TLP **14** (4–5): 755–770, 2014. © Cambridge University Press 2014
doi:10.1017/S1471068414000325

Anytime Computation of Cautious Consequences in Answer Set Programming

MARIO ALVIANO, CARMINE DODARO and FRANCESCO RICCA
Department of Mathematics and Computer Science, University of Calabria, 87036 Rende (CS), Italy
(*e-mail:* {alviano,dodaro,ricca}@mat.unical.it)

submitted 1 January 2003; revised 1 January 2003; accepted 1 January 2003

Abstract

Query answering in Answer Set Programming (ASP) is usually solved by computing (a subset of) the cautious consequences of a logic program. This task is computationally very hard, and there are programs for which computing cautious consequences is not viable in reasonable time. However, current ASP solvers produce the (whole) set of cautious consequences only at the end of their computation. This paper reports on strategies for computing cautious consequences, also introducing anytime algorithms able to produce sound answers during the computation.

KEYWORDS: answer set programming, query answering, cautious reasoning, anytime algorithms

1 Introduction

Answer Set Programming (ASP) is a declarative language for knowledge representation and reasoning (Niemelä 1999; Marek and Truszczyński 1999; Lifschitz 2002; Baral 2003; Gelfond and Kahl 2014). In ASP, knowledge concerning an application domain is encoded by a logic program whose semantics is given by a set of stable models (Gelfond and Lifschitz 1991), also referred to as answer sets. As for other nonmonotonic formalisms, the resulting knowledge base can be queried according to two possible modes of reasoning, usually referred to as brave (or credulous) and cautious (or skeptical). Brave reasoning provides answers to the input query that are witnessed by some stable model of the knowledge base. For cautious reasoning, instead, answers have to be witnessed by all stable models. Cautious reasoning over ASP knowledge bases has relevant applications in various fields ranging from Databases to Artificial Intelligence. Among them are consistent query answering (Arenas *et al.* 2003), data integration (Eiter 2005), and ontology-based reasoning (Eiter *et al.* 2008).

A common practice in ASP is to reduce query answering to the computation of a subset of the cautious consequences of a logic program (Leone *et al.* 2006), where cautious consequences are atoms belonging to all stable models. As an example, in the context of Consistent Query Answering (CQA), consider an inconsistent database

https://doi.org/10.1017/S1471068414000325 Published online by Cambridge University Press

D where in relation $R = \{\langle 1,1,1\rangle, \langle 1,2,1\rangle, \langle 2,2,2\rangle, \langle 2,2,3\rangle, \langle 3,2,2\rangle, \langle 3,3,3\rangle\}$ the second argument is required to functionally depend on the first. Given a query q over D, CQA amounts to computing answers of q that are true in all repairs of the original database. Roughly, a repair is a revision of the original database that is maximal and satisfies its integrity constraints. In the example, repairs can be modeled by the following ASP rules:

$$R_{out}(X, Y_1, Z_1) \leftarrow R(X, Y_1, Z_1),\ R(X, Y_2, Z_2),\ Y_1 \neq Y_2,\ {\sim}R_{out}(X, Y_2, Z_2) \quad (1)$$

$$R_{in}(X, Y, Z) \leftarrow R(X, Y, Z),\ {\sim}R_{out}(X, Y, Z) \quad (2)$$

where (1) detects inconsistent pairs of tuples and guesses tuples to remove in order to restore consistency, while (2) defines the repaired relation as the set of tuples that have not been removed. The first and third arguments of R can thus be retrieved by means of the following query rule:

$$Q(X, Z) \leftarrow R_{in}(X, Y, Z) \quad (3)$$

whose consistent answers are tuples of the form $\langle x, z\rangle$ such that $Q(x, z)$ belongs to all stable models. In this case the answer is $\{\langle 1,1\rangle, \langle 2,2\rangle, \langle 2,3\rangle\}$.

Cautious reasoning has been implemented by two ASP solvers, namely DLV (Leone *et al.* 2006) and clasp (Gebser *et al.* 2012a), as a variant of their stable model search algorithms. In a nutshell, DLV and clasp compute stable models of a given program by means of a two-phase process. The first phase is actually implemented by a possibly external instantiator producing a ground version of the input program. The ground program is then processed by the second phase, which actually searches for stable models. Cautious reasoning can be obtained by reiterating the stable model search step according to a specific solving strategy. The procedure implemented by DLV searches for stable models and computes their intersection, which eventually results in the set of cautious consequences of the input program. At each step of the computation, the intersection of the identified stable models represents an overestimate of the solution, which however is not provided as output by DLV. The procedure implemented by clasp is similar, but the overestimate is outputted and used to further constrain the computation. In fact, the overestimate is considered a constraint of the logic program, so that the next computed stable model is guaranteed to improve the current overestimate.

It is important to note that cautious reasoning is a resource demanding task, which is often not affordable to complete in reasonable time. As a matter of fact, in these cases current ASP solvers do not produce any sound cautious consequence, as also the overestimate produced by clasp can only guarantee that some atoms do not belong to the solution. It is interesting to observe that query answering is addressed differently in other logic programming languages. For example, Prolog queries having infinitely many answers are common due to the presence of uninterpreted function symbols. Prolog systems are thus designed to produce underestimates of the complete, possibly infinite solution, which actually represent sound answers to the input query. In fact, underestimates are useful in practice, especially in the cases in which waiting for termination is not affordable, and this may be the case even if

https://doi.org/10.1017/S1471068414000325 Published online by Cambridge University Press

termination is guaranteed. It is thus natural to ask whether underestimates can be computed also in the context of ASP.

The paper provides insights in this respect, showing that underestimates can actually be obtained by improving algorithms employed by ASP systems, or adapting to ASP the *iterative consistency testing* algorithm for computing backbones of propositional theories (Marques-Silva *et al.* 2010). The paper also introduces a modified version of this last algorithm that takes advantage of restarts and heuristic values for faster improvement of underestimates. An interesting aspect of the algorithms analyzed in this paper is that underestimates are produced during the computation of the complete solution. The computation can thus be stopped either when a sufficient number of cautious consequences have been produced, or when no new answer is produced after a specified amount of time. Such algorithms are referred to as anytime in the literature. The empirical comparison of these algorithms highlights that they could be combined in a parallel implementation to converge faster to the complete solution. Actually, a proof-of-concept implementation of the parallel approach is also presented in the paper to confirm this conjecture.

