Simulating Sets in Answer Set Programming[*]

Sarah Alice Gaggl[1] , **Philipp Hanisch**[2] and **Markus Krötzsch**[2]
[1]Logic Programming and Argumentation Group, TU Dresden, Germany
[2]Knowledge-Based Systems Group, TU Dresden, Germany
{sarah.gaggl, philipp.hanisch1, markus.kroetzsch}@tu-dresden.de

Abstract

We study the extension of non-monotonic disjunctive logic programs with terms that represent sets of constants, called DLP(S), under the stable model semantics. This strictly increases expressive power, but keeps reasoning decidable, though cautious entailment is $\text{coNExpTime}^{\text{NP}}$-complete, even for data complexity. We present two new reasoning methods for DLP(S): a semantics-preserving translation of DLP(S) to logic programming with function symbols, which can take advantage of lazy grounding techniques, and a ground-and-solve approach that uses non-monotonic existential rules in the grounding stage. Our evaluation considers problems of ontological reasoning that are not in scope for traditional ASP (unless $\text{ExpTime} = \Pi_2^{\text{P}}$), and we find that our new existential-rule grounding performs well in comparison with native implementations of set terms in ASP.

1 Introduction

The success of answer set programming (ASP) is in no small part due to its declarative approach to computation, which allows users to encode complex problems in a clean, direct manner [Brewka *et al.*, 2011; Lifschitz, 2019]. Plain ASP can express problems up to the second level of the polynomial hierarchy [Dantsin *et al.*, 2001], and highly optimised ASP solvers often lead to efficient solutions in such cases. However, not all problems can be expressed in this way, and ASP has therefore been extended with more complex terms that support function symbols [Bonatti, 2004], list, and sets [Calimeri *et al.*, 2009].

The ability to encode sets (of constants) in terms is of particular interest here, since it preserves – in contrast to function symbols and lists – the decidability of reasoning. Moreover, sets and related predicates like $\in$ and $\subseteq$ support a very declarative modelling style. For example, the rules in Figure 1 define maximal strongly connected components (scc) on a graph described by edges $e(X, Y)$ and vertices $v(X)$. The rules for c show how sets can be constructed iteratively using suitable functions, while the last two rules provide an

$$\begin{aligned}
e^+(X,Y) &\leftarrow e(X,Y)\\
e^+(X,Y) &\leftarrow e^+(X,Z) \wedge e(Z,Y)\\
c(\{X\}) &\leftarrow v(X)\\
c(S \cup \{Y\}) &\leftarrow c(S) \wedge X \in S \wedge e^+(X,Y) \wedge e^+(Y,X)\\
subc(S_1) &\leftarrow c(S_1) \wedge c(S_2) \wedge S_1 \subseteq S_2 \wedge \textbf{not}\, S_2 \subseteq S_1\\
scc(S) &\leftarrow c(S) \wedge \textbf{not}\, subc(S)
\end{aligned}$$

Figure 1: Strongly connected components in ASP with sets

elegant description of maximal sets. Such elegance presumably was the main motivation for implementations of sets in ASP solvers [Calimeri *et al.*, 2009], but sets have further computational advantages in boosting expressive power, as noted already for the case of Datalog [Ortiz *et al.*, 2010; Carral *et al.*, 2019b].

Surprisingly, the extension of ASP with sets remains poorly understood, regarding both its theoretical properties and its practical potential. Most notably, Calimeri *et al.* [2008; 2009] discuss set terms in combination with functions and lists, and provide a first implementation in *DLV-complex*. However, to our knowledge, even the complexity of reasoning in ASP with sets has not been established yet, and its utility for solving computationally hard problems has never been discussed or evaluated. One may also wonder whether formalisms that are even more powerful, such as rules with function symbols or value invention, could be exploited to implement sets, but research on feasible implementation methods is similarly scarce. Significant potential for using ASP and related tools therefore remains unexplored.

To address this, we take a closer look at the extension of ASP with set terms, investigate possible implementation methods in theory, and evaluate their practical feasibility in solving problems that are beyond the expressive power of plain ASP. Our main contributions are:

- We define DLP(S), the extension of disjunctive logic programs with set terms, and establish relevant reasoning complexities under the stable model semantics.
- We show how DLP(S) reasoning can be reduced to reasoning in disjunctive logic programs with function symbols while still ensuring finite stable models.
- We show that existing lazy ASP solvers can handle these

[*]Extended version of IJCAI'22 publication with appendix.

programs, whereas traditional ASP engines will fail.

- We develop another alternative grounding approach based on a technique for modelling sets in existential rules [Carral *et al.*, 2019b].
- We evaluate the ability of our new methods and the existing set implementation DLV-complex to solve complex real-world problems from ontology engineering.

Full proofs are in the appendix. Datasets and source code for our prototype implementations can be accessed online at https://github.com/knowsys/eval-2022-IJCAI-asp-with-sets.

2 ASP with Sets

We now introduce the extension of non-monotonic disjunctive logic programs with finite sets (DLP(S)). This will allow us to write rules as in Figure 1.

Syntax. Formally, we consider two *sorts*: a sort of *objects* $\mathbf{obj}$ and a sort of *sets* $\mathbf{set}$. A signature of DLP(S) is based on countably infinite sets of *object constants* $\mathbf{C_{obj}}$ (typically a, b, c), *object variables* $\mathbf{V_{obj}}$ (typically X, Y, Z), *set variables* $\mathbf{V_{set}}$ (typically S, U, V), and *predicate names* $\mathbf{P}$ including *special predicates* $\{\in, \subseteq\} \subseteq \mathbf{P}$. The *signature* of a predicate symbol p is a tuple $sig(p) = \langle \mathbf{S_1}, \ldots, \mathbf{S_n} \rangle$ of sorts, and n is called its *arity*. We have $sig(\in) = \langle \mathbf{obj}, \mathbf{set} \rangle$ and $sig(\subseteq) = \langle \mathbf{set}, \mathbf{set} \rangle$. An *object term* is any object constant or object variable. A *set term* is any set variable, the special constant $\emptyset$, an expression $\{t\}$ where t is an object term, or, recursively, an expression $(s_1 \cup s_2)$ where s_1, s_2 are set terms. We use $\{t_1, \ldots, t_n\}$ as abbreviation for $(\{t_1\} \cup (\{t_2\} \cup \ldots \cup \{t_n\} \ldots))$ and omit parentheses for $\cup$. An term or formula is *ground* if it contains no variables.

An *atom* is an expression $p(t_1, \ldots, t_n)$ where $p \in \mathbf{P}$ with $sig(p) = \langle \mathbf{S_1}, \ldots, \mathbf{S_n} \rangle$ and t_i is a term of sort $\mathbf{S_i}$. We write $\in(t, s)$ as $t \in s$, and $\subseteq(s_1, s_2)$ as $s_1 \subseteq s_2$. A *literal* is an atom α or a negated atom $\mathbf{not}\, \alpha$. We sometimes treat conjunctions or disjunctions of literals as sets of literals. For a formula F, $\mathrm{preds}(F)$ denotes the set of all predicates in F.

Definition 1. A *DLP(S) rule* is an expression r of the form $H \leftarrow B^+ \wedge B^-$, where the *head* H is a disjunction of atoms, the *positive body* B^+ is a conjunction of atoms, and the *negative body* B^- is a conjunction of negated atoms, such that:

1. every object variable in r occurs in B^+,
2. every set variable S in r occurs in B^+ as an argument $t_i = S$ $(1 \leq i \leq n)$ of an atom $p(t_1, \ldots, t_n) \in B^+$ with non-special predicate $p \in \mathbf{P} \setminus \{\in, \subseteq\}$, and
3. the special predicates $\in$ and $\subseteq$ do not occur in H.

A *fact* is a disjunction-free rule with empty body, i.e., a ground atom. A *DLP(S) program P* is a set of DLP(S) rules. A rule, fact, or program is DLP if it contains only object terms. We also consider the extension of DLP with arbitrary (object) function symbols, which we will denote DLP^f.

Some restrictions in Definition 1 could be relaxed without fundamentally changing the properties of DLP(S). Our selection is also motivated by practical concerns, e.g., since a rule $q(S) \leftarrow p(S \cup T)$ (which violates condition 2) would produce exponentially many derivations in the size of any set R for which $p(R)$ holds.

A substitution σ is a sort-preserving partial mapping from variables to terms. $F\sigma$ denotes the result of applying σ to all variables in the formula or term F for which it is defined.

Semantics. We define a stable model semantics for DLP(S) by interpreting set terms as finite sets over the (object) domain. Given a program P, let $\mathbf{obj}(P)$ be the set of all object constants in P (including those used in set terms), and let $\mathbf{set}(P)$ be a set of terms of the form $\{t_1, \ldots, t_n\}$ that bijectively correspond to the powerset of $\mathbf{obj}(P)$. The *grounding* $\mathrm{ground}(P)$ of P consists of all rules that can be obtained from a rule r of P by (1) uniformly replacing object and set variables in r by terms from $\mathbf{obj}(P)$ and $\mathbf{set}(P)$, respectively; and (2) replacing each of the resulting set terms s by the corresponding term from $\mathbf{set}(P)$ that represents the same set under the usual interpretation of $\emptyset$, $\{\cdot\}$, and $\cup$.

An *interpretation $\mathcal{I}$* for P is a set of ground facts that only uses terms from $\mathbf{obj}(P)$ and $\mathbf{set}(P)$, and that contains exactly those facts $c \in t$ (resp. $s \subseteq t$) for which c occurs in the set term t (resp. for which all constants in s occur in t). $\mathcal{I}$ *satisfies* (or *is a model of*) a ground atom α if $\alpha \in \mathcal{I}$. $\mathcal{I}$ satisfies a ground conjunction B (disjunction H) if $B \subseteq \mathcal{I}$ ($H \cap \mathcal{I} \neq \emptyset$), and a positive ground rule $H \leftarrow B$ if it satisfies H or does not satisfy B. The *reduct* $P^{\mathcal{I}}$ of P with respect to $\mathcal{I}$ is obtained from $\mathrm{ground}(P)$ by (1) deleting every negated atom $\mathbf{not}\, \alpha$ with $\alpha \notin \mathcal{I}$ and (2) deleting every rule with a negated atom $\mathbf{not}\, \alpha$ with $\alpha \in \mathcal{I}$. $\mathcal{I}$ is a *stable model* of P if it is a subset-minimal model of $P^{\mathcal{I}}$. A fact α is (cautiously) entailed by P if every stable model of P contains the fact α', obtained by replacing set terms in α with their representative in $\mathbf{set}(P)$.

Since there are exponentially many sets over a given object domain, groundings, reducts, and stable models can also be exponential for DLP(S), even in the size of the data:

Theorem 1. *Deciding whether P entails α is* coNExpTime$^{\mathrm{NP}}$*-complete. If P does not contain $\vee$, the problem is* coNExpTime*-complete. In either case, hardness holds even if only facts are allowed to vary while all other rules are fixed (data complexity).*

Proof sketch. The data complexities of the considered problems are Π_2^{P}-complete respectively coNP-complete for DLP [Dantsin *et al.*, 2001]. Hardness can be shown, e.g., by reducing from the word problem of polynomial-time alternating Turing machines with one quantifier alternation. The claims for DLP(S) follow by simulating an exponentially time-bounded alternating Turing machine instead.

