15-414: Bug Catching: Automated Program Verification

Lecture Notes on Resolution

Matt Fredrikson

Carnegie Mellon University
Lecture 15
Thursday, October 22, 2026

1 Introduction

In Lecture 13 we saw the sequent calculus, which has many positive attributes. Sequents are an excellent basis for communicating proof goals. With some refinement, it can serve as a foundation for goal-directed proof search procedures. It is also quite robust, which means that many logics (we have seen dynamic logic, parts of modal logic, classical logic, and a glimpse of intuitionistic logic) have formulations in the sequent calculus. On the other hand, sequent proofs are quite verbose and also lose some of their intuitive appeal if they have multiple succedents.

Many modern provers, however, are not directly based on the sequent calculus. Broadly, we can classify those as either being based on *resolution* in some form or on *cooperating decision procedures*, usually in the form of SMT (Satisfiability Modulo Theories) based on SAT (Satisfiability). In today's lecture we discuss the rudiments of resolution, which also represents a bridge between human-oriented and machine-oriented proof systems. It has the virtue that a resolution proof is quite easy to check, which is not the case for proofs carried out by a SAT or SMT solver. Therefore, SAT solvers these days produce an independently checkable proof certificate, and SMT solvers should (although implementations are not quite as far advanced). One popular format for such a certificate is in fact a resolution proof. This gives us another example of the solver and checker approach from Lecture 14.

In this lecture we present only the propositional case (no quantification) and only *binary resolution*. There are many refinements and optimizations of resolution which was one of the dominant methods of automated theorem proving for several decades. Why3 supports many provers, including provers such as Eprover, Spass, and Vampire that have their origins in variants of resolution but continue to evolve.

Learning goals. After this lecture, you should be able to:

- Carry out resolution proofs for propositional logic
- Construct a satisfying assignment from a consistent, saturated theory

2 Satisfiability

For today's lecture we restrict ourselves to propositional formulas. We use lowercase p, q for *atoms* which you can also think of as propositional variables. Then our grammar for propositions is:

$$Q, P \quad ::= \quad p \mid P \wedge Q \mid P \vee Q \mid P \to Q \mid \neg P \mid \top \mid \bot$$

A truth assignment M assigns either true or false to every propositional variable (or atom, as we say). This is analogous to the *state* ω in dynamic logic that assigns an integer to every variable. So, if you like, you can think of propositional theorem proving as deciding the (quantifier-free) theory of Booleans. We write $M \models P$ if the formula P is true given the assignment M. This is defined exactly as we did in dynamic logic on these connectives with one additional clause:

$$M \models p \quad \text{iff} \quad M(p) = \mathsf{true}$$

A convenient way to present a truth assignment is by giving a list of p if $M(p) = \mathsf{true}$ and $\neg p$ if $M(p) = \mathsf{false}$. In principle, this list would have to be infinite, but since every formula contains only finitely many atoms we can use such a finite representation.

The fundamental question is once again that of validity, that is:

> ***Validity**: decide whether $M \models P$ for* every *truth assignment M.*

For reasons that are not entirely clear to me, this is usually turned into a problem of *satisfiability*:

> ***Satisfiability**: decide whether $M \models P$ for* some *truth assignment M.*

These problems are equivalent in the sense that

> *P is valid if and only if $\neg P$ is unsatisfiable*

which you can easily verify from the definitions. So instead of proving P we try to refute $\neg P$ by searching for an assignment M such that $M \models \neg P$. This is also called a *model* for $\neg P$. If no such model exists, then $\neg P$ is *unsatisfiable* which means that P is valid. If such a model exists then $\neg P$ is *satisfiable*, which means that P is *not valid*. In this case the model M for $\neg P$ provides a counterexample to the validity of P in the sense that $M \not\models P$. Read the previous paragraph at least once more. It's crucial in understanding what follows.

3 Clausal Form

There is some redundancy among the connectives of (classical) propositional logic. Theorem proving methods are easier to think about and implement if we can put the formulas into a normal form such that P is satisfiable if and only if its normal form Q is satisfiable. Moreover, we'd like to be able to take any satisfying assignment M for Q and translate it to a satisfying assignment for P.

One such normal form is called *conjunctive normal form* (CNF). A formula is in conjunctive normal form if it is a conjunction of disjunctions of atoms and negated atoms. That is:

$$\begin{array}{lllll} \text{Literal} & L & ::= & p \mid \neg p \\ \text{Clause} & D, C & ::= & \bot \mid L \mid C \vee D \\ \text{Theory} & S, T & ::= & \top \mid C \mid S \wedge T \end{array}$$

For the algorithms, it is convenient to think of a clause as a *set of literals* $\{L_1, \ldots, L_n\}$ where we write $\bot$ for the empty set. We will still write this as $L_1 \vee \ldots \vee L_n$. Similarly, we think of a theory as a *sequence of clauses*, $C_0, C_1, \ldots, C_k$.

