2 Propositional Logic

Propositional logic

- logic of truth values,
- decidable (but NP-complete),
- can be used to describe functions over a finite domain,
- industry standard for many analysis/verification tasks (e. g., model checking).

2.1 Syntax

When we define a logic, we must define how formulas of the logic look like (syntax), and what they mean (semantics). We start with the syntax.

Propositional formulas are built from

- propositional variables,
- logical connectives (e. g., $\wedge$, $\vee$).

Propositional Variables

Let Π be a set of *propositional variables*.

We use letters P, Q, R, S, to denote propositional variables.

Propositional Formulas

F_Π is the set of propositional formulas over Π defined inductively as follows:

$$
\begin{array}{rcll}
F, G & ::= & \bot & \text{(falsum)} \\
 & \mid & \top & \text{(verum)} \\
 & \mid & P, \quad P \in \Pi & \text{(atomic formula)} \\
 & \mid & (\neg F) & \text{(negation)} \\
 & \mid & (F \wedge G) & \text{(conjunction)} \\
 & \mid & (F \vee G) & \text{(disjunction)} \\
 & \mid & (F \rightarrow G) & \text{(implication)} \\
 & \mid & (F \leftrightarrow G) & \text{(equivalence)}
\end{array}
$$

Sometimes further connectives are used, for instance

$$
\begin{array}{ll}
(F \leftarrow G) & \text{(reverse implication)} \\
(F \oplus G) & \text{(exclusive or)} \\
(\text{if } F \text{ then } G_1 \text{ else } G_0) & \text{(if-then-else)}
\end{array}
$$

Notational Conventions

As a notational convention we assume that $\neg$ binds strongest, and we remove outermost parentheses, so $\neg P \lor Q$ is actually a shorthand for $((\neg P) \lor Q)$.

Instead of $((P \land Q) \land R)$ we simply write $P \land Q \land R$ (analogously for $\lor$).

For all other logical connectives we will use parentheses when needed.

Formula Manipulation

Automated reasoning is very much formula manipulation. We perform syntactic operations on formulas in order to show semantic properties of formulas.

To precisely describe the manipulation of a formula, we introduce positions.

A *position* is a word over $\mathbb{N}$. The set of positions of a formula F is inductively defined by

$$\begin{aligned}
\mathrm{pos}(F) &:= \{\varepsilon\} \text{ if } F \in \{\top, \bot\} \text{ or } F \in \Pi \\
\mathrm{pos}(\neg F) &:= \{\varepsilon\} \cup \{\, 1p \mid p \in \mathrm{pos}(F) \,\} \\
\mathrm{pos}(F \circ G) &:= \{\varepsilon\} \cup \{\, 1p \mid p \in \mathrm{pos}(F) \,\} \cup \{\, 2p \mid p \in \mathrm{pos}(G) \,\} \\
&\ \text{where } \circ \in \{\land, \lor, \rightarrow, \leftrightarrow\}.
\end{aligned}$$

The prefix order $\le$ on positions is defined by $p \le q$ if there is some p' such that $pp' = q$.

Note that the prefix order is partial, e. g., the positions 12 and 21 are not comparable, they are "parallel", see below.

By $<$ we denote the strict part of $\le$, that is, $p < q$ if $p \le q$ but not $q \le p$.

By $\parallel$ we denote incomparable positions, that is, $p \parallel q$ if neither $p \le q$ nor $q \le p$.

We say that p is *above* q if $p \le q$, p is *strictly above* q if $p < q$, and p and q are *parallel* if $p \parallel q$.

The *size* of a formula F is given by the cardinality of $\mathrm{pos}(F)$: $|F| := |\mathrm{pos}(F)|$.

The *subformula* of F at position $p \in \mathrm{pos}(F)$ is recursively defined by

$$\begin{aligned}
F|_\varepsilon &:= F \\
(\neg F)|_{1p} &:= F|_p \\
(F_1 \circ F_2)|_{ip} &:= F_i|_p \quad \text{where } i \in \{1, 2\} \\
& \text{and } \circ \in \{\land, \lor, \rightarrow, \leftrightarrow\}.
\end{aligned}$$

Finally, the *replacement* of a subformula at position $p \in \text{pos}(F)$ by a formula G is recursively defined by

$$\begin{aligned} F[G]_\varepsilon &:= G \\ (\neg F)[G]_{1p} &:= \neg(F[G]_p) \\ (F_1 \circ F_2)[G]_{1p} &:= (F_1[G]_p \circ F_2) \\ (F_1 \circ F_2)[G]_{2p} &:= (F_1 \circ F_2[G]_p) \\ &\quad \text{where } \circ \in \{\wedge, \vee, \rightarrow, \leftrightarrow\}. \end{aligned}$$

Example 2.1 *The set of positions for the formula* $F = (P \rightarrow Q) \rightarrow (P \wedge \neg R)$ *is* $\text{pos}(F) = \{\varepsilon, 1, 11, 12, 2, 21, 22, 221\}$.