2 Preliminaries

Syntax and semantics of propositional ASP programs are briefly introduced in this section. A quick overview of the main steps of stable model search is also reported in order to provide the reader with essential knowledge on a process that is used but not substantially modified by the algorithms analyzed in this paper. (For complementary introductory material on ASP see Gelfond and Lifschitz 1991; Baral 2003; Gelfond and Kahl 2014.)

Syntax. A normal logic program consists of a set of rules of the following form:

$$a_0 \leftarrow a_1, \ldots, a_m, \sim a_{m+1}, \ldots, \sim a_n \tag{4}$$

where each a_i $(i = 0, \ldots, n)$ is a propositional atom in a fixed, countable set $\mathscr{A}$, $\sim$ denotes *negation as failure*, and $n \geqslant m \geqslant 0$. For a rule r of the form (4), atom a_0 is called *head* of r, denoted $H(r)$; conjunction $a_1, \ldots, a_m, \sim a_{m+1}, \ldots, \sim a_n$ is named *body* of r; sets $\{a_1, \ldots, a_m\}$ and $\{a_{m+1}, \ldots, a_n\}$ are denoted $B^+(r)$ and $B^-(r)$, respectively. A *constraint* is a rule of the form (4) such that $a_0 = \bot$, where $\bot$ is a fixed atom in $\mathscr{A}$.

Semantics. An interpretation I is a subset of $\mathscr{A} \setminus \{\bot\}$. I is a model of a rule r, denoted $I \models r$, if $H(r) \in I$ whenever $B^+(r) \subseteq I$ and $B^-(r) \cap I = \emptyset$. It is a model of a program P, denoted $I \models P$, if it is a model of all rules in P. The definition of stable model is based on a notion of program reduct (Gelfond and Lifschitz 1991): Let P be a normal logic program, and I an interpretation. The reduct of P w.r.t. I, denoted P^I, is obtained from P by deleting each rule r such that $B^-(r) \cap I \neq \emptyset$, and removing negated atoms in the remaining rules. An interpretation I is a stable model of P if $I \models P^I$ and there is no $J \subset I$ such that $J \models P^I$. Let $SM(P)$ denote the set of stable models of P. If $SM(P) \neq \emptyset$ then P is coherent. An atom $a \in \mathscr{A}$ is

https://doi.org/10.1017/S1471068414000325 Published online by Cambridge University Press

a cautious consequence of a program P if a belongs to all stable models of P. The set of cautious consequences of P is denoted $CC(P)$.

Example 1

The following is a (ground) program equivalent to the one reported in Section 1:

$R_{out}(1,1,1) \leftarrow \sim R_{out}(1,2,1);\ R_{in}(1,1,1) \leftarrow \sim R_{out}(1,1,1);\ Q(1,1) \leftarrow R_{in}(1,1,1);\ R_{in}(2,2,2) \leftarrow;$
$R_{out}(1,2,1) \leftarrow \sim R_{out}(1,1,1);\ R_{in}(1,2,1) \leftarrow \sim R_{out}(1,2,1);\ Q(1,1) \leftarrow R_{in}(1,2,1);\ R_{in}(2,2,3) \leftarrow;$
$R_{out}(3,2,2) \leftarrow \sim R_{out}(3,3,3);\ R_{in}(3,2,2) \leftarrow \sim R_{out}(3,2,2);\ Q(3,2) \leftarrow R_{in}(3,2,2);\ Q(2,2) \leftarrow;$
$R_{out}(3,3,3) \leftarrow \sim R_{out}(3,2,2);\ R_{in}(3,3,3) \leftarrow \sim R_{out}(3,3,3);\ Q(3,3) \leftarrow R_{in}(3,3,3);\ Q(2,3) \leftarrow.$

The program above has four stable models:

1. $I \cup \{R_{out}(1,1,1),\ R_{out}(3,2,2),\ R_{in}(1,2,1),\ R_{in}(3,3,3),\ Q(3,3)\}$;
2. $I \cup \{R_{out}(1,2,1),\ R_{out}(3,2,2),\ R_{in}(1,1,1),\ R_{in}(3,3,3),\ Q(3,3)\}$;
3. $I \cup \{R_{out}(1,1,1),\ R_{out}(3,3,3),\ R_{in}(1,2,1),\ R_{in}(3,2,2),\ Q(3,2)\}$;
4. $I \cup \{R_{out}(1,2,1),\ R_{out}(3,3,3),\ R_{in}(1,1,1),\ R_{in}(3,2,2),\ Q(3,2)\}$;

where $I = \{R_{in}(2,2,2),\ R_{in}(2,2,3),\ Q(1,1), Q(2,2),\ Q(2,3)\}$ is the set of cautious consequences.

Stable model search. Given a normal ASP program P, its stable models can be computed by means of an algorithm similar to the DPLL backtracking search algorithm (Davis *et al.* 1962) adopted by SAT solvers. In this algorithm, atoms are associated with a truth value among true, false and undefined. Moreover, atoms are associated with a nonnegative integer called *level*. Actually, the backtracking search is usually preceded by *simplification* techniques (Eén and Biere 2005). Simplifications include polynomial time algorithms that (i) identify atoms whose truth value is deterministically implied by the input program, (ii) strengthen and remove rules, and (iii) eliminate atoms by means of rule rewriting. Then, the nondeterministic search starts choosing *branching atoms* according to some heuristic, propagating consequences of these choices until either a stable model is found, or a conflict is detected. The level associated with an atom a is the depth of the search tree in which a has been either chosen or determined, where atoms assigned by (i) have level 0. The propagation step is polynomial, and corresponds to unit propagation in SAT solvers. When a conflict is found, previous choices and their consequences are unrolled until consistency is restored (backjumping; Gaschnig 1979). Modern solvers analyze conflicts in order to learn constraints that are implicit in the original program and inhibit future explorations of the same (conflictual) branch of the search tree. This learning step corresponds to clause learning in SAT (Zhang *et al.* 2001), and is usually complemented with heuristic techniques that control the number of learned constraints, and possibly restart the computation in order to explore different branches of the search tree. Restart policies are based on specific sequences of thresholds that guarantee termination of the algorithm (Gomes *et al.* 1998; Luby *et al.* 1993).

