

Domain-Specific Heuristics in Answer Set Programming: A Declarative Non-Monotonic Approach

Richard Comploi-Taupe (*corresponding author*)
*Siemens AG Österreich and
Alpen-Adria-Universität Klagenfurt, Austria*

RICHARD.TAUPE@SIEMENS.COM

Gerhard Friedrich
Alpen-Adria-Universität Klagenfurt, Austria

GERHARD.FRIEDRICH@AAU.AT

Konstantin Schekotihin
Alpen-Adria-Universität Klagenfurt, Austria

KONSTANTIN.SCHEKOTIHIN@AAU.AT

Antonius Weinzierl
TU Wien, Vienna, Austria

WEINZIERL@KR.TUWIEN.AC.AT

Abstract

Domain-specific heuristics are an essential technique for solving combinatorial problems efficiently. Current approaches to integrate domain-specific heuristics with Answer Set Programming (ASP) are unsatisfactory when dealing with heuristics that are specified non-monotonically on the basis of partial assignments. Such heuristics frequently occur in practice, for example, when picking an item that has not yet been placed in bin packing. Therefore, we present novel syntax and semantics for declarative specifications of domain-specific heuristics in ASP. Our approach supports heuristic statements that depend on the partial assignment maintained during solving, which has not been possible before. We provide an implementation in ALPHA that makes ALPHA the first lazy-grounding ASP system to support declaratively specified domain-specific heuristics. Two practical example domains are used to demonstrate the benefits of our proposal. Additionally, we use our approach to implement informed search with A*, which is tackled within ASP for the first time. A* is applied to two further search problems. The experiments confirm that combining lazy-grounding ASP solving and our novel heuristics can be vital for solving industrial-size problems.

1. Introduction

Answer Set Programming (ASP) (Baral, 2003; Gebser et al., 2012; Gelfond & Kahl, 2014; Lifschitz, 2019) is a declarative knowledge representation formalism that has been applied successfully in a variety of industrial and scientific applications (Erdem et al., 2016; Falkner et al., 2018) such as configuration (Falkner et al., 2016), team building (Ricca et al., 2012), routing (Gebser et al., 2018), or scheduling (Dodaro et al., 2019).

The *ground-and-solve* approach is the predominant method employed by state-of-the-art ASP solvers. In *ground-and-solve*, solvers first instantiate the given non-ground program and then apply various strategies to find answer sets of the obtained ground program (Gebser et al., 2018). *Ground-and-solve* is applied by systems such as CLINGO (Gebser et al., 2019), DLV (Adrian et al., 2018; Leone et al., 2006), or DLV2 (Alviano et al., 2017).

However, modern applications manifested two issues with the *ground-and-solve* approach. The first issue is the so-called *grounding bottleneck*: Large problem instances in

industrial applications often cannot be grounded by modern grounders like GRINGO (Gebser et al., 2011) or I-DLV (Calimeri et al., 2017) in acceptable time and space (Eiter et al., 2007). The second issue is that, even if the problem can be grounded, computation of answer sets might take considerable time, as indicated by ASP Competition reports (Calimeri et al., 2016; Gebser et al., 2017, 2020).

Both issues were recently addressed by many researchers. First, to overcome the grounding bottleneck, lazy-grounding ASP systems interleave grounding and solving to instantiate and store only relevant parts of the ground program in memory. Lazy grounding is implemented by systems such as GASP (Palù et al., 2009), ASPERIX (Lefèvre et al., 2017), OMIGA (Dao-Tran et al., 2012), and ALPHA (Weinzierl, 2017). The second performance-related issue is tackled by modern solvers using various techniques among which domain-specific heuristics play a central role. Examples of such approaches include both declarative and procedural approaches. A seamless and declarative integration of heuristics into ASP encodings has been proposed by Gebser et al. (2013), whereas an approach using procedural heuristics that interact directly with the internal decision-making procedures is discussed by Dodaro et al. (2016).

However, there is no ASP system that addresses both issues simultaneously. That is, previous research has not addressed declarative heuristics for the lazy-grounding case. In addition, the existing integrations of domain-specific heuristics with ASP solving are unsatisfactory: Procedural heuristics counteract the declarative nature of ASP, and the existing declarative approach makes it quite tedious and counter-intuitive to model *dynamic heuristics* for reasoning about partial assignments, which is an essential feature for practical applications. For example, heuristics in product configuration or for bin packing may need to compute the amount of space left after placing a component or an item. So far, such dynamic reasoning has been supported by the existing procedural approach, but not by the declarative one.

We tackle the challenge of finding a satisfying solution addressing both the grounding bottleneck and the runtime performance issue. To this end, we extend the existing declarative approach by Gebser et al. (2013) with dynamic heuristics while at the same time keeping the language simple and easy to use. Second, we integrate our approach into a lazy-grounding system, which requires non-trivial adaptations due to the different solving mechanisms in effect.

1.1 Contributions

In this work, we present a novel approach to dynamic declarative domain-specific heuristics for ASP. We combine this approach with lazy grounding to facilitate the solving of large and complex problems. In summary, our work provides the following contributions:

- We present novel semantics that make declarative specifications of domain-specific heuristics more intuitive by allowing heuristics to depend non-monotonically on the partial assignment maintained during solving (i.e., heuristics may be applicable in one partial assignment and cease to be applicable at a later stage).
- We propose a language for declarative specifications of domain-specific heuristics within answer-set programs that can be seen as a variant of the one introduced by Gebser et al. (2013), and we formally define the language by an EBNF grammar.

- We show how to integrate our language into a lazy-grounding ASP system and provide a reference implementation within the well-known lazy-grounding system ALPHA.
- We demonstrate how to use our approach to model domain-specific heuristics for two practical example domains: the House Reconfiguration Problem (HRP) and the Partner Units Problem (PUP). Additionally, we use our approach to implement state-space search with A*, which is tackled within ASP for the first time.¹ Integrating A* and ASP allows, on the one hand, the declarative specification of states and their successor states employing ASP and, on the other hand, the application of informed search strategies exploiting heuristic functions. We apply A* to two further search problems.
- Finally, we demonstrate how solving performance can profit from our approach by presenting experimental results in these practical domains.

1.2 Organization

Preliminaries are covered by Section 2, before Section 3 outlines the state of the art in domain-specific heuristics. On this basis, we introduce our novel semantics in Section 4. Section 5 explains how to integrate our approach into a lazy-grounding ASP system. Applications and experimental results are presented in Section 6, and Section 7 concludes the article.

2. Preliminaries

Answer Set Programming (ASP) (Baral, 2003; Gebser et al., 2012; Gelfond & Kahl, 2014; Lifschitz, 2019) is an approach to declarative programming. Instead of stating how to solve a problem, the programmer formulates the problem as a logic program specifying the search space and the properties of valid solutions. An ASP solver then finds models (so-called *answer sets*) for this logic program, which correspond to solutions for the original problem.

2.1 Syntax

ASP offers a rich input language, of which we introduce only the core concepts needed in this paper. For a comprehensive definition of ASP’s syntax and semantics, we refer to Calimeri et al. (2020).

Let $\langle \mathcal{V}, \mathcal{C}, \mathcal{F}, \mathcal{P} \rangle$ define a first-order language, where \mathcal{V} is a set of variable symbols, \mathcal{C} is a set of constant symbols, \mathcal{F} is a set of function symbols, and \mathcal{P} is a set of predicate symbols.

A classical *atom* is of the form $p(t_1, \dots, t_n)$, where $p \in \mathcal{P}$ is a predicate symbol and t_1, \dots, t_n are *terms*. Each variable $v \in \mathcal{V}$ and each constant $c \in \mathcal{C}$ is a term. Furthermore, for $f \in \mathcal{F}$, $f(t_1, \dots, t_n)$ is a function term.

An answer-set program P is a finite set of (disjunctive) rules of the form

$$h_1 \vee \dots \vee h_k \leftarrow b_1, \dots, b_m, \text{ not } b_{m+1}, \dots, \text{ not } b_n. \quad \langle 1 \rangle$$

1. Cabalar et al. (2019) present an A* algorithm which employs ASP through repeated calls to an ASP solver for computing states and their successors.

where h_1, \dots, h_k and b_1, \dots, b_n are atoms and *not* is negation as failure (a.k.a. default negation), which refers to the absence of information, i.e., an atom is assumed to be false as long as it is not derived by some rule. A *literal* is either an atom a or its negation $\text{not } a$. Given a rule r of the form $\langle l \rangle$, $\text{head}(r) = \{h_1, \dots, h_k\}$ is called the *head* of r , and $\text{body}(r) = \{b_1, \dots, b_m, \text{not } b_{m+1}, \dots, \text{not } b_n\}$ is called the *body* of r . By $\text{body}^+(r) = \{b_1, \dots, b_m\}$ and $\text{body}^-(r) = \{b_{m+1}, \dots, b_n\}$ we denote the positive and negative atoms in the body of r , respectively. A rule r where $\text{head}(r) = \emptyset$, e.g., $\leftarrow b$, is called *integrity constraint*, or simply *constraint*. A rule r where $\text{body}(r) = \emptyset$, e.g., $h \leftarrow \cdot$, is called *fact*. A rule is ground if all its atoms are variable-free. A ground program comprises only ground rules.

2.2 Semantics

Given a program P , the *Herbrand universe* of P , denoted by U_P , consists of all integers and of all ground terms constructible from constant symbols and function symbols appearing in P . The *Herbrand base* of P , denoted by B_P , is the set of all ground classical atoms that can be built by combining predicates appearing in P with terms from U_P as arguments (Calimeri et al., 2020).

A substitution σ is a mapping from variables \mathcal{V} to elements of the Herbrand universe U_P of a program P . Let O be a rule, an atom, or a literal, then by $O\sigma$ we denote a rule, atom, or literal obtained by replacing each variable $v \in \text{vars}(O)$ by $\sigma(v)$. The function vars maps any rule, atom, literal, or any other object containing variables to the set of variables it contains. For instance, $\text{vars}(a(X)) = \{X\}$ and for a rule $r_1 : a(X) \leftarrow b(X, Y)$, $\text{vars}(r_1) = \{X, Y\}$.

As usual, we assume rules to be *safe*, which is the case for a rule r if $\text{vars}(r) \subseteq \bigcup_{a \in \text{body}^+(r)} \text{vars}(a)$.

The (ground) instantiation of a rule r equals $r\sigma$ for some substitution σ , which maps all variables in r to ground terms. The (ground) instantiation $\text{grd}(P)$ of a program P is the set of all possible instantiations of the rules in P (Calimeri et al., 2020; Faber et al., 2011). Function symbols may cause the Herbrand base and the full grounding of a program to be infinite (Alviano et al., 2011). By restricted usage of function symbols, answer-set programs can be designed in a way that reasoning is decidable.

An *Herbrand interpretation* for a program P is a set of ground classical atoms $I \subseteq B_P$. A ground classical atom a is true w.r.t. an interpretation I , denoted $I \models a$, iff $a \in I$. A ground literal $\text{not } a$ is true w.r.t. an interpretation I , denoted $I \models \text{not } a$, iff $I \not\models a$. A rule r is *satisfied* w.r.t. I , denoted $I \models r$, if some head atom is true w.r.t. I ($\exists h \in \text{head}(r) : I \models h$) whenever all body literals are true w.r.t. I ($\forall b \in \text{body}(r) : I \models b$). An interpretation I is a model of P , denoted $I \models P$, if $I \models r$ for all rules $r \in \text{grd}(P)$ (Faber et al., 2011).

Given a ground program P and an interpretation I , let P^I denote the transformed program obtained from P by deleting rules in which a body literal is false w.r.t. I : $P^I = \{r \mid r \in P, \forall b \in \text{body}(r) : I \models b\}$ (Faber et al., 2011).

An interpretation I of a program P is an *answer set* of P if it is a subset-minimal model of $\text{grd}(P)^I$, i.e., I is a model of $\text{grd}(P)^I$ and there exists no $I' \subsetneq I$ that is a model of $\text{grd}(P)^I$ (Faber et al., 2011).

2.3 Notation

In this section, we introduce some notation that will be used later in the article.

An *assignment* A over B_P is a set of signed literals $\mathbf{T} a$, $\mathbf{F} a$, or $\mathbf{M} a$, where $\mathbf{T} a$ and $\mathbf{F} a$ express that an atom a is true and false, respectively, and $\mathbf{M} a$ indicates that a “must-be-true”. \mathbf{M} means that an atom must eventually become true by derivation in a correct solution extending the current partial assignment, but no derivation has yet been found that would make the atom true. Intuitively, $\mathbf{T} b \in A$ means that b is true and justified, i.e., derived by a rule that fires under A , while $\mathbf{M} b \in A$ only indicates that b is true but potentially not derived. Let $A_s = \{a \mid s a \in A\}$ for $s \in \{\mathbf{F}, \mathbf{M}, \mathbf{T}\}$ denote the set of atoms occurring with a specific sign in assignment A . We assume assignments to be *consistent*, i.e., no negative literal may also occur positively ($A_{\mathbf{F}} \cap (A_{\mathbf{M}} \cup A_{\mathbf{T}}) = \emptyset$), and every positive literal must also occur with must-be-true ($A_{\mathbf{T}} \subseteq A_{\mathbf{M}}$). The latter condition ensures that assignments are monotonically growing (w.r.t. set inclusion) in case an atom that was must-be-true becomes justified by a rule deriving it and hence changes to true.

An assignment A is *complete* if every atom in the Herbrand base is assigned true or false ($\forall a \in B_P : a \in A_{\mathbf{F}} \cup A_{\mathbf{T}}$). An assignment that is not complete is *partial*.

The function $\text{truth}_A : B_P \rightarrow \{\mathbf{T}, \mathbf{M}, \mathbf{F}, \mathbf{U}\}$ for a (partial) assignment A maps an atom to the truth value that the atom is currently assigned in the given assignment, or to \mathbf{U} if the atom is currently unassigned:

$$\text{truth}_A(a) = \begin{cases} \mathbf{F} & a \in A_{\mathbf{F}}, \\ \mathbf{M} & a \in A_{\mathbf{M}} \setminus A_{\mathbf{T}}, \\ \mathbf{T} & a \in A_{\mathbf{T}}, \\ \mathbf{U} & \text{otherwise.} \end{cases}$$

For example, if $A = \{\mathbf{M} a, \mathbf{T} a, \mathbf{M} b\}$, then $A_{\mathbf{F}} = \emptyset$, $A_{\mathbf{M}} = \{a, b\}$, $A_{\mathbf{T}} = \{a\}$, $\text{truth}_A(a) = \mathbf{T}$, and $\text{truth}_A(b) = \mathbf{M}$.

Note that historically, the concept of a truth value “must-be-true” was introduced to ASP solving by the DLV solver for efficiency (Eiter et al., 2003; Faber et al., 1999). The same efficiency was later realized without needing “must-be-true” by use of so-called source pointers, together with unfounded-set checks, and completion (Gebser et al., 2012). These techniques ensure that in a final assignment A , all atoms assigned true are derived by a rule that fires in A . The distinction between an atom being true or must-be-true, however, is still present in most modern ASP solvers. The only difference to our approach is that other solvers represent must-be-true implicitly via the internal state of source-pointers and the result of unfounded-set propagation, while we chose to represent “must-be-true” explicitly as a truth value.

Expressing “must-be-true” explicitly also allows for more fine-grained domain-specific heuristics, as they may now distinguish between whether a goal g has been reached in an assignment A by being assigned true (i.e., $\text{truth}_A(g) = \mathbf{T}$) or still must be reached (i.e., $\text{truth}_A(g) = \mathbf{M}$). The former case may indicate that a rule checking the goal condition fired and derived g , while the latter may arise from a constraint $\leftarrow \text{not } g$, which states that the goal must be met in any answer set. Our approach to domain-specific heuristics allows heuristics to distinguish both cases.

Many useful language constructs have been introduced to extend the basic language of ASP defined in Sections 2.1 and 2.2. We discuss such extensions only briefly and refer to Calimeri et al. (2020) and Gebser et al. (2019) for full details.

A *cardinality atom* is of the form $lb \{a_1 : l_{1_1}, \dots, l_{m_1}; \dots; a_n : l_{1_n}, \dots, l_{m_n}\} ub$, where, for $1 \leq i \leq n$, $a_i : l_{1_i}, \dots, l_{m_i}$ represents a *conditional literal* in which a_i (the head of the conditional literal) is a classical atom and all l_{j_i} are literals, and lb and ub are integer terms indicating a lower and an upper bound, respectively. If one or both of the bounds are not given, their defaults are used, i.e., 0 for lb and ∞ for ub . A cardinality atom is satisfied if $lb \leq |C| \leq ub$ holds, where C is the set of head atoms in the cardinality atom that are satisfied together with their conditions (e.g., l_{1_i}, \dots, l_{m_i} for a_i).

A *choice rule* is a rule r with $\text{head}(r)$ consisting of a cardinality atom. For example, consider the choice rule $1 \{a; b\} 2 \leftarrow$. A program consisting only of this rule has three answer sets: $\{a\}$, $\{b\}$, and $\{a, b\}$. Subset minimality usually required by answer sets is here circumvented by a rewriting; for details, see Calimeri et al. (2020).

As an extension of cardinality atoms, ASP also supports aggregate atoms that apply aggregate functions like *max*, *min* or *sum* to sets of literals. An aggregate atom is satisfied if the value computed by the aggregate function respects the given bounds, e.g., $1 = \# \text{sum}\{1 : a; 2 : b\}$ is satisfied if a but not b is true.

Built-in predicates with a fixed meaning, such as $=, \neq, \leq, \geq, <, >$, and arithmetic operations such as $+, -, \times, /$, and \backslash (modulo) are also used in infix notation, and $|t|$ denotes the absolute value of an integer term t .

2.4 Lazy Grounding

The grounding bottleneck is a well-known issue of traditional ASP solving, where the input program is grounded first before the ground, i.e., variable-free, program is solved. There have been many attempts to mitigate this issue with more optimised and clever grounding procedures (see, for example, Kaufmann et al., 2016; Calimeri et al., 2017; Gebser et al., 2011; Faber et al., 2012) or through formulations of answer sets in formalisms that do not need grounding per se, e.g., query-driven evaluation (cf. Cat et al., 2015; Marple et al., 2017; Arias et al., 2018) or circumscriptions (cf. Ferraris, Lee, & Lifschitz, 2011).

Lazy grounding is an approach that interleaves the solving and grounding phases, such that computations are guaranteed to yield all answer sets. The foundation for lazy grounding is known as the computation sequence and has been developed by Liu et al. (2007). A computation sequence $\mathbf{S} = \langle S_0, S_1, \dots, S_n \rangle$ is a sequence of partial assignments that is monotonically growing (w.r.t. set inclusion). Every element S_i of the sequence represents the state of the computation at step i . The first element of the sequence is empty ($S_0 = \emptyset$), and every other element S_i contains the signed literals that can be derived from the preceding partial assignment S_{i-1} in the program P .

Since each element of a computation sequence is a partial assignment containing signed literals, and the sequence is monotonically growing, each S_i contains atoms assigned \mathbf{T} that will remain true in all extensions of S_i , and atoms assigned \mathbf{F} that will definitely remain false in all extensions of S_i .

A rule r is said to be *applicable* in S_i if $\{\mathbf{T} a \mid a \in \text{body}^+(r)\} \subseteq S_i$ and $\{\mathbf{M} a \mid a \in \text{body}^-(r)\} \cap S_i = \emptyset$, i.e., if the positive body is satisfied and S_i does not contradict the

negative body. An applicable rule is said to *fire* when its negative body is chosen to be false by the solver’s search procedure. Firing a rule allows deriving its head. For each applicable rule the computation sequence may split in two, one S_{i+1} where the rule is assumed to fire and one S'_{i+1} where the rule does not fire. This property allows capturing the guessing mechanism of ASP.

Based on the fact that the computation sequence only needs to know those ground rules that are applicable, lazy-grounding ASP solvers can ground a rule lazily whenever the ground instance becomes applicable. Thus, only those rules are grounded, whose positive body holds in the current partial assignment.

The first lazy-grounding ASP solvers based on the computation sequence principle were GASP by Palù et al. (2009) and ASPERIX by Lefèvre and Nicolas (2009), Lefèvre et al. (2017). Later Dao-Tran et al. (2012) built the OMIGA solver that uses a RETE network for efficient grounding. Unfortunately, all these solvers suffered from a lack of efficient solving techniques widely used in traditional ground-and-solve systems.

Efficient ground-and-solve systems for ASP usually employ a technique called conflict-driven clause-learning (CDCL) or a closely related variant called conflict-driven nogood-learning (CDNL). This is employed in most SAT solvers (Silva et al., 2009; Zhang et al., 2001) and modern ASP solvers like CLINGO (Gebser et al., 2012) or DLV2 (Alviano et al., 2017). As CDCL works on clauses, a given ASP program is first transformed into an equivalent set of clauses. Then the models of these clauses are computed using a DPLL-style algorithm, which at its core gains insight into the given problem by learning new clauses from conflicts that are encountered during search. Note that CDNL only differs from CDCL by using the dual form of clauses, which are called nogoods. More details about CDNL and nogood representations are given in Section 5, where the implementation of our domain-specific heuristics is presented.

The ALPHA system by Weinzierl (2017) has combined the most important technique for efficient solving, CDNL, with lazy grounding. It is the most recent lazy-grounding ASP system available.

Computation sequences require a normal logic program as input (i.e., rules of the form $\langle 1 \rangle$ without disjunction, cardinality atoms, and aggregate atoms, (cf. Liu et al., 2007; Lefèvre et al., 2017; Weinzierl, 2017). Hence lazy-grounding systems usually only accept normal logic programs or, in the case of ALPHA, rewrite enhanced ASP constructs like aggregates or choice rules into normal rules. Disjunctive heads can only be rewritten if they are head-cycle-free (Ben-Eliyahu & Dechter, 1994). Therefore, in our examples, each rule contains at most one head atom. Although this appears to be rather limiting, ALPHA is close to supporting the full range of the ASP-Core-2 language through program transformations, including aggregates, choice rules, and integer arithmetics, which are very useful constructs in practical applications. In many cases, program transformations result in first-order encodings, which a lazy-grounding system only needs to instantiate for those ground instances actually encountered during search. So these transformations enable significant performance benefits in some cases (Bomanson, Janhunen, & Weinzierl, 2019).

Furthermore, restricting the input to head-cycle-free programs is not really limiting, as programs containing disjunction with head-cycles are usually employed only when problems on the second level of the polynomial hierarchy, i.e., in Σ_2^P , are to be solved. The worst-case complexity of normal logic programs without disjunction but with variables, however,

is already well beyond Σ_2^P and **PSPACE**, in fact it is at least **NEXPTIME**-hard in the worst-case as shown by Dantsin, Eiter, Gottlob, and Voronkov (1997). Hence, problems in Σ_2^P can be tackled with ALPHA.