The subformula at position 22 *is* $F|_{22} = \neg R$ *and replacing this formula by* $P \leftrightarrow Q$ *results in* $F[P \leftrightarrow Q]_{22} = (P \rightarrow Q) \rightarrow (P \wedge (P \leftrightarrow Q))$.

2.2 Semantics

In *classical logic* (dating back to Aristotle) there are "only" two truth values "true" and "false" which we shall denote, respectively, by 1 and 0.

There are *multi-valued logics* having more than two truth values.

Valuations

A propositional variable has no intrinsic meaning. The meaning of a propositional variable has to be defined by a valuation.

A Π-*valuation* is a function $\mathcal{A} : \Pi \rightarrow \{0, 1\}$ where $\{0, 1\}$ is the set of *truth values*.

Truth Value of a Formula in $\mathcal{A}$

Given a Π-valuation $\mathcal{A} : \Pi \rightarrow \{0, 1\}$, its extension to formulas $\mathcal{A}^* : \mathrm{F}_\Pi \rightarrow \{0, 1\}$ is defined inductively as follows:

$$\begin{aligned} \mathcal{A}^*(\bot) &= 0 \\ \mathcal{A}^*(\top) &= 1 \\ \mathcal{A}^*(P) &= \mathcal{A}(P) \\ \mathcal{A}^*(\neg F) &= 1 - \mathcal{A}^*(F) \\ \mathcal{A}^*(F \wedge G) &= \min(\mathcal{A}^*(F), \mathcal{A}^*(G)) \\ \mathcal{A}^*(F \vee G) &= \max(\mathcal{A}^*(F), \mathcal{A}^*(G)) \\ \mathcal{A}^*(F \rightarrow G) &= \max(1 - \mathcal{A}^*(F), \mathcal{A}^*(G)) \\ \mathcal{A}^*(F \leftrightarrow G) &= \text{if } \mathcal{A}^*(F) = \mathcal{A}^*(G) \text{ then } 1 \text{ else } 0 \end{aligned}$$

For simplicity, the extension $\mathcal{A}^*$ of $\mathcal{A}$ is usually also denoted by $\mathcal{A}$.

Note that formulas and truth values are disjoint classes of objects. Statements like $P = 1$ or $F \wedge G = 0$ that equate formulas and truth values are non-sensical. A formula is never *equal to* a truth value, but it *has* a truth value in some valuation $\mathcal{A}$.

2.3 Models, Validity, and Satisfiability

Let F be a Π-formula.

We say that F is *true* in $\mathcal{A}$ ($\mathcal{A}$ is a *model* of F; F *is valid* in $\mathcal{A}$; F *holds* in $\mathcal{A}$), written $\mathcal{A} \models F$, if $\mathcal{A}(F) = 1$.

We say that F is *valid* or that F is a *tautology*, written $\models F$, if $\mathcal{A} \models F$ for all Π-valuations $\mathcal{A}$.

F is called *satisfiable* if there exists an $\mathcal{A}$ such that $\mathcal{A} \models F$. Otherwise F is called *unsatisfiable* (or *contradictory*).

Entailment and Equivalence

F *entails* (*implies*) G (or G *is a consequence of* F), written $F \models G$, if for all Π-valuations $\mathcal{A}$ we have

$$\text{if } \ \mathcal{A} \models F \ \text{ then } \ \mathcal{A} \models G,$$

or equivalently

$$\mathcal{A}(F) \leq \mathcal{A}(G).$$

F and G are called *equivalent*, written $F \models\!\!\!\!\!\!= G$, if for all Π-valuations $\mathcal{A}$ we have

$$\mathcal{A} \models F \ \text{ if and only if } \ \mathcal{A} \models G,$$

or equivalently

$$\mathcal{A}(F) = \mathcal{A}(G).$$

F and G are called *equisatisfiable*, if either both F and G are satisfiable, or both F and G are unsatisfiable.

The notions defined above for formulas, such as satisfiability, validity, or entailment, are extended to sets of formulas N by treating sets of formulas analogously to conjunctions of formulas, e. g.:

$\mathcal{A} \models N$ if $\mathcal{A} \models G$ for all $G \in N$.

$N \models F$ if for all Π-valuations $\mathcal{A}$: if $\mathcal{A} \models N$, then $\mathcal{A} \models F$.

Note: Formulas are always finite objects; but sets of formulas may be infinite. Therefore, it is in general not possible to replace a set of formulas by the conjunction of its elements.