3 Computation of Cautious Consequences

Several strategies for computing cautious consequences of a given program are reported in this section. Some of these strategies aim at solving the problem

https://doi.org/10.1017/S1471068414000325 Published online by Cambridge University Press

producing overestimates of the solution, which are improved and eventually result in the set of cautious consequences of the input program. Among them are the algorithms implemented by the ASP solvers DLV (Alviano *et al.* 2011) and clasp (Gebser *et al.* 2012a), respectively called *enumeration of models* and *overestimate reduction* in the following. Other strategies can in addition produce sound answers during the computation of the complete solution, thus providing underestimates also when termination is not affordable in reasonable time. One of these strategies is *iterative coherence testing*, an adaptation of an algorithm computing backbones of propositional formulas (Marques-Silva *et al.* 2010). To the best of our knowledge, no previous attempt to bring such an algorithm in ASP is reported in the literature. A variant of this algorithm, namely *iterative partial coherence testing*, is also introduced here. Finally, a strategy for obtaining underestimates from enumeration of models and overestimate reduction is presented, which can also be used to improve the other algorithms. More in detail, the algorithms considered here have a common skeleton, reported as Algorithm 1. They receive as input a program P and a set of atoms Q representing answer candidates of a query, and produce as output either the largest subset of Q that only contains cautious consequences of P, in case P is coherent, or $\bot$ when P is incoherent. Initially, the underestimate U and the overestimate O are set to $\emptyset$ and Q, respectively (line 1). A coherence test of P is then performed (lines 2–4) by calling function ComputeStableModel, which actually implements stable model search as described in Section 2. (To simplify the presentation, branching atoms are assumed to be assigned the false truth value.) The first argument of the function is a program P. The second argument is a set of learned constraints, which is initially empty. The third argument is a set C of atoms used to restrict branching atoms of level 1. The function returns either I in case a stable model I of P is found, or $\bot$ otherwise. Note that $\bot$ is returned not only when P is incoherent, but in general when each stable model M of P is such that $C \subseteq M$. Similarly, when I is returned, stable model I satisfies $C \nsubseteq I$. When $C = \mathscr{A}$, the condition $C \nsubseteq I$ is trivially satisfied because $I \subseteq \mathscr{A} \setminus \{\bot\}$ by definition of interpretation. When $C = \{a\}$ for some atom a, instead, this function results in an incremental stable model search in which a is forced to be false (Eén and Sörensson 2003b).

The first stable model found improves the overestimate (line 5). At this point, estimates are improved according to different strategies until they are equal (line 6). EnumerationOfModels adds to P a constraint that eliminates the last stable model found (line 1). In fact, function Constraint($\{a_1, \ldots, a_n\}$) returns a singleton of the form $\{\bot \leftarrow a_1, \ldots, a_n\}$. The algorithm then searches for a new stable model (line 2) to improve the overestimate (line 6). If no new stable model exists, the underestimate is set equal to the overestimate (lines 3–4), thus terminating the computation. OverestimateReduction is similar, but the constraint added is obtained from the current overestimate (line 1). In this way, when a new stable model is found, an improvement of the overestimate is guaranteed, and the constraint can be reduced accordingly (lines 6 and 1).

The strategy implemented by IterativeCoherenceTesting can also improve the underestimate many times during its computation. In fact, one cautious consequence candidate is selected by calling function OneOf (line 1). This candidate is then

https://doi.org/10.1017/S1471068414000325 Published online by Cambridge University Press

Algorithm 1: CautiousReasoning

Input : a program P and a set of atoms Q
Output: atoms in Q that are cautious consequences of P, or $\bot$

```
1 U := ∅;   O := Q;   L := ∅;
2 I := ComputeStableModel(P, L, 𝒜);
3 if I = ⊥ then
4 |   return ⊥;
5 O := O ∩ I;
6 while U ≠ O do
  |   // EnumerationOfModels or other procedure
7 return U;
```

Procedure EnumerationOfModels (A1)

```
1 P := P ∪ Constraint(I);
2 I := ComputeStableModel(P, L, 𝒜);
3 if I = ⊥ then
4 |   U := O;
5 else
  |
6 |   O := O ∩ I;
```

Procedure OverestimateReduction(A2)

```
1 P := P ∪ Constraint(O);
2 I := ComputeStableModel(P, L, 𝒜);
3 if I = ⊥ then
4 |   U := O;
5 else
6 |   P := P \ Constraint(O);
7 |   O := O ∩ I;
```

Procedure IterativeCoherenceTesting(A3)

```
1 a := OneOf(O \ U);
2 I := ComputeStableModel(P, L, {a});
3 if I = ⊥ then
4 |   U := U ∪ {a};
5 else
6 |   O := O ∩ I;
```

Procedure IterativePartialCoherTest(A4)

```
1 a := OneOf(O \ U);
2 I := ComputeUpToNextRestart(P,L,{a});
3 if I = ⊥ then
4 |   U := U ∪ {a};
5 else if I ≠ RESTART then
6 |   O := O ∩ I;
```

Function ComputeStableModel*(P: **program**, L: **learned constraints**, C: **set of atoms**)

Global variables: the underestimate U

```
1 repeat
2 |   U := U ∪ {a ∈ Q | L contains ← ~a};
3 |   I := ComputeUpToNextRestart(P, L, OneOf(C));
4 until I ≠ RESTART;
5 return I;
```

constrained to be false and a stable model is searched (line 2). If none is found then the underestimate can be increased (lines 3–4). Otherwise, the overestimate can be improved (lines 5–6). IterativePartialCoherenceTesting is similar, but forces falsity of a candidate only up to the next restart (lines 1–2). In fact, ComputeStableModel is replaced by ComputeUpToNextRestart, a function that searches for a stable model but also terminates when a restart occurs, in which case it returns the value *RESTART*. In this way, the algorithm can select the most promising candidate after each restart.

https://doi.org/10.1017/S1471068414000325 Published online by Cambridge University Press

Variants of these four algorithms can be obtained by replacing function ComputeStableModel with function ComputeStableModel*, which actually implements stable model search, but also improves the current underestimate after each restart (line 2).

Theorem 1
Let P be a program and $Q \subseteq \mathscr{A}$ a set of atoms. CautiousReasoning(P,Q) terminates after finitely many steps and returns $Q \cap CC(P)$ if P is coherent; otherwise, it returns $\bot$. Moreover, $U \subseteq Q \cap CC(P) \subseteq O$ holds at each step of computation. The claim holds for all variants of Algorithm 1.

4 Implementation and Experiment

We implemented the algorithms introduced in the previous section in order to analyze their performances. Details on the implementation, on the tested benchmarks, and on the obtained results are reported in this section.