3. State of the Art in Domain-Specific Heuristics

State-of-the-art ASP solvers are well suited to solve a wide range of problems, as shown in ASP competitions, experiments, and (industrial) applications reported in the literature (Gebser et al., 2020; Erdem et al., 2016; Falkner et al., 2018; Leone & Ricca, 2015). However, applying general ASP solvers to large instances of industrial problems often requires sophisticated encodings or solver tuning methods, e.g., portfolio solvers like CLASPFOLIO (Hoos et al., 2014) or ME-ASP (Maratea et al., 2015), to achieve satisfactory performance.

Depending on the problem and instances, tuning of search parameters and encodings may not be sufficient to meet runtime requirements. Domain-specific heuristics were needed to achieve breakthroughs in solving industrial configuration problems with ASP. Several approaches have implemented embedding heuristic knowledge into the ASP solving process.

HWASP (Dodaro et al., 2016) extends WASP (Alviano et al., 2019) by facilitating external procedural heuristics that are consulted at specific points during the solving process via an API. As a result, HWASP can find solutions for all published instances of the Partner Units Problem (PUP) by exploiting external heuristics formulated in C++.

A declarative approach to formulating domain-specific heuristics in ASP was suggested by Gebser et al. (2013). The CLINGO system supports `#heuristic` directives described in detail by Gebser et al. (2019). Heuristic directives extend the ASP language to enable declarative specification of weights determining atom and sign orders in the corresponding internal decision heuristics of a solver. An atom’s weight influences the order in which atoms are considered by the solver when making a decision. A sign modifier instructs whether the selected atom must be assigned true or false. Atoms with a higher weight are assigned a value before atoms with a lower weight.

The following (non-ground) meta-statement defines domain-specific heuristics in CLINGO, where ha is an atom, hB is a conjunction of literals representing the heuristic body, and w , p , and m are terms (Gebser et al., 2019).

$$\text{\#heuristic } ha : hB. \quad [w@p, m] \quad \langle 2 \rangle$$

The optional term p gives a preference between heuristic values for the same atom (preferring those with higher p). The term m specifies the type of heuristic information and can take the following values: `sign`, `level`, `true`, `false`, `init` and `factor`. For instance, heuristics for $m=\text{init}$ and $m=\text{factor}$ allow modifying initial and actual atom scores evaluated by the solver’s decision heuristics (e.g., VSIDS). The $m=\text{sign}$ modifier forces the decision heuristics to assign an atom ha a required sign, i.e., **T** or **F**, and $m=\text{level}$ allows for the definition of an order in which the atoms are assigned—the larger the value of w , the earlier an atom must be assigned. Finally, $m=\text{true}$ specifies that ha should be mapped to **T**, i.e., $\text{truth}_A(ha) = \mathbf{T}$, with weight w if hB is satisfied, $m=\text{false}$ is the analogue heuristics that maps ha to **F**.

Gebser et al. (2013) provide a formal specification of a heuristic predicate and the effects of heuristic atoms on the selection of unassigned atoms and their truth assignment in the

search process. Gebser et al. (2019) present syntax and semantics for heuristic directives. However, the implementation of heuristic directives in CLINGO (v. 5.3) might result in a counter-intuitive interpretation of the default negation in heuristic directives since for an assignment A , a heuristic statement h of the form $\langle 2 \rangle$ is applicable and affects solving iff $\text{body}(h) = \{b_1, \dots, b_m, \text{not } b_{m+1}, \dots, \text{not } b_n\}$ is assigned true, i.e., $\text{truth}_A(b_i) = \mathbf{T}$ for $1 \leq i \leq m$ and $\text{truth}_A(b_j) = \mathbf{F}$ for $m+1 \leq j \leq n$. That is, the literal $\text{not } b_i$ is not evaluated to true in the absence of a truth assignment to b_i (as one would expect for default negation). Instead, the literal is evaluated to a truth-value iff b_i is assigned to a truth-value. We illustrate this issue in the following example:

Example 1. Consider the following program containing two heuristic directives:

```

{ a(2) ; a(4) ; a(6) ; a(8) ; a(5) } ← .
← #sum { X : a(X) } = S, S \ 2 ≠ 0.
#heuristic a(5).                               [1, true]
#heuristic a(4) : not a(5).                     [2, true]
    
```

The program guesses a subset of $\{2, 4, 6, 8, 5\}$, the sum of which must be even, i.e., $a(5)$ must not be chosen. The heuristic statements specify that $a(5)$ shall be set to true with weight 1, and that $a(4)$ shall be set to true if $\text{not } a(5)$ is true with weight 2.

In solving the program in the preceding example, CLINGO (v. 5.3) first assigns $a(5)$ to true in our experiments, although $a(4)$ has a higher weight and $a(5)$ is not known to be true in the beginning. Next, $a(8)$ is chosen to be false,² the solver backtracks and only $a(5)$ stays assigned. Finally, $a(8)$ is chosen to be true, and a conflict is learned that makes $a(5)$ false after backtracking. Now that $\text{not } a(5)$ is satisfied, the second heuristic chooses $a(4)$ to be true, and we obtain the answer set $\{a(4), a(8)\}$ after a few more guesses on the yet unassigned atoms. The second heuristic becomes active only later because $\text{not } a(5)$ is evaluated as true only if $a(5)$ is assigned false.

Heuristics evaluated over partial assignments appear quite often in practice. For example, in configuring technical systems, we might prefer to assign, in the current search state, the most relevant yet unplaced electronic component to a free slot of a motherboard. In scheduling, we might prefer the assignment of lots to machines that are not assigned in the current search state and whose delivery deadline is most urgent. CLINGO’s semantics for heuristic directives allows one to model heuristics over partial assignments as in Example 1 using priorities. However, modelling of prioritized directives can be tedious and an automatised translation might require introducing unnecessary auxiliary variables.

To overcome this issue we propose, in the following section, to evaluate negation as failure (i.e., *not*) in heuristic statements *w.r.t. the current partial assignment* in the solver. This partial assignment represents the search state. As a consequence, $\text{not } X$ is true if X is false *or unassigned* during the search. The following example shows the application of default negation to formulate a well-known heuristic for pathfinding.

Example 2. Let Pathfinding be the problem of finding a path from one square to another on a rectangular grid by moving horizontally and vertically and avoiding obstacles. For

2. Choices not determined by the heuristic directives may vary from one implementation to another.

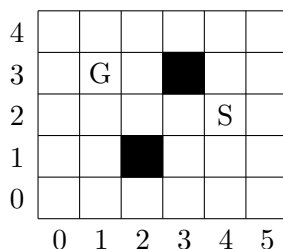


Figure 1: A sample Pathfinding instance

example, in Figure 1, there are two shortest paths from the start (S) to the goal (G), both of length 4.

This instance is specified by the following self-explanatory facts:

```
xmin(0).  ymin(0).  start(4,2).  obstacle(3,3).
xmax(5).  ymax(4).  goal(1,3).  obstacle(2,1).
```

Assume adequate definitions of the following predicates:

- $\text{at}(X, Y, T)$ means the agent's position is (X, Y) at time T ;
- $\text{neighbour}(X1, Y1, X2, Y2)$ signifies that the agent can move from $(X1, Y1)$ to $(X2, Y2)$ and vice versa;
- and $\text{max_time}(\text{MaxT})$ gives the time horizon.

Then, the following program encodes the problem:

```
at(X, Y, 0) ← start(X, Y).
reached_goal(T) ← goal(X, Y), at(X, Y, T).
1 { move(X, Y, T) : neighbour(AtX, AtY, X, Y), not obstacle(X, Y) } 1
  ← at(AtX, AtY, T), not reached_goal(T),
  max_time(MaxT), T < MaxT.
at(X, Y, T + 1) ← move(X, Y, T).
moved(T) ← move(X, Y, T).
visited(X, Y) ← at(X, Y, T).
```

A heuristic preferring to move to squares with the least Manhattan distance to the goal (“greedy best-first search”, cf. Russell & Norvig, 2010; Pearl, 1984) can be encoded by the following heuristic directive, following CLINGO’s syntax given in ⟨2⟩:

```
#heuristic move(X, Y, T) : at(AtX, AtY, T), neighbour(AtX, AtY, X, Y),
  not obstacle(X, Y), goal(GoalX, GoalY), max_time(MaxT).
  [MaxT - (|X - GoalX| + |Y - GoalY|), true]
```

But how to restrict suggestions by this heuristic to move only at time T if the agent has not already moved at time T , or to move only to squares not already visited? We would

expect to achieve such a restriction by adding one or both of the following literals to the heuristic's body:

not moved(T), not visited(X, Y)

However, the heuristic is rendered unusable if either of the two literals is added. The reason is that CLINGO does not interpret not within bodies of heuristics w.r.t. the current partial assignment, and thus the heuristic is not applicable unless all atoms within negative literals are assigned false.

4. A Novel Semantics for Declarative Domain-Specific Heuristics

Declaratively specifying domain-specific heuristics in ASP plays a vital role in enabling ASP to solve large-scale industrial problems. CLINGO has been the only ASP system to support such heuristics so far. Although language and semantics of heuristic directives in CLINGO have shown to be beneficial in many cases, dynamic aspects of negation as failure in heuristic conditions have not been addressed satisfactorily.

We present novel syntax and semantics for heuristic directives in ASP that improve this situation. We assume that the underlying solver can assign one of three values to any atom: *true* (denoted with **T**), *false* (**F**), and *must-be-true* (**M**) (cf. Weinzierl, 2017). The following definitions can be used without modification for solvers that do not use the third truth value **M**. The set of atoms assigned must-be-true will be empty in this case.

Definition 1 (Heuristic Directive). *A heuristic directive is of the form $\langle \mathcal{H} \rangle$, where ha_i ($0 \leq i \leq n$) are heuristic atoms of the form $s_i a_i$, in which $s_0 \in \{\{\mathbf{F}\}, \{\mathbf{T}\}\}$ and $s_i \subseteq \{\mathbf{F}, \mathbf{M}, \mathbf{T}\}$ are sets of sign symbols and a_i is an atom, and w and l are integer terms.*

$$\text{\#heuristic } ha_0 : ha_1, \dots, ha_k, \text{not } ha_{k+1}, \dots, \text{not } ha_n. \quad [w@l] \quad \langle \mathcal{H} \rangle$$

The heuristics' head is given by ha_0 and its condition by $\{ha_1, \dots, ha_k, \text{not } ha_{k+1}, \dots, \text{not } ha_n\}$, which is similar to a rule body.

Where the meaning is clear from the context, we may omit all symbols except sign symbols themselves in a set of sign symbols, e.g., we write **TM** instead of $\{\mathbf{T}, \mathbf{M}\}$.

Similarly to a rule (cf. Section 2.2), a heuristic directive can contain variables and must be *safe*. Since the exact definition of safety of a heuristic directive depends on implementation matters, we will address it in more detail in Section 5.4.

The textual syntax of a heuristic directive to be used in answer-set programs is defined by the following EBNF grammar, where $\langle \textit{classical_atom} \rangle$ stands for a classical atom, and $\langle \textit{term} \rangle$ stands for a term as defined in Section 2.1:

$$\langle \textit{heuristic_directive} \rangle ::= \text{'\#heuristic'} \langle \textit{head_atom} \rangle [\langle \textit{body} \rangle] \text{'.'} [\langle \textit{annotation} \rangle]$$

$$\langle \textit{head_atom} \rangle ::= [\langle \textit{head_sign} \rangle] \langle \textit{classical_atom} \rangle$$

$$\langle \textit{head_sign} \rangle ::= \text{'T'} \mid \text{'F'}$$

$$\langle \textit{body} \rangle ::= \text{'.'} \langle \textit{body_literals_list} \rangle$$

$$\begin{aligned}
 \langle \text{body_literals_list} \rangle & ::= \langle \text{body_literal} \rangle [\text{' , ' } \langle \text{body_literals_list} \rangle] \\
 \langle \text{body_literal} \rangle & ::= [\text{' not ' }] \langle \text{body_atom} \rangle \\
 \langle \text{body_atom} \rangle & ::= ([\langle \text{body_sign_list} \rangle] \langle \text{classical_atom} \rangle) \\
 \langle \text{body_sign_list} \rangle & ::= [\text{' T ' }] [\text{' M ' }] [\text{' F ' }] \\
 \langle \text{annotation} \rangle & ::= [\text{' [' } \langle \text{weight} \rangle [\text{' @ ' } \langle \text{level} \rangle] \text{'] ' } \\
 \langle \text{weight} \rangle & ::= \langle \text{term} \rangle \\
 \langle \text{level} \rangle & ::= \langle \text{term} \rangle
 \end{aligned}$$

Note that the definition of $\langle \text{body_sign_list} \rangle$ imposes a specific order in which heuristic signs have to appear (first **T**, then **M**, and finally **F**). However, we assume that an implementation allows heuristic signs to appear in any order (e.g., **FT**).

Example 3. Consider the following heuristic directive d :

$$\# \text{heuristic } \mathbf{F} \ a : \mathbf{TM} \ b, \mathbf{T} \ c, \text{not } \mathbf{TMF} \ e.$$

This directive means that the atom a shall be assigned **F** if b is assigned **T** or **M**, c is assigned **T**, and e is not assigned.

We now introduce some notation that will be used in further definitions. The function `atom` maps a heuristic atom ha_i of the form $s_i \ a_i$ to a_i by removing the sign, and a set of heuristic atoms to the set of atoms occurring in them (e.g., $\text{atom}(\mathbf{M} \ a) = a$, $\text{atom}(\{\mathbf{M} \ a, \mathbf{T} \ b\}) = \{a, b\}$). The function `signs` maps a heuristic atom ha_i to s_i by removing the atom (e.g., $\text{signs}(\mathbf{M} \ a) = \{\mathbf{M}\}$).

The *head* of a heuristic directive d of the form $\langle \mathcal{B} \rangle$ is denoted by $\text{head}(d) = ha_0$, its *weight* by $\text{weight}(d) = w$ if given, else 0, and its *level* by $\text{level}(d) = l$ if given, else 0. The (*heuristic condition*) of a heuristic directive d is denoted by $\text{cond}(d) := \{ha_1, \dots, ha_k, \text{not } ha_{k+1}, \dots, \text{not } ha_n\}$, the *positive condition* is $\text{cond}^+(d) := \{ha_1, \dots, ha_k\}$ and the *negative condition* is $\text{cond}^-(d) := \{ha_{k+1}, \dots, ha_n\}$.

Let HA be a set of heuristic atoms. Then, for $s \subseteq \{\mathbf{F}, \mathbf{M}, \mathbf{T}\}$, $HA|_s = \{a \mid s \ a \in HA\}$ denotes the set of atoms in HA whose set of sign symbols equals s .

Example 4. Consider the heuristic directive d from Example 3 again:

$$\# \text{heuristic } \mathbf{F} \ a : \mathbf{TM} \ b, \mathbf{T} \ c, \text{not } \mathbf{TMF} \ e.$$

Here, $\text{cond}^+(d)|_{\mathbf{MT}} = \{b\}$; $\text{cond}^+(d)|_{\mathbf{T}} = \{c\}$; and $\text{cond}^+(d)|_{\mathbf{F}} = \emptyset$.

Furthermore, $\text{cond}^-(d)|_{\mathbf{FMT}} = \{d\}$; note that the order of sign symbols does not matter due to set semantics.

Our proposal differs from CLINGO's in the following ways, apart from the syntactic differences between $\langle \mathcal{B} \rangle$ and $\langle \mathcal{S} \rangle$:

- Each heuristic atom contains a set of sign symbols. Each sign symbol represents one of the truth values \mathbf{F} (*false*), \mathbf{T} (*true*), and \mathbf{M} (*must-be-true*).
- In the condition, sign symbols provide a richer way of controlling when the condition is satisfied. A positive literal in the condition is satisfied if the truth value currently assigned to its atom is contained in its set of sign symbols (which is $\{\mathbf{M}, \mathbf{T}\}$ by default if not explicitly given). A negative literal in the condition is satisfied if the truth value currently assigned to its atom is *not* contained in its set of sign symbols or if its atom is currently not assigned any truth value.
- In the heuristic head, the sign symbol is used to determine the truth value to be chosen by the heuristic. If s_0 is \mathbf{T} or empty, the heuristics makes the solver guess a_0 to be true; if s_0 is \mathbf{F} , a_0 will be made false.³ We do not use the modifier m .
- Instead of weight w and tie-breaking priority p , we use terms w and l denoting weight and level as familiar from optimise-statements in ASP-Core-2 (Calimeri et al., 2020) or weak constraints in DLV (Leone et al., 2006). The level is more important than the weight; both default to 0, and together they are called priority.

We now describe our semantics more formally, beginning with the condition under which a heuristic atom is satisfied.

Definition 2 (Satisfying a Heuristic Atom). *Given a ground heuristic atom ha and a partial assignment A , ha is satisfied w.r.t. A iff: $\text{truth}_A(\text{atom}(ha)) \in \text{signs}(ha)$, i.e., if its atom is assigned a truth value that is included in the heuristic atom’s sign set.⁴*

Whether a heuristic directive is satisfied depends on whether the atoms occurring in the directive are satisfied.

Definition 3 (Satisfying a Heuristic Directive). *Given a ground heuristic directive d and a partial assignment A , $\text{cond}(d)$ is satisfied w.r.t. A iff: every $ha \in \text{cond}^+(d)$ is satisfied and no $ha \in \text{cond}^-(d)$ is satisfied.*

Intuitively, a heuristic condition is satisfied iff its positive part is fully satisfied and none of its default-negated literals is contradicted.

Definition 4 (Applicability of a Heuristic Directive – Semantics). *A ground heuristic directive d is applicable w.r.t. a partial assignment A and a ground program P iff: $\text{cond}(d)$ is satisfied, $\exists r \in P$ s.t. $\text{head}(r) = \text{atom}(\text{head}(d))$ and $\{\mathbf{T} a \mid a \in \text{body}^+(r)\} \subseteq A$ and $\{\mathbf{M} a \mid a \in \text{body}^-(r)\} \cap A = \emptyset$, and $\text{truth}_A(\text{atom}(\text{head}(d))) \in \{\mathbf{U}, \mathbf{M}\}$.*

Intuitively, a heuristic directive is applicable iff its condition is satisfied, there exists a currently applicable rule that can derive the atom in the heuristic directive’s head, and the atom in its head is assigned neither \mathbf{T} nor \mathbf{F} . If the atom in the head is assigned \mathbf{M} , the

3. In the head, we only support truth values \mathbf{T} and \mathbf{F} because, from a user’s point of view, it does not make sense to assign \mathbf{M} to an atom heuristically.

4. Note that the function truth maps to only one truth value even though $\mathbf{T} a \in A$ implies $\mathbf{M} a \in A$, so $\text{truth}_A(a) = \mathbf{M}$ iff $\mathbf{M} a \in A$ and $\mathbf{T} a \notin A$.

heuristic directive may still be applicable, because any atom with the non-final truth value \mathbf{M} must be either \mathbf{T} or \mathbf{F} in any answer set.

Definitions 2 to 4 reveal the main difference between the semantics proposed here and the one implemented by CLINGO: In our approach, heuristic signs composed of truth values \mathbf{T} , \mathbf{M} , and \mathbf{F} can be used in heuristic conditions to reason about atoms that are already assigned specific truth values in a partial assignment. Furthermore, default negation can be used to reason about atoms that are assigned *or still unassigned*. Our semantics truly means default negation in the current partial assignment, while the one implemented by CLINGO amounts to strong negation in the current search state and a sophisticated modelling approach with priorities must be used to implement required heuristics. This difference is crucial since reasoning about incomplete information is essential in many cases. An example is a heuristic for a configuration problem that only applies to components not yet placed.

What remains to be defined is the semantics of weight and level. Given a set of applicable heuristic directives, one directive with the highest weight will be chosen from the highest level. Suppose there are several heuristic directives with the same maximum priority (i.e., weight and level). In that case, the solver can use domain-independent heuristics like VSIDS (Moskewicz et al., 2001) as a fallback to break the tie.

Definition 5 (The Subset of Heuristics Eligible for Immediate Choice). *Given a set D of applicable ground heuristic directives, the subset eligible for immediate choice is defined as $\text{maxpriority}(D)$ in two steps:*

$$\begin{aligned} \text{maxlevel}(D) &:= \{d \mid d \in D \text{ and } \text{level}(d) = \max_{d \in D} \text{level}(d)\} \\ \text{maxpriority}(D) &:= \{d \mid d \in \text{maxlevel}(D) \text{ and } \text{weight}(d) = \max_{d \in \text{maxlevel}(D)} \text{weight}(d)\} \end{aligned}$$

After choosing a heuristic using maxpriority , a solver makes a decision on the directive's head atom. Other solving procedures, e.g., deterministic propagation, are unaffected by processing heuristics. In case no heuristic directive is applicable, the solver's default heuristic (e.g., VSIDS) makes a choice as usual.

Example 5. *Consider the program given in Example 1. When converted to the syntax proposed in Definition 1, its heuristic directives look like directives $\langle 4 \rangle$ and $\langle 5 \rangle$ in the following program. Consider also the newly introduced directives $\langle 6 \rangle$ and $\langle 7 \rangle$ in this program.*

```
#heuristic a(5). [1] <4>
#heuristic a(4) : not a(5). [2] <5>
#heuristic F a(5) : a(4). [2] <6>
#heuristic a(6) : F a(5), T a(4). [2] <7>
```

Intuitively, directive $\langle 4 \rangle$ unconditionally prefers to make $a(5)$ \mathbf{T} with weight 1. All other directives have a higher weight, 2, but they become applicable at different times. Directive $\langle 5 \rangle$ prefers to make $a(4)$ \mathbf{T} if $a(5)$ is neither \mathbf{T} nor \mathbf{M} , directive $\langle 6 \rangle$ prefers to make $a(5)$ \mathbf{F} if $a(4)$ is \mathbf{T} or \mathbf{M} , and $\langle 7 \rangle$ prefers to make $a(6)$ \mathbf{T} if $a(5)$ is \mathbf{F} and $a(4)$ is \mathbf{T} .

