Automated Reasoning*

Uwe Waldmann

Winter Term 2025/2026

Topics of the Course

Preliminaries

abstract reduction systems
well-founded orderings

Propositional logic

syntax, semantics
calculi: CDCL-procedure
implementation: Two watched literals

First-order predicate logic

syntax, semantics, model theory, ...
calculi: resolution, tableaux
implementation: sharing, indexing

First-order predicate logic with equality

term rewriting systems
calculi: Knuth-Bendix completion, dependency pairs

Emphasis on:

logics and their properties,

proof systems for these logics and their properties:
soundness, completeness, implementation

*This document contains the text of the lecture slides (almost verbatim) plus some additional information, in particular proofs of theorems that are presented on the blackboard during the course. It is not a full script and does not contain the examples and additional explanations given during the lecture. Moreover it should not be taken as an example how to write a research paper – neither stylistically nor typographically.
Parts of this document are based on lecture notes by Harald Ganzinger, Christoph Weidenbach, and Sophie Tourret.

1 Preliminaries

Literature:
Franz Baader and Tobias Nipkow: *Term rewriting and all that*,
Cambridge Univ. Press, 1998, Chapter 2.

Before we start with the main subjects of the lecture, we repeat some prerequisites from mathematics and computer science and introduce some tools that we will need throughout the lecture.

1.1 Mathematical Prerequisites

$\mathbb{N} = \{0, 1, 2, \ldots\}$ is the set of natural numbers (including 0).

$\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ denote the integers, rational numbers and the real numbers, respectively.

$\emptyset$ is the empty set.

If M and M' are sets, then $M \cap M'$, $M \cup M'$, and $M \setminus M'$ denote the intersection, union, and set difference of M and M'.

The subset relation is denoted by $\subseteq$. The strict subset relation is denoted by $\subset$ (i.e., $M \subset M'$ if and only if $M \subseteq M'$ and $M \neq M'$).

Relations

Let M be a set, let $n \geq 2$. We write M^n for the n-fold cartesian product $M \times \cdots \times M$.

In order to handle the cases $n \geq 2$, $n = 1$, and $n = 0$ simultaneously, we also define $M^1 = M$ and $M^0 = \{()\}$. (We do not distinguish between an element m of M and a 1-tuple (m) of an element of M.)

An n-ary *relation* R over some set M is a subset of M^n: $R \subseteq M^n$.

We often use predicate notation for relations:

Instead of $(m_1, \ldots, m_n) \in R$ we write $R(m_1, \ldots, m_n)$, and say that $R(m_1, \ldots, m_n)$ holds or is true.

For binary relations, we often use infix notation, so
$(m, m') \in <$ $\Leftrightarrow$ $<(m, m')$ $\Leftrightarrow$ $m < m'$.

Since relations are sets, we can use the usual set operations for then.

Example: Let $R = \{(0, 2), (1, 2), (2, 2), (3, 2)\} \subseteq \mathbb{N} \times \mathbb{N}$.
Then $R \cap < \; = \; R \cap \{\, (n, m) \in \mathbb{N} \times \mathbb{N} \mid n < m \,\} \; = \; \{(0, 2), (1, 2)\}$.

A relation Q is a *subrelation* of a relation R if $Q \subseteq R$.

Words

Given a non-empty set (also called *alphabet*) Σ, the set Σ^* of *finite words* over Σ is defined inductively by

(i) the empty word ε is in Σ^*,

(ii) if $u \in \Sigma^*$ and $a \in \Sigma$ then ua is in Σ^*.

The set of *non-empty finite words* Σ^+ is $\Sigma^* \setminus \{\varepsilon\}$.

The *concatenation* of two words $u, v \in \Sigma^*$ is denoted by uv.

The length $|u|$ of a word $u \in \Sigma^*$ is defined by

(i) $|\varepsilon| := 0$,

(ii) $|ua| := |u| + 1$ for any $u \in \Sigma^*$ and $a \in \Sigma$.

1.2 Abstract Reduction Systems

Throughout the lecture, we will have to work with reduction systems,

on the object level, in particular in the section on equality,

and on the meta level, i.e., to describe deduction calculi.

An *abstract reduction system* is a pair $(A, \to)$, where

A is a non-empty set,

$\to \subseteq A \times A$ is a binary relation on A.

The relation $\to$ is usually written in infix notation, i.e., $a \to b$ instead of $(a, b) \in \to$.

Let $\to' \subseteq A \times A$ and $\to'' \subseteq A \times A$ be two binary relations. Then the *composition of* $\to'$ *and* $\to''$ is the binary relation $(\to' \circ \to'') \subseteq A \times A$ defined by

$a \; (\to' \circ \to'') \; c$ if and only if there exists some $b \in A$ such that $a \to' b$ and $b \to'' c$.

For a binary relation $\to \subseteq A \times A$, we define:

$\to^0$	$= \{ (a,a) \mid a \in A \}$	*identity*
$\to^{i+1}$	$= \to^i \circ \to$	$i+1$*-fold composition*
$\to^+$	$= \bigcup_{i>0} \to^i$	*transitive closure*
$\to^*$	$= \bigcup_{i\geq 0} \to^i = \to^+ \cup \to^0$	*reflexive transitive closure*
$\to^=$	$= \to \cup \to^0$	*reflexive closure*
$\leftarrow$	$= \to^{-1} = \{ (b,c) \mid c \to b \}$	*inverse*
$\leftrightarrow$	$= \to \cup \leftarrow$	*symmetric closure*
$\leftrightarrow^+$	$= (\leftrightarrow)^+$	*transitive symmetric closure*
$\leftrightarrow^*$	$= (\leftrightarrow)^*$	*reflexive transitive symmetric closure* or *equivalence closure*

$b \in A$ is *reducible*, if there is a c such that $b \to c$.

b is *in normal form* (or *irreducible*), if it is not reducible.

c is a *normal form of* b, if $b \to^* c$ and c is in normal form.
Notation: $c = b{\downarrow}$ (if the normal form of b is unique).

A relation $\to$ is called

terminating, if there is no infinite descending chain $b_0 \to b_1 \to b_2 \to \ldots$.

normalizing, if every $b \in A$ has a normal form.

Lemma 1.1 *If $\to$ is terminating, then it is normalizing.*

Proof. Suppose that $\to$ is terminating. Let $b \in A$ be an arbitrary element. Since $\to$ is terminating, there must exist some $n \geq 0$ such that $b \to^n c$ and c is in normal form. Therefore c is a normal form of b. As we can do that for every element $b \in A$, the relation $\to$ is normalizing. □

Note: The reverse implication does not hold. Counterexample: Let $A = \{b, c\}$, let $\to = \{(b,b), (b,c)\}$. Then both b and c have the normal form c, but $\to$ is clearly not terminating.

1.3 Orderings

Important properties of binary relations:

Let $M \neq \emptyset$. A binary relation $R \subseteq M \times M$ is called

reflexive, if $R(x,x)$ for all $x \in M$,

irreflexive, if $\neg R(x,x)$ for all $x \in M$,

antisymmetric, if $R(x,y)$ and $R(y,x)$ imply $x = y$ for all $x, y \in M$,

transitive, if $R(x,y)$ and $R(y,z)$ imply $R(x,z)$ for all $x, y, z \in M$,

total, if $R(x,y)$ or $R(y,x)$ or $x = y$ for all $x, y \in M$.

A *strict partial ordering* $\succ$ on a set $M \neq \emptyset$ is a transitive and irreflexive binary relation on M.

Notation:
$\prec$ for the inverse relation $\succ^{-1}$
$\succeq$ for the reflexive closure $(\succ \cup =)$ of $\succ$

Let $\succ$ be a strict partial ordering on M; let $M' \subseteq M$.

$a \in M'$ is called *minimal in* M', if there is no $b \in M'$ with $a \succ b$.

$a \in M'$ is called *smallest in* M', if $b \succ a$ for all $b \in M' \setminus \{a\}$.

Analogously:

$a \in M'$ is called *maximal in* M', if there is no $b \in M'$ with $a \prec b$.

$a \in M'$ is called *largest in* M', if $b \prec a$ for all $b \in M' \setminus \{a\}$.

Notation:
$M^{\prec x} = \{\, y \in M \mid y \prec x \,\}$,
$M^{\preceq x} = \{\, y \in M \mid y \preceq x \,\}$.

A subset $M' \subseteq M$ is called *downward closed*, if $x \in M'$ and $x \succ y$ implies $y \in M'$.

Well-Foundedness

Termination of reduction systems is strongly related to the concept of well-founded orderings.

A strict partial ordering $\succ$ on M is called *well-founded (or Noetherian)*, if there is no infinite descending chain $a_0 \succ a_1 \succ a_2 \succ \ldots$ with $a_i \in M$.

Well-Foundedness and Termination

Lemma 1.2 *If $\succ$ is a well-founded partial ordering and $\to \subseteq \succ$, then $\to$ is terminating.*

Proof. Suppose that $\to \subseteq \succ$ for some partial ordering $\succ$ and that $\to$ is not terminating. Then there exists an infinite descending chain $b_0 \to b_1 \to b_2 \to \ldots$. Since $\to \subseteq \succ$, we have an infinite descending chain $b_0 \succ b_1 \succ b_2 \succ \ldots$, hence $\succ$ is not well-founded. $\square$

Lemma 1.3 *If $\to$ is a terminating binary relation over A, then $\to^+$ is a well-founded partial ordering.*

Proof. Transitivity of $\to^+$ is obvious; irreflexivity and well-foundedness follow from termination of $\to$. □

Well-Founded Orderings: Examples

Natural numbers: $(\mathbb{N}, >)$

Lexicographic orderings: Let $(M_1, \succ_1), (M_2, \succ_2)$ be well-founded orderings. Define their *lexicographic combination*

$$\succ = (\succ_1, \succ_2)_{\text{lex}}$$

on $M_1 \times M_2$ by

$$(a_1, a_2) \succ (b_1, b_2) \quad :\Leftrightarrow \quad a_1 \succ_1 b_1 \text{ or } (a_1 = b_1 \text{ and } a_2 \succ_2 b_2)$$

(analogously for more than two orderings). This again yields a well-founded ordering (proof below).

Length-based ordering on words: For alphabets Σ with a well-founded ordering $>_\Sigma$, the relation $\succ$ defined as

$$w \succ w' \quad :\Leftrightarrow \quad |w| > |w'| \text{ or } (|w| = |w'| \text{ and } w >_{\Sigma,\text{lex}} w')$$

is a well-founded ordering on the set Σ^* of finite words over the alphabet Σ (Exercise).

Counterexamples:
$(\mathbb{Z}, >)$
$(\mathbb{N}, <)$
the lexicographic ordering on Σ^*

Basic Properties of Well-Founded Orderings

Lemma 1.4 *$(M, \succ)$ is well-founded if and only if every non-empty $M' \subseteq M$ has a minimal element.*

Proof. "$\Leftarrow$": Suppose that $(M, \succ)$ is not well-founded. Then there is an infinite descending chain $a_0 \succ a_1 \succ a_2 \succ \ldots$ with $a_i \in M$. Consequently, the subset $M' = \{\, a_i \mid i \in \mathbb{N} \,\}$, does not have a minimal element.

"$\Rightarrow$": Suppose that the non-empty subset $M' \subseteq M$ does not have a minimal element. Choose $a_0 \in M'$ arbitrarily. Since for every $a_i \in M'$ there is a smaller $a_{i+1} \in M'$ (otherwise a_i would be minimal in M'), there is an infinite descending chain $a_0 \succ a_1 \succ a_2 \succ \ldots$ □

Lemma 1.5 *$(M_1, \succ_1)$ and $(M_2, \succ_2)$ are well-founded if and only if $(M_1 \times M_2, \succ)$ with $\succ = (\succ_1, \succ_2)_{\text{lex}}$ is well-founded.*

Proof. "⇒": Suppose $(M_1 \times M_2, \succ)$ is not well-founded. Then there is an infinite sequence $(a_0, b_0) \succ (a_1, b_1) \succ (a_2, b_2) \succ \ldots$.

Let $A = \{ a_i \mid i \geq 0 \} \subseteq M_1$. Since $(M_1, \succ_1)$ is well-founded, A has a minimal element a_n. But then $B = \{ b_i \mid i \geq n \} \subseteq M_2$ can not have a minimal element, contradicting the well-foundedness of $(M_2, \succ_2)$.

"⇐": obvious. □

Monotone Mappings

Let $(M, \succ)$ and $(M', \succ')$ be strict partial orderings. A mapping $\varphi : M \to M'$ is called *monotone*, if $a \succ b$ implies $\varphi(a) \succ' \varphi(b)$ for all $a, b \in M$.

Lemma 1.6 *If φ is a monotone mapping from $(M, \succ)$ to $(M', \succ')$ and $(M', \succ')$ is well-founded, then $(M, \succ)$ is well-founded.*

Proof. Suppose that $(M, \succ)$ is not well-founded, then there exists an infinite descending chain $a_0 \succ a_1 \succ a_2 \succ \ldots$. Since $a_i \succ a_{i+1}$ implies $\varphi(a_i) \succ' \varphi(a_{i+1})$, we obtain an infinite descending chain $\varphi(a_0) \succ' \varphi(a_1) \succ' \varphi(a_2) \succ' \ldots$, contradicting the well-foundedness of $(M', \succ')$. □

Well-founded Induction

Well-founded induction generalizes the usual induction over natural numbers or data structures.

Theorem 1.7 (Well-founded (or Noetherian) Induction) *Let $(M, \succ)$ be a well-founded ordering, let Q be a property of elements of M.*

If for all $m \in M$ the implication

> *if $Q(m')$ for all $m' \in M$ such that $m \succ m'$,*[1]
> *then $Q(m)$.*[2]

is satisfied, then the property $Q(m)$ holds for all $m \in M$.

[1] induction hypothesis
[2] induction step

Proof. Let $X = \{\, m \in M \mid Q(m) \text{ false} \,\}$. Suppose that $X \neq \emptyset$. Since $(M, \succ)$ is well-founded, X has a minimal element m_0. Hence for all $m' \in M$ with $m' \prec m_0$ the property $Q(m')$ holds. On the other hand, the implication which is presupposed for this theorem holds in particular also for m_0, hence $Q(m_0)$ must be true. Therefore m_0 cannot be in X, contradicting the assumption. □

Well-founded Recursion

Similarly, well-founded recursion generalizes the usual recursion over natural numbers or data structures. We will need this concept only once in this lecture (and once more in Automated Reasoning II), but in one of the main theorems.

Let M and S be sets, let $N \subseteq M$, and let $f : M \to S$ be a function. Then the *restriction* of f to N, denoted by $f|_N$, is a function from N to S with $f|_N(x) = f(x)$ for all $x \in N$.

Theorem 1.8 (Well-founded (or Noetherian) Recursion) *Let $(M, \succ)$ be a well-founded ordering, let S be a set. Let ϕ be a binary function that takes two arguments x and g and maps them to an element of S, where $x \in M$ and g is a function from $M^{\prec x}$ to S.*

Then there exists exactly one function $f : M \to S$ such that for all $x \in M$

$$f(x) = \phi(x, f|_{M^{\prec x}})$$

Proof. The proof consists of four parts.

Part 1: For every downward closed subset $N \subseteq M$ there is *at most one* function $f : N \to S$ such that $f(x) = \phi(x, f|_{N^{\prec x}}) = \phi(x, f|_{M^{\prec x}})$.

Proof: First observe that if $N \subseteq M$ is downward closed and $x \in N$, then $N^{\prec x} = M^{\prec x}$. Assume that there exist a downward closed subset $N \subseteq M$ and two *different* functions f_1 and f_2 from N to S with the property. Therefore, the set $N' := \{\, x \in N \mid f_1(x) \neq f_2(x) \,\}$ is non-empty. By well-foundedness, N' has a minimal element y. By minimality of y, $f_1|_{M^{\prec y}} = f_2|_{M^{\prec y}}$. Therefore $f_1(y) = \phi(y, f_1|_{M^{\prec y}}) = \phi(y, f_2|_{M^{\prec y}}) = f_2(y)$, contradicting the assumption.

Part 2: If N_1 and N_2 are downward closed subsets of M and the functions $f_1 : N_1 \to S$ and $f_2 : N_2 \to S$ satisfy $f_i(x) = \phi(x, f_i|_{M^{\prec x}})$ for all $x \in N_i$ $(i = 1, 2)$, then $f_1(x) = f_2(x)$ for all $x \in N_1 \cap N_2$.

Proof: Define $N_0 := N_1 \cap N_2$ and $f_i' = f_i|_{N_0}$ for $i = 1, 2$. Clearly N_0 is downward closed and for all $x \in N_0$ and $i = 1, 2$ we have $f_i'(x) = f_i(x) = \phi(x, f_i|_{M^{\prec x}}) = \phi(x, f_i'|_{M^{\prec x}})$. By part 1, there is at most one function from N_0 to S with this property, so $f_1' = f_2'$, and therefore $f_1(x) = f_2(x)$ for all $x \in N_1 \cap N_2$.

Part 3: For every $y \in M$ there exists a function $f_y : M^{\preceq y} \to S$ such that $f_y(x) = \phi(x, f_y|_{M^{\prec x}})$ for all $x \in M^{\preceq y}$.

Proof: We use well-founded induction over $\succ$. Let $y \in M$. By the induction hypothesis, for every $z \prec y$ there exists a function $f_z : M^{\preceq z} \to S$ such that $f_z(x) = \phi(x, f_z|_{M^{\prec x}})$ for all $x \in M^{\preceq z}$. By part 2, all functions f_z agree on the intersections of their domains. Define the function $f_y : M^{\preceq y} \to S$ by $f_y(x) = f_x(x)$ for $x \prec y$ and by $f_y(y) = \phi(y, f_y|_{M^{\prec y}})$. The function f_y has the desired property for $x = y$ by construction and for all $x \prec y$ by the induction hypothesis (since $f_y(x) = f_x(x)$ for $x \prec y$ and f_x has the desired property).

Part 4: There exists a function $f : M \to S$ such that $f(x) = \phi(x, f|_{M^{\prec x}})$ for all $x \in M$.

Proof: Define $f : M \to S$ by $f(x) = f_x(x)$.

The claim of the theorem follows now from part 1 (for $N := M$) and part 4. □

The well-founded recursion scheme generalizes terminating recursive programs.

Note that functions defined by well-founded recursion are mathematically well-defined; nevertheless they need *not* be computable, in particular since for many well-founded orderings the sets $M^{\prec x}$ may be infinite.

1.4 Multisets

Let M be a set. A *multiset* S over M is a mapping $S : M \to \mathbb{N}$. We interpret $S(m)$ as the number of occurrences of elements m of the base set M within the multiset S.

Example. $S = \{a, a, a, b, b\}$ is a multiset over $\{a, b, c\}$, where $S(a) = 3$, $S(b) = 2$, $S(c) = 0$.

We say that m is an *element* of S, if $S(m) > 0$.

We use set notation ($\in$, $\subseteq$, $\cup$, $\cap$, etc.) with analogous meaning also for multisets, e.g.,

$$
\begin{aligned}
m \in S &:\Leftrightarrow S(m) > 0 \\
(S_1 \cup S_2)(m) &:= S_1(m) + S_2(m) \\
(S_1 \cap S_2)(m) &:= \min\{S_1(m), S_2(m)\} \\
(S_1 - S_2)(m) &:= \begin{cases} S_1(m) - S_2(m) & \text{if } S_1(m) \geq S_2(m) \\ 0 & \text{otherwise} \end{cases} \\
S_1 \subseteq S_2 &:\Leftrightarrow S_1(m) \leq S_2(m) \text{ for all } m \in M
\end{aligned}
$$

A multiset S is called *finite*, if the set $\{\, m \in M \mid S(m) > 0 \,\}$ is finite.

From now on we only consider finite multisets.

Multiset Orderings

Let $(M, \succ)$ be an abstract reduction system. The *multiset extension* of $\succ$ to multisets over M is defined by

$S_1 \succ_{\text{mul}} S_2$ if and only if
there exist multisets X and Y over M such that
$\emptyset \neq X \subseteq S_1$,
$S_2 = (S_1 - X) \cup Y$,
$\forall y \in Y\ \exists x \in X\colon x \succ y$

Lemma 1.9 (König's Lemma) *Every finitely branching tree with infinitely many nodes contains an infinite path.*

Theorem 1.10
(a) If $\succ$ is transitive, then $\succ_{\text{mul}}$ is transitive.

(b) If $\succ$ is irreflexive and transitive, then $\succ_{\text{mul}}$ is irreflexive.

(c) If $\succ$ is a well-founded ordering, then $\succ_{\text{mul}}$ is a well-founded ordering.

(d) If $\succ$ is a strict total ordering, then $\succ_{\text{mul}}$ is a strict total ordering.

Proof. see Baader and Nipkow, page 22–24. □

The multiset extension as defined above is due to Dershowitz and Manna (1979).

There are several other ways to characterize the multiset extension of a binary relation. The following one is due to Huet and Oppen (1980):

Let $(M, \succ)$ be an abstract reduction system. The *(Huet/Oppen) multiset extension* of $\succ$ to multisets over M is defined by

$S_1 \succ_{\text{mul}}^{\text{HO}} S_2$ if and only if
$S_1 \neq S_2$ and
$\forall m \in M\colon \big(S_2(m) > S_1(m)$
$\Rightarrow\ \exists m' \in M\colon m' \succ m$ and $S_1(m') > S_2(m')\big)$

A third way to characterize the multiset extension of a binary relation $\succ$ is to define it as the transitive closure of the relation $\succ_{\text{mul}}^1$ given by

$S_1 \succ_{\text{mul}}^1 S_2$ if and only if
there exists $x \in S_1$ and a multiset Y over M such that
$S_2 = (S_1 - \{x\}) \cup Y$,
$\forall y \in Y\colon x \succ y$

For strict partial orderings all three characterizations of $\succ_{\text{mul}}$ are equivalent:

Theorem 1.11 *If $\succ$ is a strict partial ordering, then*
(a) $\succ_{\text{mul}} = \succ^{\text{HO}}_{\text{mul}}$,
(b) $\succ_{\text{mul}} = (\succ^{1}_{\text{mul}})^{+}$.

Proof. (a) see Baader and Nipkow, page 24–26. (b) Exercise. □

Note, however, that for an arbitrary binary relation $\succ$ all three relations $\succ_{\text{mul}}$, $\succ^{\text{HO}}_{\text{mul}}$, and $(\succ^{1}_{\text{mul}})^{+}$ may be different.

1.5 Complexity Theory Prerequisites

A *decision problem* is a subset $L \subseteq \Sigma^*$ for some fixed finite alphabet Σ.

The function $\text{chr}(L, x)$ denotes the *characteristic function* for some decision problem L and is defined by $\text{chr}(L, u) = 1$ if $u \in L$ and $\text{chr}(L, u) = 0$ otherwise.

P and NP

A decision problem is called *solvable in polynomial time* if its characteristic function can be computed in polynomial time. The class of all polynomial-time decision problems is denoted by P.

We say that a decision problem L is in NP if there is a predicate $Q(x, y)$ and a polynomial $p(n)$ such that for all $u \in \Sigma^*$ we have

(i) $u \in L$ if and only if there is a $v \in \Sigma^*$ with $|v| \leq p(|u|)$ and $Q(u, v)$ holds, and

(ii) the predicate Q is in P.

Intuitively, a decision problem is in P, if we can *solve* it in polynomial time, and it is in NP, if we can *verify* a solution (namely the string v in the definition of NP) in polynomial time.

Reducibility, NP-Hardness, NP-Completeness

A decision problem L is *polynomial-time reducible* to a decision problem L' if there is a function g computable in polynomial time such that for all $u \in \Sigma^*$ we have $u \in L$ iff $g(u) \in L'$.

For example, if L is polynomial-time reducible to L' and $L' \in \mathrm{P}$ then $L \in \mathrm{P}$.

A decision problem is *NP-hard* if every problem in NP is polynomial-time reducible to it.

A decision problem is *NP-complete* if it is NP-hard and in NP.

The following properties are equivalent:

(i) There exists some NP-complete problem that is in P.

(ii) P = NP.

The question whether P equals NP or not is probably the most famous unsolved problem in theoretical computer science.

All known algorithms for NP-complete problems have an exponential time complexity in the worst case.

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:

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

Sometimes further connectives are used, for instance

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

Notational Conventions

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

Instead of $((P \wedge Q) \wedge R)$ we simply write $P \wedge Q \wedge R$ (analogously for $\vee$).

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

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

The prefix order $\leq$ on positions is defined by $p \leq q$ 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 $\leq$, that is, $p < q$ if $p \leq q$ but not $q \leq p$.

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

We say that p is *above* q if $p \leq q$, 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

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

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

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

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

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

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 $\mathcal{A} : \Pi \to \{0, 1\}$ where $\{0, 1\}$ is the set of *truth values*.

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

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

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

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 $F \models\!\!\models G$, 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 $\models (F \to G)$.*

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

($\Leftarrow$) 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, $\mathcal{A}(F \to G) = \max(1 - \mathcal{A}(F), \mathcal{A}(G)) = \max(1 - 1, 0) = 0$, so $(F \to G)$ does not hold in $\mathcal{A}$. □

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

Proof. Analogously to Prop. 2.2. □

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. □

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. □

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

Some Important Equivalences

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

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

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:

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

Step 2: Eliminate implications:

$$H[F \to G]_p \;\Rightarrow_{\text{CNF}}\; H[\neg F \vee G]_p$$

Step 3: Push negations downward:

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

Step 4: Eliminate multiple negations:

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

Step 5: Push disjunctions downward:

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

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

$$H[F \wedge \top]_p \;\Rightarrow_{\text{CNF}}\; H[F]_p$$
$$H[F \wedge \bot]_p \;\Rightarrow_{\text{CNF}}\; H[\bot]_p$$
$$H[F \vee \top]_p \;\Rightarrow_{\text{CNF}}\; H[\top]_p$$
$$H[F \vee \bot]_p \;\Rightarrow_{\text{CNF}}\; H[F]_p$$
$$H[\neg\bot]_p \;\Rightarrow_{\text{CNF}}\; H[\top]_p$$
$$H[\neg\top]_p \;\Rightarrow_{\text{CNF}}\; H[\bot]_p$$

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 $\mu_1(\bot) = \mu_1(\top) = \mu_1(P) = 1$, $\mu_1(\neg F) = \mu_1(F)$, $\mu_1(F \wedge G) = \mu_1(F \vee G) = \mu_1(F \to G) = \mu_1(F) + \mu_1(G)$, 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 $\mu_2(\bot) = \mu_2(\top) = \mu_2(P) = 1$, $\mu_2(\neg F) = 2\mu_2(F)$, $\mu_2(F \wedge G) = \mu_2(F \vee G) = \mu_2(F \to G) = \mu_2(F \leftrightarrow G) = \mu_2(F) + \mu_2(G) + 1$. 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 $\mu_3(\bot) = \mu_3(\top) = \mu_3(P) = 1$, $\mu_3(\neg F) = \mu_3(F) + 1$, $\mu_3(F \wedge G) = \mu_3(F \to G) = \mu_3(F \leftrightarrow G) = \mu_3(F) + \mu_3(G) + 1$, 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 $\to$ 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.

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 $\mathcal{A}'(H[F]_p) = \mathcal{A}'(H[Q]_p) = 1$. □

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.

Optimization

Typically, introducing new propositional variables for only some of the subformulas yields a smaller number of clauses.

For instance, if we introduce a new variable for every binary subformula in

$$(\cdots((P_1 \wedge P_2) \wedge P_3) \cdots \wedge P_n),$$

this yields in the end $3n - 3$ clauses in the CNF.

But if we introduce only one new variable for the entire conjunction, this yields only $n + 1$ clauses in the CNF.

Further reductions of the number of clauses are possible, for instance by taking the number of negations above a subformula into account.

However, this does not necessarily improve the performance of the satisfiability checking calculus that is applied afterwards:

CDCL (this section): slightly worse performance.

Ordered resolution (next section): much better performance.

Therefore we will discuss this in Section 3.

2.6 The DPLL Procedure

Goal:
Given a propositional formula in CNF (or alternatively, a finite set N of clauses), check whether it is satisfiable (and optionally: output *one* solution, if it is satisfiable).

Preliminaries

Recall:

$\mathcal{A} \models N$ if and only if $\mathcal{A} \models C$ for all clauses C in N.

$\mathcal{A} \models C$ if and only if $\mathcal{A} \models L$ for some literal $L \in C$.

Assumptions:

Clauses contain neither duplicated literals nor complementary literals.

The order of literals in a clause is irrelevant.

$\Rightarrow$ Clauses behave like *sets* of literals.

Notation:

We use the notation $C \vee L$ to denote a clause with some literal L and a clause rest C. Here L need *not* be the last literal of the clause and C may be empty.

$\overline{L}$ is the complementary literal of L, i.e., $\overline{P} = \neg P$ and $\overline{\neg P} = P$.

Partial Valuations

Since we will construct satisfying valuations incrementally, we consider *partial valuations* (that is, partial mappings $\mathcal{A} : \Pi \to \{0, 1\}$).

Every partial valuation $\mathcal{A}$ corresponds to a set M of literals that does not contain complementary literals, and vice versa:

$\mathcal{A}(L)$ is true, if $L \in M$.

$\mathcal{A}(L)$ is false, if $\overline{L} \in M$.

$\mathcal{A}(L)$ is undefined, if neither $L \in M$ nor $\overline{L} \in M$.

We will use $\mathcal{A}$ and M interchangeably.

A clause is true in a partial valuation $\mathcal{A}$ (or in a set M of literals) if one of its literals is true; it is false (or *"conflicting"*) if all its literals are false; otherwise it is undefined (or *"unresolved"*).

Unit Clauses

Observation:
Let $\mathcal{A}$ be a partial valuation. If the set N contains a clause C, such that all literals but one in C are false in $\mathcal{A}$, then the following properties are equivalent:

- there is a valuation that is a model of N and extends $\mathcal{A}$.
- there is a valuation that is a model of N and extends $\mathcal{A}$ and makes the remaining literal L of C true.