The encoding is standard once an exponentially long chain (of time points or tape cells) is inferred. Given input facts $succ(1, 2), \ldots, succ(\ell - 1, \ell)$, a chain of length 2^ℓ is characterised by a predicate c, defined as follows:

$$succ^+(X, Y) \leftarrow succ(X, Y) \tag{1}$$

$$succ^+(X, Z) \leftarrow succ^+(X, Y) \wedge succ(Y, Z) \tag{2}$$

$$n(\emptyset, \{1\}, 1, \emptyset) \leftarrow \tag{3}$$

$$\begin{aligned} n(U, \{X\} \cup \grave{V}, X, \grave{V}) \leftarrow\ & n(\_, U, X, \grave{U}) \wedge n(\grave{U}, \grave{V}, \grave{X}, \_) \\ & \wedge succ^+(\grave{X}, X) \end{aligned} \tag{4}$$

$$\begin{aligned} n(U, \{Y\}, Y, \emptyset) \leftarrow\ & n(\_, U, X, \grave{U}) \wedge n(\grave{U}, \_, X, \_) \\ & \wedge succ(X, Y) \end{aligned} \tag{5}$$

$$c(S,T) \leftarrow n(S,T,\_,\_) \tag{6}$$

Here we use _ for anonymous variables. The encoding is adapted from Carral *et al.* [2019b]. Predicates $n(S,T,X,V)$ express that (1) T is the successor of S if both sets are taken as binary numbers with elements from $\{1,\ldots,\ell\}$ encoding active digits, (2) X is the largest number (most significant bit) in T, and (3) $T \setminus \{X\} = V$. With this in mind, (4) and (5) increment numbers with the highest bit staying the same or being increased by one, respectively.

Membership is also derived from known results by reducing the DLP(S) problems to propositional logic programming problems via grounding, which incurs an exponential blowup in the presence of set terms. □

3 Reducing DLP(S) to DLP

Answer Set Programming over DLP(S) could be implemented by similar techniques as normal ASP, taking the semantics of sets into account when applying rules (operational extension APIs like clingo's *@-terms* or DLV's *external atoms* could be used). However, this is rarely done so far, and we therefore explore alternative approaches. A first idea is to reduce DLP(S) to DLP with arbitrary function symbols (DLP^f), which is known to be computationally very powerful [Bonatti, 2004; Eiter and Simkus, 2010]. In this section, we show how this can be done and establish the correctness of the translation. Its computational properties and possible implementation in current solvers are discussed later.

We encode sets using a constant $c_\emptyset$ (the empty set) and a function $f_\cup$, such that, e.g., $f_\cup(a, f_\cup(b, c_\emptyset))$ represents the set $\{a,b\}$. However, we cannot generally represent $\{c\} \cup s$ by $f_\cup(c,s)$, since this would lead to redundant representations like $f_\cup(a, f_\cup(a, c_\emptyset))$. Instead, we use facts of the form $su(c,s,t)$ – where su stands for *singleton union* – to express $\{c\} \cup s = t$, where t might be different from $f_\cup(c,s)$. In particular, $su(c,s,s)$ means $c \in s$, which we abbreviate by $in(c,s)$. Finally, to ensure that only necessary set encodings are derived, we use facts $get\_su(c,s)$ to express that a representation of $\{c\} \cup s$ is needed. The following DLP^f rules implement these ideas:

$$su(X,S,f_\cup(X,S)) \leftarrow get\_su(X,S) \wedge \textbf{not}\, in(X,S) \tag{7}$$
$$su(X,U,U) \leftarrow su(X,S,U) \tag{8}$$
$$su(Y,U,U) \leftarrow su(X,S,U) \wedge in(Y,S) \tag{9}$$
$$in(X,S) \leftarrow su(X,S,S) \tag{10}$$

We can exercise this machinery to define rules that compute arbitrary unions using analogous predicates u and get_u:

$$u(S,c_\emptyset,S) \leftarrow get\_u(S,c_\emptyset) \tag{11}$$
$$u(S,f_\cup(X,T),U) \leftarrow get\_u(S,f_\cup(X,T)) \wedge su(X,S,\acute{S}) \wedge u(\acute{S},T,U) \tag{12}$$
$$get\_su(X,S) \leftarrow get\_u(S,f_\cup(X,T)) \tag{13}$$
$$get\_u(\acute{S},T) \leftarrow get\_u(S,f_\cup(X,T)) \wedge su(X,S,\acute{S}) \tag{14}$$

Here, (11) is the base and (12) the recursive case, and (13) and (14) ensure the computation of necessary auxiliary facts.

We are now ready to transform a single DLP(S) rule $r = H \leftarrow B$ to a set of DLP^f rules $\mathrm{dlp^f}(r)$. For every DLP(S) predicate p of arity n, let $\hat{p}$ be a unique fresh predicate of arity n and signature $sig(p) = \langle \mathbf{obj}, \ldots, \mathbf{obj} \rangle$, and for every set term s, let V_s be a fresh object variable. We construct H' and B' from H and B, respectively, by replacing each atom $t \in s$ by $in(t,s)$, each atom $s \subseteq u$ by $sub(s,u)$, every predicate p by $\hat{p}$, and every set term s by V_s. Moreover, let B^+ be the conjunction of positive literals in B'. Now let $s_1,\ldots,s_k$ be a list of all terms and subterms in r of the form $\{\!\!\{t\}\!\!\}$ or $s \Cup u$, ordered such that subterms occur before their superterms. Then $\mathrm{dlp^f}(r)$ consists of the following rules:

$$H' \leftarrow B' \wedge \textstyle\bigwedge_{i=1}^{k} \beta(s_i) \tag{15}$$
$$\alpha(s_{j+1}) \leftarrow B^+ \wedge \textstyle\bigwedge_{i=1}^{j} \beta(s_i) \quad j \in \{0,\ldots,k-1\} \tag{16}$$

where we define, for $v = \{\!\!\{t\}\!\!\}$, $\alpha(v) = get\_su(t, c_\emptyset)$ and $\beta(v) = su(t, c_\emptyset, V_v)$; and, for $v = s \Cup u$, $\alpha(v) = get\_u(V_s, V_u)$ and $\beta(v) = u(V_s, V_u, V_v)$.

Example 1. Consider the DLP(S) rule $r : p(S) \leftarrow q(S,x) \wedge r(S \Cup \{\!\!\{x\}\!\!\}) \wedge \textbf{not}\, x \in S$. The required list of set terms is $s_1 = \{\!\!\{x\}\!\!\}, s_2 = S \Cup \{\!\!\{x\}\!\!\}$. We have $H' = \hat{p}(V_S)$, $B^+ = \hat{q}(V_S,x) \wedge \hat{r}(V_{S\Cup\{\!\{x\}\!\}})$, and $B' = B^+ \wedge \textbf{not}\, in(x,V_S)$. Now $\mathrm{dlp^f}(r)$ consists of the rules:

$$H' \leftarrow B' \wedge su(x,c_\emptyset,V_{\{\!\{x\}\!\}}) \wedge u(V_S, V_{\{\!\{x\}\!\}}, V_{S\Cup\{\!\{x\}\!\}})$$
$$get\_su(x,c_\emptyset) \leftarrow B^+$$
$$get\_u(V_S, V_{\{\!\{x\}\!\}}) \leftarrow B^+ \wedge su(x,c_\emptyset,V_{\{\!\{x\}\!\}})$$

Note how B^+ is used to ensure that the auxiliary rules that define only some of the variables V_s are safe in the sense of Definition 1 (1).

As Example 1 suggests, we could sometimes use get_su and su instead of get_u and u. This optimisation can be useful in practice since it may make rules (11)–(12) obsolete, but we omit it from our formal description for simplicity.

The transformation needs one more ingredient, since our set representations are not unique. For example, $f_\cup(a, f_\cup(b, c_\emptyset))$ and $f_\cup(b, f_\cup(a, c_\emptyset))$ both represent $\{a,b\}$. We therefore explicitly compute equality of sets as follows:

$$sub(c_\emptyset, c_\emptyset) \leftarrow \tag{17}$$
$$sub(c_\emptyset, S) \leftarrow in(X,S) \tag{18}$$
$$sub(T,U) \leftarrow su(X,S,T) \wedge sub(S,U) \wedge in(X,U) \tag{19}$$
$$eq(S,T) \leftarrow sub(S,T) \wedge sub(T,S) \tag{20}$$

Definition 2. Given a DLP(S) program P, the DLP^f program $\mathrm{dlp^f}(P)$ consists of the following rules:

- the rules (7)–(12) and (17)–(20);
- for every predicate $\hat{p}$ in P of arity n, the rules

  $$\hat{p}(X_1,\cdots,X_n)[X_i/Y] \leftarrow \hat{p}(X_1,\cdots,X_n) \wedge eq(X_i,Y)$$

  where $[X_i/Y]$ is the substitution replacing X_i by Y;
- for every rule $r \in P$, the rules $\mathrm{dlp^f}(r)$.

The next result is a pre-condition for the practical utility of our translation. It can be shown directly but also follows from Theorem 5 below.

Theorem 2. *All stable models of* $\mathrm{dlp^f}(P)$ *are finite.*

We can convert ground terms s that contain $c_\emptyset$ or $f_\cup$ to canonical set terms $[s] \in \mathbf{set}(P)$ in the obvious way: $[c_\emptyset] = \emptyset$ and $[f_\cup(t, s)]$ is the representation of $\{t\} \cup [s]$ in $\mathbf{set}(P)$. Moreover, let $[t] = t$ for other terms (not representing sets). Then, for every stable model $\mathcal{J}$ of $\mathrm{dlp^f}(P)$, we find an interpretation $[\mathcal{J}] = \{p([t_1], \cdots, [t_n]) \mid \hat{p}(t_1, \cdots, t_n) \in \mathcal{J}\}$.

Theorem 3. *The set of stable models of a DLP(S) program P is exactly the set $\{[\mathcal{J}] \mid \mathcal{J}$ is a stable model of $\mathrm{dlp^f}(P)\}$.*

4 Lazy Grounding

A popular approach towards reasoning in ASP is "ground & solve", where one first computes a set of ground instances of rules that are then turned into a propositional logic problem to which an ASP solver can be applied [Gebser *et al.*, 2018; Faber, 2020]. To ensure that the grounding stage produces enough rule instances, it is common to disregard negative body literals when applying rules during grounding. Optimisations can further reduce the number of rule applications, e.g., using predicate dependencies to compute a *stratification*. A classical algorithm that combines several such ideas was proposed by Calimeri *et al.* [2008], who defined $\mathcal{FG}$ as the class of all DLP^f programs on which their grounding is finite. However, although our programs $\mathrm{dlp^f}(P)$ have finite stable models (Theorem 2), such classical grounding approaches will usually not work for them:

Proposition 4. *If P contains a fact $p(\emptyset)$ and a rule $p(S \cup \{a\}) \leftarrow p(S)$, then $\mathrm{dlp^f}(P) \notin \mathcal{FG}$.*

Indeed, it can be verified in practice that grounders such as *gringo* and *iDLV* do not terminate on $\mathrm{dlp^f}(P)$ for programs P as in Proposition 4. A challenge that grounders are facing in this case is that negation in the rules (7)–(10) of $\mathrm{dlp^f}(P)$ is not stratified. To avoid groundings that are too small, most classical grounders will therefore ignore the negated literal in rule (7), which leads to a program that has no finite model.

