

A General Framework for Automatic Termination Analysis of Logic Programs*

Nachum Dershowitz¹, Naomi Lindenstrauss², Yehoshua Sagiv²,
Alexander Serebrenik³

¹ School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel.
e-mail: nachumd@math.tau.ac.il

² Institute for Computer Science, The Hebrew University, Jerusalem 91904, Israel.
e-mail: {naomil,sagiv}@cs.huji.ac.il

³ Department of Computer Science, K.U. Leuven, Celestijnenlaan 200A, B-3001
Heverlee, Belgium. e-mail: Alexander.Serebrenik@cs.kuleuven.ac.be

The date of receipt and acceptance will be inserted by the editor

Abstract This paper describes a general framework for automatic termination analysis of logic programs, where we understand by “termination” the finiteness of the LD-tree constructed for the program and a given query. A general property of mappings from a certain subset of the branches of an infinite LD-tree into a finite set is proved. From this result several termination theorems are derived, by using different finite sets. The first two are formulated for the predicate dependency and atom dependency graphs. Then a general result for the case of the query-mapping pairs relevant to a program is proved (cf. [29,21]). The correctness of the *TermiLog* system described in [22] follows from it. In this system it is not possible to prove termination for programs involving arithmetic predicates, since the usual order for the integers is not well-founded. A new method, which can be easily incorporated in *TermiLog* or similar systems, is presented, which makes it possible to prove termination for programs involving arithmetic predicates. It is based on combining a finite abstraction of the integers with the technique of the query-mapping pairs, and is essentially capable of dividing a termination proof into several cases, such that a simple termination function suffices for each case. Finally several possible extensions are outlined.

Key words termination of logic programs – abstract interpretation – constraints

* This research has been partially supported by grants from the Israel Science Foundation

1 Introduction

The results of applying the ideas of abstract interpretation to logic programs (cf. [10]) seem to be especially beautiful and useful, because we are dealing in this case with a very simple language which has only one basic construct—the clause. Termination of programs is known to be undecidable, but again things are simpler for logic programs, because the only possible cause for their non-termination is infinite recursion, so it is possible to prove termination automatically for a large class of programs. For a formal proof of the undecidability of the termination of general logic programs see [1].

The kind of termination we address is the termination of the computation of all answers to a goal, given a program, when we use Prolog’s computation rule (cf. [24]). This is equivalent to finiteness of the LD-tree constructed for the program and query (the LD-tree is the SLD-tree constructed with Prolog’s computation rule—cf. [2]). Even if one is interested only in a single answer, it is important to know that computation of all answers terminates, since the solved query may be backtracked into (cf. [27]).

One of the difficulties when dealing with the LD-derivation of a goal, given a logic program, is that infinitely many non-variant atoms may appear as subgoals. The basic idea is to abstract this possibly infinite structure to a finite one. We do this by mapping partial branches of the LD-tree to the elements of a finite set of abstractions \mathcal{A} . By using the basic lemma of the paper and choosing different possibilities for \mathcal{A} , we get different results about termination. The first two results are formulated for the predicate dependency and atom dependency graphs.

Then we get, by using the query-mapping pairs of [29,21], first a termination condition that cannot be checked effectively and then a condition that can. The latter forms the core of the *TermiLog* system (cf. [22]), a quite powerful system we have developed for checking termination of logic programs.

Then a new method, which can be easily incorporated in the *TermiLog* or similar systems, is presented for showing termination for logic programs with arithmetic predicates. Showing termination in this case is not easy, since the usual order for the integers is not well-founded. The method consists of the following steps: First, a finite abstract domain for representing the range of integers is deduced automatically. Based on this abstraction, abstract interpretation is applied to the program. The result is a finite number of atoms abstracting answers to queries, which are used to extend the technique of query-mapping pairs. For each query-mapping pair that is potentially non-terminating, a bounded (integer-valued) termination function is guessed. If traversing the pair decreases the value of the termination function, then termination is established. Usually simple functions suffice for each query-mapping pair, and that gives our approach an edge over the classical approach of using a single termination function for all loops, which must inevitably be more complicated and harder to guess automatically.

It is worth noting that the termination of McCarthy's 91 function can be shown automatically using our method.

Finally generalizations of the algorithms presented are pointed out, which make it possible to deal successfully with even more cases.

2 Preliminaries

Consider the LD-tree determined by a program and goal.

Definition 2.1 Let $\leftarrow r_1, \dots, r_n$ and $\leftarrow s_1, \dots, s_m$ be two nodes on the same branch of the LD-tree, with the first node being above the second. We say $\leftarrow s_1, \dots, s_m$ is a direct offspring of $\leftarrow r_1, \dots, r_n$ if s_1 is, up to a substitution, one of the body atoms of the clause with which $\leftarrow r_1, \dots, r_n$ was resolved. We define the offspring relation as the irreflexive transitive closure of the direct offspring relation. We call a path between two nodes in the tree such that one is the offspring of the other a call branch.

Take for example the **add-mult** program given in Figure 2.1 and the goal $\text{mult}(\text{s}(\text{s}(0)), \text{s}(0), \text{Z})$.

- (i) $\text{add}(0, 0, 0)$.
- (ii) $\text{add}(\text{s}(X), Y, \text{s}(Z)) \text{ :- add}(X, Y, Z)$.
- (iii) $\text{add}(X, \text{s}(Y), \text{s}(Z)) \text{ :- add}(X, Y, Z)$.
- (iv) $\text{mult}(0, X, 0)$.
- (v) $\text{mult}(\text{s}(X), Y, Z) \text{ :- mult}(X, Y, Z1), \text{add}(Z1, Y, Z)$.

Fig. 2.1 add-mult example

The LD-tree is given in Figure 2.2. In this case node (2) and node (6) are, for instance, direct offspring of node (1), because the first atoms in their respective goals come from the body of clause (v), with which the goal of node (1) was resolved. Note that we add to the predicate of each atom in the LD-tree a subscript that denotes who its 'parent' is, i.e., the node in the LD-tree that caused this atom to be called as the result of resolution. A graphical representation of the direct offspring relation is given in Figure 2.3.

The following theorem holds:

Theorem 2.1 *If there is an infinite branch in the LD-tree corresponding to a program and query then there is an infinite sequence of nodes N_1, N_2, \dots such that for each i , N_{i+1} is an offspring of N_i .*

Proof Straightforward.

The main idea of the paper is to find useful finite sets of abstractions of call branches and to formulate termination results in terms of them. An effort has been made to make the presentation as simple and self-contained as possible.

- (1) $\leftarrow \text{mult}(s(s(0)), s(0), Z)$
(2) $\leftarrow \text{mult}_{(1)}(s(0), s(0), Z1), \text{add}_{(1)}(Z1, s(0), Z)$
(3) $\leftarrow \text{mult}_{(2)}(0, s(0), Z2), \text{add}_{(2)}(Z2, s(0), Z1), \text{add}_{(1)}(Z1, s(0), Z)$
 $\{Z2 \mapsto 0\}$
(4) $\leftarrow \text{add}_{(2)}(0, s(0), Z1), \text{add}_{(1)}(Z1, s(0), Z)$
 $\{Z1 \mapsto s(Z3)\}$
(5) $\leftarrow \text{add}_{(4)}(0, 0, Z3), \text{add}_{(1)}(s(Z3), s(0), Z)$
 $\{Z3 \mapsto 0\}$
(6) $\leftarrow \text{add}_{(1)}(s(0), s(0), Z)$
 $\{Z \mapsto s(Z4)\}$ $\{Z \mapsto s(Z5)\}$
(7) $\leftarrow \text{add}_{(6)}(0, s(0), Z4)$ (8) $\leftarrow \text{add}_{(6)}(s(0), 0, Z5)$
 $\{Z4 \mapsto s(Z6)\}$ $\{Z5 \mapsto s(Z7)\}$
(9) $\leftarrow \text{add}_{(7)}(0, 0, Z6)$ (10) $\leftarrow \text{add}_{(8)}(0, 0, Z7)$
 $\{Z6 \mapsto 0\}$ $\{Z7 \mapsto 0\}$
(11) \leftarrow (12) \leftarrow

Fig. 2.2 LD-tree

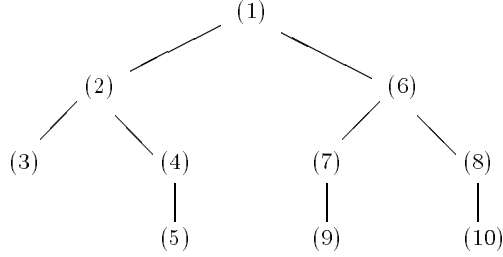


Fig. 2.3 The offspring relation

3 The basic lemma

Given an LD-tree we define a shadow of it as a mapping from its set of call branches to a finite set of abstractions.

Definition 3.1 (Shadow) *Let an LD-tree for a query and program and a finite set \mathcal{A} be given. A shadow of the LD-tree into \mathcal{A} is a mapping α that assigns to each call branch of the tree an element of \mathcal{A} .*

Then the following basic lemma holds

Lemma 3.1 (Basic Lemma) *Suppose the LD-tree for a program and a query has an infinite branch. Let α be a shadow mapping from the call branches of the tree into a finite set \mathcal{A} . Then there is a sequence of nodes*

M_1, M_2, \dots and an element $A \in \mathcal{A}$, such that for each i , M_{i+1} is an offspring of M_i , and for each j, k the call branch from M_j to M_k is mapped by α to A .

Proof By Theorem 2.1, there is an infinite sequence of nodes N_1, N_2, \dots , such that for each i , N_{i+1} is an offspring of N_i . To each call branch from N_i to an N_j the mapping α assigns one of the elements of the finite set \mathcal{A} . By Ramsey's theorem [17] we get that there is a subsequence N_{k_1}, N_{k_2}, \dots , such that for each i, j the mapping α assigns to the branch from N_{k_i} to N_{k_j} the same element.

There is some structure in the set of call branches. If we have two call branches, one going from N_1 to N_2 and one going from N_3 to N_4 , we can if $N_2 = N_3$ define their composition, which is the branch from N_1 to N_4 . This operation is associative. In accordance with the nomenclature in algebra we can call a set S with a partial associative operation $* : S \times S \rightarrow S$ a semi-groupoid. We may want the finite set \mathcal{A} to be a semi-groupoid too and the mapping α to be a homomorphism. This brings us to the definition of a structured shadow.

Definition 3.2 (Structured Shadow) *Let an LD-tree for a query and program and a finite semi-groupoid \mathcal{A} be given. A structured shadow of the LD-tree into \mathcal{A} is a mapping α that assigns to each call branch of the tree an element of \mathcal{A} so that for any two call branches B_1 and B_2 that can be composed we have that $\alpha(B_1) * \alpha(B_2)$ is defined and*

$$\alpha(B_1 * B_2) = \alpha(B_1) * \alpha(B_2)$$

When defining a structured shadow it is enough to give the value of α for call branches between nodes and their direct offspring. This is the reason for the name.

The element A whose existence is proved in the basic lemma is, in the case of a structured shadow, an element that can be composed with itself. We call such an element a *circular element*. Moreover, it is *idempotent*.

4 Two simple applications of the basic lemma

In the following sections we'll give applications of the basic lemma. In each case we'll give the set of abstractions \mathcal{A} which will always be finite and the mapping α from call branches to elements of \mathcal{A} . In the first two applications we use the absence of circular elements in \mathcal{A} to derive termination.