C is called a *unit clause;* L is called a *unit literal.*

Pure Literals

One more observation:
Let $\mathcal{A}$ be a partial valuation and P a variable that is undefined in $\mathcal{A}$. If P occurs only positively (or only negatively) in the unresolved clauses in N, then the following properties are equivalent:

- there is a valuation that is a model of N and extends $\mathcal{A}$.
- there is a valuation that is a model of N and extends $\mathcal{A}$ and assigns 1 (0) to P.

P is called a *pure literal.*

The Davis-Putnam-Logemann-Loveland Procedure

```
boolean DPLL(literal set M, clause set N) {
    if (all clauses in N are true in M) return true;
    elsif (some clause in N is false in M) return false;
    elsif (N contains unit literal P) return DPLL(M ∪ {P}, N);
    elsif (N contains unit literal ¬P) return DPLL(M ∪ {¬P}, N);
    elsif (N contains pure literal P) return DPLL(M ∪ {P}, N);
    elsif (N contains pure literal ¬P) return DPLL(M ∪ {¬P}, N);
    else {
        let P be some undefined variable in N;
        if (DPLL(M ∪ {¬P}, N)) return true;
        else return DPLL(M ∪ {P}, N);
    }
}
```

Initially, DPLL is called with an empty literal set and the clause set N.

2.7 From DPLL to CDCL

The DPLL procedure can be improved significantly:

The pure literal check is only done while preprocessing (otherwise is too expensive).

If a conflict is detected, information is reused by conflict analysis and learning.

The algorithm is implemented iteratively ⇒ the backtrack stack is managed explicitly (it may be possible and useful to backtrack more than one level).

The branching variable is not chosen randomly.

Under certain circumstances, the procedure is restarted.

Literature:

Lintao Zhang and Sharad Malik: The Quest for Efficient Boolean Satisfiability Solvers, Proc. CADE-18, LNAI 2392, pp. 295–312, Springer, 2002.

Robert Nieuwenhuis, Albert Oliveras, Cesare Tinelli: Solving SAT and SAT Modulo Theories – From an abstract Davis-Putnam-Logemann-Loveland procedure to DPLL(T), pp. 937–977, Journal of the ACM, 53(6), 2006.

Armin Biere, Marijn Heule, Hans van Maaren, Toby Walsh (eds.): Handbook of Satisfiability, IOS Press, 2009

Daniel Le Berre's slides at VTSA'09: `http://www.mpi-inf.mpg.de/vtsa09/`.

Conflict Analysis and Learning

Conflict analysis serves a dual purpose:

Backjumping (non-chronological backtracking): If we detect that a conflict is independent of some earlier branch, we can skip over that backtrack level.

Learning: By deriving a new clause from a conflict that is added to the current set of clauses, we can reuse information that is obtained in one branch in further branches. (Note: This may produce a large number of new clauses; therefore it may become necessary to delete some of them afterwards to save space.)

These ideas are implemented in all modern SAT solvers.

Because of the importance of clause learning the algorithm is now called CDCL: Conflict Driven Clause Learning.

Formalizing CDCL

We model the improved DPLL procedure by a transition relation $\Rightarrow_{\text{CDCL}}$ on a set of states.

States:

- *fail*
- $M \parallel N$,

where M is a *list of annotated literals ("trail")* and N is a set of clauses.

Annotated literals:

- L^C: deduced literal, due to unit propagation using clause C.
- L^{d}: decision literal (guessed literal).

Alternative notation: $L^{C_i} \quad \equiv \quad \begin{matrix} L \\ C_i \end{matrix} \quad \equiv \quad \begin{matrix} L \\ i \end{matrix}$

Unit Propagate:

$M \parallel N \cup \{C \vee L\} \;\Rightarrow_{\text{CDCL}}\; M\; L^{C \vee L} \parallel N \cup \{C \vee L\}$

if C is false in M and L is undefined in M.

Decide:

$M \parallel N \;\Rightarrow_{\text{CDCL}}\; M\; L^{\text{d}} \parallel N$

if L is undefined in M and contained in N.

Fail:

$M \parallel N \cup \{C\} \;\Rightarrow_{\text{CDCL}}\; \mathit{fail}$

if C is false in M and M contains no decision literals.

Backjump:

$M'\; K^{\text{d}}\; M'' \parallel N \;\Rightarrow_{\text{CDCL}}\; M'\; L^{C \vee L} \parallel N$

if some clause $D \in N$ is false in $M'\; K^{\text{d}}\; M''$
and if there is some "backjump clause" $C \vee L$ such that
$N \models C \vee L$,
C is false in M', and
L is undefined in M'.

We will see later that the Backjump rule is always applicable, if the list of literals M contains at least one decision literal and some clause in N is false in M.

There are many possible backjump clauses. One candidate is $\overline{L_1} \vee \ldots \vee \overline{L_n}$, where the L_i are all the decision literals in $M'\ L^{\mathrm{d}}\ M''$. With this backjump clause, CDCL simulates DPLL. (But usually there are better choices.)

Reasonable strategy:

- Use "Fail", if applicable.
- Otherwise use "Backjump", if applicable.
  Choose M' as short as possible. ($\Rightarrow$ Go back to the earliest position where guessing can be replaced by knowledge.)
- Otherwise use "Unit Propagate", if applicable.
- Otherwise use "Decide".

A trail M defines an ordering $\succ_M$ on literals: $L \succ_M K$ if L or $\overline{L}$ occurs in M after (i.e., right of) K or $\overline{K}$ (with any annotation d or C).

Lemma 2.13 *If $\varepsilon \parallel N \ \Rightarrow^*_{\mathrm{CDCL}}\ M \parallel N$, then:*

(1) M contains neither duplicated nor complementary literals (regardless of their annotations).

(2) If L^C is a deduced literal in M, then C has the form $C' \vee L$.

(3) If $L^{C' \vee L}$ is a deduced literal in M, then $L \succ_M K$ for every literal K of C'.

(4) If $L^{C' \vee L}$ is a deduced literal in M, then C' is false in M.

(5) Every literal L in M follows from N and decision literals in M that are smaller than or equal to L w.r.t. $\succ_M$.

Proof. By induction on the length of the derivation. □

Lemma 2.14 *Every derivation starting from $\varepsilon \parallel N$ terminates.*

Proof. Let $n \in \mathbb{N}$ be the number of propositional variables in N. We define a function ϕ that maps every CDCL state $M \parallel N$ in the derivation to a list of annotated literals S such that S is obtained from M by deleting all the C-superscripts of deduced literals in M. (The d-superscripts of decision literals in M are kept.) Since each of the n propositional variables can occur at most once in M, each of them can occur at most once in S (positively or negatively, with or without a d-superscript), therefore the number of possible lists that can be returned by ϕ is bounded by $n! \cdot 5^n$.

Let $M \parallel N$ and $M' \parallel N'$ be two CDCL states, such that

$$\phi(M \parallel N) = S = S_0 L_1^{\mathrm{d}} S_1 \dots L_k^{\mathrm{d}} S_k$$

and

$$\phi(M' \parallel N') = S' = S'_0 {L'_1}^{\mathrm{d}} S'_1 \dots {L'_{k'}}^{\mathrm{d}} S'_{k'},$$

where the lists S_i and S'_i contain no decision literals. We define a relation $\succ$ on lists of annotated literals by $S \succ S'$ if and only if

(i) there is some j such that $0 \leq j \leq \min(k, k')$, $|S_i| = |S'_i|$ for all $0 \leq i < j$, and $|S_j| < |S'_j|$, or

(ii) $|S_i| = |S'_i|$ for all $0 \leq i \leq k < k'$ and $|S| < |S'|$.

It is routine to check that $\succ$ is irreflexive and transitive, hence a strict partial ordering, and that for every CDCL step $M \parallel N \;\Rightarrow_{\mathrm{CDCL}}\; M' \parallel N'$ we have $\phi(M \parallel N) \succ \phi(M' \parallel N')$. Now assume that there is an infinite CDCL derivation

$$\varepsilon \parallel N_0 \;\Rightarrow_{\mathrm{CDCL}}\; M_1 \parallel N_1 \;\Rightarrow_{\mathrm{CDCL}}\; M_2 \parallel N_2 \;\Rightarrow_{\mathrm{CDCL}}\; \dots$$

Since ϕ can return only finitely many lists, there must exist indices $i < j$ such that

$$\phi(M_i \parallel N_i) \;\succ^+\; \phi(M_j \parallel N_j) = \phi(M_i \parallel N_i).$$

By transitivity, this implies $\phi(M_i \parallel N_i) \succ \phi(M_i \parallel N_i)$, but that contradicts the irreflexivity of $\succ$.

Lemma 2.15 *Suppose that we reach a state $M \parallel N$ starting from $\varepsilon \parallel N$ such that some clause $D \in N$ is false in M. Then:*

(1) If M does not contain any decision literal, then "Fail" is applicable.

(2) Otherwise, "Backjump" is applicable.

Proof. (1) Obvious.

(2) Let $L_1, \dots, L_n$ be the decision literals occurring in M (in this order). By part (5) of Lemma 2.13, every literal in M follows from $N \cup \{L_1, \dots, L_n\}$, and since $M \models \neg D$, we obtain $N \cup \{L_1, \dots, L_n\} \models \neg D$. By Prop. 2.6, this is equivalent to $N \cup \{L_1, \dots, L_n\} \cup D \models \bot$, and since $D \in N$, it is equivalent to $N \cup \{L_1, \dots, L_n\} \models \bot$. Consequently, $N \models \overline{L_1} \vee \dots \vee \overline{L_n}$. Now let $C = \overline{L_1} \vee \dots \vee \overline{L_{n-1}}$, $L' = \overline{L_n}$, $L = L_n$, and let M' be the list of all literals of M occurring before L_n, then the condition of "Backjump" is satisfied. □

Theorem 2.16 *Suppose that we reach a final state starting from $\varepsilon \parallel N$.*

(1) If the final state is $M \parallel N$, then N is satisfiable and M is a model of N.

(2) If the final state is fail, then N is unsatisfiable.

Proof. (1) Observe that the "Decide" rule is applicable as long as literals in N are undefined in M. Hence, in a final state, all literals must be defined. Furthermore, in a final state, no clause in N can be false in M, otherwise "Fail" or "Backjump" would be applicable. Hence M is a model of every clause in N.

(2) If we reach *fail*, then in the previous step we must have reached a state $M \parallel N$ such that some $C \in N$ is false in M (i.e., $M \models \neg C$), and M contains no decision literals. By part (5) of Lemma 2.13, every literal in M follows from N, therefore $N \models \neg C$, and therefore $N \cup \{C\} \models \bot$. On the other hand, $C \in N$, so $N \models \bot$. □

Getting Better Backjump Clauses

Lemma 2.17 *Suppose that $\varepsilon \parallel N \Rightarrow^*_{\text{CDCL}} M \parallel N$. Let D is a clause such that $N \models D$ and D is false in M. Let $\overline{L}$ be the largest literal of D w.r.t. $\succ_M$ and let $D = D' \vee \overline{L}$. Suppose that $L^{C \vee L}$ is a deduced literal in M. Let $D_0 = C \vee D'$. Then $N \models D_0$; D_0 is false in M; and all literals of D_0 are smaller than $\overline{L}$.*

Proof. If D is false in M, then every literal of D is the complement of a literal in M. Let $\overline{L}$ be the largest literal of D w.r.t. $\succ_M$ and let $D = D' \vee \overline{L}$. Clearly, D' is false in M. If L is a deduced literal, then M contains $L^{C \vee L}$ and all literals of C are smaller than L and $\overline{L}$. Moreover, C is false in M, and by construction of M, we have $N \models C \vee L$. Therefore, $D_0 = C \vee D'$ is false in M; all literals of D_0 are smaller than $\overline{L}$; and $N \models D_0$ by (generalized) resolution. □

The clause $C \vee D'$ is called a *resolvent* of $C \vee L$ and $D' \vee \overline{L}$; this is denoted by

$$\frac{C \vee L \qquad D' \vee \overline{L}}{C \vee D'}$$

Note that the resolvent $C \vee D'$ is again entailed by N and false in M. If its largest literal is the complement of a deduced literal, we can therefore repeat the process with $C \vee D'$.

Lemma 2.18 *Suppose that* $\varepsilon \parallel N \Rightarrow^*_{\text{CDCL}} M \parallel N$. *Let* D *is a clause such that* $N \models D$ *and* D *is false in* M. *If* D *has the form* $D = D' \vee \overline{L}$, *where* L *is larger than or equal to the largest decision literal of* M *and all literals in* D' *are smaller than the largest decision literal of* M, *then* M *can be written as* $M' \, K^{\text{d}} \, M''$ *such that* D' *is false in* M' *and* L *is undefined in* M' *(that is,* D *is a backjump clause).*

Proof. By assumption, there exists a decision literal of M that is larger than every literal of D'. Choose K as the smallest decision literal with this property. □

Suppose that $\varepsilon \parallel N \Rightarrow^*_{\text{CDCL}} M \parallel N$, where L^d is the largest decision literal in M and some $D \in N$ is false in M.

If we have used a reasonable strategy, then D must contain two literals whose complements are larger than or equal to L.

By repeated resolution steps as described in Lemma 2.17, we must eventually reach a clause that satisfies Lemma 2.18.

This clause is a backjump clause
⇒ *1UIP (first unique implication point) strategy.*

Learning Clauses

Backjump clauses are good candidates for learning.

To model learning, the CDCL system is extended by the following two rules:

Learn:

$M \parallel N \Rightarrow_{\text{CDCL}} M \parallel N \cup \{C\}$

if $N \models C$.

Forget:

$M \parallel N \cup \{C\} \Rightarrow_{\text{CDCL}} M \parallel N$

if $N \models C$.

If we ensure that no clause is learned infinitely often, then termination is guaranteed.

The other properties of the basic CDCL system hold also for the extended system.

Restart

Runtimes of CDCL-style procedures depend extremely on the choice of branching variables.

If no solution is found within a certain time limit, it can be useful to *restart* from scratch with an adapted variable selection heuristics. Learned clauses, however, are kept.

In addition, it is useful to restart after a unit clause has been learned.

The restart rule is typically applied after a certain number of clauses have been learned or a unit is derived:

Restart:

$$M \parallel N \;\Rightarrow_{\text{CDCL}}\; \varepsilon \parallel N$$

If Restart is only applied finitely often, termination is guaranteed.

2.8 Implementing CDCL

The formalization of CDCL that we have seen so far leaves many aspects unspecified.

To get a fast solver, we must use good heuristics, for instance to choose the next undefined variable, and we must implement basic operations efficiently.

Variable Order Heuristic

Choosing the right undefined variable to branch is important for efficiency, but the branching heuristics may be expensive itself.

State of the art: Use branching heuristics that need not be recomputed too frequently.

In general: Choose variables that occur frequently; after a restart prefer variables from recent conflicts.

The VSIDS (Variable State Independent Decaying Sum) heuristic:

- We associate a positive *score* to every propositional variable P_i. At the start, k_i is the number of occurrences of P_i in N.
- The variable order is then the descending ordering of the P_i according to the k_i.

The scores k_i are adjusted during a CDCL run.

- Every time a learned clause is computed after a conflict, the propositional variables in the learned clause obtain a bonus b, i.e., $k_i := k_i + b$.
- Periodically, the scores are leveled: $k_i := k_i/l$ for some l.
- After each restart, the variable order is recomputed, using the new scores.

The purpose of these mechanisms is to keep the search focused. The parameter b directs the search around the conflict,

Further refinements:

- Add the bonus to all literals in the clauses that occur in the resolution steps to generate a backjump clause.
- If the score of a variable reaches a certain limit, all scores are rescaled by a constant.
- Occasionally (with low probability) choose a variable at random, otherwise choose the undefined variable with the highest score.

Implementing Unit Propagation Efficiently

For applying the unit rule, we need to know the number of literals in a clause that are not false.

Maintaining this number is expensive, however.

Better approach: *"Two watched literals":*

In each clause, select two (currently undefined) "watched" literals.

For each variable P, keep a list of all clauses in which P is watched and a list of all clauses in which $\neg P$ is watched.

If an undefined variable P is set to 0 (or to 1), check all clauses in which P (or $\neg P$) is watched and watch another literal (that is true or undefined) in this clause if possible.

Watched literal information need not be restored upon backtracking.

2.9 Preprocessing and Inprocessing

Some operations are only needed once at the beginning of the CDCL run.

(i) Deletion of tautologies

(ii) Deletion of duplicated literals

Some operations are useful, but expensive. They are performed only initially and after restarts (before computation of the variable order heuristics), possibly with time limits.

Note: Some of these operations are only satisfiability-preserving; they do not yield equivalent clause sets.

Literature:

Matti Järvisalo, Marijn J. H. Heule, and Armin Biere: Inprocessing Rules, Proc. IJCAR 2012, LNAI 7364, pp. 355–370, Springer, 2012

Examples:

(i) *Subsumption*

$$N \cup \{C\} \cup \{D\} \;\Rightarrow\; N \cup \{C\}$$

if $C \subseteq D$ considering C, D as multisets of literals.

(ii) *Purity deletion*

Delete all clauses containing a literal L where $\overline{L}$ does not occur in the clause set.

(iii) *Merging replacement resolution*

$$N \cup \{C \vee L\} \cup \{D \vee \overline{L}\} \;\Rightarrow\; N \cup \{C \vee L\} \cup \{D\}$$

if $C \subseteq D$ considering C, D as multisets of literals.

(iv) *Bounded variable elimination*

Compute all possible resolution steps

$$\frac{C \vee L \qquad D \vee \overline{L}}{C \vee D}$$

on a literal L with premises in N; add all non-tautological conclusions to N; then throw away all clauses containing L or $\overline{L}$; repeat this as long as $|N|$ does not grow.

(v) *RAT ("Resolution asymmetric tautologies")*

C is called an *asymmetric tautology* w.r.t. N, if its negation can be refuted by unit propagation using clauses in N.

C has the *RAT property* w.r.t. N, if it is an asymmetric tautology w.r.t. N, or if there is a literal L in C such that $C = C' \vee L$ and all clauses $D' \vee C'$ for $D' \vee \overline{L} \in N$ are asymmetric tautologies w.r.t. N.

RAT elimination:

$$N \cup \{C\} \;\Rightarrow\; N$$

if C has the RAT property w.r.t. N.

2.10 Other Calculi

Ordered resolution
Tableau calculus
Hilbert calculus
Sequent calculus
Natural deduction

see next chapter

OBDDs (Ordered binary decision diagrams)
FRAIGs (Functionally reduced and-inverter graphs)

graph representations of equivalence classes of propositional formulas

satisfiability checking as a side effect

3 First-Order Logic

First-order logic

- is expressive:
  can be used to formalize mathematical concepts,
  can be used to encode Turing machines,
  but cannot axiomatize natural numbers or uncountable sets,
- has important decidable fragments,
- has interesting logical properties (model and proof theory).

First-order logic is also called (first-order) *predicate logic*.

3.1 Syntax

Syntax:

- non-logical symbols (domain-specific)
  $\Rightarrow$ terms, atomic formulas
- logical connectives (domain-independent)
  $\Rightarrow$ Boolean combinations, quantifiers

Signatures

A signature $\Sigma = (\Omega, \Pi)$ fixes an alphabet of non-logical symbols, where

- Ω is a set of *function symbols* f with *arity* $n \geq 0$, written $\text{arity}(f) = n$,
- Π is a set of *predicate symbols* P with arity $m \geq 0$, written $\text{arity}(P) = m$.

Function symbols are also called *operator symbols*.
If $n = 0$ then f is also called a *constant (symbol)*.
If $m = 0$ then P is also called a *propositional variable*.

We will usually use

b, c, d for constant symbols,

f, g, h for non-constant function symbols,

P, Q, R, S for predicate symbols.

Convention: We will usually write $f/n \in \Omega$ instead of $f \in \Omega$, $\text{arity}(f) = n$ (analogously for predicate symbols).

Refined concept for practical applications:
many-sorted signatures (corresponds to simple type systems in programming languages); no big change from a logical point of view.

Variables

Predicate logic admits the formulation of abstract, schematic assertions. (Object) variables are the technical tool for schematization.

We assume that X is a given countably infinite set of symbols which we use to denote *variables*.

Terms

Terms over Σ and X (Σ-terms) are formed according to these syntactic rules:

$$\begin{array}{lcll} s,t,u,v & ::= & x & , x \in X \quad \text{(variable)} \\ & | & f(s_1,...,s_n) & , f/n \in \Omega \quad \text{(functional term)} \end{array}$$

By $\mathrm{T}_\Sigma(X)$ we denote the set of Σ-terms (over X). A term not containing any variable is called a *ground term*. By T_Σ we denote the set of Σ-ground terms.

Atoms

Atoms (also called atomic formulas) over Σ are formed according to this syntax:

$$\begin{array}{lcll} A,B & ::= & P(s_1,\ldots,s_m) & , P/m \in \Pi \quad \text{(non-equational atom)} \\ & \left[| \right. & (s \approx t) & \quad \text{(equation)} \left. \right] \end{array}$$

Whenever we admit equations as atomic formulas we are in the realm of *first-order logic with equality*. Admitting equality does not really increase the expressiveness of first-order logic (see next chapter). But deductive systems where equality is treated specifically are much more efficient.

Literals

$$\begin{array}{lcll} L & ::= & A & \text{(positive literal)} \\ & | & \neg A & \text{(negative literal)} \end{array}$$

Clauses

$$\begin{array}{lcll} C,D & ::= & \bot & \text{(empty clause)} \\ & | & L_1 \vee \ldots \vee L_k, \ k \geq 1 & \text{(non-empty clause)} \end{array}$$

General First-Order Formulas

$\mathrm{F}_\Sigma(X)$ is the set of *first-order formulas* over Σ defined as follows:

$$
\begin{array}{rcll}
F, G, H & ::= & \bot & \text{(falsum)} \\
& | & \top & \text{(verum)} \\
& | & A & \text{(atomic formula)} \\
& | & \neg F & \text{(negation)} \\
& | & (F \wedge G) & \text{(conjunction)} \\
& | & (F \vee G) & \text{(disjunction)} \\
& | & (F \to G) & \text{(implication)} \\
& | & (F \leftrightarrow G) & \text{(equivalence)} \\
& | & \forall x\, F & \text{(universal quantification)} \\
& | & \exists x\, F & \text{(existential quantification)}
\end{array}
$$

Notational Conventions

We omit parentheses according to the conventions for propositional logic.

$\forall x_1, \ldots, x_n\, F$ and $\exists x_1, \ldots, x_n\, F$ abbreviate $\forall x_1 \ldots \forall x_n\, F$ and $\exists x_1 \ldots \exists x_n\, F$.

We use infix-, prefix-, postfix-, or mixfix-notation with the usual operator precedences.

Examples:

$$
\begin{array}{rcl}
s + t * u & \text{for} & +(s, *(t, u)) \\
s * u \leq t + v & \text{for} & \leq (*(s, u), +(t, v)) \\
-s & \text{for} & -(s) \\
s! & \text{for} & !(s) \\
|s| & \text{for} & |_|(s) \\
0 & \text{for} & 0()
\end{array}
$$

Example: Peano Arithmetic

$$
\begin{array}{rcl}
\Sigma_{\mathrm{PA}} & = & (\Omega_{\mathrm{PA}},\ \Pi_{\mathrm{PA}}) \\
\Omega_{\mathrm{PA}} & = & \{0/0,\ +/2,\ */2,\ s/1\} \\
\Pi_{\mathrm{PA}} & = & \{</2\}
\end{array}
$$

Examples of formulas over this signature are:

$\forall x, y\, ((x < y \vee x \approx y) \leftrightarrow \exists z\, (x + z \approx y))$
$\exists x \forall y\, (x + y \approx y)$
$\forall x, y\, (x * s(y) \approx x * y + x)$
$\forall x, y\, (s(x) \approx s(y) \to x \approx y)$
$\forall x \exists y\, (x < y \wedge \neg \exists z (x < z \wedge z < y))$

Positions in Terms and Formulas

The set of positions is extended from propositional logic to first-order logic:

The *positions* of a term s (formula F):

$$\begin{aligned}
&\mathrm{pos}(x) = \{\varepsilon\},\\
&\mathrm{pos}(f(s_1, \ldots, s_n)) = \{\varepsilon\} \cup \textstyle\bigcup_{i=1}^{n} \{\, ip \mid p \in \mathrm{pos}(s_i) \,\},\\
&\mathrm{pos}(P(t_1, \ldots, t_n)) = \{\varepsilon\} \cup \textstyle\bigcup_{i=1}^{n} \{\, ip \mid p \in \mathrm{pos}(t_i) \,\},\\
&\mathrm{pos}(\forall x\, F) = \{\varepsilon\} \cup \{\, 1p \mid p \in \mathrm{pos}(F) \,\},\\
&\mathrm{pos}(\exists x\, F) = \{\varepsilon\} \cup \{\, 1p \mid p \in \mathrm{pos}(F) \,\}.
\end{aligned}$$

The prefix order $\leq$, the subformula (subterm) operator, the formula (term) replacement operator and the size operator are extended accordingly. See the definitions in Sect. 2.

Variables

The *set of variables* occurring in a term t is denoted by $\mathrm{var}(t)$ (and analogously for atoms, literals, clauses, and formulas).

Bound and Free Variables

In $\mathsf{Q}x\, F$, $\mathsf{Q} \in \{\exists, \forall\}$, we call F the *scope* of the quantifier $\mathsf{Q}x$. An *occurrence* of a variable x is called *bound*, if it is inside the scope of a quantifier $\mathsf{Q}x$. Any other occurrence of a variable is called *free*.

Formulas without free variables are called *closed formulas* (or *sentential forms*).

Formulas without variables are called *ground*.

Example:

$$\overbrace{\forall y \quad (\overbrace{(\forall x \; P(x)}^{\text{scope of } \forall x} \;) \quad \rightarrow \quad R(x, y))}^{\text{scope of } \forall y}$$

The occurrence of y is bound, as is the first occurrence of x. The second occurrence of x is a free occurrence.

Substitutions

Substitution is a fundamental operation on terms and formulas that occurs in all inference systems for first-order logic.

Substitutions are mappings

$$\sigma : X \to \mathrm{T}_\Sigma(X)$$

such that the *domain* of σ, that is, the set

$$\mathrm{dom}(\sigma) = \{\, x \in X \mid \sigma(x) \neq x \,\},$$

is finite. The set of variables *introduced* by σ, that is, the set of variables occurring in one of the terms $\sigma(x)$, with $x \in \mathrm{dom}(\sigma)$, is denoted by $\mathrm{codom}(\sigma)$.

Substitutions are often written as $\{x_1 \mapsto s_1, \ldots, x_n \mapsto s_n\}$, with x_i pairwise distinct, and then denote the mapping

$$\{x_1 \mapsto s_1, \ldots, x_n \mapsto s_n\}(y) = \begin{cases} s_i, & \text{if } y = x_i \\ y, & \text{otherwise} \end{cases}$$

We also write $x\sigma$ for $\sigma(x)$.

The *modification* of a substitution σ at x is defined as follows:

$$\sigma[x \mapsto t](y) = \begin{cases} t, & \text{if } y = x \\ \sigma(y), & \text{otherwise} \end{cases}$$

Why Substitution is Complicated

We define the application of a substitution σ to a term t or formula F by structural induction over the syntactic structure of t or F by the equations below.

In the presence of quantification it is surprisingly complex: We must not only ensure that bound variables are not replaced by σ. We must also make sure that the (free) variables in the codomain of σ are not *captured* upon placing them into the scope of a quantifier $\mathsf{Q}y$. Hence the bound variable must be renamed into a "fresh", that is, previously unused, variable z.

Application of a Substitution

"Homomorphic" extension of σ to terms and formulas:

$$\begin{aligned}
f(s_1, \ldots, s_n)\sigma &= f(s_1\sigma, \ldots, s_n\sigma) \\
\bot\sigma &= \bot \\
\top\sigma &= \top \\
P(s_1, \ldots, s_n)\sigma &= P(s_1\sigma, \ldots, s_n\sigma) \\
(u \approx v)\sigma &= (u\sigma \approx v\sigma) \\
\neg F\sigma &= \neg(F\sigma) \\
(F \circ G)\sigma &= (F\sigma \circ G\sigma) \quad \text{for each binary connective } \circ \\
(\mathsf{Q}x\, F)\sigma &= \mathsf{Q}z\, (F\, \sigma[x \mapsto z]) \quad \text{with } z \text{ a fresh variable}
\end{aligned}$$

If $s = t\sigma$ for some subsitution σ, we call the term s an *instance* of the term t, and we call t a *generalization* of s (analogously for formulas).

3.2 Semantics

To give semantics to a logical system means to define a notion of truth for the formulas. The concept of truth that we will now define for first-order logic goes back to Tarski.

As in the propositional case, we use a two-valued logic with truth values "true" and "false" denoted by 1 and 0, respectively.

Algebras

A Σ-*algebra* (also called Σ-interpretation or Σ-structure) is a triple

$$\mathcal{A} = (U_\mathcal{A},\ (f_\mathcal{A} : U_\mathcal{A}^n \to U_\mathcal{A})_{f/n\in\Omega},\ (P_\mathcal{A} \subseteq U_\mathcal{A}^m)_{P/m\in\Pi})$$

where $U_\mathcal{A} \neq \emptyset$ is a set, called the *universe* of $\mathcal{A}$.

By Σ-Alg we denote the class of all Σ-algebras.

Σ-algebras generalize the valuations from propositional logic.

Assignments

A variable has no intrinsic meaning. The meaning of a variable has to be defined externally (explicitly or implicitly in a given context) by an assignment.

A *(variable) assignment* (over a given Σ-algebra $\mathcal{A}$), is a function $\beta : X \to U_{\mathcal{A}}$.

Variable assignments are the semantic counterparts of substitutions.

