Online appendix for the paper

Solving Horn Clauses on Inductive Data Types Without Induction

published in Theory and Practice of Logic Programming

EMANUELE DE ANGELIS, FABIO FIORAVANTI

DEC, 'G. d'Annunzio' University of Chieti-Pescara, Pescara, Italy
(*e-mail:* {emanuele.deangelis, fabio.fioravanti}@unich.it)

ALBERTO PETTOROSSI

DICII, University of Rome Tor Vergata, Rome, Italy
(*e-mail:* pettorossi@info.uniroma2.it)

MAURIZIO PROIETTI

CNR-IASI, Rome, Italy
(*e-mail:* maurizio.proietti@iasi.cnr.it)

submitted 20 February 2018; revised 5 April 2018; accepted 15 April 2018

Appendix

In order to prove Theorem 2, we first recall some notions and results regarding the transformation rules and their correctness.

A *transformation sequence* is a sequence $S_0, S_1, \ldots, S_n$ of sets of CHCs, whose constraints are in $LIA \cup Bool$, where, for $i=0,\ldots,n-1$, S_{i+1} is derived from S_i by applying one of the following rules R1–R5.

Let $Defs_i$ denote the set of all the clauses, called *definitions*, introduced by rule R1 during the construction of the transformation sequence $S_0, S_1, \ldots, S_i$. In particular, $Defs_0 = \emptyset$.

(R1) *Define.* We introduce a clause D: $newp(X_1,\ldots,X_k) \leftarrow c, G$, where: (i) $newp$ is a predicate symbol not occurring in the sequence $S_0, S_1, \ldots, S_i$, (ii) c is a constraint in $LIA \cup Bool$, (iii) G is a non-empty conjunction of atoms whose predicate symbols occur in S_0, and (iv) $X_1,\ldots,X_k$ are distinct variables occurring in (c, G). Then, we derive the new set $S_{i+1} = S_i \cup \{D\}$ and $Defs_{i+1} = Defs_i \cup \{D\}$.

(R2) *Unfold.* Let C: $H \leftarrow c, L, A, R$ be a variant of a clause in S_i. Let $K_1 \leftarrow c_1, B_1$, ..., $K_m \leftarrow c_m, B_m$ be all clauses of S_i (without loss of generality, we assume $vars(S_i) \cap vars(C) = \emptyset$) such that, for $j=1,\ldots,m$, (1) there exists a most general unifier ϑ_j of A and K_j, and (2) the constraint $(c, c_j)\vartheta_j$ is satisfiable. By *unfolding the atom A in C using S_i* we derive the new set $S_{i+1} = (S_i \setminus \{C\}) \cup \{(H \leftarrow c, c_1, L, B_1, R)\vartheta_1, \ldots, (H \leftarrow c, c_m, L, B_m, R)\vartheta_m\}$.

(R3) *Fold.* Let C: $H \leftarrow c, L, Q, R$ be a clause in S_i, where Q is a non-empty conjunction of atoms, and let D: $K \leftarrow d, B$ be (a variant of) a clause in $Defs_i$ with $vars(C) \cap vars(D) = \emptyset$. Suppose that there exist a substitution ϑ and a constraint e such that: (i) $Q = B\vartheta$, (ii) $LIA \cup Bool \models \forall(c \leftrightarrow (e \wedge d\vartheta))$, and (iii) for every variable $X \in vars(d, B) \setminus vars(K)$,

the following conditions hold: (iii.1) $X\vartheta$ is a variable not occurring in $\{H, c, L, R\}$, and (iii.2) $X\vartheta$ does not occur in the term $Y\vartheta$, for any variable Y occurring in (d, B) and different from X. By *folding* C *using the definition* D, we derive clause E: $H \leftarrow e, L, K\vartheta, R$. In this case we also say that E is derived *by folding* Q *in* C. We derive the new set $S_{i+1} = (S_i \setminus \{C\}) \cup \{E\}$.

