[k] does not halt at k". Then TM[k] does not halt at k (by the action of TM = TM[k]).

Note that we have proved:

There is no proof in PA that "TM[k] does not halt at k". TM[k] does not halt at k.

These two lines give us a form of Gödel's 1st Incompleteness Theorem for PA.

But note, that the proof was done within PA + Con(PA), which we now exploit.

If PA were to prove Con(PA), then PA would prove

There is no proof in PA that "TM[k] does not halt at k". TM[k] does not halt at k.

From this, we see that PA would prove

There is no proof in PA that "TM[k] does not halt at k". PA proves "TM[k] does not halt at k".

Hence PA would be INCONSISTENT.

Thus PA cannot prove its own consistency. This is Gödel's 2nd incompleteness theorem.