4.1 Implementation

Algorithms described in Section 3 are implemented in an experimental branch of the ASP solver WASP (Alviano et al. 2013), distributed under the Apache 2.0 license. Source codes can be downloaded from the branch *queries* of the public GIT repository *https://github.com/alviano/wasp.git*.

WASP implements ASP solving with backjumping (Gaschnig 1979), learning (Zhang *et al.* 2001) and restarts (Gomes *et al.* 1998). More in detail, the branch of WASP used in this experiment implements support propagation via program completion (Lierler and Maratea 2004), branching heuristics and deletion strategy inspired by MiniSAT (Eén and Sörensson 2003a), and simplifications via subsumption and atom elimination techniques as described by Eén and Biere (2005). (Actually, atom elimination, called variable elimination in SAT, is not applied on atoms involved in queries.) In the following, *A*2, *A*3, *A*4 will denote WASP running Algorithm 1 with procedures OverestimateReduction, IterativeCoherenceTesting, and IterativePartialCoherenceTesting, respectively. *A*2*, *A*3*, *A*4* will instead denote the variants using procedure ComputeStableModel*. Procedure EnumerationOfModels is not considered in the analysis since it is significantly outperformed by the other strategies in general.

We also implemented a proof-of-concept prototype of a parallel system, in the following referred to as *multi*. It consists of a master controller implemented in Python that coordinates the execution and the exchange of information of two instances of WASP. In particular, estimates and learned constraints of size at most two are exchanged. In our experiment, *multi* runs *A*2* and *A*4*, but *A*4* in this case does not perform the first coherence check (lines 2–4 of Algorithm 1) in order to avoid a redundant computation. Other combinations of algorithms are possible, but not considered in our analysis. Results for *multi* average real time over three runs.

https://doi.org/10.1017/S1471068414000325 Published online by Cambridge University Press

4.2 Benchmark settings

We compared the implemented algorithms on three benchmarks, corresponding to different applications of cautious reasoning, briefly described below.

Multi-Context Systems Querying (MCS). Multi-context systems (Brewka and Eiter 2007) are a formalism for interlinking heterogeneous knowledge bases, called contexts, using bridge rules that model the flow of information among contexts. Testcases in this benchmark are roughly those of the third ASP competition (Calimeri *et al.* 2014), where each context is modeled by a normal logic program under the stable model semantics. We actually made the testcases harder by requiring the computation of all pairs of the form $\langle c, a\rangle$ such that atom a is true in context c, while in the original testcases a single pair of that form was involved in the query. The benchmark contains 53 of the 73 instances submitted to the third ASP competition. We in fact excluded instances corresponding to incoherent theories, which are solved by the first coherence test in around 6 seconds on the average, and always in less than 14 seconds.

Consistent Query Answering (CQA) is a well-known application of ASP (Arenas *et al.* 2003; Manna *et al.* 2013) described in Section 1. We considered the benchmark proposed by Kolaitis *et al.* (2013), and in particular query *Q3* encoded according to the rewritings by Manna *et al.*. The benchmark contains 13 randomly-generated databases of increasing size ranging from 1000 to 7000 tuples per relation. Each relation contains around 30% of primary key violations.

SAT Backbones (SBB). The backbone of a propositional formula φ is the set of literals that are true in all models of φ. When φ is a set of clauses over variables $v_1, \ldots, v_n$ ($n \geqslant 1$), satisfiability of φ can be modeled in ASP by rules $t_i \leftarrow \sim f_i$ and $f_i \leftarrow \sim t_i$ ($i = 1, \ldots, n$), and introducing a constraint for each clause in φ. Backbone computation thus corresponds to the computation of cautious consequences of an ASP program. The benchmark contains 20 industrial instances used in the SAT Challenge 2012 (Järvisalo *et al.* 2012).

The experiment was run on a Mac Pro equipped with two 3 GHz Intel Xeon X5365 (quad core) processors, with 4 MB of L2 cache and 16 GB of RAM, running Debian Linux 7.3 (kernel ver. 3.2.0-4-amd64). Binaries were generated with the GNU C++ compiler 4.7.3-4 shipped by Debian. The parallel controller was interpreted by Python 3.3.2. Time and memory limits were set to 600 seconds and 8 GB, respectively. Performance was measured using the tool RunLim (*http://fmv.jku.at/runlim/*). All instances were grounded by gringo 3.0.5 (Gebser *et al.* 2011), whose execution time is not included in our analysis because our focus is on propositional programs. We however report that the grounding time was often less than 1 second, with a peak of around 5 seconds for the largest 10 instances of MCS.

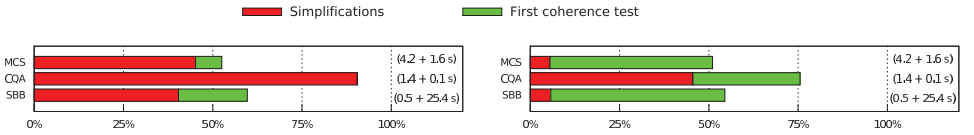
4.3 Discussion of the results

The performance of the algorithms for computing cautious consequences introduced in Section 3 can be studied from several perspectives. On the one hand, we want to

https://doi.org/10.1017/S1471068414000325 Published online by Cambridge University Press

Table 1. *Average running time and number of solved instances*

			A2*			**A3***			**A4***			**multi**		
	#	**$\#_{all}$**	**sol.**	**t**	**t_{all}**	**sol.**	**t**	**t_{all}**	**sol.**	**t**	**t_{all}**	**sol.**	**t**	**t_{all}**
MCS	53	23	23	181.9	181.9	39	254.1	102.6	40	261.6	102.4	40	177.9	69.7
CQA	13	12	12	118.8	118.8	13	89.5	62.1	13	89.2	61.8	13	137.2	51.0
SBB	20	14	15	53.4	41.9	14	65.7	65.7	14	52.1	52.1	17	98.8	51.2
Total	**86**	**49**	**50**	**118.0**	**114.2**	**66**	**136.4**	**76.8**	**67**	**134.3**	**72.1**	**70**	**138.0**	**57.3**



