CO519 - Theory of Computing - Logic

Part D : First-order logic and its natural deduction proofs

Dominic Orchard
School of Computing, University of Kent

Last updated on October 9, 2017 (incomplete)

If you spot any errors or have suggested edits, the notes are written in LaTeX and are available on GitHub at `http://github.com/dorchard/co519-logic`. Please fork and submit a pull request with any suggested changes.

1 Natural deduction for first-order logic

First-order logic (also known as *predicate logic*) extends propositional logic with quantification: existential quantification $\exists$ ("*there exists....*") and universal quantification $\forall$ ("*for all...*"). First-order logic also adds relations, predicates (unary relations or *classifiers*), and functions. The relations, predicates, and functions are domain-specific for whatever purpose the logic is being used and may be defined externally. Usually some underlying "universe" is fixed over which quantified variables range and on which predicates, relations, and functions are defined.

Consider the following sentence:

Not all birds can fly

We can capture this in first-order logic using quantification and unary predicates. Let's define abstractly two predicates:

$$\begin{aligned}\mathsf{B}(x) &: x \textit{ is a bird}\\ \mathsf{F}(x) &: x \textit{ can fly}\end{aligned}$$

Our universe here might be "animals" or just general objects. As with propositional logic, we are studying the process and framework of logic rather than any connections of certain logical statements to physical reality; it is up to us how we assign the semantics of B and F above, but the semantics of quantification and logical operators is fixed by the definition of first-order logic.

We can then express the above sentence in predicate logic as:

$$\neg(\forall x.\mathsf{B}(x) \rightarrow \mathsf{F}(x)) \tag{1}$$

We can read this exactly as *it is not true that for all x, if x is a bird then x can fly.* Another way to write this is that there are some birds which cannot fly:

$$\exists x.\mathsf{B}(x) \wedge \neg\mathsf{F}(x) \tag{2}$$

If we have a semantics for B and F that includes, for example, penguins then both (1) and (2) will be true.

We can prove that these two statements ((1) and (2)) are equivalent in predicate logic via two proofs, one for:

$$\neg(\forall x.\mathsf{B}(x) \rightarrow \mathsf{F}(x)) \vdash \exists x.\mathsf{B}(x) \wedge \neg\mathsf{F}(x)$$

and one for:

$$\exists x.\mathsf{B}(x) \wedge \neg\mathsf{F}(x) \vdash \neg(\forall x.\mathsf{B}(x) \rightarrow \mathsf{F}(x))$$

We will do this later once we have explained more about the meta-theory of the logic, as well as the introduction and elimination rules for quantification.

1.1 Names and binding

In propositional logic, variables range over propositions, *i.e.*, their "type" is a proposition. For example, $x \wedge y$ has two propositional variables x and y which could be replaced with true or false, or with any other formula. In predicate logic, universal and existential quantifiers provide *variable bindings*, which introduce variables ranging over objects in some universe rather than over propositions. For example, the formula $\forall x.P$ binds a variable x in the *scope* of P. That is, x is available within P, but not outside of it. A variable which does not have a binding in scope is called *free*.

For example, the following formula has free variables x and y and bound variables u and v:

$$\mathsf{P}(x) \vee \forall u.(\mathsf{Q}(y) \wedge \mathsf{R}(u) \rightarrow \exists v.(\mathsf{P}(v) \wedge \mathsf{Q}(x)))$$

The following repeats the formula and highlights the binders in yellow, the bound variables in green, and the free variables in red:

$$\mathsf{P}(x) \vee \forall u.(\mathsf{Q}(y) \wedge \mathsf{R}(u) \rightarrow \exists v.(\mathsf{P}(v) \wedge \mathsf{Q}(x)))$$

In the following formula, there are two syntactic occurrences of a variable called x, but semantically these are different variables:

$$\mathsf{Q}(x) \wedge (\forall x.\mathsf{P}(x))$$

The x on the left (used with a predicate Q) is free, whilst the x used with the predicate P is bound by the universal quantifier. Thus, these are semantically two different variables.

Alpha renaming The above formula is semantically equivalent to the following formula obtained by consistently renaming bound variables:

$$\mathsf{Q}(x) \wedge (\forall y.\mathsf{P}(y))$$

Renaming variables is a meta-level operation we can apply to any formula: we can rename a bound variable as long as we do not rename it to clash with any other free or bound variable names, and as long as we rename the variable consistently. This principle is more generally known as α-renaming (alpha renaming) and equality up-to renaming (equality that accounts for renaming) is known as α-equality. For example, writing α-equality as $=_\alpha$ the following equality and inequality hold:

$$\exists x.\mathsf{P}(x) \to \mathsf{P}(y) \quad =_\alpha \quad \exists z.\mathsf{P}(z) \to \mathsf{P}(y) \quad \neq_\alpha \quad \exists y.\mathsf{P}(y) \to \mathsf{P}(y)$$

The middle formula can be obtained from the left formula by renaming x to a fresh variable z. However, in the right-hand formula, we have renamed x to y which conflates the bound variable with the free variable y on the right of the implication. The formula on the right has a different meaning to the left two.