(R4) *Replace Equivalent Constraints.* Let us consider a subset of S_i of the form $\{(H \leftarrow c_1, G), \ldots, (H \leftarrow c_k, G)\}$. Suppose that, for some constraints $d_1, \ldots, d_m$,

$$LIA \cup Bool \models \forall (\exists Y_1 \ldots \exists Y_r \ (c_1 \vee \ldots \vee c_k) \leftrightarrow \exists Z_1 \ldots \exists Z_s \ (d_1 \vee \ldots \vee d_m))$$

where $\{Y_1, \ldots, Y_r\} = vars(c_1 \vee \ldots \vee c_k) \setminus vars(\{H, G\})$ and $\{Z_1, \ldots, Z_s\} = vars(d_1 \vee \ldots \vee d_m) \setminus vars(\{H, G\})$. Then, we derive the new set $S_{i+1} = (S_i \setminus \{(H \leftarrow c_1, G), \ldots, (H \leftarrow c_k, G)\}) \cup \{(H \leftarrow d_1, G), \ldots, (H \leftarrow d_m, G)\}$.

Note that rule R4 enables the deletion of a clause with an inconsistent constraint in its body. Indeed, if c_1 is unsatisfiable, then $LIA \cup Bool \models \forall (c_1 \leftrightarrow d_1 \vee \ldots \vee d_m)$ with $m = 0$.

(R5) *Replace Functional Predicates.* Let C: $H \leftarrow c, G_1, p(t, u), G_2, p(t, w), G_3$, be a clause in S_i and let $p(X, Y)$ be functional in S_i (see Definition 3). Then, we derive the new set $S_{i+1} = (S_i \setminus \{C\}) \cup \{(H \leftarrow c, G_1, p(t, u), G_2, G_3)\vartheta\}$, where ϑ is the most general unifier of u and w.

The following theorem sums up various results presented in the literature (Etalle and Gabbrielli 1996, Tamaki and Sato 1984).

Theorem 4 (*Equivalence with respect to satisfiability*)
Let $S_0, S_1, \ldots, S_n$ be a transformation sequence such that every definition in $Defs_n$ is unfolded during the construction of this sequence. Then, $S_0 \cup LIA \cup Bool$ is satisfiable if and only if $S_n \cup LIA \cup Bool$ is satisfiable.

Now, we prove Theorems 2 and 3 of Section 5.

Theorem 2 (*Partial Correctness*)
Let *Cls* be a set of definite clauses and let *Gs* be a set of goals. If Algorithm $\mathcal{E}$ terminates for the input clauses $Cls \cup Gs$, returning a set *TransfCls* of clauses, then (1) $Cls \cup Gs$ is satisfiable iff *TransfCls* is satisfiable, and (2) all clauses in *TransfCls* have basic types.
Proof
Point (1) follows from Theorem 4 by taking into account that: (i) Algorithm $\mathcal{E}$ can be viewed as a particular sequence of applications of Rules R1–R5, and (ii) every definition in *Defs* is unfolded during the execution of $\mathcal{E}$.
Point (2) follows from the fact that, by construction, every clause introduced in *TransfCls* has basic types. To see this, let us consider a clause C in *TransfCls*. Clause C belongs to the set *FldCls* of clauses obtained by a *Define-Fold* step. Looking at the *Define-Fold* procedure, we have that: (i) the head of C is either *false* (because $C \in Gs$) or its head predicate has been introduced by the *Define* step, and hence, by construction, has basic types, and (ii) the body of C has the form: $c, newp_1(V_1), \ldots, newp_n(V_n)$, where c is a constraint which has basic types (because it belongs to $LIA \cup Bool$) and $newp_1, \ldots, newp_n$ are predicates that, by construction, have basic types. □

Theorem 3 (*Termination*)
Let *Cls* be a set of definite clauses such that every clause in *Cls* has a disjoint, quasi-descending slice decomposition. Let *Gs* be a set of goals such that, for each goal $G \in Gs$,

(i) G is of the form $\mathit{false} \leftarrow c, A_1, \ldots, A_m$, where for $i = 1, \ldots, m$, A_i is an atom whose arguments are distinct variables, and (ii) G has no sharing cycles. Then Algorithm $\mathcal{E}$ terminates for the input clauses $\mathit{Cls} \cup \mathit{Gs}$.

Proof (Sketch)
Without loss of generality, we assume that for every clause $H \leftarrow c, B$ every variable of basic type has at most one occurrence in H, B.