Proposition 2.2 *$F \models G$ if and only if $\models (F \rightarrow G)$.*

Proof. ($\Rightarrow$) Suppose that F entails G. Let $\mathcal{A}$ be an arbitrary Π-valuation. We have to show that $\mathcal{A} \models F \rightarrow G$. If $\mathcal{A}(F) = 1$, then $\mathcal{A}(G) = 1$ (since $F \models G$), and hence $\mathcal{A}(F \rightarrow G) = \max(1-1, 1) = 1$. Otherwise $\mathcal{A}(F) = 0$, then $\mathcal{A}(F \rightarrow G) = \max(1-0, \mathcal{A}(G)) = 1$ independently of $\mathcal{A}(G)$. In both cases, $\mathcal{A} \models F \rightarrow G$.

($\Leftarrow$) Suppose that F does not entail G. Then there exists a Π-valuation $\mathcal{A}$ such that $\mathcal{A} \models F$, but not $\mathcal{A} \models G$. Consequently, $\mathcal{A}(F \rightarrow G) = \max(1 - \mathcal{A}(F), \mathcal{A}(G)) = \max(1-1, 0) = 0$, so $(F \rightarrow G)$ does not hold in $\mathcal{A}$. $\square$

Proposition 2.3 *$F \models\!\!\mid G$ if and only if $\models (F \leftrightarrow G)$.*

Proof. Analogously to Prop. 2.2. $\square$

Validity vs. Unsatisfiability

Validity and unsatisfiability of formulas are just two sides of the same medal as explained by the following proposition.

Proposition 2.4 *F is valid if and only if $\neg F$ is unsatisfiable.*

Proof. ($\Rightarrow$) If F is valid, then $\mathcal{A}(F) = 1$ for every valuation $\mathcal{A}$. Hence $\mathcal{A}(\neg F) = 1 - \mathcal{A}(F) = 0$ for every valuation $\mathcal{A}$, so $\neg F$ is unsatisfiable.

($\Leftarrow$) Analogously. $\square$

Hence in order to design a theorem prover (validity checker) it is sufficient to design a checker for unsatisfiability.

In a similar way, entailment can be reduced to unsatisfiability and vice versa:

Proposition 2.5 *$G \models F$ if and only if $G \wedge \neg F$ is unsatisfiable.*
$N \models F$ if and only if $N \cup \{\neg F\}$ is unsatisfiable.

Proposition 2.6 *$G \models \bot$ if and only if G is unsatisfiable.*
$N \models \bot$ if and only if N is unsatisfiable.

Checking Unsatisfiability

Every formula F contains only finitely many propositional variables. Obviously, $\mathcal{A}(F)$ depends only on the values of those finitely many variables in F in $\mathcal{A}$.

If F contains n distinct propositional variables, then it is sufficient to check 2^n valuations to see whether F is satisfiable or not $\Rightarrow$ truth table.

So the satisfiability problem is clearly decidable (but, by Cook's Theorem, NP-complete).

Nevertheless, in practice, there are (much) better methods than truth tables to check the satisfiability of a formula.

Replacement Theorem

Proposition 2.7 *Let $\mathcal{A}$ be a valuation, let F and G be formulas, and let $H = H[F]_p$ be a formula in which F occurs as a subformula at position p.*

If $\mathcal{A}(F) = \mathcal{A}(G)$, then $\mathcal{A}(H[F]_p) = \mathcal{A}(H[G]_p)$.

Proof. The proof proceeds by induction over the length of p.

If $p = \varepsilon$, then $H[F]_\varepsilon = F$ and $H[G]_\varepsilon = G$, so $\mathcal{A}(H[F]_p) = \mathcal{A}(F) = \mathcal{A}(G) = \mathcal{A}(H[G]_p)$ by assumption.

If $p = 1q$ or $p = 2q$, then $H = \neg H_1$ or $H = H_1 \circ H_2$ for $\circ \in \{\wedge, \vee, \rightarrow, \leftrightarrow\}$. Assume that $p = 1q$ and that $H = H_1 \wedge H_2$, hence $H[F]_p = H[F]_{1q} = H_1[F]_q \wedge H_2$. By the induction hypothesis, $\mathcal{A}(H_1[F]_q) = \mathcal{A}(H_1[G]_q)$. Hence $\mathcal{A}(H[F]_{1q}) = \mathcal{A}(H_1[F]_q \wedge H_2) = \min(\mathcal{A}(H_1[F]_q), \mathcal{A}(H_2)) = \min(\mathcal{A}(H_1[G]_q), \mathcal{A}(H_2)) = \mathcal{A}(H_1[G]_q \wedge H_2) = \mathcal{A}(H[G]_{1q})$.

The case $p = 2q$ and the other boolean connectives are handled analogously. $\square$

Theorem 2.8 *Let F and G be equivalent formulas, let $H = H[F]_p$ be a formula in which F occurs as a subformula at position p.*

Then $H[F]_p$ is equivalent to $H[G]_p$.

Proof. We have to show that $\mathcal{A}(H[F]_p) = \mathcal{A}(H[G]_p)$ for every Π-valuation $\mathcal{A}$.