There are a number of methods to put a propositional formula into an equisatisfiable clausal form. A not particularly efficient one would first push in negations and then use the laws of distributivity, but this could blow up the size of the formula exponentially.

In order to satisfy a formula in conjunctive normal form we have to satisfy each conjunct. On the other hand, it is sufficient to satisfy a single literal in each clause. If a clause is empty (representing $\bot$) then it can not be satisfied by any assignment. Consequently, a theory containing the empty clause is unsatisfiable.

As a simple running example, consider

$$P = (p \to (q \to r)) \to ((p \wedge q) \to r)$$

with atoms p, q, and r, where we have written some redundant parentheses for clarity. As a first step, we'll negate this, repeatedly using the laws $(P \to Q) \leftrightarrow (\neg P \vee Q)$ and $\neg(P \to Q) \leftrightarrow (P \wedge \neg Q)$.

$$\begin{array}{rcl} \neg P & = & \neg((p \to (q \to r)) \to ((p \wedge q) \to r)) \\ & \leftrightarrow & (p \to (q \to r)) \wedge \neg((p \wedge q) \to r) \\ & \leftrightarrow & (\neg p \vee (\neg q \vee r)) \wedge (p \wedge q) \wedge \neg r \\ & \leftrightarrow & (\neg p \vee \neg q \vee r) \wedge p \wedge q \wedge \neg r \end{array}$$

At this point we have reached a conjunctive normal form. Written as a theory, labeling each clause:

$$\begin{array}{ll} \neg p \vee \neg q \vee r & C_0 \\ p & C_1 \\ q & C_2 \\ \neg r & C_3 \end{array}$$

It is easy to see that this theory is *unsatisfiable*. Since it should be read as a conjunction, all of C_1, C_2, C_3 must be true, forcing that for any model, p and q would have to be true, and r false. But if p and q are true and r false, then C_0 will be false, preventing us from simultaneously satisfying all clauses.

4 Binary Resolution

Resolution is both the name of a rule of inference and a (nondeterministic) algorithm searching for a *refutation* of a theory T. Such a *refutation* is evidence that the theory is unsatisfiable. In brief, we add more and more consequences to a theory in the hope of reaching the empty clause. If we do, we conclude that the original theory was in fact *unsatisfiable* because our notion of consequence preserves satisfiability. The other outcome is that we reach *saturation*, that is, any further inference would only add clauses already in the theory. In that case, it will turn out, the theory is *satisfiable*.

The single rule of inference we use can be written as follows:

$$\frac{p \vee C \quad \neg p \vee D}{C \vee D}\ \text{resolution}$$

The two premises of the rule are clauses (really: sets of literals), even if we write them using disjunction, so C has no copy of p and D has no copy of $\neg p$.

Thinking of the theory as a sequence of clauses $C_0, \ldots, C_{k-1}$, the rule looks like this:

$$\begin{array}{ll} p \vee C & C_i \quad (i < k) \\ \neg p \vee D & C_j \quad (j < k) \\ \hline C \vee D & C_k = C_i \bowtie_p C_j \end{array}$$

Here we have invented a notation for the justification of the new clause C_k as the result of *resolving* C_i with C_j.

It is easy to see that this rule is *sound* in the sense that it preserves the set of models. To see that, consider an assignment M such that $M \models p \vee C$ and $M \models \neg p \vee D$. We consider two cases:

1. $M(p) = \mathsf{true}$. Then $M \models D$ since $M \models \neg p \vee D$. Therefore $M \models C \vee D$.
2. $M(p) = \mathsf{false}$. Then $M \models C$ since $M \models p \vee C$. Therefore $M \models C \vee D$.

Either way, M is a model of $C \vee D$. Since we just extend the theory, any model of the extended theory will automatically be also a model of the original theory.

If we can obtain the empty clause $\bot$ by repeated application of this rule, we know the original theory cannot be satisfiable. That's because all models are preserved, and the empty clause has no models.

Let's apply resolution in our example. We draw a line here between our original sequence of clauses and the further clauses inferred by resolution. You should check you understand the justification of each new clause.

$$\begin{array}{ll} \neg p \vee \neg q \vee r & C_0 \\ p & C_1 \\ q & C_2 \\ \neg r & C_3 \\ \hline \neg q \vee r & C_4 = C_1 \bowtie_p C_0 \\ r & C_5 = C_2 \bowtie_q C_4 \\ \bot & C_6 = C_5 \bowtie_r C_3 \end{array}$$

Since we have derived the empty clause we know the original theory is unsatisfiable. This in turn means that the original formula we started with $(p \to (q \to r)) \to (p \wedge q \to r)$ is valid.

