

# *A General Framework for Static Profiling of Parametric Resource Usage* \*

published in Theory and Practice of Logic Programming

P. LOPEZ-GARCIA<sup>1,2</sup> M. KLEMEN<sup>1</sup> U. LIQAT<sup>1</sup> M.V. HERMENEGILDO<sup>1,3</sup>

<sup>1</sup>*IMDEA Software Institute*

(e-mail: {pedro.lopez,maximiliano.klemen,umer.liqat,manuel.hermenegildo}@imdea.org)

<sup>2</sup>*Spanish Council for Scientific Research (CSIC)*

<sup>3</sup>*Technical University of Madrid (UPM)*

*submitted April 30, 2016; revised July 10, 2016; accepted July 22, 2016*

## Appendix A Additional Examples

### *Example 1*

Consider the following program to determine whether a list is a sublist of another. A sublist can be specified in terms of prefixes and suffixes: a suffix of a prefix, or a prefix of a suffix. The following program uses the latter to implement the `sublist` predicate.

```

1  sublist(L, L).
2  sublist(Sub, List) :- suffix(List, Suf), prefix(Suf, Sub).
3
4  suffix(List, Suffix):- append(_, Suffix, List).
5
6  prefix(List, Prefix):- append(Prefix, _, List).
7
8  append([], L, L).
9  append([X|Xs], L, [X|Zs]):- append(Xs, L, Zs).

```

Assume we are going to perform the analysis, to infer upper bounds on both the standard and accumulated costs, in terms of resolution steps, for the calling pattern `sublist(list, list)`, i.e., for the case where `sublist` is called with both of its arguments bound to lists. Given such calling pattern, CiaoPP infers the unique calling patterns `suffix(list, var)` and `prefix(var, list)`, for `suffix` and `prefix`, where `var` represents an unbound variable. However, two different calling patterns are inferred for `append`: `append(var, var, list)`, when it is called from `suffix`, and `append(list, var, list)`, when it is called from `prefix`.

Assume that size relations have been inferred for the different arguments in a clause, and that the size metric used is the *list length* of an argument, since all arguments are lists. The size of the output (second) argument of `suffix` is inferred as a function on

\* This research has received funding from the EU FP7 agreement no 318337, *ENTRA*, Spanish MINECO TIN2012-39391 *StrongSoft* and TIN2015-67522-C3-1-R *TRACES* projects, and the Madrid M141047003 *N-GREENS* program. Special thanks are due to John Gallagher for many fruitful and inspiring discussions and to the anonymous reviewers for their careful and useful comments.

its input (first) argument, and is represented by  $Sz_{\text{suffix}}^2(n)$ . Such inference sets up the following size relations:

$$\begin{aligned} Sz_{\text{suffix}}^2(n) &= Sz_{\text{append}}^2(n) \\ Sz_{\text{append}}^2(n) &= 1 && \text{if } n = 1 \\ Sz_{\text{append}}^2(n) &= n && \text{if } n > 1 \\ Sz_{\text{append}}^2(n) &= Sz_{\text{append}}^2(n-1) && \text{if } n > 1 \end{aligned}$$

and finds the closed form  $Sz_{\text{suffix}}^2(n) = n$ .

In order to infer the standard cost of this program, the analysis sets up the following cost relations for **sublist**, **suffix**, **prefix** and **append**:

$$\begin{aligned} C_{\text{sublist}}(n, m) &= C_{\text{suffix}}(n) + Sol_{\text{suffix}}(n) * C_{\text{prefix}}(Sz_{\text{suffix}}^2(n), m) + 2 \\ C_{\text{suffix}}(n) &= C_{\text{append}}(n) + 1 \\ C_{\text{prefix}}(n, m) &= C_{\text{append}}(n, m) + 1 \end{aligned}$$

Note that the size of the input to the call to **prefix** is given by the size of the output of **suffix**, represented by  $Sz_{\text{suffix}}^2(n)$ .