Let $A_0 = \emptyset$ be the empty partial assignment before any decision has been made. W.r.t. A_0 , $\langle 4 \rangle$ is applicable because its condition is empty and its head is still unassigned. Directive

$\langle 5 \rangle$ is also applicable because $a(5)$ is still unassigned. Directives $\langle 6 \rangle$ and $\langle 7 \rangle$ are not applicable w.r.t. A_0 . Directive $\langle 5 \rangle$ is chosen because it has the highest priority among applicable directives. Thus, $a(4)$ is assigned \mathbf{T} , updating our assignment to $A_1 = \{\mathbf{M} a(4), \mathbf{T} a(4)\}$. This makes $\langle 6 \rangle$ applicable, $a(5)$ is assigned \mathbf{F} and our assignment is $A_2 = \{\mathbf{M} a(4), \mathbf{T} a(4), \mathbf{F} a(5)\}$. Note that the condition of $\langle 5 \rangle$ was still satisfied at this point, but it was not applicable because its head was already assigned. Now, $\langle 4 \rangle$ is also not applicable anymore, and the only directive that remains is $\langle 7 \rangle$. Since $\langle 7 \rangle$ is applicable, $a(6)$ is made \mathbf{T} and added to the assignment. Next, the atoms that remained unassigned are guessed by the default heuristics until an answer set is found.

Example 6. Continuing from Example 2, now using heuristic directives of the form $\langle 3 \rangle$, it becomes possible to restrict the condition of the Pathfinding heuristics as desired.

```
#heuristic move(X, Y, T) : at(AtX, AtY, T), neighbour(AtX, AtY, X, Y),
                           not obstacle(X, Y), goal(GoalX, GoalY), max_time(MaxT),
                           not T visited(X, Y), not T moved(T).
                           [(|X - GoalX| + |Y - GoalY|) * -1]
```

Compared to Example 2, now we can use the additional conditions $\text{not } \mathbf{T} \text{ visited}(X, Y)$ and $\text{not } \mathbf{T} \text{ moved}(T)$ that relate to the current partial assignment. Through these conditions, the heuristic only suggests moves that do not visit squares repeatedly, and it does not suggest to move to several squares at the same time.

Note that the annotation looks different: Since negative weights make sense in our semantics, we can use $(|X - \text{GoalX}| + |Y - \text{GoalY}|) * -1$ instead of $\text{MaxT} - (|X - \text{GoalX}| + |Y - \text{GoalY}|)$. Thus, the higher the Manhattan score of a square, the lower (i.e., nearer to $-\infty$) is the heuristic weight, and the less attractive it will be to move to that square. Furthermore, the modifier “true” has been dropped because we directly encode the heuristic’s polarity in the directive’s head.

Distinguishing between \mathbf{TM} and \mathbf{T} facilitates fine-grained control of when a heuristic is active. \mathbf{T} is used when it is essential that the \mathbf{T} assignment is already justified (or not yet justified, in case default negation is used), and \mathbf{TM} is used when this is not important, as long as the atom must be true in a valid answer set. In Example 6, $\text{not } \mathbf{T} \text{ visited}(X, Y)$ is used to switch off the heuristic when the agent *has moved* to (X, Y) . The condition $\text{not } \mathbf{TM} \text{ visited}(X, Y)$, instead, would cause the heuristic to be switched off also when the solver only knows that the agent *needs to move* to (X, Y) at some point. This situation can be caused by a constraint forcing the agent to move to a specific square (X, Y) at some point, because then the solver would propagate \mathbf{M} to $\text{visited}(X, Y)$. On the other hand, in many cases we want the condition to cover both \mathbf{T} and \mathbf{M} . Section 6 will include many cases where the sign set \mathbf{TM} is used (recall that this is the default sign set used whenever none is given).

5. Integration into a Lazy-Grounding ASP Solver

This section presents how the above domain-specific heuristics can be realised within a lazy-grounding ASP system. We chose ALPHA as the basis for the implementation as it currently

is the most efficient lazy-grounding ASP system.⁵ The presentation, therefore, contains parts specific to ALPHA. Our declarative domain-specific heuristics may also be realised in traditional ground-and-solve systems. Note, however, that traditional ground-and-solve systems, in contrast to lazy-grounding systems, will have to instantiate all heuristic directives. This additional overhead may cause domain-specific heuristics to perform worse if implemented on top of a ground-and-solve system.

5.1 Transforming Sign Sets

In order to simplify the implementation, heuristic atoms with certain sign sets are transformed in a preprocessing step before grounding. Every heuristic directive of the form $\langle \mathcal{P} \rangle$ is equivalent to a set of heuristic directives containing only sign sets **F**, **T**, and **MT**. Note that every heuristic directive is also equivalent to a set of heuristic directives containing only singleton sign sets **F**, **M**, and **T**; however, we are including **MT** instead of **M** to allow for a more efficient implementation. Using **MT** instead of **M** is necessary for re-using ALPHA's nogood propagation capabilities, which will be introduced in Sections 5.3 and 5.4.

Since heuristic directives can be reduced to directives using only sign sets **F**, **T**, **MT**, all other sign sets (**M**, **FM**, **FT**, and **FMT**) can be transformed while retaining the semantics of the whole set of heuristic directives in the program.

We define a helper function `heuristic` that constructs a heuristic directive by replacing the positive and negative condition within a given heuristic directive: $\text{heuristic}(d, c^+, c^-) \mapsto d'$, s.t. $\text{head}(d') = \text{head}(d)$, $\text{cond}^+(d') = c^+$, $\text{cond}^-(d') = c^-$, $\text{weight}(d') = \text{weight}(d)$, and $\text{level}(d') = \text{level}(d)$.

We then define the function `tf` that transforms one heuristic directive into a set of equivalent directives. The function non-deterministically picks and transforms one heuristic atom with a sign set $s \in \{\mathbf{M}, \mathbf{FM}, \mathbf{FT}, \mathbf{FMT}\}$, or returns the unmodified input d if no such atom exists.

$$\text{tf}(d) = \begin{cases} \text{tf1}(d, s, a) & \text{for some } s \ a \in \text{cond}^+(d) \text{ and } s \in \{\mathbf{FM}, \mathbf{FT}, \mathbf{FMT}\} & \langle 8a \rangle \\ \text{tf2}(d, s, a) & \text{for some } s \ a \in \text{cond}^-(d) \text{ and } s \in \{\mathbf{FM}, \mathbf{FT}, \mathbf{FMT}\} & \langle 8b \rangle \\ \text{tf3}(d, a) & \text{for some } \mathbf{M} \ a \in \text{cond}^+(d) & \langle 8c \rangle \\ \text{tf4}(d, a) & \text{for some } \mathbf{M} \ a \in \text{cond}^-(d) & \langle 8d \rangle \\ \{d\} & \text{if } \nexists s \ a \in \text{cond}^+(d) \cup \text{cond}^-(d) \text{ s.t. } s \in \{\mathbf{M}, \mathbf{FM}, \mathbf{FT}, \mathbf{FMT}\}. & \langle 8e \rangle \end{cases}$$

$$\text{tf1}(d, s, a) = \{\text{heuristic}(d, \text{cond}^+(d) \cup \{s \setminus \mathbf{F} \ a\} \setminus \{s \ a\}, \text{cond}^-(d)), \\ \text{heuristic}(d, \text{cond}^+(d) \cup \{\mathbf{F} \ a\} \setminus \{s \ a\}, \text{cond}^-(d))\}$$

$$\text{tf2}(d, s, a) = \{\text{heuristic}(d, \text{cond}^+(d), \text{cond}^-(d) \cup \{s \setminus \mathbf{F} \ a, \mathbf{F} \ a\} \setminus \{s \ a\})\}$$

$$\text{tf3}(d, a) = \{\text{heuristic}(d, \text{cond}^+(d) \cup \{\mathbf{MT} \ a\} \setminus \{\mathbf{M} \ a\}, \text{cond}^-(d) \cup \{\mathbf{T} \ a\})\}$$

$$\text{tf4}(d, a) = \{\text{heuristic}(d, \text{cond}^+(d) \cup \{\mathbf{FT} \ a\}, \text{cond}^-(d) \setminus \{\mathbf{M} \ a\}), \\ \text{heuristic}(d, \text{cond}^+(d), \text{cond}^-(d) \cup \{\mathbf{FMT} \ a\} \setminus \{\mathbf{M} \ a\})\}$$

5. ALPHA sources and binaries can be found on <https://github.com/alpha-asp/Alpha>. Features described in this section have been implemented on the `domspec_heuristics_extended` branch.

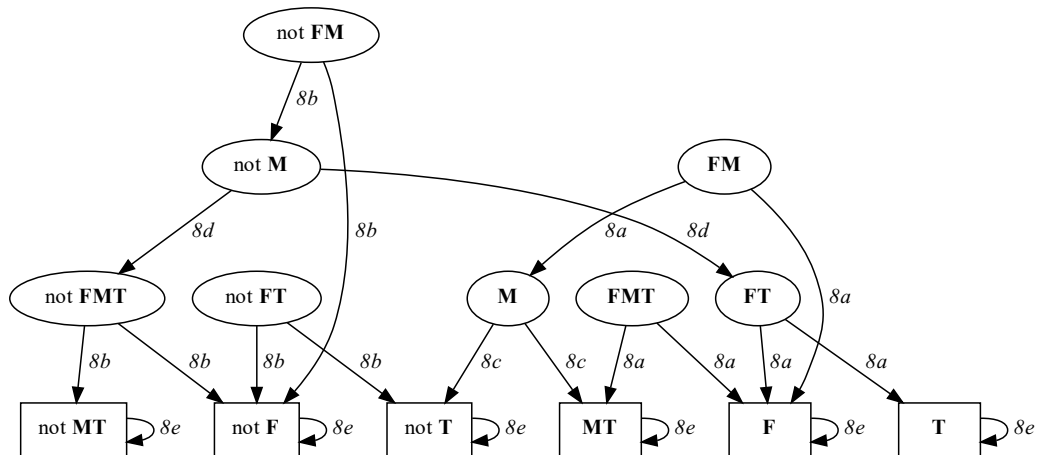


Figure 2: Visualisation of the transformation steps done by `tf` (see also Lemma 1)

Recall that \mathbf{F} is an abbreviation for $\{\mathbf{F}\}$, thus $s \setminus \mathbf{F}$ is the same as $s \setminus \{\mathbf{F}\}$.

Intuitively, the transformations work as follows:

$\langle \delta a \rangle$ transforms one directive into two new directives. In this process, one sign set containing \mathbf{F} is split into two partitions, \mathbf{F} and the remaining signs. The original directive is satisfied iff one of the new directives is satisfied.

$\langle \delta b \rangle$ transforms one directive into one new directive. This time, one heuristic atom in the negative condition whose sign set contains \mathbf{F} is split into two. The original heuristic atom is satisfied iff one of the two new heuristic atoms is satisfied.

$\langle \delta c \rangle$ represents the informal equivalence that “ \mathbf{M} ” is the same as “ \mathbf{M} or \mathbf{T} , but not \mathbf{T} ”, while $\langle \delta d \rangle$ represents that “not \mathbf{M} ” is the same as “ \mathbf{F} or \mathbf{T} or unassigned”.

Finally, due to $\langle \delta e \rangle$, a heuristic directive is unaffected by this transformation if it contains only the sign sets \mathbf{F} , \mathbf{T} , and \mathbf{MT} .

Figure 2 visualises all possible transformation steps done by `tf`. Every possible sign set is represented by two nodes—one for positive and one for negative literals. Sign sets unaffected by the transformation are shown as rectangular nodes, the others as ellipses. Each arc corresponds to a transformation step and is labelled with a symbol corresponding to one of the cases in $\langle \delta \rangle$.

Example 7. Consider the following heuristic directive:

`#heuristic h : FMT a, not FT b.`

Transformation $\langle \delta b \rangle$ replaces it by the following directive, removing \mathbf{FT} in the negative condition:

`#heuristic h : FMT a, not F b, not T b.`

Then, transformation $\langle 8a \rangle$ removes **FMT** in the positive condition and thus produces two new directives:

$$\begin{aligned} \#\text{heuristic } h &: \mathbf{F} a, \text{ not } \mathbf{F} b, \text{ not } \mathbf{T} b. \\ \#\text{heuristic } h &: \mathbf{MT} a, \text{ not } \mathbf{F} b, \text{ not } \mathbf{T} b. \end{aligned}$$

These two directives together are semantically equivalent to the original one.

Example 8. Consider the following heuristic directive containing **M** in its positive and negative condition:

$$\#\text{heuristic } h : \mathbf{M} a, \text{ not } \mathbf{M} b.$$

Transforming **M a** by $\langle 8c \rangle$ in the positive condition results in the following directive:

$$\#\text{heuristic } h : \mathbf{MT} a, \text{ not } \mathbf{T} a, \text{ not } \mathbf{M} b.$$

Transforming **M b** in the negative condition of the new directive, employing transformation $\langle 8d \rangle$, yields two further directives:

$$\begin{aligned} \#\text{heuristic } h &: \mathbf{MT} a, \mathbf{FT} b, \text{ not } \mathbf{T} a. \\ \#\text{heuristic } h &: \mathbf{MT} a, \text{ not } \mathbf{T} a, \text{ not } \mathbf{FMT} b. \end{aligned}$$

Both these conditions need to be transformed further (by $\langle 8a \rangle$ and $\langle 8b \rangle$), thus obtaining three directives in total:

$$\begin{aligned} \#\text{heuristic } h &: \mathbf{MT} a, \mathbf{F} b, \text{ not } \mathbf{T} a. \\ \#\text{heuristic } h &: \mathbf{MT} a, \mathbf{T} b, \text{ not } \mathbf{T} a. \\ \#\text{heuristic } h &: \mathbf{MT} a, \text{ not } \mathbf{T} a, \text{ not } \mathbf{F} b, \text{ not } \mathbf{MT} b. \end{aligned}$$

These three directives together are semantically equivalent to the original one.

Let D be a set of heuristic directives, and let $D^1 = \bigcup_{d \in D} \text{tf}(d)$ denote the result of replacing each heuristic directive by the result of heuristic directive transformation. Furthermore, let $D^{n+1} = \bigcup_{d \in D^n} \text{tf}(d)$ and $D^* = \lim_{n \rightarrow \infty} D^n$ denote the result of the repeated application of tf until a fixpoint is reached.

Lemma 1. For any set of heuristic directives D , D^* contains only sign sets **F**, **T**, and **MT**.

One can show that, no matter with which sign set and polarity the transformation starts, it always ends up with one of the sign sets **F**, **T**, and **MT**. This is obvious from the visualisation of the transformation steps in Figure 2.

Lemma 2. For any heuristic directive d , it holds that d is applicable w.r.t. a partial assignment A iff $\text{tf}(d)$ contains a heuristic directive d' that is applicable w.r.t. A .

A proof sketch for Lemma 2 can be found in Appendix A.

5.2 Enhancing Heuristic Directives

Recall from Definition 4 that for a heuristic directive to be applicable it is necessary that there exists an applicable rule that derives the atom in the heuristic directive’s head, which enforces that atoms chosen by the heuristics have support. This condition can be implemented by an additional transformation, also applied before grounding, that works as follows:

The condition of every heuristic directive is enhanced by adding the body of the corresponding rule. If there are several rules deriving the same head, one copy of the directive is created for each of them. To ensure consistency of variable names, directive and rule are first standardised apart (so that they do not share any variable names) and the unifier of the two heads is applied to both the directive and the rule.

Since a rule is applicable if every positive body literal is assigned \mathbf{T} and no atom in the negative body is assigned \mathbf{M} or \mathbf{T} (cf. Section 2.4), sign set \mathbf{T} is used when copying a positive literal and sign set \mathbf{MT} is used when copying a negative literal. Duplicate literals are not added to the heuristic condition.

Example 9. *Consider the following program:*

$$\begin{aligned} h(X) &\leftarrow a(X), b(X), \text{not } c(X). \\ h(X) &\leftarrow a(X), b(X), \text{not } d(X). \\ \#heuristic\ h(N) &: \mathbf{T}\ a(N). \end{aligned}$$

The transformation described above results in the following modified program:

$$\begin{aligned} h(X) &\leftarrow a(X), b(X), \text{not } c(X). \\ h(X) &\leftarrow a(X), b(X), \text{not } d(X). \\ \#heuristic\ h(N) &: \mathbf{T}\ a(N), \mathbf{T}\ b(N), \text{not } \mathbf{MT}\ c(N). \\ \#heuristic\ h(N) &: \mathbf{T}\ a(N), \mathbf{T}\ b(N), \text{not } \mathbf{MT}\ d(N). \end{aligned}$$

In this example, $h(N)$ (the head of the heuristic directive) has been unified with $h(X)$ (the head of the deriving rules). The resulting unifier $\{X \mapsto N\}$ has been applied to the rules before adding sign sets to their body literals and copying the resulting heuristic literals to the heuristic conditions. Duplicate literals (in this case, a duplicate of $\mathbf{T}\ a(N)$) have been omitted during this process.

5.3 Lazy Grounding in the Alpha System

To understand how domain-specific heuristics are integrated into the ALPHA system, we recapitulate some of its inner workings first. The ALPHA system at its core contains a search loop similar to conflict-driven clause-learning (CDCL) employed in most SAT solvers (Silva et al., 2009; Zhang et al., 2001) and modern ASP solvers like CLINGO (Gebser et al., 2012) or DLV2 (Alviano et al., 2017). In SAT, the encoding of the problem usually is given by clauses (a clause is a set of ground literals where at least one must be satisfied). ASP uses nogoods instead (a nogood is a set of ground literals where all together must never be satisfied). Note that nogoods are dual to clauses, and they can represent precisely the same information.

In CDCL solvers, the primary mode of propagation is unit propagation which works as follows: Given a partial assignment A and a nogood $g = \{l_1, \dots, l_n\}$ where each l_i , $1 \leq i \leq n$ is a signed literal $\mathbf{T} a$ or $\mathbf{F} a$ with a being a ground atom, g is *unit* w.r.t. A if there exists $1 \leq i \leq n$ such that $l_i = s b$ with $\text{truth}_A(b) = \mathbf{U}$ and $g \setminus \{l_i\} \subseteq A$. Intuitively, a nogood g is unit w.r.t. A if it is already violated except for one remaining literal l_i . To avoid violation of the nogood g , the atom b of the remaining literal $l_i = s b$ must be assigned to the opposite truth value of l_i in the end. So, whenever g is unit, one can extend A with the negation of l_i .

In addition to unit propagation, many ASP solvers also employ some form of unfounded-set propagation and source pointers, which ensure that the assignment constructed by the solver is also well-founded (that is, free of self-founding positive cycles). ALPHA does not use such propagation because ALPHA's computation sequence, like in other lazy-grounding ASP solvers, already guarantees well-foundedness. Specifically, the solver can only guess on an applicable rule whether it fires or not (cf. Section 2.4), and not guess the truth value of an arbitrary unassigned atom. So, in ALPHA, an atom b is only assigned \mathbf{T} if there is a well-founded rule firing whose head atom is b .

Moreover, the ALPHA system uses a third truth value \mathbf{M} , called *must-be-true*, which ALPHA can assign to atoms considered true but not yet justified (cf. Bogaerts & Weinzierl, 2018). This value allows further propagation in many circumstances and thus shrinks the search space. For example, the constraint $\leftarrow \text{not } d$. states that any answer set must assign true to the atom d , but it gives no justification why d should hold, i.e., d must-be-true. The solver needs to distinguish between a fact and a constraint because a fact makes d true and justifies it, while a constraint requires d to be true but does not justify it. To achieve this distinction, the notion of a nogood is slightly enhanced in ALPHA: One literal in a nogood may be indicated as the nogood's head.

Given an assignment A and a nogood $g = \{l_1, \dots, l_n\}$, we say that g is

- **F-unit** on l_i w.r.t. A if $1 \leq i \leq n$, $l_i = \mathbf{T}b$, $\text{truth}_A(b) = \mathbf{U}$, and for all $l_j \in g$ with $j \neq i$ holds $\mathbf{F}c \in A$ if $l_j = \mathbf{F}c$ and $\mathbf{M}c \in A$ if $l_j = \mathbf{T}c$.
- **M-unit** on l_i w.r.t. A if $1 \leq i \leq n$, $l_i = \mathbf{F}b$, $\text{truth}_A(b) = \mathbf{U}$, and for all $l_j \in g$ with $j \neq i$ holds $\mathbf{F}c \in A$ if $l_j = \mathbf{F}c$ and $\mathbf{M}c \in A$ if $l_j = \mathbf{T}c$.
- **T-unit** on l_i w.r.t. A if $1 \leq i \leq n$, $l_i = \mathbf{F}b$ is the indicated head of g , $\text{truth}_A(b) = \mathbf{U}$ or $\text{truth}_A(b) = \mathbf{M}$ and for all $l_j \in g$ with $j \neq i$ holds $\mathbf{F}c \in A$ if $l_j = \mathbf{F}c$ and $\mathbf{T}c \in A$ if $l_j = \mathbf{T}c$.

Given an assignment A and a nogood $g = \{l_1, \dots, l_n\}$, we denote by $\text{prop}(g, A)$ the truth value assigned by unit propagation, that is:

$$\text{prop}(g, A) = \begin{cases} \{\mathbf{F}b\} & \text{if } g \text{ is } \mathbf{F}\text{-unit on } l_i \text{ w.r.t. } A \text{ and } l_i = \mathbf{T}b, \\ \{\mathbf{M}b\} & \text{if } g \text{ is } \mathbf{M}\text{-unit but not } \mathbf{T}\text{-unit on } l_i \text{ w.r.t. } A \text{ and } l_i = \mathbf{F}b, \\ \{\mathbf{M}b, \mathbf{T}b\} & \text{if } g \text{ is } \mathbf{T}\text{-unit on } l_i \text{ w.r.t. } A \text{ and } l_i = \mathbf{F}b, \\ \{\} & \text{otherwise.} \end{cases}$$

Finally, given a set of nogoods Δ and an assignment A , we denote by $A^\Delta = A \cup \bigcup_{g \in \Delta} \text{prop}(g, A)$ the assignment extended with all truth values that follow by unit propagation.

Example 10. Consider the constraint $\leftarrow \text{not } d$, which is translated into the nogood $g_1 = \{\mathbf{F} \, d\}$, and compare it with the fact d , which is translated into the nogood $g_2 = \{\mathbf{F} \, \underline{d}\}$. Here, g_2 is a nogood where the first and only literal is indicated as the head, so if g_2 is unit (which it always is unless d is already assigned), then g_2 will propagate the atom d to truth value \mathbf{T} . On the other hand, g_1 does not have a head, so if g_1 is unit, d is propagated only to \mathbf{M} as g_1 does not justify that d is true but merely requires d to be derived by some rule.