Choose $\mathcal{A}$ arbitrarily. Since F and G are equivalent, we know that $\mathcal{A}(F) = \mathcal{A}(G)$. Hence, by the previous proposition, $\mathcal{A}(H[F]_p) = \mathcal{A}(H[G]_p)$. $\square$

Some Important Equivalences

Proposition 2.9 *The following equivalences hold for all formulas F, G, H:*

$$
\begin{array}{rcll}
(F \wedge F) & \models\!\!\!\!\!= & F & \\
(F \vee F) & \models\!\!\!\!\!= & F & \textit{(Idempotency)} \\[1ex]
(F \wedge G) & \models\!\!\!\!\!= & (G \wedge F) & \\
(F \vee G) & \models\!\!\!\!\!= & (G \vee F) & \textit{(Commutativity)} \\[1ex]
(F \wedge (G \wedge H)) & \models\!\!\!\!\!= & ((F \wedge G) \wedge H) & \\
(F \vee (G \vee H)) & \models\!\!\!\!\!= & ((F \vee G) \vee H) & \textit{(Associativity)} \\[1ex]
(F \wedge (G \vee H)) & \models\!\!\!\!\!= & ((F \wedge G) \vee (F \wedge H)) & \\
(F \vee (G \wedge H)) & \models\!\!\!\!\!= & ((F \vee G) \wedge (F \vee H)) & \textit{(Distributivity)} \\[1ex]
(F \wedge (F \vee G)) & \models\!\!\!\!\!= & F & \\
(F \vee (F \wedge G)) & \models\!\!\!\!\!= & F & \textit{(Absorption)} \\[1ex]
(\neg\neg F) & \models\!\!\!\!\!= & F & \textit{(Double Negation)} \\[1ex]
\neg(F \wedge G) & \models\!\!\!\!\!= & (\neg F \vee \neg G) & \\
\neg(F \vee G) & \models\!\!\!\!\!= & (\neg F \wedge \neg G) & \textit{(De Morgan's Laws)} \\[1ex]
(F \wedge G) & \models\!\!\!\!\!= & F, \textit{ if } G \textit{ is a tautology} & \\
(F \vee G) & \models\!\!\!\!\!= & \top, \textit{ if } G \textit{ is a tautology} & \\
(F \wedge G) & \models\!\!\!\!\!= & \bot, \textit{ if } G \textit{ is unsatisfiable} & \\
(F \vee G) & \models\!\!\!\!\!= & F, \textit{ if } G \textit{ is unsatisfiable} & \textit{(Tautology Laws)} \\[1ex]
(F \leftrightarrow G) & \models\!\!\!\!\!= & ((F \rightarrow G) \wedge (G \rightarrow F)) & \\
(F \leftrightarrow G) & \models\!\!\!\!\!= & ((F \wedge G) \vee (\neg F \wedge \neg G)) & \textit{(Equivalence)} \\
(F \rightarrow G) & \models\!\!\!\!\!= & (\neg F \vee G) & \textit{(Implication)}
\end{array}
$$

An Important Entailment

Proposition 2.10 *The following entailment holds for all formulas F, G, H:*

$$(F \vee H) \wedge (G \vee \neg H) \models F \vee G \qquad \textit{(Generalized Resolution)}$$

2.4 Normal Forms

Many theorem proving calculi do not operate on arbitrary formulas, but only on some restricted class of formulas.

We define *conjunctions* of formulas as follows:

$\bigwedge_{i=1}^{0} F_i = \top.$

$\bigwedge_{i=1}^{1} F_i = F_1.$

$\bigwedge_{i=1}^{n+1} F_i = \bigwedge_{i=1}^{n} F_i \wedge F_{n+1}.$

and analogously *disjunctions:*

$\bigvee_{i=1}^{0} F_i = \bot.$

$\bigvee_{i=1}^{1} F_i = F_1.$

$\bigvee_{i=1}^{n+1} F_i = \bigvee_{i=1}^{n} F_i \vee F_{n+1}.$

Literals and Clauses

A *literal* is either a propositional variable P or a negated propositional variable $\neg P$.

A *clause* is a (possibly empty) disjunction of literals.

CNF and DNF

A formula is in *conjunctive normal form (CNF, clause normal form),* if it is a conjunction of disjunctions of literals (or in other words, a conjunction of clauses).

A formula is in *disjunctive normal form (DNF),* if it is a disjunction of conjunctions of literals.

Warning: definitions in the literature differ:

are complementary literals permitted?
are duplicated literals permitted?
are empty disjunctions/conjunctions permitted?

Checking the validity of CNF formulas or the unsatisfiability of DNF formulas is easy:

A formula in CNF is valid, if and only if each of its disjunctions contains a pair of complementary literals P and $\neg P$.

Conversely, a formula in DNF is unsatisfiable, if and only if each of its conjunctions contains a pair of complementary literals P and $\neg P$.

On the other hand, checking the unsatisfiability of CNF formulas or the validity of DNF formulas is known to be coNP-complete.

