1 Introduction

Optimizing loops is particularly important in compilation, since loops (and in particular the inner loops) account for much of the execution times of many programs. Since tail-recursive functions are usually also turned into loops, the importance of loop optimizations is further magnified. In this lecture we will discuss two main ones: hoisting loop-invariant computation out of a loop, and optimizations based on induction variables.

2 What Is a Loop?

Before we discuss loop optimizations, we should discuss what we identify as a loop. In our source language, this is rather straightforward, since loops are formed with while or for, where it is convenient to just elaborate a for loop into its corresponding while form.

The key to a loop is a back edge in the control-flow graph from a node \( l \) to a node \( h \) that dominates \( l \). We call \( h \) the header node of the loop. The loop itself then consists of the nodes on a path from \( h \) to \( l \). It is convenient to organize the code so that a loop can be identified with its header node. We then write \( \text{loop}(h, l) \) if line \( l \) is in the loop with header \( h \).

When loops are nested, we generally optimize the inner loops before the outer loops. For one, inner loops are likely to be executed more often. For another, it could move computation to an outer loop from which it is hoisted further when the outer loop is optimized and so on.
3 Hoisting Loop-Invariant Computation

An (pure) expression is loop invariant if its value does not change throughout the loop. We can then define the predicate \( \text{inv}(h, p) \), where \( p \) is a pure expression, as follows:

\[
\begin{align*}
\text{inv}(h, c) & \quad \text{constant} \\
\text{inv}(h, x) & \quad \text{def}(l, x) \\
\text{inv}(h, x) & \quad \neg \text{loop}(h, l) \\
\text{inv}(h, s_1) & \quad \text{inv}(h, s_2) \\
\text{inv}(h, s_1 \oplus s_2) & \quad \text{inv}(h, s_1) \quad \text{inv}(h, s_2)
\end{align*}
\]

Since we are concerned only with programs in SSA form, it is easy to see that variables are loop invariant if they are not parameters of the header label. However, the definition above does not quite capture this for definitions \( t \leftarrow p \) where \( p \) is loop-invariant but \( t \) is not part of the label parameters. So we add a second propagation rule.

\[
\begin{align*}
\text{inv}(h, p) & \quad \text{loop}(h, l) \\
\text{inv}(h, t) & \quad l : t \leftarrow p \\
\end{align*}
\]

Note that we do not consider memory references of function calls to be loop invariant, although under some additional conditions they may be hoisted as well.

In order to hoist loop invariant computations out of a loop we should have a loop preheader in the control-flow graph, which immediately dominates the loop header. When then move all the loop invariant computations to the preheader, in order.

Some care must be taken with this optimization. For example, when the loop body is never executed the code could become significantly slower. Another problem if we have conditionals in the body of the loop: values computed only on one branch or the other will be loop invariant, but depending on the boolean condition one or the other may never be executed.

In some cases, when the loop guard is inexpensive and effect-free but the loop-invariant code is expensive, we might consider duplicating the test so that instead of

\[
\text{seq}(\text{pre}, \text{while}(e, s))
\]

we generate code for

\[
\text{seq}(\text{if}(e, \text{seq}(\text{pre}, \text{while}(e, s)), \text{nop}))
\]

where \( \text{pre} \) is the hoisted computation in the loop pre-header.

A typical example of hoisting loop invariant computation would be a loop to initialize all elements of a two-dimensional array:

\[
\begin{align*}
\text{for} \ (\text{int} \ i = 0; \ i < \text{width} \times \text{height}; \ i++) \\
\quad A[i] = 1;
\end{align*}
\]
We show the relevant part of the abstract assembly on the left. In the right is the result of hoisting the multiplication, enabled because both width and height are loop invariant and therefore their product is.

\[
i_0 \leftarrow 0
\]
\[
\text{goto } \text{loop}(i_0)
\]
\[
t \leftarrow \text{width} \times \text{height}
\]
\[
\text{goto } \text{loop}(i_0)
\]
\[
\text{loop}(i_1) :
\]
\[
t \leftarrow \text{width} \times \text{height}
\]
\[
\text{if } (i_1 \geq t) \text{ goto } \text{exit}
\]
\[
\ldots
\]
\[
i_2 \leftarrow i_1 + 1
\]
\[
\text{goto } \text{loop}(i_2)
\]
\[
\text{exit} :
\]

4 Induction Variables

Hoisting loop invariant computation is significant; optimizing computation which changes by a constant amount each time around the loop is probably even more important. We call such variables basic induction variables. The opportunity for optimization arises from derived induction variables, that is, variables that are computed from basic induction variables.