For any tree t, by $\mathit{height}(t)$ we denote the *height* of t, that is, the maximal length of a path from the root of t to one of its leaves. We extend the function *height* to terms and atoms, viewed as trees. First of all, we observe that Algorithm $\mathcal{E}$ terminates iff there exists two non-negative integers H and N such that, for every definition $\mathit{newp}(V) \leftarrow A_1, \ldots, A_n$ added to *Defs* during the execution of $\mathcal{E}$, we have that, for $i = 1, \ldots n$, $\mathit{height}(A_i) \leq H$, and $n \leq N$. Indeed, the existence of H and N implies the finiteness of the set of definitions introduced by $\mathcal{E}$, and hence the finiteness of the number of iterations of the body of the while-do loop of $\mathcal{E}$ itself.

Let us consider a clause D: $\mathit{newp}(V) \leftarrow A_1, \ldots, A_n$ added to *Defs* during the execution of $\mathcal{E}$. Then D satisfies the following properties:

- P1. The goal $\mathit{false} \leftarrow A_1, \ldots, A_n$ obtained from D by replacing $\mathit{newp}(V)$ by *false*, has no sharing cycles;
- P2. All atoms in the body of D are linear;
- P3. For any two distinct atoms A_i and A_j in the body of D, if A_i and A_j share a non-basic variable, then they are of the form $p(\ldots, t_i, \ldots)$ and $q(\ldots, t_j, \ldots)$, where either $t_i \preceq t_j$ or $t_j \preceq t_i$, and A_i and A_j do not share any variable besides the ones in $\mathit{vars}(t_i) \cap \mathit{vars}(t_j)$;
- P4. For any atom A_i in the body of D, $\mathit{height}(A_i) \leq H$, where H is the maximal height of an atom in $\mathit{Cls} \cup \mathit{Gs}$;
- P5. Let V_G be the number of occurrences of non-basic variables in a goal G, and let M be $\max\{V_G \mid G \in \mathit{Gs}\}+1$. Then, in the body of D, (P5.1) every non-basic variable has at most M occurrences, and (P5.2) there exist $K \leq M$ predicate arguments such that every non-basic variable that occurs more than once, also occurs in one of those K arguments.

Property P1 holds for each clause D initially in *Defs* by the hypothesis that *Gs* is a set of goals that have no sharing cycles. This property, when referred to the body of the clauses, is preserved by the *Unfold* and *Replace* steps, due to the hypothesis on the clauses in *Cls*, and hence it also holds for each new definition added to *Defs* by the *Define* step after *Unfold* and *Replace*.

Property P2 holds for each clause D initially in *Defs* by Hypothesis (i) on *Gs*. This property is preserved by the *Unfold* and *Replace* steps, due to the hypothesis on the clauses in *Cls*. Note, in particular, that the existence of a disjoint, quasi-descending slice decomposition for all clauses in *Cls* implies that each atom in the body of a clause in *Cls* is linear, and hence only linear atoms are introduced by *Unfold* steps. The linearity of the atoms different from the one replaced by an *Unfold* step is enforced by the existence of a disjoint, quasi-descending slice decomposition for all clauses in *Cls*. Linearity is also preserved by *Replace* steps. Thus, Property P2 follows from the fact that the body of D consists of atoms taken from the body of a clause derived by *Unfold* and *Replace* steps.

Property P3 holds for each clause D initially in *Defs* by Hypothesis (i) on *Gs*. Property P3 also holds for each clause derived by the *Unfold* and *Replace* steps by Property P1

and by the hypothesis that every clause in *Cls* has a disjoint, quasi-descending slice decomposition. Then, Property P3 follows from the fact that the body of D consists of atoms taken from the body of clauses derived by *Unfold* and *Replace* steps.