While a constraint can be represented as one nogood, rules are represented using multiple nogoods. Given a non-ground rule r and a grounding substitution σ for r (i.e., a substitution that covers all variables in $\text{vars}(r)$), let

$$r\sigma = h\sigma \leftarrow b_1\sigma, \dots, b_m\sigma, \text{not } b_{m+1}\sigma, \dots, \text{not } b_n\sigma.$$

be the rule where σ is applied to every atom. Similarly to CLINGO, the body of every ground rule in ALPHA is represented with a fresh atom, denoted $\beta(r, \sigma)$. The set of nogoods representing $r\sigma$ then is:

$$\{\mathbf{F} \, \beta(r, \sigma), \mathbf{T} \, b_1\sigma, \dots, \mathbf{T} \, b_m\sigma, \mathbf{F} \, b_{m+1}\sigma, \dots, \mathbf{F} \, b_n\sigma\} \quad \langle 9 \rangle$$

$$\{\mathbf{F} \, h\sigma, \mathbf{T} \, \beta(r, \sigma)\} \quad \langle 10 \rangle$$

$$\{\mathbf{T} \, \beta(r, \sigma), \mathbf{F} \, b_1\sigma\}, \dots, \{\mathbf{T} \, \beta(r, \sigma), \mathbf{F} \, b_m\sigma\} \quad \langle 11 \rangle$$

$$\{\mathbf{T} \, \beta(r, \sigma), \mathbf{T} \, b_{m+1}\sigma\}, \dots, \{\mathbf{T} \, \beta(r, \sigma), \mathbf{T} \, b_n\sigma\} \quad \langle 12 \rangle$$

Notice that $\langle 9 \rangle$ ensures that the atom $\beta(r, \sigma)$ becomes true by unit propagation if the full body of $r\sigma$ is satisfied. If $\beta(r, \sigma)$ holds, then the rule's head also becomes true due to unit propagation on $\langle 10 \rangle$. Finally, the nogoods of $\langle 11 \rangle$ and $\langle 12 \rangle$ ensure that the atom representing the body of the rule, i.e., $\beta(r, \sigma)$ is false whenever one literal in the body is not satisfied.

According to the notion of a computation sequence, any rule whose negative body is non-empty offers a choice point when becoming applicable. To aid in detecting this applicability condition, such rules also yield the following nogoods:

$$\{\mathbf{F} \, \underline{\text{ChoiceOn}}(r, \sigma), \mathbf{T} \, b_1, \dots, \mathbf{T} \, b_m\} \quad \langle 13 \rangle$$

$$\{\mathbf{F} \, \underline{\text{ChoiceOff}}(r, \sigma), \mathbf{T} \, b_{m+1}\}, \dots, \{\mathbf{F} \, \underline{\text{ChoiceOff}}(r, \sigma), \mathbf{T} \, b_n\} \quad \langle 14 \rangle$$

Intuitively, the nogood $\langle 13 \rangle$ makes $\text{ChoiceOn}(r, \sigma)$ true whenever the positive body of $r\sigma$ is satisfied, and $\text{ChoiceOff}(r, \sigma)$ is true whenever the current assignment contradicts the negative body of $r\sigma$ due to nogoods of $\langle 14 \rangle$. Whenever $\text{ChoiceOn}(r, \sigma)$ is true, and $\text{ChoiceOff}(r, \sigma)$ is not true, the ground rule $r\sigma$ is applicable. Using these nogoods, ALPHA can detect which ground rules are applicable, i.e., which choice points are allowed by the computation sequence.

For a more detailed description of nogoods and propagation using nogoods with heads in ALPHA, we refer to Weinzierl (2017) and Leutgeb and Weinzierl (2017).

5.4 Generating Nogoods for Heuristic Directives

Similarly to normal rules, heuristic directives are represented using nogoods. To obtain ground instances of heuristic directives, each heuristic directive is transformed into a specifically marked rule such that the lazy-grounding procedures of ALPHA generate all relevant

ground instances of the marked rule. Since the rule is marked specifically, each ground instance is not translated into nogoods like an ordinary rule but translated as follows.

To evaluate sign symbols correctly, the following fact is exploited: ALPHA uses unit propagation to assign a truth value to the head of the nogood—the head is always a negative literal—when all other literals in the nogood are already satisfied. The head atom will be assigned \mathbf{T} if all positive literals in the nogood are assigned \mathbf{T} , and \mathbf{M} if some positive literals are assigned \mathbf{M} (cf. Section 5.3 for details).

Six solver-internal atoms are created for each ground heuristic directive. Let d be a ground heuristic directive. Then, the following atoms are created to be used in nogoods as described below: $\text{HeuOnT}(d)$, $\text{HeuOnMT}(d)$, $\text{HeuOnF}(d)$, $\text{HeuOffT}(d)$, $\text{HeuOffMT}(d)$, and $\text{HeuOffF}(d)$. These atoms are used similarly to the atoms $\text{ChoiceOn}(r, \sigma)$ and $\text{ChoiceOff}(r, \sigma)$, which are used to detect when an ordinary rule is applicable. Here, they serve the purpose of detecting when a heuristic directive becomes active.

The following nogoods are generated:⁶

For ...	this nogood is generated.	
$\{a_1, \dots, a_n\} = \text{atom}(\text{cond}^+(d) _{\mathbf{T}})$	$\{\mathbf{F} \text{HeuOnT}(d), \mathbf{T} a_1, \dots, \mathbf{T} a_n\}$	⟨15⟩
$\{a_{n+1}, \dots, a_m\} = \text{atom}(\text{cond}^+(d) _{\mathbf{MT}})$	$\{\mathbf{F} \text{HeuOnMT}(d), \mathbf{T} a_{n+1}, \dots, \mathbf{T} a_m\}$	⟨16⟩
$\{a_{l+1}, \dots, a_k\} = \text{atom}(\text{cond}^+(d) _{\mathbf{F}})$	$\{\mathbf{F} \text{HeuOnF}(d), \mathbf{F} a_{l+1}, \dots, \mathbf{F} a_k\}$	⟨17⟩

And for every ...	this nogood is generated.	
$\mathbf{T} a \in \text{cond}^-(d)$	$\{\mathbf{F} \text{HeuOffT}(d), \mathbf{T} a\}$	⟨18⟩

$\mathbf{MT} a \in \text{cond}^-(d)$	$\{\mathbf{F} \text{HeuOffMT}(d), \mathbf{T} a\}$	⟨19⟩
--------------------------------------	---	------

$\mathbf{F} a \in \text{cond}^-(d)$	$\{\mathbf{F} \text{HeuOffF}(d), \mathbf{F} a\}$	⟨20⟩
-------------------------------------	--	------

Propagation treats these nogoods just as any other nogood. For example, nogood ⟨15⟩ has the effect that $\text{HeuOnT}(d)$ will propagate to \mathbf{T} if all $\{a_1, \dots, a_n\}$ are assigned \mathbf{T} because a heuristic is potentially switched on if all atoms in positive literals with sign set \mathbf{T} are true.

Likewise, nogood ⟨18⟩ has the effect that $\text{HeuOffT}(d)$ will propagate to \mathbf{T} if a is assigned \mathbf{T} because a heuristic is switched off if an atom occurring negatively with \mathbf{T} is true.

To prove the correctness of the representation of heuristic directives using nogoods, we introduce the notion of a *nogood representation* of a program P , which is denoted by Δ_P and defined as the set of all nogoods of the form ⟨9⟩ to ⟨20⟩ that can be obtained by grounding all rules and heuristic directives of P . For readability of the argument, we assume that Δ_P is fully grounded while in a lazy-grounding system, such as ALPHA, it is grounded lazily.

In the following, we also assume that any assignment A is compatible with the notion of a computation sequence, i.e., there exists a computation sequence that yields A . Note that any assignment that is not compatible cannot be an answer set (cf. Liu et al., 2007).

Lemma 3. *Given a ground heuristic directive d , an assignment A , and an atom $h \in \{\text{HeuOnT}(d), \text{HeuOnMT}(d), \text{HeuOnF}(d), \text{HeuOffT}(d), \text{HeuOffMT}(d), \text{HeuOffF}(d)\}$. Then, $\text{truth}_A(h) = \mathbf{T}$, respectively $\text{truth}_A(h) = \mathbf{M}$, iff there is one nogood $ng \in \Delta_P$ of the*

6. Recall from Section 5.1 that we can assume heuristic directives to contain only sign sets \mathbf{F} , \mathbf{T} , \mathbf{MT} .

form $\langle 15 \rangle$ - $\langle 20 \rangle$ and ng is \mathbf{T} -unit, respectively \mathbf{M} -unit, on \mathbf{Fh} w.r.t. $A' = A \setminus \{\mathbf{Th}, \mathbf{Mh}\}$, i.e., $\text{prop}(ng, A') = \{\mathbf{Th}, \mathbf{Mh}\}$, respectively $\text{prop}(ng, A') = \{\mathbf{Mh}\}$.

Lemma 3 is proven in Appendix A.

To be able to re-use existing grounding strategies (cf. Taupe, Weinzierl, & Friedrich, 2019b), we require every heuristic directive to be *safe*.

Definition 6 (Safety of a Heuristic Directive). *A heuristic directive of the form $\langle \beta \rangle$ is safe if every variable occurring in it also occurs in heuristic atoms inside its positive condition whose set of sign symbols is either \mathbf{T} or \mathbf{MT} . More formally, a heuristic directive d is safe if $\text{vars}(d) \subseteq \text{vars}(\text{cond}^+(d)|_{\mathbf{T}} \cup \text{cond}^+(d)|_{\mathbf{MT}})$.*

Since every variable must appear in a positive body literal that does not use sign \mathbf{F} , the safety of a heuristic directive is more restrictive than the safety of a rule (cf. Leone et al., 2006). The reason for this restriction is that ALPHA grounds heuristic directives using the same techniques as when grounding normal rules. A rule must be grounded no later than when all atoms in its positive body are assigned \mathbf{T} or \mathbf{M} because then the rule is potentially applicable.⁷ Therefore, grounding of a rule is triggered when one of its positive body literals becomes satisfied. Since ALPHA does not trigger grounding when an atom is assigned false, only literals without \mathbf{F} can be used to bind variables. Systems that do not impose such a restriction on when a rule is grounded may define heuristic directives' safety more liberally. A heuristic directive is grounded, at the latest, when all atoms in its positive body whose sign set is \mathbf{T} or \mathbf{MT} are assigned \mathbf{T} or \mathbf{M} . This is necessary for information on heuristic conditions to become known to the solver in time.

5.5 Evaluating Heuristics During Solving

To process heuristics, the solver needs to recognise when a heuristic d is applicable in the current partial assignment A . To do this, the solver inspects the truth values of the HeuOn and HeuOff atoms that are propagated using the nogoods introduced in Section 5.4.

Definition 7 (Satisfying a Heuristic Directive in Solving). *A ground heuristic directive d is solving-satisfied in a partial assignment A iff all of the following conditions hold:*

$$\begin{array}{ll} \text{truth}_A(\text{HeuOnT}(d)) = \mathbf{T}, & \text{truth}_A(\text{HeuOffT}(d)) \neq \mathbf{T}, \\ \text{truth}_A(\text{HeuOnMT}(d)) \in \{\mathbf{M}, \mathbf{T}\}, & \text{truth}_A(\text{HeuOffMT}(d)) \notin \{\mathbf{M}, \mathbf{T}\}, \\ \text{truth}_A(\text{HeuOnF}(d)) = \mathbf{T}, & \text{truth}_A(\text{HeuOffF}(d)) \neq \mathbf{T}. \end{array}$$

Intuitively, a heuristic directive is satisfied if it is enabled by all HeuOn atoms and not disabled by any HeuOff atom.

For the following correctness statements, we need the notion of an assignment that is deductively closed and consistent with a given program: Let P be a ground program, an assignment A is *deductively consistent* with P if for every rule $r \in P$ such that the body of r is satisfied w.r.t. A it holds that the head of r is satisfied w.r.t. A . Note that this also implies that A does not violate any constraint of P .

⁷ A rule may also be grounded earlier (cf. Taupe et al., 2019b).

Lemma 4. *Given a ground program P and its nogood representation Δ_P , let d be a ground heuristic directive of P and A be an assignment deductively consistent with P . Then, $\text{cond}(d)$ is satisfied w.r.t. A iff d is solving-satisfied w.r.t. A^{Δ_P} , the assignment extended with all truth values that follow by unit propagation from Δ_P .*

Lemma 4 is proven in Appendix A.

Definition 8 (Applicability of a Heuristic Directive in Solving). *A ground heuristic directive d is solving-applicable w.r.t. an assignment A iff d is solving-satisfied and $\text{truth}_A(\text{atom}(\text{head}(d))) \in \{\mathbf{U}, \mathbf{M}\}$.*

Intuitively, a heuristic directive is applicable if it is satisfied and the atom in its head is not yet assigned \mathbf{T} or \mathbf{F} . Note the similarity to Definitions 3 and 4, where the semantics of a heuristic directive are defined in terms of a directive being satisfied and applicable. The existence of a currently applicable rule that derives the atom in the heuristic directive’s head is not required here because this condition is already enforced by the transformation described in Section 5.2.

Theorem 1. *Given a ground program P and its nogood representation Δ_P . Let d be a ground heuristic directive of P and A be an assignment deductively consistent with P . Then, d is applicable w.r.t. A and P iff d is solving-applicable w.r.t. A^{Δ_P} .*

Proof. “ \Rightarrow ”: Let d be applicable w.r.t. A , i.e., (i) $\text{cond}(d)$ is satisfied, (ii) $\exists r \in P$ s.t. $\text{head}(r) = \text{atom}(\text{head}(d))$ and $\{\mathbf{T} a \mid a \in \text{body}^+(r)\} \subseteq A$ and $\{\mathbf{M} a \mid a \in \text{body}^-(r)\} \cap A = \emptyset$, and (iii) $\text{truth}_A(\text{atom}(\text{head}(d))) \in \{\mathbf{U}, \mathbf{M}\}$. From (i) and (ii) together with Lemma 4 we directly conclude that d is solving-satisfied. Since $\text{truth}_A(\text{atom}(\text{head}(d))) \in \{\mathbf{U}, \mathbf{M}\}$ holds due to (iii), it follows from Definition 8 that d is solving-applicable w.r.t. A^{Δ_P} .

“ \Leftarrow ”: Let d be solving-applicable w.r.t. A^{Δ_P} , i.e., d is solving-satisfied and $\text{truth}_A(\text{atom}(\text{head}(d))) \in \{\mathbf{U}, \mathbf{M}\}$. Being solving-satisfied implies by Lemma 4 that $\text{cond}(d)$ is satisfied w.r.t. A and P . Since the transformation described in Section 5.2 ensured that $\exists r \in P$ s.t. $\text{head}(r) = \text{atom}(\text{head}(d))$ and $\{\mathbf{T} a \mid a \in \text{body}^+(r)\} \subseteq A$ and $\{\mathbf{M} a \mid a \in \text{body}^-(r)\} \cap A = \emptyset$ all hold, it follows from Definition 4 that d is *applicable* w.r.t. A and P . \square

Recall that the solver determines information on weight, level, and head of a heuristic directive d from $\text{weight}(d)$, $\text{level}(d)$, and $\text{head}(d)$, respectively.

Finding the applicable heuristic directives with the highest priority is aided by efficient data structures like a heap. When an applicable heuristic directive exists, the one with the highest priority is *fired*, i.e., the solver uses it to make a choice. However, the atom in the head of the heuristic directive cannot be directly chosen since ALPHA cannot choose arbitrary atoms but only guess applicable rules due to its reliance on computation sequences. Therefore, an applicable rule that derives the atom in the head of the heuristic directive must be identified. This rule is then guessed to fire or set to not fire, depending on the heuristic sign.

Notice that there may be more than one applicable rule deriving the head of the heuristic directive that was fired, so it may be ambiguous which rule should be picked. Instead of picking a rule arbitrarily, we decided for our implementation to terminate with an error to

disallow such occurrences where a firing heuristic directive finds multiple applicable rules deriving the same head. Other solutions are also possible, including ones where, through some rewriting, always at most one applicable rule is available. However, the occurrence of this issue likely hints at some underlying problems with the heuristics; therefore, we think it may be best to warn the user by raising an error dynamically.

Example 11. *Consider the following program P :*

```
x(1..2).
{a(X) : x(X)}.
b(X) ← x(X), not c(X).
c(X) ← x(X), not b(X).
#heuristic b(X) : x(X), not a(X). [X@2]
```

Since the positive body of every rule in P is satisfied, the full grounding of P is immediately produced. Under the initial partial assignment consisting just of facts $A_0 = \{\mathbf{M} \ x(1), \mathbf{T} \ x(1), \mathbf{M} \ x(2), \mathbf{T} \ x(2)\}$, both ground heuristic directives are applicable since both their positive bodies are satisfied and neither $a(1)$ nor $a(2)$ is assigned yet. The directive in which X has been substituted by 2 has the higher weight, however. For this reason, it is chosen, and the solver finds the only rule that can make the heuristic directive's head $b(2)$ true: $b(2) \leftarrow x(2), \text{not } c(2)$. The choice point representing the body of this rule is assigned true and, after some propagation, the new partial assignment will contain $\mathbf{T} \ b(2)$ (amongst other consequences of propagation).

6. Applications and Experimental Results

We tested our approach to declarative domain-specific heuristics by creating such heuristics for several example domains and solving these problem domains using our implementation in the ALPHA system. Two concrete domains under investigation were the House Reconfiguration Problem (HRP) and the Partner Units Problem (PUP), and state-space search with A^* . These two configuration problems are abstracted variants of typical configuration problems experienced in more than 25 years of applying AI technology in the area of automated configuration of electronic systems (Falkner et al., 2016). For all these applications, heuristics exist which allow the efficient generation of satisfying solutions. Furthermore, we realised general state-space search in our approach, using the well-known informed search method A^* , by which we abstract from concrete domains. A^* allows the incorporation of a heuristic evaluation function. By integrating A^* into ASP, we can show that, on the one hand, we can exploit the power of informed (heuristic) search methods within ASP. On the other hand, we can employ the knowledge representation capabilities of ASP to specify states, actions, and their successor states declaratively. Depending on the quality of the heuristic function an optimal solution is generated and the number of generated nodes is reduced. We applied A^* to two specific search problems. All these applications are presented in Sections 6.2 to 6.4 below, along with heuristics in the language proposed in this work.

To put ASP systems under stress, we used problem encodings and instances of varying sizes, where the larger instances were challenging to ground and solve. More precisely,

traditional grounders excessively consumed space or time when grounding these instances, and solving also was infeasible without the aid of domain-specific heuristics. Additionally, for PUP, instances from the ASP competitions have been used, which are not challenging to ground but hard to solve.

6.1 Experimental Setup

Encodings (including heuristics) and instances, and the ALPHA binaries used for our experiments, are available in the online appendix accompanying our article and on our website.⁸ Details on the sources of the encodings are mentioned in the sections describing the domains. Optimisation statements were not used since ALPHA does not support them yet. However, heuristic directives can be written in a way that optimal or near-optimal solutions are preferably found.⁹

Problem instances were selected by first defining an instance-generating algorithm and then exploring instance sizes to find a set in which all systems could solve some instances under consideration within a time limit of 15 minutes, and some instances could be solved by none (or very few) of these systems. For PUP, competition instances were used additionally.

Taupe et al. (2019b) have introduced *degrees of laziness in grounding*, which vary the conditions under which information about a ground rule is communicated from the grounder to the solver. Traditionally in lazy grounding, rules are grounded when their positive body is fully satisfied. In the concept of degrees of laziness, this is called *strict* grounding. *Permissive* grounding, on the other hand, enables rules to be grounded if their positive body is not fully satisfied, as long as all variables can be bound by positive body literals that are already satisfied.

For example, consider the following non-ground constraint:

$$\leftarrow a(X, Y), b(X).$$

Under the partial assignment $A = \{\mathbf{M} a(1, 2), \mathbf{T} a(1, 2)\}$, the ground constraint

$$\leftarrow a(1, 2), b(1).$$

will only be produced if permissive grounding of constraints is enabled.

Due to experimental results (which did not consider domain-specific heuristics), Taupe et al. (2019b) suggested grounding constraints permissively and other rules strictly by default.¹⁰ In our experiments, we investigated the performance of ALPHA with both strict and permissive grounding of constraints.

Furthermore, ALPHA was used without justification analysis (Bogaerts & Weinzierl, 2018) and without support for negative integers in aggregates because we observed these features to deteriorate performance in some cases. Apart from that, ALPHA was used in its default configuration. The JVM running ALPHA was called with command-line parameters `-Xms1G -Xmx32G`, thus initially allocating 1 GiB for Java’s heap and setting the maximum heap size to 32 GiB.

8. <https://ainf.aau.at/dynacon>

9. However, for some problems, the optimum is unknown.

10. In Section 6.5 we address the question of whether this suggestion is also helpful in the presence of domain-specific heuristics.

For comparison, CLINGO¹¹ (Gebser et al., 2019) was used in version 5.4.0 and DLV2¹² (Alviano et al., 2017) in version 2.1.0.

Each of the machines used to run the experiments was equipped with two Intel[®] Xeon[®] E5-2650 v4 @ 2.20GHz CPUs with 12 cores. Furthermore, each machine had 251 GiB of memory and ran Ubuntu 16.04.1 LTS Linux. Scheduling of benchmarks was done with HTCCondor[™] together with the ABC Benchmarking System (Redl, 2016).¹³ Time and memory consumption were measured by PYRUNLIM,¹⁴ which was also used to limit time consumption to 15 minutes per instance, memory to 40 GiB and swapping to 0. Care was taken to avoid side effects between CPUs, e.g., by requesting exclusive access to an entire machine for each benchmark from HTCCondor.

All solvers were configured to search for the first answer set of each problem instance. Finding one or only a few solutions is often sufficient in industrial use cases since solving large instances can be challenging (Falkner et al., 2016). Therefore, the domain-specific heuristics used in the experiments are designed to help the solver find one answer set that is “good enough”, even though it may not be optimal.

6.2 Case Study 1: The House Reconfiguration Problem (HRP)

The House Reconfiguration Problem (HRP) (Friedrich et al., 2011) is an abstracted version of industrial (re)configuration problems, e.g., rack configuration.

6.2.1 PROBLEM DEFINITION

Formally, HRP is defined as a modification of the House Configuration Problem (HCP).

Definition 9 (HCP). *The input for the House Configuration Problem (HCP) is given by four sets of constants P , T , C , and R representing persons, things, cabinets, and rooms, respectively, and an ownership relation $PT \subseteq P \times T$ between persons and things.*

The task is to find an assignment of things to cabinets $TC \subseteq T \times C$ and cabinets to rooms $CR \subseteq C \times R$, such that: (1) each thing is stored in a cabinet; (2) a cabinet contains at most five things; (3) every cabinet is placed in a room; (4) a room contains at most four cabinets; and (5) a room may only contain cabinets storing things of one person.