4.1 The Predicate Dependency Graph

Take as \mathcal{A} elements of the form $(p \rightarrow q)$ where p and q are predicate symbols of the program. Define composition as

$$(p \rightarrow q) * (q \rightarrow r) = (p \rightarrow r)$$

If a call branch goes from a node $\leftarrow p(X_1, \dots, X_n), \dots$ to another node $\leftarrow q(Y_1, \dots, Y_m), \dots$ we'll define the value of α on it as $(p \rightarrow q)$. It is not difficult to see that α is a structured shadow.

The *predicate dependency graph* of a program is a graph whose vertices are the predicate symbols of the program and such that for each clause $A :- B_1, \dots, B_n$ it has an arc from the predicate of A to the predicate of B_i , for $i = 1, \dots, n$ (cf. [28]). If N_1, N_2 are nodes in the LD-tree such that one is the direct offspring of the other then the value of α for the call branch between them can be seen as an arc in the predicate dependency graph of the program.

From the basic lemma we get that if there is non-termination then there must be a circular element in the image under α of all the call branches of the LD-tree, that is, an element of the form $p \rightarrow p$. This means that there is a non-trivial strongly connected component in the predicate dependency graph (a trivial strongly connected component is one that consists of a single vertex with no arc going from it to itself). Consequently, the following well-known theorem follows from the basic lemma:

Theorem 4.1 *If there is no non-trivial strongly connected component in the predicate dependency graph of a program any query to it terminates.*

It is easy to find examples of programs such that every query to them terminates and yet their predicate dependency graph has non-trivial strongly connected components. Take the program

```
at(jerusalem,mary).
at(X,jane) :- at(X,mary).
```

where the predicate dependency graph has the single vertex at with an arc $at \rightarrow at$.

4.2 The Atom Dependency Graph

Define two atoms to be equivalent if they are variants of each other. For an atom At denote by $[At]$ its equivalence class under variance. Take as \mathcal{A} elements of the form $[p(T_1, \dots, T_n)] \rightarrow [q(S_1, \dots, S_m)]$ where $p(T_1, \dots, T_n)$ is an atom that appears in the head of a clause in the program and $q(S_1, \dots, S_m)$ is an atom that appear in the body of a clause. Composition is defined for pairs

$[p(T_1, \dots, T_n)] \rightarrow [q(S_1, \dots, S_m)]$ and $[q(R_1, \dots, R_m)] \rightarrow [r(W_1, \dots, W_k)]$
 such that representatives that are named apart of $[q(S_1, \dots, S_m)]$ and $[q(R_1, \dots, R_m)]$ can be unified. In that case the result of the composition is $[p(T_1, \dots, T_n)] \rightarrow [r(W_1, \dots, W_k)]$.

Now suppose a node $\leftarrow p(T_1, \dots, T_n), \dots$ has as direct offspring a node $\leftarrow q(S_1, \dots, S_m), \dots$ and suppose the clause used for resolution with the first node was $A :- B_1, \dots, B_l$ and that the atom $q(S_1, \dots, S_m)$ originates

in B_j . Then we will take α to map the call branch between the two nodes to $[A] \rightarrow [B_j]$. For call branches between nodes that are not direct offspring of each other we define the value of α by composition.

We can define the *atom dependency graph* of a program as follows. Consider a graph whose vertices are equivalence classes of atoms that appear in the program. If there is a rule $A :- B_1, B_2, \dots, B_n$ then we put in the graph arcs $[A] \rightarrow [B_i]$ ($i = 1, \dots, n$). We call these arcs “arcs of the first kind”. Now if there are arcs of the first kind $[A_1] \rightarrow [A_2]$ and $[B_1] \rightarrow [B_2]$ and named apart variants of A_2 and B_1 can be unified, we also add an arc $[A_2] \rightarrow [B_1]$. Such an arc we call “an arc of the second kind”. The graph we get we call the atom dependency graph (note the similarity to the U-graph of [38]). For the example at the end of the previous subsection the atom dependency graph consists of the two vertices $[at(X, jane)], [at(X, mary)]$ and an arc from the first to the second.

From the basic lemma we get that if there is an infinite branch in the LD-tree there must be a circular element in the image under α of the call branches of the LD-tree.

So we get the following conclusion of the basic lemma:

Theorem 4.2 *If there is no non-trivial strongly connected component in the atom dependency graph of a program any query to it terminates.*

Again it is not difficult to find programs such that every query to them terminates and yet their atom dependency graph has non-trivial strongly connected components. Take the following program

$$\begin{aligned} p(\mathbf{X}, \mathbf{f}(\mathbf{Z})) & :- q(\mathbf{X}, \mathbf{f}(\mathbf{Z})). \\ q(\mathbf{g}(\mathbf{Y}), \mathbf{W}) & :- r(\mathbf{g}(\mathbf{Y}), \mathbf{W}). \\ r(\mathbf{X}, \mathbf{X}) & :- p(\mathbf{X}, \mathbf{X}). \end{aligned}$$

for which every SLD-tree is finite, but for which there is a strongly connected component consisting of 6 nodes in its atom dependency graph—if we denote by a, b, c, d, e, f respectively the atom dependency graph nodes

$$[p(X, f(Y))], [q(X, f(Y))], [q(g(X), Y)], [r(g(X), Y)], [r(X, X)], [p(X, X)]$$

then there are arcs of the first kind from a to b , from c to d and from e to f and arcs of the second kind from b to c , from d to e and from f to a , so the nodes a, b, c, d, e, f form a strongly connected component.

5 The Abstraction to Query-Mapping Pairs

We will now consider a more complex abstraction and take as \mathcal{A} the set of query-mapping pairs determined by the program. In this case termination will follow not from the absence of circular elements in the image of the

shadow mapping, but from the absence of circular elements of a certain kind.

We start with a formal definition of query-mapping pairs. The meaning of the pairs will be clarified later.

Definition 5.1 (Mixed Graph) *A mixed graph is a graph with both edges and arcs. (We use the usual terminology—edges are undirected, while arcs are directed.)*

A query-mapping pair consists of two parts, both of which are mixed graphs, however a different notation is used for each.

Definition 5.2 (Query-Mapping Pair) *A query-mapping pair (π, μ) consists of two parts:*

- *The query π , that is a mixed graph whose nodes correspond to argument positions of some predicate in the program and are either black, denoted by \mathbf{b} , or white, denoted by \mathbf{f} . An edge from the i 'th to the j 'th position will be denoted by $eq(i, j)$. An arc from the i 'th to the j 'th position will be denoted by $gt(i, j)$. As an example of a query for the `add-mult` program take `mult(b,b,f) [gt(1,2),eq(2,3)]`.*
- *The mapping μ , that is a mixed graph whose nodes correspond to the argument positions of the head of some rule (the domain) and the argument positions of some body atom of that rule (the range). Again nodes can be black or white. In this case we depict the graph pictorially, as in Figure 5.1.*

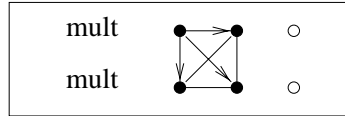


Fig. 5.1 Example of mapping

For examples of query-mapping pairs see Figures 5.2, 5.3, 5.4.

Clearly the number of query-mapping pairs that can be created by using the predicate symbols of a program is finite.

The means for proving termination is choosing a well-founded order on terms and using it to show that the LD-tree constructed for the program and query cannot have infinite branches. Different orders may be defined (cf. [14]). One of the ways an order can be given is by defining a norm on terms. For example, one can use symbolic linear norms, which include as special cases the term-size norm and the list-size norm. These symbolic linear norms will be linear expressions, which we will be able to use in the termination proof when they become integers. Essentially edges and arcs will denote equality and inequality of norms and a node will become black if its symbolic linear norm is an integer.

Definition 5.3 (Symbolic Linear Norm) *A symbolic linear norm for the terms created from an alphabet consisting of function symbols and variables is defined for non-variable terms by*

$$\|f(X_1, \dots, X_n)\| = c + \sum_{i=1}^n a_i \|X_i\|$$

where c and a_1, \dots, a_n are non-negative integers that depend only on f/n . This also defines the norm of constants if we consider them as function symbols with 0 arity. With each logical variable we associate an integer variable to represent its norm (we use the same name for both, since the meaning of the variable is clear from the context).

Definition 5.4 (Instantiated Enough) *A term is instantiated enough with respect to a symbolic linear norm if the expression giving its symbolic norm is an integer.*

In this way of defining symbolic norms we follow [35]. Some authors define the norm of a variable to be 0 and then use the norm only for terms that are *rigid* with respect to it (cf. [13]). In our context it is more convenient to use the symbolic norm. If the symbolic linear norm of a term is an integer then we know that the term is rigid with respect to this particular norm.

We get the term-size norm, which can be defined for a ground term as the number of edges in its representation as a tree, or alternatively as the sum of the arities of its functors, by setting for every f/n

$$c = n, \quad a_1 = \dots = a_n = 1$$

So, for instance, the symbolic term-size norm of $f(g(X, X, Y), X)$ is $5 + 3X + Y$. The symbolic term-size norm of a term is an integer exactly when the term is ground.

To get the list-size norm we set for the list functor

$$\|[H|T]\| = 1 + \|T\|$$

that is $c = 1, a_1 = 0, a_2 = 1$, and for all other functors equate the norm of a term with them as head functor to 0. In this case the norm is a positive integer exactly for lists that have finite positive length, irrespective of whether their elements are fully instantiated or not.

This is perhaps the place to note that, since for the term-size norm all the a_i 's are nonzero, a term is instantiated enough with respect to it only if it is ground, while for other symbolic norms a term may be instantiated enough without being ground.

Given the LD-tree of a program we define the shadow mapping α_{real} as follows.

With each call branch between nodes that are offspring of each other we associate a query-mapping pair in the following way:

If

- the node nearest to the root among the branch nodes is $\leftarrow p_1, \dots, p_m$,
- the node farthest from the root is $\leftarrow q_1, \dots, q_n$,
- the substitution θ is the composition of the substitutions associated with the branch,
- abs is the abstraction function which associates with each atom the same atom with its arguments replaced by \mathbf{b} for arguments that are instantiated enough for the norm used and \mathbf{f} otherwise,

then

- the query of the pair is $abs(p_1)$, with the constraints that hold between the arguments of p_1 ,
- the domain of the mapping is $abs(p_1\theta)$,
- the range is $abs(q_1)$,
- edges connect elements in the domain and range for which the corresponding elements in the tree (i.e. the arguments of $p_1\theta$ and q_1) have the same norm,
- arcs connect elements in the domain and range for which the corresponding elements in the tree are instantiated enough and for which a norm inequality can be inferred.

(The reader might be puzzled why we introduce arcs between elements for which a norm inequality can be inferred *only* if the arguments are instantiated enough—the term-size of $s(X)$ will always be larger than that of X , whatever the substitution for X will be. However, to prove termination we use the well foundedness of the non-negative integers, so will use the fact that there cannot be an infinite path of arcs, since in our case they connect elements with integer norm.)

Take for example the `add-mult` program given in Figure 2.1 and the goal `mult(s(s(0)),s(0),Z)` and use the term-size norm (recall that for a ground term its term-size is the number of edges in its representation as a tree).