(a) Sound answers (b) Candidates reduction

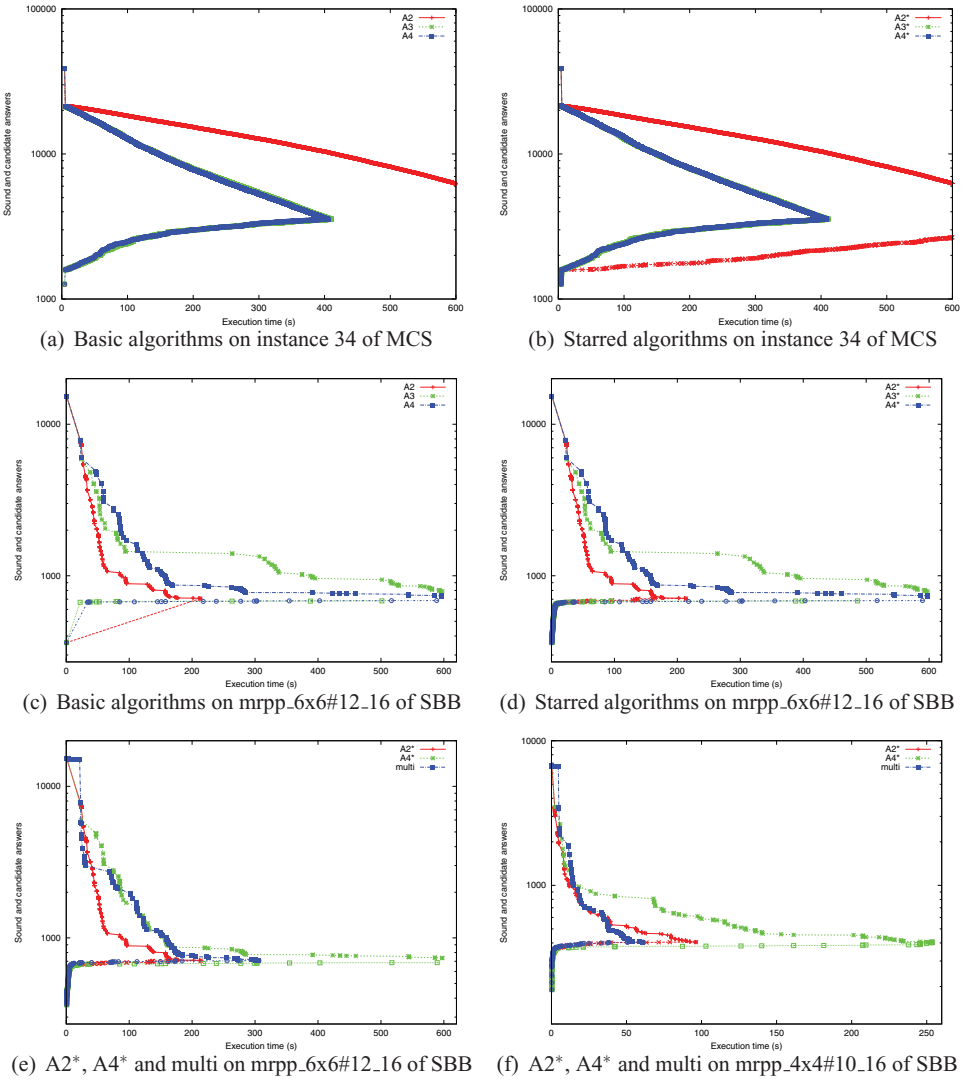
Fig. 1. (Colour online) Sound answers and candidates reduction from simplifications and from first coherence test

know which solution performs better and in which cases. On the other hand, we are interested in analyzing the rate at which each algorithm produces sound answers.

Overall performance. Table 1 summarizes the number of solved instances and average running times. In particular, the first two columns report the total number of instances (#) and the number of instances that are solved by all solvers ($\#_{all}$), respectively; the remaining columns report the number of solved instances within the time-out (sol.), the average running times on solved instances (t) and on instances solved by all algorithms (t_{all}). Comparing the single-process approaches, we note that $A4^*$ solves more instances and is also the fastest on average in the instances solved by all algorithms, even if the performance of $A3^*$ is comparable. $A4^*$ and $A3^*$ outperform $A2^*$ in MCS, and are faster also in CQA. On the other hand, $A2^*$ performs well in SBB, solving one instance more than $A3^*$ and $A4^*$, and being faster on the average. Note that, as expected, if one considers both the number of solved instances and running time, $A2$, $A3$, and $A4$ perform as $A2^*$, $A3^*$, and $A4^*$, respectively. Concerning *multi*, it provides in general the best performance.

Detailed analysis. An important feature of the algorithms analyzed in this paper is the ability to produce both sound answers and overestimates during the computation. Figure 1 reports, for each benchmark, the average percentage of (a) sound answers produced and (b) candidates reduction within the initial steps of the computation. In particular, we plot the effects of simplifications and of the first coherence test. The improvement of the overestimate reported in Figure 1(b) is significant. The first steps of the computation are able to reduce the number of candidates of at least 51% (in MCS) up to around 75% (in CQA). Simplifications are already very effective in CQA, where candidates are reduced of around 45%. It is important to note that the reduction of candidates at this stage applies to all algorithms, while sound answers are produced only by anytime algorithms. This is effective in practice, as shown in Figure 1(a). Indeed, anytime algorithms print from 40% (in SBB) to 90% (in CQA)

https://doi.org/10.1017/S1471068414000325 Published online by Cambridge University Press



(a) Basic algorithms on instance 34 of MCS

(b) Starred algorithms on instance 34 of MCS

(c) Basic algorithms on mrpp_6x6#12_16 of SBB

(d) Starred algorithms on mrpp_6x6#12_16 of SBB

(e) A2*, A4* and multi on mrpp_6x6#12_16 of SBB

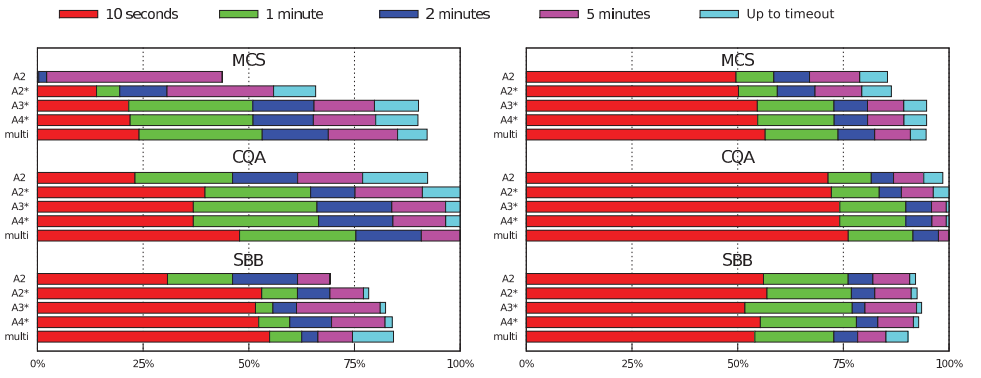
(f) A2*, A4* and multi on mrpp_4x4#10_16 of SBB