1.2 Substitution

Recall in Part C, we used the function *replace* in the DPLL algorithm where *replace*(x, Q, P) rewrites formula P such that any occurrences of variable x are replaced with formula Q. This is more generally called *substitution*.

From now on we will use a more compact syntax for substitution written

$$P[t/x]$$

which means: *replace variable x with the variable t in formula P* (akin to *replace*(x, t, P)). We will only replace variables with other variables. Note however that in predicate logic we have to careful about free and bound variables. Thus, $P[t/x]$ means replace any *free* occurrences of x in P with object t. (One way to remember this notation is to observe that the letters used in the general form above are in alphabetical order: P then t then x to give $P[t/x]$ for replacing x with t in P). This is a common notation which is also used in the course textbook.

We must be careful to replace only the free occurrences of variables, that is, those variables which are not in the scope of a variable binding of the same name. For example, in the following we have a free x and a bound x, so substitution only affects the free x as such:

$$(\mathsf{P}(x) \wedge \forall x.\mathsf{P}(x))[t/x] = \mathsf{P}(t) \wedge \forall x.\mathsf{P}(x)$$

In general, it is best practise to give bound variables a different name to all other free and bound names in a formula, in order to avoid confusion.

1.3 Natural deduction rules

As with propositional logic, we will have elimination and introduction rules for the new logical operators in first-order logic.

1.3.1 Universal quantification (*for all*)

Universal quantification essentially generalises conjunction. That is, if the objects in the universe over which we are quantifying are $a_0, a_1, \ldots a_n \in \mathcal{U}$ then universal quantification of x over a formula P is equivalent to taking the repeated conjunction of P substituting each object for x, *i.e.*

$$\forall x.P = P[a_0/x] \wedge P[a_1/x] \wedge \ldots \wedge P[a_n/x]$$

(Note that there may be an infinite number of such objects). This perspective helps us to understand elimination and introduction for universal quantification as a generalisation of elimination and introduction for conjunction:

$$\frac{P \wedge Q}{P} \wedge_{e1} \quad \frac{P \wedge Q}{Q} \wedge_{e2} \quad \frac{P \qquad Q}{P \wedge Q} \wedge_{i}$$

The elimination rule for universal quantification is then as follows:

$$\frac{\forall x.P}{P[t/x]} \forall_e \quad \textit{where } t \textit{ is free when replacing } x \textit{ in } P$$

The intuition is that we can eliminate a $\forall$ by replacing the bound variable x with an arbitrary variable t which represents *any* of things ranged over by the $\forall$. This is similar to conjunction elimination where we eliminate to either of the left or right-hand sides of the conjunction.

The side condition requires that the variable t is a free variable when it is substituted for x in P. We'll see an example of why this is needed. [1]

Consider the following statement about integers, that for every integer there exists a bigger integer:

$$\forall x.\exists y.(x < y)$$

In a natural deduction proof, we can eliminate the $\forall$ with $t = x_0$ (some fresh variable) to get $\exists y.(x_0 < y)$. We know nothing about x_0, it is just a variable representing any object (integer) in the set of things ranged over by the $\forall$. If we instead performed the elimination with the substitution where $t = y$ (violating the side condition) then we would get $\exists y.(y < y)$ which is no longer true: it says that there exists a number which is greater than itself. The problem is that by performing the substitution $(\exists y.(x < y))[y/x]$ we have "captured" the binding of y with the y we are substituting in, which then changed the meaning of the formula. This is why we have the side condition.

Next we will see introduction. This uses boxes like we used for subproofs previously, but the boxes are no longer subproofs, but instead mark out the

[1] The textbook (Huth and Ryan) uses the phrase "*t is free for x in P*" which is a little confusing, but means the same as the above side condition for $\forall$ elimination: that the variable t remains a free variable even once it replaces x inside of P.

scope of a variable. The rule is as follows:

$$\frac{\boxed{\begin{array}{l} x_0 \\ \qquad \vdots \\ \qquad P[x_0/x] \end{array}}}{\forall x.P}\ \forall_i$$

This says that there is a *scope* (not subproof) that has a variable x_0 (marked in blue in the top corner) which appears in a proof concluding with $P[x_0/x]$. If this variable x_0 is only used in this scope, then we can leave the scope and conclude $\forall x.P$. Thus, the proof inside is of P but where x is replaced by x_0. Here's an example to make this idea more clear:

Example 1. Prove $\forall x.(\mathsf{P}(x) \to \mathsf{Q}(x)), \forall y.\mathsf{P}(y) \vdash \forall z.\mathsf{Q}(z)$ is valid.

That is, for all x such that $\mathsf{P}(x)$ implies $\mathsf{Q}(x)$ and for all y that $\mathsf{P}(y)$ then $\mathsf{Q}(z)$ holds for all z. Note how I have used different names for the bound variables for the sake of clarity, but the formula would have the exact same meaning if each bound variable was called x here.

