Computer Science 331
Analysis of Algorithms

Mike Jacobson

Department of Computer Science
University of Calgary

Lectures #7-8

Outline

Measuring Efficiency

What sorts of measures could we use? The following are all (sometimes) important:

- **Running Time** — no one wants to wait too long for programs to execute
- **Memory Used by Data (Storage Space)** — time is (sort of) unconstrained, but any computer can run out of memory
- **Memory Used by Code** — an issue if a program is to be stored on a low-memory device (like a smart card)
- **Time to Code** —- programmers must be paid and software development usually has deadlines!

Our focus will be on *running time* and *storage space*.

How Do We Measure Efficiency?

How can we compare algorithms or programs?

1. **Run the Code and Time the Execution.**
   Problem: Execution time is influenced by many factors:
   - *Hardware* (How fast is the CPU? How many of them?)
   - *Compiler and System Software* (Which OS?)
   - *Simultaneous User Activity* (Potentially affected by the time of day when the program was executed)
   - *Choice of Input Data* (Running times can vary on inputs, even inputs of the same "size")
   - *Programmer's Skill*
2. **Analyze the Code**
   Advantage: Only influenced by choice of data
   Disadvantage: Can be quite difficult!

We typically try to do *both* (analysis supported by execution timings).

What Will We Measure?

Most of the time, in this course, running time and storage space will be measured in an abstract *machine-independent* way.

Running Time:

- Number of primitive operations or "steps" (programming language statements) used
- Ignores:

Different costs of operations: multiplication.

Storage Space:

- Number of words of machine memory used, assuming each word can store the same (fixed) number of bits
- Ignores:

Memory hierarchy differences, cache vs main memory.

How Do We Wish To Measure Resources?

We will try to measure the amount of resources (time or space) used as a function of the "input size." (defined in various ways, depending on the type of input considered).

Example: if the input is an *array*, the appropriate measure of input size is (usually):

- array length, i.e., number of elements

Example: if the input is a single *integer*, which can be virtually as large as we want, the appropriate measure of input size is:

- the bit-length of the integer, i.e., number of bits in its binary representation

Complication: executions of a program on different inputs *with the same size* frequently have different costs!

Worst-Case Analysis

Consider the *maximal* amount of resources (such as *longest* running time) used by the algorithm, on any input of a given size

Advantages of This Type of Analysis:

- Upper bound on running time (guarantee that the algorithm will not take any longer for any input of the given size.
- For some algorithm, worst-case occurs farily often (Ie: searching an array for an element not in it).

Disadvantage of This Type of Analysis:

- May rarely occur: (Array in reverse order is worst case for a variation of quicksort)

Average-Case Analysis

Consider the **average** (or "expected") amount of resources (such as **average** running time) used by the algorithm, for an input of a given size

Advantage of This Type of Analysis:

- Captures resource consumption for typical inputs

Disadvantages of This Type of Analysis:

- May be difficult to determine average case.
- Typically requires probabilistic analysis.

In some, but not all cases, the worst-case and average-case running times (or amount of storage space used) are approximately the same.

Other Kinds of Analysis

Best-case Analysis:

- *minimal* amount of resources (such as *shortest* running time) used by the algorithm, on any input of a given size
- occasionally of interest, but usually together with other measures (eg. see whether best and worst cases running times are close)

Amortized Analysis:

- ratio of total cost of a sequence of operations to the number of operations in the sequence
- similar to average case, except that no assumptions about input distribution are required
- mostly beyond scope of the course, but some results will be mentioned

Objective and Strategy

Objective: use code (or pseudocode) to estimate the *worst-case running time* of a program (or algorithm).

Useful Values:

- Worst-case running time (exact)
- Upper and lower bounds on worst-case running time (easier, often sufficient)

Strategy: consider subprograms . . .

- beginning with individual statements . . .
- then considering progressively larger subprograms . . .
- until the whole program has been considered.

Case: Program is a Single Statement

Example: `x := 1`

Amount to charge:

- 1 unit (eg. single arithmetic/Boolean operation, comparison, or assignment)

Example: `x := y := 1`

Amount to charge:

- 2 units (one per assignment)
- be careful with compound statements
- one line does not always equal one unit!

Case: Program is a Sequence of Subprograms

Structure to Consider: S_1; S_2

Worst-Case Running Time: If

- worst-case running time of S_1 is T_1, and
- worst-case running time of S_2 is T_2,

then

- worst-case running time of entire program is *at most:*

T1 + T2

Explanation (upper bound because...):

- Worst case of S1 may not be worst case of S2

Case: Program is a Conditional Statement

Structure to Consider:

if c **then**
  S_1
else
  S_2
end if

Worst-Case Running Time: if

- worst-case running time to evaluate c is T,
- worst-case running time of S_1 is T_1, and
- worst-case running time of S_2 is T_2,

then

- worst-case running time of program is:

Worse of the two + Condition

Case: Program is a Loop

Structure to Consider:

while G **do**
  S
end while

We need to know:

- the worst-case cost to evaluate G
- the worst-case cost to execute S
- the maximum number of executions of the loop body

Problem:

- it is not even clear that this will halt!

