The Hilbert Proof System

In secondary school, you probably took a course in plane geometry in which you were required to construct formal, step-by-step proofs which established things such as “triangle A is congruent to triangle B.” A proof system for a logic has the same flavor, but is focused on logic rather than geometry.

Example: Let L be the propositional logic with $P = \{A, B, C\}$. Let

$\Phi = \{A, (A \rightarrow B), (B \rightarrow C)\}$;

$\varphi = C$.

Suppose that it is desired to establish that $\Phi \models \varphi$ holds. A (yet to be formalized) “proof” might run as follows:

1. A	(hypothesis)
2. $(A \rightarrow B)$	(hypothesis)
3. B	(2 applied to 1)
4. $(B \rightarrow C)$	(hypothesis)
5. C	(4 applied to 3)

Generally, proofs have this linear form, in which each statement is either given, or else is a consequence of previous statements.

Here is how that idea is formalized.

Definition: Let $L = (P, A, C)$ be a propositional logic. A *proof system* for L consists of the following:

(a) a set of *proof rules* for deducing new statements from existing ones, and
(b) a set of *axiom schemata* for generating tautologies.

This really is not much of a definition, because the meaning of proof rule and axiom schema have not been given. A formal definition is possible, but tedious, so we will illustrate with the key examples instead.

The most famous proof rule is *modus ponens*, denoted MP. It is written as follows:

$$\frac{\alpha_1,\ (\alpha_1 \rightarrow \alpha_2)}{\alpha_2}$$

The formulas above the line are patterns to be matched for the rule to apply. The formula below the line is the result. Thus, this rule says that if α_1 and α_2 are any two formulas whatever, and if it is known that both α_1 and $(\alpha_1 \rightarrow \alpha_2)$ are true, then it may be concluded that α_2 is true. The patterns α_1 and $(\alpha_1 \rightarrow \alpha_2)$ are called the *premises*, and α_2 is called the *conclusion.*

In the “proof” on the previous slide, it was modus ponens which was applied at steps 3 and 5.

An *axiom schema* is a pattern which yields a wff when wff's are subsitituted for its parameters. For example,

$$(\alpha_1 \rightarrow (\alpha_2 \rightarrow \alpha_1))$$

is an axiom schema. When any wff's whatever are substituted for α_1 and α_2, an *instantiation* of the axiom schema is obtained. To be useful, an axiom schema should always yield instantiations which are tautologies. Notice that since *any* wff may be substituted for α_1 and for α_2, this schema will generate an infinite number of distinct formulas.

Formally, an axiom schema may be viewed as a special case of a proof rule; that is, one with no premises. With that interpretation, one might write the above axiom schema as

$$\frac{}{(\alpha_1 \rightarrow (\alpha_2 \rightarrow \alpha_1))} \quad \text{or} \quad \frac{\varnothing}{(\alpha_1 \rightarrow (\alpha_2 \rightarrow \alpha_1))}$$

To see what axiom schemata are all about, let us modify the previous example a bit. Now let L be the propositional logic with $P = \{A, B, C, D\}$, and let

$$\Phi = \{A, (A \rightarrow B), (B \rightarrow C)\};$$
$$\psi = ((\neg D) \rightarrow C).$$

It should be clear that $\Phi \models \psi$. Indeed, $\psi \equiv (D \vee C)$, which is a weaker conclusion than $\varphi = C$. However, a proof using only modus ponens is not possible. The use of the above axiom schema is necessary.

1. A	(hypothesis)
2. $(A \rightarrow B)$	(hypothesis)
3. B	(modus ponens on 1,2)
4. $(B \rightarrow C)$	(hypothesis)
5. C	(modus ponens on 3,4)
6. $(C \rightarrow ((\neg D) \rightarrow C))$	(axiom schema)
7. $((\neg D) \rightarrow C)$	(modus ponens on 5,6) □

A *proof system* is a pair $\Gamma = (S,R)$ in which S is a (possibly empty) finite set of axiom schemata and R is a finite set of proof rules.

A *proof* in $\Gamma = (S,R)$ of the wff φ from the set of wffs Φ is a sequence $\varphi_1, \varphi_2, .., \varphi_n$ of wff's, with the following properties.

(a) Each φ_i is either:
  - (i) A member of Φ;
  - (ii) An instantiation of a member of S;, or
  - (iii) The consequent of an instantiation of a proof rule $\rho \in R$ for which each instantiation of its premises is one of the φ_j, with $j < i$.

(b) $\varphi_n = \varphi$.

$\Phi \vdash_\Gamma \varphi$ denotes that φ is provable from Φ in Γ.