By the way, one reason to think of the theory as a sequence instead of a tree is that we can reuse intermediate clauses in multiple future inferences. This is not used in this example, but it is a frequent occurrence in realistic examples and can save an exponential amount of space and time.

5 Saturation

When we are not able to reach a contradiction, then by necessity the sequence of clauses must reach *saturation*, that is, any further application of resolution will only lead to clauses already in the sequence. We must reach such a state because if we start with a finite set of clauses they contain only finitely many literals (say n) from which we can form at most 2^n distinct clauses.

As an example of saturation, we try to prove

$$(p \to (q \to r)) \to (p \to r)$$

which is *not valid*. Negating as before, we get the CNF

$$(\neg p \vee \neg q \vee r) \wedge p \wedge \neg r$$

Turning this into a sequence of clauses we get the same as before, except the previous clause C_2 is no longer available.

$\neg p \vee \neg q \vee r$	C_0
p	C_1
$\neg r$	C_2
$\neg q \vee r$	$C_3 = C_1 \bowtie_p C_0$
$\neg q$	$C_4 = C_3 \bowtie_r C_2$
$\neg p \vee \neg q$	$C_5 = C_0 \bowtie_r C_2$

Besides inferences we have already made, we could also try $C_1 \bowtie_p C_5 = \neg q$ but this is equal to C_4. In other words, we have reached saturation without deducing a contradiction and we conclude the initial theory is satisfiable.

But what is the satisfying assignment? Robinson's original paper [Rob65] proves the *completeness* of resolution by constructing a model from a saturated theory. Together with the soundness proof of the last section this means resolution is a sound and complete algorithm for determining satisfiability.

We refer you to the original paper for its correctness, but we show the algorithm. We assume we have an enumeration

$$p_0, p_1, \ldots, p_{n-1}$$

of all the atoms in the saturated theory S. We build up an assignment M by considering each p_i in turn, deciding whether we should set $M(p_i) = \mathsf{true}$ or $M(p_i) = \mathsf{false}$. We use the representation of such a partial assignment as a set of literals containing p_i in the former case and $\neg p_i$ in the latter case.

We use the notation $\overline{M}$ to denote the negation of all the literals in M. One property we need is that if $C \subset \overline{M}$ then $M \not\models C$ because M falsifies all literals in C. Furthermore, no extension of M could possibly satisfy C because the truth value for all literals in C has already been decided.

The basic strategy is to assign truth to an atom unless it is absolutely necessary to assign it false. This would be the case at stage i if there is a clause C in the saturated theory such that $C \subseteq \overline{M \cup \{p\}}$. So in that case we assign p the value false, represented by adding $\neg p$ to the partial model.

Writing out this process more formally:

$$
\begin{array}{lcll}
M_0 & = & \{\,\} & \\
M_{i+1} & = & M_i \cup \{p_i\} & \text{provided there is no } C \in S \text{ s.t. } C \subseteq \overline{M_i \cup \{p_i\}} \\
M_{i+1} & = & M_i \cup \{\neg p_i\} & \text{provided there is a } C \in S \text{ s.t. } C \subseteq \overline{M_i \cup \{p_i\}} \\
 & & & \\
M & = & M_n &
\end{array}
$$

Then $M \models S$, as proved by Robinson.

Let's apply this algorithm to our saturated theory and the order

$$p, q, r$$

so $n = 3$, $p_0 = p$, $p_1 = q$, and $p_2 = r$. For our saturated theory

$$
\begin{array}{ll}
\neg p \vee \neg q \vee r & C_0 \\
p & C_1 \\
\neg r & C_2 \\
\hline
\neg q \vee r & C_3 = C_1 \bowtie_p C_0 \\
\neg q & C_4 = C_3 \bowtie_r C_2 \\
\neg p \vee \neg q & C_5 = C_0 \bowtie_r C_2
\end{array}
$$

we obtain the following sequence:

$$
\begin{array}{lcll}
M_0 & = & \{\,\} & \\
M_1 & = & \{p\} & \text{since there is no } C \subseteq \overline{\{p\}} = \{\neg p\} \\
M_2 & = & \{p, \neg q\} & \text{since } C_4 \subseteq \overline{\{p, q\}} = \{\neg p, \neg q\} \\
M_3 & = & \{p, \neg q, \neg r\} & \text{since } C_2 \subseteq \overline{\{p, \neg q, r\}} = \{\neg p, q, \neg r\}
\end{array}
$$

In this example we can now easily verify that $M_3 \models S$. Since it is a model of every clause in the saturated set, it is of course also a model for the original set of clauses and therefore a counterexample to validity.