The proof proceeds:

1.		$\forall x.(\mathsf{P}(x) \to \mathsf{Q}(x))$	premise
2.		$\forall y.\mathsf{P}(y)$	premise
3.	x_0	$\mathsf{P}(x_0) \to \mathsf{Q}(x_0)$	$\forall e$ 1
4.		$\mathsf{P}(x_0)$	$\forall e$ 2
5.		$\mathsf{Q}(x_0)$	$\to_e$ 3, 4
6.		$\forall z.\mathsf{Q}(z)$	$\forall i$ 3-5

We start with the two premises. The proof then proceeds with a scope box on lines 3-5: *but remember this is not a subproof*, it merely serves to delimit the scope of the fresh variable x_0 which starts on line 3 when the "for all" on line 1 is eliminated.

When we eliminate $\forall y.\mathsf{P}(y)$ on line 4 we use the same x_0. Since the quantification is universal we can pick the same object x_0 to eliminate line 4 that we "picked" for the elimination on line 3. Line 5 uses modus ponens to get $\mathsf{Q}(x_0)$ which gives us the formula on which we apply $\forall i$.

Philosophically we are applying universal quantification introduction only on $\mathsf{Q}(x_0)$, but we state the lines in which x_0 is in scope. Once we close this scope box, we can't do anything with x_0. The proof has shown that if we pick an arbitrary object x_0 then we get $Q(x_0)$ and so we can conclude that for all objects z we have $\mathsf{Q}(z)$ (expressed as $\forall z.\mathsf{Q}(z)$).

Existential quantification is more restrictive.

1.3.2 Existential quantification (*there exists*)

Whilst universal quantification generalises conjunction, existential quantification generalises disjunction. If existential quantification binds variables ranging over the objects $a_0, a_1, \ldots a_n$ then:

$$\exists x.P = P[a_0/x] \vee P[a_1/x] \vee \ldots \vee P[a_n/x] \tag{3}$$

That is, existential quantification is equivalent to the disjunction of P for every object in the universe, replacing x.

Recall the introduction and elimination rules for disjunction:

$$\frac{P}{P \vee Q}\,\vee_{i1} \qquad \frac{Q}{P \vee Q}\,\vee_{i2} \qquad \frac{P \vee Q \quad \boxed{\begin{array}{c} P \\ \vdots \\ R \end{array}} \quad \boxed{\begin{array}{c} Q \\ \vdots \\ R \end{array}}}{R}\,\vee_e$$

Disjunction introduction generalises to the following existential introduction rule:

$$\frac{P[t/x]}{\exists x.P}\,\exists_i \;\; \textit{t is free for x in P}$$

That is, we can introduce $\exists x.P$ if we have P where some other variable t replaces x. This is a bottom-up reading of the rule, where the substitution is applied to the premise. Note how this rule is essentially the converse of $\forall$*elimination* and thus we can rather directly prove the following:

Example 2. Prove $\forall x.\mathsf{P}(x) \vdash \exists x.\mathsf{P}(x)$ is valid.

$$\begin{array}{lll} 1. & \forall x.\mathsf{P}(x) & \text{premise} \\ 2. & \mathsf{P}(t) & \forall_e \\ 3. & \exists x.\mathsf{P}(x) & \exists_i \qquad \square \end{array}$$

Existential elimination resembles disjunction elimination but now combines the notion of a variable scope box with a subproof box:

$$\frac{\exists x.P \quad \boxed{\begin{array}{cc} x_0 & P[x_0/x] \\ & \vdots \\ & Q \end{array}}}{Q}$$

The intuition is that the subproof here represents a case for every single part of the disjunction in (3) by using an arbitrary variable x_0 that we know nothing about to represent each atom in the universe. Similarly to disjunction elimination, if we can then conclude Q (but without using x_0 since the scope box ends here) then we can conclude Q overall.

Exercise 1.1. $\neg\forall x.\mathsf{P}(x) \vdash \exists x.\neg\mathsf{P}(x)$

2 Collected rules of natural deduction for first-order logic

	Introduction	*Elimination*
$\forall$	$\dfrac{\boxed{\begin{array}{ll} x_0 & \\ & \vdots \\ & P[x_0/x] \end{array}}}{\forall x.P}\ \forall_i$	$\dfrac{\forall x.P}{P[t/x]}\ \forall e$ *where t is free when replacing x in P*
$\exists$	$\dfrac{P[t/x]}{\exists x.P}\ \exists_i$ *where t is free when replacing x in P*	$\dfrac{\exists x.P \quad \boxed{\begin{array}{ll} x_0 & P[x_0/x] \\ & \vdots \\ & Q \end{array}}}{Q}$

3 Exercises

Exercise 1.1. $\neg\forall x.\mathsf{P}(x) \vdash \exists x.\neg\mathsf{P}(x)$