Definition: The *Hilbert System* (denoted H) for propositional logic consists of the following three axiom schemata

Ax_1: $((\alpha_1 \rightarrow (\alpha_2 \rightarrow \alpha_1))$
Ax_2: $((\alpha_1 \rightarrow (\alpha_2 \rightarrow \alpha_3)) \rightarrow ((\alpha_1 \rightarrow \alpha_2) \rightarrow (\alpha_1 \rightarrow \alpha_3)))$
Ax_3: $(((\neg\alpha_1) \rightarrow (\neg\alpha_2)) \rightarrow (\alpha_2 \rightarrow \alpha_1))$

together with the single proof rule modus ponens, repeated below.

$$\frac{\alpha_1,\ (\alpha_1 \rightarrow \alpha_2)}{\alpha_2}$$

Here is the proof within the Hilbert system, from a previous slide, with the correct justifications listed.

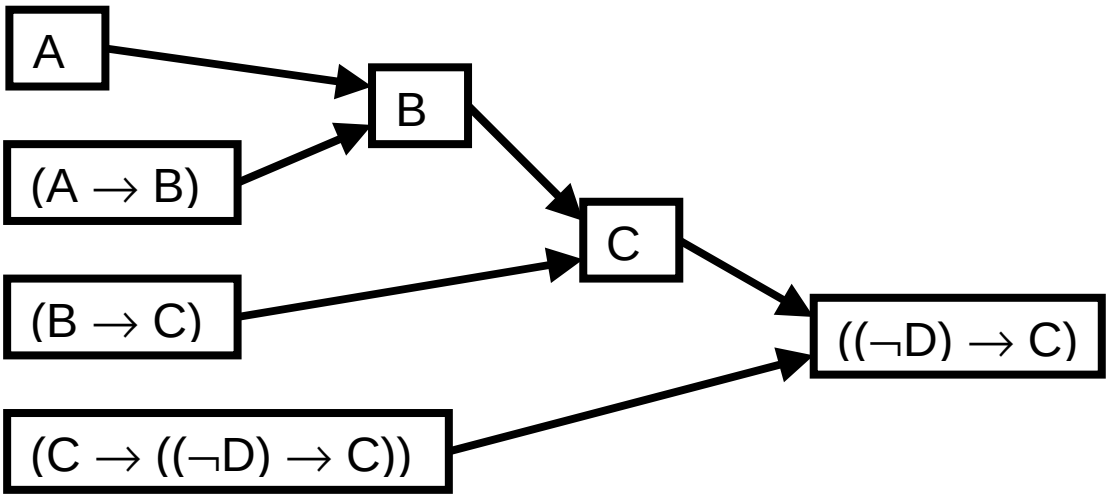
1. A	(hypothesis)
2. $(A \rightarrow B)$	(hypothesis)
3. B	(modus ponens on 1,2)
4. $(B \rightarrow C)$	(hypothesis)
5. C	(modus ponens on 3,4)
6. $(C \rightarrow ((\neg D) \rightarrow C))$	(axiom schema Ax_1 on C, $((\neg D) \rightarrow C)$)
7. $((\neg D) \rightarrow C)$	(modus ponens on 5,6) □

Note that the instantiation of the axiom schema need not be by axioms; any wff may be used.

The order is not critical, as long as the premises of each application of an inference rule precede the use of that rule. Here an alternate proof.

1. $(C \rightarrow ((\neg D) \rightarrow C))$	(axiom schema Ax_1 on C,$((\neg D) \rightarrow C))$)
2. A	(hypothesis)
3. $(B \rightarrow C)$	(hypothesis)
4. $(A \rightarrow B)$	(hypothesis)
5. B	(modus ponens on 2,4)
6. C	(modus ponens on 5,3)
7. $((\neg D) \rightarrow C)$	(modus ponens on 6,1) □

It is sometimes convenient to represent the proof with a directed acyclic graph (DAG), rather than with a linear list. This makes transparent the actual dependencies amongst the elements of the proof. Any linearization of the graph will provide a sequential proof. Shown below is the graph for this example.



Proofs in the Hilbert System:

Example: (This is (Meta-)Theorem 4.1 of the textbook.) Establish that for any wff φ,

$$\vdash_\Gamma (\varphi \to \varphi).$$

From a semantic point of view, this is obvious. That is, we know that

$$\models (\varphi \to \varphi).$$

However, we want to show that it is *provable* in the Hilbert system. This task is surprisingly nontrivial. No hypotheses are given, so the only alternative to obtain a wff to begin the proof is to use an instantiation of one of the axiom schemata. The following proof is the same as that in the textbook, except for names. Note that ψ can be any wff whatever in this proof including φ.