Property P4 holds for each clause D initially in *Defs* because the body of clause D is the set of atoms occurring in the body of a goal in *Gs*. This property also holds for each clause derived by *Unfold* steps. Indeed, suppose that we unfold an atom A in the clause C of the form $H \leftarrow c, L, A, R$ such that either (i) A is strictly maximal in L, A, R, or (ii) all atoms in L, A, R are not strictly maximal. Both in case (i) and case (ii), by Property P3, A is of the form $p(\ldots, t_i, \ldots)$ and any atom Q in L, R that shares a non-basic variable with A is of the form $q(\ldots, t_j, \ldots)$, with $t_j \preceq t_i$, and A and Q do not share any variable besides the ones in $vars(t_i) \cap vars(t_j)$. Let $K_1 \leftarrow c_1, B_1, \ldots, K_m \leftarrow c_m, B_m$ be all clauses of *Cls* (with $vars(Cls) \cap vars(C) = \emptyset$) such that, for $i = 1, \ldots, m$, (i) there exists a most general unifier ϑ_i of A and K_i, and (ii) the constraint $(c, c_i)\vartheta_i$ is satisfiable. Then, by unfolding A in C we derive the clauses $(H \leftarrow c, c_1, L, B_1, R)\vartheta_1, \ldots, (H \leftarrow c, c_m, L, B_m, R)\vartheta_m$. By Property P2 A is a linear atom, and by the hypothesis that every clause in *Cls* has a disjoint, quasi-descending slice decomposition, we have that, for $i = 1, \ldots, m$, for every atom E in L, B_i, R, $height(E\vartheta_i) \leq \max(\{height(E), height(A), height(K_i)\})$. Thus, if Property P4 holds for C, then it also holds for the clauses $(H \leftarrow c, c_1, L, B_1, R)\vartheta_1, \ldots,$ $(H \leftarrow c, c_m, L, B_m, R)\vartheta_m$. Property P4 also holds for each clause derived by *Replace* steps, and hence it also holds for D, whose body consists of atoms taken from the body of clauses derived by *Unfold* and *Replace* steps.

Property P5 holds for each clause D initially in *Defs* because the body of clause D is the set of atoms occurring in the body of a goal G in *Gs* for which Hypothesis (i) holds. Now we prove that the following two properties, which generalize Property P5, hold for each clause E derived by an *Unfold* step: in each sharing block in the body of E, (P5.1) every non-basic variable has at most M occurrences, and (P5.2) there exist $K \leq M$ predicate arguments such that every non-basic variable with more than one occurrence in the body of E, also occurs in one of those K arguments.

Suppose that P5.1 and P5.2 hold for a clause C of the form $H \leftarrow c, L, A, R$, and we unfold A in C. Let $K_1 \leftarrow c_1, B_1, \ldots, K_m \leftarrow c_m, B_m$ be all clauses of *Cls* (with $vars(Cls) \cap vars(C) = \emptyset$) such that, for $i = 1, \ldots, m$, (i) there exists a most general unifier ϑ_i of A and K_i, and (ii) the constraint $(c, c_i)\vartheta_i$ is satisfiable. Then, by unfolding A in C we derive the clauses $C_1 : (H \leftarrow c, c_1, L, B_1, R)\vartheta_1, \ldots, C_m : (H \leftarrow c, c_m, L, B_m, R)\vartheta_m$. By Properties P2 and P3, and by the existence of a disjoint, quasi-descending slice decomposition, for $i = 1, \ldots, m$, the number of occurrences of any variable with more than one occurrence in the body of C_i, is not larger than the maximal number of occurrences of any variable in the body of C, and hence Property P5.1 holds for the body of C_i.

Moreover, suppose that in every sharing block in the body of C there exist $K \leq M$ arguments $t_1, \ldots, t_K$ such that every variable variable with more than one occurrence, also occurs in one of those K arguments. By Property P3, we may assume that $t_1, \ldots, t_K$ are maximal with respect to the $\preceq$ relation and do not share any variable. Looking at the *Unfold* procedure, the atom A selected for unfolding must have one among $t_1, \ldots, t_K$ as an argument, say t_1. By our hypotheses, if by unfolding A the argument t_1 is replaced by more than one term, these new terms must appear in different sharing blocks, and hence the number of maximal arguments in each sharing block does not increase. Thus, Property P5.2 holds for $C_1, \ldots, C_m$. Similarly, we can prove that Properties P5.1 and

P5.2 hold for each clause derived by *Replace* steps, and hence Property 5 holds for D, whose body consists of a sharing block of a clause derived by *Unfold* and *Replace* steps.

Now, from Properties P4 and P5 it follows that there exists an integer J, depending on H and M, such that in the body of D there are at most J distinct variables. Thus, there exists N such that the body of D has at most N atoms of height not larger than H, and hence the thesis holds. □