Value of a Term in $\mathcal{A}$ with respect to β

By structural induction we define

$$\mathcal{A}(\beta) : \mathrm{T}_\Sigma(X) \to U_{\mathcal{A}}$$

as follows:

$$\begin{aligned} \mathcal{A}(\beta)(x) &= \beta(x), && x \in X \\ \mathcal{A}(\beta)(f(s_1, \ldots, s_n)) &= f_{\mathcal{A}}(\mathcal{A}(\beta)(s_1), \ldots, \mathcal{A}(\beta)(s_n)), && f/n \in \Omega \end{aligned}$$

In the scope of a quantifier we need to evaluate terms with respect to modified assignments. To that end, let $\beta[x \mapsto a] : X \to U_{\mathcal{A}}$, for $x \in X$ and $a \in U_{\mathcal{A}}$, denote the assignment

$$\beta[x \mapsto a](y) = \begin{cases} a & \text{if } x = y \\ \beta(y) & \text{otherwise} \end{cases}$$

Truth Value of a Formula in $\mathcal{A}$ with respect to β

$\mathcal{A}(\beta) : \mathrm{F}_\Sigma(X) \to \{0, 1\}$ is defined inductively as follows:

$$\begin{aligned} \mathcal{A}(\beta)(\bot) &= 0 \\ \mathcal{A}(\beta)(\top) &= 1 \\ \mathcal{A}(\beta)(P(s_1, \ldots, s_n)) &= \text{if } (\mathcal{A}(\beta)(s_1), \ldots, \mathcal{A}(\beta)(s_n)) \in P_{\mathcal{A}} \text{ then 1 else 0} \\ \mathcal{A}(\beta)(s \approx t) &= \text{if } \mathcal{A}(\beta)(s) = \mathcal{A}(\beta)(t) \text{ then 1 else 0} \\ \mathcal{A}(\beta)(\neg F) &= 1 - \mathcal{A}(\beta)(F) \\ \mathcal{A}(\beta)(F \wedge G) &= \min(\mathcal{A}(\beta)(F), \mathcal{A}(\beta)(G)) \\ \mathcal{A}(\beta)(F \vee G) &= \max(\mathcal{A}(\beta)(F), \mathcal{A}(\beta)(G)) \\ \mathcal{A}(\beta)(F \to G) &= \max(1 - \mathcal{A}(\beta)(F), \mathcal{A}(\beta)(G)) \\ \mathcal{A}(\beta)(F \leftrightarrow G) &= \text{if } \mathcal{A}(\beta)(F) = \mathcal{A}(\beta)(G) \text{ then 1 else 0} \\ \mathcal{A}(\beta)(\forall x\, F) &= \min_{a \in U_{\mathcal{A}}} \{\mathcal{A}(\beta[x \mapsto a])(F)\} \\ \mathcal{A}(\beta)(\exists x\, F) &= \max_{a \in U_{\mathcal{A}}} \{\mathcal{A}(\beta[x \mapsto a])(F)\} \end{aligned}$$

Example

The "Standard" interpretation for Peano arithmetic:

$$\begin{array}{rcl} U_{\mathbb{N}} & = & \{0, 1, 2, \ldots\} \\ 0_{\mathbb{N}} & = & 0 \\ s_{\mathbb{N}} & : & n \mapsto n + 1 \\ +_{\mathbb{N}} & : & (n, m) \mapsto n + m \\ *_{\mathbb{N}} & : & (n, m) \mapsto n * m \\ <_{\mathbb{N}} & = & \{\, (n, m) \mid n \text{ less than } m \,\} \end{array}$$

Note that $\mathbb{N}$ is just one out of many possible Σ_{PA}-interpretations.

Values over $\mathbb{N}$ for sample terms and formulas:

Under the assignment $\beta : x \mapsto 1, y \mapsto 3$ we obtain

$$\begin{array}{lcl} \mathbb{N}(\beta)(s(x) + s(0)) & = & 3 \\ \mathbb{N}(\beta)(x + y \approx s(y)) & = & 1 \\ \mathbb{N}(\beta)(\forall x, y\, (x + y \approx y + x)) & = & 1 \\ \mathbb{N}(\beta)(\forall z\, (z < y)) & = & 0 \\ \mathbb{N}(\beta)(\forall x \exists y\, (x < y)) & = & 1 \end{array}$$

Ground Terms and Closed Formulas

If t is a ground term, then $\mathcal{A}(\beta)(t)$ does not depend on β, that is, $\mathcal{A}(\beta)(t) = \mathcal{A}(\beta')(t)$ for every β and β'.

Analogously, if F is a closed formula, then $\mathcal{A}(\beta)(F)$ does not depend on β, that is, $\mathcal{A}(\beta)(F) = \mathcal{A}(\beta')(F)$ for every β and β'.

An element $a \in U_{\mathcal{A}}$ is called *term-generated*, if $a = \mathcal{A}(\beta)(t)$ for some ground term t.

In general, not every element of an algebra is term-generated.

3.3 Models, Validity, and Satisfiability

F is *true* in $\mathcal{A}$ under assignment β:

$$\mathcal{A}, \beta \models F \quad :\Leftrightarrow \quad \mathcal{A}(\beta)(F) = 1$$

F is *true* in $\mathcal{A}$ ($\mathcal{A}$ is a *model* of F; F is *valid* in $\mathcal{A}$):

$$\mathcal{A} \models F \quad :\Leftrightarrow \quad \mathcal{A}, \beta \models F \text{ for all } \beta \in X \to U_{\mathcal{A}}$$

F is *valid* (or is a *tautology*):

$$\models F \quad :\Leftrightarrow \quad \mathcal{A} \models F \text{ for all } \mathcal{A} \in \Sigma\text{-Alg}$$

F is called *satisfiable* if there exist $\mathcal{A}$ and β such that $\mathcal{A}, \beta \models F$. Otherwise F is called *unsatisfiable*.

Entailment and Equivalence

F *entails* (*implies*) G (or G *is a consequence of* F), written $F \models G$, if for all $\mathcal{A} \in \Sigma$-Alg and $\beta \in X \to U_{\mathcal{A}}$, we have

$$\mathcal{A}, \beta \models F \quad \Rightarrow \quad \mathcal{A}, \beta \models G$$

F and G are called *equivalent*, written $F \models\!\models G$, if for all $\mathcal{A} \in \Sigma$-Alg and $\beta \in X \to U_{\mathcal{A}}$ we have

$$\mathcal{A}, \beta \models F \quad \Leftrightarrow \quad \mathcal{A}, \beta \models G$$

Proposition 3.1 *$F \models G$ if and only if $(F \to G)$ is valid*

Proof. ($\Rightarrow$) Suppose that $(F \to G)$ is not valid. Then there exist an algebra $\mathcal{A}$ and an assignment β such that $\mathcal{A}(\beta)(F \to G) = 0$, which means that $\mathcal{A}(\beta)(F) = 1$ and $\mathcal{A}(\beta)(G) = 0$, or in other words $\mathcal{A}, \beta \models F$ but not $\mathcal{A}, \beta \models G$. Consequently, $F \models G$ does not hold.

($\Leftarrow$) Suppose that $F \models G$ does not hold. Then there exist an algebra $\mathcal{A}$ and an assignment β such that $\mathcal{A}, \beta \models F$ but not $\mathcal{A}, \beta \models G$. Therefore $\mathcal{A}(\beta)(F) = 1$ and $\mathcal{A}(\beta)(G) = 0$, which implies $\mathcal{A}(\beta)(F \to G) = 0$, so $(F \to G)$ is not valid. □

Proposition 3.2 *$F \models\!\models G$ if and only if $(F \leftrightarrow G)$ is valid.*

Extension to sets of formulas N as in propositional logic, e.g.:

$$\begin{aligned} N \models F \quad :\Leftrightarrow \quad & \text{for all } \mathcal{A} \in \Sigma\text{-Alg and } \beta \in X \to U_{\mathcal{A}}\text{:} \\ & \text{if } \mathcal{A}, \beta \models G \text{ for all } G \in N, \text{ then } \mathcal{A}, \beta \models F. \end{aligned}$$

Validity vs. Unsatisfiability

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

Proposition 3.3 *Let F and G be formulas, let N be a set of formulas. Then*

(i) *F is valid if and only if $\neg F$ is unsatisfiable.*

(ii) *$F \models G$ if and only if $F \wedge \neg G$ is unsatisfiable.*

(iii) *$N \models G$ if and only if $N \cup \{\neg G\}$ is unsatisfiable.*

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

Substitution Lemma

Lemma 3.4 *Let $\mathcal{A}$ be a Σ-algebra, let β be an assignment, let σ be a substitution. Then for any Σ-term t*

$$\mathcal{A}(\beta)(t\sigma) = \mathcal{A}(\beta \circ \sigma)(t),$$

where $\beta \circ \sigma : X \to U_{\mathcal{A}}$ is the assignment $\beta \circ \sigma(x) = \mathcal{A}(\beta)(x\sigma)$.

Proof. We use induction over the structure of Σ-terms.

If $t = x$, then $\mathcal{A}(\beta \circ \sigma)(x) = \beta \circ \sigma(x) = \mathcal{A}(\beta)(x\sigma)$ by definition of $\beta \circ \sigma$.

If $t = f(t_1, \ldots, t_n)$, then $\mathcal{A}(\beta \circ \sigma)(f(t_1, \ldots, t_n)) = f_{\mathcal{A}}(\mathcal{A}(\beta \circ \sigma)(t_1), \ldots, \mathcal{A}(\beta \circ \sigma)(t_n)) = f_{\mathcal{A}}(\mathcal{A}(\beta)(t_1\sigma), \ldots, \mathcal{A}(\beta)(t_n\sigma)) = \mathcal{A}(\beta)(f(t_1\sigma, \ldots, t_n\sigma)) = \mathcal{A}(\beta)(f(t_1, \ldots, t_n)\sigma)$ by induction. □

Proposition 3.5 *Let $\mathcal{A}$ be a Σ-algebra, let β be an assignment, let σ be a substitution. Then for every Σ-formula F*

$$\mathcal{A}(\beta)(F\sigma) = \mathcal{A}(\beta \circ \sigma)(F)\,.$$

Corollary 3.6 $\mathcal{A}, \beta \models F\sigma \quad \Leftrightarrow \quad \mathcal{A}, \beta \circ \sigma \models F$

These theorems basically express that the syntactic concept of substitution corresponds to the semantic concept of an assignment.

Two Lemmas

Lemma 3.7 *Let $\mathcal{A}$ be a Σ-algebra. Let F be a Σ-formula with free variables $x_1, \ldots, x_n$. Then*

$$\mathcal{A} \models \forall x_1, \ldots, x_n\, F \quad \textit{if and only if} \quad \mathcal{A} \models F\,.$$

Proof. ($\Rightarrow$) Suppose that $\mathcal{A} \models \forall x_1, \ldots, x_n\, F$, that is, $\mathcal{A}(\beta)(\forall x_1, \ldots, x_n\, F) = 1$ for all assignments β. By definition, that means

$$\min_{a_1,\ldots,a_n \in U_\mathcal{A}} \{\mathcal{A}(\beta[x_1 \mapsto a_1, \ldots, x_n \mapsto a_n])(F)\} = 1,$$

and therefore $\mathcal{A}(\beta[x_1 \mapsto a_1, \ldots, x_n \mapsto a_n])(F) = 1$ for all $a_1, \ldots, a_n \in U_\mathcal{A}$.

Let γ be an arbitrary assigmnment. We have to show that $\mathcal{A}(\gamma)(F) = 1$. For every $i \in \{1, \ldots, n\}$ define $a_i = \gamma(x_i)$, then $\gamma = \gamma[x_1 \mapsto a_1, \ldots, x_n \mapsto a_n]$, and therefore $\mathcal{A}(\gamma)(F) = \mathcal{A}(\gamma[x_1 \mapsto a_1, \ldots, x_n \mapsto a_n])(F) = 1$.

($\Leftarrow$) Suppose that $\mathcal{A} \models F$, that is, $\mathcal{A}(\gamma)(F) = 1$ for all assignments γ.

Then in particular $\mathcal{A}(\beta[x_1 \mapsto a_1, \ldots, x_n \mapsto a_n])(F) = 1$ for all $a_1, \ldots, a_n \in U_\mathcal{A}$ (take $\gamma = \beta[x_1 \mapsto a_1, \ldots, x_n \mapsto a_n]$). Therefore

$$\mathcal{A}(\beta)(\forall x_1, \ldots, x_n\, F) = \min_{a_1,\ldots,a_n \in U_\mathcal{A}} \{\mathcal{A}(\beta[x_1 \mapsto a_1, \ldots, x_n \mapsto a_n])(F)\} = 1.$$

□

Note that it is not possible to replace $\mathcal{A} \models \ldots$ by $\mathcal{A}, \beta \models \ldots$ in Lemma 3.7.

Lemma 3.8 *Let $\mathcal{A}$ be a Σ-algebra. Let F be a Σ-formula with free variables $x_1, \ldots, x_n$. Let σ be a substitution and let $y_1, \ldots, y_m$ be the free variables of $F\sigma$. Then*

$$\mathcal{A} \models \forall x_1, \ldots, x_n\, F \quad \textit{implies} \quad \mathcal{A} \models \forall y_1, \ldots, y_m\, F\sigma\,.$$

Proof. By the previous lemma, we have $\mathcal{A} \models \forall x_1, \ldots, x_n\, F$ if and only if $\mathcal{A} \models F$ and similarly $\mathcal{A} \models \forall y_1, \ldots, y_m\, F\sigma$ if and only if $\mathcal{A} \models F\sigma$. So it suffices to show that $\mathcal{A} \models F$ implies $\mathcal{A} \models F\sigma$. Suppose that $\mathcal{A} \models F$, that is, $\mathcal{A}(\beta)(F) = 1$ for all assignments β. Then for every assignment γ, we have by Prop. 3.5 $\mathcal{A}(\gamma)(F\sigma) = \mathcal{A}(\gamma \circ \sigma)(F) = 1$ (take $\beta = \gamma \circ \sigma$), and therefore $\mathcal{A} \models F\sigma$. □

3.4 Algorithmic Problems

Validity(F): $\models F$?

Satisfiability(F): F satisfiable?

Entailment(F,G): does F entail G?

Model($\mathcal{A}$,F): $\mathcal{A} \models F$?

Solve($\mathcal{A}$,F): find an assignment β such that $\mathcal{A}, \beta \models F$.

Solve(F): find a substitution σ such that $\models F\sigma$.

Abduce(F): find G with "certain properties" such that $G \models F$.

Theory of an Algebra

Let $\mathcal{A} \in \Sigma$-Alg. The *(first-order) theory* of $\mathcal{A}$ is defined as

$$\mathrm{Th}(\mathcal{A}) = \{\, G \in \mathrm{F}_\Sigma(X) \mid \mathcal{A} \models G \,\}$$

Problem of axiomatizability:

Given an algebra $\mathcal{A}$ (or a class of algebras) can one *axiomatize* $\mathrm{Th}(\mathcal{A})$, that is, can one write down a formula F (or a recursively enumerable set F of formulas) such that

$$\mathrm{Th}(\mathcal{A}) = \{\, G \mid F \models G \,\}?$$

Two Interesting Theories

Let $\Sigma_{\mathrm{Pres}} = (\{0/0, s/1, +/2\},\ \{<\})$ and $\mathbb{N}_+ = (\mathbb{N}, 0, s, +, <)$ its standard interpretation on the natural numbers. $\mathrm{Th}(\mathbb{N}_+)$ is called *Presburger arithmetic* (M. Presburger, 1929). (There is no essential difference when one, instead of $\mathbb{N}$, considers the integer numbers $\mathbb{Z}$ as standard interpretation.)

Presburger arithmetic is decidable in 3EXPTIME (D. Oppen, JCSS, 16(3):323–332, 1978), and in 2EXPSPACE, using automata-theoretic methods (and there is a constant $c \geq 0$ such that $\mathrm{Th}(\mathbb{Z}_+) \notin \mathrm{NTIME}(2^{2^{cn}})$).

However, $\mathbb{N}_* = (\mathbb{N}, 0, s, +, *, <)$, the standard interpretation of $\Sigma_{\mathrm{PA}} = (\{0/0, s/1, +/2, */2\}, \{<\})$, has as theory the so-called *Peano arithmetic* which is undecidable and not even recursively enumerable.

(Non-)Computability Results

1. For most signatures Σ, validity is undecidable for Σ-formulas.
   (One can easily encode Turing machines in most signatures.)
2. Gödel's completeness theorem:
   For each signature Σ, the set of valid Σ-formulas is recursively enumerable.
   (We will prove this by giving complete deduction systems.)
3. Gödel's incompleteness theorem:
   For $\Sigma = \Sigma_{\text{PA}}$ and $\mathbb{N}_* = (\mathbb{N}, 0, s, +, *, <)$, the theory $\text{Th}(\mathbb{N}_*)$ is not recursively enumerable.

These complexity results motivate the study of subclasses of formulas (*fragments*) of first-order logic

Some Decidable Fragments

Some decidable fragments:

- *Monadic class*: no function symbols, all predicates unary; validity is NEXPTIME-complete.
- Variable-free formulas without equality: satisfiability is NP-complete. (why?)
- Variable-free Horn clauses (clauses with at most one positive atom): entailment is decidable in linear time.
- Finite model checking is decidable in exponential time and PSPACE-complete.

3.5 Normal Forms and Skolemization

Study of normal forms motivated by

- reduction of logical concepts,
- efficient data structures for theorem proving.

The main problem in first-order logic is the treatment of quantifiers. The subsequent normal form transformations are intended to eliminate many of them.

Prenex Normal Form (Traditional)

Prenex formulas have the form

$$\mathsf{Q}_1 x_1 \ldots \mathsf{Q}_n x_n \; F,$$

where F is quantifier-free and $\mathsf{Q}_i \in \{\forall, \exists\}$; we call $\mathsf{Q}_1 x_1 \ldots \mathsf{Q}_n x_n$ the *quantifier prefix* and F the *matrix* of the formula.

Computing prenex normal form by the reduction system $\Rightarrow_P$:

$$\begin{array}{rcl}
H[(F \leftrightarrow G)]_p & \Rightarrow_P & H[(F \to G) \wedge (G \to F)]_p \\
H[\neg \mathsf{Q} x \, F]_p & \Rightarrow_P & H[\overline{\mathsf{Q}} x \, \neg F]_p \\
H[((\mathsf{Q} x \, F) \circ G)]_p & \Rightarrow_P & H[\mathsf{Q} y \, (F\{x \mapsto y\} \circ G)]_p, \\
 & & \quad \circ \in \{\wedge, \vee\} \\
H[((\mathsf{Q} x \, F) \to G)]_p & \Rightarrow_P & H[\overline{\mathsf{Q}} y \, (F\{x \mapsto y\} \to G)]_p, \\
H[(F \circ (\mathsf{Q} x \, G))]_p & \Rightarrow_P & H[\mathsf{Q} y \, (F \circ G\{x \mapsto y\})]_p, \\
 & & \quad \circ \in \{\wedge, \vee, \to\}
\end{array}$$

Here y is always assumed to be some fresh variable and $\overline{\mathsf{Q}}$ denotes the quantifier *dual* to Q, i.e., $\overline{\forall} = \exists$ and $\overline{\exists} = \forall$.

Skolemization

Intuition: replacement of $\exists y$ by a concrete choice function computing y from all the arguments y depends on.

Transformation $\Rightarrow_S$
(to be applied outermost, *not* in subformulas):

$$\forall x_1, \ldots, x_n \exists y \, F \quad \Rightarrow_S \quad \forall x_1, \ldots, x_n \, F\{y \mapsto f(x_1, \ldots, x_n)\}$$

where f/n is a new function symbol (*Skolem function*).

Together: $F \Rightarrow_P^* \underbrace{G}_{\text{prenex}} \Rightarrow_S^* \underbrace{H}_{\text{prenex, no } \exists}$

Theorem 3.9 *Let F, G, and H as defined above and closed. Then*

(i) *F and G are equivalent.*

(ii) *$H \models G$ but the converse is not true in general.*

(iii) *G satisfiable (w.r.t. Σ-Alg) $\Leftrightarrow$ H satisfiable (w.r.t. Σ'-Alg) where $\Sigma' = (\Omega \cup SKF, \Pi)$ if $\Sigma = (\Omega, \Pi)$.*

The Complete Picture

$$
\begin{array}{lll}
F & \Rightarrow_P^* & \mathsf{Q}_1 y_1 \ldots \mathsf{Q}_n y_n \, G \qquad (G \text{ quantifier-free}) \\
& \Rightarrow_S^* & \forall x_1, \ldots, x_m \, H \qquad (m \leq n,\ H \text{ quantifier-free}) \\
& \Rightarrow_{CNF}^* & \underbrace{\underbrace{\forall x_1, \ldots, x_m}_{\text{leave out}} \bigwedge_{i=1}^{k} \underbrace{\bigvee_{j=1}^{n_i} L_{ij}}_{\text{clauses } C_i}}_{F'}
\end{array}
$$

$N = \{C_1, \ldots, C_k\}$ is called the *clausal (normal) form* (CNF) of F.
Note: The variables in the clauses are implicitly universally quantified.

Theorem 3.10 *Let F be closed. Then $F' \models F$. (The converse is not true in general.)*

Theorem 3.11 *Let F be closed. Then F is satisfiable if and only if F' is satisfiable if and only if N is satisfiable*

Optimization

The normal form algorithm described so far leaves lots of room for optimization. Note that we only can preserve satisfiability anyway due to Skolemization.

- the size of the CNF is exponential when done naively; the transformations we introduced already for propositional logic avoid this exponential growth;
- we want to preserve the original formula structure;
- we want small arity of Skolem functions (see next section).

3.6 First-order Tseitin Transformation

We generalize the Tseitin transformation that we have discussed for propositional logic to first-order logic.

First-order Replacement Theorem

Proposition 3.12 *Let $\mathcal{A}$ be a Σ-algebra, let F and G be Σ-formulas, and let $H = H[F]_p$ be a Σ-formula in which F occurs as a subformula at position p.*

If $\mathcal{A}(\beta)(F) = \mathcal{A}(\beta)(G)$ holds for all assignments β,then $\mathcal{A}(\gamma)(H[F]_p) = \mathcal{A}(\gamma)(H[G]_p)$ holds for all assignments γ.

Proof. By induction over the length of p, analogously to the proof of Prop. 2.7. □

Note: Since F may occur in H below quantifiers (and is therefore evaluated with a modified assignment), we cannot restrict ourselves to one particular assignment β.

Tseitin Transformation

Proposition 3.13 *Let $\Sigma = (\Omega, \Pi)$. A Σ-formula $H[F]_p$ is satisfiable if and only if the Σ'-formula*

$$H[P(x_1, \ldots, x_n)]_p \wedge \forall x_1, \ldots, x_n \, (P(x_1, \ldots, x_n) \leftrightarrow F)$$

is satisfiable, where $\{x_1, \ldots, x_n\}$ are the free variables in F, P/n is a new predicate symbol, and $\Sigma' = (\Omega, \Pi \cup \{P/n\})$.

Proof. "⇒": Suppose that the Σ-formula $H[F]_p$ is satisfiable. Let $\mathcal{A}$ be a Σ-algebra and γ be an assignment such that $\mathcal{A}(\gamma)(H[F]_p) = 1$. Define a Σ'-algebra $\mathcal{A}'$ with the same universe as $\mathcal{A}$ and the same interpretation as $\mathcal{A}$ for all symbols in Σ. Define $P_{\mathcal{A}'}$ as the set of all $(a_1, \ldots, a_n) \in U_{\mathcal{A}}^n$ such that $\mathcal{A}(\beta[x_1 \mapsto a_1, \ldots, x_n \mapsto a_n])(F)$ holds. Since $\mathcal{A}$ and $\mathcal{A}'$ agree for all symbols occurring in $H[F]_p$, $\mathcal{A}'(\gamma)(H[F]_p) = \mathcal{A}(\gamma)(H[F]_p) = 1$. Furthermore $\mathcal{A}'(\beta)(\forall x_1, \ldots, x_n \, (P(x_1, \ldots, x_n) \leftrightarrow F)) = 1$ for every β, therefore by Prop. 3.12, $\mathcal{A}'(\gamma)(H[P(x_1, \ldots, x_n)]_p) = \mathcal{A}'(\gamma)(H[F]_p) = 1$.

"⇐": Let $\mathcal{A}'$ be a Σ'-algebra and γ an assignment such that $\mathcal{A}'(\gamma)(H[P(x_1, \ldots, x_n)]_p \wedge \forall x_1, \ldots, x_n \, (P(x_1, \ldots, x_n) \leftrightarrow F)) = 1$. Then $\mathcal{A}'(\beta)(P(x_1, \ldots, x_n) = \mathcal{A}'(\beta)(F)$ for every β, and therefore $\mathcal{A}'(\gamma)(H[F]_p) = \mathcal{A}'(\gamma)(H[P(x_1, \ldots, x_n)]_p) = 1$ by Prop. 3.12. Now define the Σ-algebra $\mathcal{A}$ with the same universe as $\mathcal{A}'$ and the same interpretation as $\mathcal{A}$ for all symbols in Σ, then $\mathcal{A}(\gamma)(H[F]_p) = \mathcal{A}'(\gamma)(H[F]_p) = 1$ since $\mathcal{A}$ and $\mathcal{A}'$ agree for all symbols occurring in $H[F]_p$. □

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

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

Proposition 3.14 *Let $\mathcal{A}$ be a Σ-algebra, let F and G be Σ-formulas, and let $H = H[F]_p$ be a Σ-formula in which F occurs as a subformula at position p.*

If $\mathrm{pol}(H, p) = 1$ and $\mathcal{A}(\beta)(F) \leq \mathcal{A}(\beta)(G)$ holds for all assignments β, then $\mathcal{A}(\gamma)(H[F]_p) \leq \mathcal{A}(\gamma)(H[G]_p)$ holds for all assignments γ.

If $\mathrm{pol}(H, p) = -1$ and $\mathcal{A}(\beta)(F) \geq \mathcal{A}(\beta)(G)$ holds for all assignments β, then $\mathcal{A}(\gamma)(H[F]_p) \leq \mathcal{A}(\gamma)(H[G]_p)$ holds for all assignments γ.

Proof. Exercise. □

Let P be a predicate symbol not occurring in $H[F]_p$. Let $\{x_1, \ldots, x_n\}$ be the free variables in F.

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

$$\begin{aligned}
&\forall x_1, \ldots, x_n\ (P(x_1, \ldots, x_n) \to F), \text{ if } \mathrm{pol}(H, p) = 1, \\
&\forall x_1, \ldots, x_n\ (F \to P(x_1, \ldots, x_n)), \text{ if } \mathrm{pol}(H, p) = -1, \\
&\forall x_1, \ldots, x_n\ (P(x_1, \ldots, x_n) \leftrightarrow F), \text{ if } \mathrm{pol}(H, p) = 0.
\end{aligned}$$

Proposition 3.15 *Let $\Sigma = (\Omega, \Pi)$. A Σ-formula $H[F]_p$ is satisfiable if and only if the Σ'-formula*

$$H[P(x_1, \ldots, x_n)]_p \wedge \mathrm{def}(H, p, P, F)$$

is satisfiable, where $\{x_1, \ldots, x_n\}$ are the free variables in F, P/n is a new predicate symbol, and $\Sigma' = (\Omega, \Pi \cup \{P/n\})$.

Proof. Analogously to Prop. 3.13, using Prop. 3.14 instead of Prop. 3.12. □

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 (Nonnengart and Weidenbach 2001).

We will introduce definitions for subformulas only in those cases where ν decreases ("beneficial subformulas").

G	$\nu(G)$	$\bar{\nu}(G)$
$A, \top, \bot$	1	1
$F_1 \wedge F_2$	$\nu(F_1) + \nu(F_2)$	$\bar{\nu}(F_1)\bar{\nu}(F_2)$
$F_1 \vee F_2$	$\nu(F_1)\nu(F_2)$	$\bar{\nu}(F_1) + \bar{\nu}(F_2)$
$\neg F_1$	$\bar{\nu}(F_1)$	$\nu(F_1)$
$F_1 \rightarrow F_2$	$\bar{\nu}(F_1)\nu(F_2)$	$\nu(F_1) + \bar{\nu}(F_2)$
$F_1 \leftrightarrow F_2$	$\nu(F_1)\bar{\nu}(F_2) + \bar{\nu}(F_1)\nu(F_2)$	$\nu(F_1)\nu(F_2) + \bar{\nu}(F_1)\bar{\nu}(F_2)$
$\exists x\, F,\ \forall x\, F$	$\nu(F)$	$\bar{\nu}(F)$

3.7 Getting Skolem Functions with Small Arity

A clause set that is better suited for automated theorem proving can be obtained using the following steps:

- eliminate trivial subformulas
- replace beneficial subformulas
- produce a negation normal form (NNF)
- apply miniscoping
- rename all variables
- Skolemize
- push quantifiers upward
- apply distributivity

We start with a closed formula.