Conversion to CNF/DNF

Proposition 2.11 *For every formula there is an equivalent formula in CNF (and also an equivalent formula in DNF).*

Proof. We describe a (naive) algorithm to convert a formula to CNF.

Apply the following rules as long as possible (modulo commutativity of $\wedge$ and $\vee$):

Step 1: Eliminate equivalences:

$$H[F \leftrightarrow G]_p \;\Rightarrow_{\mathrm{CNF}}\; H[(F \rightarrow G) \wedge (G \rightarrow F)]_p$$

Step 2: Eliminate implications:

$$H[F \rightarrow G]_p \;\Rightarrow_{\mathrm{CNF}}\; H[\neg F \vee G]_p$$

Step 3: Push negations downward:

$$\begin{aligned} H[\neg(F \vee G)]_p &\;\Rightarrow_{\mathrm{CNF}}\; H[\neg F \wedge \neg G]_p \\ H[\neg(F \wedge G)]_p &\;\Rightarrow_{\mathrm{CNF}}\; H[\neg F \vee \neg G]_p \end{aligned}$$

Step 4: Eliminate multiple negations:

$$H[\neg\neg F]_p \;\Rightarrow_{\mathrm{CNF}}\; H[F]_p$$

Step 5: Push disjunctions downward:

$$H[(F \wedge F') \vee G]_p \;\Rightarrow_{\mathrm{CNF}}\; H[(F \vee G) \wedge (F' \vee G)]_p$$

Step 6: Eliminate $\top$ and $\bot$:

$$\begin{aligned} H[F \wedge \top]_p &\;\Rightarrow_{\mathrm{CNF}}\; H[F]_p \\ H[F \wedge \bot]_p &\;\Rightarrow_{\mathrm{CNF}}\; H[\bot]_p \\ H[F \vee \top]_p &\;\Rightarrow_{\mathrm{CNF}}\; H[\top]_p \\ H[F \vee \bot]_p &\;\Rightarrow_{\mathrm{CNF}}\; H[F]_p \\ H[\neg\bot]_p &\;\Rightarrow_{\mathrm{CNF}}\; H[\top]_p \\ H[\neg\top]_p &\;\Rightarrow_{\mathrm{CNF}}\; H[\bot]_p \end{aligned}$$

Proving termination is easy for steps 2, 4, and 6; steps 1, 3, and 5 are a bit more complicated.

For step 1, we can prove termination in the following way: We define a function μ_1 from formulas to positive integers such that $\mu_1(\bot) = \mu_1(\top) = \mu_1(P) = 1$, $\mu_1(\neg F) = \mu_1(F)$, $\mu_1(F \wedge G) = \mu_1(F \vee G) = \mu_1(F \rightarrow G) = \mu_1(F) + \mu_1(G)$, and $\mu_1(F \leftrightarrow G) = 2\mu_1(F) + 2\mu_1(G) + 1$. Observe that μ_1 is constructed in such a way that $\mu_1(F) > \mu_1(G)$ implies $\mu_1(H[F]) > \mu_1(H[G])$ for all formulas F, G, and H. Furthermore, μ_1 has the property that swapping the arguments of some $\wedge$ or $\vee$ in a formula F does not change the value of $\mu_1(F)$. (This is important since the transformation rules can be applied modulo commutativity of $\wedge$ and $\vee$.). Using these properties, we can show that whenever a formula H' is the result of applying the rule of step 1 to a formula H, then $\mu_1(H) > \mu_1(H')$. Since μ_1 takes only positive integer values, step 1 must terminate.

Termination of steps 3 and 5 is proved similarly. For step 3, we use function μ_2 from formulas to positive integers such that $\mu_2(\bot) = \mu_2(\top) = \mu_2(P) = 1$, $\mu_2(\neg F) = 2\mu_2(F)$, $\mu_2(F \wedge G) = \mu_2(F \vee G) = \mu_2(F \rightarrow G) = \mu_2(F \leftrightarrow G) = \mu_2(F) + \mu_2(G) + 1$. Whenever a formula H' is the result of applying a rule of step 3 to a formula H, then $\mu_2(H) > \mu_2(H')$. Since μ_2 takes only positive integer values, step 3 must terminate.

For step 5, we use a function μ_3 from formulas to positive integers such that $\mu_3(\bot) = \mu_3(\top) = \mu_3(P) = 1$, $\mu_3(\neg F) = \mu_3(F) + 1$, $\mu_3(F \wedge G) = \mu_3(F \rightarrow G) = \mu_3(F \leftrightarrow G) = \mu_3(F) + \mu_3(G) + 1$, and $\mu_3(F \vee G) = 2\mu_3(F)\mu_3(G)$. Again, if a formula H' is the result of applying a rule of step 5 to a formula H, then $\mu_3(H) > \mu_3(H')$. Since μ_3 takes only positive integer values, step 5 terminates, too.

The resulting formula is equivalent to the original one and in CNF.