To give a few examples of the query-mapping pairs we get for the LD-tree of this program and goal, which is shown in Figure 2.2 (where we denote the constraints of a query by a list of elements of the form $eq(i, j)$ if the i 'th and j 'th arguments have the same term-size, and $gt(i, j)$ if the term-size of the i 'th argument is greater than that of the j 'th argument; and where there is in the mappings an edge between nodes with the same term-size and an arc from a node with larger integer term-size to a node with smaller integer term-size):

For the call branch between node (1) and its direct offspring (2) we get the pair depicted in Figure 5.2. For the call branch between node (6) and its direct offspring (8) we get the pair presented in Figure 5.3. For the call branch between node (6) and its offspring node (10) we get the pair in Figure 5.4.

Note the following properties of query-mapping pairs in the image of α_{real} . A black node corresponds to an argument position which is instanti-

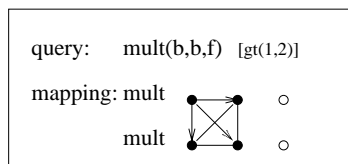


Fig. 5.2 Query-mapping pair for the branch from (1) to (2)

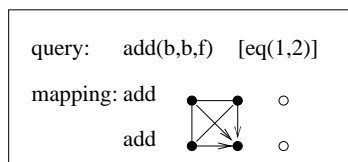


Fig. 5.3 Query-mapping pair for the branch from (6) to (8)

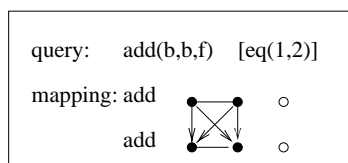


Fig. 5.4 Query-mapping pair for the branch from (6) to (10)

ated enough for the chosen symbolic norm to be an integer. A white node corresponds to an argument position that is potentially not instantiated enough. An edge connects two nodes that have equal symbolic norms, and hence must be of the same color. An arc goes from a black node to another black node that has smaller norm (recall that norms of black nodes are non-negative integers). The proof of termination uses the existence of such arcs and the well-foundedness of the non-negative integers with the usual order.

We define consistency of a mixed graph:

Definition 5.5 (Consistency) *A mixed graph is consistent if it has no positive cycle (i.e. a cycle that may contain both edges and arcs, but has at least one arc).*

Then it is clear that queries and mappings in the image must be consistent. A positive cycle may have only black nodes. This means that for each argument T represented by such a node we have that $\|T\|$ is an integer and $\|T\| < \|T\|$, which is impossible. That is, a query-mapping pair that is not consistent represents a branch that cannot really occur.

Note further that sets of edges and arcs for queries and mappings are closed under transitive closure and that the domain of a mapping of a pair is subsumed by the query, where subsumption is defined as follows.

Definition 5.6 (Subsumption) *Given two mixed graphs with black and white nodes G_1 and G_2 , we say that G_1 is subsumed by G_2 if they have the same nodes up to color, every node that is black in G_2 is also black in G_1 , and every edge or arc between nodes in G_2 also appears for the respective nodes in G_1 .*

Among all query-mapping pairs we distinguish the ‘recursive’ pairs, those for which the query is identical to the range of the mapping (the query-mapping pair given in Figure 5.4 is of this kind).

Before proceeding we need the following two definitions from [29]. We want to model recursive calls, so in the case of a query-mapping pair (π, μ) such that π is identical to the range of μ , we create what we call a *circular variant* by introducing special edges between corresponding nodes in the domain and range. These special edges behave like ordinary edges, except that they can only be traversed from range to domain. What a circular edge between a range node and domain node models, is that the range node can become unified with the domain node of another instance of the pair we considered.

Definition 5.7 (Circular Variant) *If (π, μ) is a query-mapping pair, such that π is identical to the range of μ , then the circular variant of (π, μ) is (π, μ') , where μ' is obtained from μ by connecting each pair of corresponding nodes in the domain and range with a circular edge.*

Definition 5.8 (Forward Positive Cycle) *A circular variant (π, μ) has a forward positive cycle if μ has a positive cycle, such that when this cycle is traversed, each circular edge is traversed from the range to the domain.*

From the basic lemma we get that if there is an infinite branch in the tree, there must be an infinite sequence of nodes N_1, N_2, \dots such that for each i the branch from N_i to N_{i+1} is mapped into the same recursive pair. Suppose the circular variant of this pair has a positive forward cycle. Start with a node on a forward positive cycle, in the domain of the pair corresponding to some call branch, say from N_i to N_{i+1} . Now traverse the cycle, but in the case of a circular edge go from the range of this pair to the domain of the pair for the call branch from N_{i+1} to N_{i+2} . After a number of steps equal to the number of circular arcs on the forward positive cycle we’ll return to the same node in the pair as we started from, only now corresponding to a lower call branch. From the existence of the forward positive cycle we can deduce a decrease in norm. This means that we can find an infinite sequence of arguments of atoms in the tree, such that their norms form a descending sequence of non-negative integers, which is impossible. So we get the following conclusion from the basic lemma

Theorem 5.1 *Let the LD-tree for a query and a program and a symbolic linear norm be given. Define α_{real} as above. If all the circular variants that can be created from the query-mapping pairs in the image of α_{real} have a forward positive cycle, then the tree must be finite, i.e., there is termination for the query with Prolog’s computation rule.*

Again it is easy to find an example of a program that terminates although it does not satisfy the condition of the theorem. Take the program

$$\begin{aligned} & \mathbf{p}(0). \\ & \mathbf{p}(1) \text{ :- } \mathbf{p}(0). \end{aligned}$$

with query pattern $p(f)$ and a norm that assigns the same value to 0 and 1.

Theorem 5.1 does not give us an effective way to determine for a particular program and query if there is termination, because we cannot always construct the LD-tree that may be infinite. The next section gives a way to approximate the ‘real’ query-mapping pairs. The algorithm proposed either says that there is termination, or that it is not strong enough to decide.

5.1 The Query-Mapping Pairs Algorithm

The following algorithm has been implemented in a system called *TermiLog* (cf. [21,22]), which is quite powerful and has been able to analyze correctly 82% of the 120 benchmark programs it was tested on, taken from the literature on termination and other sources. The basic idea of the algorithm is to *approximate* the set of query-mapping pairs that are associated with the LD-tree for a query and program. We will show that each ‘real’ query-mapping pair arising from the LD-tree (i.e. the image under α_{real} of a call branch) is subsumed by a query-mapping pair in the approximation, so that a sufficient condition for the finiteness of the LD-tree is that every circular variant in the approximation has a forward positive cycle.

We will define a structured shadow α_{app} , which is a *widening* (cf. [10]) of α_{real} , by giving its value for call branches between nodes that are direct offspring of each other (the *generation* step) and find its value for other call branches by composition (the *composition* step). It should be noted that in this case α_{app} associates with each call branch a query-mapping pair that depends not only—as in the previous section—on the nodes at the ends of the branch and the substitution associated with them, but also on the location of the branch in the tree. Since we are approximating the ‘real’ pairs our conclusions are sound, but there may be constraints which we have not inferred, so it may happen that the value of α_{app} for two call branches which look identical but are in different parts of the tree will be different.

The first step is constructing from each rule of the program a weighted rule graph, which extracts the information about argument norms that is in the rule.

Definition 5.9 (Weighted Rule Graph) *The weighted rule graph associated with a rule has as nodes all the argument positions of the atoms in the rule; it has edges connecting the nodes of arguments which have equal norm and has a potential weighted arc between any two nodes such that the*

difference between the norms of the respective arguments, which is a linear expression, has non-negative coefficients and a positive constant term, and this potential arc is labeled by the difference.

In our example, using the term-size norm, we get for the rule

$$\text{mult}(s(X), Y, Z) :- \text{mult}(X, Y, Z1), \text{add}(Z1, Y, Z).$$

the weighted rule graph that is shown in Figure 5.5. The potential arc is

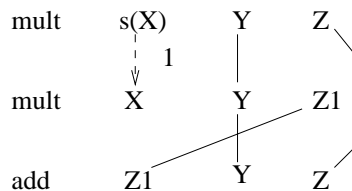


Fig. 5.5 Weighted rule graph

shown by a dashed arc. It should be explained what *potential* arcs are. In the termination proof we use the fact that the order induced by the norm on terms that are instantiated enough is well-founded (recall that for such terms the norm is a non-negative integer). Once we know that the nodes connected by a potential arc are instantiated enough, we connect them with an arc. However, we will not do this when we do not know that the arguments are instantiated enough, because we want to be sure that there cannot be an infinite path consisting of arcs. Consider for example the program

$$\begin{aligned} &\text{int}(0). \\ &\text{int}(s(X)) :- \text{int}(X). \end{aligned}$$

with the query $\text{int}(Y)$ and the term-size norm. From the rule we get the weighted rule graph that is shown in Figure 5.6, but as Figure 5.7 shows there is an infinite derivation.

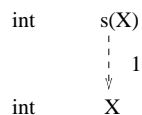


Fig. 5.6 Weighted rule graph

We return now to the original example. We have the query

$$\text{mult}(s(s(0)), s(0), Z)$$

$$\begin{array}{l}
\leftarrow \text{int}(Y) \\
\quad \{Y \mapsto s(Y1)\} \\
\leftarrow \text{int}(Y1) \\
\quad \{Y1 \mapsto s(Y2)\} \\
\leftarrow \text{int}(Y2) \\
\quad \vdots
\end{array}$$

Fig. 5.7 Infinite derivation

which we abstract to the query pattern $\text{mult}(b, b, f)$ with empty constraint list. We now construct the summaries of the augmented argument mappings associated with it, terms that will be defined presently. Our definition of augmented argument mapping differs from that in [29], for reasons that will be explained. The basic idea is that we take a rule r , for which the head predicate is the same as in the query, and a subgoal s in the body of r , and try to approximate the ‘real’ query-mapping pair corresponding to the head of r and s in the LD-tree. We do this by using information we have from the weighted rule graph of r and also, if we have them, results of **instantiation analysis** and **constraint inference** about the instantiations and constraints of the body subgoals preceding s , which we assume have succeeded before we got to s . The method used for the instantiation analysis and constraint inference is abstract interpretation. For the details see [21].

Definition 5.10 (Augmented Argument Mapping) *An augmented argument mapping, which is a mixed graph, is constructed for a rule r and a subgoal s in its body as follows.*

- *There are nodes for the argument positions of the head of r , for s , and for all subgoals that precede s .*
- *Nodes are blackened in agreement with the rule and the instantiation analysis for subgoals that precede s .¹*
- *There are all the edges and arcs that can be derived from the weighted rule graph and from the constraint inference for subgoals preceding s .*
- *In the case of disjunctive information about constraints or instantiations the augmented argument mapping will use one disjunct.*

¹ In [29] nodes that correspond to arguments that precede s are made black. This is justified there because of the assumption that any variable in the head of a program clause also appears in its body, an assumption which causes all atoms in the success set of the program to be ground, but not in our more general setting. Another difference is the use of weighted arcs. For instance, if we have the configuration with edges, arcs and weighted arcs, that is presented in Figure 5.8, we can say that in the transitive closure there is an arc from N_1 to N_5 , while we could not come to this conclusion if the weighted arcs were ordinary arcs.

- The graph is consistent.
- The domain consists of the nodes corresponding to the head of r , and the range consists of the nodes corresponding to s .

