Eberhard Karls Universität Tübingen
Mathematisch-Naturwissenschaftliche Fakultät
Wilhelm-Schickard-Institut für Informatik

Master Thesis Informatik

Investigations on Backbone Computation

Jonas Bollgrün

September 28, 2019

First Examiner

Prof. Dr. Wolfgang Küchlin
Wilhelm-Schickard-Institut für Informatik
Arbeitsbereich Symbolisches Rechnen
Universität Tübingen

Second Examiner

Prof. Dr. Michael Kaufmann
Wilhelm-Schickard-Institut für Informatik
Arbeitsbereich Algorithmen
Universität Tübingen

Supervisor

Dr. Thore Kübart
Steinbeis-Transferzentrum Objekt-
und Internet-Technologien an der
Universität Tübingen

Bollgrün, Jonas:
Investigations on Backbone Computation
Master Thesis Informatik
Eberhard Karls Universität Tübingen
Processing Period: April 1st - September 30th

Selbstständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Masterarbeit selbständig und nur mit den angegebenen Hilfsmitteln angefertigt habe und dass alle Stellen, die dem Wortlaut oder dem Sinne nach anderen Werken entnommen sind, durch Angaben von Quellen als Entlehnung kenntlich gemacht worden sind. Diese Masterarbeit wurde in gleicher oder ähnlicher Form in keinem anderen Studiengang als Prüfungsleistung vorgelegt.

Jonas Bollgrün (Matrikelnummer 3353424), September 28, 2019

Abstract

This thesis is meant as an overview on the computation of backbones of propositional formulas in conjunctive normal form, being a problem situated in the domain of SAT solving. An introduction into this general domain is given to allow understanding without deep prior knowledge of the field. Various algorithms, which are grouped into approaches on the problem, are listed and subsequently also methods on how to improve these algorithms. A special emphasis is given on the concept of preferences, where underlying SAT solving algorithms are guided in their search to speed up the computation of backbones. Various methods are proposed to resolve performance issues of the preferences approach and subsequently tested on multiple benchmarks of varying difficulty to verify their effectiveness. In closing, using an in-depth analysis of the results, specialized backbone algorithms for some of the same benchmarks are constructed.

Contents

List of Tables

List of Algorithms

1 Introduction

Imagine a discussion about a certain issue where the case gets complicated by the fact that the two parties cannot even agree on factual real world data, but can at least both agree on logical relationships between certain hypothetical concepts that may or may not be true. In this case, you have the opportunity to set up a boolean formula that expresses the issue, with one all deciding boolean variable that determines which of the parties is right. If one of the parties should be lucky, it will turn out that it would actually be impossible that that formula would judge in favor of their opponent, no matter what the facts were, as long as the formula itself was correctly stated and applied. For this you would calculate the backbone of the formula, the set of conclusions that lie in the formula itself and will not change whatever the facts were. Then you would simply check whether it contained the all deciding variable with either positive or negative assignment.

Now how would you go about calculating such a backbone. Many great papers have been written on the topic, for example [MSJL15], which discusses the common approaches exhaustively. Typically, you would collect various solutions of the formula and check for parts that always look the same and if you can't get them to work another way, then it must be part of the backbone. However an original thought not discussed there can be found much earlier in [KK01]:

> Obviously, the effect of filtering depends on the quality of the results returned by the SAT algorithm whenever the examined formula is satisfiable. [...] In order to maximize the number of variables for which this condition holds, we use a corresponding variable selection strategy in the underlying SAT checker [...] Variables that have not been classified at all have the highest priority, followed by variables that are not known as not necessary, and finally by those not recognized as not inadmissible.

At that time, this approach faired very well in the performance tests listed in the same paper, but the idea seems to have been forgotten. Over 15 years later, in [PJ18], where a very similar method is described, it is stated that...

> [...]to the best of our understanding the use of SAT with preferences has not been previously proposed in the context of backbone computation.

For this reason, I decided to reexamine this peculiar concept in this master's thesis, along other promising methods. The remainder of this chapter contains a primer on elementary knowledge of SAT solving which necessary to understand backbone computation. This includes an algorithm to find solutions to boolean formulas, a functionality that almost always lies at the heart of backbone computation.

Subsequently, in the second chapter multiple algorithms for backbone computation are introduced and grouped according to certain characteristics of them. The third chapter describes supplementary methods, that can be applied ontop of base algorithms to fine-tune their behaviour.
The fourth chapter is dedicated to experimentation on the concept of preference guided backbone computation. It also contains information for the purpose of applying the knowledge found in this thesis in the *Sat4J* library.
Finally all of these methods are tested on multiple benchmarks in chapter 5. Building on the findings about individual strategies, the final goal is to find the optimal strategy for each of the benchmarks.

1.1 Disambiguation

1.1.1 Terminology

This thesis is an investigation on the calculation of backbones for boolean formulas in conjunctive normal form (or CNF in short). A CNF formula F is a conjunction of a set of clauses $C(F)$, meaning that all of these clauses have to be satisfied to satisfy the formula. A clause c in turn is a disjunction of a set of literals, meaning at least one of said literals must be fulfilled. A literal l can be defined as the occurence of a boolean variable v which may or not be negated and to fulfill such a literal, its variable must be assigned $\bot$ for negated literals and $\top$ for those literals without negation. The same variable can occur in multiple clauses of the same formula but must have the same assignment in all occurences, either ($\bot$) or ($\top$). A complete assignment of all variables of F (written as $Var(F)$) that leads to the formula being fulfilled is called a model. A formula for which no model can be found is called unsatisfiable. Any set of assignments that is sufficient to satisfy a formula is called an implicant and if has no subset that isn't itself an implicant it is called a primeimplicant. The variables that do not occur in an implicant are called optional.

When we want to know whether a model M satisfies a formula F, clause c or literal l, we write $F\langle M\rangle \rightarrow \{\top, \bot\}$ or $c\langle M\rangle$ and $l\langle M\rangle$ respectively. The result is $\top$ if the assignment satisfies what it is applied to or $\bot$ if it doesn't. TODO mehr benutzen oder weg

The exact terminology can differ depending on the paper and project. *Formula* may be called a *problem* and the assignment of a variable might be called its *phase*. Sometimes *assignment* and *literal* are used interchangeably, as they both consist of a variable and a boolean value. *Clauses* can also be called *constraints* and sometimes *sentences*. A synonym for a formula, clause or literal being $fulfilled$ is it being $satisfied$. *Models* can also be called *solutions* of formulas. The terminology for $\top$ or $\bot$ can be (T,F),$(true, false)$ or $(1,0)$.

The backbone is a formula specific set of literals that contains all literals that occur in every model of said formula. We can also say that a variable is not part of the backbone, if neither its positive or its negative assignment is in the backbone. If we

have an unsatisfiable formula, its backbone can be considered undefinable, which is why this thesis concerns itself only with satisfiable CNF formulas.

For the context of CNF formulas, on which this thesis focuses, the term "subsumption" should to be explained. A clause c_1 subsumes another clause c_2 of the same formula, if and only if $c_1 \subseteq c_2$, in prose if all the literals that occur in c_1 also occur in c_2[1]. If c_1 subsumes c_2 in formula F then this means that you can remove c_2 from F because in terms of satisfying models, $F \backslash \{c_2\}$ is equivalent to F as $\{c_1\}$ is equivalent to $\{c_1, c_2\}$. This is because in this case an assignment that satisfies c_1 also satisfies c_2 automatically. There is no possible assignment that satisfies clause c_1 that doesn't satisfy its subsuming clause c_2.

A program that determines whether a boolean formula is satisfiable or not is called a *SAT solver*. Typically this is done by finding a model for said formula.

1.1.2 CDCL

All the methods to generate a backbone of a formula F that are described in this thesis essentially rely on calculating various models of F, so it makes sense to describe a method to do that in depth. The current state-of-the-art algorithm to do this is the *Conflict Driven Clause Learning* algorithm, or *CDCL* for short. This algorithm was first published as "*GRASP*" in [SS96]. An earlier algorithm for the SAT problem, which *CDCL* is essentially based upon is the *DPLL* algorithm [DLL62] published in 1962. However for this thesis I worked with the *CDCL* implementation available in the *Sat4J* library that is heavily based on *MiniSAT* which can be read about in [ES03a]. Understanding of the employed SAT solver will become useful in later chapters, when I explain backbone computation strategies based on it.

Storing solver state

In this algorithm, a so-called *CDCL table* is used as a special dataset to document how and why variables were assigned to either $\top$ or $\bot$. This table, which behaves like a stack, stores the succession of assignments with four values for each assignment.

- The affected variable.
- The value that the variable was assigned to.
- The reason for the assignment. This can be one of two cases, either *Unit* or *Decision*.

[1] One can filibuster whether c_1 would have to be a true subset of c_2. If a formula has two occurences of the exact same clause, then one of the occurences would be redundant, so the same rule could be applied here as well. In practice it makes more sense to filter out duplicates of clauses before running any computation on the formula. Similarly, you can safely drop all clauses where both literals of the same variable occur, as that clause would be satisfied in each and every possible model.

  - *Unit* assignments happen, when a clause has all but one of its literals unsatisfied. Since all clauses have to be satisfied for a CNF formula to be satisfied, that last literal must be assigned a value that satisfies it and its clause. Entries in the CDCL table that refer to a unit assignment also store a reference to the clause that required the assignment. A clause that fulfills the above condition and requires an assignment can be called a *unit clause* or that it *is unit*.
  - *Decisions* happen when no clause is in the unit state. In theory, in this case you are free to pick any variable and assign it either $\top$ or $\bot$.
- The *level* of the assignment. This level increases with each decision and starts at 0, where unit assignments before any decision are stored.

Reducing search space

The purpose of unit assignments is to reduce the amount of futile computational effort. The solving process for a formula can be modeled as the traversal of a search tree, where each node corresponds to an individual assignment and every leaf node to a complete assignment that might be satisfying or not[2]. However, given that you can stop to search once a single clause is unsatisfied, you can disqualify many branches of the tree early, for example when the assignments in your tree path so far require some additional assignment for some clause, which would be a *unit clause*. Going the other way in the tree at that particular node will never result in a satisfying model.

The solver will now fill the table with assignments, unit assignments if possible or decisions otherwise, until one of two things happens. Either the formula becomes satisfied[3] in which case the assignments that are stored in the CDCL table are returned as a model.

Learning from conflicts

The other possibility is that you run into a contradiction. Here a unit clause requires that a variable is assigned a value b, but it is already assigned $\neg b$. If the assignments have a reason (a unit clause), then some of the variables in these clauses must be assigned differently (since they were

F:= {(¬c),(a,e),(b,¬d),(b,c,d)}

Level	Var	Value	Reason
0	c	⊥	{¬c}
1	a	⊥	Decision
1	e	T	{a,e}
2	b	⊥	Decision
2	d	⊥	{b,¬d}
2	d	T	{b,c,d}

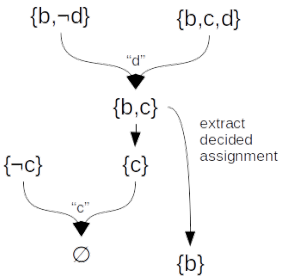

Figure 1.1: Example of a resolution. Shown is a formula, the content of the CDCL table at the point of conflict in variable d and the resolution graph. The resulting learned clause is $\{b\}$.