The conversion of a formula to DNF works in the same way, except that conjunctions have to be pushed downward in step 5. □

Negation Normal Form (NNF)

The formula after application of Step 4 is said to be in *Negation Normal Form*, i.e., it contains neither $\rightarrow$ nor $\leftrightarrow$ and negation symbols only occur in front of propositional variables (atoms).

Complexity

Conversion to CNF (or DNF) may produce a formula whose size is *exponential* in the size of the original one.

2.5 Improving the CNF Transformation

The goal

"Given a formula F, find an *equivalent* formula G in CNF"

is unpractical.

But if we relax the requirement to

"Given a formula F, find an *equisatisfiable* formula G in CNF"

we can get an efficient transformation.

Literature:

Andreas Nonnengart and Christoph Weidenbach: Computing small clause normal forms, in *Handbook of Automated Reasoning*, pages 335-367. Elsevier, 2001.

Christoph Weidenbach: Automated Reasoning (Chapter 2). Textbook draft, available for registered participants in the lecture Nextcloud (same link as for the online session recordings), 2021.

Tseitin Transformation

Proposition 2.12 *A formula $H[F]_p$ is satisfiable if and only if $H[Q]_p \wedge (Q \leftrightarrow F)$ is satisfiable, where Q is a new propositional variable that works as an abbreviation for F.*

Proof. "$\Rightarrow$": Suppose that the Π-formula $H[F]_p$ is satisfiable. Let $\mathcal{A}$ be a Π-valuation such that $\mathcal{A}(H[F]_p) = 1$. Let Q be a new propositional variable (that is, a variable that is not contained in Π). Let $\Pi' = \Pi \cup \{Q\}$ and let $\mathcal{A}'$ be the Π'-valuation defined by $\mathcal{A}'(P) = \mathcal{A}(P)$ for all $P \in \Pi$ and $\mathcal{A}'(Q) = \mathcal{A}(F)$. Since $H[F]_p$ is a Π-formula, we have $\mathcal{A}'(H[F]_p) = \mathcal{A}(H[F]_p) = 1$ and $\mathcal{A}'(F) = \mathcal{A}(F)$. Therefore $\mathcal{A}'(Q) = \mathcal{A}'(F)$ and by Prop. 2.7 $\mathcal{A}'(H[Q]_p) = \mathcal{A}'(H[F]_p) = 1$, thus $\mathcal{A}'(H[Q]_p \wedge (Q \leftrightarrow F)) = 1$.

"$\Leftarrow$": Let $\Pi' = \Pi \cup \{Q\}$. Suppose that the Π'-formula $H[Q]_p \wedge (Q \leftrightarrow F)$ is satisfiable. Let $\mathcal{A}'$ be a Π'-valuation such that $\mathcal{A}'(H[Q]_p \wedge (Q \leftrightarrow F)) = 1$. Then $\mathcal{A}'(H[Q]_p) = 1$ and $\mathcal{A}'(Q) = \mathcal{A}'(F)$, so by Prop. 2.7 $\mathcal{A}'(H[F]_p) = \mathcal{A}'(H[Q]_p) = 1$. $\square$

Satisfiability-preserving CNF transformation (Tseitin 1970):

Apply Prop. 2.12 recursively bottom-up to all subformulas F in the original formula (except $\bot$, $\top$, and literals). This introduces a linear number of new propositional variables Q and definitions $Q \leftrightarrow F$.

Convert the resulting conjunction to CNF. This increases the size only by an additional factor, since each formula $Q \leftrightarrow F$ yields at most four clauses in the CNF.

Polarity-based CNF Transformation

A further improvement is possible by taking the *polarity* of the subformula F into account (Plaisted and Greenbaum 1986):

Intuitively, if G occurs in F at the position p, then the polarity of G determines the number of "negations" starting from F down to G. It is 1 for an even number, -1 for an odd number and 0 if there is at least one equivalence connective along the path.

The *polarity* of a subformula $G = F|_p$ at position p is $\mathrm{pol}(F, p)$, where pol is recursively defined by

$$\begin{aligned}
\mathrm{pol}(F, \varepsilon) &:= 1 \\
\mathrm{pol}(\neg F, 1p) &:= -\mathrm{pol}(F, p) \\
\mathrm{pol}(F_1 \circ F_2, ip) &:= \mathrm{pol}(F_i, p) \text{ if } \circ \in \{\wedge, \vee\} \\
\mathrm{pol}(F_1 \rightarrow F_2, 1p) &:= -\mathrm{pol}(F_1, p) \\
\mathrm{pol}(F_1 \rightarrow F_2, 2p) &:= \mathrm{pol}(F_2, p) \\
\mathrm{pol}(F_1 \leftrightarrow F_2, ip) &:= 0
\end{aligned}$$

Example 2.13 *Let* $F = (P \rightarrow Q) \rightarrow (P \wedge \neg R)$. *Then* $\mathrm{pol}(F, 1) = \mathrm{pol}(F, 12) = \mathrm{pol}(F, 221) = -1$ *and* $\mathrm{pol}(F, \varepsilon) = \mathrm{pol}(F, 11) = \mathrm{pol}(F, 2) = \mathrm{pol}(F, 21) = \mathrm{pol}(F, 22) = 1$.