As an example we will use a function check if a given array is sorted in ascending order.

```c
bool is_sorted(int[] A, int n)
//@requires 0 <= n && n <= \length(A);
{
    for (int i = 0; i < n-1; i++)
        //@loop_invariant 0 <= i && i <= n-1;
        if (A[i] > A[i+1]) return false;
    return true;
}
```

Below is a possible compiled SSA version of this code, assuming that we do not
perform array bounds checks (or have eliminated them).

```c
is_sorted(A, n) :
    i_0 ← 0
    goto loop(i_0)

loop(i_1) :
    t_0 ← n - 1
    if (i_1 ≥ t_0) goto rtrue
    t_1 ← 4 * i_1
    t_2 ← A + t_1
    t_3 ← M[t_2]
    t_4 ← i_1 + 1
    t_5 ← 4 * t_4
    t_6 ← A + t_5
    t_7 ← M[t_6]
    if (t_3 > t_7) goto rfalse
    i_2 ← i_1 + 1
    goto loop(i_2)

rtrue :
    return 1

rfalse :
    return 0
```

Here, \( i_1 \) is the basic induction variable, and \( t_1 = 4 \times i_1 \) and \( t_4 = i_1 + 1 \) are the derived induction variables. In general, we consider a variable a derived induction variable if its has the form \( a \times i + b \), where \( a \) and \( b \) are loop invariant.

Let’s consider \( t_4 \) first. We see that common subexpression elimination applies. However, we would like to preserve the basic induction variable \( i_1 \) and its version.
$i_2$, so we apply code motion and then eliminate the second occurrence of $i_1 + 1$.

\begin{verbatim}
is_sorted(A, n) :  
  $i_0$ ← 0  
goto loop($i_0$)  
loop($i_1$) :  
  $t_0$ ← $n - 1$  
  if ($i_1 \geq t_0$) goto rtrue  
  $t_1$ ← $4 \times i_1$  
  $t_2$ ← $A + t_1$  
  $t_3$ ← $M[t_2]$  
  $t_4$ ← $i_1 + 1$  
  $t_5$ ← $4 \times t_4$  
  $t_6$ ← $A + t_5$  
  $t_7$ ← $M[t_6]$  
  if ($t_3 > t_7$) goto rfalse  
  goto loop($i_2$)  

Next we look at the derived induction variable $t_1 \leftarrow 4 \times i_1$. The idea is to see how we can calculate $t_1$ at a subsequent iteration from $t_1$ at a prior iteration. In order to achieve this effect, we add a new induction variable to represent $4 \times i_1$. We call this $j$ and add it to our loop variables in SSA form.

\begin{verbatim}
is_sorted(A, n) :  
  $i_0$ ← 0  
  $j_0$ ← $4 \times i_0$  
goto loop($i_0, j_0$)  
loop($i_1, j_1$) :  
  $t_0$ ← $n - 1$  
  if ($i_1 \geq t_0$) goto rtrue  
  $t_1$ ← $j_1$  
  $t_2$ ← $A + t_1$  
  $t_3$ ← $M[t_2]$  
  $i_2$ ← $i_1 + 1$  
  $j_2$ ← $4 \times i_2$  
  $t_4$ ← $i_2$  
  $t_5$ ← $4 \times t_4$  
  $t_6$ ← $A + t_5$  
  $t_7$ ← $M[t_6]$  
  if ($t_3 > t_7$) goto rfalse  
  goto loop($i_2, j_2$)
\end{verbatim}
\end{verbatim}
Crucial here is the invariant that $j_1 = 4 \times i_1$ when label $\text{loop}(i_1, j_1)$ is reached. Now we calculate