Fortunately, this problem can be avoided by *lazy grounding* approaches, which interleave grounding and solving to produce rule instances only when relevant to the search for a stable model. Notable systems of this type include *Alpha* [Weinzierl, 2017; Taupe *et al.*, 2019] and *ASPeRiX* [Lefèvre *et al.*, 2017]. We analyse the former to show that such approaches are applicable to our task:

Theorem 5. *The algorithm of Alpha terminates on every program of the form $\mathrm{dlp^f}(P)$.*

Alpha therefore produces a complete set of (necessarily finite) stable models of $\mathrm{dlp^f}(P)$, which correspond to the stable models of P by Theorem 3. The essence of our proof of this claim is the fact that Alpha eagerly applies a form of *unit propagation* where it exhaustively applies rules (8)–(10) before considering further *choice points* obtained by grounding rule (7). Since the propagation will derive facts for *in*, rule (7) is only potentially applicable to values of X that are not in the set represented by S, and sets can only be enlarged a finite number of times before no such elements are left.

5 Grounding with Existential Rules

Instead of relying on lazy grounding, an alternative approach towards reasoning in DLP(S) is to develop a set-aware grounder whose output can be combined with existing solvers. Since classical grounders are mostly Datalog engines (with some support for stratified negation), we essentially need an engine for Datalog(S), the extension of Datalog with sets. To obtain this, we follow an approach by Carral *et al.* [2019b] who used a reasoning algorithm from databases (the *standard chase*) to simulate Datalog(S) reasoning, and we argue that it can safely be combined with stratified negation, which is not immediate in this context.

We first explain grounding based on Datalog(S) with stratified negation. Given a DLP(S) program P, a *stratification* $s : P \to \mathbb{N}$ maps rules to natural numbers such that, for all pairs of rules $r_1, r_2 \in P$ of form $r_i : H_i \leftarrow B_i^+ \wedge B_i^- \in P$, and all $p \in \mathrm{preds}(H_2)$: (1) if $p \in \mathrm{preds}(B_1^+)$, then $s(r_1) \geq s(r_2)$, and (2) if $p \in \mathrm{preds}(B_1^-)$, then $s(r_1) > s(r_2)$. This induces a partitioning $P_1, \ldots, P_n$ of P such that $P_i = \{r \in P \mid s(r) = i\}$. A program P is *stratified* if it has a stratification.

Now for any program P, let $P^! \subseteq P$ be the largest disjunction-free, stratified subset of P such that predicates in heads of $P \setminus P^!$ do not occur in $P^!$.[1] A predicate is *certain* if it occurs only in $P^!$. The Datalog(S) program $\mathrm{grnd}(P)$ contains all facts in P and, for each non-fact rule $H \leftarrow B^+ \wedge B^- \in P$, the rules

$$\mathit{rule}_r(\boldsymbol{X}) \leftarrow B^+ \wedge B^! \qquad (21)$$

$$\textstyle\bigwedge_{\alpha \in H} \alpha \leftarrow \mathit{rule}_r(\boldsymbol{X}) \qquad (22)$$

where rule_r is a dedicated fresh predicate for r, $\boldsymbol{X}$ is a list of all variables in r, and $B^! = \bigwedge\{\mathbf{not}\,\alpha \in B^- \mid$ the predicate of α is certain$\}$. For a ground atom of the form $\mathit{rule}_r(\boldsymbol{c})$, the ground rule $r[\boldsymbol{c}]$ is obtained by applying the substitution $\boldsymbol{X} \mapsto \boldsymbol{c}$ to r. We get a simple but correct grounding:

Proposition 6. *For any DLP(S) program P, $\mathrm{grnd}(P)$ is a stratified Datalog(S) program. If $\mathcal{I}$ is the (unique) stable model of $\mathrm{grnd}(P)$, then the stable models of $\{r[\boldsymbol{c}] \mid \mathit{rule}_r(\boldsymbol{c}) \in \mathcal{I}\}$ are exactly the stable models of P.*

To compute the stable model of $\mathrm{grnd}(P)$, we simulate reasoning in stratified Datalog(S) by generalising an approach for *existential rules* (also know as *tuple-generating dependencies*), which we extend by negation (which will be stratified in the same sense as defined above). Such rules have the form $\exists \boldsymbol{Y}.H \leftarrow B^+ \wedge B^-$, where H and B^+ are conjunctions of atoms, B^- is a conjunction of negated atoms, $\boldsymbol{Y}$ is a list of existentially quantified variables, and all other (implicitly universally quantified) variables also occur in B^+ (*safety*). Existential quantifiers may lead to new domain elements, represented by *named nulls*, which play the role of anonymous constants. We can *apply* an existential rule $\exists \boldsymbol{Y}.H \leftarrow B^+ \wedge B^-$ to a set $\mathcal{I}$ of facts (using constants and possibly nulls) if (1) there is a substitution θ with $B^+\theta \subseteq \mathcal{I}$ and $B^- \cap \mathcal{I} = \emptyset$, and (2) the rule is not already satisfied under θ, i.e., $\mathcal{I} \not\models \exists \boldsymbol{Y}.(H\theta)$. If this holds, we apply the rule by extending $\mathcal{I}$ with $H\theta'$, where θ' is an extension of θ that maps each $Y \in \boldsymbol{Y}$ to a fresh null.

For existential rules without negation, a *(standard) chase* is a possibly infinite set $\mathcal{I}$ obtained by exhaustive, fair application of rules. For a set of existential rules with negation

[1]This can be viewed as a kind of split program in the sense of Lifschitz and Turner [1994].

that has a stratification $P_1, \ldots, P_n$, the *stratified chase* $\mathcal{I}$ is $\bigcup_{i=1}^{n} \mathcal{I}_i$, where $\mathcal{I}_0 = \emptyset$ and $\mathcal{I}_i$ $(1 \leq i \leq n)$ is the possibly infinite set obtained by exhaustive, fair application of rules P_i to $\mathcal{I}_{i-1}$. Note that this only leads to a practical (and overall fair) procedure if each $\mathcal{I}_i$ is finite, which may depend on the chosen order of rule applications.[2] We follow Carral *et al.* [2019b] and require that, within each stratum P_i, rules without existential variables are applied first (the so-called *Datalog-first chase*). This assumption simplifies presentation and is justified for the system used in our evaluation.

Example 2. Contrary to expectations in normal logic programming, the stratified chase does not yield a unique result, not even up to homomorphic renaming of nulls. Consider the fact $q(a)$ and the rules

$$\begin{aligned} r_1: &\quad r(X,X) \leftarrow q(X) \\ r_2: &\quad \exists V.r(X,V) \leftarrow q(X) \\ r_3: &\quad e(Y,Y) \leftarrow r(X,Y) \\ r_4: &\quad p() \leftarrow r(X,Y) \wedge r(X,Z) \wedge \textbf{not}\, e(Y,Z) \end{aligned}$$

A stratification s must satisfy $s(q(a)) \leq s(r_1) \leq s(r_3)$, $s(q(a)) \leq s(r_2) \leq s(r_3)$, and $s(r_3) < s(r_4)$. A valid partitioning would be $\{q(a), r_1\}, \{r_2, r_3\}, \{r_4\}$, which yields a stratified chase $\{q(a), r(a,a), e(a,a)\}$. Note that r_2 is not applicable since $r(a,a)$ is already derived before r_2 is considered. However, if we consider the partition $\{q(a), r_2\}, \{r_1, r_3\}, \{r_4\}$, which is also a valid stratification, then the stratified chase is $\{q(a), r(a,n), r(a,a), e(a,a), e(n,n), p()\}$, where n is a named null introduced when applying r_2 in the first stratum (as it is not satisfied at this point).

Now we can simulate sets in a similar way as in Definition 2, but using existential rules and nulls instead of DLP^f rules and function terms. To this end, we replace rule (7) by

$$\exists Y.su(X,S,Y) \leftarrow get\_su(X,S) \tag{23}$$

while all other auxiliary rules remain as before. The translation of Datalog(S) rules to existential rules likewise remains the same as in Definition 2. This produces existential rules with negation if the input rules contain negation but no disjunction. However, stratification can be lost if auxiliary predicates like get_su are needed in every stratum. To address this, each stratum P_i will use distinct copies of each auxiliary rule, using indexed predicates such as su_i and get_su_i. The resulting set of existential rules with negation is denoted $\mathrm{nex}(P_i)$, and for Datalog(S) program P with negation and stratification $P_1, \ldots, P_n$, we define $\mathrm{nex}(P) = \mathrm{nex}(P_1) \cup \bigcup_{i=2}^{n} \mathrm{nex}(P_i) \cup \{su_i(X,S,U) \leftarrow su_{i-1}(X,S,U)\}$. Then $\mathrm{nex}(P)$ is stratified. Since every null n that is introduced by a chase over $\mathrm{nex}(P)$ first appears in a fact $su(t,s,n)$, we can extend the notation introduced before Theorem 3 and associate n with the set $[n] \coloneqq [f_\cup(t,s)]$.

Theorem 7. *If P is a stratified Datalog(S) program with negation, then every stratified, Datalog-first chase $\mathcal{I}$ of $\mathrm{nex}(P)$ is finite and $[\mathcal{I}] \coloneqq \{p([t_1], \cdots, [t_n]) \mid \hat{p}(t_1, \cdots, t_n) \in \mathcal{I}\}$ is the unique stable model of P.*

[2]This is typical for the standard chase; Krötzsch *et al.* [2019] give an introductory discussion.

$$\begin{aligned} same(C,C) &\leftarrow class(C) \\ ind(A,C) &\leftarrow sc(A,B) \wedge sc(B,C) \wedge \textbf{not}\, same(A,B) \\ &\qquad \wedge \textbf{not}\, same(B,C) \\ sc^{-}(A,C) &\leftarrow sc(A,C) \wedge \textbf{not}\, ind(A,C) \\ &\qquad \wedge \textbf{not}\, same(A,C) \end{aligned}$$

Figure 2: Rules for transitive reduction

$$\begin{aligned} same(C,C) &\leftarrow class(C) \\ in\_chain(C) &\leftarrow class(C) \wedge \textbf{not}\, out\_chain(C) \\ out\_chain(C) &\leftarrow class(C) \wedge \textbf{not}\, in\_chain(C) \\ comp(C) &\leftarrow class(C) \wedge sc(C,D) \wedge in\_chain(D) \\ &\qquad \wedge \textbf{not}\, same(C,D) \\ comp(D) &\leftarrow class(C) \wedge sc(C,D) \wedge in\_chain(C) \\ &\qquad \wedge \textbf{not}\, same(C,D) \\ &\leftarrow in\_chain(C) \wedge comp(C) \\ &\leftarrow out\_chain(C) \wedge \textbf{not}\, comp(C) \end{aligned}$$

Figure 3: Rules for maximal antichains

The previous result is remarkable in the light of Example 2, since it identifies a case where a classical stratification can safely be applied to existential rules with negation.