Let $F' = (P \wedge Q) \leftrightarrow (P \vee Q)$. *Then* $\mathrm{pol}(F', \varepsilon) = 1$ *and* $\mathrm{pol}(F', p) = 0$ *for all* $p \in \mathrm{pos}(F')$ *different from* ε.

Proposition 2.14 *Let* $\mathcal{A}$ *be a valuation, let* F *and* G *be formulas, and let* $H = H[F]_p$ *be a formula in which* F *occurs as a subformula at position* p.

If $\mathrm{pol}(H, p) = 1$ *and* $\mathcal{A}(F) \leq \mathcal{A}(G)$, *then* $\mathcal{A}(H[F]_p) \leq \mathcal{A}(H[G]_p)$.

If $\mathrm{pol}(H, p) = -1$ *and* $\mathcal{A}(F) \geq \mathcal{A}(G)$, *then* $\mathcal{A}(H[F]_p) \leq \mathcal{A}(H[G]_p)$.

Proof. Exercise. □

Let Q be a propositional variable not occurring in $H[F]_p$.

Define the formula $\mathrm{def}(H, p, Q, F)$ by

- $(Q \rightarrow F)$, if $\mathrm{pol}(H, p) = 1$,
- $(F \rightarrow Q)$, if $\mathrm{pol}(H, p) = -1$,
- $(Q \leftrightarrow F)$, if $\mathrm{pol}(H, p) = 0$.

Proposition 2.15 *Let Q be a propositional variable not occurring in $H[F]_p$. Then $H[F]_p$ is satisfiable if and only if $H[Q]_p \wedge \mathrm{def}(H, p, Q, F)$ is satisfiable.*

Proof. ($\Rightarrow$) Since $H[F]_p$ is satisfiable, there exists a Π-valuation $\mathcal{A}$ such that $\mathcal{A} \models H[F]_p$. Let $\Pi' = \Pi \cup \{Q\}$ and define the Π'-valuation $\mathcal{A}'$ by $\mathcal{A}'(P) = \mathcal{A}(P)$ for $P \in \Pi$ and $\mathcal{A}'(Q) = \mathcal{A}(F)$. Obviously $\mathcal{A}'(\mathrm{def}(H, p, Q, F)) = 1$; moreover $\mathcal{A}'(H[Q]_p) = \mathcal{A}'(H[F]_p) = \mathcal{A}(H[F]_p) = 1$ by Prop. 2.7, so $H[Q]_p \wedge \mathrm{def}(H, p, Q, F)$ is satisfiable.

($\Leftarrow$) Let $\mathcal{A}$ be a valuation such that $\mathcal{A} \models H[Q]_p \wedge \mathrm{def}(H, p, Q, F)$. So $\mathcal{A}(H[Q]_p) = 1$ and $\mathcal{A}(\mathrm{def}(H, p, Q, F)) = 1$. We will show that $\mathcal{A} \models H[F]_p$.

If $\mathrm{pol}(H, p) = 0$, then $\mathrm{def}(H, p, Q, F) = (Q \leftrightarrow F)$, so $\mathcal{A}(Q) = \mathcal{A}(F)$, hence $\mathcal{A}(H[F]_p) = \mathcal{A}(H[Q]_p) = 1$ by Prop. 2.7.

If $\mathrm{pol}(H, p) = 1$, then $\mathrm{def}(H, p, Q, F) = (Q \rightarrow F)$, so $\mathcal{A}(Q) \leq \mathcal{A}(F)$. By Prop. 2.14, $\mathcal{A}(H[F]_p) \geq \mathcal{A}(H[Q]_p) = 1$, so $\mathcal{A}(H[F]_p) = 1$.

If $\mathrm{pol}(H, p) = -1$, then $\mathrm{def}(H, p, Q, F) = (F \rightarrow Q)$, so $\mathcal{A}(F) \leq \mathcal{A}(Q)$. By Prop. 2.14, $\mathcal{A}(H[F]_p) \geq \mathcal{A}(H[Q]_p) = 1$, so $\mathcal{A}(H[F]_p) = 1$. □

Optimized CNF

Not every introduction of a definition for a subformula leads to a smaller CNF.

The number of potentially generated clauses is a good indicator for useful CNF transformations.

The functions $\nu(F)$ and $\bar{\nu}(F)$ give us upper bounds for the number of clauses in $\mathrm{cnf}(F)$ and $\mathrm{cnf}(\neg F)$ using a naive CNF transformation.