Definition 10 (HRP). *The input for the House Reconfiguration Problem (HRP) is given by an HCP instance $H = \langle P, T, C, R, PT \rangle$, a legacy configuration $\langle TC', CR' \rangle$, and a set of things $T' \subseteq T$ that are defined as “long” (all other things are “short”).*

The task is then to find an assignment of things to cabinets $TC \subseteq T \times C$ and cabinets to rooms $CR \subseteq C \times R$, that satisfies all requirements of HCP as well as the following ones: (1) a cabinet is either small or high; (2) a long thing can only be put into a high cabinet; (3) a small cabinet occupies 1 and a high cabinet 2 of 4 slots available in a room; (4) all legacy cabinets are small.

11. <https://potassco.org/clingo/>

12. <https://dlv.demacs.unical.it/>

13. <http://research.cs.wisc.edu/htcondor>, <https://github.com/credl/abcbenchmarking>

14. <https://alviano.com/software/pyrunlim/>

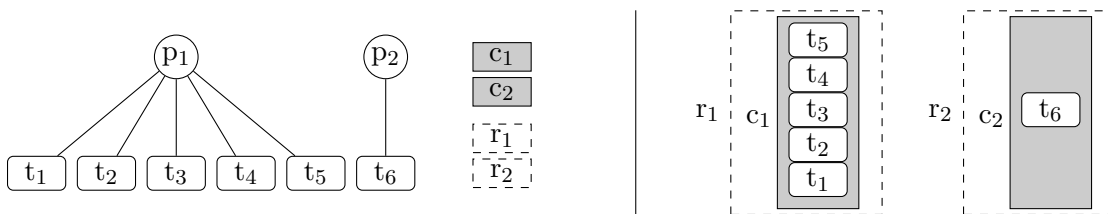


Figure 3: Sample HRP instance (left) and one of its solutions (right)

The sample HRP instance shown in Figure 3 comprises two cabinets, two rooms, five things that belong to person p_1 , and one thing that belongs to person p_2 . A legacy configuration is empty, and all things are small. In a solution, the first person’s things are placed in cabinet c_1 in the first room, and the thing of the second person is in cabinet c_2 in the second room. For this sample instance, a solution of HRP corresponds to a solution of HCP.

6.2.2 ENCODINGS AND INSTANCES

We adapted the original encoding by Friedrich et al. (2011) to conform to the current capabilities of ALPHA (i.e., restricted usage of aggregates, no optimisation).

The main two choice rules guessing the assignment of things to cabinets and the assignment of cabinets to rooms look as follows:

$$\begin{aligned} \{ \text{cabinetTOthing}(C, T) \} &\leftarrow \text{cabinetDomain}(C), \text{thing}(T). \\ \{ \text{roomTOcabinet}(R, C) \} &\leftarrow \text{roomDomain}(R), \text{cabinet}(C). \end{aligned}$$

Instances consist of facts over the following predicates:

- `cabinetDomain/1` defines potential cabinets,
- `roomDomain/1` defines potential rooms;
- `thingLong/1` defines which things are long; and
- `legacyConfig/1` defines all the other data in the legacy configuration, for example:
 - `legacyConfig(personTOthing(p1, t1))` defines that person p_1 owns thing t_1 , and
 - `legacyConfig(roomTOcabinet(r1, c1))` specifies one tuple in the legacy assignment of cabinets to rooms.

Instances for HRP were generated in the pattern of the original instances by Friedrich et al. (2011). This pattern represents four different reconfiguration scenarios encountered in practice, and the instances are abstracted real-world instances. Our instances are considerably larger than the original ones, though (ranging up to 800 things, while the original instances used at most 280 things).

6.2.3 HEURISTICS

The domain-specific heuristic for HRP implemented in our novel approach works by (1) first trying to re-use the legacy configuration; (2) then filling cabinets with things; (3) then filling rooms with cabinets; (4) and finally closing remaining choices. Long things are always assigned before short things.

By “closing remaining choices” we mean assigning **F** to choice points not yet assigned by the heuristic. The purpose of this is to avoid the default heuristics (e.g., VSIDS) from causing conflicts by choosing the wrong truth values.

We now present some selected heuristic directives. The directives use some intermediate predicates whose meaning should become evident from their names. The full encoding is available online.¹⁵

The following heuristics re-use the legacy assignment of cabinets to things and of rooms to cabinets (1):

```
#heuristic reuse(cabinetTOthing(C, T)) :
    legacyConfig(cabinetTOthing(C, T), thingLong(T).    [4@4]

#heuristic reuse(cabinetTOthing(C, T)) :
    legacyConfig(cabinetTOthing(C, T), not thingLong(T). [3@4]

#heuristic reuse(roomTOcabinet(R, C)) :
    legacyConfig(roomTOcabinet(R, C)).    [2@4]
```

The following heuristic assigns things to cabinets, preferring long over short things (2):

```
#heuristic cabinetTOthing(C, T) :
    cabinetDomain(C), not fullCabinet(C),
    not T assignedThing(T), personTOthing(P, T),
    not otherPersonTOcabinet(P, C),
    maxCabinet(MC), thingLong(T).                    [(MC-C)@3]

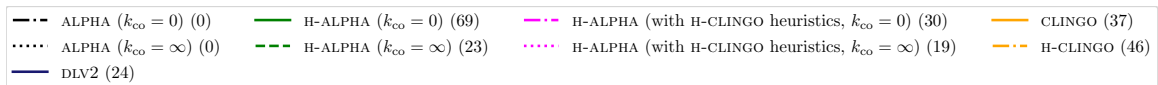
#heuristic cabinetTOthing(C, T) :
    cabinetDomain(C), not fullCabinet(C),
    not T assignedThing(T), personTOthing(P, T),
    not otherPersonTOcabinet(P, C),
    maxCabinet(MC), not thingLong(T).                [(MC-C)@2]
```

The following heuristic assigns cabinets to rooms (3):

```
#heuristic roomTOcabinet(R, C) :
    roomDomain(R), not fullRoom(R),
    cabinet(C), not T assignedCabinet(C),
    personTOcabinet(P, C), not otherPersonTOroom(P, R),
    maxRoom(MR).                                     [(MR-R)@1]
```

Finally, the following heuristics close choice points that are still unassigned (4):

15. <https://ainf.aau.at/dynacon>



(a) Solver configurations, with numbers of solved instances

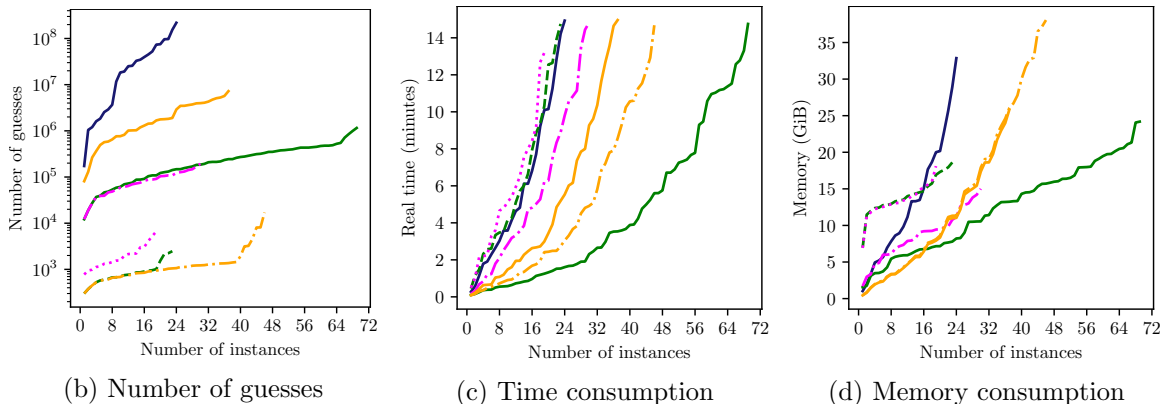


Figure 4: Resource consumption for solving each HRP instance

```

#heuristic F cabinetTOthing(C, T) :
    not cabinetTOthing(C, T), cabinetDomain(C), thing(T).

#heuristic F roomTOcabinet(R, C) :
    not roomTOcabinet(R, C), roomDomain(R), cabinet(C).
    
```

The heuristics we created for ALPHA cannot be used with CLINGO due to the usage of **T** and default negation.

An alternative encoding containing heuristic directives for CLINGO has also been created. This encoding contains heuristic directives that have been faithfully adapted: by using sign modifiers instead of sign symbols in heuristic heads; by adding $(l-1)$ times the maximum weight from the next lower level to w and omitting p ; and by removing all literals from the condition for which it only makes sense to evaluate them w.r.t. a partial assignment.

Furthermore, an ALPHA encoding with heuristics without our novel features, compliant with those employed by CLINGO, has been created.

6.2.4 RESULTS

Figure 4 shows performance data for experiments with HRP. Cactus plots were created in the usual way. In Figure 4c, the x-axis gives the number of instances solved within real (i.e., wall-clock) time, given on the y-axis. Similarly, Figure 4b shows the number of guesses needed and Figure 4d shows the memory consumed to solve the instances. In all three plots, data points are sorted by y-values. Figure 4a contains a legend with all solver configurations. The number of instances solved by each system is shown next to its name (in parentheses).

One curve was drawn for each solver configuration: ALPHA without domain-specific heuristics, with strict ($k_{co} = 0$) and permissive ($k_{co} = \infty$) grounding of constraints; ALPHA with domain-specific heuristics (H-ALPHA), with strict and permissive grounding of con-

straints; ALPHA with H-CLINGO-like domain-specific heuristics, with strict and permissive grounding of constraints; CLINGO with (H-CLINGO) and without domain-specific heuristics; and DLV2.

Substantial differences can be observed. The curves for H-ALPHA ($k_{co} = 0$) reach farthest to the right, meaning that ALPHA with domain-specific heuristics solved the highest number of instances (69 out of 94) when grounding constraints strictly. Surprisingly, with permissive grounding of constraints, ALPHA with domain-specific heuristics exhibited relatively low time and space performance.

No curves are visible at all for ALPHA without domain-specific heuristics because, in this configuration, the system could not solve any instance. The other solvers' performance was somewhere in between the ALPHA configurations at both ends of the spectrum. Notably, H-CLINGO with domain-specific heuristics solved more instances in less time compared to CLINGO without domain-specific heuristics. The largest instance solved by H-ALPHA contained 675 things, which is almost 50% more than the size of the largest instance solved by H-CLINGO (456). Recall that the time limit for solving each instance was 15 minutes.

While Figure 4c only reports overall solving time including grounding efforts, the distribution between grounding and solving time consumed by the ground-and-solve system CLINGO has also been analysed. For each of the 37 instances solved by CLINGO, the system spent between 27% and 84% of total solving time in grounding, on average 63%. When using domain-specific heuristics, H-CLINGO solved 46 instances and spent between 92% and 97% in grounding, on average 96% of overall time. DLV2 solved 24 instances and spent between 5% and 22% of total solving time in grounding, on average 10%.

6.3 Case Study 2: The Partner Units Problem (PUP)

Like HRP, the Partner Units Problem (PUP) (Teppan, 2016; Teppan, Friedrich, & Gottlob, 2016) is an abstracted version of industrial (re)configuration problems. In particular, PUP deals with the configuration of parts of railway safety systems, where the development of domain-specific heuristics was difficult.

6.3.1 PROBLEM DEFINITION

Definition 11 (PUP). *The input to the Partner Units Problem (PUP) is given by a set of units U and a bipartite graph $G = (S, Z, E)$, where S is a set of sensors, Z is a set of zones, and E is a relation between S and Z .*

The task is to find a partition of vertices $v \in S \cup Z$ into bags $u_i \in U$ such that for each bag the following requirements hold: (1) the bag contains at most UCAP vertices from S and at most UCAP vertices from Z ; and (2) the bag has at most IUCAP adjacent bags, where the bags u_1 and u_2 are adjacent whenever $v_i \in u_1$ and $v_j \in u_2$ for some $(v_i, v_j) \in E$.

Figure 5 shows an example of a PUP instance. The bipartite graph comprises six sensors and six zones. Each of the three units can be adjacent to at most two other units, and each unit can contain at most two sensors and two zones. An assignment of sensors and zones to units that satisfies all PUP requirements is also presented in Figure 5.

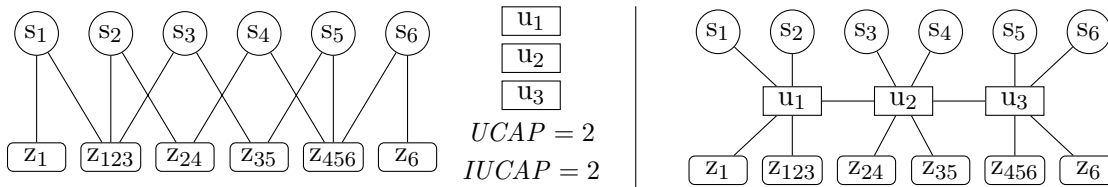


Figure 5: Sample PUP instance (left) and one of its solutions (right)

6.3.2 ENCODINGS AND INSTANCES

PUP instances consist of atoms over the predicates

- `comUnit/1` (specifying units U) and
- `zone2sensor/2` (specifying the zone-to-sensor relation E).¹⁶

We used two different PUP encodings. The most efficient ASP encoding available for PUP is the “new” encoding from the Fifth ASP competition¹⁷ (Calimeri et al., 2016). The following rules constitute the main guessing part of the encoding:¹⁸

$$\begin{aligned} \text{elem}(z, Z) &\leftarrow \text{zone2sensor}(Z, S). \\ \text{elem}(s, S) &\leftarrow \text{zone2sensor}(Z, S). \\ \{ \text{gt}(A, X, U) \} &\leftarrow \text{elem}(A, X), \text{comUnit}(U), \text{comUnit}(U1), U1 = U + 1, \text{rule}(A), U < X. \\ \{ \text{gt}(A, X, U) \} &\leftarrow \text{elem}(A, X), \text{comUnit}(U), \text{comUnit}(U1), U1 = U + 1, \text{not rule}(A). \end{aligned}$$

Other rules are used to derive `unit2zone/2` and `unit2sensor/2` under very specific conditions from `gt/3`.

This encoding is written in such a way that heuristics are encoded in the form of rules and symmetry-breaking constraints, forcing the solver into parts of the search space where a solution can be expected. This goes at the cost of readability. Thus, the encoding does not lend itself well to extension by heuristic directives interacting with the choice rules.

For this reason, a new encoding has been created for ALPHA and extended by heuristic directives. Since also in this encoding parts of the heuristics are encoded as rules, the encoding is introduced in the next section.

All 18 instances from the ASP competitions (Calimeri, Ianni, & Ricca, 2014; Calimeri et al., 2016) that are satisfiable, belong to the subclass of PUP instances that can be polynomially decided, and in which the input graphs are connected have been used for our experiments. Unsatisfiable and non-connected instances have not been used because our heuristics are tailored towards generating valid configurations for connected instances.

Additionally, 100 instances were generated that exhibit the grounding bottleneck but are easy to solve when using dedicated heuristics. Our motivation for using these instances is that problems that humans can solve easily should also be easy for ASP solvers. The generated instances share the characteristics that their sensor-zone graphs are acyclic and that each pair of neighbouring zones shares one sensor.

16. In the instances used for our experiments, both $UCAP$ and $IUCAP$ are fixed at the value 2.

17. https://www.mat.unical.it/aspcomp2014/#Participants.2C_Encodings.2C_Instance_Sets

18. Symbols like D for “Door” have been replaced by S for “Sensor” to be consistent with Definition 11.

The newly generated instances are significantly larger than the competition instances. The largest generated instance contains 300 units, 300 zones, and 597 sensors. In contrast, the largest competition instance under consideration contains 40 units, 60 zones, and 79 sensors.

6.3.3 HEURISTICS

QuickPup is a heuristic for PUP that successfully solves many hard problem instances (Teppan, Friedrich, & Falkner, 2012). Our approach supports implementing large parts of the originally procedural algorithm for QuickPup. Our encoding uses rules by Teppan (2016) and Teppan and Friedrich (2016) to derive a topological order of the zones and sensors. Heuristic directives subsequently use this topological order.

First, a start zone is determined and denoted by `startZone/1`. In our encoding, the start zone is always the first one. QuickPup should actually try to use each zone as the start zone one after the other and abort search after a certain amount of time has passed. This part of the algorithm cannot currently be represented in our framework.

QuickPup assigns zones and sensors to units in a breadth-first-order, called “topological order” because the graph is traversed level by level. First, the start zone is assigned, then the sensors connected to the start zone, then the zones connected to those sensors and so on. A helper predicate `layer/3` is introduced to compute the topological order. In an atom `layer(T, X, L)`, T denotes the type of element (“s” for sensor and “z” for zone), X is the element’s identifier, and L is its layer in the computed breadth-first order.

While experimenting with different encodings and heuristic directives, we found out that ALPHA greatly profits from encodings where the search space is expanded step-by-step according to this topological order, instead of making all choice points available to the solver at once. Therefore, in the encoding used with ALPHA in our final experiments, the topological order is represented by atoms of the `elem_n/3` predicate, e.g., `elem_n(T, X, N)` means that element (T, X) has position N in the topological order. Furthermore, `elems_assigned_up_to/1` denotes how many elements are already assigned to units, e.g., `elems_assigned_up_to(N)` means that the elements with positions 1 up to N in the topological order have already been assigned. The choice of units is also restricted: Since assigning an element to a previously unused unit does not need to differentiate between individual unused units, each unit only becomes assignable when the unit with the next-lower number is already used.

$$\begin{aligned} \text{elem}(z, Z) &\leftarrow \text{zone2sensor}(Z, S). \\ \text{elem}(s, S) &\leftarrow \text{zone2sensor}(Z, S). \\ \text{assignable_unit}(U) &\leftarrow \text{comUnit}(U), \text{used}(U-1). \\ \text{assignable}(U, T, X) &\leftarrow \text{assignable_unit}(U), \text{elem}(T, X), \text{elem_n}(T, X, N), \\ &\quad \text{elems_assigned_up_to}(N-1). \\ \{ \text{assign}(U, T, X) \} &\leftarrow \text{assignable}(U, T, X). \end{aligned}$$

Due to these rules, elements are processed one after the other and can be assigned to any of the units already used or to the unused unit with the lowest number. The selection among these units is guided by heuristic directives, which encode parts of QuickPup. We tried

many different representations of (parts of) QuickPup as heuristic directives and found out that the overhead caused by a fully faithful representation in ALPHA outweighs its benefits. The resulting directives presented below are therefore a compromise between faithfulness to the original heuristics and computational efficiency. The directives use some intermediate predicates whose meaning should become evident from their names. The full encoding is available online.¹⁹

First, already used units are tried in decreasing order:

```
#heuristic assign(U, T, X) : assignable(U, T, X), maxLayer(ML), layer(T, X, L),
                                not full(U, T), not T assigned(T, X),
                                T used(U). [U@(ML-L)]
```

A fresh unit is only touched if assignment to the preceding unit has failed:

```
#heuristic assign(U, T, X) : assignable(U, T, X),
                                maxLayer(ML), layer(T, X, L),
                                not full(U, T), not T assigned(T, X),
                                not T used(U), comUnit(U-1), T used(U-1),
                                F assign(U-1, T, X). [U@(ML-L)]
```

Note the condition \mathbf{F} assign($U-1, T, X$) in the last heuristic directive. Due to this condition, the heuristic is only applicable if the same element could not be assigned to the preceding unit $U-1$. This situation may be caused by backtracking or by the following heuristic avoiding assignments to units that are already full:

```
#heuristic F assign(U, T, X) : assignable(U, T, X),
                                maxLayer(ML), full(U, T),
                                not assign(U, T, X). [1@ML]
```

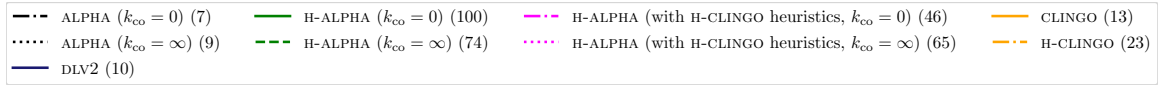
Choice points not assigned by any of these heuristics are finally assigned false by a dedicated heuristic directive, similarly as shown for HRP in Section 6.2.3.

The heuristics we created for ALPHA cannot be used with CLINGO due to the usage of \mathbf{T} , \mathbf{F} , and default negation. An alternative encoding containing heuristic directives for CLINGO has been created in a similar way as for HRP. Furthermore, an ALPHA encoding with heuristics without our novel features, compliant with those employed by CLINGO, has been created.

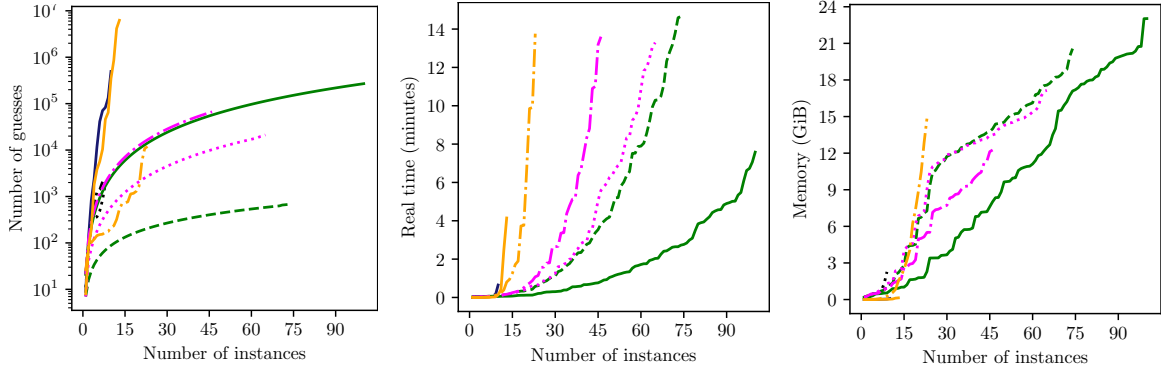
6.3.4 RESULTS

In the experiments, ALPHA used the encoding presented in Section 6.3.3; H-ALPHA additionally used the heuristics presented in Section 6.3.3; ALPHA with H-CLINGO-like domain-specific heuristics used the same encoding but different heuristics; H-CLINGO also used the ALPHA encoding but with different heuristics; and DLV2 and CLINGO used the new ASP competition encoding mentioned in Section 6.3.2.