Elimination of Trivial Subformulas

Eliminate subformulas $\top$ and $\bot$ modulo associativity/commutativity of $\wedge$ and $\vee$:

$$
\begin{aligned}
H[(F \wedge \top)]_p &\Rightarrow_{\mathrm{OCNF}} H[F]_p \\
H[(F \vee \bot)]_p &\Rightarrow_{\mathrm{OCNF}} H[F]_p \\
H[(F \leftrightarrow \bot)]_p &\Rightarrow_{\mathrm{OCNF}} H[\neg F]_p \\
H[(F \leftrightarrow \top)]_p &\Rightarrow_{\mathrm{OCNF}} H[F]_p \\
H[(F \vee \top)]_p &\Rightarrow_{\mathrm{OCNF}} H[\top]_p \\
H[(F \wedge \bot)]_p &\Rightarrow_{\mathrm{OCNF}} H[\bot]_p \\
H[\neg\top]_p &\Rightarrow_{\mathrm{OCNF}} H[\bot]_p \\
H[\neg\bot]_p &\Rightarrow_{\mathrm{OCNF}} H[\top]_p \\
H[(F \rightarrow \bot)]_p &\Rightarrow_{\mathrm{OCNF}} H[\neg F]_p \\
H[(F \rightarrow \top)]_p &\Rightarrow_{\mathrm{OCNF}} H[\top]_p \\
H[(\bot \rightarrow F)]_p &\Rightarrow_{\mathrm{OCNF}} H[\top]_p \\
H[(\top \rightarrow F)]_p &\Rightarrow_{\mathrm{OCNF}} H[F]_p \\
H[\mathsf{Q}x\,\top]_p &\Rightarrow_{\mathrm{OCNF}} H[\top]_p \\
H[\mathsf{Q}x\,\bot]_p &\Rightarrow_{\mathrm{OCNF}} H[\bot]_p
\end{aligned}
$$

Replacement of Beneficial Subformulas

Introduce top-down fresh predicates for beneficial subformulas:

$$H[F]_p \;\Rightarrow_{\mathrm{OCNF}}\; H[P(x_1,\ldots,x_n)]_p \wedge \mathrm{def}(H,p,P,F)$$

if $\nu(H[F]_p) > \nu(H[P(\ldots)]_p \wedge \mathrm{def}(H,p,P,F))$,

where $\{x_1,\ldots,x_n\}$ are the free variables in F, P/n is a predicate new to $H[F]_p$, and $\mathrm{def}(H,p,P,F)$ is defined by

$$\begin{array}{l}\forall x_1,\ldots,x_n\ (P(x_1,\ldots,x_n) \to F),\ \text{if } \mathrm{pol}(H,p) = 1,\\ \forall x_1,\ldots,x_n\ (F \to P(x_1,\ldots,x_n)),\ \text{if } \mathrm{pol}(H,p) = -1,\\ \forall x_1,\ldots,x_n\ (P(x_1,\ldots,x_n) \leftrightarrow F),\ \text{if } \mathrm{pol}(H,p) = 0.\end{array}$$

Remark: Computing ν is not practical in general. However, the test

$$\nu(H[F]_p) > \nu(H[P(\ldots)]_p \wedge \mathrm{def}(H,p,P,F))$$

can be computed in constant time without actually computing ν.

Negation Normal Form (NNF)

Apply the reduction system $\Rightarrow_{\mathrm{NNF}}$:

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

if $\mathrm{pol}(H,p) = 1$ or $\mathrm{pol}(H,p) = 0$.

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

if $\mathrm{pol}(H,p) = -1$.

$$\begin{array}{rcl} H[F \to G]_p & \Rightarrow_{\mathrm{NNF}} & H[\neg F \vee G]_p \\ H[\neg\neg F]_p & \Rightarrow_{\mathrm{NNF}} & H[F]_p \\ H[\neg(F \vee G)]_p & \Rightarrow_{\mathrm{NNF}} & H[\neg F \wedge \neg G]_p \\ H[\neg(F \wedge G)]_p & \Rightarrow_{\mathrm{NNF}} & H[\neg F \vee \neg G]_p \\ H[\neg \mathsf{Q} x\ F]_p & \Rightarrow_{\mathrm{NNF}} & H[\overline{\mathsf{Q}} x\ \neg F]_p \end{array}$$

Miniscoping

Apply the reduction system $\Rightarrow_{\mathrm{MS}}$ modulo associativity and commutativity of $\wedge$, $\vee$. For the rules below we assume that x occurs freely in F, F', but x does not occur freely in G:

$$\begin{aligned}
H[\mathsf{Q}x\,(F \wedge G)]_p &\Rightarrow_{\mathrm{MS}} H[(\mathsf{Q}x\,F) \wedge G]_p \\
H[\mathsf{Q}x\,(F \vee G)]_p &\Rightarrow_{\mathrm{MS}} H[(\mathsf{Q}x\,F) \vee G]_p \\
H[\forall x\,(F \wedge F')]_p &\Rightarrow_{\mathrm{MS}} H[(\forall x\,F) \wedge (\forall x\,F')]_p \\
H[\exists x\,(F \vee F')]_p &\Rightarrow_{\mathrm{MS}} H[(\exists x\,F) \vee (\exists x\,F')]_p \\
H[\mathsf{Q}x\,G]_p &\Rightarrow_{\mathrm{MS}} H[G]_p
\end{aligned}$$

Variable Renaming

Rename all variables in H such that there are no two different positions p, q with $H|_p = \mathsf{Q}x\,F$ and $H|_q = \mathsf{Q}'x\,G$.

Standard Skolemization

Apply the reduction system:

$$H[\exists x\,F]_p \Rightarrow_{\mathrm{SK}} H[F\{x \mapsto f(y_1, \ldots, y_n)\}]_p$$

where p has minimal length,
$\{y_1, \ldots, y_n\}$ are the free variables in $\exists x\,F$,
and f/n is a new function symbol to H.

Final Steps

Apply the reduction system modulo commutativity of $\wedge$, $\vee$ to push $\forall$ upward:

$$\begin{aligned}
H[(\forall x\,F) \wedge G]_p &\Rightarrow_{\mathrm{OCNF}} H[\forall x\,(F \wedge G)]_p \\
H[(\forall x\,F) \vee G]_p &\Rightarrow_{\mathrm{OCNF}} H[\forall x\,(F \vee G)]_p
\end{aligned}$$

Note that variable renaming ensures that x does not occur in G.

Apply the reduction system modulo commutativity of $\wedge$, $\vee$ to push disjunctions downward:

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

The resulting algorithm is close to the state of the art, but some additional redundancy tests and simplification mechanisms are still missing.

3.8 Herbrand Interpretations

From now on we shall consider FOL without equality. We assume that Ω contains at least one constant symbol.

An *Herbrand interpretation* (over Σ) is a Σ-algebra $\mathcal{A}$ such that

- $U_{\mathcal{A}} = \mathrm{T}_\Sigma$ (= the set of ground terms over Σ)
- $f_{\mathcal{A}} : (s_1, \ldots, s_n) \mapsto f(s_1, \ldots, s_n),\ f/n \in \Omega$

In other words, *values are fixed* to be ground terms and *functions are fixed* to be the *term constructors*. Only predicate symbols $P/m \in \Pi$ may be freely interpreted as relations $P_{\mathcal{A}} \subseteq \mathrm{T}_\Sigma^m$.

Proposition 3.16 *Every set of ground atoms I uniquely determines an Herbrand interpretation $\mathcal{A}$ via*

$$(s_1, \ldots, s_n) \in P_{\mathcal{A}} \quad \textit{if and only if} \quad P(s_1, \ldots, s_n) \in I$$

Thus we shall identify Herbrand interpretations (over Σ) with sets of Σ-ground atoms.

Existence of Herbrand Models

An Herbrand interpretation I is called an *Herbrand model* of F, if $I \models F$.

The importance of Herbrand models lies in the following theorem, which we will prove later in this lecture:

Let N be a set of (universally quantified) Σ-clauses. Then the following properties are equivalent:

(1) N has a model.
(2) N has an Herbrand model (over Σ).
(3) $G_\Sigma(N)$ has an Herbrand model (over Σ).

where $G_\Sigma(N) = \{\, C\sigma \text{ ground clause} \mid (\forall \vec{x}\, C) \in N,\ \sigma : X \to \mathrm{T}_\Sigma \,\}$ is the set of *ground instances* of N.

3.9 Inference Systems and Proofs

Inference systems Γ (proof calculi) are sets of tuples

$$(F_1, \ldots, F_n, F_{n+1}),\ n \geq 0,$$

called *inferences*, and written

$$\frac{\overbrace{F_1 \ \ldots \ F_n}^{\text{premises}}}{\underbrace{F_{n+1}}_{\text{conclusion}}} \quad \textit{side condition}$$

Clausal inference system: premises and conclusions are clauses. One also considers inference systems over other data structures.

Inference Systems

Inference systems Γ are shorthands for reduction systems over sets of formulas. If N is a set of formulas, then

$$\frac{\overbrace{F_1 \ \ldots \ F_n}^{\text{premises}}}{\underbrace{F_{n+1}}_{\text{conclusion}}} \quad \textit{side condition}$$

is a shorthand for

$$N \cup \{F_1, \ldots, F_n\} \quad \Rightarrow_\Gamma \quad N \cup \{F_1, \ldots, F_n\} \cup \{F_{n+1}\} \quad \text{if } \textit{side condition}$$

Proofs

A *proof* in Γ of a formula F from a set of formulas N (called *assumptions*) is a sequence $F_1, \ldots, F_k$ of formulas where

(i) $F_k = F$,

(ii) for all $1 \leq i \leq k$: $F_i \in N$ or there exists an inference

$$\frac{F_{m_1} \ \ldots \ F_{m_n}}{F_i}$$

in Γ, such that $0 \leq m_j < i$, for $1 \leq j \leq n$.

Soundness and Completeness

Provability $\vdash_\Gamma$ of F from N in Γ:

$N \vdash_\Gamma F$ if there exists a proof in Γ of F from N.

Γ is called *sound*, if

$$\frac{F_1 \ \ldots \ F_n}{F} \in \Gamma \text{ implies } F_1, \ldots, F_n \models F$$

Γ is called *complete*, if

$N \models F$ implies $N \vdash_\Gamma F$

Γ is called *refutationally complete*, if

$N \models \bot$ implies $N \vdash_\Gamma \bot$

Proposition 3.17

(i) Let Γ be sound. Then $N \vdash_\Gamma F \Rightarrow N \models F$

(ii) If $N \vdash_\Gamma F$ then there exist finitely many $F_1, \ldots, F_n \in N$ such that $F_1, \ldots, F_n \vdash_\Gamma F$

Reduced Proofs

The definition of a proof of F given above admits sequences $F_1, \ldots, F_k$ of formulas where some F_i are not ancestors of $F_k = F$ (i.e., some F_i are not actually used to derive F).

A proof is called *reduced*, if every F_i with $i < k$ is an ancestor of F_k.

We obtain a reduced proof from a proof by marking first F_k and then recursively all the premises used to derive a marked conclusion, and by deleting all non-marked formulas in the end.

Reduced Proofs as Trees

markings	$\widehat{=}$	formulas			
leaves	$\widehat{=}$	assumptions and axioms			
other nodes	$\widehat{=}$	inferences:	conclusion	$\widehat{=}$	parent node
			premises	$\widehat{=}$	child nodes

$$\dfrac{\dfrac{\dfrac{P(f(c)) \vee Q(b) \quad \neg P(f(c)) \vee \neg P(f(c)) \vee Q(b)}{\neg P(f(c)) \vee Q(b) \vee Q(b)}}{\dfrac{P(f(c)) \vee Q(b) \qquad \neg P(f(c)) \vee Q(b)}{Q(b) \vee Q(b)}}}{\dfrac{\dfrac{Q(b)}{P(f(c)) \qquad \neg P(f(c))}}{\bot}} \qquad \neg P(f(c)) \vee \neg Q(b)$$

Mandatory vs. Admissible Inferences

It is useful to distinguish between two kinds of inferences:

- Mandatory (required) inferences:

  Must be performed to ensure refutational completeness.

  The less, the better.

- Optional (admissible) inferences:

  May be performed, if useful.

We will first consider only mandatory inferences.

3.10 Ground (or propositional) Resolution

We observe that propositional clauses and ground clauses are essentially the same, as long as we do not consider equational atoms.

In this section we only deal with ground clauses.

Unlike in Section 2 we admit duplicated literals in clauses, i.e., we treat clauses like multisets of literals, not like sets.

The Resolution Calculus *Res*

Resolution inference rule:

$$\frac{D \vee A \qquad C \vee \neg A}{D \vee C}$$

Terminology: $D \vee C$: *resolvent*; A: *resolved atom*

(Positive) factorization inference rule:

$$\frac{C \vee A \vee A}{C \vee A}$$

These are *schematic inference rules*; for each substitution of the *schematic variables* C, D, and A, by ground clauses and ground atoms, respectively, we obtain an inference.

We treat "$\vee$" as associative and commutative, hence A and $\neg A$ can occur anywhere in the clauses; moreover, when we write $C \vee A$, etc., this includes unit clauses, that is, $C = \bot$.

Sample Refutation

1	$\neg P(f(c)) \vee \neg P(f(c)) \vee Q(b)$	(given)
2	$P(f(c)) \vee Q(b)$	(given)
3	$\neg P(g(b,c)) \vee \neg Q(b)$	(given)
4	$P(g(b,c))$	(given)
5	$\neg P(f(c)) \vee Q(b) \vee Q(b)$	(Res. 2 into 1)
6	$\neg P(f(c)) \vee Q(b)$	(Fact. 5)
7	$Q(b) \vee Q(b)$	(Res. 2 into 6)
8	$Q(b)$	(Fact. 7)
9	$\neg P(g(b,c))$	(Res. 8 into 3)
10	$\bot$	(Res. 4 into 9)

Soundness of Resolution

Theorem 3.18 *Ground first-order resolution is sound.*

Proof. As in propositional logic. □

Note: In ground first-order logic we have (like in propositional logic):

1. $\mathcal{B} \models L_1 \vee \ldots \vee L_n$ if and only if there exists i: $\mathcal{B} \models L_i$.
2. $\mathcal{B} \models A$ or $\mathcal{B} \models \neg A$.

This does not hold for formulas with variables!

3.11 Refutational Completeness of Resolution

How to show refutational completeness of ground resolution:

- We have to show: $N \models \bot \;\Rightarrow\; N \vdash_{Res} \bot$, or equivalently: If $N \not\vdash_{Res} \bot$, then N has a model.
- Idea: Suppose that we have computed sufficiently many inferences (and not derived $\bot$).
- Now order the clauses in N according to some appropriate ordering, inspect the clauses in ascending order, and construct a series of Herbrand interpretations.
- The limit interpretation can be shown to be a model of N.

Closure of Clause Sets under *Res*

$$
\begin{aligned}
Res(N) &= \{\, C \mid C \text{ is conclusion of an inference in } Res \\
&\qquad\qquad \text{with premises in } N \,\} \\
Res^0(N) &= N \\
Res^{n+1}(N) &= Res(Res^n(N)) \cup Res^n(N), \text{ for } n \geq 0 \\
Res^*(N) &= \textstyle\bigcup_{n\geq 0} Res^n(N)
\end{aligned}
$$

N is called *saturated* (w.r.t. resolution), if $Res(N) \subseteq N$.

Proposition 3.19

(i) $Res^*(N)$ *is saturated.*

(ii) *Res is refutationally complete, if and only if for each set* N *of ground clauses:*

$$N \models \bot \;\; \textit{implies} \;\; \bot \in Res^*(N)$$

Proof. (i): We have to show that $Res(Res^*(N)) \subseteq Res^*(N)$, or in other words, that the conclusion of every inference in Res with premises in $Res^*(N)$ is again contained in $Res^*(N)$. An inference in Res is either a resolution inference or a factorization inference. Let us first consider a resolution inference with premises $C_1 \in Res^*(N)$ and $C_2 \in Res^*(N)$ and conclusion C. Since $Res^*(N) = \bigcup_{n\geq 0} Res^n(N)$, we know that there exist $j, k \geq 0$ such that $C_1 \in Res^j(N)$ and $C_2 \in Res^k(N)$. Without loss of generality assume that $j \geq k$. It is easy to see that in this case $Res^k(N) \subseteq Res^j(N)$, hence $C_1 \in Res^j(N)$ and $C_2 \in Res^j(N)$. Consequently, $C \in Res(Res^j(N)) \subseteq Res^{j+1}(N) \subseteq Res^*(N)$.

Otherwise we have a factorization inference with premise $C_1 \in Res^*(N)$ and conclusion C. Again we conclude that $C_1 \in Res^j(N)$ for some $j \geq 0$, hence $C \in Res(Res^j(N)) \subseteq Res^{j+1}(N) \subseteq Res^*(N)$.

(ii) This part follows immediately from the fact that for every clause C we have $N \vdash_{Res} C$ if and only if $C \in Res^*(N)$. □

Clause Orderings

1. We assume that $\succ$ is any fixed ordering on ground atoms that is *total* and *well-founded*. (There exist many such orderings, e.g., the length-based ordering on atoms when these are viewed as words over a suitable alphabet.)
2. Extend $\succ$ to an *ordering* $\succ_L$ *on ground literals:*

$$\begin{array}{lll} [\neg]A & \succ_L & [\neg]B \quad \text{, if } A \succ B \\ \neg A & \succ_L & A \end{array}$$

3. Extend $\succ_L$ to an *ordering* $\succ_C$ *on ground clauses:*
   $\succ_C = (\succ_L)_{\text{mul}}$, the multiset extension of $\succ_L$.

   Notation: $\succ$ also for $\succ_L$ and $\succ_C$.

Example

Suppose $A_5 \succ A_4 \succ A_3 \succ A_2 \succ A_1 \succ A_0$. Then:

$$\begin{array}{ll} & A_1 \vee \neg A_5 \\ \succ & A_3 \vee \neg A_4 \\ \succ & \neg A_1 \vee A_3 \vee A_4 \\ \succ & A_1 \vee \neg A_2 \\ \succ & \neg A_1 \vee A_2 \\ \succ & A_1 \vee A_1 \vee A_2 \\ \succ & A_0 \vee A_1 \end{array}$$

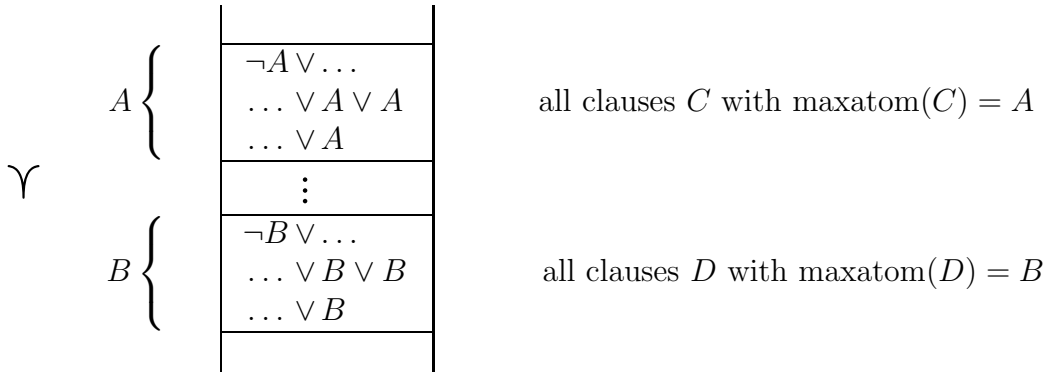
Properties of the Clause Ordering

Proposition 3.20

1. *The orderings on literals and clauses are total and well-founded.*
2. *Let C and D be clauses with $A = \text{maxatom}(C)$, $B = \text{maxatom}(D)$, where $\text{maxatom}(C)$ denotes the maximal atom in C.*
   (i) If $A \succ B$ then $C \succ D$.
   (ii) If $A = B$, A occurs negatively in C but only positively in D, then $C \succ D$.

Stratified Structure of Clause Sets

Let $A \succ B$. Clause sets are then stratified in this form:



Construction of Interpretations

Given: set N of ground clauses, atom ordering $\succ$.

Wanted: Herbrand interpretation I such that

$$I \models N \quad \text{if } N \text{ is saturated and } \bot \notin N$$

Construction according to $\succ$, starting with the smallest clause.

Main Ideas of the Construction

- Clauses are considered in the order given by $\succ$.
- When considering C, one already has an interpretation so far available (I_C). Initially $I_C = \emptyset$.
- If C is true in this interpretation, nothing needs to to be changed.
- Otherwise, one would like to change the interpretation such that C becomes true.
- Changes should, however, be *monotone*. One never deletes atoms from the interpretation, and the truth value of clauses smaller than C should not change from true to false.
- Hence, one adds $\Delta_C = \{A\}$, if and only if C is false in I_C, if A occurs positively in C (*adding A will make C become true*) and if this occurrence in C is strictly maximal in the ordering on literals (*changing the truth value of A has no effect on smaller clauses*). Otherwise, $\Delta_C = \emptyset$.

- We say that the construction fails for a clause C, if C is false in I_C and $\Delta_C = \emptyset$.
- We will show: If there are clauses for which the construction fails, then some inference with the smallest such clause (the so-called "minimal counterexample") has not been computed. Otherwise, the limit interpretation is a model of all clauses.

Construction of Candidate Interpretations

Let $N, \succ$ be given. We define sets I_C and Δ_C for all ground clauses C over the given signature inductively over $\succ$:

$$I_C := \bigcup_{C \succ D} \Delta_D$$

$$\Delta_C := \begin{cases} \{A\}, & \text{if } C \in N,\ C = C' \vee A,\ A \succ C',\ I_C \not\models C \\ \emptyset, & \text{otherwise} \end{cases}$$

We say that C *produces* A, if $\Delta_C = \{A\}$.

Note that the definitions satisfy the conditions of Thm. 1.8; so they are well-defined even if $\{ D \mid C \succ D \}$ is infinite.

The *candidate interpretation* for N (w.r.t. $\succ$) is given as $I_N^{\succ} := \bigcup_C \Delta_C$. (We also simply write I_N or I for $I_N^{\succ}$ if $\succ$ is either irrelevant or known from the context.)

Example

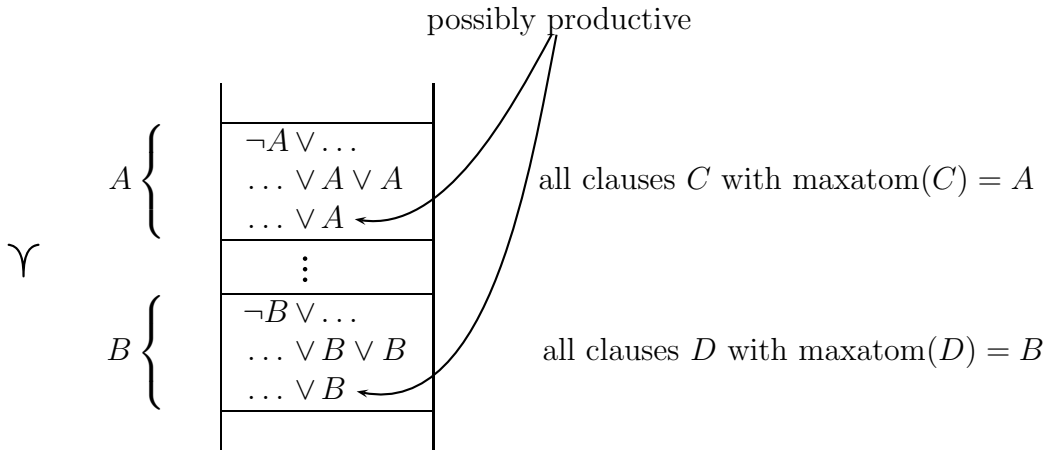
Let $A_5 \succ A_4 \succ A_3 \succ A_2 \succ A_1 \succ A_0$ (max. literals in red)

	clauses C	I_C	Δ_C	Remarks
7	$\neg A_1 \vee A_5$	$\{A_1, A_2, A_4\}$	$\{A_5\}$	
6	$\neg A_1 \vee A_3 \vee \neg A_4$	$\{A_1, A_2, A_4\}$	$\emptyset$	max. lit. $\neg A_4$ neg.; *min. counter-ex.*
5	$A_0 \vee \neg A_1 \vee A_3 \vee A_4$	$\{A_1, A_2\}$	$\{A_4\}$	A_4 maximal
4	$\neg A_1 \vee A_2$	$\{A_1\}$	$\{A_2\}$	A_2 maximal
3	$A_1 \vee A_2$	$\{A_1\}$	$\emptyset$	true in I_C
2	$A_0 \vee A_1$	$\emptyset$	$\{A_1\}$	A_1 maximal
1	$\neg A_0$	$\emptyset$	$\emptyset$	true in I_C

$I = \{A_1, A_2, A_4, A_5\}$ is not a model of the clause set
$\Rightarrow$ there exists a *counterexample*.

Structure of $N, \succ$

Let $A \succ B$. Note that producing a new atom does not change the truth value of smaller clauses.



Some Properties of the Construction

Proposition 3.21

(i) If $D = D' \vee \neg A$, then no $C \succeq D$ produces A.
(ii) If $I_D \models D$, then $I_C \models D$ for every $C \succeq D$ and $I_N^\succ \models D$.
(iii) If $D = D' \vee A$ produces A, then $I_C \models D$ for every $C \succ D$ and $I_N^\succ \models D$.
(iv) If $D = D' \vee A$ produces A, then $I_C \not\models D'$ for every $C \succeq D$ and $I_N^\succ \not\models D'$.
(v) If for every clause $C \in N$, C is productive or $I_C \models C$, then $I_N^\succ \models N$.

Proof. *(i) If C produces A, then $A \succeq L$ for every literal L of C. On the other hand, D contains $\neg A$, and $\neg A \succ A$. Since $\neg A \succ L$ for every literal L of C, we obtain $D \succ C$.*

(ii) Suppose that $I_D \models D$ and $C \succeq D$. If $I_D \models A$ for some positive literal A of D, then $A \in I_D \subseteq I_C \subseteq I_N^\succ$, so $I_C \models D$ and $I_N^\succ \models D$. Otherwise $I_D \models \neg A$ for some negative literal $\neg A$ of D, hence $A \notin I_D$. By (i), no clause that is larger than or equal to D produces A, so $A \notin I_C$ and $A \notin I_N^\succ$. Again, $I_C \models D$ and $I_N^\succ \models D$.

(iii) Obvious, since $C \succ D$ implies $A \in \Delta_D \subseteq I_C \subseteq I_N^\succ$.

(iv) If $D = D' \vee A$ produces A, then $A \succ L$ for every literal L of D' and $I_D \not\models A$. Since $I_D \not\models D$, we have $I_D \not\models L$ for every literal L of D'. Let $C \succeq D$. If L is a positive literal A', then $A' \notin I_D$. Since all atoms in $I_C \setminus I_D$ and $I_N^\succ \setminus I_D$ are larger than or equal to A, we get $A' \notin I_C$ and $A' \notin I_N^\succ$. Otherwise L is a negative literal $\neg A'$, then obviously $A' \in I_D \subseteq I_C \subseteq I_N^\succ$. In both cases L is false in I_C and $I_N^\succ$.

(v) By (ii) and (iii). □

Model Existence Theorem

Proposition 3.22 *Let $\succ$ be a clause ordering. If N is saturated w.r.t. Res and $\bot \notin N$, then for every clause $C \in N$, C is productive or $I_C \models C$.*

Proof. Let N be saturated w.r.t. *Res* and $\bot \notin N$. Assume that the proposition does not hold. By well-foundedness, there must exist a minimal clause $C \in N$ (w.r.t. $\succ$) such that C is neither productive nor $I_C \models C$. As $C \neq \bot$ there exists a maximal literal in C. There are two possible reasons why C is not productive:

Case 1: The maximal literal $\neg A$ is negative, i.e., $C = C' \vee \neg A$. Then $I_C \models A$ and $I_C \not\models C'$. So some $D = D' \vee A \in N$ with $C \succ D$ produces A, and $I_C \not\models D'$. The inference

$$\frac{D' \vee A \qquad C' \vee \neg A}{D' \vee C'}$$

yields a clause $D' \vee C' \in N$ that is smaller than C. As $I_C \not\models D' \vee C'$, we know that $D' \vee C'$ is neither productive nor $I_{D' \vee C'} \models D' \vee C'$. This contradicts the minimality of C.

Case 2: The maximal literal A is positive, but not strictly maximal, i.e., $C = C' \vee A \vee A$. Then there is an inference

$$\frac{C' \vee A \vee A}{C' \vee A}$$

that yields a smaller clause $C' \vee A \in N$. As $I_C \not\models C' \vee A$, this clause is neither productive nor $I_{C' \vee A} \models C' \vee A$. Since $C \succ C' \vee A$, this contradicts the minimality of C. □

Theorem 3.23 (Bachmair & Ganzinger 1990) *Let $\succ$ be a clause ordering. If N is saturated w.r.t. Res and $\bot \notin N$, then $I_N^{\succ} \models N$.*

Proof. By Prop. 3.22 and part (v) of Prop. 3.21. □

Corollary 3.24 *Let N be saturated w.r.t. Res. Then $N \models \bot$ if and only if $\bot \in N$.*

Compactness of Propositional Logic

Lemma 3.25 *Let N be a set of propositional (or first-order ground) clauses. Then N is unsatisfiable, if and only if some finite subset $N' \subseteq N$ is unsatisfiable.*

Proof. The "if" part is trivial. For the "only if" part, assume that N be unsatisfiable. Consequently, $Res^*(N)$ is unsatisfiable as well. By refutational completeness of resolution, $\bot \in Res^*(N)$. So there exists an $n \geq 0$ such that $\bot \in Res^n(N)$, which means that $\bot$ has a finite resolution proof. Now choose N' as the set of assumptions in this proof. □

Theorem 3.26 (Compactness for Propositional Formulas) *Let S be a set of propositional (or first-order ground) formulas. Then S is unsatisfiable, if and only if some finite subset $S' \subseteq S$ is unsatisfiable.*

Proof. The "if" part is again trivial. For the "only if" part, assume that S be unsatisfiable. Transform S into an equivalent set N of clauses. By the previous lemma, N has a finite unsatisfiable subset N'. Now choose for every clause C in N' one formula F of S such that C is contained in the CNF of F. Let S' be the set of these formulas. □

3.12 General Resolution

Propositional (ground) resolution:

- refutationally complete,
- in its most naive version: not guaranteed to terminate for satisfiable sets of clauses, (improved versions do terminate, however)
- inferior to the CDCL procedure.

But: in contrast to the CDCL procedure, resolution can be easily extended to non-ground clauses.

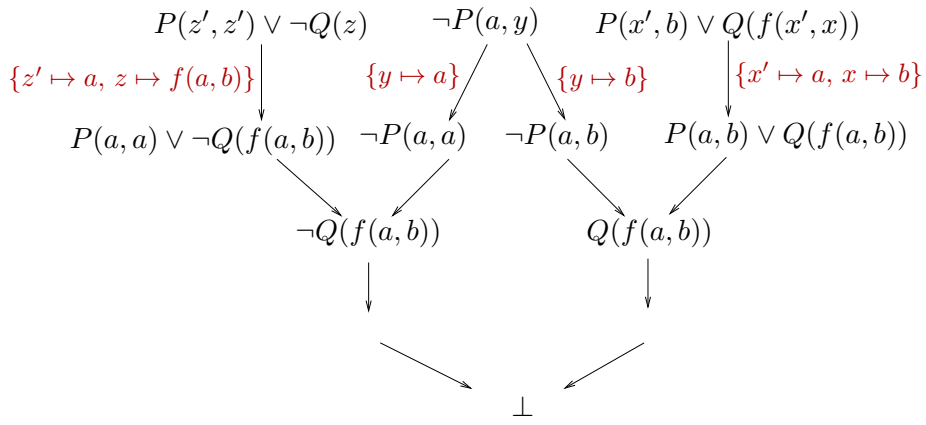
Observation

If $\mathcal{A}$ is a model of an (implicitly universally quantified) clause C, then by Lemma 3.8 it is also a model of all (implicitly universally quantified) instances $C\sigma$ of C.