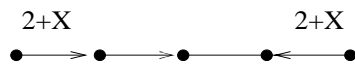


Fig. 5.8 The weighted arcs

A few words about the consistency: If, for instance, we are constructing the augmented argument mapping for the rule

$$\mathbf{a}(X, Y) :- \mathbf{b}(X, Y), \mathbf{c}(X, Y), \mathbf{d}(X, Y).$$

and the subgoal $d(X, Y)$ we cannot use for b the constraint $gt(1, 2)$ and for c the conflicting constraint $gt(2, 1)$, or for b the instantiation $b(ie, nie)$ and for c the conflicting instantiation $c(nie, ie)$ (here ie denotes an argument that is instantiated enough, and nie an argument that is not—note the difference between ie and nie , which are mutually exclusive and b and f , where the second possibility includes the first).

Definition 5.11 (Summary) *If μ is a consistent augmented argument mapping, then the summary of μ consists of the nodes in the domain and range of μ and the edges and arcs among these nodes (it is undefined if μ is inconsistent).*

Summaries of augmented argument mappings give us approximations to ‘real’ mappings constructed for nodes that are direct offspring of each other.

We return now to the query $mult(b, b, f)$ and build the augmented argument mappings derived for it from the rule

$$\mathbf{mult}(\mathbf{s}(X), Y, Z) :- \mathbf{mult}(X, Y, Z1), \mathbf{add}(Z1, Y, Z).$$

It should be noted that when we build the augmented argument mapping for a query we also take into account the instantiations and constraints of the query.

For the first subgoal we get the mapping presented in Figure 5.9. The

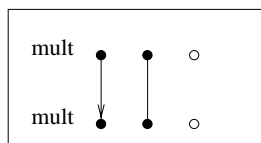


Fig. 5.9 Mapping for the first subgoal

summary is identical to the mapping, and Figure 5.10 presents the query-mapping pair obtained.

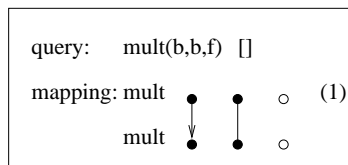


Fig. 5.10 Query-mapping pair for the first subgoal of `mult`

From the augmented argument mapping corresponding to the second subgoal of the rule we basically want to infer the relationship between $mult(s(X), Y, Z)$ and $add(Z1, Y, Z)$ assuming $mult(X, Y, Z1)$ has already been proved. This is where we use instantiation analysis and, possibly, constraint inference. If we did not use any information on $mult(X, Y, Z1)$ we would get the augmented argument mapping presented in Figure 5.11. Now

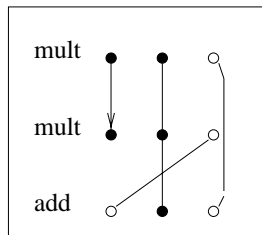


Fig. 5.11 Mapping for the second subgoal

from the instantiation analysis we will get that the two possible instantiations for $mult$ are $mult(ie, ie, ie)$ and $mult(ie, nie, ie)$ where ie denotes a ground term and nie denotes a non-ground term, i.e., a term that contains at least one variable. Since the first two arguments of the intermediate subgoal are ground, so must the third one be. So we get the augmented argument mapping that is presented in Figure 5.12, which gives rise to the query-mapping pair presented in Figure 5.13 and the new query $add(b, b, f)$ (with no constraints). If we had not been able to use the results of the instantiation analysis we would have gotten the query $add(f, b, f)$, which does not terminate, and our algorithm would just have said for the original query that there *may* be non-termination.

Using the appropriate augmented argument mappings for the new query $add(b, b, f)$, we get the new query-mapping pairs (Figures 5.14 and 5.15) and no new queries.

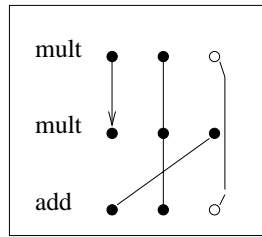


Fig. 5.12 Mapping for the second subgoal

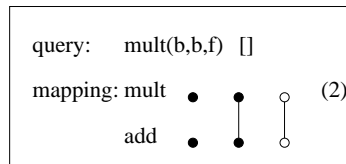


Fig. 5.13 Query-mapping pair for the second subgoal of mult

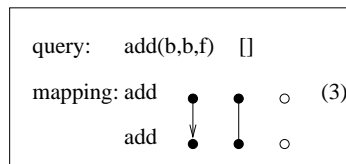


Fig. 5.14 Query-mapping pair for add

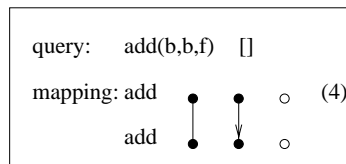


Fig. 5.15 Query-mapping pair for add

Now we have to apply composition to the query-mapping pairs we have created thus far. Recall the following definitions from [29]:

Definition 5.12 (Composition of Mappings) *If the range of a mapping μ and the domain of a mapping ν are labeled by the same predicate, then the composition of the mappings μ and ν , denoted $\mu \circ \nu$, is obtained by unifying each node in the range of μ with the corresponding node in the domain of ν . When unifying two nodes, the result is a black node if at least one of the nodes is black, otherwise it is a white node. If a node becomes black, so do all nodes connected to it with an edge. The domain of $\mu \circ \nu$ is that of μ , and*

its range is that of ν . The edges and arcs of $\mu \circ \nu$ consist of the transitive closure of the union of the edges and arcs of μ and ν .

Definition 5.13 (Composition of Query-Mapping Pairs) *Let (π_1, μ_1) and (π_2, μ_2) be query-mapping pairs, such that the range of μ_1 is identical to π_2 . The composition of (π_1, μ_1) and (π_2, μ_2) is (π_1, μ) , where μ is the summary of $\mu_1 \circ \mu_2$ (and, hence, the composition is undefined if $\mu_1 \circ \mu_2$ is inconsistent).*

By repeatedly composing the approximations we got thus far we get the following new pairs.

Composition of pairs (1) and (2) gives a new pair (5); pairs (3) and (4) give a new pair (6); (2) and (6) give (7). These new pairs are presented by Figures 5.16, 5.17 and 5.18 respectively. No more query-mapping pairs can

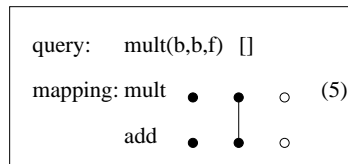


Fig. 5.16 Query-mapping pair for add

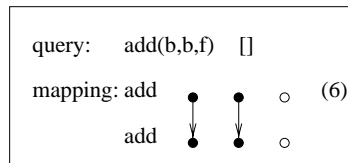


Fig. 5.17 Query-mapping pair for add

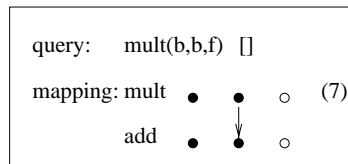


Fig. 5.18 Query-mapping pair for add

be created.

Since for each of the above query-mapping pairs, if the circular variant exists it has a forward positive cycle, it follows, since for each call branch its image under α_{app} subsumes its image under α_{real} , that queries of the form $mult(b, b, f)$ terminate.

If in the above example we would have used results of the constraint inference for the one intermediate subgoal we had, we would have gotten more query-mapping pairs, but again every circular variant would have had a forward positive cycle, so we would have been able to show termination. However, if it is possible to prove termination without constraint inference, there is no reason to use it, because the more query-mapping pairs there are the longer the termination proof takes.

The following theorem holds:

Theorem 5.2 *Let the LD-tree for a query and program and a symbolic linear norm be given. Define the structured shadow α_{app} as above. If all the circular variants that can be created from query-mapping pairs in the image of α_{app} have a forward positive cycle then the tree must be finite, i.e., there is termination for the query with Prolog's computation rule.*

This theorem follows from Theorem 5.1 if we notice that for every call branch B we have that $\alpha_{real}(B)$ is subsumed by $\alpha_{app}(B)$.

Actually, since we have here a structured shadow, the element whose existence is proved in the basic lemma is both circular and idempotent. So we can formulate a stronger theorem, which is more efficient to implement:

Theorem 5.3 *Let the LD-tree for a query and program and a symbolic linear norm be given. Define the structured shadow α_{app} as above. If all the circular idempotent query-mapping pairs in the image of α_{app} have an arc from an argument in the domain to the corresponding argument in the range, then the tree must be finite, i.e., there is termination for the query with Prolog's computation rule.*

It is not difficult to see that the condition of Theorem 5.2 implies the condition of Theorem 5.3, since if we have a circular idempotent pair for which the circular variant has a forward positive cycle, and compose it with itself the right number of times (that is, the number of circular arcs on the forward positive cycle), we get an arc from an argument in the domain to the corresponding argument in the range.

Theorem 5.3 is really stronger than Theorem 5.2, as the following program shows:

$$\begin{aligned} & p(0, 0). \\ & p(s(X), Y) :- p(f(0), X). \\ & p(X, s(Y)) :- p(Y, f(0)). \end{aligned}$$

In this case there is termination for queries of the form $p(b, b)$, but, using the term-size norm, this can only be deduced from Theorem 5.3 and not

Theorem 5.2, because there are circular variants without forward positive cycle.

It is interesting to note that the rule of composition does not hold for α_{real} , so it is not a structured shadow. For instance if we take the program

$$\begin{aligned} p(\mathbf{X}, \mathbf{a}) &:- q(\mathbf{X}). \\ q(\mathbf{X}) &:- r(\mathbf{X}, \mathbf{a}). \end{aligned}$$

and the LD-tree for $p(\mathbf{X}, \mathbf{a})$

$$\begin{aligned} (1) &\leftarrow p(X, a) \\ (2) &\leftarrow q(X) \\ (3) &\leftarrow r(X, a) \end{aligned}$$

and the term-size norm, then the image under α_{real} of the branch from (1) to (2) composed with the image under α_{real} of the branch from (2) to (3) will not contain the edge between the second arguments of p and r that is in the image under α_{real} of the branch between (1) and (3). We will always have that $\alpha_{real}(B_1) * \alpha_{real}(B_2)$ subsumes $\alpha_{real}(B_1 * B_2)$.

The following optimization of Theorem 5.2 holds: It is enough to consider only query-mapping pairs in which the predicate of the domain and the predicate of the range are in the same strongly connected component of the predicate dependency graph.

6 Logic programs containing arithmetic predicates

The algorithm we describe next would come into play only when the usual termination analyzers fail to prove termination using the structural arguments of predicates. As a first step it verifies the presence of an integer loop in the program. If no integer loop is found, the possibility of non-termination is reported, meaning that the termination cannot be proved by this technique. If integer loops are found, each of them is taken into consideration. The algorithm starts by discovering integer positions in the program, proceeds with creating appropriate abstractions, based on the integer loops, and concludes by applying an extension of the query-mapping pairs technique. The formal algorithm is presented in Subsection 6.6.

The structured shadow we define in this case assigns to the branch from node $\leftarrow r_1, \dots, r_n$ to its direct offspring $\leftarrow s_1, \dots, s_k$, where θ is the composition of the substitutions between the nodes, the following query-mapping pair: each atom is abstracted to a pair (*predicate*, *constraint*), where the constraint is one from a finite set of mutually exclusive numerical constraints (for example, $arg1 > 0$, $arg1 > arg2$, where $arg1$ and $arg2$ are respectively the first and the second arguments of the atom). The query is the abstraction of r_1 . The mapping of the query-mapping pair, is as before, a quadruple—the domain, the range, edges and arcs. The domain of the query-mapping

pair is the abstraction of $r_1\theta$, the range is the abstraction of s_1 , and there are edges and arcs between nodes of $r_1\theta$ and s_1 . Edges and arcs correspond to numerical equalities and inequalities of the respective arguments. When composing two query-mapping pairs numerical nodes are unified only if they have the same constraint (remember that the constraints are mutually exclusive). Termination is shown by means of a non-negative termination function of the arguments of an atom, that decreases from the domain to the range (cf. [15]). Note that in the numerical part of the program we will use both the query-mapping pairs relative to the norm and the new kind of numerical query-mapping pairs. As we will see later on, sometimes in order to prove termination (cf. Example 6.12) both kinds of query-mapping pairs are essential.