[2] Sadly the size of such a tree makes it impractical to actually implement SAT solving like this.

[3] In this case only an implicant is returned. If you want a complete model, you can keep making decisions until all variables are assigned and return the assignments after that. However once you have an implicant, you are free to assign the remaining variables to anything you want, as the formula is already satisfied and further assignments cannot change that.

unit then), so we can connect the reasons for our conflict to other assignments. Repeating this we sometimes meet decisions instead of unit clauses as reason. Collecting these decisions, we end with the precise combination of assignment decisions that resulted in the conflict.

This process of following assignments through their reasons is called *resolution* and can be briefly explained with the following formula:

$$r_\lambda(c_1, c_2) = (c_1 \setminus \{\lambda\}) \cup (c_2 \setminus \{\neg\lambda\})$$

The resolvent r consists of all literals of the two clauses that are resolved, where in this case c_1 contains λ and c_2 contains $\neg\lambda$. This precondition is given in this case, because c_1 is the reason for the assignment of λ to $\top$ and c_2 the reason for assigning λ to $\bot$.

Figure 1.1 shows this with an example, where a conflict in variable d happened. First the clauses $\{b, \neg d\}$ and $\{c, b, d\}$ are resolved over the variable d. Then b is extracted for the learned clause, as its reason was a decision and not a unit clause. Finally $\{c\}$ and $\{\neg c\}$ are resolved over c, resulting in the empty set. The learned clause is $\{b\}$. An alternative resolution scheme would leave literals that stem from decisions in the working set until the end, but here they are extracted on first extraction. Otherwise, these literals would be inspected multiple times.

Once collected, this set of decisions must be taken back, by reversing the assignments up until the first of these decisions, as this path through the search tree results in a conflict and will not end in models.

We also add the clause to our formula as a *learned clause*. This clause serves to prevent the particular combination of assignments that led to our conflict. The resulting formula will still be completely equivalent to before. It merely stores the information that we gained through analysis of our conflict to prevent it from happening again.

However it is also possible that we end up with reasons for assignments that do not end in a decision. This would mean that all reasons were axiomatic assignments, for example when a formula contains a clause with only a single literal in it[4] . In this case the formula implies a contradiction and the clause that we would learn from this would be empty, unsatisfiable.

Decision heuristics

Concerning the decisions, depending on the particular formula, it is possible that some assignments make it easier to solve the formula than others and some decisions

[4] A simple example: A formula contains the two clauses $\{a\}$ and $\{\neg a\}$. Resolving these two results in the empty set.

Algorithm 1: CDCL ALGORITHM

Input: A formula F in CNF
Output: A CDCL table which implies an implicant for F, or $\bot$ if F is not satisfiable

```
1  level ← 0
2  table ← emptyList
3  while 1 do
4      table.pushAll(F.getUnits())
5      if ∃ conflicting assignment then
6          if level = 0 then
7              return ⊥
8          else
9              level ← backtrackAndLearn(F, table)
10     else if F is fulfilled then
11         return table
12     else
13         level++
14         l ← any free variable
15         l.assign(either ⊤ or ⊥)
16         table.pushDecision(l)
```

might lead to a completely unsatisfied clause. To prevent this, there exists a range of heuristics that try to pick a variable and corresponding assignment that would lead to a satisfying model without complications. A typical one would be to measure the activity of a variable, which is an estimate of how often it was involved in conflicts. The heuristic would then always pick the variable with the lowest activity, since it would most likely be unproblematic to assign one or the other value to it. To choose the boolean value that should be assigned to it, a good strategy can be to remember the value that each variable was assigned to on the last occasion and this time give it the opposite one. When a variables assignment is removed through conflict resolution, going in the complete opposite direction with the assignments has a good chance to work out better.

Notation in algorithms

As is usually done in literature, calls to SAT solvers such as CDCL in code listings are written as $(outc, v) = SAT(F)$. Here, two values are returned. $outc$ is a boolean value that simply states whether F turned out to be satisfiable or not, the result of which is often written as SAT and $UNSAT$. Only if $outc$ equals to SAT, the second return parameter v can have a meaningful value, which would be the model that was found and satisfies F. In some of the algorithms listed in this thesis, one of the return parameters is not used at all. In that case I write $(\_, v) = SAT(F)$ or $(outc, \_) = SAT(F)$ to indicate that either $outc$ or v is discarded.

2 Base Algorithms

With a method to compute satisfying models explained, I will now go on to explain algorithms to compute backbones. The ones that I investigated for this thesis can be grouped very broadly into two approaches, which I will describe in the following two sections.

2.1 Enumeration algorithms

2.1.1 Model enumeration

A simple definition of the backbone is that it is the intersection of all models of it's formula. If a literal is not part of the backbone, there must exist a model that contains the negation of that literal. Therefore if we had a way to iterate over every single model of the formula and, starting with the set of both literals for every variable and removing every literal from that set that was missing in one of these models, that set would end up being the backbone of the formula. [MSJL10] as well as [MSJL15] list an algorithm that does exactly this.

Algorithm 2: Enumeration-based backbone computation

Input: A satisfiable formula F
Output: Backbone of F, ν_r

$\nu_r \leftarrow \{x | x \in Var(F)\} \cup \{\neg x | x \in Var(F)\}$
while $\nu_r \neq \emptyset$ **do**
  $(outc, \nu) \leftarrow SAT(F)$
  if $outc = \bot$ **then**
    return ν_r
  $\nu_r \leftarrow \nu_r \cap \nu$
  $\omega_B \leftarrow \bigvee_{l \in \nu} \neg l$
  $F \leftarrow F \cup \omega_B$
return ν_r

Here, found models are prevented from being found again by adding a blocking clause of said model and the algorithm terminates once all models are prohibited and the formula became unsatisfiable through this.

2.1.2 Upper bound reduction

Clearly, calculating every single model of a formula leaves room for optimization. Most models of a common boolean formula differ by small, independent differences that can just as well occur in other models. Therefore the intersection of only a handful of models can suffice to result in the backbone, as long as these models are chosen to be as different as possible. This was achieved in [MSJL15] as is described in algorithm 3.

Algorithm 3: Iterative algorithm with complement of backbone estimate

Input: A satisfiable formula F
Output: Backbone of F, v_r

```
1 (outc, v_r) ← SAT(F)
2 while v_r ≠ ∅ do
3     bc ← ⋁_{l∈v_r} ¬l
4     (outc, v) ← SAT(F ∪ {bc})
5     if outc = ⊥ then
6         return v_r
7     v_r ← v_r ∩ v
8 return v_r
```

It generates an upper bound v_r of the backbone by intersecting found models and inhibits this upper bound instead of individual models. This blocking clause is much more powerful, because it enforces not only that a new model is found, but also that this new model will reduce the upper bound estimation of the backbone in each iteration.

This is because what remains after the intersection of a handful of models, are the assignments that were the same in all these models and from that we make a blocking clause that prohibits the next model to contain that particular combination of assignments. The next model will then have to be different from all previous models for at least one of the variables in the blocking clause to satisfy it.

Eventually v_r will be reduced to the backbone. This can be easily recognized, because the blocking clause of the backbone or any of it's subsets makes the formula unsatisfiable, except in the case that the formulas backbone would be empty.

Note that it is not particularly important for the algorithm whether the blocking clauses remain in F or get replaced by the next blocking clause, because the new blocking clause bc_{i+1} always subsumes the previous one bc_i, meaning that every solution that is prohibited by bc_{i+1} is also prohibited by bc_i and $F \cup \{bc_i, bc_{i+1}\}$ is equivalent to $F \cup \{bc_{i+1}\}$ concerning the set of models.

This algorithm is implemented in the *Sat4J* library under the designation *IBB*, except that prime implicants are used instead of models. For specifics on these, see chapter 3.1.1.

2.2 Iterative algorithms

2.2.1 Testing every literal

Alternatively, you can define the backbone as all literals that occur with the same assignment in all models of it's problem, which implies that enforcing a backbone literal to it's negation should make the formula unsatisfiable. This definition already leads to a simple algorithm that can calculate the backbone, by checking both assignments of every literal for whether it would make the formula unsatisfiable, see algorithm 4. This algorithm is referenced in [MSJL10]

Algorithm 4: Iterative algorithm (two tests per variable)

Input: A satisfiable formula F in CNF
Output: All literals of the backbone of F v_r

```
1  v_r ← ∅
2  for x ∈ Var(F) do
3      (outc_1, _) ← SAT(F ∪ {x})
4      (outc_2, _) ← SAT(F ∪ {¬x})
5      assert(outc_1 = ⊤ ∨ outc_2 = ⊤)  // Otherwise F would be unsatisfiable
6      if outc_1 = ⊥ then
7          v_r = v_r ∪ {¬x}
8          F = F ∪ {¬x}
9      else if outc_2 = ⊥ then
10         v_r = v_r ∪ {x}
11         F = F ∪ {x}
12 return v_r
```

The two calls to the SAT function return a pair which consists first of whether the given function was satisfiable at all and, secondly, the found model, which in this case is discarded. There is no good algorithm that can tell whether a boolean formula is satisfiable or not without trying to find a model for said formula, but we can use it to greatly improve the algorithm above by combining this approach with that of the enumeration algorithms.

2.2.2 Combining with enumeration

First observe that any model of F would already reduce the set of literals to test by half, because for every assignment missing in the model, we know that it cannot be part of the backbone, so there is no need to test it.

This can be repeated with every further model that we find. The following algorithm 5 is another one that is listed in both [MSJL10] and [MSJL15] and is implemented in the *Sat4J* library as BB[1].

[1] With the exception of using prime implicants instead of models, similar to the *IBB* algorithm.

Algorithm 5: Iterative algorithm (one test per variable)

Input: A satisfiable formula F in CNF
Output: All literals of the backbone of F v_r

```
1  (outc, v) ← SAT(F)
2  Λ ← v
3  v_r ← ∅
4  while Λ ≠ ∅ do
5      l ← pick any literal from Λ
6      (outc, v) ← SAT(F ∪ {¬l})
7      if outc = ⊥ then
8          v_r ← v_r ∪ {l}
9          Λ ← Λ\{l}
10         F ← F ∪ {l}
11     else
12         Λ ← Λ ∩ v
13 return v_r
```

Note that both possible results of the call to the SAT solver are converted to useful information. In the else branch, the formula together with the blocked literal l was still solvable. In this case v is still a valid model for F, so we can search through it to look for more assignments that don't need to be checked. Note that here v must contain $\neg l$, as it was enforced. Therefore l will be removed from Λ in this case as well.