Consequently, if we show that some instances of clauses in a set N are unsatisfiable, then we have also shown that N itself is unsatisfiable.

General Resolution through Instantiation

Idea: instantiate clauses appropriately:



Early approaches (Gilmore 1960, Davis and Putnam 1960):

Generate ground instances of clauses.

Try to refute the set of ground instances by resolution.

If no contradiction is found, generate more ground instances.

Problems:

More than one instance of a clause can participate in a proof.

Even worse: There are infinitely many possible instances.

Observation:

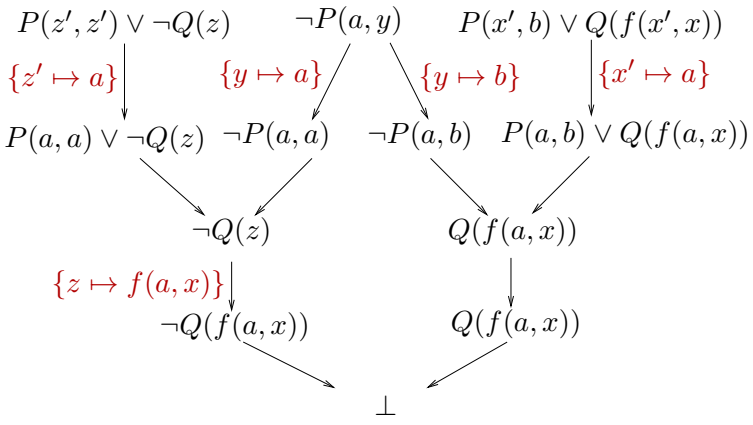
Instantiation must produce complementary literals (so that inferences become possible).

Idea (Robinson 1965):

Do not instantiate more than necessary to get complementary literals
⇒ most general unifiers (mgu).

Calculus works with non-ground clauses; inferences with non-ground clauses represent infinite sets of ground inferences which are computed simultaneously
⇒ lifting principle.

Computation of instances becomes a by-product of boolean reasoning.



Unification

Let $E = \{s_1 \doteq t_1, \ldots, s_n \doteq t_n\}$ (s_i, t_i terms or atoms) be a multiset of *equality problems*. A substitution σ is called a *unifier* of E if $s_i\sigma = t_i\sigma$ for all $1 \leq i \leq n$.

If a unifier of E exists, then E is called *unifiable*.

A substitution σ is called *more general* than a substitution τ, denoted by $\sigma \leq \tau$, if there exists a substitution ρ such that $\rho \circ \sigma = \tau$, where $(\rho \circ \sigma)(x) := (x\sigma)\rho$ is the composition of σ and ρ as mappings. (Note that $\rho \circ \sigma$ has a finite domain as required for a substitution.)

If a unifier of E is more general than any other unifier of E, then we speak of a *most general unifier* of E, denoted by $\mathrm{mgu}(E)$.

Proposition 3.27

(i) $\leq$ *is a quasi-ordering on substitutions, and* $\circ$ *is associative.*

(ii) If $\sigma \leq \tau$ *and* $\tau \leq \sigma$ *(we write* $\sigma \sim \tau$ *in this case), then* $x\sigma$ *and* $x\tau$ *are equal up to (bijective) variable renaming, for any* x *in* X*.*

A substitution σ is called *idempotent*, if $\sigma \circ \sigma = \sigma$.

Proposition 3.28 *σ is idempotent if and only if* $\mathrm{dom}(\sigma) \cap \mathrm{codom}(\sigma) = \emptyset$.

Rule-Based Naive Standard Unification

$$
\begin{array}{rcl}
t \doteq t, E & \Rightarrow_{SU} & E \\
f(s_1, \ldots, s_n) \doteq f(t_1, \ldots, t_n), E & \Rightarrow_{SU} & s_1 \doteq t_1, \ldots, s_n \doteq t_n, E \\
f(\ldots) \doteq g(\ldots), E & \Rightarrow_{SU} & \bot \\
 & & \text{if } f \neq g \\
x \doteq t, E & \Rightarrow_{SU} & x \doteq t, E\{x \mapsto t\} \\
 & & \text{if } x \in \mathrm{var}(E), x \notin \mathrm{var}(t) \\
x \doteq t, E & \Rightarrow_{SU} & \bot \\
 & & \text{if } x \neq t, x \in \mathrm{var}(t) \\
t \doteq x, E & \Rightarrow_{SU} & x \doteq t, E \\
 & & \text{if } t \notin X
\end{array}
$$

SU: Main Properties

If $E = \{x_1 \doteq u_1, \ldots, x_k \doteq u_k\}$, with x_i pairwise distinct, $x_i \notin \mathrm{var}(u_j)$, then E is called an (equational problem in) *solved form* representing the solution $\sigma_E = \{x_1 \mapsto u_1, \ldots, x_k \mapsto u_k\}$.

Proposition 3.29 *If* E *is a solved form then* σ_E *is an mgu of* E*.*

Theorem 3.30

1. *If* $E \Rightarrow_{SU} E'$ *then* σ *is a unifier of* E *if and only if* σ *is a unifier of* E'
2. *If* $E \Rightarrow^*_{SU} \bot$ *then* E *is not unifiable.*
3. *If* $E \Rightarrow^*_{SU} E'$ *with* E' *in solved form, then* $\sigma_{E'}$ *is an mgu of* E*.*

Proof. (1) We have to show this for each of the rules. Let's treat the case for the 4th rule here. Suppose σ is a unifier of $x \doteq t$, that is, $x\sigma = t\sigma$. Thus, $\sigma \circ \{x \mapsto t\} = \sigma[x \mapsto t\sigma] = \sigma[x \mapsto x\sigma] = \sigma$. Therefore, for any equation $u \doteq v$ in E: $u\sigma = v\sigma$, if and only if $u\{x \mapsto t\}\sigma = v\{x \mapsto t\}\sigma$. (2) and (3) follow by induction from (1) using Proposition 3.29. $\square$

Main Unification Theorem

Theorem 3.31 *E is unifiable if and only if there is a most general unifier σ of E, such that σ is idempotent and* $\mathrm{dom}(\sigma) \cup \mathrm{codom}(\sigma) \subseteq \mathrm{var}(E)$.

Proof. The right-to-left implication is trivial. For the left-to-right implication we observe the following:

- $\Rightarrow_{SU}$ is terminating. A suitable lexicographic ordering on the multisets E (with $\bot$ minimal) shows this. Compare in this order:
  (1) the number of variables that occur in E below a function or predicate symbol, or on the right-hand side of an equation, or at least twice;
  (2) the multiset of the sizes (numbers of symbols) of all equations in E;
  (3) the number of non-variable left-hand sides of equations in E.
- A system E that is irreducible w.r.t. $\Rightarrow_{SU}$ is either $\bot$ or a solved form.
- Therefore, reducing any E by SU will end (no matter what reduction strategy we apply) in an irreducible E' having the same unifiers as E, and we can read off the mgu (or non-unifiability) of E from E' (Theorem 3.30, Proposition 3.29).
- σ is idempotent because of the substitution in rule 4. $\mathrm{dom}(\sigma) \cup \mathrm{codom}(\sigma) \subseteq \mathrm{var}(E)$, as no new variables are generated.

□

Rule-Based Polynomial Unification

Problem: Using $\Rightarrow_{SU}$, an *exponential growth* of terms is possible.

The following unification algorithm avoids this problem, at least if the final solved form is represented as a DAG.

$$
\begin{array}{rcl}
t \doteq t, E & \Rightarrow_{PU} & E \\
f(s_1, \ldots, s_n) \doteq f(t_1, \ldots, t_n), E & \Rightarrow_{PU} & s_1 \doteq t_1, \ldots, s_n \doteq t_n, E \\
f(\ldots) \doteq g(\ldots), E & \Rightarrow_{PU} & \bot \\
 & & \text{if } f \neq g \\
x \doteq y, E & \Rightarrow_{PU} & x \doteq y, E\{x \mapsto y\} \\
 & & \text{if } x \in \mathrm{var}(E), x \neq y \\
x_1 \doteq t_1, \ldots, x_n \doteq t_n, E & \Rightarrow_{PU} & \bot \\
 & & \text{if there are positions } p_i \text{ with} \\
 & & t_i|_{p_i} = x_{i+1}, t_n|_{p_n} = x_1 \\
 & & \text{and some } p_i \neq \varepsilon
\end{array}
$$

$$
\begin{array}{rcl}
x \doteq t, E & \Rightarrow_{PU} & \bot \\
 & & \text{if } x \neq t, x \in \mathrm{var}(t) \\
t \doteq x, E & \Rightarrow_{PU} & x \doteq t, E \\
 & & \text{if } t \notin X \\
x \doteq t, x \doteq s, E & \Rightarrow_{PU} & x \doteq t, t \doteq s, E \\
 & & \text{if } t, s \notin X \text{ and } |t| \leq |s|
\end{array}
$$

Properties of PU

Theorem 3.32

1. *If $E \Rightarrow_{PU} E'$ then σ is a unifier of E if and only if σ is a unifier of E'*
2. *If $E \Rightarrow^*_{PU} \bot$ then E is not unifiable.*
3. *If $E \Rightarrow^*_{PU} E'$ with E' in solved form, then $\sigma_{E'}$ is an mgu of E.*

Note: The solved form of $\Rightarrow_{PU}$ is different from the solved form obtained from $\Rightarrow_{SU}$. In order to obtain the unifier $\sigma_{E'}$, we have to sort the list of equality problems $x_i \doteq t_i$ in such a way that x_i does not occur in t_j for $j < i$, and then we have to compose the substitutions $\{x_1 \mapsto t_1\} \circ \cdots \circ \{x_k \mapsto t_k\}$.

Resolution for General Clauses

We obtain the resolution inference rules for non-ground clauses from the inference rules for ground clauses by replacing equality by unifiabilty:

General resolution *Res*:

$$
\frac{D \vee B \qquad C \vee \neg A}{(D \vee C)\sigma} \quad \text{if } \sigma = \mathrm{mgu}(A, B) \qquad \text{[resolution]}
$$

$$
\frac{C \vee A \vee B}{(C \vee A)\sigma} \quad \text{if } \sigma = \mathrm{mgu}(A, B) \quad \text{[factorization]}
$$

For inferences with more than one premise, we assume that the variables in the premises are (bijectively) renamed such that they become different to any variable in the other premises. We do not formalize this. Which names one uses for variables is otherwise irrelevant.

Lifting Lemma

Lemma 3.33 *Let C and D be variable-disjoint clauses. If*

$$\begin{array}{cc} D & C \\ \downarrow \theta_1 & \downarrow \theta_2 \\ D\theta_1 & C\theta_2 \\ \hline & C' \end{array} \qquad \textit{[ground resolution]}$$

then there exists a substitution ρ such that

$$\begin{array}{c} \dfrac{D \qquad C}{C''} \\ \downarrow \rho \\ C' = C''\rho \end{array} \qquad \textit{[general resolution]}$$

An analogous lifting lemma holds for factorization.

Saturation of Sets of General Clauses

Corollary 3.34 *Let N be a set of general clauses saturated under Res, i.e., $Res(N) \subseteq N$. Then also $G_\Sigma(N)$ is saturated, that is,*

$$Res(G_\Sigma(N)) \subseteq G_\Sigma(N).$$

Proof. W.l.o.g. we may assume that clauses in N are pairwise variable-disjoint. (Otherwise make them disjoint, and this renaming process changes neither $Res(N)$ nor $G_\Sigma(N)$.)

Let $C' \in Res(G_\Sigma(N))$. Then either (i) there exist resolvable ground instances $D\theta_1$ and $C\theta_2$ of N with resolvent C', or else (ii) C' is a factor of a ground instance $C\theta$ of C.

Case (i): By the Lifting Lemma, D and C are resolvable with a resolvent C'' with $C''\rho = C'$, for a suitable substitution ρ. As $C'' \in N$ by assumption, we obtain that $C' \in G_\Sigma(N)$.

Case (ii): Similar. □

Soundness for General Clauses

Proposition 3.35 *The general resolution calculus is sound.*

Proof. We have to show that, if $\sigma = \mathrm{mgu}(A, B)$ then $\{\forall \vec{x}\ (D \vee B), \forall \vec{y}\ (C \vee \neg A)\} \models \forall \vec{z}\ (D \vee C)\sigma$ and $\{\forall \vec{x}\ (C \vee A \vee B)\} \models \forall \vec{z}\ (C \vee A)\sigma$.

Let $\mathcal{A}$ be a model of $\forall \vec{x}\ (D \vee B)$ and $\forall \vec{y}\ (C \vee \neg A)$. By Lemma 3.8, $\mathcal{A}$ is also a model of $\forall \vec{z}\ (D \vee B)\sigma$ and $\forall \vec{z}\ (C \vee \neg A)\sigma$ and by Lemma 3.7, $\mathcal{A}$ is also a model of $(D \vee B)\sigma$ and $(C \vee \neg A)\sigma$. Let β be an assignment. If $\mathcal{A}(\beta)(B\sigma) = 0$, then $\mathcal{A}(\beta)(D\sigma) = 1$. Otherwise $\mathcal{A}(\beta)(B\sigma) = \mathcal{A}(\beta)(A\sigma) = 1$, hence $\mathcal{A}(\beta)(\neg A\sigma) = 0$ and therefore $\mathcal{A}(\beta)(C\sigma) = 1$. In both cases $\mathcal{A}(\beta)((D \vee C)\sigma) = 1$, so $\mathcal{A} \models (D \vee C)\sigma$ and by Lemma 3.7, $\mathcal{A} \models \forall \vec{z}\ (D \vee C)\sigma$.

The proof for factorization inferences is similar. □

Herbrand's Theorem

Lemma 3.36 *Let N be a set of Σ-clauses, let $\mathcal{A}$ be an interpretation. Then $\mathcal{A} \models N$ implies $\mathcal{A} \models G_\Sigma(N)$.*

Lemma 3.37 *Let N be a set of Σ-clauses, let $\mathcal{A}$ be an* Herbrand *interpretation. Then $\mathcal{A} \models G_\Sigma(N)$ implies $\mathcal{A} \models N$.*

Proof. Let $\mathcal{A}$ be an Herbrand model of $G_\Sigma(N)$. We have to show that $\mathcal{A} \models \forall \vec{x}\ C$ for all clauses $\forall \vec{x}\ C$ in N. This is equivalent to $\mathcal{A} \models C$, which in turn is equivalent to $\mathcal{A}(\beta)(C) = 1$ for all assignments β.

Choose $\beta : X \to U_\mathcal{A}$ arbitrarily. Since $\mathcal{A}$ is an Herbrand interpretation, $\beta(x)$ is a ground term for every variable x, so there is a substitution σ such that $x\sigma = \beta(x)$ for all variables x occurring in C. Now let γ be an arbitrary assignment, then for every variable occurring in C we have $(\gamma \circ \sigma)(x) = \mathcal{A}(\gamma)(x\sigma) = x\sigma = \beta(x)$ and consequently $\mathcal{A}(\beta)(C) = \mathcal{A}(\gamma \circ \sigma)(C) = \mathcal{A}(\gamma)(C\sigma)$. Since $C\sigma \in G_\Sigma(N)$ and $\mathcal{A}$ is an Herbrand model of $G_\Sigma(N)$, we get $\mathcal{A}(\gamma)(C\sigma) = 1$, so $\mathcal{A}$ is a model of C. □

Theorem 3.38 (Herbrand) *A set N of Σ-clauses is satisfiable if and only if it has an Herbrand model over Σ.*

Proof. The "$\Leftarrow$" part is trivial. For the "$\Rightarrow$" part let $N \not\models \bot$. Since resolution is sound, this implies that $\bot \notin \mathit{Res}^*(N)$. Obviously, a ground instance of a clause has the same number of literals as the clause itself, so we can conclude that $\bot \notin G_\Sigma(\mathit{Res}^*(N))$. Since $\mathit{Res}^*(N)$ is saturated, $G_\Sigma(\mathit{Res}^*(N))$ is saturated as well by Cor. 3.34. Now $I_{G_\Sigma(\mathit{Res}^*(N))}$ is an Herbrand interpretation over Σ and by Thm. 3.23 it is a model of $G_\Sigma(\mathit{Res}^*(N))$. By Lemma 3.37, every Herbrand model of $G_\Sigma(\mathit{Res}^*(N))$ is a model of $\mathit{Res}^*(N)$. Now $N \subseteq \mathit{Res}^*(N)$, so $I_{G_\Sigma(\mathit{Res}^*(N))} \models N$. □

Corollary 3.39 *A set N of Σ-clauses is satisfiable if and only if its set of ground instances $G_\Sigma(N)$ is satisfiable.*

Proof. The "$\Rightarrow$" part follows directly from Lemma 3.36. For the "$\Leftarrow$" part assume that $G_\Sigma(N)$ is satisfiable. By Thm. 3.38 $G_\Sigma(N)$ has an Herbrand model. By Lemma 3.37, every Herbrand model of $G_\Sigma(N)$ is a model of N. □

Refutational Completeness of General Resolution

Theorem 3.40 *Let N be a set of general clauses that is saturated w.r.t. Res. Then $N \models \bot$ if and only if $\bot \in N$.*

Proof. The "$\Leftarrow$" part is trivial. For the "$\Rightarrow$" part assume that N is saturated, that is, $Res(N) \subseteq N$. By Corollary 3.34, $G_\Sigma(N)$ is saturated as well, i.e., $Res(G_\Sigma(N)) \subseteq G_\Sigma(N)$. By Cor. 3.39, $N \models \bot$ implies $G_\Sigma(N) \models \bot$. By the refutational completeness of ground resolution, $G_\Sigma(N) \models \bot$ implies $\bot \in G_\Sigma(N)$, so $\bot \in N$. □

3.13 Theoretical Consequences

We get some classical results on properties of first-order logic as easy corollaries.

The Theorem of Löwenheim-Skolem

Theorem 3.41 (Löwenheim–Skolem) *Let Σ be a countable signature and let S be a set of closed Σ-formulas. Then S is satisfiable if and only if S has a model over a countable universe.*

Proof. If both X and Σ are countable, then S can be at most countably infinite. Now generate, maintaining satisfiability, a set N of clauses from S. This extends Σ by at most countably many new Skolem functions to Σ'. As Σ' is countable, so is $T_{\Sigma'}$, the universe of Herbrand-interpretations over Σ'. Now apply Theorem 3.38. □

There exist more refined versions of this theorem. For instance, one can show that, if S has some infinite model, then S has a model with a universe of cardinality κ for every κ that is larger than or equal to the cardinalty of the signature Σ.

Compactness of Predicate Logic

Theorem 3.42 (Compactness Theorem for First-Order Logic) *Let S be a set of closed first-order formulas. S is unsatisfiable $\Leftrightarrow$ some finite subset $S' \subseteq S$ is unsatisfiable.*

Proof. The "$\Leftarrow$" part is trivial. For the "$\Rightarrow$" part let S be unsatisfiable and let N be the set of clauses obtained by Skolemization and CNF transformation of the formulas in S. Clearly $Res^*(N)$ is unsatisfiable. By Theorem 3.40, $\bot \in Res^*(N)$, and therefore $\bot \in Res^n(N)$ for some $n \in \mathbb{N}$. Consequently, $\bot$ has a finite resolution proof B of depth $\leq n$. Choose S' as the subset of formulas in S such that the corresponding clauses contain the assumptions (leaves) of B. □

3.14 Ordered Resolution with Selection

Motivation: Search space for *Res very* large.

Ideas for improvement:

1. In the completeness proof (Model Existence Theorem 3.23) one only needs to resolve and factor maximal atoms
   $\Rightarrow$ if the calculus is restricted to inferences involving maximal atoms, the proof remains correct
   $\Rightarrow$ *ordering restrictions*
2. In the proof, it does not really matter with which negative literal an inference is performed
   $\Rightarrow$ choose a negative literal don't-care-nondeterministically
   $\Rightarrow$ *selection*

Ordering Restrictions

In the completeness proof one only needs to resolve and factor maximal atoms. Therefore the proof remains correct, if we impose ordering restrictions on ground inferences.

(Ground) Ordered Resolution:

$$\frac{D \vee A \qquad C \vee \neg A}{D \vee C}$$

if $A \succ L$ for all L in D and $\neg A \succeq L$ for all L in C.

(Ground) Ordered Factorization:

$$\frac{C \vee A \vee A}{C \vee A}$$

if $A \succeq L$ for all L in C.

Problem: How to extend this to non-ground inferences?

In the completeness proof, we talk about (strictly) maximal literals of *ground* clauses.

In the non-ground calculus, we have to consider those literals that correspond to (strictly) maximal literals of ground instances.

An ordering $\succ$ on atoms (or terms) is called *stable under substitutions*, if $A \succ B$ implies $A\sigma \succ B\sigma$.

Note:

- We can not require that $A \succ B$ if and only if $A\sigma \succ B\sigma$.
- We can not require that $\succ$ is total on non-ground atoms.

Consequence: In the ordering restrictions for non-ground inferences, we have to replace $\succ$ by $\not\preceq$ and $\succeq$ by $\not\prec$.

Ordered Resolution:

$$\frac{D \vee B \qquad C \vee \neg A}{(D \vee C)\sigma}$$

if $\sigma = \mathrm{mgu}(A, B)$ and $B\sigma \not\preceq L\sigma$ for all L in D and $\neg A\sigma \not\prec L\sigma$ for all L in C.

Ordered Factorization:

$$\frac{C \vee A \vee B}{(C \vee A)\sigma}$$

if $\sigma = \mathrm{mgu}(A, B)$ and $A\sigma \not\prec L\sigma$ for all L in C.

Selection Functions

Selection functions can be used to override ordering restrictions for individual clauses.

A *selection function* is a mapping

$$\mathrm{sel} : C \quad \mapsto \quad \text{set of occurrences of } \textit{negative} \text{ literals in } C$$

Example of selection with selected literals indicated as $\boxed{X}$:

$$\boxed{\neg A} \vee \neg A \vee B$$

$$\boxed{\neg B_0} \vee \boxed{\neg B_1} \vee A$$

Intuition:

- If a clause has at least one selected literal, compute only inferences that involve a selected literal.
- If a clause has no selected literals, compute only inferences that involve a maximal literal.

Resolution Calculus $Res^{\succ}_{\text{sel}}$

The resolution calculus $Res^{\succ}_{\text{sel}}$ is parameterized by

- a selection function sel
- and a well-founded ordering $\succ$ on atoms that is total on ground atoms and stable under substitutions.

(Ground) Ordered Resolution with Selection:

$$\frac{D \vee A \qquad C \vee \neg A}{D \vee C}$$

if the following conditions are satisfied:

(i) $A \succ L$ for all L in D;

(ii) nothing is selected in $D \vee A$ by sel;

(iii) $\neg A$ is selected in $C \vee \neg A$, or nothing is selected in $C \vee \neg A$ and $\neg A \succeq L$ for all L in C.

(Ground) Ordered Factorization with Selection:

$$\frac{C \vee A \vee A}{C \vee A}$$

if the following conditions are satisfied:

(i) $A \succeq L$ for all L in C;

(ii) nothing is selected in $C \vee A \vee A$ by sel.

The extension from ground inferences to non-ground inferences is analogous to ordered resolution (replace $\succ$ by $\not\preceq$ and $\succeq$ by $\not\prec$). Again we assume that $\succ$ is stable under substitutions.

Ordered Resolution with Selection:

$$\frac{D \vee B \qquad C \vee \neg A}{(D \vee C)\sigma}$$

if the following conditions are satisfied:

(i) $\sigma = \mathrm{mgu}(A, B)$;

(ii) $B\sigma \not\preceq L\sigma$ for all L in D;

(iii) nothing is selected in $D \vee B$ by sel;

(iv) $\neg A$ is selected in $C \vee \neg A$, or nothing is selected in $C \vee \neg A$ and $\neg A\sigma \not\prec L\sigma$ for all L in C.

Ordered Factorization with Selection:

$$\frac{C \vee A \vee B}{(C \vee A)\sigma}$$

if the following conditions are satisfied:

(i) $\sigma = \mathrm{mgu}(A, B)$;

(ii) $A\sigma \not\prec L\sigma$ for all L in C.

(iii) nothing is selected in $C \vee A \vee B$ by sel.

Lifting Lemma for $Res^{\succ}_{\mathrm{sel}}$

Lemma 3.43 *Let C and D be variable-disjoint clauses. If*

$$\dfrac{\begin{array}{cc} D & C \\ \downarrow \theta_1 & \downarrow \theta_2 \\ D\theta_1 & C\theta_2 \end{array}}{C'} \qquad \textit{[ground inference in } Res^{\succ}_{\mathrm{sel}}\textit{]}$$

and if $\mathrm{sel}(D\theta_1) \simeq \mathrm{sel}(D)$, $\mathrm{sel}(C\theta_2) \simeq \mathrm{sel}(C)$ *(that is, "corresponding" literals are selected), then there exists a substitution ρ such that*

$$\dfrac{\begin{array}{cc} D & C \end{array}}{\begin{array}{c} C'' \\ \downarrow \rho \\ C' = C''\rho \end{array}} \qquad \textit{[inference in } Res^{\succ}_{\mathrm{sel}}\textit{]}$$

An analogous lifting lemma holds for factorization.

Saturation of Sets of General Clauses

Corollary 3.44 *Let N be a set of general clauses saturated under $Res^{\succ}_{\mathrm{sel}}$, i.e., $Res^{\succ}_{\mathrm{sel}}(N) \subseteq N$. Then there exists a selection function* sel' *such that* $\mathrm{sel}|_N = \mathrm{sel}'|_N$ *and $G_\Sigma(N)$ is also saturated, i.e.,*

$$Res^{\succ}_{\mathrm{sel}'}(G_\Sigma(N)) \subseteq G_\Sigma(N).$$

Proof. We first define the selection function sel' such that $\mathrm{sel}'(C) = \mathrm{sel}(C)$ for all clauses $C \in G_\Sigma(N) \cap N$. For $C \in G_\Sigma(N) \setminus N$ we choose a fixed but arbitrary clause $D \in N$ with $C \in G_\Sigma(D)$ and define $\mathrm{sel}'(C)$ to be those occurrences of literals that are ground instances of the occurrences selected by sel in D. Then proceed as in the proof of Cor. 3.34 using the lifting lemma above. □

Soundness and Refutational Completeness

Theorem 3.45 *Let $\succ$ be an atom ordering and* sel *a selection function such that $Res^{\succ}_{\text{sel}}(N) \subseteq N$. Then*

$$N \models \bot \Leftrightarrow \bot \in N$$

Proof. The "$\Leftarrow$" part is trivial. For the "$\Rightarrow$" part consider first the propositional level: Construct a candidate interpretation I_N as for unrestricted resolution, except that clauses C in N that have selected literals are never productive (even if they are false in I_C and if their maximal atom occurs only once and is positive). The result for general clauses follows using Corollary 3.44. □

What Do We Gain?

Search spaces become smaller:

1	$P \vee Q$	
2	$P \vee \boxed{\neg Q}$	
3	$\neg P \vee Q$	
4	$\neg P \vee \boxed{\neg Q}$	
5	$Q \vee Q$	Res 1, 3
6	Q	Fact 5
7	$\neg P$	Res 6, 4
8	P	Res 6, 2
9	$\bot$	Res 8, 7

we assume $P \succ Q$ and sel as indicated by $\boxed{X}$. The maximal literal in a clause is depicted in red.

In this example, the ordering and selection function even ensure that the refutation proceeds strictly deterministically.

Rotation redundancy can be avoided:

From

$$\dfrac{\dfrac{C_1 \vee A \quad C_2 \vee \neg A \vee B}{C_1 \vee C_2 \vee B} \qquad C_3 \vee \neg B}{C_1 \vee C_2 \vee C_3}$$

we can obtain by *rotation*

$$\dfrac{C_1 \vee A \qquad \dfrac{C_2 \vee \neg A \vee B \quad C_3 \vee \neg B}{C_2 \vee \neg A \vee C_3}}{C_1 \vee C_2 \vee C_3}$$

another proof of the same clause. In large proofs many rotations are possible. However, if $A \succ B$, then the second proof does not fulfill the ordering restrictions.

Craig-Interpolation

Theorem 3.46 (Craig 1957) *Let F and G be two propositional formulas such that $F \models G$. Then there exists a formula H (called the* interpolant *for $F \models G$), such that H contains only propositional variables occurring both in F and in G, and such that $F \models H$ and $H \models G$.*

Proof. Let Π_F, Π_G, and Π_{FG} be the sets of propositional variables that occur only in F, only in G, or both in F and G. Translate F and $\neg G$ into CNF; let N and M, respectively, denote the resulting clause set. Choose an atom ordering $\succ$ for which the propositional variables in Π_F are larger than those in $\Pi_{FG} \cup \Pi_G$. Saturate N into N' w.r.t. $Res^{\succ}_{\text{sel}}$ with an empty selection function sel. Then saturate $N' \cup M$ w.r.t. $Res^{\succ}_{\text{sel}}$ to derive $\bot$. As N' is already saturated, due to the ordering restrictions only inferences need to be considered where premises, if they are from N', only contain symbols from Π_{FG}. The conjunction of these premises is an interpolant H. □

The theorem also holds for first-order formulas, but in the general case, a proof based on resolution technology is complicated because of Skolemization.

3.15 Redundancy

So far: local restrictions of the resolution inference rules using orderings and selection functions.

Is it also possible to delete clauses altogether? Under which circumstances are clauses unnecessary? (e.g., if they are tautologies)

Intuition: If a clause is guaranteed to be neither a minimal counterexample nor productive, then we do not need it.