This completes our approach of using existential rule reasoners for DLP(S)-reasoning: for a DLP(S) program P, we conduct a stratified chase over $\mathrm{nex}(\mathrm{grnd}(P))$ to infer facts of the form $rule_r(\boldsymbol{t})$ and $in(c,n)$, from which we can construct ground instances of DLP(S) rules as in Proposition 6.

6 Evaluation

To evaluate the practical feasibility of our approaches, we developed prototypical implementations of the necessary procedures and combined them with existing reasoning engines to compute stable models. The first approach (LAZY) uses the transformation $\mathrm{dlp^f}(P)$ and the lazy ASP solver *Alpha*[3] [Weinzierl, 2017]. The second approach (EXRULES) implements grounding with existential rules using the *Rulewerk* library with the *VLog* reasoner[4] [Carral *et al.*, 2019a] to produce a grounded file in *aspif* format, to which we apply the ASP solver *clasp* v3.2.1 [Gebser *et al.*, 2007]. As a third scenario DLVCOMP, we use the native implementation of sets provided in *DLV-complex* [Calimeri *et al.*, 2009]. Our evaluation source code and data sets are available from https://github.com/knowsys/eval-2022-IJCAI-asp-with-sets.

As a challenging task for set-based reasoning, we use a rule-based reasoning method for the expressive description logic Horn-$\mathcal{ALC}$ by Carral *et al.* [2019b]. Given an ontology that is syntactically transformed into facts, the rules compute the inferred subclass relationships that follow from the ontology – a problem that is EXPTIME-complete in data complexity. Following Carral *et al.*, we omit auxiliary rules (11)–(20)

[3]https://github.com/alpha-asp/Alpha, *master*, 29 Aug 2021

[4]https://github.com/karmaresearch/vlog, *master*, 30 Aug 2021

ID	Name	#Classes	#scAx	#scInf
00668	vaccine	6,481	6,090	94,605
00368	bp	16,298	25,626	187,379
00371	bp×cell	17,810	27,171	198,683
00541	mf×anatomy	18,955	23,592	168,338
00375	bp×cell. comp.	27,490	44,949	352,738
00395	bp×anatomy	36,752	55,022	437,770
00533	mf×ChEBI	52,726	61,148	911,856
00477	Gazetteer	150,976	10,606	11,031

Table 1: Ontologies used in experiments with their names ("mf": molecular function; "bp": biological processes) and numbers of class names, stated subclass relations, and inferred subclass relations

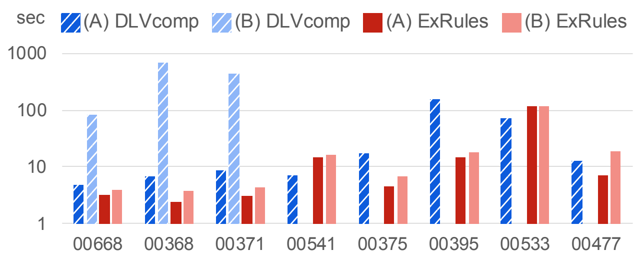

Figure 4: Evaluation results for DLVCOMP and EXRULES (time in sec, log-scale; empty columns indicate timeouts).

from $\mathrm{dlp}^{\mathrm{f}}(P)$ and the corresponding existential rules version, since they would not contribute new derivations here. The rules define a binary predicate *sc* that represents the inferred subclass hierarchy, and a unary predicate *class* that marks class names in the ontology.

On top of this basic classification task, we specified two tasks that require additional expressive power of DLP(S). First, we compute the *transitive reduction* of the class hierarchy, i.e., a directed graph that contains only direct subclass relations without the transitive bridges (Task A). The additional rules in Figure 2 were used for this task. While Task A is inherently non-monotonic, these rules are still stratified and lead to a unique stable model.

As a second task, we compute maximal antichains (i.e., sets of incomparable classes) in the class hierarchy (Task B). This task admits many solutions, each a stable model, and we asked systems to compute five stable models in this case. The rules we used are shown in Figure 3.

Experiments were executed with eight different ontologies from the Oxford Ontology Repository[5] as shown in Table 1. In each case, we deleted axioms that are not supported in the description logic Horn-$\mathcal{ALC}$ and normalised the remaining axioms as required by the classification rules [Carral *et al.*, 2019b]. Our final evaluation data sets are provided in the auxiliary material. Our evaluation computer is a mid-end server (Debian Linux 9.13; Intel Xeon CPU E5-2637v4@3.50GHz; 384GB RAM DDR4; 960GB SSD), though most experiments did not use more than 8GB of RAM. A timeout of 15min was used in all experiments. We report results of single runs, as we observed almost no time variations across runs.

[5]https://www.cs.ox.ac.uk/isg/ontologies/

The results of scenarios DLVCOMP and EXRULES on both tasks are shown in Figure 4, with missing values indicating a timeout. Scenario LAZY is omitted from the figure since it did not succeed in any of the tasks. The most difficult problem that we could solve with method LAZY was the basic classification (without any of the additional rules for tasks A or B) of the *vaccine* ontology 00668, which took 248sec (vs. <4sec needed in the other approaches).

Overall, we find that ASP with set terms can be used to solve complex problems on large real-world inputs. Both the native implementation in DLV-complex and our new existential-rule grounding performed well across a range of inputs. Contrary to our expectations, the lazy grounding approach was much less effective. Since Alpha solved instances of the base task, which already contains (7) as the only rule that can create infinitely many domain elements, we speculate that the algorithm behaves correctly in principle but is sometimes affected by additional optimisations or heuristics.

DLVCOMP and EXRULES often showed similar performance on Task A, though we can see some differences, most notably an order of magnitude advantage for VLog+clasp on 00395. Interestingly, EXRULES showed almost the same performance on Task B, whereas times for DLVCOMP increased such that only smaller problems could be solved. Again, this might be caused by a particular weakness of the implementation that is not conceptual. For EXRULES, most of the time was spent in the grounding phase, whereas solving the grounded files was relatively quick. Overall, we see promise in the performance of our existential grounding prototype, but the absolute run times should not be considered the main outcome of our experiments, given the diversity of the underlying systems in terms of maturity and recency.

7 Conclusions

Set terms are a natural and intuitive modelling construct, and clearly a useful addition to practical ASP tools. This seems to be the first work, however, that highlights their expressive advantages and puts them to concrete use on real-world problems that (due to their computational complexity) cannot be solved with plain ASP. Our work indicates that various existing ASP systems can feasibly be applied to these new kinds of tasks. The competitive performance of our prototypical existential-rule grounder encourage further research into combinations of existential rules and ASP.

We were surprised by the weak performance of lazy grounding approaches in our experiments, but we still consider them promising. Our programs can serve as benchmarks for further improving such implementations. An operational alternative to DLV-complex would to implement own set datatypes through extension points of modern ASP solvers, such as clingo's *@-terms* or DLV's *external atoms*. Finally, there is potential for investigating further uses of set terms in ASP for improving performance or readability. Candidate applications can be found, e.g., in the area of abstract argumentation, where ASP has been successfully applied [Gaggl *et al.*, 2015]. Our experiments with description logics (DLs) also suggest the use of sets for a native integration of DLs and ASP, e.g., in the style of *dl-programs* [Eiter *et al.*, 2004].

Acknowledgments

We thank an anonymous reviewer of IJCAI'22 for pointing out the potential utility of clingo *@-terms* and DLV *external atoms* for implementing complex-value extensions of ASP. This work was partly supported by *Deutsche Forschungsgemeinschaft* (DFG) in project 389792660 (TRR 248, Center for Perspicuous Systems), by the *Bundesministerium für Bildung und Forschung* (BMBF) in projects 01IS20056_NAVAS and ScaDS.AI (Center for Scalable Data Analytics and Artificial Intelligence), and by the Center for Advancing Electronics Dresden (cfaed).

References

[Bonatti, 2004] Piero A. Bonatti. Reasoning with infinite stable models. *Artif. Intell.*, 156(1):75–111, 2004.

[Brewka *et al.*, 2011] Gerhard Brewka, Thomas Eiter, and Miroslaw Truszczynski. Answer set programming at a glance. *Commun. ACM*, 54(12):92–103, 2011.

[Calimeri *et al.*, 2008] Francesco Calimeri, Susanna Cozza, Giovambattista Ianni, and Nicola Leone. Computable functions in ASP: theory and implementation. In *Proc. 24th Int. Conf. on Logic Programming (ICLP'08)*, volume 5366 of *LNCS*, pages 407–424. Springer, 2008.

[Calimeri *et al.*, 2009] Francesco Calimeri, Susanna Cozza, Giovambattista Ianni, and Nicola Leone. An ASP system with functions, lists, and sets. In *Proc. 10th Int. Conf. on Logic Programming and Nonmonotonic Reasoning (LPNMR'09)*, volume 5753 of *LNCS*, pages 483–489. Springer, 2009.

[Carral *et al.*, 2019a] David Carral, Irina Dragoste, Larry González, Ceriel Jacobs, Markus Krötzsch, and Jacopo Urbani. VLog: A rule engine for knowledge graphs. In *Proc. 18th Int. Semantic Web Conf. (ISWC'19, Part II)*, volume 11779 of *LNCS*, pages 19–35. Springer, 2019.

[Carral *et al.*, 2019b] David Carral, Irina Dragoste, Markus Krötzsch, and Christian Lewe. Chasing sets: How to use existential rules for expressive reasoning. In *Proc. 28th Int. Joint Conf. on Artificial Intelligence (IJCAI'19)*, pages 1624–1631. ijcai.org, 2019.

[Dantsin *et al.*, 2001] Evgeny Dantsin, Thomas Eiter, Georg Gottlob, and Andrei Voronkov. Complexity and expressive power of logic programming. *ACM Computing Surveys*, 33(3):374–425, 2001.

[Eiter and Simkus, 2010] Thomas Eiter and Mantas Simkus. FDNC: decidable nonmonotonic disjunctive logic programs with function symbols. *ACM Trans. Comput. Log.*, 11(2):14:1–14:50, 2010.

[Eiter *et al.*, 2004] Thomas Eiter, Thomas Lukasiewicz, Roman Schindlauer, and Hans Tompits. Combining answer set programming with description logics for the Semantic Web. In *Proc. 9th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'04)*, pages 141–151. AAAI Press, 2004.

[Faber, 2020] Wolfgang Faber. An introduction to answer set programming and some of its extensions. In *Tutorial Lectures 16th Int. Summer School on Reasoning Web. Declarative Artificial Intelligence*, volume 12258 of *LNCS*, pages 149–185. Springer, 2020.