19. <https://ainf.aau.at/dynacon>



(a) Solver configurations, with numbers of solved instances

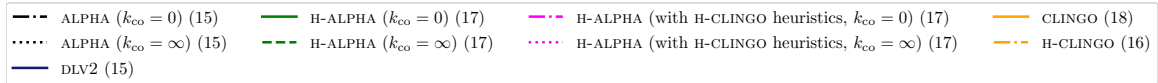


(b) Number of guesses

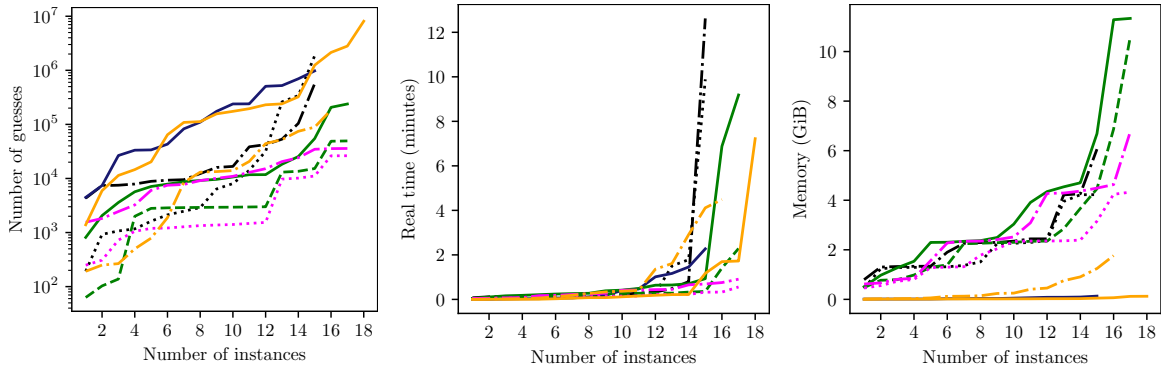
(c) Time consumption

(d) Memory consumption

Figure 6: Resource consumption for solving each generated PUP instance



(a) Solver configurations, with numbers of solved instances



(b) Number of guesses

(c) Time consumption

(d) Memory consumption

Figure 7: Resource consumption for solving each competition PUP instance

We evaluated these approaches on two sets of instances: (1) 100 newly generated instances, and (2) 18 competition instances that are polynomially decidable, connected, and satisfiable. Results obtained in both experiments are shown in cactus plots (Figures 6 and 7), which were generated in the same way as for HRP (cf. Section 6.2.4).

In the first experiment, ALPHA with domain-specific heuristics again solved the highest number of instances (all 100) when grounding constraints strictly, see Figure 6a. When grounding constraints permissively, ALPHA needed much fewer guesses but consumed more

time and memory, as shown in Figures 6b to 6d. On the other extreme, ALPHA without domain-specific heuristics could solve only 7 of the 100 instances when grounding constraints strictly, and 9 instances when grounding constraints permissively. The systems DLV2, CLINGO, and H-CLINGO performed somewhere in between those extremes. H-CLINGO with domain-specific heuristics solved many more instances than CLINGO without domain-specific heuristics.

The largest instance in our instance set contained 300 units. H-ALPHA was able to solve all these instances. In contrast, the size of the largest instance that could be solved by any other system, using the given encoding, was only 75. Recall that the time limit for solving each instance was 15 minutes. For 28 instances, H-CLINGO returned an error: “Value too large for defined data type: Id out of range”.

For each of the 13 instances solved by CLINGO, the system spent between 0% and 100% of total solving time in grounding, on average 34%. When using domain-specific heuristics, H-CLINGO solved 23 instances and spent between 63% and 100% in grounding, on average 85% of the overall time. DLV2 solved 10 instances and spent between 0% and 14% of total solving time in grounding, on average 4%.

In the second experiment, all systems used the same encoding as in the experiment above using the generated instances. The number of guesses needed to find the answer set is significantly reduced by our heuristics (Figure 7b). The effect on time and memory consumption is much lower, however (Figures 7c and 7d). Overall, CLINGO without domain-specific heuristics outperforms all other systems and solves all 18 instances.

Again, domain-specific heuristics raised CLINGO’s grounding efforts: For each of the 18 instances solved by CLINGO, the system spent between 0% and 50% of total solving time in grounding, on average 7%. When using domain-specific heuristics, H-CLINGO solved 16 instances and spent between 2% and 89% in grounding, on average 40% of the overall time. DLV2 solved 15 instances and spent between 0% and 10% of total solving time in grounding, on average 2%.

For comparison, we also ran CLINGO with ALPHA’s encoding without domain-specific heuristics. The results (which are not shown in the figures) revealed that CLINGO performed very poorly with this encoding, solving only 5 of the generated and only 7 competition instances.

6.4 Case Study 3: State-Space Search with A*

A* is a form of *best-first search* that searches a weighted graph for the optimal path from a given start node to a given goal node. Best-first search strategies expand the most promising among all the nodes encountered so far. The promise of a node n is estimated numerically by a heuristic evaluation function, traditionally denoted $f(n)$ (Hart, Nilsson, & Raphael, 1968; Russell & Norvig, 2010; Pearl, 1984).

A* evaluates nodes by combining $g(n)$, the path-cost to reach node n from the start node, and $h(n)$, the estimated cost of the cheapest path from n to the goal (Russell & Norvig, 2010):

$$f(n) = g(n) + h(n).$$

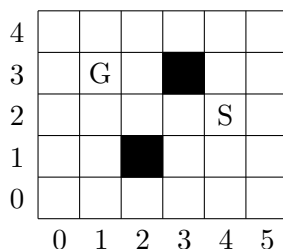


Figure 8: A sample Pathfinding instance (repeated from Figure 1)

A^* is *complete* on finite graphs with non-negative edge weights. The graph-search version of A^* , in which no state is explored repeatedly, is *optimal* if h is consistent. Heuristics discussed in this section are consistent; for details, see Russell and Norvig (2010).

6.4.1 PROBLEM DEFINITION

The formal specification of search problems will now be discussed. The *Pathfinding* problem introduced in Examples 2 and 6 will be used to give examples.²⁰ We repeat Figure 1 in Figure 8 for easier reference.

Definition 12 (Search Problem). *A problem to be solved by a search algorithm like A^* can be defined formally by five components (Russell & Norvig, 2010):*

- The **initial state**, e.g., $\text{at}(4, 2)$.
- The **actions** available. Given a state s , $\text{actions}(s)$ returns the set of actions that can be executed in s .
For example, $\text{actions}(\text{at}(4, 2)) = \{\text{move}(4, 3), \text{move}(5, 2), \text{move}(4, 1), \text{move}(3, 2)\}$.
- The **transition model** specifying the results of each action.
For example, $\text{result}(\text{at}(4, 2), \text{move}(3, 2)) = \text{at}(3, 2)$.
The **state-space** of the problem, the set of all states reachable from the initial state by any sequence of actions, is implicitly defined by the initial state, the actions, and the transition model. The state-space forms a graph in which the nodes are states, and the edges are actions.
- The **goal test**, which determines whether a given state is a goal state. The set of goal states can be defined implicitly or explicitly; in our example, it contains the single element $\text{at}(1, 3)$.
- A **cost function** that assigns a numerical cost to each edge in the graph. The step cost of reaching state s' from s by applying action a is denoted by $c(s, a, s')$. In our example, the cost of each step is the same, e.g., $c(\text{at}(4, 2), \text{move}(3, 2), \text{at}(3, 2)) = 1$. In practical routing problems, step cost might be defined as the length of a road or the time needed to travel between two cities.

²⁰. Note that we disregard time steps in the formulation of Pathfinding as a search problem.

Algorithm 1: Graph Search with A* (adapted from Figures 3.7 and 3.14 by Russell & Norvig, 2010)

Input: P , a problem defined by the components specified in Definition 12.
Output: A solution, or failure.

Initialize frontier F , a priority queue ordered by $f(\cdot)$, initially containing the initial state of P
Initialize explored set E to be empty
while F is not empty **do**
 Pop node n with least $f(n)$ from F
 if n contains a goal state **then**
 | **return** the corresponding solution
 Add state of n to E
 foreach action a available in the state of n **do**
 | Generate child node by applying action a on node n
 | **if** the child state is neither in F nor in E **then**
 | | Insert the child node into F
 | **else if** the child state is already in F , but with a higher path cost **then**
 | | Replace that node in F with child
 |
return failure

A solution for the search problem is a path, representing a sequence of actions, from the initial state to a goal state. An optimal solution is a solution with lowest path cost among all solutions.

Algorithm 1 shows an informal specification of the A* algorithm based on graph-search (adapted from Figures 3.7 and 3.14 by Russell & Norvig, 2010). For details, consult Hart et al. (1968), Russell and Norvig (2010), Pearl (1984). Note that as opposed to tree-search, graph-search maintains a set of explored states to avoid repeated work on states already generated by applying an action.

6.4.2 INSTANCES, ENCODINGS, AND HEURISTICS

We now present our ASP-based encoding of the A* search algorithm based on graph-search, which is the first of its kind to the best of our knowledge. As we apply graph-search we assume the heuristic evaluation function $h(n)$ to be consistent. Moreover, we require that actions are deterministic, i.e., the result of an action is exactly one successor state. To facilitate using our encoding with different search problems, we distinguish between the encoding of A* itself and a problem-specific encoding that can be used with different problem instances. We will demonstrate the usage of our A* encoding with two search problems: Pathfinding known from Example 2, and routing on a real-world street network. We will first present A* with the Pathfinding problem and move from most specific (problem instances) to most general (A* encoding). Afterwards, we will present the routing application. State-based search, where the problem's solution is a path, deals with pathfinding. Therefore we

focus our evaluation on pathfinding and routing, showing the benefits of lazy grounding combined with heuristic search.

Problem Instances for Pathfinding. Instances are defined as stated in Example 2, i.e., they consist of atoms over the following predicates:

- $xmin/1$,
- $xmax/1$,
- $ymin/1$,
- $ymax/1$,
- $start/2$,
- $goal/2$, and
- $obstacle/2$.

For our experiments, we have generated instances as follows. Each instance represents a square plane of width w . For each $w \in \{5, 10, \dots, 500\}$ there exists one instance, so there are 100 instances in total. The start is located in the top-left corner, and the goal is located in the bottom-right corner. Each square in the plane, except for the start and the goal, is an obstacle with a probability of 0.2. Additionally, two long vertical walls cut through the plane, one starting at the bottom and one at the top of the plane. The x -positions of these walls, as well as their lengths, are randomly determined within reasonable ranges.

Problem-Specific Encoding for Pathfinding. The problem-specific encoding defines some intermediate predicates:

- $neighbour/4$ defines which squares are adjacent to each other, and
- $manhattan/3$ defines the Manhattan distance from each square to the goal.

For example, $neighbour(4, 2, 3, 2)$ denotes that $(4, 2)$ is adjacent to $(3, 2)$, and $manhattan(3, 2, 3)$ denotes that the Manhattan distance from $(3, 2)$ to the goal at $(1, 3)$ is 3.

As input to the problem-independent A* encoding, the following predicates have to be defined:

- $init/1$ defines the initial state;
- $goal/1$ defines the goal state;
- $c/4$ defines the cost function; and
- $h/2$ defines the heuristic evaluation function h (which is here the Manhattan distance to the goal).

Actions and transition model need not be explicitly defined because this information is implicitly contained in $c/4$.

The input to A^* is realised in the problem-specific Pathfinding encoding by the following rules:

$$\begin{aligned} \text{init}(\text{at}(X, Y)) &\leftarrow \text{start}(X, Y). \\ \text{goal}(\text{at}(X, Y)) &\leftarrow \text{goal}(X, Y). \\ c(\text{at}(\text{AtX}, \text{AtY}), \text{move}(X, Y), \text{at}(X, Y), \text{StepCost}) &\leftarrow \text{neighbour}(\text{AtX}, \text{AtY}, X, Y), \\ &\quad \text{not obstacle}(X, Y), \\ &\quad \text{StepCost} = 1. \\ h(\text{at}(X, Y), H) &\leftarrow \text{manhattan}(X, Y, H). \end{aligned}$$

Problem-Independent A^* Encoding. We will now present our ASP-based A^* encoding step by step. This problem-independent encoding stays the same for any search problem to which A^* is applied.

The first step is to retrieve the information on the actions and result functions from the problem definition, implicitly contained in $c/4$.

$$\begin{aligned} \text{result}(\text{ParentState}, \text{Action}, \text{ChildState}) &\leftarrow c(\text{ParentState}, \text{Action}, \text{ChildState}, \text{StepCost}). \\ \text{action}(\text{ParentState}, \text{Action}) &\leftarrow \text{result}(\text{ParentState}, \text{Action}, \text{ChildState}). \end{aligned}$$

The $n/4$ predicate is used to represent each node n in the searched graph by state S , parent state PS (state where action A was applied), action A (applied to generate n), and path cost $G = g(n)$. There also exists an $n/1$ predicate that is used when only the state of a node is needed. Recall that $\text{init}/1$ is defined in the problem-specific encoding. The $\text{explore}/2$ predicate will be defined further below; it is used for state/action pairs that have already been explored, i.e., the action applied in a state.

$$\begin{aligned} n(S, \text{null}, \text{null}, 0) &\leftarrow \text{init}(S). \\ n(S, PS, A, G) &\leftarrow \text{result}(PS, A, S), \text{explore}(PS, A), g(PS, A, G). \\ n(S) &\leftarrow n(S, PS, A, G). \end{aligned}$$

Note that if $h(n)$ is consistent, a property of graph-search is that nodes are only explored if their smallest g -value was found (Russell & Norvig, 2010). Based on this property and the determinism of actions it follows that PS and A are uniquely identifying a tuple of $n/4$.

The f -value for each node is the sum of the node's g -value (cost from the initial state to the node) and its h -value (heuristically estimated cost from the node to the nearest goal). Recall that $h/2$ is defined in the problem-specific encoding.

$$\begin{aligned} g(S, G) &\leftarrow n(S, P, A, G). \\ f(S, G + H) &\leftarrow g(S, G), h(S, H). \end{aligned}$$

Nodes are only generated in our approach when they are explored. However, $f(n)$ is already needed before n is explored to decide which nodes to explore. Therefore, we define

f also for state/action pairs.

$$\begin{aligned} g(\text{PS}, A, \text{ParentG} + \text{StepCost}) &\leftarrow g(\text{PS}, \text{ParentG}), c(\text{PS}, A, S, \text{StepCost}). \\ h(\text{PS}, A, H) &\leftarrow \text{result}(\text{PS}, A, S), h(S, H). \\ f(S, A, G + H) &\leftarrow g(S, A, G), h(S, A, H). \end{aligned}$$

Now we can encode elements of A^* (cf. Algorithm 1) as rules. There is no one-to-one correspondence between statements in the algorithm and rules in our ASP encodings, so we present ASP rules in a logical order.

The union of frontier and explored set is represented by `frontier_and_explored_states/2`. The predicate consists of state/action pairs instead of nodes because nodes are not generated before they are explored. Since an atom's truth value cannot change from true to false without backtracking, atoms of the `frontier_and_explored_states/2` predicate also include nodes that have already been removed from the frontier. These removed nodes are in the explored set, i.e., nodes represented by `n/4` were generated. The following rule adds an explored node's children to `frontier_and_explored_states/2`:

$$\text{frontier_and_explored_states}(S, A) \leftarrow n(S), \text{action}(S, A), \text{not goal}(S).$$

The explored set E is already represented by predicates `n/4` and `n/1` because atoms of these predicates represent nodes that are only generated when explored. However, to avoid exploring the same state repeatedly, the `explored_by_other_step/3` predicate stores actions that may not be used because the resulting children have already been explored.

$$\begin{aligned} \text{explored_by_other_step}(S, \text{PS}, A) &\leftarrow n(S, \text{PS}', A', G), \text{result}(\text{PS}, A, S), \text{PS}' \neq \text{PS}. \\ \text{explored_by_other_step}(S, \text{PS}, A) &\leftarrow n(S, \text{PS}', A', G), \text{result}(\text{PS}, A, S), A' \neq A. \end{aligned}$$

When a child state is already explored, the corresponding state/action pair is still added to the frontier but marked as suboptimal.

$$\begin{aligned} \text{suboptimal_step}(\text{PS}, A) &\leftarrow \text{frontier_and_explored_states}(\text{PS}, A), \text{result}(\text{PS}, A, S), \\ &\quad \text{explored_by_other_step}(S, \text{PS}, A). \end{aligned}$$

Similarly, when a cheaper path to a child state is discovered, the existing state/action pair in the frontier is marked as suboptimal.

$$\begin{aligned} \text{suboptimal_step}(\text{PS}, A) &\leftarrow \text{frontier_and_explored_states}(\text{PS}, A), \text{result}(\text{PS}, A, S), \\ &\quad \text{frontier_and_explored_states}(\text{PS}', A'), \text{result}(\text{PS}', A', S), \\ &\quad f(\text{PS}, A, F), f(\text{PS}', A', F'), F > F'. \end{aligned}$$

The following two rules are here to recognise when a goal state is explored.

$$\begin{aligned} \text{goal_found}(S) &\leftarrow n(S), \text{goal}(S). \\ \text{goal_found} &\leftarrow \text{goal_found}(S). \end{aligned}$$

A found goal is not accepted if a node with a lower f -value is still on the frontier. This restriction ensures optimality of found solutions even without heuristic directives.²¹

$$\begin{aligned} \leftarrow & \text{goal_found}(S), \text{explore}(PS, A), \text{result}(PS, A, S), f(S, F), \\ & \text{frontier_and_explored_states}(S', A'), \\ & \text{not explore}(S', A'), \text{not suboptimal_step}(S', A'), \\ & f(S', A', F'), F' < F. \end{aligned}$$

The encoding contains a single choice rule, which realises choosing a leaf node from the frontier. Without heuristic directives, which will be defined later, this rule does not differentiate between nodes in the frontier.

$$\{ \text{explore}(S, A) \} \leftarrow \text{frontier_and_explored_states}(S, A), \text{not suboptimal_step}(S, A).$$

Note that by the consistency property of $h(n)$ nodes in graph-search are only selected for exploration if the minimum g -value for this node was found. We assure this property by the heuristic formulated for `explore/2` (see below). S, A tuples of the frontier with smallest f -value are explored first. Consequently, if there are two distinct state/action pairs (s, a) and (s', a') which were not explored and which lead to the same successor state, then our heuristic will select the state/action pair for exploration which is on an optimal path.

Failure is derived when the frontier is empty, and no goal has been found. An atom of `frontier_nonempty/0` can be derived if there is a state/action pair S, A that is not explored and whose application is not suboptimal. Conversely, `frontier_nonempty/0` cannot be derived if for all S, A pairs for which `frontier_and_explored_states/2` are true, these pairs are suboptimal or explored. If there is an answer set where neither `frontier_nonempty/0` nor `goal_found/0` can be derived then no solution for the search problem exists.

Our encoding of A^* is always satisfiable. An answer set contains a solution if the search problem has a solution, and an answer set contains the failure atom if there is no solution. We have decided this way because the alternative, having a constraint that `goal_found` must be true in an answer set, makes solving performance deteriorate sharply for instances without a solution.

$$\begin{aligned} \text{frontier_nonempty} & \leftarrow \text{frontier_and_explored_states}(S, A), \\ & \text{not suboptimal_step}(S, A), \text{not explore}(S, A). \\ \text{failure} & \leftarrow \text{not frontier_nonempty}, \text{not goal_found}. \\ & \leftarrow \text{not failure}, \text{not goal_found}. \end{aligned}$$

When the goal has been found, the path from the initial state to the goal state can be computed along with corresponding path costs.

$$\begin{aligned} \text{path_to_goal}(PS, A, S) & \leftarrow \text{explore}(PS, A), \text{result}(PS, A, S), \text{goal_found}(S). \\ \text{path_to_goal}(PS, A, S) & \leftarrow \text{explore}(PS, A), \text{result}(PS, A, S), \text{path_to_goal}(S, A', S'). \\ \text{cost_to_goal}(G) & \leftarrow \text{goal_found}(\text{GoalState}), g(\text{GoalState}, G). \end{aligned}$$

21. Using the heuristics presented below, the problem can be solved without backtracking, which means that constraints are not needed when using these heuristics. We still include constraints in the encoding to render it a correct declarative problem specification when viewed without heuristics.

This subsequent computation is necessary because A* may choose to apply more than one action on the same state if one of those actions turns out to lead to a more costly path to the goal. Our approach does not need backtracking, so the answer set includes all explore/2 atoms tried, and we need to reconstruct the path that leads to the goal.

The result computed by A* is encoded by the predicates

- path_to_goal/3,
- cost_to_goal/1, and
- failure/0,

so an answer set solver can be instructed to display only atoms of those predicates in an answer set. For ALPHA, this can be done with the command-line arguments `--filter path_to_goal --filter cost_to_goal --filter failure`.

Of course, our A* encoding also needs heuristic directives to work correctly. Without heuristics, the encoding would just search arbitrarily through the state-space and ignore the f -value of nodes. Therefore, our primary heuristic directive prefers to choose those nodes from the frontier whose f -value is the lowest.

```
#heuristic explore(PS, A) : frontier_and_explored_states(PS, A),
                             not suboptimal_step(PS, A),
                             f(PS, A, PathCost), not T goal_found. [-PathCost@3]
```

A second heuristic directive closes unassigned choice points when the goal is found.

```
#heuristic F explore(PS, A) : frontier_and_explored_states(PS, A),
                             T goal_found. [1@2]
```

The heuristics we created for ALPHA cannot be used with CLINGO due to the usage of **T** and default negation. To assess the performance of CLINGO, we have also created an encoding omitting unsupported literals in heuristic conditions. Furthermore, an ALPHA encoding with heuristics without our novel features, compliant with those employed by CLINGO, has been created.

Generating States on Demand. In many practical domains, the number of states is vast, thus prohibiting the upfront generation of all states. For example, in the well-known 8-puzzle, 181,440 distinct states are reachable. While this is a manageable number, the corresponding number for the 15-puzzle is already about 10^{13} (Russell & Norvig, 2010).

Since only a small fraction of all states is usually explored by A* (depending on the quality of the heuristic function), a crucial memory-saving feature is to generate only the states that A* needs. Our approach can generate states on demand using the interface between problem-specific encoding and A* encoding bidirectionally instead of just passing information from the problem-specific encoding to A*.