The technique we present in this section allows us to analyze correctly on the one hand common examples of Prolog programs (such as *factorial* [12], *Fibonacci*, *Hanoi* [7], *odd_even* [28], *between* [2], *Ackermann* [32]), and on the other hand more difficult examples, such as *gcd* and *mc_carthy_91* [25, 20, 16]. Note that some of these examples were previously considered in the literature on termination. However, they were always assumed to be given in the *successor notation*, thus solving the problem of well-foundedness. Moreover, the analysis of some of these examples, such as *gcd*, required special techniques [23].

6.1 The 91 function

We start by illustrating informally the use of our algorithm for proving the termination of the 91 function. This convoluted function was invented by John McCarthy for exploring properties of recursive programs, and is considered to be a good test case for automatic verification systems (cf. [25, 20, 16]). The treatment here is on the intuitive level. Formal details will be given in subsequent sections.

Consider the clauses:

Example 6.1

$$\begin{aligned} \text{mc_carthy_91}(X, Y) &:- X > 100, Y \text{ is } X - 10. \\ \text{mc_carthy_91}(X, Y) &:- X \leq 100, Z \text{ is } X + 11, \text{mc_carthy_91}(Z, Z1), \\ &\quad \text{mc_carthy_91}(Z1, Y). \end{aligned}$$

and assume that a query of the form $\text{mc_carthy_91}(i, f)$ is given, that is, a query in which the first argument is bound to an integer, and the second is free. This program computes the same answers as the following one:

$$\begin{aligned} \text{mc_carthy_91}(X, Y) &:- X > 100, Y \text{ is } X - 10. \\ \text{mc_carthy_91}(X, 91) &:- X \leq 100. \end{aligned}$$

with the same query. Note, however, that while the termination of the latter program is obvious, since there is no recursion in it, the termination of the

first one is far from being trivial and a lot of effort was dedicated to find termination proofs for it ([25, 20, 16]).

Our algorithm starts off by **discovering numerical arguments**. This step is based on abstract interpretation, and as a result both arguments of `mc_carthy_91` are proven to be numerical. Moreover, they are proven to be of integer type. The importance of knowledge of this kind and techniques for its discovery are discussed in Subsection 6.2.2.

The next step of the algorithm is the inference of the (finite) **integer abstraction domain** which will help overcome difficulties caused by the fact that the (positive and negative) integers with the usual (greater-than or less-than) order are not well-founded. Integer abstractions are derived from arithmetic comparisons in the bodies of rules. However, a simplistic approach may be insufficient and the more powerful techniques presented in Section 6.3 are sometimes essential. In our case the domain

$$\{(-\infty, 89], [90, 100], [101, \infty)\}$$

of intervals is deduced. For the sake of convenience we denote this tripartite domain by $\{small, med, big\}$.

In the next step, we **use abstract interpretation to describe answers to queries**. This allows us to infer numerical inter-argument relations of a novel type. In Section 6.4 the technique for inference of constraints of this kind is presented. For our running example we get the following abstract atoms:

$$\begin{array}{ll} mc_carthy_91(big, big) & mc_carthy_91(med, med) \\ mc_carthy_91(big, med) & mc_carthy_91(small, med) \end{array}$$

These abstract atoms characterize the answers of the program.

The concluding step creates **query-mapping pairs** in the fashion of [21]. This process uses the abstract descriptions of answers to queries and is described in Section 6.5. In our case, we obtain among others, the query-mapping pair having the query `mc_carthy_91(i, f)`, where i denotes an integer argument and f an unrestricted one, and the mapping presented in Figure 6.1. The upper nodes correspond to argument positions of the head of the recursive clause, and the lower nodes—to argument positions of the second recursive subgoal in the body. Circled black nodes denote integer argument positions, and white nodes denote positions on which no assumption is made. The arc denotes an increase of the first argument, in the sense that the first argument in the head is less than the first argument in the second recursive subgoal. Each set of nodes is accompanied by a set of constraints. Some could be inter-argument relations of the type considered in [21]. In our example this subset is empty. The rest are constraints based on the integer abstraction domain. In this case, that set contains the constraint that the first argument is in *med*. The query-mapping pair presented is circular (upper and lower nodes are the same), but the termination tests of [21, 9] fail. Thus, a termination function must be guessed. For this loop we can use the

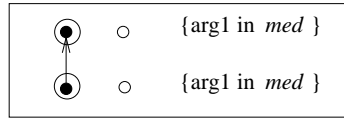


Fig. 6.1 Mapping for McCarthy’s 91 function

function $100 - \text{arg1}$, where arg1 denotes the first argument of the atom. The value of this function decreases while traversing the given query-mapping pair from the upper to the lower nodes. Since it is also bounded from below ($100 \geq \text{arg1}$), this query-mapping pair may be traversed only finitely many times. The same holds for the other circular query-mapping pair in this case. Thus, termination is proved.

6.2 Arithmetic Loops

We start our discussion on termination of numerical computations by providing a formal definition of numerical loop, analyzing the problems one discovers when reasoning about termination of numerical loops and explain why we restrict ourselves to integer loops. In the end of this section we discuss a technique for discovering numerical argument positions that we’ll base our termination analysis on.

6.2.1 Numerical and integer loops Our notion of numerical loop is based on the predicate dependency graph (cf. [28]):

Definition 6.1 *Let P be a program and let Π be a strongly connected component in its predicate dependency graph. Let $S \subseteq P$ be the set of program clauses, associated with Π (i.e. those clauses that have the predicates of Π in their head). S is called loop if there is a cycle through predicates of Π .*

Definition 6.2 *A loop S is called numerical if there is a clause*

$$H :- B_1, \dots, B_n$$

in S , such that for some i , $B_i \equiv \text{Var is Exp}$, and either Var is equal to some argument of H or Exp is an arithmetic expression involving a variable that is equal to some argument of H .

However, termination of numerical loops that involve numbers that are not integers often depends on the specifics of implementation and hardware, so we limit ourselves to “integer loops”, rather than all numerical loops. The following examples illustrate actual behavior that contradicts intuition—a loop that should not terminate terminates, while a loop that should terminate does not. We checked the behavior of these examples on UNIX, with the CLP(Q,R) library [18] of SICStus Prolog [31], CLP(R) [19] and XSB [30].

Example 6.2 Consider the following program. The goal $p(1.0)$ terminates although we would expect it not to terminate. On the other hand the goal $q(1.0)$ does not terminate, although we would expect it to terminate.

$$\begin{aligned} p(0.0) &:- !. \\ p(X) &:- X1 \text{ is } X/2, p(X1). \\ \\ q(0.0) &:- !. \\ q(X) &:- X1 \text{ is } X - 0.1, q(X1). \end{aligned}$$

One may suggest that assuming that the program does not contain division and non-integer constants will solve the problem. The following example shows that this is not the case:

Example 6.3

$$\begin{aligned} r(0). \\ r(X) &:- X > 0, X1 \text{ is } X - 1, r(X1). \end{aligned}$$

The predicate r may be called with a real, non-integer argument, and then its behavior is implementation dependent. For example, one would expect that $r(0.0)$ will succeed and $r(0.000000001)$ will fail. However, in SICStus Prolog both goals fail, while in CLP(R) both of them succeed!

Therefore, we limit ourselves to integer loops, that is numerical loops involving integer constants and arithmetical calculations over integers:

Definition 6.3 *A program P is integer-based if, given a query such that all numbers appearing in it are integers, all subqueries that arise have this property as well.*

Definition 6.4 *A numerical loop S in a program P is called an integer loop if P is integer-based.*

Termination of a query may depend on whether its argument is an integer, as the following example shows:

$$\begin{aligned} p(0). \\ p(N) &:- N > 0, N1 \text{ is } N - 1, p(N1). \\ p(a) &:- p(a). \end{aligned}$$

For this program, $p(X)$ for integer X terminates, while $p(a)$ does not.

So we extend our notion of query pattern. Till now a query pattern was an atom with the same predicate and arity as the query, and arguments b (denoting an argument that is instantiated enough with respect to the norm) or f (denoting an argument on which no assumptions are made). Here, we extend the notion to include arguments of the form i , denoting an argument that is an integer (or integer expression). Note that b includes the possibility of i in the same way that f includes the possibility b . In the

diagrams to follow we denote i -arguments by circled black nodes, and as before, b -arguments by black nodes and f -arguments by white nodes.

Our termination analysis is always performed with respect to a given program and a query pattern. A positive response guarantees termination of every query that matches the pattern.

6.2.2 Discovering integer arguments Our analysis that will be discussed in the subsequent sections is based on the size relationships between “integer arguments”. So we have to discover which arguments are integer arguments. In simple programs this is immediate, but there may be more complicated cases.

The inference of integer arguments is done in two phases—bottom-up and top-down. Bottom-up inference is similar to type analysis (cf. [8, 5]), using the abstract domain $\{int, not_int\}$ and the observation that an argument may become int only if it is obtained from $is/2$ or is bound to an integer expression of arguments already found to be int (i.e. the abstraction of $int + int$ is int). Top-down inference is query driven and is similar to the “blackening” process, described in [21], only in this case the information propagated is being an integer expression instead of “instantiated enough”.

Take for example the program

```
p(0).
p(N) :- N > 0, N1 is N - 1, p(N1).
p(a) :- p(a).
q(X) :- r(X,Y), p(Y).
r(b,a).
r(X,X).
```

Denoting by int an integer argument, by gni an argument that is ground but not integer, and by ng an argument that is not ground, we get from bottom-up instantiation analysis that the only pattern possible for $r(X, Y)$ atoms that are logical consequences of the program are

$$r(int, int), r(gni, gni), r(ng, ng)$$

Now we get from top down analysis that a query $q(i)$ gives rise to the query $p(i)$ and hence terminates.

The efficiency of discovering numerical arguments may be improved by a preliminary step of guessing the numerical argument positions. The guessing is based on the knowledge that variables appearing in comparisons or $is/2$ -atoms should be numerical. Instead of considering the whole program it is sufficient in this case to consider only clauses of predicates having clauses with the guessed arguments and clauses of predicates on which they depend. The guessing as a preliminary step becomes crucial when considering “real-world” programs that are large, while their numerical part is usually small.

6.3 Integer Abstraction Domain

In this subsection we present a technique that allows us to overcome the difficulties caused by the fact that the integers with the usual order are not well-founded. Given a program P we introduce a finite abstraction domain, representing integers. The integer abstractions are derived from the subgoals involving integer arithmetic positions.

Let S be a set of clauses in P , consisting of an integer loop and all the clauses for predicates on which the predicates of the integer loop depend. As a first step for defining the abstract domain for each recursive predicate p in S we obtain the set of comparisons \mathcal{C}_p . If p is clear from the context we omit the index.