[Gaggl *et al.*, 2015] Sarah Alice Gaggl, Norbert Manthey, Alessandro Ronca, Johannes Peter Wallner, and Stefan Woltran. Improved answer-set programming encodings for abstract argumentation. *Theory Pract. Log. Program.*, 15(4-5):434–448, 2015.

[Gebser *et al.*, 2007] Martin Gebser, Benjamin Kaufmann, André Neumann, and Torsten Schaub. clasp: A conflict-driven answer set solver. In *LPNMR*, volume 4483 of *Lecture Notes in Computer Science*, pages 260–265. Springer, 2007.

[Gebser *et al.*, 2018] Martin Gebser, Nicola Leone, Marco Maratea, Simona Perri, Francesco Ricca, and Torsten Schaub. Evaluation techniques and systems for answer set programming: a survey. In *Proc. 27th Int. Conf. on Artificial Intelligence, (IJCAI'18)*, pages 5450–5456. ijcai.org, 2018.

[Krötzsch *et al.*, 2019] Markus Krötzsch, Maximilian Marx, and Sebastian Rudolph. The power of the terminating chase. In *Proc. 22nd Int. Conf. on Database Theory (ICDT'19)*, volume 127 of *LIPIcs*, pages 3:1–3:17. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2019.

[Lefèvre *et al.*, 2017] Claire Lefèvre, Christopher Béatrix, Igor Stéphan, and Laurent Garcia. Asperix, a first-order forward chaining approach for answer set computing. *Theory Pract. Log. Program.*, 17(3):266–310, 2017.

[Lifschitz and Turner, 1994] Vladimir Lifschitz and Hudson Turner. Splitting a logic program. In *Proc. 11th Int. Conf. on Logic Programming (ICLP'94)*, pages 23–37. MIT Press, 1994.

[Lifschitz, 2019] Vladimir Lifschitz. *Answer Set Programming*. Springer, 2019.

[Ortiz *et al.*, 2010] Magdalena Ortiz, Sebastian Rudolph, and Mantas Simkus. Worst-case optimal reasoning for the Horn-DL fragments of OWL 1 and 2. In *Proc. 12th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'10)*, pages 269–279. AAAI Press, 2010.

[Taupe *et al.*, 2019] Richard Taupe, Antonius Weinzierl, and Gerhard Friedrich. Degrees of laziness in grounding – effects of lazy-grounding strategies on ASP solving. In *Proc. 15th Int. Conf. on Logic Programming and Nonmonotonic Reasoning (LPNMR'19)*, volume 11481 of *LNCS*, pages 298–311. Springer, 2019.

[Weinzierl, 2017] Antonius Weinzierl. Blending lazy-grounding and CDNL search for answer-set solving. In *Proc. 14th Int. Conf. on Logic Programming and Nonmonotonic Reasoning (LPNMR'17)*, volume 10377 of *LNCS*, pages 191–204. Springer, 2017.

A Proofs for Section 3

The claimed finiteness of stable models of $\mathrm{dlp}^{\mathrm{f}}(P)$ can be obtained from Theorem 5 together with the correctness of the Alpha algorithm. However, we prefer an alternative route that avoids the dependency on the complicated Alpha approach and its correctness.

Theorem 2. *All stable models of* $\mathrm{dlp}^{\mathrm{f}}(P)$ *are finite.*

Proof. Suppose for a contradiction that $\mathrm{dlp}^{\mathrm{f}}(P)$ has an infinite stable model $\mathcal{I}$. Then $\mathcal{I}$ must contain infinitely many distinct ground terms. Since $f_\cup$ is the only function symbol, $\mathcal{I}$ contains arbitrarily large terms using only $f_\cup$ and constants from the input.

Rule (7) is the only rule that can be used to infer the existence of $f_\cup$-terms in a minimal model. It is easy to verify that, whenever (7) is applicable in a minimal model, variable X is mapped to a constant: $f_\cup$-terms can only occur in set positions, which are kept separate from other positions in all rules. Hence, we conclude that $\mathcal{I}$ contains arbitrarily large terms of the form $f_\cup(c_1, f_\cup(c_2, \ldots f_\cup(c_n, c_\emptyset)\ldots))$. Since the number of distinct constants c_i is finite, there must be a term t of this form such that $c_1 = c_k$ for some $k \in \{2, \ldots, n\}$.

We show that this implies $\mathit{in}(c_1, t) \in \mathcal{I}$. Let $t_{n+1} = c_\emptyset$ and $t_i = f_\cup(c_i, t_{i+1})$ for all $i \in \{1, \ldots, n\}$. Since all such terms are inferred by (7), we have $\mathit{su}(c_i, t_{i+1}, t_i) \in \mathcal{I}$ for all $i \in \{1, \ldots, n\}$. For $i = k$, we can use rules (8) and (10) (whose ground instances occur in any reduct $\mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{I}}$), to get $\mathit{in}(c_k, t_k) = \mathit{in}(c_1, t_k) \in \mathcal{I}$. Similarly, from $\mathit{su}(c_{k-1}, t_k, t_{k-1}) \in \mathcal{I}$, we get $\mathit{in}(c_k, t_{k-1}) = \mathit{in}(c_1, t_{k-1}) \in \mathcal{I}$ by (9) and (10). Therefore, by induction, we find $\mathit{in}(c_1, t_2) \in \mathcal{I}$. But then, no ground instance of (7) that has $X \mapsto c_1$ and $S \mapsto t_2$ occurs in $\mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{I}}$, and therefore no rule in $\mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{I}}$ mentions t_1 in its head. This contradicts the assumption that t_1 occurs in a minimal model of $\mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{I}}$. □

For establishing Theorem 3, we first establish the basic correctness of our set simulation using function symbols.

Lemma 8. *Let $\mathcal{J}$ be a minimal model of some reduct $\mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{I}}$, and let s, s_1, s_2, t be ground terms.*

(1) If $\mathcal{J} \models \mathit{in}(t, s)$ then $t \in [s]$.

(2) If $\mathcal{J} \models \mathit{su}(t, s_1, s_2)$ then $[s_2] = [s_1] \cup \{t\}$.

(3) If $\mathcal{J} \models u(s_1, s_2, s)$ then $[s_1] \cup [s_2] = [s]$.

(4) If $\mathcal{J} \models \mathit{sub}(s_1, s_2)$ then $[s_1] \subseteq [s_2]$.

(5) If $\mathcal{J} \models \mathit{eq}(s_1, s_2)$ then $[s_1] = [s_2]$.

Here we use the usual set operations and relations on canonical set terms with the intuitive interpretation.

Proof. Since $\mathcal{J}$ is a minimal model of $\mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{I}}$, all atoms that are satisfied are entailed by some (ground) rule in $\mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{I}}$, and one can establish the claims by an easy inductive argument that shows that the claims must hold true for a rule head whenever they are true for its body.

For example, suppose that $\mathcal{J} \models u(s_1, s_2, s)$ for item (3) was entailed by rule (12), where $s_2 = f_\cup(u, s_2')$ (hence $[s_2] = [s_2'] \cup \{u\}$ by the definition of $[\cdot]$) and the variable $\acute{S}$ was grounded by a term $\acute{s}$. Then, by the hypothesis and the required body atoms, we find that $[\acute{s}] = [s_1] \cup \{u\}$ and $[\acute{s}] \cup [s_2'] = [s]$, from which $[s_1] \cup [s_2] = [s_1] \cup [s_2'] \cup \{u\} = [\acute{s}] \cup [s_2'] = [s]$ follows. The other cases are similar. □

The converse of Lemma 8 also holds, though restricted to function terms for which the respective relationships are relevant in the program.

Lemma 9. *Let $\mathcal{J}$ be a minimal model of some reduct $\mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{I}}$, and let s, s_1, s_2, t be ground terms that occur in $\mathcal{J}$.*

(1) If $t \in [s]$ then $\mathcal{J} \models \mathit{in}(t, s)$.

(2) If $\mathcal{J} \models \mathit{get\_su}(t, s_1)$ then $\mathcal{J} \models \mathit{su}(t, s_1, s')$ for some s' with $[s'] = [s_1] \cup \{t\}$.

(3) If $\mathcal{J} \models \mathit{get\_u}(s_1, s_2)$ then $\mathcal{J} \models u(s_1, s_2, s')$ for some s' with $[s_1] \cup [s_2] = [s']$.

(4) If $[s_1] \subseteq [s_2]$ then $\mathcal{J} \models \mathit{sub}(s_1, s_2)$.

(5) If $[s_1] = [s_2]$ then $\mathcal{J} \models \mathit{eq}(s_1, s_2)$.

Proof. Item (1) follows by a similar argument as in the proof of Theorem 2. Note that all relevant rules (8)–(10) are guaranteed to be in any reduct $\mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{I}}$.

Item (2) requires a case distinction. If $t \in [s_1]$, then we can take $s' = s_1$ and obtain $\mathcal{J} \models \mathit{su}(t, s_1, s_1)$ from rule (8) (since s_1 must have been introduced as a term by some use of rule (7)). If $t \notin [s_1]$, then $\mathcal{J} \not\models \mathit{in}(t, s)$ by Lemma 8, and hence rule (7) is applicable to produce the required $\mathcal{J} \models \mathit{su}(t, s_1, s')$. The case for item (3) is similar.

For item (4), we show the claim by induction on the size of $[s_1]$. For $[s_1] = \emptyset$, the claim follows from rules (17) and (18). If s_1 has the form $f_\cup(t, s_1')$ for some $u \notin [s_1']$, then it must have been introduced by (7), and we have $\mathcal{J} \models \mathit{su}(u, s_1', s_1)$. Since $[s_1'] \subseteq [s_2]$ and $[s_1']$ is smaller than $[s_1]$, we find $\mathcal{J} \models \mathit{sub}(s_1', s_2)$ by induction. Moreover, $u \in [s_1] \subseteq [s_2]$ implies $\mathcal{J} \models \mathit{in}(u, s_2)$ by item (1). Hence the claim follows by applying rule (19).

Item (5) is immediate from item (4) and rule (20). □

Theorem 3. *The set of stable models of a DLP(S) program P is exactly the set $\{[\mathcal{J}] \mid \mathcal{J}$ is a stable model of $\mathrm{dlp}^{\mathrm{f}}(P)\}$.*

Proof. Let $\mathcal{J}$ be a stable model of $\mathrm{dlp}^{\mathrm{f}}(P)$. We show that $[\mathcal{J}]$ is a minimal model of $P^{[\mathcal{J}]}$. Since $\mathcal{J}$ is a minimal model of $\mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{J}}$, there is a sequence $\mathcal{J}_0, \mathcal{J}_1, \ldots$ of interpretations such that $\mathcal{J}_0 = \emptyset$ and, for every $i > 0$, $\mathcal{J}_i = \mathcal{J}_{i-1} \cup \{\alpha\}$ where α is a ground atom that occurs in the head of a ground rule $H \leftarrow B \in \mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{J}}$ such that $B \subseteq \mathcal{J}_i$ (this condition always holds for rules that are facts).