$$j_2 = 4 \times i_2 = 4 \times (i_1 + 1) = 4 \times i_1 + 4 = j_1 + 4$$

so we can express $j_2$ in terms of $j_1$ without multiplication. This is an example of *strength reduction* since addition is faster than multiplication. Recall that all the laws we used are valid for modular arithmetic. Similarly:

$$j_0 = 4 \times i_0 = 0$$

since $i_0 = 0$, which is an example of constant propagation followed by constant folding.

```plaintext
is_sorted(A, n) :
    i_0 ← 0
    j_0 ← 0
    goto loop(i_0, j_0)

loop(i_1, j_1) :
    t_0 ← n - 1
    if (i_1 ≥ t_0) goto rtrue
    t_1 ← j_1
    t_2 ← A + t_1
    t_3 ← M[t_2]
    i_2 ← i_1 + 1
    j_2 ← j_1 + 4
    t_4 ← i_2
    t_5 ← 4 \times t_4
    t_6 ← A + t_5
    t_7 ← M[t_6]
    if (t_3 > t_7) goto rfalse
    goto loop(i_2, j_2)
```
With some copy propagation, and noticing that \( n - 1 \) is loop invariant, we next get:

```plaintext
is_sorted(A, n) :
  i_0 ← 0
  j_0 ← 0  \@ensures j_0 = 4 * i_0
  t_0 ← n - 1
  goto loop(i_0, j_0)

loop(i_1, j_1) :
  \@requires j_1 = 4 * i_1
  if (i_1 ≥ t_0) goto rtrue
  t_2 ← A + j_1
  t_3 ← M[t_2]
  i_2 ← i_1 + 1
  j_2 ← j_1 + 4  \@ensures j_2 = 4 * i_2
  t_5 ← 4 * i_2
  t_6 ← A + t_5
  t_7 ← M[t_6]
  if (t_3 > t_7) goto rfalse
  goto loop(i_2, j_2)
```

With common subexpression elimination (noting the additional assertions we are aware of), we can replace \( 4 * i_2 \) by \( j_2 \). We combine this with copy propagation.

```plaintext
is_sorted(A, n) :
  i_0 ← 0
  j_0 ← 0  \@ensures j_0 = 4 * i_0
  t_0 ← n - 1
  goto loop(i_0, j_0)

loop(i_1, j_1) :
  \@requires j_1 = 4 * i_1
  if (i_1 ≥ t_0) goto rtrue
  t_2 ← A + j_1
  t_3 ← M[t_2]
  i_2 ← i_1 + 1
  j_2 ← j_1 + 4  \@ensures j_2 = 4 * i_2
  t_6 ← A + j_2
  t_7 ← M[t_6]
  if (t_3 > t_7) goto rfalse
  goto loop(i_2, j_2)
```

We observe another derived induction variable, namely \( t_2 = A + j_1 \). We give this a new name \( k_1 = A + j_1 \) and introduce it into our function. Again we just calculate:
Loop Optimizations

\[ k_2 = A + j_2 = A + j_1 + 4 = k_1 + 4 \text{ and } k_0 = A + j_0 = A. \]

```plaintext
is_sorted(A, n):
    i_0 ← 0
    j_0 ← 0
    k_0 ← A + j_0
    t_0 ← n - 1
    goto loop(i_0, j_0, k_0)
loop(i_1, j_1, k_1):
    @requires j_1 = 4 * i_1 ∧ k_1 = A + j_1
    if (i_1 ≥ t_0) goto rtrue
    t_2 ← k_1
    t_3 ← M[t_2]
    i_2 ← i_1 + 1
    j_2 ← j_1 + 4
    k_2 ← k_1 + 4
    t_6 ← A + j_2
    t_7 ← M[t_6]
    if (t_3 > t_7) goto rfalse
    goto loop(i_2, j_2, k_2)
```