More formally, we consider as a *comparison*, an atom of the form $t_1 \rho t_2$, such that t_1 and t_2 are either variables or constants and $\rho \in \{<, \leq, \geq, >\}$. Our aim in restricting ourselves to these atoms is to ensure the finiteness of \mathcal{C} . Other decisions can be made as long as finiteness is ensured. Note that by excluding \neq and $=$ we do not limit the generality of the analysis. Indeed if $t_1 \neq t_2$ appears in a clause it may be replaced by two clauses containing $t_1 > t_2$ and $t_1 < t_2$ instead of $t_1 \neq t_2$, respectively. Similarly, if the clause contains a subgoal $t_1 = t_2$, the subgoal may be replaced by two subgoals $t_1 \geq t_2, t_1 \leq t_2$. Thus, the equalities we use in the examples to follow should be seen as a brief notation as above.

In the following subsections we present a number of techniques to infer \mathcal{C} from the clauses of S .

We define \mathcal{D}_p as the set of pairs (p, c) , for all satisfiable $c \in 2^{\mathcal{C}_p}$. Here we interpret $c \in 2^{\mathcal{C}_p}$ as a conjunction of the comparisons in c and the negations of the comparisons in $\mathcal{C}_p \setminus c$. The abstraction domain \mathcal{D} is taken as the union of the sets \mathcal{D}_p for the recursive predicates p in S . Simplifying the domain may improve the running time of the analysis, however it may make it less precise.

6.3.1 The simple case—collecting comparisons The simplest way to obtain \mathcal{C} from the clauses of S is to consider the comparisons appearing in the bodies of recursive clauses and restricting integer positions in their heads (we limit ourselves to the recursive clauses, since these are the clauses that can give rise to circular pairs).

We would like to view \mathcal{C} as a set of comparisons of head argument positions. Therefore we assume in the simple case that S is *partially normalized*, that is, all head *integer* argument positions in clauses of S are occupied by distinct variables. This assumption holds for all the examples considered so far. This assumption will not be necessary with the more powerful technique presented in the next subsection.

Example 6.4 Consider

$$t(X) :- X > 5, X < 8, X < 2, X1 \text{ is } X + 1, X1 < 5, t(X1).$$

Let $\mathbf{t}(i)$ be a query pattern for the program above. In this case, the first argument of \mathbf{t} is an integer argument. Since $\mathbf{X1}$ does not appear in the head of the first clause $\mathbf{X1} < 5$ is not considered and, thus, $\mathcal{C} = \{X > 5, X < 8, X < 2\}$. We have in this example only one predicate and the union is over the singleton set. So, $\mathcal{D} = \{X < 2, 2 \leq X \leq 5, 5 < X < 8, X \geq 8\}$.

The following example evaluates the *mod* function.

Example 6.5

$$\begin{aligned} \text{mod}(\mathbf{A}, \mathbf{B}, \mathbf{C}) &:- \mathbf{A} \geq \mathbf{B}, \mathbf{B} > 0, \mathbf{D} \text{ is } \mathbf{A} - \mathbf{B}, \text{mod}(\mathbf{D}, \mathbf{B}, \mathbf{C}). \\ \text{mod}(\mathbf{A}, \mathbf{B}, \mathbf{C}) &:- \mathbf{A} < \mathbf{B}, \mathbf{A} \geq 0, \mathbf{A} = \mathbf{C}. \end{aligned}$$

Here we ignore the second clause since it is not recursive. Thus, by collecting comparisons from the first clause, $\mathcal{C}_{\text{mod}} = \{arg1 \geq arg2, arg2 > 0\}$ and thus, by taking all the conjunctions of comparisons of \mathcal{C} and their negations, we obtain $\mathcal{D}_{\text{mod}} = \{(\text{mod}, arg1 \geq arg2 \ \& \ arg2 > 0), (\text{mod}, arg1 \geq arg2 \ \& \ arg2 \leq 0), (\text{mod}, arg1 < arg2 \ \& \ arg2 > 0), (\text{mod}, arg1 < arg2 \ \& \ arg2 \leq 0)\}$.

However, sometimes the abstract domain obtained in this way is insufficient for proving termination, and thus, should be refined. The domain may be refined by enriching the underlying set of comparisons. Possible ways to do this are using inference of comparisons instead of collecting them, or performing an unfolding, and applying the collecting or inference techniques to the unfolded program.

6.3.2 Inference of Comparisons As mentioned above, sometimes the abstraction domain obtained from comparisons appearing in S is insufficient. Instead of collecting comparisons, appearing in bodies of clauses, we collect certain comparisons that are *implied* by bodies of clauses. For example, $\mathbf{X} \text{ is } \mathbf{Y} + \mathbf{Z}$ implies the constraint $X = Y + Z$ and $\text{functor}(\text{Term}, \text{Name}, \text{Arity})$ implies $\text{Arity} \geq 0$.

As before, we restrict ourselves to recursive clauses and comparisons that constrain integer argument positions of heads. Since a comparison that is contained in the body is implied by it, we always get a superset of the comparisons obtained by the collecting technique, presented previously. The set of comparisons inferred depends on the power of the inference engine used (e.g. CLP-techniques may be used for this purpose).

We define the abstract domain \mathcal{D} as above. Thus, the granularity of the abstract domain also depends on the power of the inference engine.

6.3.3 Unfolding Unfolding (cf. [33, 4, 2, 23]) allows us to generate a sequence of abstract domains, such that each refines the previous.

More formally, let P be a program and let $H :- B_1, \dots, B_n$ be a recursive rule in P . Let P_1 be the result of unfolding an atom B_i in $H :- B_1, \dots, B_n$ in P . Let S_1 be a set of clauses in P_1 , consisting of an integer loop and the clauses of the predicates on which the integer loop predicates depend.

Obtain \mathcal{D} for the clauses of S_1 either by collecting comparisons from rule bodies or by inferring them, and use it as a new abstraction domain for the original program. If the algorithm still fails to prove termination, the process of unfolding can be repeated.

Example 6.6 Unfolding `mc_carthy_91(Z1,Z2)` in the recursive clause we obtain a new program for the query `mc_carthy_91(i,f)`

```
mc_carthy_91(X,Y) :- X > 100, Y is X - 10.
mc_carthy_91(X,Y) :- X ≤ 100, Z1 is X + 11, Z1 > 100,
                    Z2 is Z1 - 10, mc_carthy_91(Z2,Y).
mc_carthy_91(X,Y) :- X ≤ 100, Z1 is X + 11, Z1 ≤ 100, Z3 is Z1 + 11,
                    mc_carthy_91(Z3,Z4), mc_carthy_91(Z4,Z2),
                    mc_carthy_91(Z2,Y).
```

Now if we use an inference engine that is able to discover that `X is Y+Z` implies the constraint $X=Y+Z$, we obtain the following constraints on the bound head integer variable X (for convenience we omit redundant ones): From the second clause we obtain: $X \leq 100$, and since $X + 11 > 100$ we get $X > 89$. Similarly, from the third clause: $X \leq 89$. Thus, $\mathcal{C} = \{X \leq 89, X > 89 \wedge X \leq 100\}$ Substituting this in the definition of \mathcal{D} , and removing inconsistencies and redundancies, we obtain $\mathcal{D} = \{X \leq 89, X > 89 \wedge X \leq 100, X > 100\}$.

6.3.4 Propagating domains The comparisons we obtain by the techniques presented above may restrict only some subset of integer argument positions. However, for the termination proof, information on integer arguments outside of this subset may be needed. For example, as we will see shortly, in order to analyze correctly `mc_carthy_91` we need to determine the domain for the second argument, while the comparisons we have constrain only the first one. Thus, we need some technique of *propagating* abstraction domains that we obtained for one subset of integer argument positions to another subset of integer argument positions. Clearly, this technique may be seen as a heuristic and it is inapplicable if there is no interaction between argument positions.

To capture this interaction we draw a graph for each recursive numerical predicate, that has the numerical argument positions as vertices and edges between vertices that can influence each other. In the case of `mc_carthy_91` we get the graph having an edge between the first argument position and the second one.

Let π be a permutation of the vertices of a connected component of this graph. Define $\pi\mathcal{D}$ to be the result of replacing each occurrence of arg_i in \mathcal{D} by $arg_{\pi(i)}$. Consider the Cartesian product of all abstract domains $\pi\mathcal{D}$ thus obtained, discarding unsatisfiable conjunctions. We will call this Cartesian product the *extended domain* and denote it by \mathcal{ED} . In the case of

`mc_carthy_91` we get as \mathcal{ED} the set of elements `mc_carthy_91(A, B)`, such that A and B are in $\{\textit{small}, \textit{med}, \textit{big}\}$.

More generally, when there are arithmetic relations (e.g. `Y is X+1`) between argument positions, \mathcal{ED} can contain new subdomains that can be inferred from those in \mathcal{D} .

6.4 Abstract interpretation

In this section we use the integer abstractions obtained earlier to classify, in a finite fashion, all possible answers to queries. This analysis can be skipped in simple cases (just as in TermiLog constraint inference can be skipped when not needed), but is necessary in more complicated cases, like `mc_carthy_91`. Most examples encountered in practice do not need this analysis.

The basic idea is as follows: define an abstraction domain and perform a bottom-up constraints inference.

The abstraction domain that should be defined is a refinement of the abstraction domain we defined in Subsection 6.3. There we considered only recursive clauses, since non-recursive clauses do not affect the query-mapping pairs. On the other hand, when trying to infer constraints that hold for answers of the program we should consider non-recursive clauses as well. In this way using one of the techniques presented in the previous subsection both for the recursive and the non-recursive clauses an abstraction domain $\tilde{\mathcal{D}}$ is obtained. Clearly, $\tilde{\mathcal{D}}$ is a refinement of \mathcal{D} .

Example 6.7 For `mc_carthy_91` we obtain that the elements of $\tilde{\mathcal{D}}$ are the intersections of the elements in \mathcal{ED} (see the end of Subsection 6.3.4) with the constraint in the non-recursive clause and its negation.

Example 6.8 Continuing the `mod`-example we considered in Example 6.5 and considering the non-recursive clause for `mod` as well, we obtain by collecting comparisons $\tilde{\mathcal{C}} = \{\textit{arg1} \geq \textit{arg2}, \textit{arg2} > 0, \textit{arg1} < \textit{arg2}, \textit{arg3} < \textit{arg2}, \textit{arg1} \geq 0, \textit{arg1} \leq \textit{arg3}, \textit{arg1} \geq \textit{arg3}\}$ and, thus, $\tilde{\mathcal{D}}$ consists of all pairs (\textit{mod}, c) for c a satisfiable element of $2^{\tilde{\mathcal{C}}}$.

Given a program P , let \mathcal{B} be the corresponding extended Herbrand base, where we assume that arguments in numerical positions are integers. Let T_P be the immediate consequence operator. Consider the Galois connection provided by the abstraction function $\alpha : \mathcal{B} \rightarrow \tilde{\mathcal{D}}$ and the concretization function $\gamma : \tilde{\mathcal{D}} \rightarrow \mathcal{B}$ defined as follows: The abstraction α of an element in \mathcal{B} is the pair from $\tilde{\mathcal{D}}$ that characterizes it. The concretization γ of an element in $\tilde{\mathcal{D}}$ is the set of all atoms in \mathcal{B} that satisfy it. Note that α and γ form a Galois connection due to the disjointness of the elements of $\tilde{\mathcal{D}}$.

Using the Fixpoint Abstraction Theorem (cf. [10]) we get that

$$\alpha \left(\bigcup_{n=1}^{\infty} T_P^n(\emptyset) \right) \subseteq \bigcup_{n=1}^{\infty} (\alpha \circ T_P \circ \gamma)^n(\emptyset)$$

We will take a map $w : \tilde{\mathcal{D}} \rightarrow \tilde{\mathcal{D}}$, that is a *widening* [10] of $\alpha \circ TP \circ \gamma$ and compute its fixpoint. Because of the finiteness of $\tilde{\mathcal{D}}$ this fixpoint may be computed bottom-up.