Fig. 2. (Colour online) Overestimate and underestimate improvement during execution

of sound answers already after simplifications, which requires few seconds on the average. The first coherence test further improves the underestimate, which ranges from 52% (in MCS) to around 91% (in CQA). However, we observe that the first coherence test may require some time (25s on the average for SBB instances, with a peak of 193s), which motivated the starred variants. In fact, starred variants can produce underestimates at each restart, not only when a coherence test is completed. Actually, $A2^*$, $A3^*$ and $A4^*$ improve progressively the underestimate up to around an additional 30% before the first stable model is found, which is desirable on hard instances.

In order to further confirm the above observations, we analyze in detail the behavior of the algorithms after simplifications. In particular, Figure 2 plots both

https://doi.org/10.1017/S1471068414000325 Published online by Cambridge University Press



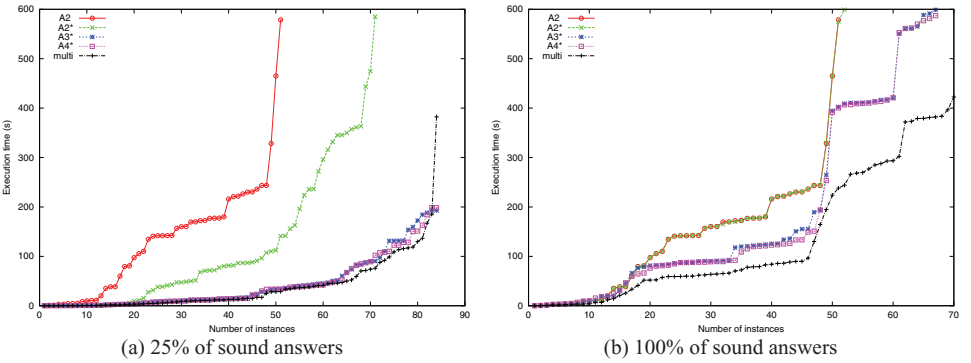
(a) Sound answers (b) Candidates reduction

Fig. 3. (Colour online) Sound answers and candidates reduction after simplification

the number of sound answers (line below) and the number of candidate answers (line above) over time. In particular, Figure 2(b) is devoted to the starred algorithms on an instance of MCS, whereas Figure 2(a) plots the behavior of the basic algorithms on the same instance. First, we note that $A3$ and $A4$ perform similarly and outperform $A2$, which timed out. Notably, $A2^*$ can produce the underestimate (see the bottom line in Figure 2(b)) whereas $A2$ can only print the overestimate (there is no underestimate line for $A2$ in Figure 2(a)). In general, $A3$ and $A4$ are able to improve their estimates better than $A2$. Note that there is a point in the plots for each improvement of estimates, and lines are very dense on MCS instances. This confirms that MCS instances have a huge number of stable models that can be rapidly computed. We observed an analogous behavior for CQA. Plots for SBB instances on Figure 2(c) and Figure 2(d) have, instead, sparse lines, confirming that stable model search is harder for this benchmark. Nevertheless, the starred algorithms can rapidly produce most of the sound answers. A deeper look at Figure 2(c) suggests that $A2$ is much faster than both $A3$ and $A4$ in solving this instance. In fact, it improves the overestimate faster than any other algorithm. Figure 2(e) and Figure 2(f) focus on *multi* and its components. In Figure 2(f), *multi* is faster and improves estimates better than $A2^*$ and $A4^*$. In contrast, *multi* is slower than $A2^*$ in Figure 2(e). A possible cause is the information exchange among processes that modifies the program handled by $A2^*$ with constraints produced by $A3^*$, which in this case results in an harder instance also for the $A2^*$ process. Note that *multi* has a non-deterministic behavior due to its parallel nature. Indeed, information exchanged may be different in different runs.

More insights on the general behavior of the algorithms in the non-deterministic part of the computation can be obtained by looking at Figure 3. In particular, Figure 3(a) reports the average percentage of sound answers produced *after the simplification step*, while candidates reduction is shown on Figure 3(b). $A3$ is not shown in the figure because it performs similarly to $A3^*$ in this perspective. The same holds for $A4$ and $A4^*$. We point out that all bars refer to sound answers and candidates remaining after simplifications, also for $A2$. As a general observation, $A2$

https://doi.org/10.1017/S1471068414000325 Published online by Cambridge University Press



(a) 25% of sound answers (b) 100% of sound answers

Fig. 4. (Colour online) Time performance of algorithms for computing underestimates

prints sound answers only at the end of the computation, while other algorithms are anytime. Consequently, $A2$ does not provide sound answers as soon as the other algorithms, as shown on Figure 3(a). Basically, $A2$ can print something in the first 10s only for easy instances, while $A2^*$ improves a lot in this respect. For example, $A2^*$ outputs around 14% of sound answers already in the first 10s of computation in MCS, while $A2$ produces no output. Nonetheless, $A3^*$ and $A4^*$ perform generally better than $A2^*$. The difference between $A3^*$ and $A4^*$ emerges only in SBB, where finding stable models is harder. In particular, by looking at Figure 3(a), $A4^*$ produces more sound answers than $A3^*$, whereas $A3^*$ is more effective in reducing the number of candidates on Figure 3(b). Note that $A4^*$ may change the candidate to test at each restart, and in our implementation it selects the one with the largest value of activity (which very roughly means the one that was involved more often in conflicts). On the other hand, $A3^*$ insists on the same candidate until the end of a stable model search. As a consequence, $A4^*$ has more chances to find inconsistent branches and, therefore, to improve the underestimate. On the contrary, $A3^*$ has more chances to find a stable model and, thus, to improve the overestimate. This suggests that $A4^*$ should be preferred, since it outputs sound answers more frequently, and also because, as discussed above, $A4^*$ is faster than $A3^*$ on the average (see Table 1). Concerning *multi*, we observe that it outperforms the alternatives on CQA and SBB, while it is less effective in printing sound answers on the average for SBB instances. This is due to the presence of an outlier in this benchmark that spoils the average. That outlier can be explained by the already discussed non-deterministic behavior of *multi*. Analogous considerations can be done for the reduction of candidates on Figure 3(b).