1. $((\varphi \to ((\psi \to \varphi) \to \varphi))$
   $[Ax_1: \alpha_1 \leftarrow \varphi; \alpha_2 \leftarrow (\psi \to \varphi)]$

2. $((\varphi \to ((\psi \to \varphi) \to \varphi)) \to$
   $((\varphi \to (\psi \to \varphi) \to (\varphi \to \varphi)))$
   $[Ax_2: \alpha_1 \leftarrow \varphi; \alpha_2 \leftarrow (\psi \to \varphi); \alpha_3 \leftarrow \varphi]$

3. $(((\varphi \to (\psi \to \varphi)) \to (\varphi \to \varphi))$
   $[MP: \alpha_1 \leftarrow (1.); \alpha_2 \leftarrow (2.)]$

4. $(\varphi \to (\psi \to \varphi))$
   $[Ax_1: \alpha_1 \leftarrow \varphi; \alpha_2 \leftarrow (\psi \to \varphi)]$

5. $(\varphi \to \varphi)$ $[MP: \alpha_1 \leftarrow (4.); \alpha_2 \leftarrow (3.)]$ □

Some FAQ’s about the Hilbert system:

Q: How does one know which axiom schemata to use, and which substitutions to make? Since there are infinitely many possibilities, it is not possible to try them all, even in princple.

A: There is no algorithm; at least no simple one. Rather, one has to be clever. In pure mathematics, this is not viewed as a problem, since one is most concerned about the existence of a proof. However, in computer science applications, one is interested in automating the deduction process, so this is a fatal flaw. The Hilbert system is not normally used in automated theorem proving.

Q: So, why do people care about the Hilbert system?

A: With modus ponens as its single deductive rule, it provides a palatable model of how humans devise mathematical proofs. As we shall see, methods which are more amenable to computer implementation produce proofs which are less “human like.”

Despite these shortcomings, it is possible to simplify proof in the Hilbert system somewhat by generating meta-theorems which may later be used in proofs.

(Meta-)Theorem 4.3 of the textbook:

$$\{(\varphi_1 \rightarrow \varphi_2), (\varphi_2 \rightarrow \varphi_3)\} \vdash_H (\varphi_1 \rightarrow \varphi_3)$$

Proof:

1. $(\varphi_2 \rightarrow \varphi_3)$ [hypothesis]

2. $((\varphi_2 \rightarrow \varphi_3) \rightarrow (\varphi_1 \rightarrow (\varphi_2 \rightarrow \varphi_3)))$
   [Ax_1: $\alpha_1 \leftarrow (\varphi_2 \rightarrow \varphi_3)$; $\alpha_2 \leftarrow \varphi_1$]

3. $(\varphi_1 \rightarrow (\varphi_2 \rightarrow \varphi_3))$
   [MP: $\alpha_1 \leftarrow$ (1.); $\alpha_2 \leftarrow$ (2.)]

4. $((\varphi_1 \rightarrow (\varphi_2 \rightarrow \varphi_3)) \rightarrow ((\varphi_1 \rightarrow \varphi_2) \rightarrow (\varphi_1 \rightarrow \varphi_3)))$
   [Ax_2: $\alpha_1 \leftarrow \varphi_1$; $\alpha_2 \leftarrow \varphi_2$; $\alpha_3 \leftarrow \varphi_3$]

5. $((\varphi_1 \rightarrow \varphi_2) \rightarrow (\varphi_1 \rightarrow \varphi_3))$
   [MP: $\alpha_1 \leftarrow$ (3.); $\alpha_2 \leftarrow$ (4.)]

6. $(\varphi_1 \rightarrow \varphi_2)$ [hypothesis]

7. $(\varphi_1 \rightarrow \varphi_3)$ [MP: $\alpha_1 \leftarrow$ (6.); $\alpha_2 \leftarrow$ (5.)] □

(Meta-)Theorem 4.2 of the textbook:

$$\vdash_H \ ((\neg\varphi_1) \rightarrow (\varphi_1 \rightarrow \varphi_2))$$

The textbook provides two proofs of this statement. One is a pure proof, using only the axiom schemata and modus ponens. The second, which is much shorter, makes use of (Meta-)Theorem, which is proven above. Here is the second proof:

1. $((\neg\varphi_1) \rightarrow ((\neg\varphi_2) \rightarrow (\neg\varphi_1)))$
   [Ax_1: $\alpha_1 \leftarrow (\neg\varphi_1)$; $\alpha_2 \leftarrow (\neg\varphi_2)$]

2. $(((\neg\varphi_2) \rightarrow (\neg\varphi_1)) \rightarrow (\varphi_1 \rightarrow \varphi_2))$
   [Ax_3: $\alpha_1 \leftarrow \varphi_2$; $\alpha_2 \leftarrow \varphi_1$]

3. $((\neg\varphi_1) \rightarrow (\varphi_1 \rightarrow \varphi_2))$
   [Th. 4.3 on (1.) and (2.)]