The abstraction domain $\tilde{\mathcal{D}}$ describes all possible atoms in the extended Herbrand base \mathcal{B} . However, it is sufficient for our analysis to describe only computed answers of the program, i.e., a subset of \mathcal{B} . Thus, in practice, the computation of the fixpoint can sometimes be simplified as follows: We start with the constraints of the non-recursive clauses. Then we repeatedly apply the recursive clauses to the set of the constraints obtained thus far, but abstract the conclusions to elements of \mathcal{D} . In this way we obtain a CLP program that is an abstraction of the original one. This holds in the next example. The abstraction corresponding to the predicate p is denoted p_w .

Example 6.9 Consider once more `mc_carthy_91`. As claimed above we start from the non-recursive clause, and obtain that

$$\text{mc_carthy_91}_w(A, B) :- \{A > 100, B = A - 10\}.$$

By substituting in the recursive clause of `mc_carthy_91` we obtain the following

$$\begin{aligned} \text{mc_carthy_91}(X, Y) :- & X \leq 100, Z1 \text{ is } X + 11, Z1 > 100, \\ & Z2 \text{ is } Z1 - 10, Z2 > 100, Y \text{ is } Z2 - 10. \end{aligned}$$

By simple computation we discover that in this case X is 100, and Y is 91. However, in order to guarantee the termination of the inference process we do not infer the precise constraint $\{X = 100, Y = 91\}$, but its abstraction, i.e., an atom `mc_carthy_91_w(med, med)`. Repeatedly applying the procedure described, we obtain an additional answer `mc_carthy_91_w(small, med)`.

More formally, the following SICStus Prolog CLP(R) program performs the bottom-up construction of the abstracted program, as described above. We use the auxiliary predicate `in/2` to denote a membership in \mathcal{D} and the auxiliary predicate `e_in/2` to denote a membership in the extended domain \mathcal{ED} .

```

:- use_module(library(clpr)).
:- use_module(library(terms)).
:- dynamic(mc_carthy_91_w/2).

in(X, big) :- {X > 100}.
in(X, med) :- {X > 89, X ≤ 100}.
in(X, small) :- {X ≤ 89}.

e_in((X, Y), (XX, YY)) :- in(X, XX), in(Y, YY).

```

```

mc_carthy_91_w(X, Y) :- {X > 100, Y = X - 10}.

assert_if_new((H :- B)) :- \+ (clause(H1, B1),
                               unify_with_occurs_check((H, B), (H1, B1))),
               assert((H :- B)).

deduce :- {X ≤ 100, Z = X + 11}, mc_carthy_91_w(Z, Z1),
          mc_carthy_91_w(Z1, Y), e_in((X, Y), (XX, YY)),
          assert_if_new((mc_carthy_91_w(A, B) :- e_in((A, B), (XX, YY)))),
          deduce.

deduce.

```

The resulting abstracted program is

```

mc_carthy_91_w(A, B) :- {A > 100, B = A - 10}.
mc_carthy_91_w(A, B) :- e_in((A, B), (med, med)).
mc_carthy_91_w(A, B) :- e_in((A, B), (small, med)).

```

Since we assumed that the query was of the form $\text{mc_carthy_91}(i, f)$ we can view these abstractions as implications of constraints like $\text{arg1} \leq 89$ implies $89 < \text{arg2} \leq 100$. We also point out that the resulting abstracted program coincides with the results obtained by the theoretic reasoning above.

As an additional example consider the computation of the *gcd* according to Euclid's algorithm. Proving termination is not trivial, even if the successor notation is used. In [23] only applying a special technique allowed to do this.

Example 6.10 Consider the following program and the query $\text{gcd}(i, i, f)$.

```

gcd(X, 0, X) :- X > 0.
gcd(X, Y, Z) :- Y > 0, mod(X, Y, U), gcd(Y, U, Z).

mod(A, B, C) :- A ≥ B, B > 0, D is A - B, mod(D, B, C).
mod(A, B, C) :- A < B, A ≥ 0, A = C.

```

In this example we have two nested integer loops represented by the predicates `mod` and `gcd`. We would like to use the information obtained from the abstract interpretation of `mod` to find the relation between the `gcd`-atoms in the recursive clause. Thus, during the bottom-up inference process we abstract the conclusions to elements of \tilde{D}_{mod} , as it was evaluated in Example 6.8. Using this technique we get that if $\text{mod}(X, Y, Z)$ holds then always $Z < Y$ holds, and this is what is needed to prove the termination of *gcd*.

6.5 Query-mapping pairs

In this subsection we extend the query-mapping pairs technique to programs having numerical arguments. We assume that a norm is defined for all arguments.

We start with the construction of the original query-mapping pairs, but for atoms (in the query, domain or range) that are part of integer loops we also add the appropriate numerical constraints from the integer abstraction domain (remember that there is only a finite number of elements in the integer abstraction domain).

We also add numerical arcs and edges between numerical argument positions. These arcs and edges are added if numerical inequalities and equalities between the arguments can be deduced. Deduction of numerical edges and arcs is usually done by considering the clauses. However, if a subquery q unifies with a head of a clause of the form $A :- B_1, \dots, B_k, \dots, B_n$ and we want to know the relation between q and B_k (under appropriate substitutions), we *may* use the results of the abstract interpretation to conclude numerical constraints for B_1, \dots, B_{k-1} . The reason is that if we arrive at B_k , this means that we have proved B_1, \dots, B_{k-1} (under appropriate substitutions). All query-mapping pairs deduced in this way are then repeatedly composed. The process terminates because there is a finite number of query-mapping pairs.

A query-mapping pair is called *circular* if the query coincides with the range. The initial query terminates if for every circular query-mapping pair one of the following conditions holds:

- The circular pair meets the requirements of the termination test of Theorem 5.2.
- There is a non-negative termination function for which we can prove a decrease from the domain to the range using the numerical edges and arcs and the constraints of the domain and range.

Two questions remain: how does one automate the guessing of the function, and how does one prove that it decreases. Our heuristic for guessing a termination function is based on the inequalities appearing in the abstract constraints. Each inequality of the form $Exp1 \rho Exp2$ where ρ is one of $\{>, \geq\}$ suggests a function $Exp1 - Exp2$.

The common approach to termination analysis is to find *one* termination function that decreases over all possible execution paths. This leads to complicated termination functions. Our approach allows one to guess a number of relatively simple termination functions, each suited to its query-mapping pair. When termination functions are simple to find, the guessing process can be performed automatically.

After the termination function is guessed, its decrease must be proved. Let V_1, \dots, V_n denote numerical argument positions in the domain and U_1, \dots, U_n the corresponding numerical argument positions in the range of the query-mapping pair. First, edges of the query-mapping pair are translated to equalities and arcs, to inequalities between these variables. Second,

the atom constraints for the V 's and for the U 's are added. Third, let φ be a termination function. We would like to check that $\varphi(V_1, \dots, V_n) > \varphi(U_1, \dots, U_n)$ is implied by the constraints. Thus, we add the negation of this claim to the collection of the constraints and check for unsatisfiability. Since termination functions are linear, CLP-techniques, such as CLP(R) [19] and CLP(Q,R) [18], are robust enough to obtain the desired contradiction. Note however, that if more powerful constraints solvers are used, non-linear termination functions may be considered.

To be more concrete:

Example 6.11 Consider the following program with query $p(i, i)$.

```

p(0, -).
p(X, Y) :- X > 0, X < Y, X1 is X + 1, p(X1, Y).
p(X, Y) :- X > 0, X ≥ Y, X1 is X - 5, Y1 is Y - 1, p(X1, Y1).

```

We get, among others, the circular query-mapping pair having the query $(p(i, i), \{arg1 > 0, arg1 < arg2\})$ and the mapping given in Figure 6.2. The termination function derived for the circular query-mapping pair is $arg2 - arg1$. In this case, we get from the arc and the edge the constraints: $V_1 < U_1, V_2 = U_2$. We also have that $V_1 > 0, U_1 > 0, V_1 < V_2, U_1 < U_2$. We would like to prove that $V_2 - V_1 > U_2 - U_1$ is implied. Thus, we add $V_2 - V_1 \leq U_2 - U_1$ to the set of constraints and CLP-tools easily prove unsatisfiability, and thus, that the required implication holds.

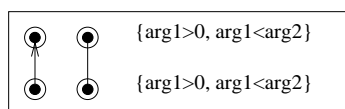


Fig. 6.2 Mapping for p

In the case of the 91-function the mappings are given in Figure 6.3. (We omit the queries from the query-mapping pairs, since they are identical to the corresponding domains.)

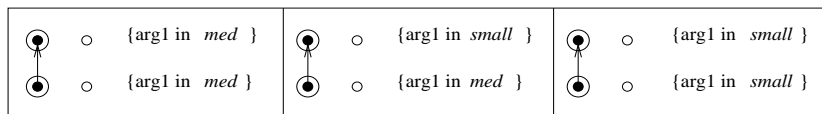


Fig. 6.3 Mappings for McCarthy's 91 function

In the examples above there were no interargument relations of the type considered in [21]. However, this need not be the case in general.

Example 6.12 Consider the following program with the query $q(b, b, i)$ and the term-size norm.

- (1) $q(s(X), X, -)$.
- (2) $q(s(X), X, N) :- N > 0, N1 \text{ is } N - 1, q(s(X), X, N1)$.
- (3) $q(s(s(X)), Z, N) :- N = 0, N1 \text{ is } N - 1, q(s(X), Y, N1), q(Y, Z, N1)$.

Note that constraint inference is an essential step for proving termination—in order to infer that there is a norm decrease in the first argument between the head of (3) and the second recursive call (i.e. $\|s(s(X))\| > \|Y\|$), one should infer that the second argument in q is less than the first with respect to the norm (i.e. $\|s(X)\| > \|Y\|$). We get among others circular query-mapping pairs having the mappings presented in Figure 6.4. The queries of the mappings coincide with the corresponding domains. In the first mapping termination follows from the decrease in the third argument and the termination function $\text{arg3} > 0$. In the second mapping termination follows from the norm decrease in the first argument.

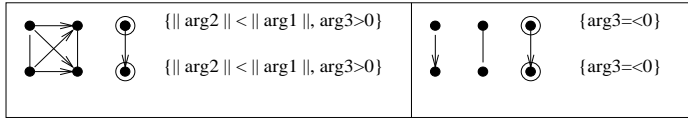


Fig. 6.4 Mappings for q

6.6 The Extended Algorithm

In this section we combine all the techniques suggested so far. The complete algorithm `Analyze_Termination` is presented in Figure 6.5. Each step corresponds to one of the previous sections.

Note that Step 3, computing the abstractions of answers to queries, is optional. If the algorithm returns `NO` it may be re-run either with Step 3 included or with a different integer abstraction domain.

The `Analyze_Termination` algorithm is sound:

Theorem 6.1 *Let P be a program and q a query pattern.*

- `Analyze_Termination`(P, q) *terminates.*
- *If `Analyze_Termination`(P, q) reports YES then, for every query Q matching the pattern q , the LD-tree of Q w.r.t. P is finite.*

Work is being done now to implement the ideas in this section, and thus to be able to deal with programs for which termination depends on the behavior of arithmetic predicates.