In the other case, we identified l as a backbone literal. In that case it will be added to the returned set, removed from the set of literals to test and, lastly, added to the problem F, which increases performance in subsequent solving steps. However it would be even better, not only to reuse the learned backbone literals, but all learned clauses.

It is possible to apply the concept of preferences described in section 2.3 to the iterative algorithms listed here. However, this is less beneficial than adding it to the enumeration approach of *IBB*, because in the *BB* algorithm many SAT calls are supposed to return *UNSAT* to positively identify an assignment as backbone. However, this case does not give us a model, so the extra effort of trying to find a more valuable model is often wasted here.

Another difficulty with the *BB* algorithm in comparison to *IBB* is that here it is more difficult to reuse learned clauses. These are based on a subset of the clauses in F. Their existence is virtual so to speak, implied through these base clauses but difficult to find. Adding more clauses to the formula does not remove a learned clause, as the set of base clauses is untouched. At most it might be possible to subsume it with another learned clause. However if a clause is removed from F, it might be that the learned clause is no longer implied through the formula, therefore making it invalid.

Example: Using the iterative approach on formula F which contains the clause $\{(a,b,c)\}$ we temporarily add the blocking clause $(\neg a)$. In this case, the clause (b,c) is implied. CDCL could learn this new clause by making the resolution of $\{(a,b,c)\}$ with the blocking clause $\{\neg a\}$. Should we now change the blocking clause, this learned clause must be discarded. Otherwise when we test with a different blocking clause $(\neg b)$, together with (b,c) it would imply (c), making c a backbone of F, which it clearly isn't.

As a sidenote, the general approach of the *BB* algorithm can be extended to test multiple backbone literals simultaneously in one single SAT call with a concept called "Chunking". Whereas in algorithm 5 you would pass only one possible backbone literal in negation, here you would instead pass a clause that contained multiple negated candidates. To make the formula unsatisfiable, each literal of that clause would have to be impossible to satisfy in any way, meaning that all of them are negated backbone literals. The downside is that in the other case, you still don't know whether an individual literal of that temporarily added clause could be in the backbone after all, so you might end up having to make additional tests. However, this thesis contains no experimental results for this concept, because it was not supplied in the *Sat*4*J* library and I focused more on the concept of preferences as described in the next section.

2.3 Preferences

The previous approaches still leave much of their efficiency to chance. Theoretically the solver might return models with only the slightest differences from each other, when other models could reduce the set of backbone candidates much more. For example the blocking clause can be satisfied with only one literal in it being satisfied, but if we were to find a model that satisfies all literals in the blocking clause, we can immediately tell that the backbone is empty and we would be finished. So it would be a good approach for backbone computation if we could direct our SAT solver to generate models that disprove as many of the literals in the blocking clause as possible. Precisely this has been described by [PJ18], but has also been proposed much earlier by [KK01].

[PJ18] describe an algorithm called $BB-PREF$ or *Prefbones*, which makes use of a slightly modified SAT solver based on CDCL, which is called $prefSAT$ in the algorithm below. It can be configured with a set of preferred literals $prefs$. As already stated, when the CDCL algorithm reaches the point where it has the freedom to decide the assignment of a variable, it consults a heuristic that tries to predict the best choice of variable and assignment to reach a model. Instead, $prefSAT$ uses two separate instances of these heuristics h_{pref} and h_{tail}, which by themselves may work just as the single heuristic used in the ordinary CDCL solver. The key difference in $prefSAT$ is, that h_{pref}, which contains the literals in $prefs$, is consulted first for decisions, and only when all variables with a preferred assignment are already assigned, h_{tail} is used to pick the most agreeable literal, meaning that it would most

likely not result in a conflict.

Algorithm 6: BB-pref: Backbone computation using pref-SAT

Input: A satisfiable formula F in CNF
Output: All literals of the backbone of F, v_r

```
1 (_, v_r) ← SAT(F)
2 Repeat
3     prefs ← {¬l : l ∈ v_r}
4     (_, v) ← prefSAT(F, prefs)
5     if v ⊇ v_r then
6         return v_r                    // No preference was applied
7     v_r ← v_r ∩ v
```

This algorithm also differs from *IBB* in that it does not add blocking clauses, and that is also why it cannot use the case when F becomes unsatisfiable to terminate the algorithm. Instead it relies on the preferences to be taken into account. Except for the case where a formula has only one singular and immediately obvious model, CDCL must make at least one decision. That decision must come from h_{pref}, except for the case that CDCL learned axiomatic assignments for all variables in $prefs$. Depending on whether the learned value for the variables in $prefs$ contradicts all preferences it may take another call to $prefSAT$, but at the latest then no more changes will happen to v_r and the algorithm terminates. The return condition also covers the case when the backbone turns out to be empty, because then v_r was reduced to $\emptyset$ and that is a subset of every set. This algorithm stands out from all the others because it has the unique property, that the formula of which the backbone is calculated is never modified in any way, not even temporarily. Later in section 5.2.4, I will point out how this trait can be useful.

Note that the algorithm was written slightly different from what is listed in [PJ18] to make the relation with common enumeration algorithms more apparent and also make it easier to read. For the purpose of this thesis, it is further called *PB*0.

The return condition makes this algorithm inflexible, as the preferences have to be taken into account without exception. If not, a model might be returned that terminates the algorithm prematurely, because it did not properly test a variable assignment, instead taking a shortcut to save time in the calculation of a model. The purpose of this thesis was to experiment with solvers and I was interested in the concrete effects of preferences by themselves on the backbone computation. This is why I created a variation of Prefbones, designated *PB*1, that uses the previous approach of upper bound reduction, adding a blocking clause to F in every iteration and terminating when F would become unsatisfiable, see algorithm 7. This made the preferences algorithmically completely optional and allowed to experiment with many variations on the concept. Coincidentally, the added blocking clause happens to be the exact same set as that of the preferred variables.

Later in the results section I will show how this variant faired against the previous

Algorithm 7: BB-pref: Backbone computation using pref-SAT and blocking clauses

Input: A formula F in CNF
Output: All literals of the backbone of F, v_r

```
1 (_, v_r) ← SAT(F)
2 Repeat
3     bc ← ⋁_{l∈v_r} ¬l
4     F ← F ∪ {bc}
5     prefs ← {¬l : l ∈ v_r}
6     (outc, v) ← prefSAT(F, prefs)
7     if outc = ⊥ then
8         return v_r
9     v_r ← v_r ∩ v
```

algorithm from [PJ18].

3 Optimizations

The following chapter elaborates on methods to enhance the algorithms that were described in the previous chapter. Depending on the particular combination of algorithm and enhancement, applying the optimization can be be considered a no brainer. However, experimental results show that this is not true without exception and in individual cases we will give thoughts why that is. Other improvements can only be applied to some algorithms due to the data that is available.

3.1 Model reduction

The algorithms described in section 2 all boil down to testing for each variable whether there exist two models where one assigns the variable to $\top$ and the other to $\bot$. This section describes two strategies to reduce the model that is returned by the SAT solver to a subset that tells us more for the purpose of calculating a backbone. Both methods are essentially about using implicants instead of models.

Having a single implicant I that leaves some variables optional immediately tells us, that every possible combination of assignments of the optional literals can make a model, if we just add the assigned literals in I. You could say that an implicant implies a large set of models.

Without any further information, a single implicant I tells us, that every one of the optional variables in it cannot be part of the backbone.

Starting with complete models, implicants can be subsets of other implicants, by removing more and more assignments that are not essential. This way you will eventually reach an implicant where all of it's assignments are required and removing any further literal from it, would leave some clause unsatisfied. This would be called a prime implicant I_π.

3.1.1 Prime implicant

[DFLBM13] describes an algorithm that allows to calculate the prime implicant from a model in linear time over the number of literal occurrences in the formula. This algorithm works best if you generate the model of a variable with the CDCL algorithm for multiple reasons.

First, it takes advantage of data structures that you also need in a good implementation of CDCL, namely a lookup from each variable to all clauses that contain either literal of that variable. In CDCL this lookup table improves the performance of unit propagation, because you can check exactly the set of clauses that might be affected

by the assignment to determine whether one of the clauses has become exhaustively unsatisfied or implies another assignment[1]. Here, the lookup is used to determine whether a literal is required in the implicant that you are in the process of generating, by looking for clauses that only contain a single satisfying literal anymore, which then must be part of the prime implicant. You can even reuse the watched literals[2] of the clauses, however you would have to change the way in which the propagate, since you take assignments away instead of adding them.

The second fit with CDCL is that it not only generates a model, but also a table containing information on how that model was generated. This information can be used in this algorithm to reduce the number of literals that need to be checked on whether they are required in the resulting prime implicant. You can quickly generate the input I_r by going through the table generated by CDCL and noting down every assignment that happened through unit propagation. These must be part of the prime implicant because to have been assigned through unit propagation at some point in time there must have been a clause that required that particular assignment.

The only assignments that you really have to test here are those that were decisions. Additionally, you can avoid many decided assignments if you configure your CDCL SAT solver to stop once the formula is satisfied instead of stopping once every variable was assigned. Once the formula is satisfied, no more assignment is necessary, so all further ones are arbitrary decisions that are not necessary in the implicant.

It is sufficient to go over the remaining set of assignments only once each. After having determined that a literal is required to satisfy the formula, it cannot become optional later. After all, for this the algorithm would have to add assignments instead of taking them away. And after having discarded a literal from our implicant it cannot become required again. For this to happen we would have to drop a literal from our implicant where that was the last one that satisfied some particular clause. This fits the description of a required literal and we do not drop those.

Algorithm 8: Base approach to compute a prime implicant

Input: A formula F, a model I_m, I_r containing some required literals in I_m
Output: I_r, reduced to a primeimplicant of F, being a subset of I_m

```
1 while ∃l ∈ I_m\I_r do
2     if ∃c ∈ F : Req(I_m, l, c) then
3         I_r ← I_r ∪ {l}
4     else
5         I_m ← I_m\{l}
6 return I_r
```

[1] The unit case

[2] These are used in CDCL to determine and store the information on whether a clause is satisfied, unsatisfied, unit or neither of them. This is done by having two pointers that rest on unassigned literals in each clause. In this algorithm you can let these pointers rest on satisfying literals.

Since the same model could be calculated in different ways by CDCL, depending on the particular order in which the variables in it were assigned values, the partition between decided assignments and unit assignments can be different for the same model. Therefore, this method can give you different prime implicants for the same model depending on the CDCL table that you use. Additionally, the order in which you check the decided literals can also make a difference.

The function $Req(I_m, l, c)$ in line 2 of algorithm 8 tells, whether the assignment l in the implicant I_m is required to satisfy c. In other words, is l the only literal in c that also occurs in I_m.

3.1.2 Rotations

There is a model reduction method that is even more powerful than calculating the prime implicant. Even better, the concept is much simpler than calculating the prime implicant.

Any model of a CNF formula can contain multiple implicants and even prime implicants[3] , so if we could find out from only a single model M_0 of F, which of these literals occured in all implicants that that model covered, with each model we could reduce the set of backbone candidates even more than with just one of it's prime implicants.