Several other (Meta-)Theorems are presented in the textbook, the most important of which is the following:

The Deduction Theorem:

$$\Phi \cup \{\varphi_1\} \vdash_H \varphi_2 \ \text{ iff } \ \Phi \ \vdash_H (\varphi_1 \rightarrow \varphi_2)$$

The proof is an induction on the length of the proof. It will not be reproduced here.

You should read through, and understand the statements of, the other meta-theorems in Chapter 4 of the textbook.

Remark on notation: The deduction theorem may also be written as

$$\Phi \cup \{\varphi_1\} \vdash_H \varphi_2 \;\Leftrightarrow\; \Phi \vdash_H (\varphi_1 \rightarrow \varphi_2)$$

provided that $\Leftrightarrow$ is interpreted as a meta-symbol. The textbook author is inconsistent in his notation on such issues. Note, however, that it is **not** correct to write

$$\Phi \cup \{\varphi_1\} \vdash_H \varphi_2 \;\leftrightarrow\; \Phi \vdash_H (\varphi_1 \rightarrow \varphi_2).$$

In the notation used in the text and these notes, $\leftrightarrow$ is a logical connective (an alternate to $\equiv$). It is not a meta-symbol.

Properties of proof systems:

Soundness: A proof system Γ is *sound* if, for any wff φ,

$$\vdash_\Gamma \varphi \quad \text{implies} \quad \models \varphi.$$

In words, a proof system is sound if everything which can be proven is true. To be useful, a proof system should always be sound.

Completeness: A proof system Γ is *complete* if, for any wff φ,

$$\models \varphi \quad \text{implies} \quad \vdash_\Gamma \varphi.$$

In words, a proof system is sound if everything which is true can be proven. This is a very desirable, although it may be compromised in the name of efficiency.

Theorem: The Hilbert proof system H is both sound and complete. □

Decidability:

- Decidability is a property of the logic, not of the proof system. (The textbook is unclear on this issue.)

An *algorithm* is a program which always halts. It never loops forever looking for an answer.

A logic $\mathcal{L}$ is *decidable* if there is an algorithm which takes as input an arbitrary wff φ in the logic and provides as output the answer to the question:

"Does $\models \varphi$ hold?"

Theorem: Propositional logic is decidable.
Proof: If the number of propositions in the logic is finite, this is trivial. All one need do is construct a truth table. Suppose that the number of propositions is infinite. Even in this case, if we are asked to determine whether or not $\models \varphi$ holds, the formula φ can contain only a finite number of proposition names. Restricting the truth table to these proposition names is sufficient, so the truth table method may be employed in this case as well. □

Algorithms for deduction:

- The truth-table method, and often even the semantic-tableaux method, can be very inefficient in certain circumstances. One of the principal reasons that one employs a proof system is to obtain a more efficient proof procedure.

- Unfortunately, the Hilbert proof system fails to provide an algorithm for generating proofs. It is clear that if “intelligent” control on the application of the axiom schemata is not imposed, the process of generating new elements in a proof sequence can proceed endlessly.

- Next, we will examine another proof system (resolution), which overcomes this difficulty.

Deduction of non-tautologies:

It might seem restrictive that attention is paid solely to establishing that a formula is a tautology; that is, to establishing that

$$\models \varphi.$$

For general theorem proving, the more general situation

$$\Phi \models \varphi$$

must be established. However, as demonstrated in the notes on semantic tableaux, $\Phi \models \varphi$ holds iff $(\Phi_\wedge \wedge (\neg\varphi))$ is unsatisfiable, where $\Phi_\wedge$ is the conjunction of all formulas in Φ. However, $(\Phi_\wedge \wedge (\neg\varphi))$ is unsatisfiable iff $(\neg(\Phi_\wedge \wedge (\neg\varphi)))$ is a tautology. However, the latter is equivalent to $(\neg\Phi_\wedge \vee \varphi)$ and so to $(\Phi_\wedge \rightarrow \varphi)$. Thus,

$$\Phi \models \varphi$$

is equivalent to

$$\models (\Phi_\wedge \rightarrow \varphi).$$

Notice that the deduction theorem is a syntactic version of this idea.