A Formal Notion of Redundancy[3]

Let N be a set of ground clauses and C a ground clause (not necessarily in N). C is called *redundant* w.r.t. N, if there exist $C_1, \ldots, C_n \in N$, $n \geq 0$, such that $C_i \prec C$ and $C_1, \ldots, C_n \models C$.

Redundancy for general clauses: C is called *redundant* w.r.t. N, if all ground instances $C\sigma$ of C are redundant w.r.t. $G_\Sigma(N)$.

Intuition: If a ground clause C is redundant and all clauses smaller than C hold in I_C, then C holds in I_C (so C is neither a minimal counterexample nor productive).

[3]The notations in this subsection differ significantly from the 2021/2022 lecture. Keep that in mind when you use online lecture recordings or read exercises or exam questions from previous years.

Notation: The set of all clauses that are redundant w.r.t. N is denoted by $Red(N)$.

Note: The same ordering $\succ$ is used for ordering restrictions and for redundancy (and for the completeness proof).

Examples of Redundancy

In general, redundancy is undecidable. Decidable approximations are sufficient for us, however.

Proposition 3.47 *Some redundancy criteria:*

- *C tautology (i.e., $\models C$) $\Rightarrow$ C redundant w.r.t. any set N.*
- *$C\sigma \subset D \Rightarrow$ D redundant w.r.t. $N \cup \{C\}$.*

(Under certain conditions one may also use non-strict subsumption, but this requires a slightly more complicated definition of redundancy.)

Saturation up to Redundancy

N is called *saturated up to redundancy* (w.r.t. $Res^{\succ}_{\mathrm{sel}}$) if

$$Res^{\succ}_{\mathrm{sel}}(N \setminus Red(N)) \subseteq N \cup Red(N)$$

Theorem 3.48 *Let N be saturated up to redundancy. Then*

$$N \models \bot \Leftrightarrow \bot \in N$$

Proof (Sketch).
(i) Ground case: Consider the construction of the candidate interpretation $I^{\succ}_N$ for $Res^{\succ}_{\mathrm{sel}}$.

If a clause $C \in N$ is redundant, then there exist $C_1, \ldots, C_n \in N$, $n \geq 0$, such that $C_i \prec C$ and $C_1, \ldots, C_n \models C$.

If $I_C \models C_i$ by minimality, then $I_C \models C$.

In particular, C is not productive.

$\Rightarrow$ Redundant clauses are not used as premises for "essential" inferences.

By saturation, the conclusion $D' \vee C'$ of a resolution inference is contained in N (as before) or in $Red(N)$. In the first case, minimality of C ensures that $D' \vee C'$ is productive or $I_{D' \vee C'} \models D' \vee C'$; in the second case, it ensures that $I_{D' \vee C'} \models D' \vee C'$. So in both cases we get a contradiction (analogously for factorization). The rest of the proof works as before.

(ii) Lifting: no additional problems over the proof of Theorem 3.45. $\square$

Monotonicity Properties of Redundancy

When we want to delete redundant clauses during a derivation, we have to ensure that redundant clauses *remain redundant* in the rest of the derivation.

Theorem 3.49

(i) $N \subseteq M \Rightarrow Red(N) \subseteq Red(M)$

(ii) $M \subseteq Red(N) \Rightarrow Red(N) \subseteq Red(N \setminus M)$

Proof. (i) Obvious.

(ii) For ground clause sets N, the well-foundedness of the multiset extension of the clause ordering implies that every clause in $Red(N)$ is entailed by smaller clauses in N that are themselves not in $Red(N)$.

For general clause sets N, the result follows from the fact that every clause in $G_\Sigma(N) \setminus Red(G_\Sigma(N))$ is an instance of a clause in $N \setminus Red(N)$. □

Recall that $Red(N)$ may include clauses that are not in N.

Computing Saturated Sets

Redundancy is preserved when, during a theorem proving derivation one adds new clauses or deletes redundant clauses. This motivates the following definitions:

A *run* of the resolution calculus is a sequence $N_0 \vdash N_1 \vdash N_2 \vdash \ldots$, such that
(i) $N_i \models N_{i+1}$, and
(ii) all clauses in $N_i \setminus N_{i+1}$ are redundant w.r.t. N_{i+1}.

In other words, during a run we may add a new clause if it follows from the old ones, and we may delete a clause, if it is redundant w.r.t. the remaining ones.

For a run, we define $N_\infty = \bigcup_{i\geq 0} \bigcap_{j\geq i} N_j$.The set N_∞ of all *persistent* clauses is called the *limit* of the run.

Lemma 3.50 *Let* $N_0 \vdash N_1 \vdash N_2 \vdash \ldots$ *be a run. Then* $Red(N_i) \subseteq Red(\bigcup_{i\geq 0} N_i)$ *and* $Red(N_i) \subseteq Red(N_\infty)$ *for every* i.

Proof. Exercise. □

Corollary 3.51 $N_i \subseteq N_\infty \cup Red(N_\infty)$ *for every* i.

Proof. If $C \in N_i \setminus N_\infty$, then there is a $k \geq i$ such that $C \in N_k \setminus N_{k+1}$, so C must be redundant w.r.t. N_{k+1}. Consequently, C is redundant w.r.t. N_∞. □

Even if a set N is inconsistent, it could happen that $\bot$ is never derived, because some required inference is never computed.

The following definition rules out such runs:

A run is called *fair*, if the conclusion of every inference from clauses in $N_\infty \setminus \mathit{Red}(N_\infty)$ is contained in some $N_i \cup \mathit{Red}(N_i)$.

Lemma 3.52 *If a run is fair, then its limit is saturated up to redundancy.*

Proof. If the run is fair, then the conclusion of every inference from non-redundant clauses in N_∞ is contained in some $N_i \cup \mathit{Red}(N_i)$, and therefore contained in $N_\infty \cup \mathit{Red}(N_\infty)$. Hence N_∞ is saturated up to redundancy. □

Theorem 3.53 (Refutational Completeness: Dynamic View) *Let $N_0 \vdash N_1 \vdash N_2 \vdash \ldots$ be a fair run, let N_∞ be its limit. Then N_0 has a model if and only if $\bot \notin N_\infty$.*

Proof. ($\Leftarrow$): By fairness, N_∞ is saturated up to redundancy. If $\bot \notin N_\infty$, then it has an Herbrand model. Since every clause in N_0 is contained in N_∞ or redundant w.r.t. N_∞, this model is also a model of $G_\Sigma(N_0)$ and therefore a model of N_0.

($\Rightarrow$): Obvious, since $N_0 \models N_\infty$. □

Simplifications

In theory, the definition of a run permits to add arbitrary clauses that are entailed by the current ones.

In practice, we restrict to two cases:

- We add conclusions of $\mathit{Res}^{\succ}_{\mathrm{sel}}$-inferences from non-redundant premises.
  $\rightsquigarrow$ necessary to guarantee fairness
- We add clauses that are entailed by the current ones if this *makes* other clauses redundant:

  $$N \cup \{C\} \;\vdash\; N \cup \{C, D\} \;\vdash\; N \cup \{D\}$$
  $$\text{if } N \cup \{C\} \models D \text{ and } C \in \mathit{Red}(N \cup \{D\}).$$

  Net effect: C is *simplified* to D
  $\rightsquigarrow$ useful to get easier/smaller clause sets

Notation for simplification rules:

$$\frac{C_1 \ \ldots \ C_n}{D_1 \ \ldots \ D_m}$$

means

$$N \cup \{C_1, \ldots, C_n\} \ \vdash \ N \cup \{D_1, \ldots, D_m\}$$

Examples of simplification techniques:

- Deletion of duplicated literals:

$$\frac{C \vee L \vee L}{C \vee L}$$

- Subsumption resolution:

$$\frac{D \vee L \qquad C \vee D\sigma \vee \overline{L}\sigma}{D \vee L \qquad C \vee D\sigma}$$

3.16 Hyperresolution

There are *many* variants of resolution.

One well-known example is hyperresolution (Robinson 1965):

> Assume that several negative literals are selected in a clause C. If we perform an inference with C, then one of the selected literals is eliminated.
>
> Suppose that the remaining selected literals of C are again selected in the conclusion. Then we must eliminate the remaining selected literals one by one by further resolution steps.

Hyperresolution replaces these successive steps by a single inference. As for $Res^{\succ}_{\mathrm{sel}}$, the calculus is parameterized by an atom ordering $\succ$ and a selection function sel.

$$\frac{D_1 \vee B_1 \quad \ldots \quad D_n \vee B_n \qquad C \vee \neg A_1 \vee \ldots \vee \neg A_n}{(D_1 \vee \ldots \vee D_n \vee C)\sigma}$$

with $\sigma = \mathrm{mgu}(A_1 \doteq B_1, \ldots, A_n \doteq B_n)$, if

(i) $B_i\sigma$ strictly maximal in $D_i\sigma$, $1 \leq i \leq n$;

(ii) nothing is selected in D_i;

(iii) the indicated occurrences of the $\neg A_i$ are exactly the ones selected by sel, or nothing is selected in the right premise and $n = 1$ and $\neg A_1\sigma$ is maximal in $C\sigma$.

Similarly to resolution, hyperresolution has to be complemented by a factorization inference.

As we have seen, hyperresolution can be simulated by iterated binary resolution.

However this yields intermediate clauses which HR might not derive, and many of them might not be extendable into a full HR inference.

3.17 Summary: Resolution Theorem Proving

- Resolution is a machine-oriented calculus.
- Using unification, the enumeration of instances becomes a by-product of inference computation.
- Parameters: atom ordering $\succ$ and selection function sel. On the non-ground level, ordering constraints can (only) be solved approximatively.
- Completeness proof by constructing candidate interpretations from productive clauses $C \vee A$, $A \succ C$.
- *Local* restrictions of inferences via $\succ$ and sel
  $\Rightarrow$ fewer proof variants.
- *Global* restrictions of the search space via redundancy
  $\Rightarrow$ computing with "smaller"/"easier" clause sets.
  (In practice: simplification and detection of redundant clauses uses 90% of the prover runtime.)
- Termination on many decidable fragments.
- However: not good enough for dealing with orderings, equality and more specific algebraic theories (lattices, abelian groups, rings, fields)
  $\Rightarrow$ further specialization of inference systems required.

3.18 Semantic Tableaux

Literature:

M. Fitting: First-Order Logic and Automated Theorem Proving, Springer-Verlag, New York, 1996, chapters 3, 6, 7.

R. M. Smullyan: First-Order Logic, Dover Publ., New York, 1968, revised 1995.

Like resolution, semantic tableaux were developed in the sixties, independently by Zbigniew Lis and Raymond Smullyan on the basis of work by Gentzen in the 30s and of Beth in the 50s.

Idea

Idea (for the propositional case):

A set $\{F \wedge G\} \cup N$ of formulas has a model if and only if $\{F \wedge G, F, G\} \cup N$ has a model.

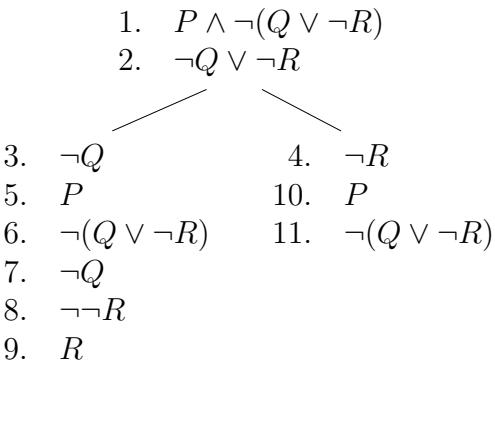
A set $\{F \vee G\} \cup N$ of formulas has a model if and only if $\{F \vee G, F\} \cup N$ or $\{F \vee G, G\} \cup N$ has a model.

(and similarly for other connectives).

To avoid duplication, represent sets as paths of a tree.

Continue splitting until two complementary formulas are found $\Rightarrow$ inconsistency detected.

A Tableau for $\{P \wedge \neg(Q \vee \neg R),\ \neg Q \vee \neg R\}$



This tableau is not "maximal", however the first "path" is. This path is not "closed", hence the set $\{1, 2\}$ is satisfiable. (These notions will all be defined below.)

Properties

Properties of tableau calculi:

analytic: inferences correspond closely to the logical meaning of the symbols.

goal oriented: inferences operate directly on the goal to be proved.

global: some inferences affect the entire proof state (set of formulas), as we will see later.

Propositional Expansion Rules

Expansion rules are applied to the formulas in a tableau and expand the tableau at a leaf. We append the conclusions of a rule (horizontally or vertically) at a *leaf*, whenever the premise of the expansion rule matches a formula appearing *anywhere* on the path from the root to that leaf.

Negation Elimination

$$\frac{\neg\neg F}{F} \qquad \frac{\neg\top}{\bot} \qquad \frac{\neg\bot}{\top}$$

α-Expansion

(for formulas that are essentially conjunctions: append subformulas α_1 and α_2 one on top of the other)

$$\frac{\alpha}{\begin{array}{l}\alpha_1\\ \alpha_2\end{array}}$$

β-Expansion

(for formulas that are essentially disjunctions:
append β_1 and β_2 horizontally, i.e., branch into β_1 and β_2)

$$\frac{\beta}{\beta_1 \mid \beta_2}$$

Classification of Formulas

conjunctive			disjunctive		
α	α_1	α_2	β	β_1	β_2
$F \wedge G$	F	G	$\neg(F \wedge G)$	$\neg F$	$\neg G$
$\neg(F \vee G)$	$\neg F$	$\neg G$	$F \vee G$	F	G
$\neg(F \to G)$	F	$\neg G$	$F \to G$	$\neg F$	G

We assume that the binary connective $\leftrightarrow$ has been eliminated in advance.

Tableaux: Notions

A *semantic tableau* is a marked (by formulas), finite, unordered tree and inductively defined as follows: Let $\{F_1, \ldots, F_n\}$ be a set of formulas.

(i) The tree consisting of a single path

$$\begin{array}{l} F_1 \\ \vdots \\ F_n \end{array}$$

is a tableau for $\{F_1, \ldots, F_n\}$. (We do not draw edges if nodes have only one successor.)

(ii) If T is a tableau for $\{F_1, \ldots, F_n\}$ and if T' results from T by applying an expansion rule then T' is also a tableau for $\{F_1, \ldots, F_n\}$.

Note: We may also consider the *limit tableau* of a tableau expansion; this can be an *infinite* tree.

A *path* (from the root to a leaf) in a tableau is called *closed*, if it either contains $\bot$, or else it contains both some formula F and its negation $\neg F$. Otherwise the path is called *open*.

A tableau is called *closed*, if all paths are closed.

A *tableau proof* for F is a closed tableau for $\{\neg F\}$.

A path π in a tableau is called *maximal*, if for each formula F on π that is neither a literal nor $\bot$ nor $\top$ there exists a node in π at which the expansion rule for F has been applied.

In that case, if F is a formula on π, π also contains:

(i) α_1 and α_2, if F is a α-formula,

(ii) β_1 or β_2, if F is a β-formula, and

(iii) F', if F is a negation formula, and F' the conclusion of the corresponding elimination rule.

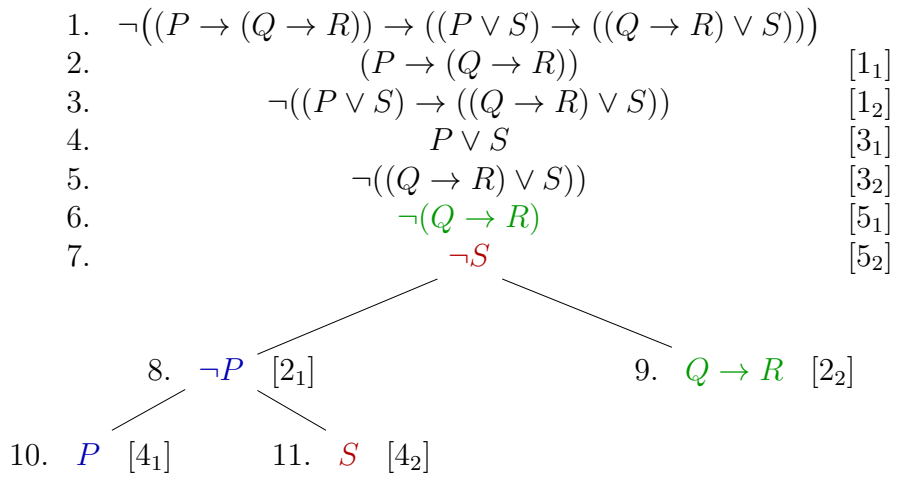
A tableau is called *maximal*, if each path is closed or maximal.

A tableau is called *strict*, if for each formula the corresponding expansion rule has been applied at most once on each path containing that formula.

A tableau is called *clausal*, if each of its formulas is a clause.

A Sample Proof

One starts out from the negation of the formula to be proved.



There are three paths, each of them closed.

Properties of Propositional Tableaux

We assume that T is a tableau for $\{F_1, \ldots, F_n\}$.

Theorem 3.54 *$\{F_1, \ldots, F_n\}$ satisfiable $\Leftrightarrow$ some path (i.e., the set of its formulas) in T is satisfiable.*

Proof. ($\Leftarrow$) Trivial, since every path contains in particular $F_1, \ldots, F_n$.
($\Rightarrow$) By induction over the structure of T. □

Corollary 3.55 *T closed $\Rightarrow$ $\{F_1, \ldots, F_n\}$ unsatisfiable*

Theorem 3.56 *Every strict propositional tableau expansion is finite.*

Proof. New formulas resulting from expansion are either $\bot$, $\top$ or subformulas of the expanded formula (modulo de Morgan's law), so the number of formulas that can occur is finite. By strictness, on each path a formula can be expanded at most once. Therefore, each path is finite, and a finitely branching tree with finite paths is finite by Lemma 1.9. □

Conclusion: Strict and maximal tableaux can be effectively constructed.

Refutational Completeness

A set $\mathcal{H}$ of propositional formulas is called a Hintikka set, if

(1) there is no $P \in \Pi$ with $P \in \mathcal{H}$ and $\neg P \in \mathcal{H}$;

(2) $\bot \notin \mathcal{H}$, $\neg\top \notin \mathcal{H}$;

(3) if $\neg\neg F \in \mathcal{H}$, then $F \in \mathcal{H}$;

(4) if $\alpha \in \mathcal{H}$, then $\alpha_1 \in \mathcal{H}$ and $\alpha_2 \in \mathcal{H}$;

(5) if $\beta \in \mathcal{H}$, then $\beta_1 \in \mathcal{H}$ or $\beta_2 \in \mathcal{H}$.

Lemma 3.57 (Hintikka's Lemma) *Every Hintikka set is satisfiable.*

Proof. Let $\mathcal{H}$ be a Hintikka set. Define a valuation $\mathcal{A}$ by $\mathcal{A}(P) = 1$ if $P \in \mathcal{H}$ and $\mathcal{A}(P) = 0$ otherwise. Then show that $\mathcal{A}(F) = 1$ for all $F \in \mathcal{H}$ by induction over the size of formulas. □

Theorem 3.58 *Let π be a maximal open path in a tableau. Then the set of formulas on π is satisfiable.*

Proof. We show that set of formulas on π is a Hintikka set: Conditions (3), (4), (5) follow from the fact that π is maximal; conditions (1) and (2) follow from the fact that π is open and from maximality for the second negation elimination rule. □

Note: The theorem holds also for infinite trees that are obtained as the limit of a tableau expansion.

Theorem 3.59 $\{F_1, \ldots, F_n\}$ *satisfiable* $\Leftrightarrow$ *there exists no closed strict tableau for* $\{F_1, \ldots, F_n\}$.

Proof. ($\Rightarrow$) Clear by Cor. 3.55.
($\Leftarrow$) Let T be a strict maximal tableau for $\{F_1, \ldots, F_n\}$ and let π be an open path in T. By the previous theorem, the set of formulas on π is satisfiable, and hence by Theorem 3.54 the set $\{F_1, \ldots, F_n\}$, is satisfiable. □

Consequences

The validity of a propositional formula F can be established by constructing a strict maximal tableau for $\{\neg F\}$:

- T closed $\Leftrightarrow$ F valid.
- It suffices to test complementarity of paths w.r.t. atomic formulas (cf. reasoning in the proof of Theorem 3.58).
- Which of the potentially many strict maximal tableaux one computes does not matter. In other words, tableau expansion rules can be applied don't-care non-deterministically ("*proof confluence*").
- The expansion strategy, however, can have a dramatic impact on the tableau size.

A Variant of the β-Rule

Since $F \vee G \models\!\!\!\models F \vee (G \wedge \neg F)$, the β expansion rule

$$\frac{\beta}{\beta_1 \mid \beta_2}$$

can be replaced by the following variant:

$$\frac{\beta}{\begin{array}{l|l}\beta_1 & \beta_2 \\ & \neg\beta_1\end{array}}$$

The variant β-rule can lead to much shorter proofs, but it is not always beneficial.

In general, it is most helpful if $\neg\beta_1$ can be at most (iteratively) α-expanded.

3.19 Semantic Tableaux for First-Order Logic

There are two ways to extend the tableau calculus to quantified formulas:

- using ground instantiation,
- using free variables.

Tableaux with Ground Instantiation

Classification of quantified formulas:

universal		existential	
γ	$\gamma(t)$	δ	$\delta(t)$
$\forall x F$	$F\{x \mapsto t\}$	$\exists x F$	$F\{x \mapsto t\}$
$\neg\exists x F$	$\neg F\{x \mapsto t\}$	$\neg\forall x F$	$\neg F\{x \mapsto t\}$

Idea:

Replace universally quantified formulas by appropriate ground instances.

γ-expansion

$$\frac{\gamma}{\gamma(t)} \quad \text{where } t \text{ is some ground term}$$

δ-expansion

$$\frac{\delta}{\delta(c)} \quad \text{where } c \text{ is a new Skolem constant}$$

Skolemization becomes part of the calculus and needs not necessarily be applied in a preprocessing step. Of course, one could do Skolemization beforehand, and then the δ-rule would not be needed.

Note:

Skolem *constants* are sufficient:
In a δ-formula $\exists x\, F$, $\exists$ is the outermost quantifier and x is the only free variable in F.

Problems:

Having to guess ground terms is impractical.

Even worse, we may have to guess *several* ground instances, as strictness for γ is incomplete. For instance, constructing a closed tableau for

$$\{\forall x\,(P(x) \to P(f(x))),\ P(b),\ \neg P(f(f(b)))\}$$

is impossible without applying γ-expansion twice on one path.

Free-Variable Tableaux

An alternative approach:

Delay the instantiation of universally quantified variables.

Replace universally quantified variables by new free variables.

Intuitively, the free variables are universally quantified *outside* of the entire tableau.

γ-expansion

$$\frac{\gamma}{\gamma(x)} \quad \text{where } x \text{ is a new free variable}$$

δ-expansion

$$\frac{\delta}{\delta(f(x_1, \ldots, x_n))}$$

where f is a new Skolem function, and the x_i are the free variables in δ

Application of expansion rules has to be supplemented by a *substitution rule:*

(iii) If T is a tableau for $\{F_1, \ldots, F_n\}$ and if σ is a substitution, then $T\sigma$ is also a tableau for $\{F_1, \ldots, F_n\}$.

The substitution rule may, potentially, modify all the formulas of a tableau. This feature is what makes the tableau method a *global proof method.* (Resolution, by comparison, is a local method.)

One can show that it is sufficient to consider substitutions σ for which there is a path in T containing two *literals* $\neg A$ and B such that $\sigma = \text{mgu}(A, B)$. Such tableaux are called *AMGU-Tableaux*.

Example

1.	$\neg\big(\exists w \forall x\ P(x, w, f(x, w)) \to \exists w \forall x \exists y\ P(x, w, y)\big)$	
2.	$\exists w \forall x\ P(x, w, f(x, w))$	1_1 $[\alpha]$
3.	$\neg\exists w \forall x \exists y\ P(x, w, y)$	1_2 $[\alpha]$
4.	$\forall x\ P(x, c, f(x, c))$	$2(c)$ $[\delta]$
5.	$\neg\forall x \exists y\ P(x, v_1, y)$	$3(v_1)$ $[\gamma]$
6.	$\neg\exists y\ P(b(v_1), v_1, y)$	$5(b(v_1))$ $[\delta]$
7.	$P(v_2, c, f(v_2, c))$	$4(v_2)$ $[\gamma]$
8.	$\neg P(b(v_1), v_1, v_3)$	$6(v_3)$ $[\gamma]$

7. and 8. are complementary (modulo unification):

$$\{v_2 \doteq b(v_1),\ c \doteq v_1,\ f(v_2, c) \doteq v_3\}$$

is solvable with an mgu $\sigma = \{v_1 \mapsto c,\ v_2 \mapsto b(c),\ v_3 \mapsto f(b(c), c)\}$, and hence, $T\sigma$ is a closed (linear) tableau for the formula in 1.

Problem:

Strictness for γ is still incomplete. For instance, constructing a closed tableau for

$$\{\forall x\,(P(x) \to P(f(x)))),\ P(b),\ \neg P(f(f(b)))\}$$

is impossible without applying γ-expansion twice on one path.

Semantic Tableaux vs. Resolution

- Tableaux: global, goal-oriented, "backward".
- Resolution: local, "forward".
- Goal-orientation is a clear advantage if only a small subset of a large set of formulas is necessary for a proof. (Note that resolution provers saturate also those parts of the clause set that are irrelevant for proving the goal.)
- Resolution can be combined with more powerful redundancy elimination methods; because of its global nature this is more difficult for the tableau method.
- Resolution can be refined to work well with equality; for tableaux this seems to be impossible.
- On the other hand tableau calculi can be easily extended to other logics; in particular tableau provers are very successful in modal and description logics.

3.20 Other Deductive Systems

- Instantiation-based methods
  Resolution-based instance generation
  Disconnection calculus
  ...
- Natural deduction
- Sequent calculus/Gentzen calculus
- Hilbert calculus

Instantiation-Based Methods for FOL

Idea:

Overlaps of complementary literals produce instantiations (as in resolution);

However, contrary to resolution, clauses are not recombined.

Instead: treat remaining variables as constant and use efficient propositional proof methods, such as CDCL.

There are both saturation-based variants, such as partial instantiation (Hooker et al. 2002) or resolution-based instance generation (Inst-Gen) (Ganzinger and Korovin 2003), and tableau-style variants, such as the disconnection calculus (Billon 1996; Letz and Stenz 2001).

Successful in practice for problems that are "almost propositional" (i.e., no non-constant function symbols, no equality).

Natural Deduction

Idea:

Model the concept of proofs from assumptions as humans do it.

To prove $F \to G$, assume F and try to derive G.

Initial ideas: Jaśkowski (1934), Gentzen (1934); extended by Prawitz (1965).

Popular in interactive proof systems.

Sequent Calculus

Idea:

Assumptions internalized into the data structure of sequents

$$F_1, \ldots, F_m \vdash G_1, \ldots, G_k$$

meaning

$$F_1 \wedge \cdots \wedge F_m \rightarrow G_1 \vee \cdots \vee G_k$$

Inferences rules, e.g.:

$$\frac{\Gamma \vdash \Delta}{\Gamma, F \vdash \Delta}\ (WL) \qquad \frac{\Gamma, F \vdash \Delta \qquad \Sigma, G \vdash \Pi}{\Gamma, \Sigma, F \vee G \vdash \Delta, \Pi}\ (\vee L)$$

$$\frac{\Gamma \vdash \Delta}{\Gamma \vdash F, \Delta}\ (WR) \qquad \frac{\Gamma \vdash F, \Delta \qquad \Sigma \vdash G, \Pi}{\Gamma, \Sigma \vdash F \wedge G, \Delta, \Pi}\ (\wedge R)$$

Initial idea: Gentzen 1934.

Perfect symmetry between the handling of assumptions and their consequences; interesting for proof theory.

Can be used both backwards and forwards.

Allows to simulate both natural deduction and semantic tableaux.

Hilbert Calculus

Idea:

Direct proof method (proves a theorem from axioms, rather than refuting its negation)

Axiom schemes, e.g.,

$$F \rightarrow (G \rightarrow F)$$
$$(F \rightarrow (G \rightarrow H)) \rightarrow ((F \rightarrow G) \rightarrow (F \rightarrow H))$$

plus Modus ponens:

$$\frac{F \qquad F \rightarrow G}{G}$$

Unsuitable for finding or reading proofs, but sometimes used for *specifying* (e.g., modal) logics.

4 First-Order Logic with Equality

Equality is the most important relation in mathematics and functional programming.

In principle, problems in first-order logic with equality can be handled by any prover for first-order logic without equality:

4.1 Handling Equality Naively

Proposition 4.1 *Let F be a closed first-order formula with equality. Let $\sim \notin \Pi$ be a new predicate symbol. The set $Eq(\Sigma)$ contains the formulas*

$$\begin{gathered}
\forall x\,(x \sim x) \\
\forall x, y\,(x \sim y \to y \sim x) \\
\forall x, y, z\,(x \sim y \wedge y \sim z \to x \sim z) \\
\forall \vec{x}, \vec{y}\,(x_1 \sim y_1 \wedge \cdots \wedge x_n \sim y_n \to f(x_1, \ldots, x_n) \sim f(y_1, \ldots, y_n)) \\
\forall \vec{x}, \vec{y}\,(x_1 \sim y_1 \wedge \cdots \wedge x_m \sim y_m \wedge P(x_1, \ldots, x_m) \to P(y_1, \ldots, y_m))
\end{gathered}$$

for every $f/n \in \Omega$ and $P/m \in \Pi$. Let $\tilde{F}$ be the formula that one obtains from F if every occurrence of $\approx$ is replaced by $\sim$. Then F is satisfiable if and only if $Eq(\Sigma) \cup \{\tilde{F}\}$ is satisfiable.

Proof. Let $\Sigma = (\Omega, \Pi)$, let $\Sigma_1 = (\Omega, \Pi \cup \{\sim/2\})$.