Doing this is actually pretty simple. Instead of generating a small set of various prime implicants, you check for each assignment a in M_0, whether it is required in M_0, by checking whether $(M_0 \backslash a)$ is still an implicant that satisfies F. If so, we know that a is not part of the backbone, because we found an implicant without it. This approach was described in [MSJL15].

You can do this reduction faster if you make use of the lookup table described in the previous section, where each literal is mapped to the set of clauses where it occurs. After all, you already know that M_0 satisfies F, so it is sufficient to check only the clauses that contain the tested literal a, whether they are satisfied in a different way in $M_0 \backslash \{a\}$. Unaffected clauses must still be satisfied without a.

Additionally, if you happen to use a computation strategy, where you have a set of positively identifed backbone literals (for example an iterative algorithm as described in section 2.2, or when you make use of axiomatic literals as described in the next section), then of course these don't have to be tested either.

The performance benefit of this approach depends on the particular example. If the individual models of a formula contain very few primeimplicants or even only one, then looking for rotatable literals might not give a benefit over calculating the prime implicants. However formulas where individual models contain many prime implicants with large differences, the benefit can be extreme. See section 5 for experimental results.

[3] As an example imagine any formula with only clauses of size two, where the first literals are disjunct from the second literals. This formula has at least two prime implicants I_1 and I_2 that you can generate by collecting either the first literal of every clause for I_1 or the second literal for I_2. Since the literals in I_1 are disjunct from those in I_2, you can build a model of the combination of I_1 and I_2.

Algorithm 9: Literal rotation in models

Input: A formula F, a model M
Output: R, the required literals of all implicants in M

```
1 R ← ∅
2 for a ∈ M do
3     I_a ← M\a
4     if ∃c ∈ F : c⟨I_a⟩ = ⊥ then
5         R ← R ∪ {a}
6 return R
```

3.2 Cheap identification of backbone literals

This section describes various ways that allow you to recognize backbone literals without an additional satisfiability check. Knowing these backbone literals early can speed up the individual calls to the SAT solver, because enforcing the backbone literals prevents the solver from trying to find solutions containing the negation of backbone literals, which by definition don't exist. Specifically for algorithms that apply preferences, knowing backbone literals explicitely can be helpful, because then you can remove some of those preferences that go against the backbone and could never be satisfied anyway.

3.2.1 Learned literals

The most straightforward method to quickly identify backbone literals is to scan the formula for clauses with only a single literal. Since these clauses have only one possible way to become satisfied, that assignment must be used in every model of the formula and is therefore backbone.

Additionally you should check the CDCL table that your solver creates. If you look at this table you might find variable assignments through unit implication which happened before any decision. This includes all assignments from the paragraph above, but also those that are implied through these[4].

It makes sense to do this check after every SAT computation, as every different way that leads to a different model brings different learned clauses, that may sometimes consist of a single literal. The number of these clauses depends on the structure of your formula. If it is easy to solve, very few conflicts will occur and in turn, very few clauses will be learned.

An expansion on this would be to look for pairs of clauses (a, b), $(a, \neg b)$ for any two variables of the formula. The only way to fulfill this pair of clauses is to assign a to $\top$[5]. This scheme can theoretically be applied to any clause size, but then you would require a quadratically increasing number of clauses to determine a backbone which

[4] Example: Formula $\{\{a\}, \{\neg a, b\}\}$ has the backbone $\{a, b\}$, but only a is immediately obvious.
[5] Which fits the resolution formula described in section 1.1.2

would first increase computation time and secondly decrease the chance that the necessary set of clauses was available.

However, if you employ a backbone algorithm that adds clauses during its course, then learned literals may not be guaranteed to be valid backbone literals, just as learned clauses may not be valid beyond the running backbone computation[6]. In theory, you could track for each learned clause which original clauses were used to create it. This way, you could tell for each individual learned clause (and literal), whether it would also be valid for the original formula. However, as far I investigated it, *Sat4J* does not support this feature. This means that the *PB*1 and *IBB* algorithms should not be combined with this optimization, except for their very first SAT call since it is done before a blocking clause is added. It also means that the same algorithms have to discard all their learned literals and clauses together with the inserted blocking clauses before they return.

The other algorithms, *BB* and *PB*0 on the other hand do not have this issue. *PB*0 does not modify the formula at all and the *BB* algorithm can use a different mechanism for its temporary modifications which is to solve it under *assumptions*. This means that you start the SAT computation by putting decisions that correspond to the assignments that you want to enforce at the very beginning of the CDCL table. Then once you have to backtrack these decisions, you know that your formula is unsatisfiable with your assumptions. The difference that this makes regarding learned clauses is that the content of them will differ. When implemented as temporary clauses, your assumption is taken as ground truth, as part of your formula. The resolution step for the variable associated with the assumption will remove it from the learned clause. Without the clause that implemented the assumption, that learned clause would be too strong.
However as decisions, where this assumption is involved with a conflict a corresponding literal will simply be added to the learned clause. That way the learned clause will later only come into effect under the circumstance that the variable that has now an assumption is then assigned the same value.

For this reason, both *BB* and *PB*0 are also able to keep their learned clauses after a backbone computation, even under assumptions. This will become relevant later in the *Results* section.

3.2.2 Unit implication

When you have some backbone literals identified, there are some methods that you can apply on this set, which can potentially expand it without a complete model calculation.

One method becomes obvious, if you recall the CDCL algorithm, specifically unit propagation. Suppose you have a clause where all but one of its variables turned out to be part of the backbone, however all of them with the exact opposite sign from

[6] A simple example: You find an implicant $\{a\}$ with only a single literal. Then you add the blocking clause $\{\neg a\}$. But in $F \cup \{\neg a\}$ $\neg a$ is suddenly a backbone literal.

that in the clause, so that the clause is still not satisfied by them. In this case you are forced to assign the remaining literal in a way that it does satisfy the clause and you have to do this in every possible model, since what forces you to do that are backbone literals. Therefore, this unit implied assignment is in the backbone.

An efficient algorithm for this would be as follows: Keep a counter for each of the clauses in your formula that indicates the number of not satisfying backbone literals in said clause. When you identify a new backbone literal, increment all the counters where that literal occurs in negation. You can drop the counter completely for clauses where the newly identified backbone literal occurs with the same sign so that it satisfies the clause[7]. If the counter reaches the length of its clause minus one, then you know that to satisfy this particular clause, the remaining literal in it has to be in the backbone. In this case you can compare its literals with your current inventory of backbone literals to identify the remaining one. This algorithm should be done inbetween every SAT computation with only those backbone literals that were identified in the last iteration.

Without an efficient implementation, this search for unit implied backbone literals might consume a lot of time. If you try to find the backbone of a formula where the computation takes many SAT calls which each return relatively quickly, the time spent to search for unit implied backbone literals can actually exceed the time spent in the SAT solver if it's implemented inefficiently.

I have tested this method in two solvers, *BB* and a variant of *PB*1 that identifies backbone literals only through the learned literals[8]. The tested files were from a SAT competition. Here it showed, that for the method described in this section it is important to supply a sufficient number of already known backbone literals for it to have a positive impact on runtime. The learned backbone literals alone could rarely supply this, before the *PrefBones* algorithm terminated from other conditions. However the iterative approach of the *BB* solver, testing every yet unidentified literal individually, was much faster at providing positively identified backbone literals that you would need to imply other backbone literals through unit implication.

For further investigation on the effects of this method see section 5.2.2.

3.2.3 Implication through cooccurrence

Another method to recognize backbone literals from other ones is described in [WBX+05] and the rationale goes as follows:

Lemma 1 Given a backbone literal a and another literal b. If b occurs in all clauses that also contain a, then $\neg b$ must be part of the backbone.

[7] In fact, you could remove the whole clause from your formula, since it will always be satisfied through the backbone literal. You could say, it get's subsumed by a clause with only the backbone literal.

[8] In the previous section I stated, that the *PB*1 solver should not be combined with learning backbone literals, but I only realized this constraint much later. During testing, the computed backbones were always verified and an incorrectly identified backbone literal never occurred. The subsequent observations are still valid, independently of this.

Proof: Assuming $\neg b$ was not in the backbone, then there would have to exist a model that contained b. Given that all clauses that contain b also contain a, a cannot be part of the backbone.

In other words, if we determine a new backbone literal, and all clauses that contain it also contain another literal, then we can add the negation of the latter to our backbone set.

However this strategy does not seem to be very effective, at least for the benchmarks that I tested it on. Table 3.1 shows the number of literals that were identified through this method in comparison to those that were identifed through unit implication and the number of variables in the formula. It is probably very rare that a variable should always occur together with another one.

File	n_{unit}	n_{coocc}	n_{BB}
dimacs-hanoi5.cnf	1465	47	1931
grieu-vmpc-s05-25.cnf	565	0	625
grieu-vmpc-s05-27.cnf	142	0	677
fla-barthel-200-2.cnf	33	0	78
fla-barthel-200-3.cnf	43	0	97
fla-barthel-220-1.cnf	149	0	184
fla-barthel-220-2.cnf	0	0	4
fla-barthel-220-4.cnf	0	0	6
fla-barthel-240-2.cnf	61	0	128
fla-qhid-280-1.cnf	1	0	15
fla-qhid-280-3.cnf	251	2	274
fla-qhid-300-1.cnf	260	0	291
fla-qhid-300-4.cnf	270	0	293
fla-qhid-320-1.cnf	296	0	318
fla-qhid-320-2.cnf	0	0	2
fla-qhid-320-5.cnf	283	2	312
fla-qhid-340-2.cnf	310	0	332
fla-qhid-340-3.cnf	309	3	333
fla-qhid-340-4.cnf	301	4	335
fla-qhid-360-1.cnf	326	0	335
fla-qhid-360-5.cnf	321	0	349
fla-qhid-380-1.cnf	344	2	372
fla-qhid-400-3.cnf	366	0	393
fla-qhid-400-4.cnf	359	0	392
smallSatBomb.cnf	0	0	9

Table 3.1: Comparison of number of backbone literals identified through cooccurrence in comparison to the number identified through unit implication and the overall number of backbone literals in the formula.

4 Experiments with Preferences

4.1 Tested modifications

As is mentioned in [PJ18], the concept of preferences has not experienced much research as of yet. That is why I experimented a lot with various modifications on the concept for the purpose of this thesis.

All of the following modifications were implemented ontop of algorithm 7 (*PB*1). Algorithm 6 (PB0) depends on that all available preferences are taken into account strictly, so there is not much space to experiment there without making the algorithm unreliable.

4.1.1 Forgetting preferences

During the course of solving a CNF formula with the CDCL algorithm, many assignment conflicts occur which must be resolved through backtracking. This means that some assignments need to be taken back. Looking at this from the other direction, it can very well happen, that the same variable can occur in decisions multiple times over such a calculation. If a variable assignment is even slightly involved in a conflict, then, lacking any further knowledge, it is a good strategy to simply try it's negation in the future. It is not guaranteed, but simply more likely that the other assignment resolves the conflict that occurred with the previous assignment.