Algorithm	Analyze_Termination
Input	A query pattern q and a Prolog program P
Output	YES, if termination is guaranteed NO, if no termination proof was found
(1)	Guess and verify numerical argument positions;
(2)	Compute the integer abstraction domain;
(3)	Compute abstractions of answers to queries (optional);
(4)	Compute ordinary and numerical query-mapping pairs;
(5)	For each circular query-mapping pair do :
(6)	If its circular variant has a forward positive cycle then
(7)	Continue ;
(8)	If the query-mapping pair is numerical then
(9)	Guess bounded termination function;
(10)	Traverse the query-mapping pair and compute values of the termination function;
(11)	If the termination function decreases monotonically then
(12)	Continue ;
(13)	Return NO;
(14)	Return YES.

Fig. 6.5 Termination Analysis Algorithm

7 Conclusion and Generalizations

We have seen the usefulness of the query mapping-pairs approach for proving termination of queries to logic programs by using symbolic linear norm relations between arguments and also by comparing numerical arguments.

In the query-mapping pairs method as outlined above there are two crucial elements:

1. There is a finite number of abstractions of atoms in subgoals of the LD-tree. This ensures that \mathcal{A} is finite.
2. Arcs represent an order.

This suggests two directions for generalization—using different abstractions and using different orders.

7.1 Using Different Abstractions of Terms and a Linear Norm

We can use the original query-mapping pairs as before with the only difference that we'll abstract nodes not to just black and white ones but to a larger, though finite, set. For instance if we have a program

$$\begin{aligned} p(1) &:- \{\text{infinite loop}\}. \\ p(0) &.\end{aligned}$$

and take the term-size norm and a query $p(\text{bound})$, the query-mapping algorithm will say that there may be non-termination. However, we can use the abstractions $1, g, f$, where g means any ground term that is not 1 and f means any term, and apply the above algorithm, with the only difference being in the unification of the abstractions. In the algorithm in Subsection 5.1 we used unification of abstractions in two places—when adjusting the weighted rule graph to the instantiation pattern of the query and when composing query-mapping pairs. In those cases the result of the unification of two nodes of which at least one was black resulted in a black node, and the unification of two white nodes resulted in a white node. In the present case g and 1 will not unify so we will be able to prove that a query $p(g)$ terminates.

Observation 7.1 *The original query-mapping pairs algorithm remains valid if we abstract arguments of atoms in the LD-tree to elements of any finite set of abstractions, as long as we include a sound procedure for unification of these abstractions.*

7.2 Using Norms that Involve Ordinal Numbers

There are programs for which the use of linear norms is not sufficient. The following program performs repeated differentiation.

```

d(deriv(t), 1).
d(deriv(A), 0) :- number(A).
d(deriv(X + Y), L + M) :- d(deriv(X), L), d(deriv(Y), M).
d(deriv(X * Y), (X * L + Y * M)) :- d(deriv(X), M), d(deriv(Y), L).
d(deriv(deriv(X)), L) :- d(deriv(X), M), d(deriv(M), L).

```

In this case one can show that for no choice of constants in the definition of the linear norm will it be possible to prove termination of $d(\text{ground}, \text{free})$. However, we can use the query-mapping pairs method with the abstraction of arguments to ground and non-ground, but use a norm that associates with each term an ordinal number in the following way (ω denotes, of course, the first infinite ordinal):

$$\begin{aligned} \|\text{deriv}(X)\| &= \omega + \|X\| \\ \|X + Y\| &= \|X\| \oplus \|Y\| + 2 \\ \|X * Y\| &= \|X\| \oplus \|Y\| + 2 \end{aligned}$$

where $(n_1\omega + k_1) \oplus (n_2\omega + k_2)$ for non-negative integers n_1, n_2, k_1, k_2 is defined as $\max(n_1, n_2)\omega + (k_1 + k_2)$ and $+$ is a usual addition of ordinal numbers.

Observation 7.2 *The original query-mapping algorithm remains valid if we replace, in the computation of norms, integers by ordinal numbers with the operations defined above.*

Acknowledgement: We are very grateful to the anonymous referees for their careful reading and helpful suggestions.

References

1. Krzysztof R. Apt. Logic programming. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B: Formal Models and Semantics, chapter 15, pages 493–574. MIT Press, 1990.
2. Krzysztof R. Apt. *From Logic Programming to Prolog*. Prentice-Hall International Series in Computer Science. Prentice Hall, 1997.
3. Florence Benoy and Andy King. Inferring argument size relationships with CLP(R). In J. Gallagher, editor, *Logic Program Synthesis and Transformation. Proceedings*, pages 204–223. Springer Verlag, 1996. Lecture Notes in Computer Science, volume 1207.
4. Annalisa Bossi and Nicoletta Cocco. Preserving universal termination through unfold/fold. In Giorgio Levi and Mario Rodríguez-Artalejo, editors, *Algebraic and Logic Programming*, pages 269–286. Springer Verlag, 1994. Lecture Notes in Computer Science, volume 850.
5. Johan Boye and Jan Maluszyński. Directional types and the annotation method. *Journal of Logic Programming*, 33(3):179–220, 1997.
6. Alex Brodsky and Yehoshua Sagiv. Inference of inequality constraints in logic programs. In *Proceedings of the Tenth ACM SIGACT-SIGART-SIGMOD Symposium on Principles of Database Systems*, pages 227–240, 1991.
7. Francisco Bueno, Maria J. García de la Banda, and Manuel V. Hermenegildo. Effectiveness of global analysis in strict independence-based automatic parallelization. In Maurice Bruynooghe, editor, *Logic Programming, Proceedings of the 1994 International Symposium*, pages 320–336. MIT Press, 1994.
8. Michael Codish and Vitaly Lagoon. Type dependencies for logic programs using ACI-unification. In *Proceedings of the 1996 Israeli Symposium on Theory of Computing and Systems*, pages 136–145. IEEE Press, June 1996.
9. Michael Codish and Cohavit Taboch. A semantic basis for termination analysis of logic programs. *The Journal of Logic Programming*, 41(1):103–123, 1999.
10. Patrick Cousot and Radhia Cousot. Abstract interpretation and application to logic programs. *The Journal of Logic Programming*, 13:103–180, 1992.
11. Danny De Schreye and Stefaan Decorte. Termination of logic programs: The never-ending story. *The Journal of Logic Programming*, 19/20:199–260, May/July 1994.
12. Saumya K. Debray. A simple code improvement scheme for prolog. *The Journal of Logic Programming*, 13:103–180, 1992.
13. Stefaan Decorte, Danny De Schreye, and Massimo Fabris. Automatic inference of norms: a missing link in automatic termination analysis. In D. Miller, editor, *Proceedings of the 1993 International Logic Programming Symposium*, pages 420–436, 1993.
14. Nachum Dershowitz. Orderings for term-rewriting systems. *Theoretical Computer Science*, 17(3):279–301, 1982.
15. Robert W. Floyd. Assigning meanings to programs. In J. Schwartz, editor, *Mathematical Aspects of Computer Science*, pages 19–33. American Mathematical Society, 1967. Proceedings of Symposia in Applied Mathematics; v. 19.
16. Jürgen Giesl. Termination of nested and mutually recursive algorithms. *Journal of Automated Reasoning*, 19:1–29, 1997.
17. Ronald L. Graham. *Rudiments of Ramsey theory*. Number 45 in Regional conference series in mathematics. American Mathematical Society, 1980.

18. Christian Holzbaur. OFAI CLP(Q,R) Manual. Technical Report TR-95-09, Austrian Research Institute for Artificial Intelligence, Vienna, 1995.
19. Joxan Jaffar and Michael J. Maher. Constraint logic programming: A survey. *The Journal of Logic Programming*, 19/20:503–582, May/July 1994.
20. Donald E. Knuth. Textbook examples of recursion. In Vladimir Lifschitz, editor, *Artificial Intelligence and Mathematical Theory of Computation. Papers in Honor of John McCarthy*, pages 207–229. Academic Press, Inc., 1991.
21. N. Lindenstrauss and Yehoshua Sagiv. Automatic termination analysis of logic programs. In Lee Naish, editor, *Proceedings of the Fourteenth International Conference on Logic Programming*, pages 63–77. MIT Press, July 1997.
22. Naomi Lindenstrauss, Yehoshua Sagiv, and Alexander Serebrenik. *TermiLog*: A system for checking termination of queries to logic programs. In Orna Grumberg, editor, *Computer Aided Verification, 9th International Conference*, pages 63–77. Springer Verlag, June 1997. Lecture Notes in Computer Science, volume 1254.
23. Naomi Lindenstrauss, Yehoshua Sagiv, and Alexander Serebrenik. Unfolding mystery of the *mergesort*. In Norbert Fuchs, editor, *Proceedings of the Seventh International Workshop on Logic Program Synthesis and Transformation*. Springer Verlag, 1998. Lecture Notes in Computer Science, volume 1463.
24. John W. Lloyd. *Foundations of Logic Programming*. Symbolic Computation. Springer Verlag, 1984.
25. Zohar Manna and John McCarthy. Properties of Programs and partial function logic. *Machine Intelligence*, 5:27–37, 1970.
26. Fred Mesnard and J. Ganascia. CLP(Q) for proving interargument relations. In A. Petrossi, editor, *Proceedings META '92*, pages 309–320. Springer Verlag, 1992.
27. Richard A. O'Keefe. *The Craft of Prolog*. MIT Press, 1990.
28. Lutz Plümer. *Termination Proofs for Logic Programs*. Lecture Notes in Artificial Intelligence, volume 446. Springer Verlag, 1990.
29. Yehoshua Sagiv. A termination test for logic programs. In *International Logic Programming Symposium*. MIT Press, 1991.
30. Konstatinos F. Sagonas, Terrance Swift, and David S. Warren. *The XSB Programmer's Manual. Version 1.3 (β)*. Department of Computer Science, SUNY @ Stony Brook, U.S.A., September 1993.
31. SICS. *User Manual. Version 3.7.1*. Swedish Institute of Computer Science, 1998.
32. Leon Sterling and Ehud Shapiro. *The Art of Prolog*. The MIT Press, second edition, 1994.
33. Hisao Tamaki and Taisuke Sato. Unfold/Fold transformation of logic programs. In Sten-Åke Tärnlund, editor, *Proceedings of the Second International Logic Programming Conference*, pages 127–138. Uppsala University, 1984.
34. Jeffrey D. Ullman and Allen van Gelder. Efficient tests for top-down termination of logical rules. *Journal of the Association for Computing Machinery*, 35(2):345–373, April 1988.
35. Allen van Gelder. Deriving constraints among argument sizes in logic programs. *Annals of Mathematics and Artificial Intelligence*, 3:361–392, 1991.
36. Kristof Verschaetse and Danny De Schreye. Deriving termination proofs for logic programs, using abstract procedures. In Koichi Furukawa, editor, *Logic Programming, Proceedings of the Eighth International Conference*, pages 301–315. MIT Press, 1991.

37. Kristof Verschaetse and Danny De Schreye. Deriving of linear size relations by abstract interpretation. In Maurice Bruynooghe and Martin Wirsing, editors, *Programming Language Implementation and Logic Programming, 4th International Symposium*, pages 296–310. Springer Verlag, 1992. Lecture Notes in Computer Science, volume 631.
38. Bal Wang and R. K. Shyamasundar. A methodology for proving termination of logic programs. *The Journal of Logic Programming*, 21:1–30, 1994.