Another perspective on the behavior of the various algorithms can be obtained by looking at Figure 4. Here are reported, for each benchmark, two variants of the classical cactus plot. Recall that, in a cactus plot, the x-axis reports the number of instances that are solved within the time reported on the y-axis. Here we consider variants where on the x-axis is the number of instances for which an algorithm printed 25% (resp. 100%) of sound answers within the time reported on the y-axis. We point out that anytime solvers can print 100% of sound answers before the timeout, even if termination is not reached within the allotted time. Figure 4 confirms

https://doi.org/10.1017/S1471068414000325 Published online by Cambridge University Press

that $A2$ is slower than anytime algorithms in printing sound answers. It is interesting to note that the anytime $A2^*$ improves sensibly $A2$ in all benchmarks, especially at the beginning of the computation. We also observe that $A4^*$ is slightly preferable to $A3^*$, and *multi* is the fastest solution. Note that the differences are more evident in the plots on the left that focus on the first 25% of sound answers. Finally, we confirm that $A2^*$ is the fastest single-process implementation in SBB. Nonetheless, $A3^*$ and $A4^*$ print more sound answers also in non-terminating instances.

To sum up, anytime algorithms are convenient in practice because they determine several sound answers already in the first steps of computation. In particular, the starred variants are preferable ($A4^*$ leading the group) because they provide sound answers as soon as possible and are thus effective also in non-terminating instances. Finally, the parallel system combining $A4^*$ with $A2^*$ is the best variant overall.

5 Related Work

The computation of cautious consequences in ASP is a feature available in two solvers, namely DLV (Maratea *et al.* 2008) and clasp (Gebser *et al.* 2012a). The algorithm implemented by DLV is enumeration of models, while clasp implements overestimate reduction. Our implementation differs from these solvers especially with respect to the output produced during the computation of cautious consequences. In fact, DLV does not print any form of estimation during the computation, and clasp only prints overestimates. Our implementation, instead, is anytime and thus prints both underestimates and overestimates during the computation. Underestimates provide sound answers also when termination is not affordable in reasonable time, and are thus of practical importance for hard problems. It is interesting to observe that among the strategies supported by our implementation there is $A2^*$, an anytime variant of the algorithm used by clasp that performed very well on two of our three benchmarks. We also note that DLV and clasp feature brave reasoning, which is not currently supported by our implementation.

Clasp, being a parallel ASP solver (Gebser *et al.* 2012b), also supports the parallel computation of cautious consequences by means of the overestimate reduction algorithm. Our proposal is different, as it is based on the combination of two different algorithms, namely iterative partial coherence testing and overestimate reduction, for reducing both estimates at the same time. However, we observe that our parallel implementation is a proof-of-concept prototype obtained by combining two instances of WASP (properly modified to share short learned constraints and answer estimates) controlled by a Python script. It is devised to show the benefits of combining two different algorithms, while a more efficient implementation is subject of future work.

The computation of cautious consequences of a ground program is related to the problem of backbone computations of propositional formulas (Marques-Silva *et al.* 2010; Slaney and Walsh 2001). In fact, the backbone of a propositional formula φ is the set of literals that are true in all models of φ. Several algorithms for computing backbones of propositional formulas are based on variants of the iterative consistency testing algorithm (Marques-Silva *et al.* 2010; Janota *et al.* 2014),

https://doi.org/10.1017/S1471068414000325 Published online by Cambridge University Press

which essentially corresponds to the iterative coherence testing algorithm analyzed in this paper. Backbone search algorithms usually feature additional techniques for removing candidates to be tested, such as *implicant reduction* and *core-based chunking* (Ravi and Somenzi 2004). Most of the implicant reduction techniques are not applicable to normal ASP programs because of the intrinsic minimality of stable models. For example, backbone search algorithms can reduce their overestimate by removing all unassigned variables when a (partial) model is found; in our setting, ASP solvers always terminate with a complete assignment. Core-based chunking, instead, requires a portfolio of algorithms (Janota *et al.* 2014) in order to be effective, which is beyond the scope of this paper.

Note that all considered algorithms work on ground programs. Combinations with query optimization techniques such as magic sets (Greco 2003; Alviano and Faber 2011) are possible but not the focus of the paper.

6 Conclusion

Several algorithms for computing cautious consequences of ASP programs were analyzed in this paper. At the time of this writing, ASP solvers do not implement anytime algorithms, which means that computation must terminate in order to obtain *some* cautious consequences. On the other hand, the computation of cautious consequences is similar to the computation of backbones of propositional theories, for which anytime algorithms do exist. We adapted one of these algorithms to cautious reasoning, showing that underestimates can be effectively obtained in reasonable time also for hard instances. Moreover, we introduced a general strategy to obtain anytime variants of existing algorithms such as those implemented by DLV and clasp. All algorithms as well as a proof-of-concept parallel implementation were implemented in the solver WASP. Our empirical evaluation highlights that sound answers are computable within the first seconds of computation in many cases. Moreover, the performance of the parallel system is encouraging and leaves space for future work on this subject.

Acknowledgements

This research has been partly supported by Regione Calabria under the EU Social Fund and project PIA KnowRex POR FESR 2007–2013, and the Italian Ministry of University and Research under PON project "Ba2Know S.I.-LAB" n. PON03PE_0001.

References

Alviano, M., Dodaro, C., Faber, W., Leone, N., and Ricca, F. 2013. WASP: A native ASP solver based on constraint learning. In *LPNMR*, P. Cabalar and T. C. Son, Eds. LNCS, vol. 8148. Springer, 54–66.

Alviano, M. and Faber, W. 2011. Dynamic magic sets and super-coherent answer set programs. *AI Communications*. IOS Press *24,* 2, 125–145.

https://doi.org/10.1017/S1471068414000325 Published online by Cambridge University Press

Alviano, M., Faber, W., Leone, N., Perri, S., Pfeifer, G., and Terracina, G. 2011. The disjunctive datalog system DLV. In *Datalog 2.0*, G. Gottlob, Ed. Vol. 6702. Springer Berlin/Heidelberg, 282–301.

Arenas, M., Bertossi, L. E., and Chomicki, J. 2003. Answer sets for consistent query answering in inconsistent databases. *Theory and Practice of Logic Programming 3,* 4–5, 393–424.