For the "only if" part assume that F is satisfiable and let $\mathcal{A}$ be a Σ-model of F. Then we define a Σ_1-algebra $\mathcal{B}$ in such a way that $\mathcal{B}$ and $\mathcal{A}$ have the same universe, $f_\mathcal{B} = f_\mathcal{A}$ for every $f \in \Omega$, $P_\mathcal{B} = P_\mathcal{A}$ for every $P \in \Pi$, and $\sim_\mathcal{B}$ is the identity relation on the universe. It is easy to check that $\mathcal{B}$ is a model of both $\tilde{F}$ and of $Eq(\Sigma)$.

For the "if" part assume that the Σ_1-algebra $\mathcal{B} = (U_\mathcal{B},\ (f_\mathcal{B} : U_\mathcal{B}^n \to U_\mathcal{B})_{f\in\Omega},\ (P_\mathcal{B} \subseteq U_\mathcal{B}^m)_{P\in\Pi\cup\{\sim\}})$ is a model of $Eq(\Sigma) \cup \{\tilde{F}\}$. Then the interpretation $\sim_\mathcal{B}$ of $\sim$ in $\mathcal{B}$ is a congruence relation on $U_\mathcal{B}$ with respect to the functions $f_\mathcal{B}$ and the predicates $P_\mathcal{B}$.

We will now construct a Σ-algebra $\mathcal{A}$ from $\mathcal{B}$ and the congruence relation $\sim_\mathcal{B}$. Let $[a]$ be the congruence class of an element $a \in U_\mathcal{B}$ with respect to $\sim_\mathcal{B}$. The universe $U_\mathcal{A}$ of $\mathcal{A}$ is the set $\{\,[a] \mid a \in U_\mathcal{B}\,\}$ of congruence classes of the universe of $\mathcal{B}$. For a function symbol $f \in \Omega$, we define $f_\mathcal{A}([a_1], \ldots, [a_n]) = [f_\mathcal{B}(a_1, \ldots, a_n)]$, and for a predicate symbol $P \in \Pi$, we define $([a_1], \ldots, [a_n]) \in P_\mathcal{A}$ if and only if $(a_1, \ldots, a_n) \in P_\mathcal{B}$. Observe that this is well-defined: If we take different representatives of the same congruence class, we get the same result by congruence of $\sim_\mathcal{B}$. For any $\mathcal{A}$-assignment γ choose some $\mathcal{B}$-assignment β such that $\mathcal{B}(\beta)(x) \in \mathcal{A}(\gamma)(x)$ for every x, then for every Σ-term t we have $\mathcal{A}(\gamma)(t) = [\mathcal{B}(\beta)(t)]$, and analogously for every Σ-formula G, $\mathcal{A}(\gamma)(G) = \mathcal{B}(\beta)(\tilde{G})$. Both properties can easily shown by structural induction. Therefore, $\mathcal{A}$ is a model of F. □

An analogous proposition holds for *sets* of closed first-order formulas with equality.

By giving the equality axioms explicitly, first-order problems with equality can in principle be solved by a standard resolution or tableaux prover.

But this is unfortunately not efficient (mainly due to the transitivity and congruence axioms).

Equality is theoretically difficult: First-order functional programming is Turing-complete.

But: resolution theorem provers cannot even solve equational problems that are intuitively easy.

Consequence: to handle equality efficiently, knowledge must be integrated into the theorem prover.

Roadmap

How to proceed:

- This semester: Equations (unit clauses with equality).

  Term rewrite systems.
  Expressing semantic consequence syntactically.
  Knuth-Bendix-Completion.
  Entailment for equations.

- Next semester: Equational clauses.

  Combining resolution and KB-completion. → Superposition.
  Entailment for clauses with equality.

4.2 Rewrite Systems

Let E be a set of (implicitly universally quantified) equations.

The *rewrite relation* $\to_E \subseteq \mathrm{T}_\Sigma(X) \times \mathrm{T}_\Sigma(X)$ is defined by

$$\begin{aligned} s \to_E t \quad \text{if and only if} \quad & \text{there exist } (l \approx r) \in E,\ p \in \mathrm{pos}(s), \\ & \text{and } \sigma : X \to \mathrm{T}_\Sigma(X), \\ & \text{such that } s|_p = l\sigma \text{ and } t = s[r\sigma]_p. \end{aligned}$$

An instance of the lhs (left-hand side) of an equation is called a *redex* (reducible expression). *Contracting* a redex means replacing it with the corresponding instance of the rhs (right-hand side) of the rule.

An equation $l \approx r$ is also called a *rewrite rule*, if l is not a variable and $\mathrm{var}(l) \supseteq \mathrm{var}(r)$.

Notation: $l \to r$.

A set of rewrite rules is called a *term rewrite system (TRS)*.

We say that a set of equations E or a TRS R is terminating, if the rewrite relation $\to_E$ or $\to_R$ has this property.

(Analogously for other properties of abstract reduction systems).

Note: If E is terminating, then it is a TRS.

E-Algebras

Let E be a set of universally quantified equations. A model of E is also called an *E-algebra*.

If $E \models \forall\vec{x}(s \approx t)$, i.e., $\forall\vec{x}(s \approx t)$ is valid in all E-algebras, we write this also as $s \approx_E t$.

Goal:
Use the rewrite relation $\to_E$ to express the semantic consequence relation syntactically:

$s \approx_E t$ if and only if $s \leftrightarrow_E^* t$.

Let E be a set of equations over $\mathrm{T}_\Sigma(X)$. The following inference system allows to derive consequences of E:

$E \vdash t \approx t$ for every $t \in \mathrm{T}_\Sigma(X)$ — *(Reflexivity)*

$$\frac{E \vdash t \approx t'}{E \vdash t' \approx t}$$ — *(Symmetry)*

$$\frac{E \vdash t \approx t' \qquad E \vdash t' \approx t''}{E \vdash t \approx t''}$$ — *(Transitivity)*

$$\frac{E \vdash t_1 \approx t_1' \quad \ldots \quad E \vdash t_n \approx t_n'}{E \vdash f(t_1, \ldots, t_n) \approx f(t_1', \ldots, t_n')}$$ — *(Congruence)*

$E \vdash t\sigma \approx t'\sigma$ if $(t \approx t') \in E$ and $\sigma : X \to \mathrm{T}_\Sigma(X)$ — *(Instance)*

Lemma 4.2 *The following properties are equivalent:*

(i) $s \leftrightarrow_E^* t$

(ii) $E \vdash s \approx t$ *is derivable.*

Proof. (i)⇒(ii): $s \leftrightarrow_E t$ implies $E \vdash s \approx t$ by induction on the depth of the position where the equation is applied; then $s \leftrightarrow_E^* t$ implies $E \vdash s \approx t$ by induction on the number of rewrite steps in $s \leftrightarrow_E^* t$.

(ii)⇒(i): By induction on the size (number of symbols) of the derivation for $E \vdash s \approx t$. □

Constructing a *quotient algebra:*

Let X be a set of variables.

For $t \in \mathrm{T}_\Sigma(X)$ let $[t] = \{\, t' \in \mathrm{T}_\Sigma(X) \mid E \vdash t \approx t' \,\}$ be the *congruence class* of t.

Define a Σ-algebra $\mathrm{T}_\Sigma(X)/E$ (abbreviated by $\mathcal{T}$) as follows:

$U_\mathcal{T} = \{\, [t] \mid t \in \mathrm{T}_\Sigma(X) \,\}$.

$f_\mathcal{T}([t_1], \ldots, [t_n]) = [f(t_1, \ldots, t_n)]$ for $f/n \in \Omega$.

Lemma 4.3 *$f_\mathcal{T}$ is well-defined: If $[t_i] = [t_i']$, then $[f(t_1, \ldots, t_n)] = [f(t_1', \ldots, t_n')]$.*

Proof. Follows directly from the *Congruence* rule for $\vdash$. □

Lemma 4.4 *$\mathcal{T} = \mathrm{T}_\Sigma(X)/E$ is an E-algebra.*

Proof. Let $\forall x_1 \ldots x_n (s \approx t)$ be an equation in E; let β be an arbitrary assignment.

We have to show that $\mathcal{T}(\beta)(\forall \vec{x}(s \approx t)) = 1$, or equivalently, that $\mathcal{T}(\gamma)(s) = \mathcal{T}(\gamma)(t)$ for all $\gamma = \beta[\, x_i \mapsto [v_i] \mid 1 \le i \le n \,]$ with $[v_i] \in U_\mathcal{T}$.

Let $\sigma = \{x_1 \mapsto v_1, \ldots, x_n \mapsto v_n\}$, then we get by structural induction that $u\sigma \in \mathcal{T}(\gamma)(u)$ for every $u \in \mathrm{T}_\Sigma(\{x_1, ..., x_n\})$. In particular, $s\sigma \in \mathcal{T}(\gamma)(s)$ and $t\sigma \in \mathcal{T}(\gamma)(t)$.

By the *Instance* rule, $E \vdash s\sigma \approx t\sigma$ is derivable, hence $\mathcal{T}(\gamma)(s) = [s\sigma] = [t\sigma] = \mathcal{T}(\gamma)(t)$. □

Lemma 4.5 *Let X be a countably infinite set of variables; let $s, t \in \mathrm{T}_\Sigma(Y)$. If $\mathrm{T}_\Sigma(X)/E \models \forall\vec{x}(s \approx t)$, then $E \vdash s \approx t$ is derivable.*

Proof. Without loss of generality, we assume that all variables in $\vec{x}$ are contained in X. (Otherwise, we rename the variables in the equation. Since X is countably infinite, this is always possible.) Assume that $\mathcal{T} \models \forall\vec{x}(s \approx t)$, i.e., $\mathcal{T}(\beta)(\forall\vec{x}(s \approx t)) = 1$. Consequently, $\mathcal{T}(\gamma)(s) = \mathcal{T}(\gamma)(t)$ for all $\gamma = \beta[\, x_i \mapsto [v_i] \mid 1 \leq i \leq n\,]$ with $[v_i] \in U_\mathcal{T}$.

Choose $v_i := x_i$, then by structural induction $[u] = \mathcal{T}(\gamma)(u)$ for every $u \in \mathrm{T}_\Sigma(\{x_1, ..., x_n\})$, so $[s] = \mathcal{T}(\gamma)(s) = \mathcal{T}(\gamma)(t) = [t]$. Therefore $E \vdash s \approx t$ is derivable by definition of $\mathcal{T}$. □

Theorem 4.6 ("Birkhoff's Theorem") *Let X be a countably infinite set of variables, let E be a set of (universally quantified) equations. Then the following properties are equivalent for all $s, t \in \mathrm{T}_\Sigma(X)$:*

(i) $s \leftrightarrow_E^* t$.

(ii) $E \vdash s \approx t$ is derivable.

(iii) $s \approx_E t$, i.e., $E \models \forall\vec{x}(s \approx t)$.

(iv) $\mathrm{T}_\Sigma(X)/E \models \forall\vec{x}(s \approx t)$.

Proof. (i)⇔(ii): Lemma 4.2.

(ii)⇒(iii): By induction on the size of the derivation for $E \vdash s \approx t$.

(iii)⇒(iv): Obvious, since $\mathcal{T} = \mathrm{T}_\Sigma(X)/E$ is an E-algebra.

(iv)⇒(ii): Lemma 4.5. □

Universal Algebra

$\mathrm{T}_\Sigma(X)/E = \mathrm{T}_\Sigma(X)/{\approx_E} = \mathrm{T}_\Sigma(X)/{\leftrightarrow_E^*}$ is called the *free E-algebra* with generating set $X/{\approx_E} = \{\, [x] \mid x \in X \,\}$:

Every mapping $\varphi : X/{\approx_E} \to \mathcal{B}$ for some E-algebra $\mathcal{B}$ can be extended to a homomorphism $\hat{\varphi} : \mathrm{T}_\Sigma(X)/E \to \mathcal{B}$.

$\mathrm{T}_\Sigma(\emptyset)/E = \mathrm{T}_\Sigma(\emptyset)/{\approx_E} = \mathrm{T}_\Sigma(\emptyset)/{\leftrightarrow_E^*}$ is called the *initial E-algebra*.

$\approx_E = \{\, (s,t) \mid E \models s \approx t \,\}$ is called the *equational theory* of E.

$\approx_E^I = \{\, (s,t) \mid \mathrm{T}_\Sigma(\emptyset)/E \models s \approx t \,\}$ is called the *inductive theory* of E.

Example:

Let $E = \{\forall x(x + 0 \approx x),\ \forall x \forall y(x + s(y) \approx s(x + y))\}$. Then $x + y \approx_E^I y + x$, but $x + y \not\approx_E y + x$.

4.3 Confluence

Let $(A, \to)$ be an abstract reduction system.

b and $c \in A$ are *joinable*, if there is a a such that $b \to^* a \leftarrow^* c$.
Notation: $b \downarrow c$.

The relation $\to$ is called

Church-Rosser, if $b \leftrightarrow^* c$ implies $b \downarrow c$.

confluent, if $b \leftarrow^* a \to^* c$ implies $b \downarrow c$.

locally confluent, if $b \leftarrow a \to c$ implies $b \downarrow c$.

convergent, if it is confluent and terminating.

Theorem 4.7 *The following properties are equivalent:*

(i) $\to$ *has the Church-Rosser property.*

(ii) $\to$ *is confluent.*

Proof. (i)⇒(ii): trivial.

(ii)⇒(i): by induction on the number of peaks in the derivation $b \leftrightarrow^* c$. □

Lemma 4.8 *If* $\to$ *is confluent, then every element has at most one normal form.*

Proof. Suppose that some element $a \in A$ has normal forms b and c, then $b \leftarrow^* a \to^* c$. If $\to$ is confluent, then $b \to^* d \leftarrow^* c$ for some $d \in A$. Since b and c are normal forms, both derivations must be empty, hence $b \to^0 d \leftarrow^0 c$, so b, c, and d must be identical. □

Corollary 4.9 *If* $\to$ *is normalizing and confluent, then every element* b *has a unique normal form.*

Proposition 4.10 *If* $\to$ *is normalizing and confluent, then* $b \leftrightarrow^* c$ *if and only if* $b{\downarrow} = c{\downarrow}$.

Proof. Either using Thm. 4.7 or directly by induction on the length of the derivation of $b \leftrightarrow^* c$. □

Confluence and Local Confluence

Theorem 4.11 ("Newman's Lemma") *If a terminating relation $\to$ is locally confluent, then it is confluent.*

Proof. Let $\to$ be a terminating and locally confluent relation. Then $\to^+$ is a well-founded ordering. Define $\phi(a) \;\Leftrightarrow\; \big(\forall b, c : b \leftarrow^* a \to^* c \Rightarrow b \downarrow c\big)$.

We prove $\phi(a)$ for all $a \in A$ by well-founded induction over $\to^+$:

Case 1: $b \leftarrow^0 a \to^* c$: trivial.

Case 2: $b \leftarrow^* a \to^0 c$: trivial.

Case 3: $b \leftarrow^* b' \leftarrow a \to c' \to^* c$: use local confluence, then use the induction hypothesis. □

Rewrite Relations

Corollary 4.12 *If E is convergent (i.e., terminating and confluent), then $s \approx_E t$ if and only if $s \leftrightarrow_E^* t$ if and only if $s{\downarrow_E} = t{\downarrow_E}$.*

Corollary 4.13 *If E is finite and convergent, then $\approx_E$ is decidable.*

Reminder:
If E is terminating, then it is confluent if and only if it is locally confluent.

Problems:

Show local confluence of E.

Show termination of E.

Transform E into an equivalent set of equations that is locally confluent and terminating.

4.4 Critical Pairs

Showing local confluence (Sketch):

Problem: If $t_1 \leftarrow_E t_0 \rightarrow_E t_2$, does there exist a term s such that $t_1 \rightarrow_E^* s \leftarrow_E^* t_2$?

If the two rewrite steps happen in different subtrees (disjoint redexes): yes.

If the two rewrite steps happen below each other (overlap at or below a variable position): yes.

If the left-hand sides of the two rules overlap at a non-variable position: needs further investigation.

Question:
Are there rewrite rules $l_1 \rightarrow r_1$ and $l_2 \rightarrow r_2$ such that some subterm $l_1|_p$ and l_2 have a common instance $(l_1|_p)\sigma_1 = l_2\sigma_2$?

Observation:
If we assume w.l.o.g. that the two rewrite rules do not have common variables, then only a single substitution is necessary: $(l_1|_p)\sigma = l_2\sigma$.

Further observation:
The mgu of $l_1|_p$ and l_2 subsumes all unifiers σ of $l_1|_p$ and l_2.

Let $l_i \rightarrow r_i$ $(i = 1, 2)$ be two rewrite rules in a TRS R whose variables have been renamed such that $\mathrm{var}(l_1) \cap \mathrm{var}(l_2) = \emptyset$. (Remember that $\mathrm{var}(l_i) \supseteq \mathrm{var}(r_i)$.)

Let $p \in \mathrm{pos}(l_1)$ be a position such that $l_1|_p$ is not a variable and σ is an mgu of $l_1|_p$ and l_2.

Then $r_1\sigma \leftarrow l_1\sigma \rightarrow (l_1\sigma)[r_2\sigma]_p$.

$\langle r_1\sigma, (l_1\sigma)[r_2\sigma]_p \rangle$ is called a *critical pair* of R.

The critical pair is *joinable* (or: converges), if $r_1\sigma \downarrow_R (l_1\sigma)[r_2\sigma]_p$.

Theorem 4.14 ("Critical Pair Theorem") *A TRS R is locally confluent if and only if all its critical pairs are joinable.*

Proof. "only if": obvious, since joinability of a critical pair is a special case of local confluence.

"if": Suppose s rewrites to t_1 and t_2 using rewrite rules $l_i \rightarrow r_i \in R$ at positions $p_i \in \mathrm{pos}(s)$, where $i = 1, 2$. Without loss of generality, we can assume that the two rules are variable disjoint, hence $s|_{p_i} = l_i\theta$ and $t_i = s[r_i\theta]_{p_i}$.

We distinguish between two cases: Either p_1 and p_2 are in disjoint subtrees ($p_1 \parallel p_2$), or one is a prefix of the other (w.l.o.g., $p_1 \leq p_2$).

Case 1: $p_1 \parallel p_2$.

Then $s = s[l_1\theta]_{p_1}[l_2\theta]_{p_2}$, and therefore $t_1 = s[r_1\theta]_{p_1}[l_2\theta]_{p_2}$ and $t_2 = s[l_1\theta]_{p_1}[r_2\theta]_{p_2}$.

Let $t_0 = s[r_1\theta]_{p_1}[r_2\theta]_{p_2}$. Then clearly $t_1 \to_R t_0$ using $l_2 \to r_2$ and $t_2 \to_R t_0$ using $l_1 \to r_1$.

Case 2: $p_1 \leq p_2$.

Case 2.1: $p_2 = p_1q_1q_2$, where $l_1|_{q_1}$ is some variable x.

In other words, the second rewrite step takes place at or below a variable in the first rule. Suppose that x occurs m times in l_1 and n times in r_1 (where $m \geq 1$ and $n \geq 0$).

Then $t_1 \to_R^* t_0$ by applying $l_2 \to r_2$ at all positions $p_1q'q_2$, where q' is a position of x in r_1.

Conversely, $t_2 \to_R^* t_0$ by applying $l_2 \to r_2$ at all positions p_1qq_2, where q is a position of x in l_1 different from q_1, and by applying $l_1 \to r_1$ at p_1 with the substitution θ', where $\theta' = \theta[x \mapsto (x\theta)[r_2\theta]_{q_2}]$.

Case 2.2: $p_2 = p_1p$, where p is a non-variable position of l_1.

Then $s|_{p_2} = l_2\theta$ and $s|_{p_2} = (s|_{p_1})|_p = (l_1\theta)|_p = (l_1|_p)\theta$, so θ is a unifier of l_2 and $l_1|_p$.

Let σ be the mgu of l_2 and $l_1|_p$, then $\theta = \tau \circ \sigma$ and $\langle r_1\sigma, (l_1\sigma)[r_2\sigma]_p\rangle$ is a critical pair.

By assumption, it is joinable, so $r_1\sigma \to_R^* v \leftarrow_R^* (l_1\sigma)[r_2\sigma]_p$.

Consequently, $t_1 = s[r_1\theta]_{p_1} = s[r_1\sigma\tau]_{p_1} \to_R^* s[v\tau]_{p_1}$ and $t_2 = s[r_2\theta]_{p_2} = s[(l_1\theta)[r_2\theta]_p]_{p_1} = s[(l_1\sigma\tau)[r_2\sigma\tau]_p]_{p_1} = s[((l_1\sigma)[r_2\sigma]_p)\tau]_{p_1} \to_R^* s[v\tau]_{p_1}$.

This completes the proof of the Critical Pair Theorem. □

Note: Critical pairs between a rule and (a renamed variant of) itself must be considered – except if the overlap is at the root (i.e., $p = \varepsilon$).

Corollary 4.15 *A terminating TRS R is confluent if and only if all its critical pairs are joinable.*

Proof. By Newman's Lemma and the Critical Pair Theorem. □

Corollary 4.16 *For a finite terminating TRS, confluence is decidable.*

Proof. For every pair of rules and every non-variable position in the first rule there is at most one critical pair $\langle u_1, u_2\rangle$.

Reduce every u_i to some normal form u_i'. If $u_1' = u_2'$ for every critical pair, then R is confluent, otherwise there is some non-confluent situation $u_1' \leftarrow_R^* u_1 \leftarrow_R s \to_R u_2 \to_R^* u_2'$. □

4.5 Termination

Termination problems:

Given a finite TRS R and a term t, are all R-reductions starting from t terminating?

Given a finite TRS R, are all R-reductions terminating?

Proposition 4.17 *Both termination problems for TRSs are undecidable in general.*

Proof. Encode Turing machines using rewrite rules and reduce the (uniform) halting problems for TMs to the termination problems for TRSs. □

Consequence:

Decidable criteria for termination are not complete.

Two Different Scenarios

Depending on the application, the TRS whose termination we want to show can be

(i) fixed and known in advance, or

(ii) evolving (e.g., generated by some saturation process).

Methods for case (ii) are also usable for case (i). Many methods for case (i) are not usable for case (ii).

We will first consider case (ii); additional techniques for case (i) will be considered later.

Reduction Orderings

Goal:

Given a finite TRS R, show termination of R by looking at finitely many rules $l \to r \in R$, rather than at infinitely many possible replacement steps $s \to_R s'$.

A binary relation $\sqsupset$ over $\mathrm{T}_\Sigma(X)$ is called *compatible with Σ-operations*, if $s \sqsupset s'$ implies $f(t_1, \ldots, s, \ldots, t_n) \sqsupset f(t_1, \ldots, s', \ldots, t_n)$ for all $f \in \Omega$ and $s, s', t_i \in \mathrm{T}_\Sigma(X)$.

Lemma 4.18 *The relation $\sqsupset$ is compatible with Σ-operations, if and only if $s \sqsupset s'$ implies $t[s]_p \sqsupset t[s']_p$ for all $s, s', t \in \mathrm{T}_\Sigma(X)$ and $p \in \mathrm{pos}(t)$.*

Note: *compatible with Σ-operations = compatible with contexts.*

A binary relation $\sqsupset$ over $\mathrm{T}_\Sigma(X)$ is called *stable under substitutions*, if $s \sqsupset s'$ implies $s\sigma \sqsupset s'\sigma$ for all $s, s' \in \mathrm{T}_\Sigma(X)$ and substitutions σ.

A binary relation $\sqsupset$ is called a *rewrite relation*, if it is compatible with Σ-operations and stable under substitutions.

Example: If R is a TRS, then $\to_R$ is a rewrite relation.

A strict partial ordering over $\mathrm{T}_\Sigma(X)$ that is a rewrite relation is called *rewrite ordering*.

A well-founded rewrite ordering is called *reduction ordering*.

Theorem 4.19 *A TRS R terminates if and only if there exists a reduction ordering $\succ$ such that $l \succ r$ for every rule $l \to r \in R$.*

Proof. "if": $s \to_R s'$ if and only if $s = t[l\sigma]_p$, $s' = t[r\sigma]_p$. If $l \succ r$, then $l\sigma \succ r\sigma$ and therefore $t[l\sigma]_p \succ t[r\sigma]_p$. This implies $\to_R \subseteq \succ$. Since $\succ$ is a well-founded ordering, $\to_R$ is terminating.

"only if": Define $\succ = \to_R^+$. If $\to_R$ is terminating, then $\succ$ is a reduction ordering. □

The Interpretation Method

Proving termination by interpretation:

Let $\mathcal{A}$ be a Σ-algebra; let $\succ$ be a well-founded strict partial ordering on its universe.

Define the ordering $\succ_\mathcal{A}$ over $\mathrm{T}_\Sigma(X)$ by $s \succ_\mathcal{A} t$ if and only if $\mathcal{A}(\beta)(s) \succ \mathcal{A}(\beta)(t)$ for all assignments $\beta : X \to U_\mathcal{A}$.

Is $\succ_\mathcal{A}$ a reduction ordering?

Lemma 4.20 *$\succ_\mathcal{A}$ is stable under substitutions.*

Proof. Let $s \succ_\mathcal{A} s'$, that is, $\mathcal{A}(\beta)(s) \succ \mathcal{A}(\beta)(s')$ for all assignments $\beta : X \to U_\mathcal{A}$. Let σ be a substitution. We have to show that $\mathcal{A}(\gamma)(s\sigma) \succ \mathcal{A}(\gamma)(s'\sigma)$ for all assignments $\gamma : X \to U_\mathcal{A}$. Choose $\beta = \gamma \circ \sigma$, then by the substitution lemma, $\mathcal{A}(\gamma)(s\sigma) = \mathcal{A}(\beta)(s) \succ \mathcal{A}(\beta)(s') = \mathcal{A}(\gamma)(s'\sigma)$. Therefore $s\sigma \succ_\mathcal{A} s'\sigma$. □

A function $\phi : U_\mathcal{A}^n \to U_\mathcal{A}$ is called *monotone* (with respect to $\succ$), if $a \succ a'$ implies $\phi(b_1, \ldots, a, \ldots, b_n) \succ \phi(b_1, \ldots, a', \ldots, b_n)$ for all $a, a', b_i \in U_\mathcal{A}$.

Lemma 4.21 *If the interpretation $f_\mathcal{A}$ of every function symbol f is monotone w.r.t. $\succ$, then $\succ_\mathcal{A}$ is compatible with Σ-operations.*

Proof. Let $s \succ_{\mathcal{A}} s'$, that is, $\mathcal{A}(\beta)(s) \succ \mathcal{A}(\beta)(s')$ for all $\beta : X \to U_{\mathcal{A}}$. Let $\beta : X \to U_{\mathcal{A}}$ be an arbitrary assignment. Then

$$\begin{aligned}\mathcal{A}(\beta)(f(t_1, \ldots, s, \ldots, t_n)) &= f_{\mathcal{A}}(\mathcal{A}(\beta)(t_1), \ldots, \mathcal{A}(\beta)(s), \ldots, \mathcal{A}(\beta)(t_n)) \\ &\succ f_{\mathcal{A}}(\mathcal{A}(\beta)(t_1), \ldots, \mathcal{A}(\beta)(s'), \ldots, \mathcal{A}(\beta)(t_n)) \\ &= \mathcal{A}(\beta)(f(t_1, \ldots, s', \ldots, t_n))\end{aligned}$$

Therefore $f(t_1, \ldots, s, \ldots, t_n) \succ_{\mathcal{A}} f(t_1, \ldots, s', \ldots, t_n)$. □

Theorem 4.22 *If the interpretation $f_{\mathcal{A}}$ of every function symbol f is monotone w.r.t. $\succ$, then $\succ_{\mathcal{A}}$ is a reduction ordering.*

Proof. By the previous two lemmas, $\succ_{\mathcal{A}}$ is a rewrite relation. If there were an infinite chain $s_1 \succ_{\mathcal{A}} s_2 \succ_{\mathcal{A}} \ldots$, then it would correspond to an infinite chain $\mathcal{A}(\beta)(s_1) \succ \mathcal{A}(\beta)(s_2) \succ \ldots$ (with β chosen arbitrarily). Thus $\succ_{\mathcal{A}}$ is well-founded. Irreflexivity and transitivity are proved similarly. □

Polynomial Orderings

Polynomial orderings:

Instance of the interpretation method:

The carrier set $U_{\mathcal{A}}$ is $\mathbb{N}$ or some subset of $\mathbb{N}$.

To every function symbol f/n we associate a polynomial $P_f(X_1, \ldots, X_n) \in \mathbb{N}[X_1, \ldots, X_n]$ with coefficients in $\mathbb{N}$ and indeterminates $X_1, \ldots, X_n$. Then we define $f_{\mathcal{A}}(a_1, \ldots, a_n) = P_f(a_1, \ldots, a_n)$ for $a_i \in U_{\mathcal{A}}$.

Requirement 1:

If $a_1, \ldots, a_n \in U_{\mathcal{A}}$, then $f_{\mathcal{A}}(a_1, \ldots, a_n) \in U_{\mathcal{A}}$. (Otherwise, $\mathcal{A}$ would not be a Σ-algebra.)

Requirement 2:

$f_{\mathcal{A}}$ must be monotone (w.r.t. $\succ$).

From now on:

$U_{\mathcal{A}} = \{\, n \in \mathbb{N} \mid n \geq 1 \,\}$.

If $\text{arity}(f) = 0$, then P_f is a constant ≥ 1.

If $\text{arity}(f) = n \geq 1$, then P_f is a polynomial $P(X_1, \ldots, X_n)$, such that every X_i occurs in some monomial $m \cdot X_1^{j_1} \cdots X_k^{j_k}$ with exponent at least 1 and non-zero coefficient $m \in \mathbb{N}$.

$\Rightarrow$ Requirements 1 and 2 are satisfied.

The mapping from function symbols to polynomials can be extended to terms: A term t containing the variables $x_1, \ldots, x_n$ yields a polynomial P_t with indeterminates $X_1, \ldots, X_n$ (where X_i corresponds to $\beta(x_i)$).

Example:

$\Omega = \{b/0,\ f/1,\ g/3\}$
$P_b = 3, \quad P_f(X_1) = X_1^2, \quad P_g(X_1, X_2, X_3) = X_1 + X_2 X_3.$

Let $t = g(f(b), f(x), y)$, then $P_t(X, Y) = 9 + X^2 Y$.