First Objective: Counting Executions of the Loop Body

Recall that a *Loop Variant* is an integer-valued function f_L of variables such that

- the value of f_L decreases by at least 1 each time loop body is executed;
- the test G is **false** if the value of f_L is ≤ 0

The *existence* of a loop variant implies that the loop terminates if each evaluation of G and each execution of the loop body terminates.

Useful fact: number of executions of loop body is *less than or equal to* the value of f_L immediately before execution of the loop begins

Next Objective: Bounding Total Running Time

Suppose:

- Loop body is executed at most k times
- Worst-case cost for each evaluation of the loop test G is $\leq T_1$
- Worst-case cost for each execution of the loop body S is $\leq T_2$

Then:

- *Total* cost for *all* evaluations of test G is at most: (k+1)T1
- *Total* cost for *all* executions of loop body is at most: kT2
- Therefore, the *total* cost to execute the loop is at most: (k+1)T1 + kT2

If cost of jth iteration of S is $T_2(j)$:

(k+1)T1 + Sum from j = 1 to k of T2(j)

Example

Suppose A is an integer array with length n, *key* is an integer, and the following code is executed.

$i := 0$
while $((i < n)$ and $(A[i] <> key))$ **do**
  $i := i + 1$
end while

Loop Variant for this program's loop:

i increaees after each iteration, so f(n,i) decreases

-
- f(n, i)<=0 if i >= n and the loop terminates if i >= n

What about 2nd condition in test?

Example, Continued

Maximum number of executions of the loop body:

- f(n,0) = n-0=n

Worst-case cost to evaluate test:

- 3 units (two comparisons, one boolean operation), or constant c

Worst-case cost for an execution of the loop body:

- 2(addition and assignment), or c2

Upper bound on worst-case cost to execute the loop:

-
- 3(n+1) + 2n = 5n + 3

Case: Program is a Nested Loop

Structure to Consider:

while G_1 **do**
  while G_2 **do**
    S
  end while
end while

Method:

- Compute worst case of inside loop.
- Compute cost of outer loop using computed inner loop cost as worst case cost of outer loop body.

Case: Program Calls Itself a Constant Number of Times

Example: Fibonacci Number Program

`int Fib(n)`
  if $n == 0$ **then**
    return 0
  else if $n == 1$ **then**
    return 1
  else
    return $\text{Fib}(n-1) + \text{Fib}(n-2)$
  end if

n==0, n==1, n-1, n-2, Fib+Fib, return

Objective: Writing an Expression for the Running Time

Let $T(n)$ be the number of steps used on input n. Then

$T(n) \leq$
2 when n = 0
3 when n = 1
6 + T(n-1)+T(n-2) when n >= 2
(Add steps from previous if/elses)

This is an example of a *recurrence relation*:

- $T(n)$ expressed using the same function T evaluated at **smaller** inputs
- Explicit (non-recursive) values of T given for small inputs n (base cases)

Analysis of Recursive Programs

The following exercises on computing bounds on $T(n)$ can be solved using *mathematical induction*.

Exercises:

1. Use the above information to prove that

$$T(n) \leq 6 \times 2^n - 6$$

for every integer $n \geq 0$.

2. Use the above information to prove that

$$T(n) \leq 6 \times \text{fib}(n+1) - 6$$

for every integer $n \geq 0$.

Finding a *Lower Bound*

In order to prove that the worst-case running time of a program P is *at least* T, for input size N (for a fixed N):

- Find a valid input I of size N (where "valid" means that P's precondition is satisfied)
- Count the number of steps used by P on input I
- *If* this number is greater than or equal to T *then* you have proved what we want!

Why This Works:

- worst-case cannot be less than the running time of any particular input

Finding a *Lower* Bound, Continued

In order to prove that the worst-case running time of a program P is *at least* $T(n)$, for a function $T(n)$:

- Find a collection $I_1, I_2, I_3, I_4, \ldots$ of inputs, where I_i is a valid input of size i for all $i \geq 1$
- Show that the number of steps used by P on input I_i is greater than or equal to $T(i)$, for every integer $i \geq 1$

A Common Mistake

Some people try to prove that the worst-case running time of a program P is *at most* $T(n)$, for a function $T(n)$, by doing the following:

- They give a collection $I_1, I_2, I_3, \ldots$ of inputs, where I_i is a valid input of size i for all $i \geq 1$
- They show (generally, correctly) that the number of steps used by P on input I_i is less than or equal to $T(i)$, for every integer $i \geq 1$.
- They then conclude that the worst-case running time of P on inputs of size n is at most $T(n)$ (for all n)

Why This is Incorrect:

- This does not prove that there are no inputs for which the running time is larger

Further Reading . . .

Introduction to Algorithms, Sections 2.2-2.3

- includes *much* more material about this topic