The cost relations for the two variants of **append** are:

$$\begin{aligned} C_{\text{append}}(n) &= 1 && \text{if } n = 1 \\ C_{\text{append}}(n) &= C_{\text{append}}(n-1) + 2 && \text{if } n > 1 \\ \\ C_{\text{append}}(n, m) &= 1 && \text{if } m = 1, n = 1 \\ C_{\text{append}}(n, m) &= 1 && \text{if } m = 1 \\ C_{\text{append}}(n, m) &= C_{\text{append}}(n-1, m-1) + 1 && \text{if } m > 1 \end{aligned}$$

and their closed forms are  $C_{\text{append}}(n) = 2n - 1$  and  $C_{\text{append}}(n, m) = m$  respectively.

Note that in this program **suffix** produces multiple solutions. For each solution of **suffix** the **prefix** predicate is executed on backtracking.

The cost relations for the inference of the number of solutions are:

$$\begin{aligned} Sol_{\text{suffix}}(n) &= Sol_{\text{append}}(n) \\ Sol_{\text{append}}(n) &= 1 && \text{if } n = 1 \\ Sol_{\text{append}}(n) &= Sol_{\text{append}}(n-1) + 1 && \text{if } n > 1 \end{aligned}$$

and the closed form is  $Sol_{\text{suffix}}(n) = n$ .

After composing all the closed forms, the analysis obtains the following function representing an upper bound on the resource usage of all the predicates:

$$\begin{aligned} C_{\text{suffix}}(n) &= 2n \\ C_{\text{prefix}}(n) &= m + 1 \\ C_{\text{append}}(n) &= 2n - 1 \\ C_{\text{append}}(n, m) &= m \\ C_{\text{sublist}}(n) &= nm + 3n + 2 \end{aligned}$$

Assume now that we declare **sublist** and **append** as cost centers to infer the accumulated costs in them. The cost relations set up for **sublist** are:

$$\begin{aligned} C_{\text{sublist}}^{\text{sublist}}(n, m) &= C_{\text{suffix},1}^{\text{sublist}}(n) + Sol_{\text{suffix}}(n) * C_{\text{prefix},1}^{\text{sublist}}(Sz_{\text{suffix}}^2(n), m) + 2 \\ C_{\text{sublist}}^{\text{append}}(n, m) &= C_{\text{suffix},0}^{\text{append}}(n) + Sol_{\text{suffix}}(n) * C_{\text{prefix},0}^{\text{append}}(Sz_{\text{suffix}}^2(n), m) \end{aligned}$$

Replacing the functions for sizes ( $Sz$ ) and solutions ( $Sol$ ) from the previous step we get:

$$\begin{aligned} C_{\text{sublist}}^{\text{sublist}}(n, m) &= C_{\text{suffix},1}^{\text{sublist}}(n) + n * C_{\text{prefix},1}^{\text{sublist}}(n, m) + 2 \\ C_{\text{sublist}}^{\text{append}}(n, m) &= C_{\text{suffix},0}^{\text{append}}(n) + n * C_{\text{prefix},0}^{\text{append}}(n, m) \end{aligned}$$

Furthermore, the intermediate cost relations are set up as follows:

$$\begin{aligned} C_{\text{suffix},1}^{\text{sublist}}(n) &= C_{\text{append}}^{\text{sublist}}(n) + 1 \\ &= 1 && \text{(Lemma 3)} \\ C_{\text{prefix},1}^{\text{sublist}}(n, m) &= C_{\text{append}}^{\text{sublist}}(n, m) + 1 \\ &= 1 && \text{(Lemma 3)} \\ C_{\text{suffix},0}^{\text{append}}(n) &= C_{\text{append}}^{\text{append}}(n) \\ C_{\text{prefix},0}^{\text{append}}(n, m) &= C_{\text{append}}^{\text{append}}(n, m) \\ \\ C_{\text{append}}^{\text{append}}(n) &= C_{\text{append}}^{\text{append}}(n-1) + 2 && \text{if } n > 1 \\ C_{\text{append}}^{\text{append}}(n) &= 1 && \text{if } n = 1 \\ \\ C_{\text{append}}^{\text{append}}(n, m) &= 1 && \text{if } m = 1, n = 1 \\ C_{\text{append}}^{\text{append}}(n, m) &= 1 && \text{if } m = 1 \\ C_{\text{append}}^{\text{append}}(n, m) &= C_{\text{append}}^{\text{append}}(n-1) + 1 && \text{if } m > 1 \end{aligned}$$