Baral, C. 2003. *Knowledge Representation, Reasoning and Declarative Problem Solving*. Cambridge University Press.

Brewka, G. and Eiter, T. 2007. Equilibria in heterogeneous nonmonotonic multi-context systems. In *AAAI*. AAAI Press, 385–390.

Calimeri, F., Ianni, G., and Ricca, F. 2014. The third open answer set programming competition. *Theory and Practice of Logic Programming 14,* 1, 117–135.

Davis, M., Logemann, G., and Loveland, D. 1962. A machine program for theorem proving. *Commun. ACM 5*, 394–397.

Eén, N. and Biere, A. 2005. Effective preprocessing in SAT through variable and clause elimination. In *SAT*. LNCS, vol. 3569. Springer, 61–75.

Eén, N. and Sörensson, N. 2003a. An extensible SAT-solver. In *SAT*, E. Giunchiglia and A. Tacchella, Eds. LNCS, vol. 2919. Springer, 502–518.

Eén, N. and Sörensson, N. 2003b. Temporal induction by incremental sat solving. *Electr. Notes Theor. Comput. Sci. 89,* 4, 543–560.

Eiter, T. 2005. Data integration and answer set programming. In *LPNMR*, C. Baral, G. Greco, N. Leone, and G. Terracina, Eds. LNCS, vol. 3662. Springer, 13–25.

Eiter, T., Ianni, G., Lukasiewicz, T., Schindlauer, R., and Tompits, H. 2008. Combining answer set programming with description logics for the semantic web. *Artif. Intell. 172,* 12–13, 1495–1539.

Gaschnig, J. 1979. Performance measurement and analysis of certain search algorithms. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, USA. Technical Report CMU-CS-79–124.

Gebser, M., Kaminski, R., König, A., and Schaub, T. 2011. Advances in *gringo* series 3. In *LPNMR*, J. P. Delgrande and W. Faber, Eds. LNCS, vol. 6645. Springer, 345–351.

Gebser, M., Kaufmann, B., and Schaub, T. 2012a. Conflict-driven answer set solving: From theory to practice. *Artif. Intell. 187*, 52–89.

Gebser, M., Kaufmann, B., and Schaub, T. 2012b. Multi-threaded ASP solving with clasp. *Theory and Practice of Logic Programming 12,* 4–5, 525–545.

Gelfond, M. and Kahl, Y. 2014. *Knowledge Representation, Reasoning, and the Design of Intelligent Agents: The Answer-Set Programming Approach*. Cambridge University Press.

Gelfond, M. and Lifschitz, V. 1991. Classical negation in logic programs and disjunctive databases. *New Generation Computing 9*, 365–385.

Gomes, C. P., Selman, B., and Kautz, H. A. 1998. Boosting combinatorial search through randomization. In *Proceedings of AAAI/IAAI 1998*. AAAI Press, 431–437.

Greco, S. 2003. Binding propagation techniques for the optimization of bound disjunctive queries. *IEEE Transactions on Knowledge and Data Engineering 15,* 2 (March/April), 368–385.

Janota, M., Lynce, I., and Marques-Silva, J. 2014. Algorithms for computing backbones of propositional formulae. *AI Commun.*. To appear.

Järvisalo, M., Berre, D. L., Roussel, O., and Simon, L. 2012. The international SAT solver competitions. *AI Magazine 33,* 1.

Kolaitis, P. G., Pema, E., and Tan, W.-C. 2013. Efficient querying of inconsistent databases with binary integer programming. *PVLDB 6,* 6, 397–408.

https://doi.org/10.1017/S1471068414000325 Published online by Cambridge University Press

Leone, N., Pfeifer, G., Faber, W., Eiter, T., Gottlob, G., Perri, S., and Scarcello, F. 2006. The DLV system for knowledge representation and reasoning. *ACM Transactions on Computational Logic 7,* 3 (July), 499–562.

Lierler, Y. and Maratea, M. 2004. Cmodels-2: SAT-based answer set solver enhanced to non-tight programs. In *Proceedings of the 7th International Conference on Logic Programming and Non-Monotonic Reasoning (LPNMR-7)*, V. Lifschitz and I. Niemelä, Eds. LNAI, vol. 2923. Springer, 346–350.

Lifschitz, V. 2002. Answer set programming and plan generation. *Artificial Intelligence 138*, 39–54.

Luby, M., Sinclair, A., and Zuckerman, D. 1993. Optimal speedup of Las Vegas algorithms. *Inf. Process. Lett. 47*, 173–180.

Manna, M., Ricca, F., and Terracina, G. 2013. Consistent query answering via ASP from different perspectives: Theory and practice. *Theory and Practice of Logic Programming 13,* 2, 227–252.

Maratea, M., Ricca, F., Faber, W., and Leone, N. 2008. Look-back techniques and heuristics in dlv: Implementation, evaluation, and comparison to qbf solvers. *J. Algorithms 63,* 1–3, 70–89.

Marek, V. W. and Truszczyński, M. 1999. Stable models and an alternative logic programming paradigm. In *The Logic Programming Paradigm – A 25-Year Perspective*, K. R. Apt, V. W. Marek, M. Truszczyński, and D. S. Warren, Eds. Springer Verlag, 375–398.

Marques-Silva, J., Janota, M., and Lynce, I. 2010. On computing backbones of propositional theories. In *ECAI*, H. Coelho, R. Studer, and M. Wooldridge, Eds. Frontiers in Artificial Intelligence and Applications, vol. 215. IOS Press, 15–20.

Niemelä, I. 1999. Logic programming with stable model semantics as constraint programming paradigm. *Annals of Mathematics and Artificial Intelligence 25,* 3–4, 241–273.

Ravi, K. and Somenzi, F. 2004. Minimal assignments for bounded model checking. In *TACAS*, K. Jensen and A. Podelski, Eds. LNCS, vol. 2988. Springer, 31–45.

Slaney, J. K. and Walsh, T. 2001. Backbones in optimization and approximation. In *IJCAI*, B. Nebel, Ed. Morgan Kaufmann, 254–259.

Zhang, L., Madigan, C. F., Moskewicz, M. W., and Malik, S. 2001. Efficient conflict driven learning in boolean satisfiability solver. In *Proceedings of the International Conference on Computer-Aided Design (ICCAD 2001)*. 279–285.

https://doi.org/10.1017/S1471068414000325 Published online by Cambridge University Press