As described above, predicates init/1, goal/1, c/4, and h/2 are defined in the problem-specific encoding and accessed by the A* encoding. In the other direction, the problem-specific encoding can access the frontier_and_explored_states/2 predicate defined in the

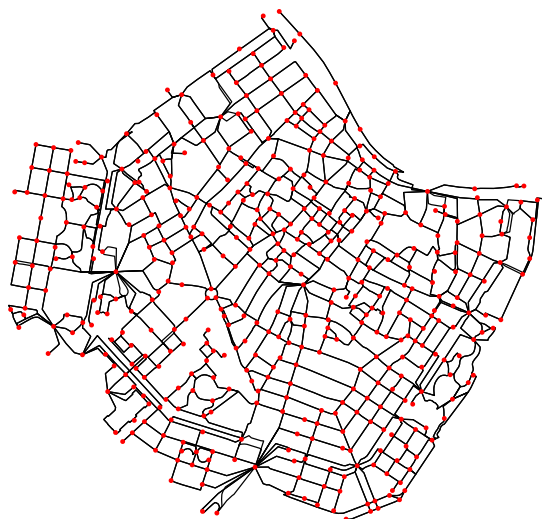


Figure 9: Vienna’s first district

problem-independent A* encoding to generate states on demand, for example, by introducing a new state/1 predicate:

```
state(InitState) ← init(InitState).
state(GoalState) ← goal(GoalState).
state(ChildState) ← frontier_and_explored_states(ParentState, Action),
                  result(ParentState, Action, ChildState).
```

This new state/1 predicate can then be used in the body of rules deriving c/4 and h/2 to derive also information on step cost and heuristic values only on-demand, utilising lazy grounding. Deriving this information on-demand is not possible with ground-and-solve.

Real-World Routing. To test our A* encoding on real world routing problems, we obtained a graph representation of the walkable street network for the first district of Vienna, Austria, from OpenStreetMap,²² employing the OSMnx Python library (Boeing, 2017).²³ We then pre-processed the graph by consolidating intersections (merging nodes located within 15m of each other) and removing parallel edges. Figure 9 shows the resulting graph, which contains 462 nodes and 1632 edges.

We have generated 100 instances, each consisting of the following atoms:

- 1632 atoms of the predicate edge/3, in which the first two arguments denote two nodes connected by an edge and the third argument denotes the length of the edge in meters,

22. <https://www.openstreetmap.org>

23. <https://github.com/gboeing/osmnx>

--- ALPHA ($k_{co} = 0$) (13)	— H-ALPHA ($k_{co} = 0$) (72)	- - - H-ALPHA (with H-CLINGO heuristics, $k_{co} = 0$) (69)	— CLINGO (8)
⋯ ALPHA ($k_{co} = \infty$) (12)	- - - H-ALPHA ($k_{co} = \infty$) (20)	⋯ H-ALPHA (with H-CLINGO heuristics, $k_{co} = \infty$) (19)	- - - H-CLINGO (10)
— DLV2 (8)			

(a) Solver configurations, with numbers of solved instances

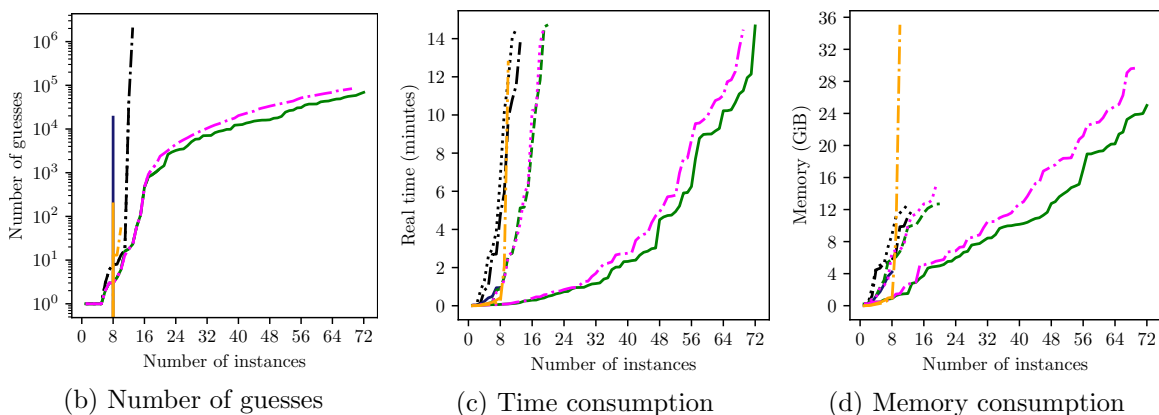


Figure 10: Resource consumption for solving each Pathfinding instance with A*

- 106,953 atoms²⁴ of the predicate distance/3, in which the first two arguments denote nodes and the third argument denotes the euclidean distance between them in meters,
- one atom of the predicate origin/1 denoting the origin node of the route to be computed, and
- one atom of the predicate destination/1 denoting the destination node of the route to be computed.

The set of edge/3 and distance/3 atoms encodes the street network graph and is the same in each of the 100 instances. Origin and destination nodes, however, have been chosen randomly for each instance.

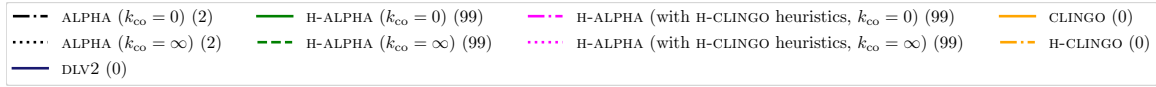
The problem-specific encoding for street network routing then consists of only the following five rules:

$$\begin{aligned}
 \text{init}(\text{at}(\text{N})) &\leftarrow \text{origin}(\text{N}). \\
 \text{goal}(\text{at}(\text{N})) &\leftarrow \text{destination}(\text{N}). \\
 \text{c}(\text{at}(\text{X}), \text{move}(\text{Y}), \text{at}(\text{Y}), \text{Length}) &\leftarrow \text{edge}(\text{X}, \text{Y}, \text{Length}). \\
 \text{h}(\text{at}(\text{X}), \text{Distance}) &\leftarrow \text{destination}(\text{Y}), \text{distance}(\text{X}, \text{Y}, \text{Distance}). \\
 \text{distance}(\text{Y}, \text{X}, \text{Distance}) &\leftarrow \text{distance}(\text{X}, \text{Y}, \text{Distance}).
 \end{aligned}$$

6.4.3 RESULTS

Cactus plots for Pathfinding with A* (Figure 10) and Routing with A* (Figure 11) were generated in the same way as for HRP (cf. Section 6.2.4).

24. Since the graph contains $n = 462$ nodes, there are $\frac{n \cdot (n+1)}{2} = 106,953$ unique pairs of nodes including reflexive pairs.



(a) Solver configurations, with numbers of solved instances

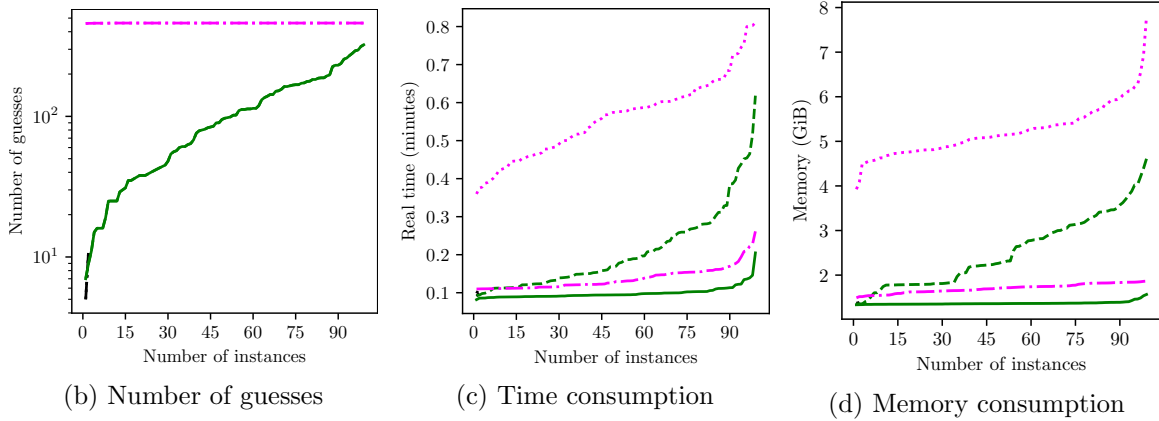


Figure 11: Resource consumption for solving each Routing instance with A*

In the case of Pathfinding (Figure 10), ALPHA with domain-specific heuristics (H-ALPHA) solved the highest number of instances (72 out of 100). Like with HRP, this performance could only be achieved when grounding constraints strictly ($k_{co} = 0$), while permissive grounding of constraints ($k_{co} = \infty$) resulted in lower performance, solving only 20 instances. Performance of ALPHA with H-CLINGO-like heuristics was a bit lower.

On the other extreme, ALPHA without domain-specific heuristics could solve only 13 of the 100 instances.

The systems DLV2, CLINGO, and H-CLINGO performed somewhere in between those extremes. Of the 72 instances solved by H-ALPHA, 18 are unsatisfiable, and of the 10 instances solved by H-CLINGO, 9 are unsatisfiable. The largest instance solved by H-ALPHA had a width of 400, and the largest instance for which H-CLINGO could derive failure had a width of 470. Note that our heuristics are designed to find solutions, not to efficiently prove unsatisfiability.

In the Routing domain (Figure 11), ALPHA with domain-specific heuristics (H-ALPHA) solved 99 of the 100 instances, while ALPHA without domain-specific heuristics could solve only two instances, and the other systems solved no instances at all. ALPHA with H-CLINGO-like heuristics solved the same number of instances as H-ALPHA, but consumed more time and memory in doing so.

For each of the 8 Pathfinding instances solved by CLINGO, the system spent between 88% and 100% of total solving time in grounding, on average 95%. When using domain-specific heuristics, H-CLINGO solved 10 Pathfinding instances and spent between 96% and 100% in grounding, on average more than 99% of overall time. DLV2 solved 8 instances and spent between 15% and 55% of total solving time in grounding, on average 43%. Neither CLINGO nor H-CLINGO nor DLV2 finished grounding for any of the Routing instances.

Note some peculiarities in Figures 10 and 11: In Figure 10b, there is a vertical line for CLINGO and DLV2. This is because for seven instances, these two systems were able to derive failure without doing any guesses. Figure 11b, on the other hand, shows a horizontal line for H-ALPHA with H-CLINGO heuristics. This is because these heuristics do not represent A* correctly and try to visit all 461 non-start nodes in almost every instance. In Figure 11b, all lines for $k_{co} = \infty$ are hidden by their respective $k_{co} = 0$ counterparts because numbers of guesses coincide.

6.5 Discussion

Our results show that we have extended the application area of ASP. By combining our novel approach to domain-specific heuristics with lazy-grounding answer set solving, we could solve large-scale problem instances that are out of reach for conventional ASP systems. This finding supports our initial hypothesis that both lazy grounding and domain-specific heuristics are crucial for solving large-scale industrial problems.

Our approach extends the one by Gebser et al. (2013), the first extension of ASP’s input language by a declarative framework for domain-specific heuristics. Gebser et al. (2013) have provided the first experimental evidence that such an approach can improve ASP solving performance and also reported grounding issues with some instances.

Our advancement consists of novel syntax and semantics for heuristic directives that make it possible to reason about the current partial assignment, facilitating heuristics based on what has or has *not yet* been decided by the solver. Although the approach by Gebser et al. (2013) has worked very well on planning problems, it seems that more flexibility in the definition of heuristics, supported by the novel features of our approach, is necessary to represent heuristics for other kinds of problems.

Our results undeniably show that domain-specific heuristics improve solving performance for the domains under consideration. This is not only true for ALPHA but also for CLINGO. However, domain-specific heuristics usually increase CLINGO’s memory consumption, thus exacerbating the grounding bottleneck from which ground-and-solve systems such as CLINGO are suffering. Domain-specific heuristics for DLV2 were out of scope because DLV2 does not support the declarative specification of heuristics.

However, we do not claim that heuristics based on partial assignments are always beneficial. Our findings cannot reject the possibility that H-CLINGO might outperform H-ALPHA when other encodings or other heuristics are used since there might be encoding optimisations that we have not thought of. We even observed this situation with most competition instances of the Partner Units Problem in our experiments. Still, we are confident that our approach’s novel features make the specification of practical heuristics more intuitive and effortless. Furthermore, in our experiments, ALPHA usually performed worse when employing CLINGO-like heuristics without our novel features.

Results for HRP (Figure 4), PUP with generated instances (Figure 6), and A* (Figures 10 and 11) indicate that permissive grounding (cf. Taupe et al., 2019b), i.e., providing the solver with more nogoods representing ground constraints than necessary, can be counterproductive when domain-specific heuristics are used. We conjecture the reason for this to be that suitable domain-specific heuristics can assist the solver even better than additional constraints while avoiding the overhead of additional nogoods (in terms of space consump-

tion and propagation efforts). This assumption is supported by the considerable increase in ALPHA’s memory consumption when grounding constraints permissively in those domains.

Closer investigation of this issue revealed that in both HRP and A*, one single constraint was the source of the performance deterioration. In HRP, the full and very large grounding for one specific constraint was produced before the solver had even made a single choice. In A*, many ground nogoods were produced for one constraint not necessary to find the optimal solution when domain-specific heuristics are used.

The Partner Units Problem (PUP) proved more challenging for our approach than the other domains under consideration. While the combination of lazy grounding and domain-specific heuristics was able to solve large instances that are out of reach of ground-and-solve systems, hard competition instances can be solved by ground-and-solve systems even without domain-specific heuristics. In these cases, domain-specific heuristics in ALPHA did more harm than good, because they caused a significant overhead and they are obviously not effective when the full grounding and efficient solving techniques are available. However, finding efficient encodings or heuristics for PUP is out of scope of this paper. Future work should investigate whether another combination of encoding and heuristic directives can help to solve instances that are both large and hard to solve.

To sum up, domain-specific heuristics implemented in our novel framework, combined with strict lazy grounding by ALPHA, outperformed all other tested systems when applied to large instances of the House Reconfiguration Problem, the Partner Units Problem, and A* used with two different search problems. Applications to other domains should be easy to put into practice and belong to future work.

7. Conclusions and Future Work

We have proposed novel syntax and semantics for declarative domain-specific heuristics in ASP that can depend non-monotonically on the partial assignment maintained during solving. Furthermore, we have demonstrated how to integrate our approach in a lazy-grounding ASP system and presented experimental results obtained with the lazy-grounding solver ALPHA.

Our semantics has proven beneficial for several practical application domains, advancing the work by Gebser et al. (2013). In experiments, our implementation exhibited convincing time and memory consumption behaviour. Thus, we extended the application area of ASP by solving large problem instances that conventional ASP systems could not solve.

Our approach’s suitability to implement other practice-oriented heuristics should be assessed by the community. Some real-world domain-specific heuristics will require extensions of our approach, such as by supporting randomness and restarts. Furthermore, adopting ideas like `init` and `factor` modifiers from CLINGO, and investigating the special role of aggregates in heuristic conditions should be addressed in future work. An adaptation to ground-and-solve systems like CLINGO (Gebser et al., 2019) or DLV2 (Alviano et al., 2017) should be investigated, also addressing the question of how heuristics interact with non-head-cycle-free disjunction.

Thinking more broadly, the question of how to generate domain-specific heuristics automatically is of great importance since, currently, such heuristics have to be invented by humans familiar with the domain (and partly also with solving technology).

Acknowledgments

This work significantly extends a conference paper by Taupe, Schekotihin, Schüller, Weinzierl, and Friedrich (2019a) by more examples, rethought language syntax, new encodings, a more detailed evaluation, and an extensive study of state-space search with A* implemented in the suggested approach.

Richard Comptoi-Taupe (formerly known as Richard Taupe) is the main author of this paper; other authors contributed equally and are listed in alphabetical order.

This work has been conducted in the scope of the research project *DynaCon* (FFG-PNr.: 861263), which was funded by the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK) under the program “ICT of the Future” (via the project *DynaCon*, FFG-PNr.: 861263) between 2017 and 2020.²⁵ This research was also supported by the Academy of Finland, project 251170, and by EU ECSEL Joint Undertaking under grant agreement no. 737459 (project Productive4.0).

We are thankful to Peter Schüller for his contributions to our earlier conference paper on this topic, Martin Gebser for proofreading this article, Andreas Falkner for his comments on an earlier version of this paper, Stephen Cummings for language editing, and the anonymous reviewers of ICLP 2019 and the Journal of Artificial Intelligence Research for their helpful reviews.

Appendix A. Proofs

Proof sketch for Lemma 2. Since heads of heuristic directives are unaffected by the transformation, it suffices to show that $\text{cond}(d)$ is satisfied iff $\exists d' \in \text{tf}(d)$ s.t. $\text{cond}(d')$ is satisfied (cf. Definitions 3 and 4). Furthermore, it suffices to consider the conditions under which the affected heuristic atoms are satisfied (cf. Definition 2).

Case $\langle 8e \rangle$ is trivial.

For $\langle 8a \rangle$ and $\langle 8d \rangle$, $\text{tf}(d)$ maps d to two new directives, i.e., $\text{tf}(d) = \{d', d''\}$ and $d' \neq d''$. Let $s a$ be the heuristic atom in $\text{cond}(d)$ affected by the transformation. In both cases, $\langle 8a \rangle$ and $\langle 8d \rangle$, $s a$ is removed from $\text{cond}(d)$, one atom is added to (the positive or negative part of) $\text{cond}(d')$, and one atom is added to (the positive or negative part of) $\text{cond}(d'')$. Let us call these new atoms $s' a$ and $s'' a$.

For $\langle 8a \rangle$, all atoms mentioned occur in the positive part of the condition, i.e., $s a \in \text{cond}^+(d)$, $s' a \in \text{cond}^+(d')$, and $s'' a \in \text{cond}^+(d'')$. So we now have to show that $s a$ is satisfied iff $s' a$ is satisfied or $s'' a$ is satisfied. For example, if $s = \mathbf{FM}$, then $s' = \mathbf{M}$ and $s'' = \mathbf{F}$. Since a heuristic atom $s a$ is satisfied w.r.t. an assignment A iff $\text{truth}_A(a) \in s$ (cf. Definition 2), it is obvious that $\mathbf{FM} a$ is satisfied iff $\mathbf{M} a$ is satisfied or $\mathbf{F} a$ is satisfied.

The other cases can be shown in a similar way. \square

Proof of Lemma 3. “ \Rightarrow ”: Let $\text{truth}_A(h) = \mathbf{T}$, resp. $\text{truth}_A(h) = \mathbf{M}$, and recall that A is compatible with the notion of a computation sequence, which means that any atom, except those that represent an applicable rule, i.e., those atoms of the form $\beta(r, \sigma)$ (cf. Section 5.3 and Weinzierl, 2017), are only assigned to \mathbf{T} , respectively \mathbf{M} , if they are propagated. Specifically, this implies that the value assigned to h comes from propagation. Observe that all

²⁵. See <https://iktderzukunft.at/en/> for more information.

nogoods from $\langle 15 \rangle$ - $\langle 20 \rangle$ are such that the heuristic atom only appears with negative polarity in it, i.e., h appears as $\mathbf{F}h \in ng$. Furthermore, since there are no other nogoods that contain h , the only way for $\text{truth}_A(h) = \mathbf{T}$, respectively $\text{truth}_A(h) = \mathbf{M}$, to hold is by ng being unit in $A' = A \setminus \{\mathbf{T}h, \mathbf{M}h\}$.

Since the polarity of h is negative in all nogoods and A is compatible with the notion of a computation sequence, it is impossible for ng to be involved in a conflict, as h is unassigned until ng is unit and then h will be assigned such that ng is not violated and by that all other nogoods containing h can no longer become violated. This implies that ng also is never directly involved in the mechanics of conflict-driven learning, hence learning will not create additional nogoods that contain h .

“ \Leftarrow ”: Let one nogood ng of the form $\langle 15 \rangle$ - $\langle 20 \rangle$ be unit w.r.t. $A' = A \setminus \{\mathbf{T}h, \mathbf{M}h\}$, then unit propagation on ng will lead to an extended assignment where h is assigned \mathbf{T} , respectively \mathbf{M} , i.e., $\text{prop}(ng, A') = \{\mathbf{T}, \mathbf{M}\}$, respectively $\text{prop}(ng, A') = \{\mathbf{M}\}$, and the resulting assignment A is such that $\text{truth}_A(h) = \mathbf{T}$, respectively $\text{truth}_A(h) = \mathbf{M}$, holds. \square

Proof of Lemma 4. “ \Rightarrow ”: Let $\text{cond}(d)$ be satisfied w.r.t. A , i.e., every $ha \in \text{cond}^+(d)$ is satisfied and no $ha \in \text{cond}^-(d)$ is satisfied. We have to show that d is solving-satisfied w.r.t. $A^{\Delta P}$, i.e., all of the following hold:

- (1) $\text{truth}_{A^{\Delta P}}(\text{HeuOnT}(d)) = \mathbf{T}$,
- (2) $\text{truth}_{A^{\Delta P}}(\text{HeuOnMT}(d)) \in \{\mathbf{M}, \mathbf{T}\}$,
- (3) $\text{truth}_{A^{\Delta P}}(\text{HeuOnF}(d)) = \mathbf{T}$,
- (4) $\text{truth}_{A^{\Delta P}}(\text{HeuOffT}(d)) \neq \mathbf{T}$,
- (5) $\text{truth}_{A^{\Delta P}}(\text{HeuOffMT}(d)) \notin \{\mathbf{M}, \mathbf{T}\}$, and
- (6) $\text{truth}_{A^{\Delta P}}(\text{HeuOffF}(d)) \neq \mathbf{T}$.

(1) Since every $ha \in \text{cond}^+(d)$ is satisfied, it holds by Definition 2 that $\text{truth}_A(\text{atom}(ha)) \in \text{signs}(ha)$. This specifically implies for every $a \in \text{atom}(\text{cond}^+(d)|_{\mathbf{T}})$ that $\text{truth}_A(a) = \mathbf{T}$. Hence, the nogood g of the form $\langle 15 \rangle$ is unit w.r.t. A , i.e., $\mathbf{T} \text{HeuOnT}(d) \in \text{prop}(g, A)$ and consequently $\text{truth}_{A^{\Delta P}}(\text{HeuOnT}(d)) = \mathbf{T}$.

(2) The proof is analogous to (1) relying on the nogood of the form $\langle 16 \rangle$ to show that $\text{truth}_{A^{\Delta P}}(\text{HeuOnMT}(d)) \in \{\mathbf{M}, \mathbf{T}\}$.