After composing all the intermediate cost relations, the analysis obtains the following functions representing upper bounds on the accumulated resource usage of `sublist` and `append`:

$$\begin{aligned} C_{\text{sublist}}^{\text{sublist}}(n) &= n + 3 \\ C_{\text{sublist}}^{\text{append}}(n) &= nm + 2n - 1 \end{aligned}$$

### Example 2

Consider the following program  $\mathcal{P}$ :

```

1 p(X, Y) :- i1, i2, q(X, Z), s(Z, Y).
2
3 q(0, 0).
4 q(X, Y) :- X1 is X - 1, q(X1, Z), i1, i3, s(Z, W), Y is W + 1.
5
6 s(0, 0) :- i1.
7 s(X, Y) :- i2, i4, X1 is X - 1, s(X1, Z), Y is Z + 1.

```

Assume that `im`, for  $m \in \{1, 2, 3, 4\}$ , are builtin/library predicates and that their standard costs are given by means of trust assertions: for simplicity we assume that  $\Psi(\text{im}) = C_{\text{im}} = 1$ , for  $m \in \{1, 2, 3, 4\}$ , and that the (standard) cost of the `is/2` arithmetic predicate is given as zero. Assume also that  $\varphi(\text{p}) = 0$  for all predicates  $\text{p} \in \mathcal{P}$ .

Assume that for all the predicates, the first argument is an input argument and the second one is output, and that the type of all arguments is the set of natural numbers. Assume that the following size relationships, expressing the size of the output argument as a function of the size of the input argument, have already been inferred for all of them:

- $Sz_{\text{p}}^2(n) = n$ , which means that the size (under the natural value metric) of the second argument of predicate `p` is  $n$ , the size of the input argument.
- Similarly,  $Sz_{\text{q}}^2(n) = n$ , and  $Sz_{\text{s}}^2(n) = n$ .

Assume that the set of cost centers is  $\diamond = \{\mathbf{p}, \mathbf{q}, \mathbf{s}\}$ , and that we want to estimate (upper bounds on) the cost accumulated in all the cost centers for the predicates  $\mathbf{p}$ ,  $\mathbf{q}$ , and  $\mathbf{s}$ . Let  $\mathbf{C}_p^q(\bar{x})$  denote an upper bound on the accumulated cost in cost center  $\mathbf{q}$  corresponding to a call  $\mathbf{p}(\bar{x})$ .

Assume we process each strongly-connected component of the call graph of the program in reverse topological order. We start by inferring the costs accumulated in cost center  $\mathbf{s}$ . The accumulated cost in  $\mathbf{s}$  corresponding to a call to  $\mathbf{s}$  is expressed by the following cost relation:

$$\begin{aligned} \mathbf{C}_s^s(0) &= \mathbf{C}_{i1,1}^s = \mathbf{C}_{i1} = 1 \\ \mathbf{C}_s^s(n) &= \mathbf{C}_{i2,1}^s + \mathbf{C}_{i4,1}^s + \mathbf{C}_s^s(n-1) \end{aligned}$$

which can be written as:

$$\begin{aligned} \mathbf{C}_s^s(0) &= 1 \\ \mathbf{C}_s^s(n) &= 2 + \mathbf{C}_s^s(n-1) \end{aligned}$$

