Recurrences

David Kauchak
cs140
Spring 2024

1

Administrative

How was assignment 0?

Mentor hours posted

Group assignment: must attend mentor hours on Thursday or Friday and submit group assignment

Assignment 1 (due Sunday): must work with different partner

2

Big O: Upper bound

$O(g(n))$ is the set of functions:

$O(g(n)) = \left\{ f(n) : \text{there exists positive constants } c \text{ and } n_0 \text{ such that } 0 \leq f(n) \leq cg(n) \text{ for all } n \geq n_0 \right\}$

3

Proving bounds: find constants that satisfy inequalities

Show that $5n^2 - 15n + 100$ is $O(n^2)$

Step 1: Prove $O(n^2)$ – Find constants $c$ and $n_0$ such that

$5n^2 - 15n + 100 \leq cn^2$ for all $n > n_0$

$c n^2 \geq 5n^2 - 15n + 100$

$c \geq 5 - 15/n + 100/n^2$

Let $n_0 = 1$ and $c = 5 + 100 = 105$. $100/n^2$ only gets smaller as $n$ increases and we ignore $-15/n$ since it only varies between $-15$ and 0.

4
Proving bounds

Step 2: Prove $\Omega(n^2)$ – Find constants $c$ and $n_0$ such that $5n^2 - 15n + 100 \geq cn^2$ for all $n > n_0$

- Find constants $c$ and $n_0$ such that $5n^2 - 15n + 100 \geq cn^2$ for all $n > n_0$
- Let $n_0 = 4$ and $c = 5 - 15/4 = 1.25$ (or anything less than 1.25).
- $-15/n$ is always increasing and we ignore $100/n^2$ since it is always between 0 and 100.

Bounds

Is $5n^2 \in \Omega(n)$? No

How would we prove it?

$\Omega(g(n)) = \left\{ f(n) : \text{there exists positive constants } c \text{ and } n_0 \text{ such that } 0 \leq f(n) \leq cg(n) \text{ for all } n \geq n_0 \right\}$

Disproving bounds

Is $5n^2 \notin \Omega(n)$?

Assume it’s true.

That means there exists some $c$ and $n_0$ such that

- $5n^2 \leq cn$ for $n > n_0$
- $5n \leq c$ contradiction

Divide and Conquer

Divide: Break the problem into smaller sub-problems

Conquer: Solve the sub-problems. Generally, this involves waiting for the problem to be small enough that it is trivial to solve (i.e. 1 or 2 items)

Combine: Given the results of the solved sub-problems, combine them to generate a solution for the complete problem
Divide and Conquer: some thoughts

Often, the sub-problem is the same as the original problem

Dividing the problem in half frequently does the job

May have to get creative about how the data is split

Splitting tends to generate run times with log n in them

Divide and conquer

One approach:
- Pretend like you have a working version of your function, but it only works on smaller sub-problems

- If you split up the current problem in some way (e.g. in half) and solved those sub-problems, how could you then get the solution to the larger problem?

MergeSort

\[
\text{MERGE-SORT}(A) \\
1 \text{ if } \text{length}[A] == 1 \\
2 \quad \text{return } A \\
3 \text{ else } \\
4 \quad q \leftarrow \lfloor \text{length}[A] / 2 \rfloor \\
5 \quad \text{create arrays } L[1..q] \text{ and } R[q+1..\text{length}[A]] \\
6 \quad \text{copy } A[1..q] \text{ to } L \\
7 \quad \text{copy } A[q+1..\text{length}[A]] \text{ to } R \\
8 \quad \text{LS } \leftarrow \text{MERGE-SORT}(L) \\
9 \quad \text{RS } \leftarrow \text{MERGE-SORT}(R) \\
10 \quad \text{return } \text{MERGE}(LS, RS)
\]

MergeSort: Merge

Assuming L and R are sorted already, merge the two to create a single sorted array

\[
\text{MERGE}(L, R) \\
1 \quad \text{create array } B \text{ of length } \text{length}[L] + \text{length}[R] \\
2 \quad i \leftarrow 1 \\
3 \quad j \leftarrow 1 \\
4 \quad \text{for } k \leftarrow 1 \text{ to } \text{length}[B] \\
5 \quad \quad \text{if } j > \text{length}[R] \text{ or } (i \leq \text{length}[L] \text{ and } L[i] \leq R[j]) \\
6 \quad \quad \quad B[k] \leftarrow L[i] \\
7 \quad \quad \quad i \leftarrow i + 1 \\
8 \quad \quad \text{else} \\
9 \quad \quad \quad B[k] \leftarrow R[j] \\
10 \quad \quad \quad j \leftarrow j + 1 \\
11 \quad \text{return } B
\]
Merge

L: 1 3 5 8   R: 2 4 6 7

```plaintext
Merge(L, R)
1 create array B of length length[L] + length[R]
2 i ← 1
3 j ← 1
4 for k ← 1 to length[L]
5 if j > length[R] or (j ≤ length[R] and L[i] ≤ R[j])
6 B[k] ← L[i]
7 i ← i + 1
8 else
9 B[k] ← R[j]
10 j ← j + 1
11 return B
```

14

Merge

L: 1 3 5 8   R: 2 4 6 7

```plaintext
Merge(L, R)
1 create array B of length length[L] + length[R]
2 i ← 1
3 j ← 1
4 for k ← 1 to length[L]
5 if j > length[R] or (j ≤ length[R] and L[i] ≤ R[j])
6 B[k] ← L[i]
7 i ← i + 1
8 else
9 B[k] ← R[j]
10 j ← j + 1
11 return B
```