The default way in that preferences are implemented stands in the way of this. If you tell the SAT solver to always assign the same boolean value to a specific variable, then the strategy above cannot be applied. That is why I have implemented an option to *PB*1 where preferences are forgotten after the first time when they are taken into account. This means that when a variable is selected for a decision, it is removed from the primary heap that is consulted first for this selection[1].

The secondary heap always contains all used variables for selection, but in the default configuration where preferences are not forgotten the preferred variables would already be assigned at the point in time when it is consulted. Here this means, that once a preference was forgotten, the associated variables can still occur in decisions after they were reverted through backtracking and can potentially be assigned a different value.

[1] In the *Sat*4*J* library, what decides the order of decisions and what decides the value assigned in these decisions are actually two separate data structures. This means that this removal must be written in two separate places.

The most important effect of this strategy is, that the solver quickly falls back to his default behaviour when it turns out that the formula is difficult to solve when preferences for the resulting model are to be taken into account.

It is quite likely that a formula contains models that are relevant for the backbone, because they contain the only occurence of some literal. However, if they contradict the current set of preferred literals in many other places, giving these preferences to the SAT solver can actually make it difficult to find said model.

4.1.2 Subsets

During benchmarking it showed itself, that all solvers that did not employ any preferences at all were the fastest ones when it came to difficult formulas. So it might be interesting to test some algorithm that can be configured with a floating point number, where a ground value of 0.0 would be equivalent to having no preferences at all and then allowing to enable preferences in very small increments.

The simplest way to do this would be to restrict the number of preferences that could be set for each SAT call. The parameter above is multiplied with the number of variables in the formula and then the number of preferred literals is pruned in each iteration to have at most that size, without of course discarding the set of assignments that has yet to be tested.

Aside from the straightforward implementation, I also created a variant, where extra care is taken to use different subsets for the preferences for each SAT call. To do this, all preferences given to the SAT solver are remembered and not picked again until all of them were tried once. At this point the memory of previously used preferences is cleared so that it builds up again.

4.1.3 Nudging

The previous approach to scalable preferences still works against the decision heuristic because before the SAT solver even starts, the preferred variables are chosen. What subset of the preferred variables would be easiest for the solver to take into account cannot be said upfront. Therefore it should be much more efficient to "nudge" the existing decision heuristic in the direction of the preferences while it's running.

To do this, I implemented a different kind of preference strategy. Here, similar to the previous method, you can pass a floating point number f between 0.0 and 1.0 in addition to the set of preferred variables. When a decision happens in the CDCL algorithm, the selection heuristic sorts all variables by an activity value, which is typically based on how often it was involved in conflicts. Then the resulting array is scanned, beginning with the variable with the lowest activity until one is found that is not yet assigned.

Here this activity value is multiplied with $1 - f$ during this sorting step where a variable is preferred. Therefore, if you pass 0 as value for f, the order that is returned from sorting the activities would be completely unaffected. Very small values of f

would only affect the selection heuristic if a preferred value was already deemed very important but only in second or third place with a small distance to the first one.

However, this factor f can only influence the order of variables for decisions. Which value is then assigned can not be weighted[2]. That is why the default strategy for deciding the phase was left in place.

The purpose of this scheme was to see, whether benefits of preferences would rise up faster than the drawbacks with very low values of f.

4.1.4 Approaching lower and upper bounds

I also tried an idea that was shortly mentioned in [MSJL10, Chapter 3.4]:

> Another approach consists of executing enumeration and iteration based algorithms in parallel, since enumeration refines upper bounds on the size of the backbone, and iteration refines lower bounds. Such algorithm could terminate as soon as both bounds become equal.

You could argue that *BB* (see algorithm 5) matches this concept, since it reduces the set of literals to further examine with every found model, while simultaneously checking individual literals. It removes both the identified backbone literals and the found models from the set of literals to test and once that becomes empty, it terminates.

However, since [MSJL10] also lists the *BB* algorithm, the authors probably ment to run two different algorithms concurrently. Still, making use of both approaches makes sense, so I tried this out in combination with preferences, by checking the set of learned literals in each iteration, as proposed in section 3.2.1.

4.2 Implementation notes for Sat4J

TODO mal schaun ob kann weg

It is advisable to habitually verify that a function does exactly what you expect it to do, because *Sat*4*J* has a habit of sometimes only doing something very similar. For example, it contains a class for a dynamically resizing array called `VecInt`. This is a stack that you can push integers into. Internally it consists of an array that will double in size, when it's capacity is exceeded while the actual size of the `VecInt` is stored explicitely. If you want to use it's content in functions that expect an ordinary integer array, then you would be tempted to use it's `toArray()` function. However this directly returns a reference to the internal array, including values that were removed with the `pop()` function before. Throughout the library, you often see the line `java.util.Arrays.copyOf(ints.toArray(), ints.size())` when such a `VecInt` is converted to an integer array.

- wo kommen spezial selektionsheuristiken hin (Klassenname)

[2] Except if there was a heuristic for that, which can give a confidence value

- genaue adresse der Backbone klasse
- internal/dimacs literals
- vecint toArray bug
- keine doku, beste hinweise geben unit tests über howto use
- prime implicant bug bei Prefbones
- menge an gelernten klauseln kann von der maschine abhängen, vorher nochmal mit synchronisiertem repo ausprobieren

TODO PB1x besser als IBB, unterschied ?????

habe irgendwo footnote dass beim benchmarken der erste call immer schlecht ist, wegen der hardware, der muss hierhin.

keine explizit auftauchenden unit klauseln, musste selber handeln und zum resetten rausnehmen

quote thore: ich dachte mehr daran, dass Du den relevanten Teil von SAT4J grob in seinen Dimensionen darstellst. Vielleicht mit sowas wie einem Klassendiagramm? Und dann könntest Du auf einer "hohen Flughöhe" beschreiben/zusammenfassen, was Du neu implementiert hast. Das ganze aber knapp halten. Ist nur eine Idee, um eine wesentliche Leistung von Dir zu betonen. Kein Muss.

Sehr interessant wäre sicher auch ein Abschnitt, mit Problemen die Du beim Umgang mit Sat4j erlebt hast. Aber nur falls dafür noch Zeit bleibt.

addClause gibt nullpointer zurück wenn eine äquivalente

5 Results

5.1 Tested backbone algorithms

This chapter contains a couple of benchmarks that were all tested against a series of backbone algorithms. There is a slight focus on the preferences approach. Since this is not as thoroughly examined as other approaches in literature, most of these variants are slight modifications of algorithm 7, where preferences are combined with blocking clauses.

For clarity, the used benchmarks are listed here. All algorithms unless otherwise stated reduce the models that they find to prime implicants with the method described in section 3.1.1.

- **BB**: *Sat*4*J* implementation of algorithm 5.
- **IBB**: *Sat*4*J* implementation of algorithm 3.
- **KBB**: Algorithm *BB* in combination with unit implication as described in section 3.2.2.
- **PB0**: *PrefBones* after [PJ18]. See algorithm 6.
- **PB1**: *PrefBones* with blocking clause. See algorithm 7.
- **PB1(model)**: *PB*1 without any model reduction.
- **PB1(forget)**: *PB*1 with forgetting preferences as described in section 4.1.1.
- **PB1b**: *PB*1 where preferences can only affect the selection order. Given for the purpose of comparison with *PB*2.
- **PB1c**: *PB*1 with a limited number of preferences, as described in section 4.1.2. Listed with three different size restrictions.
- **PB1d**: Like *PB*1*c*, but with more diverse preferences in each iteration.
- **PB1e**: *PB*1 in combination with approaching upper and lower bounds as described in section 4.1.4, where the lower bound is made up of learned literals. As previously mentioned, this method should not be applied once a formula is modified. Results of *PB*1*e* serve the sole purpose of indicating that looking for learned literals is a useful strategy.
- **PB2**: Uses scalable preferences as described in section 4.1.3.
- **PB3**: *PB*1 with model reduction through rotation. TODO rename

5.2 SAT competition benchmark

For the first benchmark, I collected a set of 64 files from the 2017 SAT competition[1]. The SAT competitions generally use problems that are difficult to solve compared to other problems of the same file size. This is in order to encourage development of solving strategies that reliably have good performance and not only for most of them. To save time during benchmarking, files that took longer than around one minute for a single model computation were excluded, resulting in files that are generally around 30 kilobytes large. Additionally, problems where the duration for backbone computation averaged below one second were excluded, because here the small differences in the measurements could just as well be explained with external factors like the CPU throttling for a short time. To get meaningful testresults for such files, multiple testpasses should be conducted.

	t_{full}	t_{sat}	t_{last}	n_{sat}
BB	8.63	8.628	-	254
IBB	4.946	4.944	1.911	9
KBB	8.713	8.687	-	36
PB0	31.78	31.779	1.17	4
PB1	17.49	17.488	6.998	5
PB1(model)	25.794	25.793	8.677	5
PB1(forget)	10.867	10.865	4.781	6
PB1x	5.064	5.062	1.562	9
PB1a	11.241	11.239	5.351	8
PB1b	9.785	9.783	1.921	6
PB1c(50%)	25.312	25.311	12.909	5
PB1c(5%)	1064.097	1064.095	797.244	6
PB1d(50%)	26.746	26.745	11.284	5
PB1e	17.187	16.591	6.796	6
PB1f	15.971	15.97	4.333	6
PB2(50%)	10.135	10.133	6.428	10
PB2(5%)	5.232	5.229	1.737	9
PB2(0.5%)	5.251	5.249	1.816	9
PB3	20.927	20.922	8.912	4

Table 5.1: Averages of 64 testfiles taken from sat competitions. The columns indicate: The full time that the calculation took in seconds; The time that was spent in the sat solver; The time that the last sat computation took; The number of sat calls (all values are averages).

The averaged time to compute the backbones of these 64 testfiles can be seen in 5.1[2]. The second column shows the time that was only spent in the SAT solver. This gives a hint on whether a particular algorithm configures it's SAT solver well and whether it looses computation time in things besides the SAT calls, for example through model reduction.