If P, Q are polynomials in $\mathbb{N}[X_1, \ldots, X_n]$, we write $P > Q$ if $P(a_1, \ldots, a_n) > Q(a_1, \ldots, a_n)$ for all $a_1, \ldots, a_n \in U_{\mathcal{A}}$.

Clearly, $s \succ_{\mathcal{A}} t$ if and only if $P_s > P_t$ if and only if $P_s - P_t > 0$.

Question: Can we check $P_s - P_t > 0$ automatically?

Hilbert's 10th Problem:

Given a polynomial $P \in \mathbb{Z}[X_1, \ldots, X_n]$ with integer coefficients, is $P = 0$ for some n-tuple of natural numbers?

Theorem 4.23 *Hilbert's 10th Problem is undecidable.*

Proposition 4.24 *Given a polynomial interpretation and two terms s, t, it is undecidable whether $P_s > P_t$.*

Proof. By reduction of Hilbert's 10th Problem. □

One easy case:

If we restrict to linear polynomials, deciding whether $P_s - P_t > 0$ is trivial:

$\sum k_i a_i + k > 0$ for all $a_1, \ldots, a_n \geq 1$ if and only if

$k_i \geq 0$ for all $i \in \{1, \ldots, n\}$,

and $\sum k_i + k > 0$

Another possible solution:

Test whether $P_s(a_1, \ldots, a_n) > P_t(a_1, \ldots, a_n)$ for all $a_1, \ldots, a_n \in \{\, x \in \mathbb{R} \mid x \geq 1 \,\}$.

This is decidable (but hard). Since $U_{\mathcal{A}} \subseteq \{\, x \in \mathbb{R} \mid x \geq 1 \,\}$, it implies $P_s > P_t$.

Alternatively:

Use fast overapproximations.

Simplification Orderings

The *proper subterm ordering* $\rhd$ is defined by $s \rhd t$ if and only if $s|_p = t$ for some position $p \neq \varepsilon$ of s.

A rewrite ordering $\succ$ over $\mathrm{T}_\Sigma(X)$ is called *simplification ordering,* if it has the *subterm property:* $s \rhd t$ implies $s \succ t$ for all $s, t \in \mathrm{T}_\Sigma(X)$.

Example:

Let R_{emb} be the rewrite system $R_{\mathrm{emb}} = \{\, f(x_1, \ldots, x_n) \to x_i \mid f/n \in \Omega,\, 1 \leq i \leq n \,\}$.

Define $\rhd_{\mathrm{emb}} = \to^+_{R_{\mathrm{emb}}}$ and $\trianglerighteq_{\mathrm{emb}} = \to^*_{R_{\mathrm{emb}}}$ *("homeomorphic embedding relation").*

$\rhd_{\mathrm{emb}}$ is a simplification ordering.

Lemma 4.25 *If $\succ$ is a simplification ordering, then $s \rhd_{\mathrm{emb}} t$ implies $s \succ t$ and $s \trianglerighteq_{\mathrm{emb}} t$ implies $s \succeq t$.*

Proof. Since $\succ$ is transitive and $\succeq$ is transitive and reflexive, it suffices to show that $s \to_{R_{\mathrm{emb}}} t$ implies $s \succ t$. By definition, $s \to_{R_{\mathrm{emb}}} t$ if and only if $s = s[l\sigma]$ and $t = s[r\sigma]$ for some rule $l \to r \in R_{\mathrm{emb}}$. Obviously, $l \rhd r$ for all rules in R_{emb}, hence $l \succ r$. Since $\succ$ is a rewrite relation, $s = s[l\sigma] \succ s[r\sigma] = t$. □

Goal:

Show that every simplification ordering is well-founded (and therefore a reduction ordering).

Note: This works only for *finite* signatures!

To fix this for infinite signatures, the definition of simplification orderings and the definition of embedding have to be modified.

Theorem 4.26 ("Kruskal's Theorem") *Let Σ be a finite signature, let X be a finite set of variables. Then for every infinite sequence $t_1, t_2, t_3, \ldots$ there are indices $j > i$ such that $t_j \trianglerighteq_{\mathrm{emb}} t_i$. ($\trianglerighteq_{\mathrm{emb}}$ is called a* well-partial-ordering (wpo)*.)*

Proof. See Baader and Nipkow, page 113–115. □

Theorem 4.27 (Dershowitz) *If Σ is a finite signature, then every simplification ordering $\succ$ on $\mathrm{T}_\Sigma(X)$ is well-founded (and therefore a reduction ordering).*

Proof. Suppose that $t_1 \succ t_2 \succ t_3 \succ \ldots$ is an infinite descending chain.

First assume that there is an $x \in \mathrm{var}(t_{i+1}) \setminus \mathrm{var}(t_i)$. Let $\sigma = \{x \mapsto t_i\}$, then $t_{i+1}\sigma \trianglerighteq x\sigma = t_i$ and therefore $t_i = t_i\sigma \succ t_{i+1}\sigma \succeq t_i$, contradicting irreflexivity.

Consequently, $\mathrm{var}(t_i) \supseteq \mathrm{var}(t_{i+1})$ and $t_i \in \mathrm{T}_\Sigma(V)$ for all i, where V is the finite set $\mathrm{var}(t_1)$. By Kruskal's Theorem, there are $i < j$ with $t_i \trianglelefteq_{\mathrm{emb}} t_j$. Hence $t_i \preceq t_j$, contradicting $t_i \succ t_j$. □

There are reduction orderings that are not simplification orderings and terminating TRSs that are not contained in any simplification ordering.

Example:

Let $R = \{f(f(x)) \to f(g(f(x)))\}$.

R terminates and $\to_R^+$ is therefore a reduction ordering.

Assume that $\to_R$ were contained in a simplification ordering $\succ$. Then $f(f(x)) \to_R f(g(f(x)))$ implies $f(f(x)) \succ f(g(f(x)))$, and $f(g(f(x))) \trianglerighteq_{\mathrm{emb}} f(f(x))$ implies $f(g(f(x))) \succeq f(f(x))$, hence $f(f(x)) \succ f(f(x))$.

Path Orderings

Let $\Sigma = (\Omega, \Pi)$ be a finite signature, let $\succ$ be a strict partial ordering *("precedence")* on Ω.

The *lexicographic path ordering* $\succ_{\mathrm{lpo}}$ on $\mathrm{T}_\Sigma(X)$ induced by $\succ$ is defined by: $s \succ_{\mathrm{lpo}} t$ if

(1) $t \in \mathrm{var}(s)$ and $t \neq s$, or

(2) $s = f(s_1, \ldots, s_m)$, $t = g(t_1, \ldots, t_n)$, and

  (a) $s_i \succeq_{\mathrm{lpo}} t$ for some i, or

  (b) $f \succ g$ and $s \succ_{\mathrm{lpo}} t_j$ for all j, or

  (c) $f = g$, $s \succ_{\mathrm{lpo}} t_j$ for all j, and $(s_1, \ldots, s_m)\ (\succ_{\mathrm{lpo}})_{\mathrm{lex}}\ (t_1, \ldots, t_n)$.

where $(\succ_{\mathrm{lpo}})_{\mathrm{lex}}$ is the m-fold lexicographic combination of $\succ_{\mathrm{lpo}}$
(note that $f = g$ implies $m = n$).

Lemma 4.28 *$s \succ_{\mathrm{lpo}} t$ implies* $\mathrm{var}(s) \supseteq \mathrm{var}(t)$.

Proof. By induction on $|s| + |t|$ and case analysis. □

Theorem 4.29 *$\succ_{\text{lpo}}$ is a simplification ordering on $\mathrm{T}_\Sigma(X)$.*

Proof. Show transitivity, subterm property, stability under substitutions, compatibility with Σ-operations, and irreflexivity, usually by induction on the sum of the term sizes and case analysis. Details: Baader and Nipkow, page 119/120. □

Theorem 4.30 *If the precedence $\succ$ is total, then the lexicographic path ordering $\succ_{\text{lpo}}$ is total on ground terms, i.e., for all $s, t \in \mathrm{T}_\Sigma(\emptyset)$: $s \succ_{\text{lpo}} t \vee t \succ_{\text{lpo}} s \vee s = t$.*

Proof. By induction on $|s| + |t|$ and case analysis. □

Recapitulation:

Let $\Sigma = (\Omega, \Pi)$ be a finite signature, let $\succ$ be a strict partial ordering *("precedence")* on Ω. The *lexicographic path ordering* $\succ_{\text{lpo}}$ on $\mathrm{T}_\Sigma(X)$ induced by $\succ$ is defined by: $s \succ_{\text{lpo}} t$ if

(1) $t \in \mathrm{var}(s)$ and $t \neq s$, or

(2) $s = f(s_1, \ldots, s_m)$, $t = g(t_1, \ldots, t_n)$, and

  (a) $s_i \succeq_{\text{lpo}} t$ for some i, or

  (b) $f \succ g$ and $s \succ_{\text{lpo}} t_j$ for all j, or

  (c) $f = g$, $s \succ_{\text{lpo}} t_j$ for all j, and $(s_1, \ldots, s_m)\ (\succ_{\text{lpo}})_{\text{lex}}\ (t_1, \ldots, t_n)$.

There are several possibilities to compare subterms in (2)(c):

- compare list of subterms lexicographically left-to-right (*"lexicographic path ordering (lpo)"*, Kamin and Lévy)
- compare list of subterms lexicographically right-to-left (or according to some permutation π)
- compare multiset of subterms using the multiset extension (*"multiset path ordering (mpo)"*, Dershowitz)
- to each function symbol $f/n \in \Omega$ with $n \geq 1$ associate a status $\in \{\text{mul}\} \cup \{\, \text{lex}_\pi \mid \pi : \{1, \ldots, n\} \to \{1, \ldots, n\} \,\}$ and compare according to that status (*"recursive path ordering (rpo) with status"*)

The Knuth-Bendix Ordering

Let $\Sigma = (\Omega, \Pi)$ be a finite signature, let $\succ$ be a strict partial ordering *("precedence")* on Ω, let $w : \Omega \cup X \to \mathbb{R}_0^+$ be a *weight function*, such that the following admissibility conditions are satisfied:

$w(x) = w_0 \in \mathbb{R}^+$ for all variables $x \in X$; $w(c) \geq w_0$ for all constants $c \in \Omega$.

If $w(f) = 0$ for some $f/1 \in \Omega$, then $f \succ g$ for all $g/n \in \Omega$ with $f \neq g$.

The weight function w can be extended to terms recursively:

$$w(f(t_1, \ldots, t_n)) = w(f) + \sum_{1 \leq i \leq n} w(t_i)$$

or alternatively

$$w(t) = \sum_{x \in \mathrm{var}(t)} w(x) \cdot \#(x, t) + \sum_{f \in \Omega} w(f) \cdot \#(f, t).$$

where $\#(a, t)$ is the number of occurrences of a in t.

The *Knuth-Bendix ordering* $\succ_{\mathrm{kbo}}$ on $\mathrm{T}_\Sigma(X)$ induced by $\succ$ and w is defined by: $s \succ_{\mathrm{kbo}} t$ if

(1) $\#(x, s) \geq \#(x, t)$ for all variables x and $w(s) > w(t)$, or

(2) $\#(x, s) \geq \#(x, t)$ for all variables x, $w(s) = w(t)$, and

  (a) $t = x$, $s = f^n(x)$ for some $n \geq 1$, or

  (b) $s = f(s_1, \ldots, s_m)$, $t = g(t_1, \ldots, t_n)$, and $f \succ g$, or

  (c) $s = f(s_1, \ldots, s_m)$, $t = f(t_1, \ldots, t_m)$, and $(s_1, \ldots, s_m)$ $(\succ_{\mathrm{kbo}})_{\mathrm{lex}}$ $(t_1, \ldots, t_m)$.

Theorem 4.31 *The Knuth-Bendix ordering induced by $\succ$ and w is a simplification ordering on $\mathrm{T}_\Sigma(X)$.*

Proof. Baader and Nipkow, pages 125–129. □

Remark

If $\Pi \neq \emptyset$, then all the term orderings described in this section can also be used to compare non-equational atoms by treating predicate symbols like function symbols.

4.6 Knuth-Bendix Completion

Completion:

Goal: Given a set E of equations, transform E into an equivalent convergent set R of rewrite rules.
(If R is finite: decision procedure for E.)

Knuth-Bendix Completion: Idea

How to ensure termination?

Fix a reduction ordering $\succ$ and construct R in such a way that $\to_R \subseteq \succ$ (i.e., $l \succ r$ for every $l \to r \in R$).

How to ensure confluence?

Check that all critical pairs are joinable.

Note: Every critical pair $\langle s, t\rangle$ can be *made* joinable by adding $s \to t$ or $t \to s$ to R.

(Actually, we first add $s \approx t$ to E and later try to turn it into a rule that is contained in $\succ$; this gives us some additional degree of freedom.)

Knuth-Bendix Completion: Inference Rules

The completion procedure is presented as a set of inference rules working on a set of equations E and a set of rules R: $E_0, R_0 \vdash E_1, R_1 \vdash E_2, R_2 \vdash \ldots$

At the beginning, $E = E_0$ is the input set and $R = R_0$ is empty. At the end, E should be empty; then R is the result.

For each step $E, R \vdash E', R'$, the equational theories of $E \cup R$ and $E' \cup R'$ agree: $\approx_{E \cup R} = \approx_{E' \cup R'}$.

Notations:

The formula $s \mathrel{\dot{\approx}} t$ denotes either $s \approx t$ or $t \approx s$.

$\mathrm{CP}(R)$ denotes the set of all critical pairs between rules in R.

Orient:

$$\frac{E \cup \{s \mathrel{\dot{\approx}} t\},\quad R}{E,\quad R \cup \{s \to t\}} \qquad \text{if } s \succ t$$

Note: There are equations $s \approx t$ that cannot be oriented, i.e., neither $s \succ t$ nor $t \succ s$.

Trivial equations cannot be oriented – but we don't need them anyway:

Delete:

$$\frac{E \cup \{s \approx s\},\quad R}{E,\quad R}$$

Critical pairs between rules in R are turned into additional equations:

Deduce:

$$\frac{E,\quad R}{E \cup \{s \approx t\},\quad R} \qquad \text{if } \langle s, t\rangle \in \mathrm{CP}(R).$$

Note: If $\langle s, t\rangle \in \mathrm{CP}(R)$ then $s \leftarrow_R u \to_R t$ and hence $R \models s \approx t$.

The following inference rules are not absolutely necessary, but very useful (e.g., to get rid of joinable critical pairs and to deal with equations that cannot be oriented):

Simplify-Eq:

$$\frac{E \cup \{s \mathrel{\dot{\approx}} t\},\quad R}{E \cup \{u \approx t\},\quad R} \qquad \text{if } s \to_R u.$$

Simplification of the right-hand side of a rule is unproblematic:

R-Simplify-Rule:

$$\frac{E,\quad R \cup \{s \to t\}}{E,\quad R \cup \{s \to u\}} \qquad \text{if } t \to_R u.$$

Simplification of the left-hand side may influence orientability and orientation. Therefore, it yields an *equation*:

L-Simplify-Rule:

$$\frac{E,\quad R \cup \{s \to t\}}{E \cup \{u \approx t\},\quad R} \qquad \begin{array}{l}\text{if } s \to_R u \text{ using a rule } l \to r \in R\\ \text{such that } s \sqsupset l \text{ (see below).}\end{array}$$

For technical reasons, the lhs of $s \to t$ may only be simplified using a rule $l \to r$, if $l \to r$ *cannot* be simplified using $s \to t$, that is, if $s \sqsupset l$, where the *encompassment quasi-ordering* $\supseteq$ is defined by

$$s \supseteq l \text{ if } s|_p = l\sigma \text{ for some } p \text{ and } \sigma$$

and $\sqsupset = \supseteq \setminus \sqsubseteq$ is the strict part of $\supseteq$.

Lemma 4.32 *$\sqsupset$ is a well-founded strict partial ordering.*

Lemma 4.33 *If $E, R \vdash E', R'$, then $\approx_{E \cup R} = \approx_{E' \cup R'}$.*

Lemma 4.34 *If $E, R \vdash E', R'$ and $\to_R \subseteq \succ$, then $\to_{R'} \subseteq \succ$.*

Note: Like in ordered resolution, simplification should be preferred to deduction:

- Simplify/delete whenever possible.
- Otherwise, orient an equation, if possible.
- Last resort: compute critical pairs.

Knuth-Bendix Completion: Correctness Proof[4]

What can happen if we run the completion procedure on a set E of equations?

(1) We reach a state where no more inference rules are applicable and E is not empty. $\Rightarrow$ Failure (try again with another ordering?)

(2) We reach a state where E is empty and all critical pairs between the rules in the current R have been checked.

(3) The procedure runs forever.

In order to treat these cases simultaneously, we need some definitions.

A (finite or infinite sequence) $E_0, R_0 \vdash E_1, R_1 \vdash E_2, R_2 \vdash \ldots$ with $R_0 = \emptyset$ is called a *run* of the completion procedure with input E_0 and $\succ$.

For a run, $E_\cup = \bigcup_{i \geq 0} E_i$ and $R_\cup = \bigcup_{i \geq 0} R_i$.

The sets of *persistent equations or rules* of the run are $E_\infty = \bigcup_{i \geq 0} \bigcap_{j \geq i} E_j$ and $R_\infty = \bigcup_{i \geq 0} \bigcap_{j \geq i} R_j$.

Note: If the run is finite and ends with E_n, R_n, then $E_\infty = E_n$ and $R_\infty = R_n$.

[4]The notations in this subsection differ significantly from the 2021/2022 lecture. Keep that in mind when you use online lecture recordings or read exercises or exam questions from previous years.

A run is called *fair*, if $CP(R_\infty) \subseteq E_\cup$ (i.e., if every critical pair between persisting rules is computed at some step of the derivation).

Goal:

Show: If a run is fair and E_∞ is empty, then R_∞ is convergent and equivalent to E_0.

In particular: If a run is fair and E_∞ is empty, then $\approx_{E_0} = \approx_{E_\cup \cup R_\cup} = \leftrightarrow^*_{E_\cup \cup R_\cup} = \downarrow_{R_\infty}$.

General assumptions from now on:

$E_0, R_0 \vdash E_1, R_1 \vdash E_2, R_2 \vdash \ldots$ is a fair run.

R_0 and E_∞ are empty.

A *proof* of $s \approx t$ in $E_\cup \cup R_\cup$ is a finite sequence $(s_0, \ldots, s_n)$ such that $s = s_0$, $t = s_n$, and for all $i \in \{1, \ldots, n\}$:

(1) $s_{i-1} \leftrightarrow_{E_\cup} s_i$, or

(2) $s_{i-1} \rightarrow_{R_\cup} s_i$, or

(3) $s_{i-1} \leftarrow_{R_\cup} s_i$.

The pairs (s_{i-1}, s_i) are called *proof steps*.

A proof is called a *rewrite proof in* R_∞, if there is a $k \in \{0, \ldots, n\}$ such that $s_{i-1} \rightarrow_{R_\infty} s_i$ for $1 \leq i \leq k$ and $s_{i-1} \leftarrow_{R_\infty} s_i$ for $k+1 \leq i \leq n$

Idea (Bachmair, Dershowitz, Hsiang):

Define a well-founded ordering on proofs, such that for every proof that is not a rewrite proof in R_∞ there is an equivalent smaller proof.

Consequence: For every proof there is an equivalent rewrite proof in R_∞.

We associate a *cost* $c(s_{i-1}, s_i)$ with every proof step as follows:

(1) If $s_{i-1} \leftrightarrow_{E_\cup} s_i$, then $c(s_{i-1}, s_i) = (\{s_{i-1}, s_i\}, -, -)$, where the first component is a multiset of terms and $-$ denotes an arbitrary (irrelevant) term.

(2) If $s_{i-1} \rightarrow_{R_\cup} s_i$ using $l \rightarrow r$, then $c(s_{i-1}, s_i) = (\{s_{i-1}\}, l, s_i)$.

(3) If $s_{i-1} \leftarrow_{R_\cup} s_i$ using $l \rightarrow r$, then $c(s_{i-1}, s_i) = (\{s_i\}, l, s_{i-1})$.

Proof steps are compared using the lexicographic combination of the multiset extension of the reduction ordering $\succ$, the encompassment ordering $\sqsupset$, and the reduction ordering $\succ$.

The cost $c(P)$ of a proof P is the multiset of the costs of its proof steps.

The *proof ordering* $\succ_C$ compares the costs of proofs using the multiset extension of the proof step ordering.

Lemma 4.35 *$\succ_C$ is a well-founded ordering.*

Lemma 4.36 *Let P be a proof in $E_\cup \cup R_\cup$. If P is not a rewrite proof in R_∞, then there exists an equivalent proof P' in $E_\cup \cup R_\cup$ such that $P \succ_C P'$.*

Proof. If P is not a rewrite proof in R_∞, then it contains

(a) a proof step that is in $E_\cup$, or
(b) a proof step that is in $R_\cup \setminus R_\infty$, or
(c) a subproof $s_{i-1} \leftarrow_{R_\infty} s_i \rightarrow_{R_\infty} s_{i+1}$ (peak).

We show that in all three cases the proof step or subproof can be replaced by a smaller subproof:

Case (a): A proof step using an equation $s \mathrel{\dot{\approx}} t$ is in $E_\cup$. This equation must be deleted during the run.

If $s \mathrel{\dot{\approx}} t$ is deleted using *Orient*:
$\ldots s_{i-1} \leftrightarrow_{E_\cup} s_i \ldots \quad \Longrightarrow \quad \ldots s_{i-1} \rightarrow_{R_\cup} s_i \ldots$

If $s \mathrel{\dot{\approx}} t$ is deleted using *Delete*:
$\ldots s_{i-1} \leftrightarrow_{E_\cup} s_{i-1} \ldots \quad \Longrightarrow \quad \ldots s_{i-1} \ldots$

If $s \mathrel{\dot{\approx}} t$ is deleted using *Simplify-Eq*:
$\ldots s_{i-1} \leftrightarrow_{E_\cup} s_i \ldots \quad \Longrightarrow \quad \ldots s_{i-1} \rightarrow_{R_\cup} s' \leftrightarrow_{E_\cup} s_i \ldots$

Case (b): A proof step using a rule $s \rightarrow t$ is in $R_\cup \setminus R_\infty$. This rule must be deleted during the run.

If $s \rightarrow t$ is deleted using *R-Simplify-Rule*:
$\ldots s_{i-1} \rightarrow_{R_\cup} s_i \ldots \quad \Longrightarrow \quad \ldots s_{i-1} \rightarrow_{R_\cup} s' \leftarrow_{R_\cup} s_i \ldots$

If $s \rightarrow t$ is deleted using *L-Simplify-Rule*:
$\ldots s_{i-1} \rightarrow_{R_\cup} s_i \ldots \quad \Longrightarrow \quad \ldots s_{i-1} \rightarrow_{R_\cup} s' \leftrightarrow_{E_\cup} s_i \ldots$

Case (c): A subproof has the form $s_{i-1} \leftarrow_{R_\infty} s_i \rightarrow_{R_\infty} s_{i+1}$.

If there is no overlap or a non-critical overlap:
$\ldots s_{i-1} \leftarrow_{R_\infty} s_i \rightarrow_{R_\infty} s_{i+1} \ldots \Longrightarrow \ldots s_{i-1} \rightarrow^*_{R_\infty} s' \leftarrow^*_{R_\infty} s_{i+1} \ldots$

If there is a critical pair that has been added using *Deduce*:
$\ldots s_{i-1} \leftarrow_{R_\infty} s_i \rightarrow_{R_\infty} s_{i+1} \ldots \Longrightarrow \quad \ldots s_{i-1} \leftrightarrow_{E_\cup} s_{i+1} \ldots$

In all cases, checking that the replacement subproof is smaller than the replaced subproof is routine. □

Theorem 4.37 *Let $E_0, R_0 \vdash E_1, R_1 \vdash E_2, R_2 \vdash \ldots$ be a fair run and let R_0 and E_∞ be empty. Then*

(1) every proof in $E_\cup \cup R_\cup$ is equivalent to a rewrite proof in R_∞,

(2) R_∞ is equivalent to E_0, and

(3) R_∞ is convergent.

Proof. (1) By well-founded induction on $\succ_C$ using the previous lemma.

(2) Clearly $\approx_{E_\cup \cup R_\cup} = \approx_{E_0}$. Since $R_\infty \subseteq R_\cup$, we get $\approx_{R_\infty} \subseteq \approx_{E_\cup \cup R_\cup}$. On the other hand, by (1), $\approx_{E_\cup \cup R_\cup} \subseteq \approx_{R_\infty}$.

(3) Since $\to_{R_\infty} \subseteq \succ$, R_∞ is terminating. By (1), R_∞ is confluent. □

4.7 Unfailing Completion

Classical completion:

Try to transform a set E of equations into an equivalent convergent TRS.

Fail, if an equation can neither be oriented nor deleted.

Unfailing completion (Bachmair, Dershowitz and Plaisted):

If an equation cannot be oriented, we can still use *orientable instances* for rewriting.

Note: If $\succ$ is total on ground terms, then every *ground instance* of an equation is trivial or can be oriented.

Goal: Derive a *ground convergent* set of equations.

Let E be a set of equations, let $\succ$ be a reduction ordering.

We define the relation $\to_{E^\succ}$ by

$$s \to_{E^\succ} t \quad \text{if} \quad \begin{array}{l}\text{there exist } (u \approx v) \in E \text{ or } (v \approx u) \in E,\\ p \in \mathrm{pos}(s), \text{ and } \sigma : X \to \mathrm{T}_\Sigma(X),\\ \text{such that } s|_p = u\sigma \text{ and } t = s[v\sigma]_p \text{ and } u\sigma \succ v\sigma.\end{array}$$

Note: $\to_{E^\succ}$ is terminating by construction.

From now on let $\succ$ be a reduction ordering that is total on ground terms.

E is called ground convergent w.r.t. $\succ$, if for all ground terms s and t with $s \leftrightarrow_E^* t$ there exists a ground term v such that $s \to_{E^\succ}^* v \leftarrow_{E^\succ}^* t$. (Analogously for $E \cup R$.)

As for standard completion, we establish ground convergence by computing critical pairs.

However, the ordering $\succ$ is not total on non-ground terms. Since $s\theta \succ t\theta$ implies $s \not\preceq t$, we approximate $\succ$ on ground terms by $\not\preceq$ on arbitrary terms.

Let $u_i \stackrel{\cdot}{\approx} v_i$ $(i = 1, 2)$ be equations in E whose variables have been renamed such that $\mathrm{var}(u_1 \stackrel{\cdot}{\approx} v_1) \cap \mathrm{var}(u_2 \stackrel{\cdot}{\approx} v_2) = \emptyset$. Let $p \in \mathrm{pos}(u_1)$ be a position such that $u_1|_p$ is not a variable, σ is an mgu of $u_1|_p$ and u_2, and $u_i\sigma \not\preceq v_i\sigma$ $(i = 1, 2)$. Then $\langle v_1\sigma, (u_1\sigma)[v_2\sigma]_p\rangle$ is called a *semi-critical pair* of E with respect to $\succ$.

The set of all semi-critical pairs of E is denoted by $\mathrm{SP}_\succ(E)$.

Semi-critical pairs of $E \cup R$ are defined analogously. If $\to_R \subseteq \succ$, then $\mathrm{CP}(R)$ and $\mathrm{SP}_\succ(R)$ agree.

Note: In contrast to critical pairs, it may be necessary to consider overlaps of an equation with itself at the top. For instance, if $E = \{f(x) \approx g(y)\}$, then $\langle g(y), g(y')\rangle$ is a non-trivial semi-critical pair.

The *Deduce* rule takes now the following form:

Deduce:

$$\frac{E, \quad R}{E \cup \{s \approx t\}, \quad R} \qquad \text{if } \langle s, t\rangle \in \mathrm{SP}_\succ(E \cup R).$$

Moreover, the fairness criterion for runs is replaced by

$$\mathrm{SP}_\succ(E_\infty \cup R_\infty) \subseteq E_\cup$$

(i.e., if every semi-critical pair between persisting rules or equations is computed at some step of the derivation).

Analogously to Thm. 4.37 we obtain now the following theorem:

Theorem 4.38 *Let $E_0, R_0 \vdash E_1, R_1 \vdash E_2, R_2 \vdash \ldots$ be a fair run; let $R_0 = \emptyset$. Then*

(1) $E_\infty \cup R_\infty$ is equivalent to E_0, and

(2) $E_\infty \cup R_\infty$ is ground convergent.

Moreover one can show that, whenever there exists a *reduced* convergent R such that $\approx_{E_0} = \downarrow_R$ and $\to_R \in \succ$, then for every fair *and simplifying* run $E_\infty = \emptyset$ and $R_\infty = R$ up to variable renaming.

Here R is called reduced, if for every $l \to r \in R$, both l and r are irreducible w.r.t. $R \setminus \{l \to r\}$. A run is called simplifying, if R_∞ is reduced, and for all equations $u \approx v \in E_\infty$, u and v are incomparable w.r.t. $\succ$ and irreducible w.r.t. R_∞.

Unfailing completion is refutationally complete for equational theories:

Theorem 4.39 *Let E be a set of equations, let $\succ$ be a reduction ordering that is total on ground terms. For any two terms s and t, let $\hat{s}$ and $\hat{t}$ be the terms obtained from s and t by replacing all variables by Skolem constants. Let $eq/2$, $true/0$ and $false/0$ be new operator symbols, such that $true$ and $false$ are smaller than all other terms. Let $E_0 = E \cup \{eq(\hat{s}, \hat{t}) \approx true,\ eq(x, x) \approx false\}$. If $E_0, \emptyset \vdash E_1, R_1 \vdash E_2, R_2 \vdash \ldots$ be a fair run of unfailing completion, then $s \approx_E t$ if and only if some $E_i \cup R_i$ contains $true \approx false$.*

Outlook:

Combine ordered resolution and unfailing completion to get a calculus for equational clauses:

compute inferences between (strictly) maximal literals as in ordered resolution,
compute overlaps between maximal sides of equations as in unfailing completion

$\Rightarrow$ Superposition calculus.