Carnegie Mellon

School of Computer Science

Scheme Flow Analysis Note 2 4/27/90

Useless-Variable Elimination

Olin Shivers
shivers@cs.cmu.edu

1 Intro

In my 1988 SIGPLAN paper "Control-Flow Analysis in Scheme," I promised to cover several items in more detail in a "forthcoming tech report." The tech report has yet to happen. Most of the promised items have been covered in other papers ("The Semantics of Scheme Control-Flow Analysis" and "Data-Flow Analysis and Type Recovery in Scheme"). Useless-variable elimination, however, is still described only in the T source code of my analysis system.

I am writing this informal note to get an outline of UVE set down on paper. This text will reappear in my dissertation. This note is intended for limited distribution to interested parties who don't want to wait for the dissertation.

2 Useless variables

A *useless variable* is one whose value contributes nothing to the final outcome of the computation. We can remove a useless variable and the computation producing its value from our program. For example, consider the following code fragment:

```
(let ((sum  (+ a b))              (let ((sum (+ a b)))
      (prod (* a b)))     ⇒         (f sum a))
  (f sum a))
```

`Prod` is never used in the program, so we can remove it and its binding computation `(* a b)`. This example is fairly easy to detect; most Scheme compilers would find and optimise it.

On the other hand, some useless variables involve circular dependencies or multiple (join) dependencies. The simple lexical analysis that suffices for the above example won't spot these cases. For example, consider the factorial loop:

```
(letrec ((lp (λ (ans j bogus)
               (if (= 0 j) ans
                   (lp (* ans j)
                       (- j 1)
                       (sqrt bogus))))))
  (lp 1 n n))
```

Although the variable `bogus` doesn't contribute anything at all to the final result, it appears to be used in the loop.

UVE is often useful to clean up after applying other code transformations, such as copy propagation or induction-variable elimination. For example, when we introduce a new variable to track an induction function on some basic induction variable, the basic variable frequently becomes useless (*e.g.*, compare parts (d) and (e) of figure 4 in "Control-Flow Analysis in Scheme").

3 Finding useless variables

Detecting these cases requires a simple backwards flow analysis on the CPS intermediate form. We actually compute a conservative approximation to the inverse problem — finding the set of all useful variables. We start with variables that must be assumed useful (*e.g.*, a variable whose value is returned as the value of the procedure being analysed). Then we trace backwards through the control-flow structure of the program. If a variable's value contributes to the computation of a useful variable, then it, too, is marked useful. When we're done, all unmarked variables are useless.

This gives us a mark-and-sweep algorithm for a sort of "computational gc."

To be specific, in the implementation of UVE I have implemented, a variable is *useful* if it appears

- in the function position of a call:
  `(f 5 0)`
- as the predicate in a conditional primop:
  `(if p (λ () ...) (λ () ...))`
- as the continuation of a primop:
  `(+ 3 5 c)`, `(if #t c a)`

- as an argument in a call to
  - a side-effecting operation (output or store): `(print a)`, `(set-car! x y)`
  - an external procedure, or a primop whose continuation is an external procedure.
  - a primop whose continuation binds a useful variable: `(+ metoo 3 (λ (used) ...))`
  - a lambda whose corresponding parameter is useful: `((λ (x used y) ...) 3 metoo 7)`

The first three conditions spot variables used for control-flow. The next two mark variables whose value escapes, and must therefore be assumed useful. The final two recursive conditions are the ones that cause the analysis to chain backwards through the control-flow graph: if a variable is useful, then all the variables used to compute its value are useful.

4 Optimising useless variables

Once we have found the useless variables in a program, we can optimise the program in two steps. In the first step, we remove all references to useless variables and eliminate all useless computations (primop calls). In the second step, we remove useless variables from their lambda lists where possible.

4.1 Removing useless variables from calls

In this phase, we globally remove all references to useless variables in our program. If a useless variable appears as an argument to a primop call, we remove the entire primop computation. For instance, suppose the useless variable `x` appears as an argument in the primop call `(+ x y k)`. By the definition of a useless variable, we know that the continuation `k` must bind the result of the addition to a useless variable, as well (if it didn't, we'd have marked `x` as useful). This renders the addition operation useless, so we can remove it, replacing `(+ x y k)` with `(k #f)`. The actual value passed to the continuation (we used `#f` here) is not important — remember that we are globally deleting all references to useless variables. Since the continuation `k` must bind the value `#f` to a useless variable, the value is guaranteed never to be referenced.