Taking a look at this table, we can quickly see that for these instances, all solvers that employed

File	t_{keep}	$t_{discard}$
brock400-2.cnf	0.233	0.252
dimacs-hanoi5.cnf	1.41	1.596
grieu-vmpc-s05-25.cnf	71.945	78.964
grieu-vmpc-s05-27.cnf	554.52	648.697
fla-barthel-200-2.cnf	0.684	6.019
fla-barthel-200-3.cnf	0.619	2.16
fla-barthel-220-1.cnf	2.511	9.572
fla-barthel-220-2.cnf	7.497	17.759
fla-barthel-220-4.cnf	2.24	14.113
fla-barthel-240-2.cnf	3.632	50.552
fla-qhid-320-1.cnf	6.61	6.575
fla-qhid-320-2.cnf	9.227	101.17
fla-qhid-320-5.cnf	7.861	7.962
fla-qhid-340-2.cnf	11.922	13.688
fla-qhid-340-3.cnf	11.796	17.157
fla-qhid-340-4.cnf	9.454	9.297
fla-qhid-360-1.cnf	39.998	39.797

[1] To be more precise the *essential* folder from the *incremental* package, currently available at `https://baldur.iti.kit.edu/sat-competition-2017/benchmarks/incremental.zip`

[2] The hardware used for all benchmarks listed in this thesis was an Intel Core i7-8809G with 16 gigabytes of DDR4-2667 memory in dual channel and a 1TB hardrive on OSX

preferences as part of their algorithm performed much worse than those that did not. These would be *BB*,*IBB*,*KBB* and *PB*1x. This impression is further supported by the performance results of the three configurations of the *PB*2 algorithm. Here the effect of preferences was reduced incrementally, eventually reaching a computation time close to that of *IBB* where no preferences were set.

5.2.1 Importance of reusing learned clauses

Table 5.2 shows a comparison of individual benchmarks to highlight the importance of reusing learned clauses. While working with the *Sat*4*J* library I noticed that the *IBB* backbone algorithm was accidentally configured in a way that learned clauses would always be discarded between SAT computations. However, these learned clauses are still valid in later iterations of the *IBB* algorithm. The only difference that the formula goes through during this algorithm, is that the blocking clause that ensures a new model repeatedly looses some of it's literals. As long as the set of solutions for a formula is only reduced, the learned clauses of that formula stay valid.[3]

The information contained in learned clauses is very valuable, as it prevents the solver from repeating invalid combinations of assignments that might even be likely to occur again. But if already learned clauses can guide the SAT solver away from possible conflicts, it could ideally return a model without any backtracking.

5.2.2 Benefits of unit implication

This section evaluates the effects of trying to recognize backbone literals through unit implication as described in section 3.2.2. Table 5.1 interestingly shows no

[3] Example: With a formula without any clauses but three variables a,b and c you can create eight models. With only the clause $\{a\}$ in your formula you can have four models, with only the clause $\{a \vee b\}$ you can have 6 models, with only the clause $\{a \vee b \vee c\}$ you can have 7 models.

File	$t_{BB}[n_{sat}]$	$t_{KBB}(t_{sat})[n_{sat}]$	n_{unit}	$n_{backbone}$
brock400-2.cnf	0.04[254]	0.04(0.03)[254]	0	0
fla-komb-400-3.cnf	1276.71[381]	1241.12(1241.11)[45]	336	379
dimacs-hanoi5.cnf	1.94[1931]	1.84(1.23)[281]	1650	1931
vonThore42.cnf	0.02[398]	0.01(0.01)[51]	345	346
grieu-vmpc-s05-27.cnf	261.09[678]	281.08(277.1)[536]	142	677
fla-komb-360-4.cnf	14.94[337]	14.94(14.9)[45]	292	333
fla-qhid-360-1.cnf	74.96[355]	76.03(76.03)[29]	326	355
grieu-vmpc-s05-25.cnf	182.84[625]	179.91(179.44)[51]	574	625
fla-qhid-360-5.cnf	20.68[350]	20.89(20.89)[29]	321	349
fla-barthel-220-4.cnf	3.55[32]	2.74(2.73)[32]	0	6
9012345.cnf	4.18[1483]	6.73(5.92)[813]	670	1478
fla-barthel-220-2.cnf	8.18[23]	8.11(8.11)[23]	0	4
1098765.cnf	1.21[938]	1.29(0.83)[493]	444	921
smallSatBomb.cnf	0.01[26]	0.01(0.01)[19]	7	9

Table 5.3: Benchmark results for a selection of files with a focus on the benefit of unit implication. Rows indicate: Calculation time and number of SAT calls for the *BB* solver ; Calculation time, time spent in SAT solver and number of SAT calls for the *KBB* solver ; Number of backbone literals identified through unit implication (in *KBB*) ; Number of backbone variables in formula.

performance benefit for this technique, even though the number of sat calls is much smaller.

Table 5.3[4] shows individual performance differences between the *BB* solver supplied by *Sat*4*J* and my own *KBB* solver for comparison, as well as the backbone's size and the number of backbone literals identified through unit implication. The time columns also contain the number of sat calls that were comitted. It's content matches the results of table 5.1. Occasional cases where *KBB* trumps in performance over *BB* get balanced out by cases where this is the other way round, but the number of sat calls is always better with the *KBB* algorithm. And this is true in spite of the fact that almost all identifications happened through unit implication.

This indicates that the effect from scanning for unit implied backbone literals also appears in the *BB* solver, only not quite as obvious. When you look at line 10 of algorithm 5, you see that *BB* actually takes identified backbone literals up into the formula to speed up the solving process. Now imagine what happens, if the unit implication case happens during the course of the *BB* algorithm. You have a clause that is not satisfied, even though all but one of it's literals are already assigned through the learned backbone literals. When *BB* tests the last of that clauses literals

[4]In this table the pure sat calculation time for the *BB* column is missing. The *BB* algorithm actually spends almost no time outside of the SAT solver in the case of the listed formulas.

with an assumption against it, the clause becomes immediately unsatisfied[5], proving that the last remaining literal in the clause is part of the backbone.

The results shown in the following chapter (5.3) paint a slightly different picture. Here the pure SAT time of *KBB* is actually smaller than that of *BB*, however even *BB* looses around a third of it's calculation time outside of pure SAT calls. This could be explained by the much higher number of SAT calls for this benchmark and the overhead that comes with preparations and cleanups for a SAT call, which appearantly became relevant at that point. Given good conditions (like in that benchmark) *KBB* can identify many backbone literals in a single search, which also requires less data structures around it compared to a complete SAT computation.

In summary, searching for backbone literals through unit implication can be a more efficient strategy than testing each individual literal through a SAT call, however only if the search for unit literals is implemented in a very efficient way. Also note that the problem of having too many SAT calls can just as well be resolved by using a enumeration algorithm like *IBB* instead of an iterative approach.

5.2.3 Effect of subset preferences

The algorithms where the set of preferences was restricted in size (*PB*1*c* and *PB*1*d*) compared relatively bad to the base algorithm *PB*1 and this is most pronounced in the case of *PB*1*c*(5%) where the restriction is the strongest. This means that, at least for this benchmark, restricting the number of preferences completely backfired. The decrease in performance can be traced beginning with *PB*1 which is equivalent to *PB*1(100%) towards the 50% and 5% configuration.

For a possible explanation I should begin with a reminder about the exact way in which preferences are implemented in *PB*1. Here two decision heuristics coexist, h_{pref} and h_{tail}, whereas under normal circumstances a *CDCL* SAT solver would only use one to pick a literal for a decision. In *PB*1, h_{pref} is always consulted first, and only if all the variables that it offers for an assignment are already assigned to a value, h_{tail} is queried. Both h_{pref} and h_{tail} contain a heuristic to choose the optimal variable for a decision, but in h_{pref} the set to choose from is restricted and what that variable would then be assigned to, is fixed. In contrast, h_{tail} is free to choose any remaining free variable and give it either boolean value,

	t_{last}/t_{avg}
IBB	3.479
PB0	0.147
PB1	2.000
PB1(model)	1.682
PB1(forget)	2.640
PB1x	2.777
PB1a	3.808
PB1b	1.178
PB1c(50%)	2.550
PB1c(5%)	4.495
PB1d(50%)	2.109
PB1e	2.457
PB1f	1.627
PB2(50%)	6.343
PB2(5%)	2.989
PB2(0.5%)	3.113
PB3	1.703

Table 5.4: Ratio between the average duration of a SAT call and the duration of the last SAT call. Calculation is based on the values in table 5.1

[5] Remember that *BB* tests by trying to disprove a potential backbone literal.

depending on what it deems better to satisfy the formula.

When a decided literal is involved with a conflict, it will be pushed back in it's decision heuristic. That way an opportunity is given to variables that might not be so difficult to be assigned a good value and the problematic variables might be assigned a necessary value through unit implication. However this can only happen if enough other variables are available in the same heuristic to take it's place and with it's strong size restriction, $PB1c(5\%)$ doesn't have them.

Additionally, the more that you restrict the size of the preferences, the more likely it becomes that all preferred literals are negated backbone literals. In this case, all decisions made based on h_{pref} prevent you from finding a model and must eventually be reverted.

Even worse, since the preferred assignments have to be done before all the others, they occur at the beginning of the CDCL table. This means that in case of a conflict with other assignments, the preferred decisions will not be reverted if there is any other decision available to be reverted, since the strategy of CDCL is usually to revert only the youngest decision involved with the conflict. But if this preference actually goes against a backbone literal of the formula, it must eventually be reverted to reach a valid model. This can only happen by a unit propagation, since otherwise a decision would just take the preference into account again. And since no decision can happen before the preferred decisions, to counter a preference a unit propagation requires the maximum amount of information so that it can happen without a prior decision. You would need to do an amount of learning that is actually equivalent to directly identifying a backbone literal.

5.2.4 Duration of last SAT call

A similar effect to the previous is hinted by the performance results listed in the third column t_{last} of table 5.1. Many of these timings are larger than the average computation would take. The observation coincides exactly with whether the last SAT call would be unsatisfiable. The *PB*0 solver does not apply blocking clauses and still performs well in the last iteration, whereas the *IBB* solver that does not employ preferences performs badly.[6] This indicates that the problem lies in the fact, that in this last sat call, the formula is unsatisfiable.

That *PB*0 performs so well in the last SAT call is actually quite interesting, as here the last SAT call is about seven times faster than it's own average. This could be explained by the fact that since the formula is never really changed in the case of *PB*0, the clauses that it learns never become outdated. In contrast, in the typical model enumeration schemes, they may not become invalid but at least less useful over time, because they might only affect the generation of implicants that are later excluded

[6]The *BB* and *KBB* solvers are excluded from this table, because as they are iterative algorithms. Algorithmically the first and last SAT calls do not differ that much.

from the formula through a stronger blocking clause. The information stored in these learned clauses can be put to use especially well in the last iteration. Theoretically, the preferences at that point in time are all impossible to implement and should slow down that particular calculation. But the solver has already spent a lot of time with conflicts that stem from them and should thus have learned many clauses that involve the backbone literals, possibly even directly identifying them as such. In that particular case, the corresponding preference is overridden immediately and *PB*0 merely needs to get the remainder of the formula to satisfy, just like any other backbone algorithm.

TODO alte these ausgraben: backbone literale stehen in CDCL tabelle vor erster decision, beweisen, oder einfach weg

5.2.5 Benefits of forgetting preferences

Table 5.5 lists a comparison of multiple backbone algorithms once with ordinary, permanent preferences and in a forgetting scheme as described in section 4.1.1. Preferences in general resulted in worse performance for this benchmark. However, *PB*1($forget$), where the preferences are configured to immediately drop those that were involved in a conflict, reduced the penalty to an acceptable level in all examples. [7]

	Permanent	Forgetting
PB1	20.251	12.426
PB1c(50%)	25.708	13.109
PB1c(5%)	1080.724	14.483
PB2(50%)	13.023	7.025
PB2(5%)	6.283	6.348
PB3	21.255	13.154

Table 5.5: Average of the complete backbone computation for variants of *PrefBones* with and without forgetting preferences.