6 Checking Certificates

If we are given a resolution refutation (that is, one deducing $\bot$), it is easy to check that all applications of the resolution rule are correct as claimed. We wouldn't even need the intermediate clauses because we can always reconstruct them by calculating $C_i \bowtie_p C_j$. One can also prune some intermediate clauses by working backwards from the contradiction and keeping only those clauses involved in its generation.

If we are given a saturated theory *without* an empty clause, we could easily check that all rule applications are correct and that any further rule applications will not lead to new consequences.

An alternative would be for the prover to apply Robinson's algorithm to construct a satisfying assignment and use that as a small certificate of satisfiability. A checker can just take the original theory T and quickly verify that there is a true literal in each clause. With such a certificate we no longer need the saturated theory at all.

7 Modern Solvers

While resolution is a powerful and elegant method for propositional theorem proving, modern solvers do not implement it directly. We will explore this more in future lectures, but resolution is not entirely absent in the most efficient solvers today.

One of the most important tools in modern solvers is *Boolean constraint propagation* (BCP), also known as *unit propagation*. It is summarized by the following rule, which is just a special case of resolution:

$$\frac{p \vee C \quad \neg p}{C}$$

Consider the following CNF formula:

$$\underbrace{(p_1 \vee \neg p_3 \vee \neg p_5)}_{C_1} \wedge \underbrace{(\neg p_1 \vee p_2)}_{C_2} \wedge \underbrace{(\neg p_1 \vee \neg p_3 \vee p_4)}_{C_3} \wedge \underbrace{(\neg p_1 \vee \neg p_2 \vee p_3)}_{C_5} \wedge \underbrace{(\neg p_4 \vee \neg p_2)}_{C_6} \quad (1)$$

Suppose that `sat` begins by choosing to assign p_1 to true. This leaves us with:

$$\begin{aligned}
&(p_1 \vee \neg p_3 \vee \neg p_5) \wedge (\neg p_1 \vee p_2) \wedge (\neg p_1 \vee \neg p_3 \vee p_4) \wedge (\neg p_1 \vee \neg p_2 \vee p_3) \wedge (\neg p_4 \vee \neg p_2) \\
\leftrightarrow\ &(\top \vee \neg p_3 \vee \neg p_5) \wedge (\bot \vee p_2) \wedge (\bot \vee \neg p_3 \vee p_4) \wedge (\bot \vee \neg p_2 \vee p_3) \wedge (\neg p_4 \vee \neg p_2) \\
\leftrightarrow\ &\top \wedge p_2 \wedge (\neg p_3 \vee p_4) \wedge (\neg p_2 \vee p_3) \wedge (\neg p_4 \vee \neg p_2) \\
\leftrightarrow\ &p_2 \wedge (\neg p_3 \vee p_4) \wedge (\neg p_2 \vee p_3) \wedge (\neg p_4 \vee \neg p_2)
\end{aligned}$$

Notice the clause C_2, which was originally $\neg p_1 \vee p_2$, is now simply p_2. It is obvious that any satisfying interpretation must assign p_2 true, so there is really no choice to make given this formula. We say that p_2 is a *unit literal*, which simply means that it occurs in a clause with no other literals.

We can immediately set p_2 to the value that satisfies its literal, and apply equivalences to remove constants from the formula.

$$\begin{aligned} & \top \wedge (\neg p_3 \vee p_4) \wedge (\neg \top \vee p_3) \wedge (\neg p_4 \vee \neg \top) \\ \leftrightarrow\ & (\neg p_3 \vee p_4) \wedge (\bot \vee p_3) \wedge (\neg p_4 \vee \bot) \\ \leftrightarrow\ & (\neg p_3 \vee p_4) \wedge p_3 \wedge \neg p_4 \end{aligned}$$

After simplifying, we again have two unit literals p_3 and $\neg p_4$. We can continue by picking p_3, assigning it a satisfying value, and simplifying.

$$\begin{aligned} & (\neg \top \vee p_4) \wedge \top \wedge \neg p_4 \\ \leftrightarrow\ & (\bot \vee p_4) \wedge \neg p_4 \\ \leftrightarrow\ & p_4 \wedge \neg p_4 \end{aligned}$$

Now all clauses are unit, and it is clear that if we assign p_1 to true then the resulting formula is not satisfiable. Notice that once we assigned p_1 to true, we were able to determine that the resulting formula was unsatisfiable without making any further decisions. All of the resulting simplifications were a logical consequence of this original choice. The process of carrying this to its conclusion is called *Boolean constraint propagation* (BCP), or sometimes *unit propagation* for short.

References

[Rob65] J. A. Robinson. A machine-oriented logic based on the resolution principle. *Journal of the ACM*, 12(1):23–41, January 1965.