An analogous sequence of rule applications from $P^{[\mathcal{J}]}$ shows that $[\mathcal{J}]$ is a minimal model of $P^{[\mathcal{J}]}$. We construct a sequence $\mathcal{I}_0, \mathcal{I}_1, \ldots$ that represents the derivations on $P^{[\mathcal{J}]}$, where we allow $\mathcal{I}_{i+1} = \mathcal{I}_i$ in cases where $\mathcal{J}_{i+1}$ differs from $\mathcal{J}_i$ in an atom that uses a predicate that is not of the form $\hat{p}$, to keep the indices synchronised for readability.

Initially, we have $\mathcal{I}_0 = \emptyset$. We proceed by induction to show that, for all $i > 0$, $\hat{p}(t_1, \ldots, t_n) \in \mathcal{J}_i$ implies that we can infer $p([t_1], \ldots, [t_n])$ by applying a rule in $P^{[\mathcal{J}]}$ to $\mathcal{I}_{i-1}$. Inference steps that produce atoms with predicate $\hat{p}$ must use (a reduct of) a rule of the form (15) or an equality rule from the second item of Definition 2. The latter can be ignored, since its body contains an atom that is equal to $p([t_1], \ldots, [t_n])$, where we apply Lemma 8 (5).

Therefore, let $\bar{r}' = \bar{H}' \leftarrow \bar{B^{+}}' \wedge \bar{C}$ be the ground rule in $\mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{J}}$ that was applied in step i, where $\bar{B^{+}}'$ is a conjunction of positive literals with predicates of the form $\hat{q}$, and $\bar{C}$ is a conjunction of ground instantiations of atoms of the form $\beta(v)$ as used in the definition of rule (15). Rule $\bar{r}'$ was obtained as the reduct of a ground instantiation of a rule $r' = H' \leftarrow B' \wedge \bigwedge_{i=1}^{k} \beta(s_i) \in \mathrm{dlp}^{\mathrm{f}}(P)$ of form (15). We use $\bar{B^{-}}'$ to denote the ground instance of the negative literals in B' that corresponds to the ground instantiation used for $\bar{r}'$ (by safety, all variables in negative literals occur in positive literals as well). Let $r = H \leftarrow B \in P$ denote the rule $r \in P$ that $r' \in \mathrm{dlp}^{\mathrm{f}}(P)$ was based on.

By the induction hypothesis, for every atom $\hat{q}(u_1, \ldots, u_m) \in \bar{B^{+}}'$, there is an atom $q([u_1], \ldots, [u_m]) \in \mathcal{I}_{i-1}$. Since all variables of r occur in positive literals of B, this induces a ground instance $\bar{r} = \bar{H} \leftarrow \bar{B}$ of r with respect to P. Suppose for a contradiction that some negative literal **not** $\hat{q}(u_1, \ldots, u_m) \in \bar{B^{-}}'$ such that the corresponding **not** $q([u_1], \ldots, [u_m]) \in \hat{B}$ occurs in positive form in $[\mathcal{J}]$. Then $q([u_1], \ldots, [u_m]) \in [\mathcal{J}]$ implies that $\hat{q}(u_1', \ldots, u_m') \in \mathcal{J}$ for some terms u_i' such that $[u_i'] = [u_i]$ for all $i \in \{1, \ldots, m\}$. By Lemma 9 (5) and the equality rules of Definition 2, we obtain $\hat{q}(u_1, \ldots, u_m) \in \mathcal{J}$, which implies that $\bar{r}' \notin \mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{J}}$ – contradiction.

Hence, negative literals of $\bar{r}$ do not occur in $[\mathcal{J}]$ and we find the positive rule $\bar{H} \leftarrow \bar{B}^{+}$ in $P^{[\mathcal{J}]}$. Since the positive literals $\bar{B}^{+}$ occur in $\mathcal{I}_{i-1}$ by the hypothesis, the rule is applicable and can be used to produce $\mathcal{I}_i$ such that $p([t_1], \ldots, [t_n]) \in \mathcal{I}_i$ as required.

This finishes a construction of a minimal sequence of inferences $\mathcal{I}_0, \mathcal{I}_1, \ldots$ with $\bigcup_{i \geq 0} \mathcal{I}_i = [\mathcal{J}]$, where minimality follows from the minimality of $\mathcal{J}$. It remains to show that this is a model of $P^{[\mathcal{J}]}$, i.e., that all rules in $P^{[\mathcal{J}]}$ whose body is satisfied in $[\mathcal{J}]$ also have their heads satisfied. Indeed, suppose that a rule $\bar{r} \in P^{[\mathcal{J}]}$ is not satisfied in $[\mathcal{J}]$. Then, using the same notation as before, we can find a rule $\bar{r}' \in \mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{J}}$ of form (15). The sought after contradiction follows if this rule is applicable, which hinges upon the truth of the grounded version of the second part of the body of the form $\bigwedge_{i=1}^{k} \beta(s_i)$. This can be obtained from the rules (16). Indeed, for $j = 0$, the corresponding grounding of (16) is applicable, since it only requires the positive body atoms to be satisfied. For remaining values $j \in \{1, \ldots, k-1\}$, the rules become applicable since every entailment of the form $\mathcal{J} \models \alpha(v)$ for some ground term v leads to the entailment of the corresponding $\mathcal{J} \models \beta(v)$ by Lemma 9 (2) and (3). We therefore conclude that $[\mathcal{J}]$ is a stable model of P.

For the converse, let $\mathcal{I}$ be a stable model of P. We can show that here is a suitable minimal model $\mathcal{J}$ of $\mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{I}}$ such that $[\mathcal{J}] = \mathcal{I}$. The proof proceeds as above by tracing the derivation of $\mathcal{I}$ in $P^{\mathcal{I}}$, which allows us to select a suitable representative term s for every set $[s]$ that occurs in atoms of $\mathcal{I}$. Importantly, not every representation of $[s]$ may occur in $\mathcal{J}$. The remaining arguments are analogous to the above. $\square$

B Proofs for Section 4

We start by proving Proposition 4, for which we will also need to recall some essential definitions regarding $\mathcal{FG}$.

Definition 3. [Calimeri *et al.*, 2008] Given a rule r and a set S of ground atoms, an *S-restricted* instance of r is a ground instance r' of r such that $B^{+}(r') \subseteq S$. The set of all S-restricted instances of a program P is denoted as $Inst_P(S)$.

Lemma 10. *Let* $S = \{\hat{p}(\emptyset)\}$ *and* $P = \{p(\emptyset); p(S \cup \{a\}) \leftarrow p(S)\}$. *Then* $Inst_{\mathrm{dlp}^{\mathrm{f}}(P)}(S)$ *is infinite.*

Proof. It is easy to see that the rules in $\mathrm{dlp}^{\mathrm{f}}(P)$ lead to infinite S-restricted instances, as the following rule of $\mathrm{dlp}^{\mathrm{f}}(P)$

$$\hat{p}(V_{S \cup \{a\}}) \leftarrow \hat{p}(V_S), su(a, c_\emptyset, V_{\{a\}}), u(V_S, V_{\{a\}}, V_{S \cup \{a\}})$$

together with the rules (7)–(12) would have infinitely many ground rules with $B^{+} \subseteq S$ of the form

$$\begin{aligned} \hat{p}(f_\cup(a, c_\emptyset)) \leftarrow{} & \hat{p}(c_\emptyset), su(a, c_\emptyset, f_\cup(a, c_\emptyset)), \\ & u(c_\emptyset, f_\cup(a, c_\emptyset), f_\cup(a, c_\emptyset)) \\ \hat{p}(f_\cup(a, f_\cup(a, c_\emptyset))) \leftarrow{} & \hat{p}(f_\cup(a, c_\emptyset)), su(a, c_\emptyset, f_\cup(a, c_\emptyset)), \\ & u(f_\cup(a, c_\emptyset), f_\cup(a, c_\emptyset), f_\cup(a, f_\cup(a, c_\emptyset))) \\ & \vdots \end{aligned}$$

$\square$

Consider the SCCs of the dependency graph of P. The *component graph* of P, denoted by $\mathcal{G}^{C}(P)$, is a directed labelled graph having a node for each SCC of the dependency graph and: (i) an edge (B, A), labelled "+", if there is a rule r in P s.t. there is a predicate $q \in A$ occurring in the head of r and a predicate $p \in B$ occurring in the positive body of r; (ii) an edge (B, A), labelled "-", if there is a rule r in P s.t. there is a predicate $q \in A$ occurring in the head of r and a predicate $p \in B$ occurring in the negative body of r, and there is no edge (B, A), with label "+". Self-cycles are not considered. A component ordering for a given program P is a total ordering $\langle C_1, \ldots C_n, \rangle$ of all components of P s.t., for any C_i, C_j with $i < j$, both conditions hold: (i) there is no path labelled with "+" in the component graph from C_j to C_i; (ii) if there is a path labelled with "-" in the component graph from C_j to C_i, then there must also be a path labelled with "-" in the component graph from C_i to C_j. The *module* of a component C_i, denoted by $P(C_i)$, is the set of all rules r where some predicate $p \in C_i$ appears in $H(r)$.

Definition 4. [Calimeri *et al.*, 2008] Given a program P, a component ordering $\langle C_1, \ldots C_n, \rangle$, a component C_i, the module $M = P(C_i)$, a set X of ground rules of M, and a set R of ground rules belonging only to facts of P or to modules of components C_j with $j < i$, let $\Phi_{M,R}(X)$ be the transformation defined as follows: $\Phi_{M,R}(X) = Simpl(Inst_M(Heads(R \cup X)), R)$.

Where $Simpl(S_i, R)$ is a simplification of a set of ground rules S_i as defined in [Calimeri *et al.*, 2008], which we omit here, as due to Lemma 10 the S-restricted instances are infinite, which is a sufficient condition to show Proposition 4. $Heads(R \cup X)$ returns all head atoms in $R \cup X$. The least fixed-point of $\Phi_{M,R}$ is denoted by $\Phi^{\infty}_{M,R}(\emptyset)$.

Definition 5. [Calimeri *et al.*, 2008] Given a program P and a component ordering $\gamma = \langle C_1, \ldots C_n, \rangle$ for P, the *intelligent instantiation* P^γ of P for γ is the last element S_n of the sequence s.t. S_0 contains all facts of P, $S_i = S_{i-1} \cup \Phi^{\infty}_{M_i, S_{i-1}}(\emptyset)$, where M_i is the program module $P(C_i)$.

Definition 6. [Calimeri *et al.*, 2008] A program P is *finitely-ground* ($\mathcal{FG}$) if P^γ is finite, for every component ordering γ for P.