and whose closed-form solution is:

$$\mathbf{C}_s^s(n) = 2n + 1.$$

Now, we analyze predicate  $\mathbf{q}$ . To this end, the accumulated cost in  $\mathbf{s}$  for a call to  $\mathbf{q}$  is expressed by:

$$\begin{aligned} \mathbf{C}_q^s(0) &= 0 \\ \mathbf{C}_q^s(n) &= \mathbf{C}_q^s(n-1) + \mathbf{C}_{i1,0}^s + \mathbf{C}_{i3,0}^s + \mathbf{C}_s^s(n-1) \end{aligned}$$

Since there is a trust assertion providing the cost of  $\mathbf{i1}$ ,  $\Psi(\mathbf{i1}) = 1$  (as already said), according to Expression 5, we have that  $\mathbf{C}_{i1,0}^s = B_\varphi(\mathbf{i1}, \mathbf{s}, 0) \times \Psi(\mathbf{i1}) = 0 \times 1 = 0$ . Note that  $B_\varphi(\mathbf{i1}, \mathbf{s}, 0) = 0$  because  $\mathbf{i1} \neq \mathbf{s}$  and the environment (third argument of  $B_\varphi$ ) is 0 (since  $\mathbf{i1}$  is called in the scope of cost center  $\mathbf{q}$ , not in the scope of the cost center where the analysis is accumulating costs in this equation, i.e.,  $\mathbf{s}$ ). Thus, the cost of  $\mathbf{i1}$  is not taken into account in this equation. The same consideration applies to  $\mathbf{i3}$ .

Since the cost function for  $\mathbf{C}_s^s(n)$  has already been computed, replacing values we have

$$\begin{aligned} \mathbf{C}_q^s(0) &= 0 \\ \mathbf{C}_q^s(n) &= \mathbf{C}_q^s(n-1) + 2(n-1) + 1 \end{aligned}$$

and

$$\begin{aligned} \mathbf{C}_q^s(0) &= 0 \\ \mathbf{C}_q^s(n) &= \mathbf{C}_q^s(n-1) + 2n - 1 \end{aligned}$$

The solution to the cost relation above is:

$$\mathbf{C}_q^s(n) = n^2.$$

We now analyze predicate  $\mathbf{p}$ , so that the accumulated cost in  $\mathbf{s}$  for a call to  $\mathbf{p}$  is expressed by:

$$\mathbf{C}_p^s(n) = \mathbf{C}_{i1,0}^s + \mathbf{C}_{i2,0}^s + \mathbf{C}_q^s(n) + \mathbf{C}_s^s(n)$$

For the same considerations as before, the costs of  $\mathbf{i1}$  and  $\mathbf{i2}$  in the body of the clause defining  $\mathbf{p}$  are not taken into account (i.e.,  $\mathbf{C}_{i1,0}^s = \mathbf{C}_{i2,0}^s = 0$ ). Replacing values we have that:

$$\mathbf{C}_p^s(n) = n^2 + 2n + 1.$$

The inference of the accumulated cost in  $\mathbf{s}$  for predicates  $\mathbf{p}$ ,  $\mathbf{q}$ , and  $\mathbf{s}$  has finished, and we start now the inference of the accumulated costs in  $\mathbf{q}$ . By Lemma 3,  $\mathbf{C}_s^q(n) = 0$ , i.e., we do not need to analyze predicate  $\mathbf{s}$ , since it does not call  $\mathbf{q}$ . However, now the costs of  $\mathbf{i1}$  and  $\mathbf{i3}$  in the body of the second clause of  $\mathbf{q}$  do have to be taken into account. To this end, the recurrence equations expressing the accumulated cost in  $\mathbf{q}$  for a call to  $\mathbf{q}$  are:

$$\begin{aligned} \mathbf{C}_q^q(0) &= 0 \\ \mathbf{C}_q^q(n) &= \mathbf{C}_q^q(n-1) + \mathbf{C}_{q,1}^{i1} + \mathbf{C}_{q,1}^{i3}. \\ \mathbf{C}_q^q(0) &= 0 \\ \mathbf{C}_q^q(n) &= \mathbf{C}_q^q(n-1) + 2. \end{aligned}$$