(3) The proof is analogous to (1) relying on the nogood of the form $\langle 17 \rangle$ to show that $\text{truth}_{A^{\Delta P}}(\text{HeuOnF}(d)) = \mathbf{T}$.

(4) Since no $ha \in \text{cond}^-(d)$ is satisfied w.r.t. A , it holds for every $\mathbf{T}a \in \text{cond}^-(d)$ that $\text{truth}_A(a) \neq \mathbf{T}$. Consequently, no nogood of the form $\langle 18 \rangle$ is unit w.r.t. A . From Lemma 3 it then follows that $\text{truth}_{A^{\Delta P}}(\text{HeuOffT}(d)) \neq \mathbf{T}$.

(5) The proof is analogous to (4) relying on the nogood(s) of the form $\langle 19 \rangle$ and Lemma 3 to show that $\text{truth}_{A^{\Delta P}}(\text{HeuOffMT}(d)) \notin \{\mathbf{M}, \mathbf{T}\}$.

(6) The proof is analogous to (4) relying on the nogood(s) of the form $\langle 20 \rangle$ and Lemma 3 to show that $\text{truth}_{A^{\Delta P}}(\text{HeuOffF}(d)) \neq \mathbf{T}$.

From the above follows that d is solving-satisfied w.r.t. $A^{\Delta P}$.

“ \Leftarrow ”: Let d be solving-satisfied w.r.t. $A^{\Delta P}$, i.e., (1) $\text{truth}_{A^{\Delta P}}(\text{HeuOnT}(d)) = \mathbf{T}$, (2) $\text{truth}_{A^{\Delta P}}(\text{HeuOnMT}(d)) \in \{\mathbf{M}, \mathbf{T}\}$, (3) $\text{truth}_{A^{\Delta P}}(\text{HeuOnF}(d)) = \mathbf{T}$, (4) $\text{truth}_{A^{\Delta P}}(\text{HeuOffT}(d)) \neq \mathbf{T}$, (5) $\text{truth}_{A^{\Delta P}}(\text{HeuOffMT}(d)) \notin \{\mathbf{M}, \mathbf{T}\}$, and (6) $\text{truth}_{A^{\Delta P}}(\text{HeuOffF}(d)) \neq \mathbf{T}$. From Lemma 3 and $\text{truth}_{A^{\Delta P}}(\text{HeuOnT}(d)) = \mathbf{T}$ it directly follows that the nogood of the form $\langle 15 \rangle$ is unit w.r.t. A . Since A is deductively consistent with P it holds for every $a \in \text{atom}(\text{cond}^+(d)|_{\mathbf{T}})$ that $\text{truth}_A(a) = \mathbf{T}$ and consequently, that every $ha \in \text{cond}^+(d)|_{\mathbf{T}}$ is satisfied w.r.t. A . By an analogous argument it also follows that that every $ha \in \text{cond}^+(d)|_{\mathbf{MT}}$ is satisfied w.r.t. A and that every $ha \in \text{cond}^+(d)|_{\mathbf{F}}$ is satisfied w.r.t. A .

Likewise, from $\text{truth}_{A^{\Delta P}}(\text{HeuOffT}(d)) \neq \mathbf{T}$, Lemma 3, and every nogood g of the form $\langle 18 \rangle$ it also follows that $\mathbf{T}a \in \text{cond}^-(d)$ is not assigned true w.r.t. A , i.e., $\text{truth}_A(a) \neq \mathbf{T}$. By Definition 2 it therefore follows that no $a \in \text{atom}(\text{cond}^-(d)|_{\mathbf{T}})$ is satisfied w.r.t. A . By analogous arguments it also follows that no $a \in \text{atom}(\text{cond}^-(d)|_{\mathbf{MT}})$ and no $a \in \text{atom}(\text{cond}^-(d)|_{\mathbf{F}})$ is satisfied w.r.t. A , i.e., no $ha \in \text{atom}(\text{cond}^-(d))$ is satisfied w.r.t. A .

Since every $ha \in \text{cond}^+(d)$ is satisfied w.r.t. A and no $ha \in \text{cond}^-(d)$ is satisfied w.r.t. A , it follows that $\text{cond}(d)$ is satisfied w.r.t. A . □

References

- Adrian, W. T., Alviano, M., Calimeri, F., Cuteri, B., Dodaro, C., Faber, W., Fuscà, D., Leone, N., Manna, M., Perri, S., Ricca, F., Veltri, P., & Zangari, J. (2018). The ASP system DLV: advancements and applications. *Künstliche Intell.*, 32(2-3), 177–179.
- Alviano, M., Amendola, G., Dodaro, C., Leone, N., Maratea, M., & Ricca, F. (2019). Evaluation of disjunctive programs in WASP. In Balduccini, M., Lierler, Y., & Woltran, S. (Eds.), *Logic Programming and Nonmonotonic Reasoning - 15th International Conference, LPNMR 2019, Philadelphia, PA, USA, June 3-7, 2019, Proceedings*, Vol. 11481 of *Lecture Notes in Computer Science*, pp. 241–255. Springer.
- Alviano, M., Calimeri, F., Dodaro, C., Fuscà, D., Leone, N., Perri, S., Ricca, F., Veltri, P., & Zangari, J. (2017). The ASP system DLV2. In Balduccini, M., & Janhunen, T. (Eds.), *Logic Programming and Nonmonotonic Reasoning - 14th International Conference, LPNMR 2017, Espoo, Finland, July 3-6, 2017, Proceedings*, Vol. 10377 of *Lecture Notes in Computer Science*, pp. 215–221. Springer.
- Alviano, M., Calimeri, F., Faber, W., Ianni, G., & Leone, N. (2011). Function symbols in ASP: Overview and perspectives. In Brewka, G., Marek, V., & Truszczyński, M. (Eds.), *Nonmonotonic Reasoning at 30*, Vol. 31 of *Studies in Logic*, pp. 1–24. College Publications.
- Arias, J., Carro, M., Salazar, E., Marple, K., & Gupta, G. (2018). Constraint answer set programming without grounding. *Theory Pract. Log. Program.*, 18(3-4), 337–354.
- Baral, C. (2003). *Knowledge Representation, Reasoning and Declarative Problem Solving*. Cambridge University Press.
- Ben-Eliyahu, R., & Dechter, R. (1994). Propositional semantics for disjunctive logic programs. *Ann. Math. Artif. Intell.*, 12(1-2), 53–87.

- Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Comput. Environ. Urban Syst.*, 65, 126–139.
- Bogaerts, B., & Weinzierl, A. (2018). Exploiting justifications for lazy grounding of answer set programs. In Lang, J. (Ed.), *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, pp. 1737–1745. ijcai.org.
- Bomanson, J., Janhunen, T., & Weinzierl, A. (2019). Enhancing lazy grounding with lazy normalization in answer-set programming. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pp. 2694–2702. AAAI Press.
- Cabalar, P., Rey, M., & Vidal, C. (2019). A complete planner for temporal answer set programming. In Oliveira, P. M., Novais, P., & Reis, L. P. (Eds.), *Progress in Artificial Intelligence, 19th EPIA Conference on Artificial Intelligence, EPIA 2019, Vila Real, Portugal, September 3-6, 2019, Proceedings, Part II*, Vol. 11805 of *Lecture Notes in Computer Science*, pp. 520–525. Springer.
- Calimeri, F., Faber, W., Gebser, M., Ianni, G., Kaminski, R., Krennwallner, T., Leone, N., Maratea, M., Ricca, F., & Schaub, T. (2020). ASP-Core-2 input language format. *Theory Pract. Log. Program.*, 20(2), 294–309.
- Calimeri, F., Fuscà, D., Perri, S., & Zangari, J. (2017). I-DLV: the new intelligent grounder of DLV. *Intelligenza Artificiale*, 11(1), 5–20.
- Calimeri, F., Gebser, M., Maratea, M., & Ricca, F. (2016). Design and results of the fifth answer set programming competition. *Artif. Intell.*, 231, 151–181.
- Calimeri, F., Ianni, G., & Ricca, F. (2014). The third open answer set programming competition. *Theory Pract. Log. Program.*, 14(1), 117–135.
- Cat, B. D., Denecker, M., Bruynooghe, M., & Stuckey, P. J. (2015). Lazy model expansion: Interleaving grounding with search. *J. Artif. Intell. Res.*, 52, 235–286.
- Dantsin, E., Eiter, T., Gottlob, G., & Voronkov, A. (1997). Complexity and expressive power of logic programming. In *Computational Complexity Conference*, pp. 82–101. IEEE Computer Society.
- Dao-Tran, M., Eiter, T., Fink, M., Weidinger, G., & Weinzierl, A. (2012). Omega : An open minded grounding on-the-fly answer set solver. In del Cerro, L. F., Herzig, A., & Mengin, J. (Eds.), *Logics in Artificial Intelligence - 13th European Conference, JELIA 2012, Toulouse, France, September 26-28, 2012. Proceedings*, Vol. 7519 of *Lecture Notes in Computer Science*, pp. 480–483. Springer.
- Dodaro, C., Galatà, G., Khan, M. K., Maratea, M., & Porro, I. (2019). An ASP-based solution for operating room scheduling with beds management. In Fodor, P., Montali, M., Calvanese, D., & Roman, D. (Eds.), *Rules and Reasoning - Third International Joint Conference, RuleML+RR 2019, Bolzano, Italy, September 16-19, 2019, Proceedings*, Vol. 11784 of *Lecture Notes in Computer Science*, pp. 67–81. Springer.

- Dodaro, C., Gasteiger, P., Leone, N., Musitsch, B., Ricca, F., & Schekotihin, K. (2016). Combining answer set programming and domain heuristics for solving hard industrial problems (application paper). *Theory Pract. Log. Program.*, 16(5-6), 653–669.
- Eiter, T., Faber, W., Fink, M., & Woltran, S. (2007). Complexity results for answer set programming with bounded predicate arities and implications. *Ann. Math. Artif. Intell.*, 51(2-4), 123–165.
- Eiter, T., Faber, W., Leone, N., Pfeifer, G., & Polleres, A. (2003). A logic programming approach to knowledge-state planning, II: the dlv^k system. *Artif. Intell.*, 144(1-2), 157–211.
- Erdem, E., Gelfond, M., & Leone, N. (2016). Applications of answer set programming. *AI Magazine*, 37(3), 53–68.
- Faber, W., Leone, N., & Perri, S. (2012). The intelligent grounder of DLV. In *Correct Reasoning*, Vol. 7265 of *Lecture Notes in Computer Science*, pp. 247–264. Springer.
- Faber, W., Leone, N., & Pfeifer, G. (1999). Pushing goal derivation in DLP computations. In Gelfond, M., Leone, N., & Pfeifer, G. (Eds.), *Logic Programming and Nonmonotonic Reasoning, 5th International Conference, LPNMR 1999, El Paso, Texas, USA, December 2-4, 1999, Proceedings*, Vol. 1730 of *Lecture Notes in Computer Science*, pp. 177–191. Springer.
- Faber, W., Pfeifer, G., & Leone, N. (2011). Semantics and complexity of recursive aggregates in answer set programming. *Artif. Intell.*, 175(1), 278–298.
- Falkner, A. A., Friedrich, G., Haselböck, A., Schenner, G., & Schreiner, H. (2016). Twenty-five years of successful application of constraint technologies at Siemens. *AI Magazine*, 37(4), 67–80.
- Falkner, A. A., Friedrich, G., Schekotihin, K., Taupe, R., & Teppan, E. C. (2018). Industrial applications of answer set programming. *Künstliche Intell.*, 32(2-3), 165–176.
- Ferraris, P., Lee, J., & Lifschitz, V. (2011). Stable models and circumscription. *Artif. Intell.*, 175(1), 236–263.
- Friedrich, G., Ryabokon, A., Falkner, A. A., Haselböck, A., Schenner, G., & Schreiner, H. (2011). (Re)configuration using answer set programming. In Shchekotykhin, K. M., Jannach, D., & Zanker, M. (Eds.), *Proceedings of the IJCAI 2011 Workshop on Configuration, Barcelona, Spain, July 16, 2011*, Vol. 755 of *CEUR Workshop Proceedings*, pp. 17–24. CEUR-WS.org.
- Gebser, M., Kaminski, R., Kaufmann, B., Lindauer, M., Ostrowski, M., Romero, J., Schaub, T., Thiele, S., & Wanko, P. (2019). *Potassco guide version 2.2.0*. Retrieved from <https://github.com/potassco/guide/releases/tag/v2.2.0>.
- Gebser, M., Kaminski, R., Kaufmann, B., & Schaub, T. (2012). *Answer Set Solving in Practice*. Morgan and Claypool Publishers.
- Gebser, M., Kaminski, R., Kaufmann, B., & Schaub, T. (2019). Multi-shot ASP solving with clingo. *Theory Pract. Log. Program.*, 19(1), 27–82.
- Gebser, M., Kaminski, R., König, A., & Schaub, T. (2011). Advances in *gringo* series 3. In Delgrande, J. P., & Faber, W. (Eds.), *Logic Programming and Nonmonotonic*

- Reasoning - 11th International Conference, LPNMR 2011, Vancouver, Canada, May 16-19, 2011. Proceedings*, Vol. 6645 of *Lecture Notes in Computer Science*, pp. 345–351. Springer.
- Gebser, M., Kaufmann, B., Romero, J., Otero, R., Schaub, T., & Wanko, P. (2013). Domain-specific heuristics in answer set programming. In desJardins, M., & Littman, M. L. (Eds.), *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence, July 14-18, 2013, Bellevue, Washington, USA*. AAAI Press.
- Gebser, M., Kaufmann, B., & Schaub, T. (2012). Conflict-driven answer set solving: From theory to practice. *Artif. Intell.*, 187, 52–89.
- Gebser, M., Leone, N., Maratea, M., Perri, S., Ricca, F., & Schaub, T. (2018). Evaluation techniques and systems for answer set programming: a survey. In Lang, J. (Ed.), *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, pp. 5450–5456. ijcai.org.
- Gebser, M., Maratea, M., & Ricca, F. (2017). The sixth answer set programming competition. *J. Artif. Intell. Res.*, 60, 41–95.
- Gebser, M., Maratea, M., & Ricca, F. (2020). The seventh answer set programming competition: Design and results. *Theory Pract. Log. Program.*, 20(2), 176–204.
- Gebser, M., Obermeier, P., Schaub, T., Ratsch-Heitmann, M., & Runge, M. (2018). Routing driverless transport vehicles in car assembly with answer set programming. *Theory Pract. Log. Program.*, 18(3-4), 520–534.
- Gelfond, M., & Kahl, Y. (2014). *Knowledge Representation, Reasoning, and the Design of Intelligent Agents: The Answer-Set Programming Approach*. Cambridge University Press.
- Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE Trans. Syst. Sci. Cybern.*, 4(2), 100–107.
- Hoos, H. H., Lindauer, M., & Schaub, T. (2014). claspfolio 2: Advances in algorithm selection for answer set programming. *Theory Pract. Log. Program.*, 14(4-5), 569–585.
- Kaufmann, B., Leone, N., Perri, S., & Schaub, T. (2016). Grounding and solving in answer set programming. *AI Magazine*, 37(3), 25–32.
- Lefèvre, C., Béatrix, C., Stéphan, I., & Garcia, L. (2017). ASPeRiX, a first-order forward chaining approach for answer set computing. *Theory Pract. Log. Program.*, 17(3), 266–310.
- Lefèvre, C., & Nicolas, P. (2009). A first order forward chaining approach for answer set computing. In Erdem, E., Lin, F., & Schaub, T. (Eds.), *Logic Programming and Nonmonotonic Reasoning, 10th International Conference, LPNMR 2009, Potsdam, Germany, September 14-18, 2009. Proceedings*, Vol. 5753 of *Lecture Notes in Computer Science*, pp. 196–208. Springer.
- Leone, N., Pfeifer, G., Faber, W., Eiter, T., Gottlob, G., Perri, S., & Scarcello, F. (2006). The DLV system for knowledge representation and reasoning. *ACM Trans. Comput. Log.*, 7(3), 499–562.

- Leone, N., & Ricca, F. (2015). Answer set programming: A tour from the basics to advanced development tools and industrial applications. In Faber, W., & Paschke, A. (Eds.), *Reasoning Web*, Vol. 9203 of *Lecture Notes in Computer Science*, pp. 308–326. Springer.
- Leutgeb, L., & Weinzierl, A. (2017). Techniques for efficient lazy-grounding ASP solving. In Seipel, D., Hanus, M., & Abreu, S. (Eds.), *Declarative Programming and Knowledge Management - Conference on Declarative Programming, DECLARE 2017, Unifying INAP, WFLP, and WLP, Würzburg, Germany, September 19-22, 2017, Revised Selected Papers*, Vol. 10997 of *Lecture Notes in Computer Science*, pp. 132–148. Springer.
- Lifschitz, V. (2019). *Answer Set Programming*. Springer.
- Liu, L., Pontelli, E., Son, T. C., & Truszczyński, M. (2007). Logic programs with abstract constraint atoms: The role of computations. In Dahl, V., & Niemelä, I. (Eds.), *Logic Programming, 23rd International Conference, ICLP 2007, Porto, Portugal, September 8-13, 2007, Proceedings*, Vol. 4670 of *Lecture Notes in Computer Science*, pp. 286–301. Springer.
- Maratea, M., Pulina, L., & Ricca, F. (2015). Advances in multi-engine ASP solving. In Gavanelli, M., Lamma, E., & Riguzzi, F. (Eds.), *AI*IA 2015, Advances in Artificial Intelligence - XIVth International Conference of the Italian Association for Artificial Intelligence, Ferrara, Italy, September 23-25, 2015, Proceedings*, Vol. 9336 of *Lecture Notes in Computer Science*, pp. 179–190. Springer.
- Marple, K., Salazar, E., & Gupta, G. (2017). Computing stable models of normal logic programs without grounding. *CoRR*, *abs/1709.00501*.
- Moskewicz, M. W., Madigan, C. F., Zhao, Y., Zhang, L., & Malik, S. (2001). Chaff: Engineering an efficient SAT solver. In *Proceedings of the 38th Design Automation Conference, DAC 2001, Las Vegas, NV, USA, June 18-22, 2001*, pp. 530–535. ACM.
- Palù, A. D., Dovier, A., Pontelli, E., & Rossi, G. (2009). GASP: answer set programming with lazy grounding. *Fundam. Inform.*, *96*(3).
- Pearl, J. (1984). *Heuristics - intelligent search strategies for computer problem solving*. Addison-Wesley series in artificial intelligence. Addison-Wesley.
- Redl, C. (2016). Automated benchmarking of KR-systems. In Bistarelli, S., Formisano, A., & Maratea, M. (Eds.), *Proceedings of the 23rd RCRA International Workshop on Experimental Evaluation of Algorithms for Solving Problems with Combinatorial Explosion 2016 (RCRA 2016), Genova, Italy, November 28, 2016*, Vol. 1745 of *CEUR Workshop Proceedings*, pp. 45–56. CEUR-WS.org.
- Ricca, F., Grasso, G., Alviano, M., Manna, M., Lio, V., Iiritano, S., & Leone, N. (2012). Team-building with answer set programming in the gioia-tauro seaport. *Theory Pract. Log. Program.*, *12*(3), 361–381.
- Russell, S. J., & Norvig, P. (2010). *Artificial Intelligence - A Modern Approach, Third International Edition*. Pearson Education.
- Silva, J. P. M., Lynce, I., & Malik, S. (2009). Conflict-driven clause learning SAT solvers. In *Handbook of Satisfiability*, pp. 131–153. IOS Press.

- Taupe, R., Schekotihin, K., Schüller, P., Weinzierl, A., & Friedrich, G. (2019a). Exploiting partial knowledge in declarative domain-specific heuristics for ASP. In Bogaerts, B., Erdem, E., Fodor, P., Formisano, A., Ianni, G., Incezan, D., Vidal, G., Villanueva, A., Vos, M. D., & Yang, F. (Eds.), *Proceedings 35th International Conference on Logic Programming (Technical Communications), ICLP 2019 Technical Communications, Las Cruces, NM, USA, September 20-25, 2019*, Vol. 306 of *EPTCS*, pp. 22–35.
- Taupe, R., Weinzierl, A., & Friedrich, G. (2019b). Degrees of laziness in grounding - effects of lazy-grounding strategies on ASP solving. In Balduccini, M., Lierler, Y., & Woltran, S. (Eds.), *Logic Programming and Nonmonotonic Reasoning - 15th International Conference, LPNMR 2019, Philadelphia, PA, USA, June 3-7, 2019, Proceedings*, Vol. 11481 of *Lecture Notes in Computer Science*, pp. 298–311. Springer.
- Teppan, E. C. (2016). Solving the partner units configuration problem with heuristic constraint answer set programming. In *Configuration Workshop*, pp. 61–68.
- Teppan, E. C., & Friedrich, G. (2016). Heuristic constraint answer set programming. In Kaminka, G. A., Fox, M., Bouquet, P., Hüllermeier, E., Dignum, V., Dignum, F., & van Harmelen, F. (Eds.), *ECAI 2016 - 22nd European Conference on Artificial Intelligence, 29 August-2 September 2016, The Hague, The Netherlands*, Vol. 285 of *Frontiers in Artificial Intelligence and Applications*, pp. 1692–1693. IOS Press.
- Teppan, E. C., Friedrich, G., & Falkner, A. A. (2012). QuickPup: A heuristic backtracking algorithm for the partner units configuration problem. In Fromherz, M. P. J., & Muñoz-Avila, H. (Eds.), *Proceedings of the Twenty-Fourth Conference on Innovative Applications of Artificial Intelligence, July 22-26, 2012, Toronto, Canada*. AAAI.
- Teppan, E. C., Friedrich, G., & Gottlob, G. (2016). Tractability frontiers of the partner units configuration problem. *J. Comput. Syst. Sci.*, 82(5), 739–755.
- Weinzierl, A. (2017). Blending lazy-grounding and CDNL search for answer-set solving. In Balduccini, M., & Janhunen, T. (Eds.), *Logic Programming and Nonmonotonic Reasoning - 14th International Conference, LPNMR 2017, Espoo, Finland, July 3-6, 2017, Proceedings*, Vol. 10377 of *Lecture Notes in Computer Science*, pp. 191–204. Springer.
- Zhang, L., Madigan, C. F., Moskewicz, M. H., & Malik, S. (2001). Efficient conflict driven learning in a boolean satisfiability solver. In *IEEE/ACM International Conference on Computer Aided Design, ICCAD 2001*, IEEE/ACM Digest of Technical Papers, pp. 279–285. IEEE.