2.7 From DPLL to CDCL

In practice, there are several changes to the procedure:

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

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

Information is reused by conflict analysis and learning.

The branching variable is not chosen randomly.

Under certain circumstances, the procedure is restarted.

Conflict Analysis and Learning

Conflict analysis serves a dual purpose:

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

Learning: By deriving a new clause from the 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 literal:

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

Unit Propagate:

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

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

Decide:

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

if L is undefined under M and contained in N.

Fail:

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

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

Backjump:

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

if there is some "backjump clause" $C \vee L'$ such that
$N \models C \vee L'$,
C is false under M', and
L' is undefined under 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 under M.

There are many possible backjump clauses. One candidate: $\overline{L_1} \vee \ldots \vee \overline{L_n}$, where the L_i are all the decision literals in $M'\ L^{\text{d}}\ M''$. (But usually there are better choices.)

Lemma 2.16 *If we reach a state $M \parallel N$ starting from $\varepsilon \parallel N$, then:*

(1) M does not contain complementary literals.

(2) Every deduced literal L in M follows from N and decision literals occurring before L in M.

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

Lemma 2.17 *Every derivation starting from $\varepsilon \parallel N$ terminates.*

Proof. Let $M \parallel N$ and $M' \parallel N'$ be two CDCL states, such that $M = M_0 L_1^{\mathrm{d}} M_1 \dots L_k^{\mathrm{d}} M_k$ and $M' = M'_0 {L'_1}^{\mathrm{d}} M'_1 \dots {L'_{k'}}^{\mathrm{d}} M'_{k'}$. We define a relation $\succ$ on lists of annotated literals by $M \succ M'$ if and only if

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

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

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 $M \succ M'$. Moreover, the set of propositional variables in N is finite, and each of these variables can occur at most once in a literal list (positively or negatively, with or without a d-superscript). So there are only finitely many literal lists that can occur in a CDCL derivation. Consequently, if there were an infinite CDCL derivation, there would be some cycle $M \parallel N \Rightarrow^{+}_{\mathrm{CDCL}} M \parallel N'$, so by transitivity $M \succ M$, but that would contradict the irreflexivity of $\succ$.

Lemma 2.18 *Suppose that we reach a state $M \parallel N$ starting from $\varepsilon \parallel N$ such that some clause $D \in N$ is false under 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). Since $M \models \neg D$, we obtain, by Lemma 2.16, $N \cup \{L_1, \dots, L_n\} \models \neg D$. Since $D \in N$, this is a contradiction, so $N \cup \{L_1, \dots, L_n\}$ is unsatisfiable. 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 2.19 *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 under M. Hence, in a final state, all literals must be defined. Furthermore, in a final state, no clause in N can be false under 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 under M and M contains no decision literals. By part (2) of Lemma 2.16, every literal in M follows from N. On the other hand, $C \in N$, so N must be unsatisfiable. □

Getting Better Backjump Clauses

Suppose that we have reached a state $M \parallel N$ such that some clause $C \in N$ (or following from N) is false under M.

Consequently, every literal of C is the complement of some literal in M.

(1) If every literal in C is the complement of a decision literal of M, then C is a backjump clause.

(2) Otherwise, $C = C' \vee \overline{L}$, such that L is a deduced literal.

For every deduced literal L, there is a (unit or backjump) clause $D \vee L$, such that $N \models D \vee L$ and D is false under M.

Then $N \models D \vee C'$ and $D \vee C'$ is also false under M. ($D \vee C'$ is a *resolvent* of $C' \vee \overline{L}$ and $D \vee L$.)

By repeating this process, we will eventually obtain a clause that satisfies the requirements of a backjump clause (or the empty clause).

Usually, one resolves the literals in the reverse order in which they were added to M and stops as soon as one obtains a clause in which all but one literal are complements of literals occurring in M before the last decision literal.
$\Rightarrow$ 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.