The solution to the recurrence above is:

$$\mathbf{C}_q^q(n) = 2n.$$

Now, the accumulated cost in  $\mathbf{q}$  for a call to  $\mathbf{p}$  is expressed as:

$$\mathbf{C}_p^q(n) = \mathbf{C}_q^q(n) + \mathbf{C}_s^q(n)$$

Replacing values we have that:

$$\mathbf{C}_p^q(n) = 2n$$

Let us compute now the accumulated cost in  $\mathbf{p}$ . Since it is not called from  $\mathbf{q}$  nor  $\mathbf{s}$ , we have that  $\mathbf{C}_q^p(n) = \mathbf{C}_s^p(n) = 0$ . The accumulated cost in  $\mathbf{p}$  for a call to  $\mathbf{p}$  is just the cost of  $\mathbf{i1}$  and  $\mathbf{i2}$ :

$$\mathbf{C}_p^p(n) = \mathbf{C}_{i1,1}^p + \mathbf{C}_{i2,1}^p = \mathbf{C}_{i1} + \mathbf{C}_{i2}.$$

Thus:

$$\mathbf{C}_p^p(n) = 2.$$

Note that the *standard cost* of  $\mathbf{p}$  ( $\mathbf{C}_p(n)$ ) can be expressed in terms of the accumulated costs in each of the cost centers:

$$\mathbf{C}_p(n) = \mathbf{C}_p^p(n) + \mathbf{C}_p^q(n) + \mathbf{C}_p^s(n).$$

### Example 3

Consider the following program where the predicates  $p$  and  $q$  are mutually recursive.

```

1 coupled(X, Y):- f(X, Y).
2
3 p(0, []).
4 p(N, [a|R]) :- N1 is N-1, q(N1, R).
5
6 q(0, []).
7 q(N, [a|R]) :- N1 is N-1, p(N1, R).
```

Assuming that we want to infer the standard cost of this program in terms of resolution steps, the analysis sets up the following cost relations for *coupled*,  $p$  and  $q$ , we have the following cost relations:

$$\begin{aligned} \mathbf{C}_p(n) &= 1 && \text{if } n = 0 \\ \mathbf{C}_p(n) &= 1 + \mathbf{C}_q(n-1) && \text{if } n > 0 \\ \mathbf{C}_q(n) &= 1 && \text{if } n = 0 \\ \mathbf{C}_q(n) &= 1 + \mathbf{C}_p(n-1) && \text{if } n > 1 \end{aligned}$$

$$\mathbf{C}_{coupled}(n) = 1 + \mathbf{C}_p(n)$$

After composing the closed forms, the analysis obtains the following function representing an upper bound on the resource usage of *coupled*, *p* and *q*:

$$\mathbf{C}_{coupled}(n) = n + 2$$

$$\mathbf{C}_p(n) = n + 1$$

$$\mathbf{C}_q(n) = n + 1$$

Notice that in this program the cost relations for *p* and *q* are mutually recursive (i.e., they are defined in terms of each other), and for this reason the cost functions representing the upper bound on the resolution steps in the two are same ( $n + 1$ ). Hence, the cost of each mutually-recursive predicate subsumes the cost of the other. However, this cost is in fact distributed between the *p* and *q* predicates. In order to identify the cost that each of these predicates contributes to this  $n + 1$  expression and to the overall cost of *coupled* ( $n + 2$ ), we perform the accumulated cost analysis, declaring all the predicates as cost centers. The instantiation of the equational framework described in Sect. 4 obtains the following accumulated costs for *coupled*, *p*, and *q*:

$$\mathbf{C}_{coupled}^{coupled}(n) = 1$$

$$\mathbf{C}_{coupled}^p(n) = \frac{n}{2} + \frac{(-1)^n}{4} + \frac{3}{4}$$

$$\mathbf{C}_{coupled}^q(n) = \frac{n}{2} - \frac{(-1)^n}{4} + \frac{1}{4}$$

It is now clear how much cost each of *coupled*, *p*, and *q* contributes to the standard cost of the whole program ( $n + 2$ ). Note that the standard cost of the mutually recursive predicates *p* and *q*, which is  $n + 1$ , is now halved among the two as accumulated costs of *p* and *q*.

In this example we have shown a hypothetical scenario highlighting that the accumulated cost information is more useful for mutually recursive parts of a program in order to identify how much each of the mutually recursive predicates contributes to the overall cost. This was not possible using only the standard cost information.

#### Example 4

Consider the following program to determine the parity of a number where predicates *even* and *odd* are mutually recursive.

```

1 even(0).
2 even(N):- N > 0, N1 is N - 1, odd(N1).
3
4 odd(1).
5 odd(N):- N > 1, N1 is N - 1, even(N1).

```

Similar to the Example 3, this program contains mutually recursive predicates *even* and *odd*. Since both are defined in terms of each other, the standard analysis obtains a same cost function for them representing an upper bound on the resource usage.

$$\mathbf{C}_{even}(n) = n + 1$$

$$\mathbf{C}_{odd}(n) = n + 1$$

In order to identify the cost that each of these predicates contributes to the overall cost of the program  $n + 1$ , we perform the accumulated cost analysis, declaring both *even* and *odd* as cost centers. The instantiation of the equational framework described in Sect. 4 obtains the following accumulated costs for *even* and *odd*:

$$C_{even}^{even}(n) = \frac{n}{2} + \frac{(-1)^n}{4} + \frac{3}{4}$$

$$C_{even}^{odd}(n) = \frac{n}{2} - \frac{(-1)^n}{4} + \frac{1}{4}$$

## Appendix B

### Additional Comments on the Relation of the Standard and Accumulated Cost

Assume that predicate  $p$  is a cost center. As already said, in this case the standard cost of a single call  $p(\bar{x})$  is the sum of its accumulated costs in all the cost centers in the program. This is formalized by Theorem 1 in (Haemmerlé et al. 2016), which holds under the assumption that  $p$  is a cost center. Intuitively, predicate  $c$  is “reachable” from predicate  $p$  if  $c = p$  or  $c$  can be invoked (either directly or indirectly) by  $p$ . If  $p$  is a cost center, Theorem 1 also holds if we restrict to the set of cost centers that are reachable from  $p$ , or to the set of cost centers that are descendants (in the call stack) of  $p$ . The reason is that if  $p$  is a cost center, and another cost center  $c$  (different from  $p$ ) is not reachable from  $p$ , then no part of the cost of a call to  $p$  is attributed to  $c$ . This is stated in Lemma 3.

Assume that  $p$  is the main (entry) predicate in a program, and that we are interested in knowing how its total (standard) cost is *distributed* over the (user-defined) cost centers. In this case,  $p$  should be declared as a cost center. This is because if  $p$  is a cost center, the residual cost (as defined in Section 4) of the call to the main predicate will be assigned to  $p$ . Otherwise the residual cost will be left unassigned to any cost center.

If  $p$  is not a cost center, the standard cost of a single call  $p(\bar{x})$  is the sum of its accumulated costs in all the cost centers that are descendants (in the call stack) of  $p$ , plus the residual cost of that call.

## References

- HAEMMERLÉ, R., LOPEZ-GARCIA, P., LIQAT, U., KLEMEN, M., GALLAGHER, J. P., AND HERMENEGILDO, M. V. 2016. A Transformational Approach to Parametric Accumulated-cost Static Profiling. In *FLOPS'16*. LNCS, vol. 9613. Springer, 163–180.