Proposition 4. *If P contains a fact $p(\emptyset)$ and a rule $p(S \cup \{a\}) \leftarrow p(S)$, then $\text{dlp}^{\text{f}}(P) \notin \mathcal{FG}$.*

Proof. To show that $\text{dlp}^{\text{f}}(P) \notin \mathcal{FG}$ we consider the program P that only contains the fact $p(\emptyset)$ and a rule $p(S \cup \{a\}) \leftarrow p(S)$, and no further rules or facts. $\text{dlp}^{\text{f}}(P)$ is obtained as given in Definition 2. The dependency graph of $\text{dlp}^{\text{f}}(P)$ consists of a single SCC C, thus the component ordering is trivially $\langle C \rangle$. As the definition of $\mathcal{FG}$ is based on the set of all S-restricted instances of a program and some additional simplifications, it follows from Lemma 10 that the intelligent instantiation of $\text{dlp}^{\text{f}}(P)$ is not finite, thus $\text{dlp}^{\text{f}}(P) \notin \mathcal{FG}$. □

Theorem 5. *The algorithm of Alpha terminates on every program of the form $\text{dlp}^{\text{f}}(P)$.*

Proof. We consider the Alpha algorithm as given in [Weinzierl, 2017]. To show termination of Alpha for any $\text{dlp}^{\text{f}}(P)$, we need to show that the rules that compute the sets don't lead to an infinite derivation of sets. The only rules that actually perform the set computations are the rules (7)-(14). The lazy grounder would always ground rule (7) before the other rules (8)-(10), as the positive part of rule (7) does not depend on rules (8)-(10), but they depend on rule (7) (for a partial assignment σ). Furthermore, only rule (7) contains negation, thus this is the only rule that will add choice nogoods in the lazy grounding step to the set of nogoods. These choice nogoods are the only parts in the Alpha algorithm where choices are performed. The rest of the computation is only performed by a repeated iteration of lazy grounding and unit-propagation.

In lazy grounding, a rule is grounded if the positive part of the rule is already fulfilled by the current assignment. Thus rule (7) will be grounded if there is a substitution σ that makes the atom $get\_su(X, S)$ true under σ, i.e. $\mathbf{T}get\_su(X, S)\sigma$ is contained in the current assignment. The necessary nogoods are added, in particular also the choice nogoods. Then, unit propagation is applied. Let us inspect the nogoods newly added for the assignment σ and rule (7). Either non of the choice nogoods is neither strongly- nor weakly-unit, thus nothing is propagated, or there might be $\mathbf{T}in(X, S)\sigma$ in the current assignment. Thus the respective choice nogood $\{\mathbf{F}cOff(r_7, \sigma), \mathbf{T}in(X, S)\sigma\}$ is weakly unit and $\mathbf{M}cOff(r_7, \sigma)$ is added to the assignment. This prevents that the body atom of this rule is selected in the respective choice part of the algorithm, thus the rule will not fire and the head of the rule will not be derived. This means that there is no way to add something to a set, that is already contained in the set. In the first case, where nothing from the choice nogoods was propagated, the respective body atom from the nogoods is added to the active choice points, and will be selected by the algorithm to be added to the assignment. Thus, the algorithm chooses to make the body of rule (7) true under the assignment σ and the decision level is increased. In case this choice leads to a conflict, Alpha would proceed as usual to resolve this.

□

C Proofs for Section 5

Proposition 6. *For any DLP(S) program P, $\text{grnd}(P)$ is a stratified Datalog(S) program. If $\mathcal{I}$ is the (unique) stable model of $\text{grnd}(P)$, then the stable models of $\{r[\boldsymbol{c}] \mid rule_r(\boldsymbol{c}) \in \mathcal{I}\}$ are exactly the stable models of P.*

We prove the parts of this result in the following two lemmas.

Lemma 11. *For any DLP(S) program P, $\text{grnd}(P)$ is a stratified Datalog(S) program.*

Proof. Every rule $r \in \text{grnd}(P)$ is of the form (21) or (22) for some corresponding rule $H \leftarrow B^+ \wedge B^- \in P$, which we denote by r^-. Now let s be a stratification of $P^\bullet$. The claimed stratification $\hat{s}$ of $\text{grnd}(P)$ is obtained by setting:

- $\hat{s}(r) = s(r^-)$ for every rule $r \in \text{grnd}(P)$ with $r^- \in P^\bullet$, and
- $\hat{s}(r) = m+1$ for every rule $r \in \text{grnd}(P)$ with $r^- \notin P^\bullet$, where m is the largest number that s assigns to any rule.

Then $\hat{s}$ is a stratification of $\text{grnd}(P)$, which we can verify by considering an arbitrary pair of rules r_1, r_2 of form $r_i : H_i \leftarrow B_i^+ \wedge B_i^- \in P$, and a predicate $p \in \text{preds}(H_2)$. We consider conditions (1) and (2) of stratifications as given in Section 5.

Condition (1) Assume $p \in \text{preds}(B_1^+)$. If $p = rule_{r_2}$, then $r_1^- = r_2^-$, where r_1 is of form (21) and r_2 of form (22). Thus $\hat{s}(r_1) = \hat{s}(r_2)$ by definition.

If p is a predicate from P and $r_2^- \in P \setminus P^\bullet$, then p does not occur in $P^\bullet$, hence $p \in \text{preds}(B_1^+)$ implies $r_1^- \in P \setminus P^\bullet$ and therefore $\hat{s}(r_1) = \hat{s}(r_2)$.

If p is a predicate from P and $r_2^- \in P^\bullet$, we consider two cases: (a) if $r_1^- \in P^\bullet$, then $\hat{s}(r_1) \geq \hat{s}(r_2)$ follows since s is a stratification of $P^\bullet$, and (b) if $r_1^- \in P \setminus P^\bullet$, then $\hat{s}(r_1) > \hat{s}(r_2)$ by definition.

Condition (2) Assume $p \in \text{preds}(B_1^-)$. Then p is a certain predicate from P and therefore occurs only in heads of rules $r_2^- \in P^\bullet$. If $r_1^- \in P^\bullet$, then $\hat{s}(r_1) > \hat{s}(r_2)$ follows again since s is a stratification. If $r_1^- \in P \setminus P^\bullet$, then $\hat{s}(r_1) > \hat{s}(r_2)$ by definition. □

Lemma 12. *For any DLP(S) program P, if $\mathcal{I}$ is the (unique) stable model of $\text{grnd}(P)$, then the stable models of $\{r[\boldsymbol{c}] \mid rule_r(\boldsymbol{c}) \in \mathcal{I}\}$ are exactly the stable models of P.*

Proof. The fact that the stable model of $\text{grnd}(P)$ is unique follows from Lemma 11. Let $G \coloneqq \{r[\boldsymbol{c}] \mid rule_r(\boldsymbol{c}) \in \mathcal{I}\}$, and note that $G = \text{ground}(G) \subseteq \text{ground}(P)$. Also note that all rules in $\text{ground}(P)$ are of the form $r[\boldsymbol{c}]$, even if they are not in G. We write $r[\boldsymbol{c}]^+$ for the corresponding rule with all negative body atoms removed, as it may occur in a reduct.

Property $(*)$ Consider an arbitrary interpretation $\mathcal{J}$ for atoms in G, such that $\mathcal{J}'$ is a minimal model of $G^{\mathcal{J}}$. We find that $\mathcal{J}' \subseteq \mathcal{I}$. Indeed, since $\mathcal{J}'$ is a minimal model of $G^{\mathcal{J}}$, every fact in $\mathcal{J}'$ is the consequence of a rule $r[\boldsymbol{c}]^+ \in G^{\mathcal{J}}$, stemming from some rule $r[\boldsymbol{c}] \in G$. Now $r[\boldsymbol{c}] \in G$ implies that $rule_r(\boldsymbol{c}) \in \mathcal{I}$, and therefore the head of $r[\boldsymbol{c}]$ is in $\mathcal{I}$ by rule (22).

Stable models of P agree with $\mathcal{I}$ on certain predicates Assume that $\mathcal{J}$ is a stable model of P. For every predicate p that is *certain*, and every fact $\alpha = p(\boldsymbol{d})$, we have $\alpha \in \mathcal{I}$ iff $\alpha \in \mathcal{J}$. We proceed by induction over the strata P_i of $P^\bullet$ (and corresponding stratum $\mathrm{grnd}(P_i)$ of $\mathrm{grnd}(P)$). The claim is immediate if α is a fact. Now assume that the claim holds for all facts α' that use a predicate that occurs only in facts or in rule heads of a stratum $< i$.

- Suppose for a contradiction that there is $\alpha \in \mathcal{I} \setminus \mathcal{J}$ that is derived in stratum i and assume w.l.o.g. that α is the first such fact in $\mathcal{I}$ to be derived in a bottom-up construction of the least model.[6] Then α was inferred by applying some rule r_2 of form (22) in $\mathrm{grnd}(P)$ for a body atom $rule_r(\boldsymbol{c})$, which was inferred from a rule r_1 of form (21). We have $r[\boldsymbol{c}] \in P$, where this rule has the same body as r_1. By our assumption on α, all positive body atoms of r_1 are in $\mathcal{J}$, and by the induction hypothesis, all negative body atoms (which must belong to a lower stratum) are not. Hence, $r[\boldsymbol{c}]^+ \in P^{\mathcal{J}}$ is applicable in $\mathcal{J}$, and entails α, contradicting our assumption.
- The case of $\alpha \in \mathcal{J} \setminus \mathcal{I}$ is analogous.

The induction covers all strata of $P^\bullet$. The remaining rules (and the top-most stratum of $\mathrm{grnd}(P)$) cannot entail facts for certain predicates, so the claim is shown.

Stable models of G agree with $\mathcal{I}$ on certain predicates Consider a stable model $\mathcal{J}$ of G. By $(*)$ above, it remains to show that $\alpha = p(\boldsymbol{d}) \in \mathcal{I}$ for a certain predicate p, then $\alpha \in \mathcal{J}$. We proceed by induction over the derivation of α in $\mathrm{grnd}(P)$. Clearly, if α is a fact, then $\alpha \in \mathcal{J}$. Now let α be inferred by applying a rule r_2 of form (22) in $\mathrm{grnd}(P)$ for a body atom $rule_r(\boldsymbol{c})$, which was inferred from a rule r_1 of form (21). Since $rule_r(\boldsymbol{c}) \in \mathcal{I}$, we have $r[\boldsymbol{c}] \in G$, where this rule has the same body as r_1. Note that $r[\boldsymbol{c}]$ is free of disjunctions since it contains a certain predicate in its head. By stratification, all positive predicates in the body of r_1 are certain, so by induction we find that the positive body is satisfied in $\mathcal{J}$. Moreover, since $\mathcal{J} \subseteq \mathcal{I}$ $(*)$, the negative body is satisfied in $\mathcal{J}$ too, and we find $r[\boldsymbol{c}]^+ \in G^{\mathcal{J}}$ and $\alpha \in \mathcal{J}$, as claimed.