If a reference to a useless variable appears in a non-primop call, we simply replace it with some constant. If `x` is useless, we convert `(f x y)` to `(f #f y)`. Similar reasoning applies in this case: if `x` is useless, it must be the case that `f`'s corresponding parameter is useless as well. All references to this parameter will be deleted, so we can pass any value we like to it.

4.2 Removing useless variables from lambda lists

Suppose we have determined that `x` is useless in lambda $\ell =$ `(`λ `(x y) ...)`. After applying the transformation of the previous subsection, we can be sure that ℓ contains no references to `x`. The only remaining appearance of `x` is in ℓ's parameter list. Consider the places this lambda is called from, *e.g.*, `(f a 7)`. We can delete both the formal parameter `x` from its lambda list and the corresponding argument `a` from its call:

`(`λ `(x y) ...)` $\Rightarrow$ `(`λ `(y) ...)`   and   `(f a 7)` $\Rightarrow$ `(f 7)`

However, we can't apply this optimisation in all cases. Suppose ℓ is called from two places, the external call and `(f a 7)`. We can't simply delete `x` from ℓ's parameter list, because the external call, which we have no control over, is going to pass a value to ℓ for `x`. Or, suppose that our lambda is only called from one place, `(f a 7)`, but that call site calls two possible lambdas. Again, we can't delete the argument `a` from the call, because the other procedure is expecting it (unless it, as well, binds a useless variable).

We have a circular set of dependencies determining when it is safe to remove a useless variable from its lambda list and its corresponding arguments from the lambda's call sites:

- We can delete a variable from a lambda list only if we can delete its corresponding argument in all the calls that could branch to that lambda.
- We delete an argument from a call only if we can delete the corresponding formal parameter from every lambda we could branch to from that call.

It is not hard to compute a maximal solution to these constraints given control-flow information. We use a simple algorithm that iterates to a fixpoint. We compute two sets: the set RV of removeable useless variables, and the set RA of removeable call arguments. Initialise RV to be the set of all useless variables. For each useless variable v , find all the call sites that could branch to v's lambda, and put the call's corresponding argument into RA. Then iterate over these sets until we converge:

```
ITERATE until no change
    FOR each argument a in RA
        IF a's call could branch to a lambda whose
           corresponding parameter is not in RV
        THEN remove a from RA
    FOR each variable v in RV
        IF v's lambda can be called from a call site whose
           corresponding argument is not in RA
        THEN remove v from RV
```

For the purposes of the first loop, the xlambda counts as a disqualifying branch target; for the purposes of the second loop, the xcall counts as a disqualifying branch source. When we're done, we're left with RV and RA sets that satisfy the circular criteria above. We can safely remove all the arguments in RA from their calls and all the variables in RV from their lambda lists.

5 Results

Applying these two phases of optimisation will eliminate useless variables and their associated computations wherever possible. For example, we can remove the useless `bogus` variable and its square-root calculation from the factorial loop of section 2, leaving only the necessary computations:

```
(letrec ((lp (λ (ans j)
               (if (= 0 j) ans
                   (lp (* ans j)
                       (- j 1))))))
  (lp 1 n))
```

Note that as a special case of UVE, all unreferenced variables in a program are useless. The second phase of the UVE optimisation will remove these variables when possible.

References

Olin Shivers. Control-flow analysis in Scheme. *Proceedings of SIGPLAN '88 Conference on Programming Language Design and Implementation* (June 1988). (Also available as Technical Report ERGO-88-60, CMU School of Computer Science, Pittsburgh, Penn.)

Olin Shivers. Data-Flow analysis and type recovery in Scheme. Technical Report CMU-CS-90-115, CMU School of Computer Science, Pittsburgh, Penn. (March 1990). (Also to appear in *Topics in Advanced Language Implementation*, ed. Peter Lee, MIT Press.)

Olin Shivers. The semantics of Scheme control-flow analysis. *(In preparation)*

Olin Shivers. The semantics of Scheme control-flow analysis (Prelim). Technical Report ERGO-90-090, CMU School of Computer Science, Pittsburgh, Penn. (November 1988).

Olin Shivers. Super-β: Copy, constant and lambda propagation in Scheme. Working note #3 (May 1, 1990).