Table 5.6 in the upcoming section shows results for a formula that is more beneficient to preferences. Here we see, that in such a case, forgetting preferences still work as intended. If these results turn out to be reproducible for other formulas as well, forgetting preferences could be a viable strategy to compute the backbone of any formula without prior knowledge about it, since the penalty for difficult formulas would be relatively low but the speedup for easy ones very high.

5.3 Industrial benchmark

Another benchmark I applied the various backbone algorithms to, was a formula for a real world application from

[7] In earlier executions of my benchmark setup, I also tested *PB*1c with 0.5% as preference size restriction, however only in combination with forgetting preferences. Without forgetting them, this configuration turned out to be completely unreasonable to test within an acceptable time frame and was dropped.

	t_{calc}	t_{sat}	n_{sat}
BB	11.117	7.235	159545
IBB	6.353	2.316	15748
KBB	13.747	2.11	15266
PB0	3.659	1.429	6531
PB1	3.918	1.611	6531
PB1(model)	2.148	1.59	6531
PB1(forget)	4.067	1.682	6531
PB1x	7.596	2.94	15748
PB1b	7.754	3.199	15248
PB1c(50%)	5.376	2.248	9680
PB1c(5%)	7.026	2.693	13955
PB1d	4.464	1.804	7386
PB1e	3.524	1.354	6387
PB1f	3.603	1.526	6369
PB2(50%)	8.341	3.457	15752
PB2(5%)	8.313	3.484	15752
PB2(0.5%)	8.364	3.529	15752

the automobile industry, a so-called "Product Overview Formula". The purpose of this formula POF_1 was to describe a product in the context of options or features available to the customer. Some of these options can be combined, others exclude or require each other. Most of the variables in POF_1 correspond to a boolean parameter that is set to $\top$ if the associated feature is requested by the customer. If the formula would become unsatisfiable under such assumptions equal to the requested configuration, then the combination of these features would be impossible. Looking at it from the other side, the set of models of POF_1 matches exactly the set of possible product configurations that are available to the customer. POF_1 further contains a small set of additional variables to model more complex relationships between options.

The particular use case that I examined was to find the implications in the formula, i.e. if a customer requests feature a, would he also have to pick feature b. A primitive way to calculate this would be to iterate over all possible pairs of features and check for satisfiability of POF_1 under the condition that a is $\top$ and b is $\bot$. If this was unsatisfiable, then a would imply b. However, with 407 literals to choose from, you would have to do 165,242[8] sat calls.

A more efficient approach is to only go over the 407 options once and for each variable a of them calculate the backbone of $POF_1 \cup \{a\}$ or in other words POF_1 under the assumption a. If a feature b occurs in all models when a is assumed, then a implies b in some way.

5.3.1 Performance measurement and discussion

Table 5.6[9] lists experimental results with the same set of algorithms that were run on the files from the SAT competition. In the following, I will list some observations about these results.

[8] 407 times 406

[9] The duration of the last SAT call is missing in this table, because it was too small to be meaningfully interpretable.

- We immediately see that preferences give a great benefit to performance, other than in the previous benchmark with the files from the SAT competition. Here, all cases where I tried to soften the effects of preferences (*PB*1(*forget*),*PB*1*c*, *PB*2) had worse results than the base algorithm *PB*1[10].

- The fastest algorithm overall (*PB*1(*model*)) is the one that does not reduce the models at all. The number of SAT calls is even exactly the same as that of that variant with prime implicant reduction. However the time that was spent in the SAT solver is very similar to that execution with prime implicant reduction (*PB*1), so the difference must be the time that was spent to reduce the model and the benefit, namely the number of optional variables in the prime implicant, was very small.
  Still, this does not mean that model reduction wouldn't be useful at all in this benchmark, as the next point will show. It merely implies that the benefit of applying any special strategy may very well be outweighed by the extra performance cost that comes with it, since here all of this work would have to be done a very large number of times.

- The fastest implementation when it comes to pure SAT solver time as well as number of SAT calls was *PB*3, which differs from the others in that it uses literal rotation to reduce models instead of prime implicant computation. Apparantly the models that occur in this formula actually do contain many optional assignments. However these are spread over many different prime implicants with little distance to the model.
  A previous variant of this algorithm took around 250 seconds to compute overall. This version did not use the lookup from literal to containing clauses (as described in later paragraphs of section 3.1.2), but iterated over all clauses in the formula. Such a drastic effect of an efficient implementation did not occur for the SAT competition benchmark.
  This has primarily two reasons. First, the number of SAT calls per instance is twice as much in this benchmark compared to those from the SAT competitions[11], therefore the model reduction happens twice as often. Secondly, the formula of this benchmark is much larger than those from the SAT competition, with the filesize being around twenty times as large. Therefore there are much more clauses to search through.

- The two algorithms in third place of overall computation time were *PB*1*e* and *PB*1*f*. Both of these make use of learned literals so appearantly the backbone literals in this formula are easy to identify by checking the set of literals that the solver learns when it computes new models.
  In fact, on inspection, all literals that later turn out to be part of the backbone occur in this set after the very first SAT call(with an exception of the forced

[10] Especially *PB*2, but note that this algorithm only affects the assignment order. A fair comparison would be between *PB*2 and *PB*1*b*, but here *PB*2 is the slower one as well.

[11] The 4471 calls span over 407 problem instances.

literal, which can be extracted by searching the formula for clauses with only one literal).

- Another noteworthy algorithm is the *KBB* algorithm that uses unit implication to identify backbone literals. The implementation of this algorithm was based on that of the *BB* algorithm and here the number of SAT calls with *KBB* is a tenth of the number in the case of *BB*. However this advantage is not expressed equally strong when it comes to the time spent in the SAT solver(around a fourth), which means that the identification through calculating a new model can sometimes be more efficient than regularly checking unit implications, as was already explained in the previous section (5.2.2).
  Finally, the overall computation time of *KBB* is actually worse than that of *BB*, meaning that the benefit through cheap backbone literal identification is outweighed by the time that it takes to check the clauses for the unit case.

5.3.2 Specialized algorithm

TODO PB0 war auch ziemlich gut

With the findings from the last section and some experimentation, I devised a specialized algorithm *PB*4 for this use case which is listed in algorithm 10. For comparison with the other implementations, the performance results can be seen in table 5.7.

	t_{calc}(sec)	t_{sat}(sec)	NSatCalls
PB4	1.482	0.863	4064

Table 5.7: Results of specialized backbone algorithm on product formula benchmark.

First of all, we should make use of the fact, that all backbone literals are learned after the very first SAT call. This can be done with the scheme of approaching upper and lower bounds as described in section 4.1.4. The lower bound bb_l of the backbone consists of all learned literals, whereas it's upper bound bb_u consists of the intersection of found models or reductions of them. Having both upper and lower bound not only allows to potentially stop the loop before the last sat call. It also accelerates the reduction of the first model to it's required literals, because if you know that a literal is in the lower bound of the backbone, you don't need to explicitely test, whether it would also be in it's upper bound. This behaviour is listed in algorithm 11. It might sound sensible at first to regularly check for new learned literals in case some might only show up later. However for the reasons described in section 3.2.1, this is not guaranteed to result in valid backbone literals.

Secondly, it turned out, that only the first model reduction in line 3 was worth it's computational effort. The reduction of subsequent models did not give performance benefits that would have outweighed the time it took to do. That is why in line 10 the upper bound is intersected with the complete model instead of a reduction of it. This property of the POF_1 would also explain why the overall computation time of *PB*3 is over three times larger than it's pure SAT time. If your instance requires a lot of SAT

calls, then it might actually be a very sound strategy to apply optimizations such as model reductions only on the first iteration where it probably has the greatest effect.

For the preferences, the typical scheme was applied, where the solver was configured to disprove as many of the backbone candidates as possible. The same set was added as a blocking clause to ensure that the loop would eventually terminate if bb_l happenend to miss some backbone literal. In line 7 the identified backbone literals are enforced as well, to prevent impossible assignments against them. However if implemented correctly in *Sat4J*, they should still be in place from the previous model computation.

Algorithm 10: Specialized algorithm for industrial application

Input: A satisfiable formula F in CNF
Output: All literals of the backbone of F

```
1  (outc, model, learnt) ← SAT(F)
2  bb_l ← learnt
3  bb_u ← required(model, F, bb_l)
4  while bb_l ≠ bb_u do
5      blocker ← ⋁_{l∈bb_u} ¬l
6      prefs ← {¬l : l ∈ bb_u}
7      (outc, model) ← prefSAT(F ∪ blocker ∪ bb_l, prefs)
8      if ¬outc then
9          return bb_u
10     bb_u = bb_u ∩ model
11 return bb_u
```

Algorithm 11: Function *required*(M, F, bb_l)

Input: A model M for formula F, a lower bound bb_l of the backbone of F
Output: All unrotatable literals in M as an upper bound of the backbone of F

```
1  R ← bb_l
2  for l ∈ M do
3      if l ∈ bb_l ∨ ∃c ∈ F : c⟨M\l⟩ = ⊥ then
4          R ← R ∪ {l}
5  return R
```

5.4 Second industrial benchmark

In the following I will analyze another product formula POF_2, similar to the one in the previous chapter. However this

	$n_{Literals}$	$n_{Clauses}$	n_{test}	Filesize
POF_1	469	66957	407	817 KiB
POF_2	5055	81267	948	1401 KiB

Table 5.8: Size comparison of the two industrial benchmarks. Contains the number of literals, clauses, tested literals and the filesize.

one is much larger than the previous, see table 5.8. I tested this formula against the same set of algorithms as the previous formula POF_1 including $PB4$, to identify the specific characteristics of it and identify the optimal backbone solver for it, just as in the previous section. Table 5.9 shows the performance result.

The first thing that shows, again, is that algorithms with preferences have a much better performance overall. $PB0$ and $PB1$ finish between two and three times faster than BB and IBB. Something similar an be seen when you compare $PB1b$ with the $PB2$ variants, however it becomes more complicated to explain the result of $PB1(forget)$, which outperformed $PB1$ slightly. Perhaps POF_2 contains easy components as well as difficult ones. If so, forgetting preferences could adjust to both dynamically during the same model computation. This would be supported by the fact that $PB1(forget)$ turned out to be the fastest when it comes to pure SAT time.

	t_{calc}	t_{sat}	t_{last}	n_{sat}
BB	964.05	680.113	-	2,063,982
IBB	1244.075	832.699	1.527	432,408
KBB	1229.866	429.248	-	417,846
PB0	427.187	276.132	2.094	129,943
PB0(model)	304.072	273.185	2.084	130,099
PB1	498.542	316.225	0.235	146,349
PB1(model)	375.146	325.547	0.221	151,119
PB1(forget)	471.434	292.868	0.243	143,406
PB1b	1333.112	818.313	0.614	460,292
PB1c	508.382	318.489	0.215	146,799
PB1c(5%)	575.371	355.767	3.06	198,242
PB1c(0.5%)	805.314	511.091	3.452	273,882
PB1d	517.259	317.346	0.208	146,430
PB1e	509.053	314.43	1.123	148,096
PB2	1515.258	1025.452	3.576	437,898
PB2(5%)	1459.938	969.117	1.921	439,641
PB3	614.978	311.232	0.286	141,126
PB4	363.885	318.21	0.243	146,498

Table 5.9: Second Industrial benchmark. Values are not averaged, but summed up over 948 different benchmarks.