Finishing the proof Now consider a stable model $\mathcal{J}_P$ of P. Since $G \subseteq \mathrm{ground}(P)$, we find $G^{\mathcal{J}_P} \subseteq P^{\mathcal{J}_P}$. Therefore, there is a minimal model $\mathcal{J}_G$ of $G^{\mathcal{J}_P}$ with $\mathcal{J}_G \subseteq \mathcal{J}_P$. We show that $\mathcal{J}_G = \mathcal{J}_P$ and that $\mathcal{J}_P$ therefore is a stable model of G. Suppose for a contradiction that there is $\alpha \in \mathcal{J}_P \setminus \mathcal{J}_G$. Without loss of generality, we can chose α that occurs in the head of a rule $r[\boldsymbol{c}]^+ \in P^{\mathcal{J}_P}$ such that:

(i) α is necessary to satisfy $r[\boldsymbol{c}]^+$, i.e., no other atom in the head of $r[\boldsymbol{c}]^+$ is satisfied in $\mathcal{J}_P$, and

(ii) α is the "first" such atom in $\mathcal{J}_P \setminus \mathcal{J}_G$, i.e., all positive body atoms of $r[\boldsymbol{c}]^+$ are in $\mathcal{J}_P \cap \mathcal{J}_G$.

Indeed, if no such α occurs, then $\mathcal{J}_P = \mathcal{J}_G$. By (ii) and $(*)$, the positive body of $r[\boldsymbol{c}]$ is satisfied by $\mathcal{I}$. No negative body atoms of $r[\boldsymbol{c}]$ are in $\mathcal{J}_P$, so in particular no negative body atoms for certain predicates are in $\mathcal{I}$, since $\mathcal{J}_P$ and $\mathcal{I}$ agree on certain-predicate atoms. Using similar reasoning as before, we find that $rule_r(\boldsymbol{c}) \in \mathcal{I}$ and hence $r[\boldsymbol{c}]^+ \in G^{\mathcal{J}_P}$ must be satisfied by $\mathcal{J}_G$. By (i) and $\mathcal{J}_G \subseteq \mathcal{J}_P$, all head atoms other than α are not in $\mathcal{J}_G$, so this requires that $\alpha \in \mathcal{J}_G$, which yields the required contradiction $\alpha \in \mathcal{J}_G$.

The argument for showing that every stable model $\mathcal{J}_G$ of G is also a stable model for P is analogous, with the main difference in the initial quantifiers (for every stable model $\mathcal{J}_G$ of G, we find some stable model $\mathcal{J}_P$ of P). Note that $\mathcal{J}_G \subseteq \mathcal{J}_P$ as above, so the remaining arguments are indeed the same, with the exception that we now use that $\mathcal{J}_G$ agrees with $\mathcal{I}$ on certain-predicate atoms to show that $r[\boldsymbol{c}]^+ \in P^{\mathcal{J}_G}$ implies $rule_r(\boldsymbol{c}) \in \mathcal{I}$. $\square$

To prepare our proof of Theorem 7 below, we first establish several lemmas below.

Lemma 13. *If P is a stratified Datalog(S) program with negation, then every stratified, Datalog-first chase $\mathcal{I}$ of $\mathrm{nex}(P)$ is finite.*

Proof. The claim follows from analogous arguments as given for Theorem 2, using nulls instead of function terms. Instead of arbitrarily large terms $f_\cup(c_1, f_\cup(c_2, \ldots f_\cup(c_m, c_\emptyset) \ldots))$, we now consider arbitrarily long chains of entailments $su(c_m, c_\emptyset, n_m), su(c_{m-1}, n_m, n_{m-1}), \ldots, su(c_1, n_2, n_1)$, where n_i are nulls. All other arguments are as before, where the use of a Datalog-first strategy ensures that the required (stratified copies of) rules (8)–(10) are applied before considering the creation of further nulls in (7). $\square$

Lemma 14. *If P is a stratified Datalog(S) program with negation, then it has a unique stable model that can be obtained applying rules bottom up and stratum-by-stratum, and which is independent of the chosen stratification.*

Proof. Indeed, similar results are well known for Datalog with negation (but without sets). To transfer them to stratified Datalog(S) with negation, it suffices to note that we can simulate sets as follows:

- Introduce a fresh constant for every possible set (there are only finitely many).
- Create all ground facts for predicates in and sub that are true for the given set-constants and normal constants.
- Add predicates to express unions, similar to our predicate u, which are similarly populated.
- Replace all set unions and singletons that occur in rules with fresh variables and add body atoms for u to axiomatise their relationship to their subterms.

[6]Note that a *minimal* model in this case is indeed the unique *least* model, since the strata of $P^\bullet$ do not contain disjunctions

It is not hard to see that this transformation leads to an (exponentially large) Datalog program with negation whose stable models correspond to those of the original program P. Moreover, the transformed program admits the same stratifications as the original, since there is a bijective correspondence between the non-fact rules. Hence the classical result that stratified Datalog with negation leads to a unique, stratification-independent stable model implies that the same is true for Datalog(S). □

The next result draws concrete benefit from the fact that $\mathrm{nex}(P)$ is very similar to $\mathrm{dlp}^{\mathrm{f}}(P)$.

Lemma 15. *If P is a stratified Datalog(S) program with negation, then for every stratified, Datalog-first chase $\mathcal{I}$ of $\mathrm{nex}(P)$, we can define an interpretation $\mathrm{dlp}^{\mathrm{f}}(\mathcal{I})$, which is a stable model of $\mathrm{dlp}^{\mathrm{f}}(P)$.*

Proof. First, we recursively associate a ground term $\bar{t}$ of $\mathrm{dlp}^{\mathrm{f}}(P)$ to every term t in $\mathcal{I}$. For all constants c (including $c_\emptyset$), we set $\bar{c} := c$. Every null n in $\mathcal{I}$ must be introduced by the unique existential quantifier of (a copy of) rule (23), which produces some fact $su_i(t, s, n)$ where i is the stratum that the rule was used in. We recursively define $\bar{n} := f_\cup(t, \bar{s})$. This definition is well since every null is only introduced exactly once.

By definition, this mapping is compatible with the representation of sets in the sense that $[t] = [\bar{t}]$. Moreover, $\bar{\cdot}$ is injective. Suppose for a contradiction that there are two distinct nulls n_1 and n_2 such that $\bar{n}_1 = \bar{n}_2$. Then we have two corresponding facts $su_i(t, s, n_1), su_j(t, s, n_2) \in \mathcal{I}$. Assume w.l.o.g. that $su_i(t, s, n_1)$ is the one that was derived first in the chase, and hence $i \leq j$. Using rules of the form $su_{k+1}(X, S, U) \leftarrow su_k(X, S, U)$, we get $su_j(t, s, n_1) \in \mathcal{I}$, which by the Datalog-first strategy must be inferred before considering applications of rule (7) in stratum j. But then $su_j(t, s, n_2)$ cannot be inferred, since the rule is already satisfied. The interpretation $\mathrm{dlp}^{\mathrm{f}}(\mathcal{I})$ will be defined over the elements $\{\bar{t} \mid t \text{ occurs in } \mathcal{I}\}$, so that we obtain a bijection of the two domains.

Now for every atom $p(t_1, \ldots, t_n) \in \mathcal{I}$, $\mathrm{dlp}^{\mathrm{f}}(\mathcal{I})$ contains the following atom α:

- If p is of the form q_i for some $q \in \{in, su, u, sub, eq\}$ and number i, then $\alpha = q(\bar{t}_1, \ldots, \bar{t}_n)$.
- If p is any other predicate, then $\alpha = p(\bar{t}_1, \ldots, \bar{t}_n)$.

For readability, let $\mathcal{J} := \mathrm{dlp}^{\mathrm{f}}(\mathcal{I})$. We show that $\mathcal{J}$ is a stable model of $\mathrm{dlp}^{\mathrm{f}}(P)$. We first show that every rule $r \in \mathrm{nex}(P)$ that was applied in the chase that produced $\mathcal{I}$ is also represented in the reduct $\mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{J}}$ in the sense that there is a rule $r' \in \mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{J}}$ that is obtained from r by (i) instantiating variables as in the application of r in the chase, (ii) removing all negative literals, and (iii) replacing predicates q_i for some $q \in \{in, su, u, sub, eq\}$ by q. This is immediate from the definitions for all rules in $\mathrm{dlp}^{\mathrm{f}}(P)$ without negative atoms.

If r is a rule of the form (15) (using the stratified set predicates of $\mathrm{nex}(P)$), then all predicates used in negative literals of r occur in strictly lower strata of $\mathrm{nex}(P)$. Since the negative literals of (the respective instance of) r had not been inferred in the chase when r was applied, they are therefore never inferred in $\mathcal{I}$. Their images are therefore not in $\mathcal{J}$, and hence we find a representation of r in $\mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{J}}$.

The only remaining type of rule in $\mathrm{nex}(P)$ is (23) (resp. its copies for each stratum). Whenever such a rule is applied to a fact $get\_su_i(t, s)$ in the chase, we find that $in_j(t, s) \notin \mathcal{I}$ for any j. Otherwise, if $in_j(t, s)$ were inferred from $su_j(t, s, s)$ using (10), where $su_j(t, s, s)$ was inferred using rules (8) or (9). Since s is necessarily initialised in a stratum $\ell \leq i$, the same inferences are also possible to obtain $su_\ell(t, s, s)$, and hence $su_i(t, s, s)$ (using rules $su_{k+1}(X, S, U) \leftarrow su_k(X, S, U)$). In a Datalog-first chase, $su_i(t, s, s)$ would therefore be inferred before applying the rule r as assumed previously, contradicting the assumption that this rule was applicable. Hence, $\mathcal{I}$ contains no fact of the form $in_j(t, s)$, and we find that the instance of r is represented by a ground instance of (7) in $\mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{J}}$.

Therefore, every rule applied in the chase over $\mathrm{nex}(P)$ has a corresponding rule in $\mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{J}}$. By a simple induction, we can verify that there is a sequence of applications of rules in $\mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{J}}$ that follows the chase over $\mathrm{nex}(P)$ and leads to $\mathcal{J}$. Hence, $\mathcal{J}$ is contained in a minimal model of $\mathrm{nex}(P)$. Finally, we can verify that $\mathcal{J}$ satisfies all rules of $\mathrm{dlp}^{\mathrm{f}}(P)^{\mathcal{J}}$, so that $\mathcal{J}$ is a minimal model and hence a stable model of $\mathrm{dlp}^{\mathrm{f}}(P)$. □

Theorem 7. *If P is a stratified Datalog(S) program with negation, then every stratified, Datalog-first chase $\mathcal{I}$ of $\mathrm{nex}(P)$ is finite and $[\mathcal{I}] := \{p([t_1], \cdots, [t_n]) \mid \hat{p}(t_1, \cdots, t_n) \in \mathcal{I}\}$ is the unique stable model of P.*

Proof. Consider P and $\mathcal{I}$ as in the claim. Finiteness was established in Lemma 13. By Lemma 15, we obtain a stable model $\mathcal{J} = \mathrm{dlp}^{\mathrm{f}}(\mathcal{I})$ of $\mathrm{dlp}^{\mathrm{f}}(P)$. By Theorem 3, $[\mathcal{J}]$ is a stable model of P, and, by construction, $[\mathcal{J}] = [\mathcal{I}]$, so that $[\mathcal{I}]$ is a stable model of P. By Lemma 14, P has only one stable model, and therefore $[\mathcal{I}]$ is this unique model for every possible stratification and Datalog-first chase sequence. □