G	$\nu(G)$	$\bar{\nu}(G)$
$P, \top, \bot$	1	1
$F_1 \wedge F_2$	$\nu(F_1) + \nu(F_2)$	$\bar{\nu}(F_1)\bar{\nu}(F_2)$
$F_1 \vee F_2$	$\nu(F_1)\nu(F_2)$	$\bar{\nu}(F_1) + \bar{\nu}(F_2)$
$\neg F_1$	$\bar{\nu}(F_1)$	$\nu(F_1)$
$F_1 \rightarrow F_2$	$\bar{\nu}(F_1)\nu(F_2)$	$\nu(F_1) + \bar{\nu}(F_2)$
$F_1 \leftrightarrow F_2$	$\nu(F_1)\bar{\nu}(F_2) + \bar{\nu}(F_1)\nu(F_2)$	$\nu(F_1)\nu(F_2)+\bar{\nu}(F_1)\bar{\nu}(F_2)$

A better CNF transformation (Nonnengart and Weidenbach 2001):

Step 1: Exhaustively apply modulo commutativity of $\leftrightarrow$ and associativity/commutativity of $\land$, $\lor$:

$$\begin{aligned}
H[(F \land \top)]_p &\Rightarrow_{\mathrm{OCNF}} H[F]_p \\
H[(F \lor \bot)]_p &\Rightarrow_{\mathrm{OCNF}} H[F]_p \\
H[(F \leftrightarrow \bot)]_p &\Rightarrow_{\mathrm{OCNF}} H[\neg F]_p \\
H[(F \leftrightarrow \top)]_p &\Rightarrow_{\mathrm{OCNF}} H[F]_p \\
H[(F \lor \top)]_p &\Rightarrow_{\mathrm{OCNF}} H[\top]_p \\
H[(F \land \bot)]_p &\Rightarrow_{\mathrm{OCNF}} H[\bot]_p
\end{aligned}$$

$$\begin{aligned}
H[(F \land F)]_p &\Rightarrow_{\mathrm{OCNF}} H[F]_p \\
H[(F \lor F)]_p &\Rightarrow_{\mathrm{OCNF}} H[F]_p \\
H[(F \land (F \lor G))]_p &\Rightarrow_{\mathrm{OCNF}} H[F]_p \\
H[(F \lor (F \land G))]_p &\Rightarrow_{\mathrm{OCNF}} H[F]_p \\
H[(F \land \neg F)]_p &\Rightarrow_{\mathrm{OCNF}} H[\bot]_p \\
H[(F \lor \neg F)]_p &\Rightarrow_{\mathrm{OCNF}} H[\top]_p \\
H[\neg \top]_p &\Rightarrow_{\mathrm{OCNF}} H[\bot]_p \\
H[\neg \bot]_p &\Rightarrow_{\mathrm{OCNF}} H[\top]_p
\end{aligned}$$

$$\begin{aligned}
H[(F \rightarrow \bot)]_p &\Rightarrow_{\mathrm{OCNF}} H[\neg F]_p \\
H[(F \rightarrow \top)]_p &\Rightarrow_{\mathrm{OCNF}} H[\top]_p \\
H[(\bot \rightarrow F)]_p &\Rightarrow_{\mathrm{OCNF}} H[\top]_p \\
H[(\top \rightarrow F)]_p &\Rightarrow_{\mathrm{OCNF}} H[F]_p
\end{aligned}$$

Note: Applying the absorption laws exhaustively modulo associativity/commutativity of $\land$ and $\lor$ is expensive. In practice, it is sufficient to apply them only in those cases that are easy to detect.

Step 2: Introduce top-down fresh variables for beneficial subformulas:

$$H[F]_p \ \Rightarrow_{\mathrm{OCNF}} \ H[Q]_p \land \mathrm{def}(H, p, Q, F)$$

where Q is new to $H[F]_p$ and $\nu(H[F]_p) > \nu(H[Q]_p \land \mathrm{def}(H, p, Q, F))$.

Remark: Although computing ν is not practical in general, the test $\nu(H[F]_p) > \nu(H[Q]_p \land \mathrm{def}(H, p, Q, F))$ can be computed in constant time.

Step 3: Eliminate equivalences dependent on their polarity:

$$H[F \leftrightarrow G]_p \;\Rightarrow_{\text{OCNF}}\; H[(F \rightarrow G) \wedge (G \rightarrow F)]_p$$

if $\text{pol}(F, p) = 1$ or $\text{pol}(F, p) = 0$.

$$H[F \leftrightarrow G]_p \;\Rightarrow_{\text{OCNF}}\; H[(F \wedge G) \vee (\neg F \wedge \neg G)]_p$$

if $\text{pol}(F, p) = -1$.

Step 4: Apply steps 2, 3, 4, 5 of $\Rightarrow_{\text{CNF}}$

Remark: The $\Rightarrow_{\text{OCNF}}$ algorithm is already close to a state of the art algorithm, but some additional redundancy tests and simplification mechanisms are missing.