After more round of constant propagation, common subexpression elimination, and dead code elimination we get:

```plaintext
is_sorted(A, n):
    i_0 ← 0
    j_0 ← 0
    k_0 ← A
    t_0 ← n - 1
    goto loop(i_0, j_0, k_0)
loop(i_1, j_1, k_1):
    @requires j_1 = 4 * i_1 ∧ k_1 = A + j_1
    if (i_1 ≥ t_0) goto rtrue
    t_3 ← M[k_1]
    i_2 ← i_1 + 1
    j_2 ← j_1 + 4
    k_2 ← k_1 + 4
    t_7 ← M[k_2]
    if (t_3 > t_7) goto rfalse
    goto loop(i_2, j_2, k_2)
```

With neededness analysis we can say that \( j_0, j_1, \) and \( j_2 \) are no longer needed and
can be eliminated.

\[
is\_sorted(A, n) :
\]

\[
\begin{align*}
i_0 &\leftarrow 0 \\
k_0 &\leftarrow A \\
t_0 &\leftarrow n - 1 \\
&\text{goto loop}(i_0, k_0)
\end{align*}
\]

\[
\text{loop}(i_1, k_1) : \quad \text{\@requires } k_1 = A + 4 * i_1
\]

\[
\begin{align*}
&\text{if } (i_1 \geq t_0) \text{ goto } rtrue \\
&t_3 \leftarrow M[k_1] \\
i_2 &\leftarrow i_1 + 1 \\
k_2 &\leftarrow k_1 + 4 \\
t_7 &\leftarrow M[k_2] \\
&\text{if } (t_3 > t_7) \text{ goto } rfalse \\
&\text{goto loop}(i_2, k_2)
\end{align*}
\]

Unfortunately, \( i_1 \) is still needed, since it governs a conditional jump. In order to eliminate that we would have to observe that

\[
i_1 \geq t_0 \text{ iff } A + 4 * i_1 \geq A + 4 * t_0
\]

This holds the the addition here is a on 64 bit quantities where the second operand is 32 bits, so no overflow can occur. If we exploit this we obtain:

\[
is\_sorted(A, n) :
\]

\[
\begin{align*}
i_0 &\leftarrow 0 \\
k_0 &\leftarrow A \\
t_0 &\leftarrow n - 1 \\
&\text{goto loop}(i_0, k_0)
\end{align*}
\]

\[
\text{loop}(i_1, k_1) : \quad \text{\@requires } k_1 = A + 4 * i_1
\]

\[
\begin{align*}
&\text{if } (k_1 \geq A + 4 * t_0) \text{ goto } rtrue \\
&t_3 \leftarrow M[k_1] \\
i_2 &\leftarrow i_1 + 1 \\
k_2 &\leftarrow k_1 + 4 \\
t_7 &\leftarrow M[k_2] \\
&\text{if } (t_3 > t_7) \text{ goto } rfalse \\
&\text{goto loop}(i_2, k_2)
\end{align*}
\]

Now \( i_0, i_1, \) and \( i_2 \) are no longer needed and can be eliminated. Moreover, \( A + 4 * t_0 \)
is loop invariant and can be hoisted.

\[
is\text{\_sorted}(A, n) : \\
k_0 \leftarrow A \\
t_0 \leftarrow n - 1 \\
t_8 \leftarrow 4 \times t_0 \\
t_9 \leftarrow A + t_8 \\
goto \text{loop}(k_0) \\
\text{loop}(k_1) : \\
\text{if } (k_1 \geq t_9) \text{ goto rtrue} \\
t_3 \leftarrow M[k_1] \\
k_2 \leftarrow k_1 + 4 \\
t_7 \leftarrow M[k_2] \\
\text{if } (t_3 > t_7) \text{ goto rfalse} \\
goto \text{loop}(k_2) \\
rtrue : \\
\text{return 1} \\
rfalse : \\
\text{return 0}
\]

It was suggested that we can avoid two memory accesses per iteration by unrolling the loop once. This make sense, but this optimization is beyond the scope of this lecture.

We have carried out the optimizations here on concrete programs and values, but it is straightforward to generalize them to arbitrary induction variables \(x\) that are updated with \(x_2 \leftarrow x_1 \pm c\) for a constant \(c\), and derived variables that arise from constant multiplication with or addition to a basic induction variable.