Additionally, it seems that model reduction does not give any benefit in this instance. $PB1(model)$ outperforms the base configuration with prime implication reduction as well as $PB3$ and the same goes for the comparison between $PB0$ and $PB0(model)$. Looking for required literals in the model[12] does give a performance benefit visible in the number of SAT calls and the pure calculation time, as is to be expected from the strongest model reduction scheme. It is however negligible. A more interesting case is $PB4$. Here the overall computation time is comparatively close to the pure SAT time, which coincides with the fact that $PB4$ reduces it's model only once. However, $PB4$ is still outperformed by $PB0(model)$. I also tried out the combination of forgetting preferences and not employing model reduction, but their benefits did not add up to result in a better method than $PB0(model)$.

[12]Meaning $PB3$

Further we can see that the observation about unit implied backbone literals can be observed again in this benchmark. The number of SAT calls and the time spent in the SAT solver are much better in *KBB* compared to *BB*. But this does not reflect in better overall performance. Also, learnable literals don't seem to be particular valuable in this formula as can be deducted from the results of *PB*1*e*.

Judging from this, the best backbone computation strategy for POF_2 would be the *PB*0 algorithm without any model reduction.

PB4 bringt nix
präferenzen immer besser
gelernte literale bringen nichts.
rotieren < primimplikante < model bzgl calcTime, in pureSat andersrum, aber logisch, weil stärkere reduktion. model reduktion hier anscheinend völlig nutzlos und verschwendet zeit
PB0 besser als PB1.
PB0 komischerweise schlechtere last sat time TODO entweder begründen oder einfach weglassen
PB1(forget) geht auch gut
aber: basisformel hat backbones und lernbare klauseln

PB0 geht mit (model), leider nicht auch mit (forget) Basis formel hat backbone literale, also de

ad
ad
ad
ad
ad
ad
ad
ad
ad
ad
ad
ad
ad
ad
ad
ad

wichtig: die literale nicht per klausel erzwingen sondern per assumption, dann werden gelernte klauseln nicht weggeworfen. geht nur mit BB und PB0, und bringt nix in vonThore, weil da nix gelernt wird.

5.4.1 Efficiency over all assumptions

	t_{calc}	t_{sat}	t_{last}	n_{sat}
BB	1275.055	829.091	-	2,116,264
IBB	1312.134	878.177	2.198	431,354
PB0	487.614	287.116	1.939	129,626
PB0(model)	316.511	285.347	1.914	129,668

Table 5.10: Results with assumptions instead of formula modification

Since for this benchmark, the almost same formula is worked on hundreds of times, it makes sense to think about ways how to make this outer loop efficient as well.

eigentlich erstes: kopieren rausoptimieren

The first consideration was to first compute information of the base formula and reuse it in all subsequent computations, where the backbone under an assumption was calculated. Two interesting pieces of this would be the clauses that would be learned by doing this and secondly the backbone of the base formula.

The gain of this depends on the formula. POF_1 by itself actually had neither backbone literals nor was it necessary for the SAT solver to learn any new clauses to find a model for it, and even when I tried to compute it's backbone with either *PB*0 or *BB*[13] in an effort to learn as much about the formula as possible, not a single clause was learned.

POF_2 on the other hand actually had quite a large backbone and many clauses to learn. Table 5.11 shows performance results where the solver object was always copied from either the original formula or one where learned clauses and backbone literals were added explicitely. The third ssection shows how long the preparations took with different solvers and how many clauses were learned and additionally the size of POF_2's backbone.

learned lits übernehmen in POF ist schlechter ansatz, weil wird relativ schnell schlechter als immer die formel übernehmen, womöglich nutzlose learned clauses, die nur verlangsamen und nichts aussagen (weil von anderen assumptions)

The first consideration was to first compute information of the base formula and reuse it in all subsequent computations, where the backbone under an assumption was calculated. However this was not possible.
One interesting piece of information would have been the set of clauses that were learned during the computation of a model of the

Computation	t_{full}	t_{sat}	t_{last}	n_{sat}
BB	980.539	672.579	-	2,080,370
IBB	1227.487	776.105	0.855	461,746
PB0	374.425	219.909	1.93	129,912
PB0(model)	250.473	219.202	1.951	129,912
PB1	412.914	238.005	0.244	137,985
PB1(model)	290.615	244.659	0.248	140,790
Preparation	t_{calc}	$n_{learned}$	n_{BB}	
BB	1.928	1898	1449	
PB0	0.918	2214	1449	
PB0(model)	0.505	2244	1449	

Table 5.11: Comparison of benchmark results depending on reuse of learned clauses and backbones of the base formula POF_2. Subsequently number of learned clauses and backbone literals through preparation.

[13] Both algorithms, that can leave their learned clauses in place after backbone computation.

base formula POF_1. However this
set was empty. Appearantly the formula is simple enough, that the model can be found without a single conflict. The solver returned without having learned any new clause for the formula.

Another thing I thought about was the backbone of the base formula. If you can determine backbone literals of POF_1, then you can safely assume, that all of these are part of any restricted formula $POF_1 \cup \{a\}$. This is because, when a clause gets added to a CNF formula, it's set of models can only shrink, no new model can be induced through an additional clause[14]. This implies that an assignment that didn't occur in the models before will also not appear in the new smaller set of models, in turn implying that if a variable always occured with the same assignment in POF_1, it will also do so in $POF_1 \cup \{a\}$.
However this set was empty as well and for a good reason in this particular case. The backbone of a formula consists of all assignments that have to be done to end up in a model. In the context of a product formula, which we have here, this would mean, that a specific feature of the product would always have to be selected. In other words, the product has an option that is not optional.

What remained was to see, if the loop over the assumptions could be sped up. In fact whereas the backbone computations[15] took a sum of 1.5 seconds, this loop overall took 3.9 seconds. The difference consisted of primarily the copying of the formula(2.2 seconds) and writing down benchmark information(around 150 milliseconds).

To do this in *Sat*4*J* you must first of all reset the formula object[16]. In the case of model enumeration algorithms like *IBB* or *PB*1, this incorporates removing all the blocking clauses because they modify the formula. If you don't do this, all backbone computations after the first one will immediately fail, because the final blocking clause makes the formula unsatisfiable. Additionally, you must flush the set of learned clauses and literals because some of these may be based on the blocking clauses. Therefore, if you would leave them in place, they could reproduce some of the effects that the blocking clause had. Take care that all of this happens in a way that during the individual backbone computations, the learned clauses stay in place. *Sat*4*J* supplies a special function to remove a clause under the expectation that it will immediately become subsumed or already is, which you can use to remove old blocking clauses during the loop without flushing the learned clauses. Alternatively you could remove all of them at the end. The ordinary function to remove a clause in *Sat*4*J* triggers the cleanup automatically.

Additionally, it makes sense to revert the decision heuristic if you replaced it with something that allows to implement preferences and you want to reuse the formula object with a backbone solver without preferences. The typical behaviour throughout the *Sat*4*J* library, including the *BB* and *IBB* backbone solvers, is to assume that such a data structure is already set up correctly.

[14] Assuming that no new variable was introduced.
[15] Of only the new *PB*4 solver
[16] The class name in *Sat*4*J* is *Solver* or *ISolver*.

One last bit of state inside the formula object that I noticed were two floating point numbers called *claInc* and *claDecay*. These influence the scale in which the activity values inside the decision heuristic grow and shrink. They do not influence the calculation result, but merely the exact behaviour of the solver.

With all of this taken care of, the duration of the outermost loop over the assumptions took just as long as the sum of the durations of the backbone computations.

6 Conclusion

TODO prefbones ist besser für echtwelt beispiele, ohne prefs zuverlässiger bei komplizierten beispielen

kombination aus strategien auf fallbeispiel optimieren(?)

Guiding the behaviour of the SAT solver using preferences can be very beneficial in the case of real world applications and where it would be disadvantageous, the penalty can be kept under control.

future work dazu machen

Bibliography

[DFLBM13] David Déharbe, Pascal Fontaine, Daniel Le Berre, and Bertrand Mazure. Computing prime implicant. *2013 Formal Methods in Computer-Aided Design, FMCAD 2013*, pages 46–52, 10 2013.

[DLL62] Martin Davis, George Logemann, and Donald Loveland. A machine program for theorem-proving. *Commun. ACM*, 5(7):394–397, July 1962.

[ES03a] Niklas Een and Niklas Sörensson. An extensible sat-solver [ver 1.2]. 2003.

[ES03b] Niklas Een and Niklas Sörensson. Temporal induction by incremental sat solving. *Electronic Notes in Theoretical Computer Science*, 89:543–560, 12 2003.

[KK01] Andreas Kaiser and Wolfgang Küchlin. Detecting inadmissible and necessary variables in large propositional formulae. In *Proceedings of International Joint Conference on Automated Reasoning*, IJCAR 2001, 2001.

[MSJL10] João Marques-Silva, Mikoláš Janota, and Inês Lynce. On computing backbones of propositional theories. *Frontiers in Artificial Intelligence and Applications*, 215:15–20, 01 2010.

[MSJL15] João Marques-Silva, Mikoláš Janota, and Inês Lynce. Algorithms for computing backbones of propositional formulae. *AI Commun.*, 28(2):161–177, April 2015.

[PJ18] Alessandro Previti and Matti Järvisalo. A preference-based approach to backbone computation with application to argumentation. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, SAC '18, pages 896–902, New York, NY, USA, 2018. ACM.

[SS96] João P. Marques Silva and Karem A. Sakallah. *GRASP - a New Search Algorithm for Satisfiability*, pages 220–227. ICCAD '96. IEEE Computer Society, Washington, DC, USA, 1996.

[WBX+05] Zipeng Wang, Yiping Bao, Junhao Xing, Xinyu Chen, and Sihan Liu. Algorithm for computing backbones of propositional formulae based on solved backbone information. In *2016 International Conference on Applied Mathematics, Simulation and Modelling*. Atlantis Press, 2016/05.

[WKS01] Jesse Whittemore, Joonyoung Kim, and Karem Sakallah. Satire: A new incremental satisfiability engine. *Proceedings - Design Automation Conference*, pages 542 – 545, 02 2001.