15

Merge

L: 1 3 5 8   R: 2 4 6 7

```plaintext
Merge(L, R)
1 create array B of length length[L] + length[R]
2 i ← 1
3 j ← 1
4 for k ← 1 to length[L]
5 if j > length[R] or (j ≤ length[R] and L[i] ≤ R[j])
6 B[k] ← L[i]
7 i ← i + 1
8 else
9 B[k] ← R[j]
10 j ← j + 1
11 return B
```

16

Merge

L: 1 3 5 8   R: 2 4 6 7

```plaintext
Merge(L, R)
1 create array B of length length[L] + length[R]
2 i ← 1
3 j ← 1
4 for k ← 1 to length[L]
5 if j > length[R] or (j ≤ length[R] and L[i] ≤ R[j])
6 B[k] ← L[i]
7 i ← i + 1
8 else
9 B[k] ← R[j]
10 j ← j + 1
11 return B
```

17
Merge

L: 1 3 5 8  R: 2 4 6 7

B: 1 2 3 4

Merge

L: 1 3 5 8  R: 2 4 6 7

B: 1 2 3 4
merge(L, R)
1. create array B of length length(L) + length(R)
2. i = 1
3. j = 1
4. for k = 1 to length(B)
5. if j > length(R) or (i ≤ length(L) and L[i] ≤ R[j])
   6. B[k] = L[i]
   7. i = i + 1
8. else
10. j = j + 1
11. return B
Merge

L: 1 3 5 8

R: 2 4 6 7

B: 1 2 3 4 5 6 7

Running time?

36
**Merge**

Running time? $\Theta(n)$ - linear

```plaintext
Merge(L, R)
1 create array B of length $length[L] + length[R]
2 i ← 1
3 j ← 1
4 for k ← 1 to $length[B]
5   if $j > length[R]$ or ($i \leq length[L] and L[i] \leq R[j]$)
6       $B[k] ← L[i]
7       i ← i + 1
8   else
9       $B[k] ← R[j]
10      j ← j + 1
11 return B
```

**MergeSort**

Running time?

```plaintext
MergeSort(A)
1 if $length[A] = 1$
2 return $A$
3 else
4   $q ← \frac{length[A]}{2}$
5   create arrays $L[1..q]$ and $R[q+1..length[A]]$
6   copy $A[1..q]$ to $L$
7   copy $A[q+1..length[A]]$ to $R$
8   $LS ← MergeSort(L)$
9   $RS ← MergeSort(R)$
10 return Merge(LS, RS)
```

**Merge-Sort**

Running time?

$$T(n) = \begin{cases} 
  c & \text{if } n \text{ is small} \\
  2T(n/2) + D(n) + C(n) & \text{otherwise}
\end{cases}$$

$D(n)$: cost of splitting (dividing) the data

$C(n)$: cost of merging/combining the data

**Merge-Sort**

Running time?

$$T(n) = \begin{cases} 
  c & \text{if } n \text{ is small} \\
  2T(n/2) + D(n) + C(n) & \text{otherwise}
\end{cases}$$

$D(n)$: cost of splitting (dividing) the data - linear $\Theta(n)$

$C(n)$: cost of merging/combining the data – linear $\Theta(n)$
Merge-Sort

Running time?

\[ T(n) = \begin{cases} 
  c & \text{if } n \text{ is small} \\
  2T(n/2) + cn & \text{otherwise}
\end{cases} \]

Which is?
Merge-Sort

\[ T(n) = \begin{cases} 
  c & \text{if } n \text{ is small} \\
  \frac{c}{2T(n/2) + cn} & \text{otherwise} 
\end{cases} \]

Depth?

We can calculate the depth, by determining when the recursion gets to down to a small problem size, e.g. 1

At each level, we divide by 2

\[ \frac{n}{2^d} = 1 \]

\[ 2^d = n \]

\[ \log 2^d = \log n \]

\[ d \log 2 = \log n \]

\[ d = \log_2 n \]
Merge-Sort

Running time?
- Each level costs $cn$
- $\log n$ levels

$cn \log n = \Theta(n \log n)$

Why don’t we write it as $n \log_2 n$?

Log properties

Logarithm properties:

$\log_a b = \frac{\log b}{\log a}$

$n \log_2 n = \frac{n \log n}{\log 2}$

$n \log_2 n = \frac{n \log n}{c} = \Theta(n \log n)$

Recurrence

A function that is defined with respect to itself on smaller inputs

$T(n) = 2T(n/2) + n$

$T(n) = 16T(n/4) + n$

$T(n) = 2T(n-1) + n^2$

Why are we interested in recurrences?

Computational cost of divide and conquer algorithms

$T(n) = aT(n/b) + D(n) + C(n)$

- $a$ subproblems of size $n/b$
- $D(n)$ the cost of dividing the data
- $C(n)$ the cost of recombining the subproblem solutions

In general, the runtimes of most recursive algorithms can be expressed as recurrences.
The challenge

Recurrences are often easy to define because they mimic the structure of the program

But... they do not directly express the computational cost, i.e. $n$, $n^2$, ...

We want to remove self-recurrence and find a more understandable form for the function

Three approaches

Substitution method: when you have a good guess of the solution, prove that it’s correct

Recursion-tree method: If you don’t have a good guess, the recursion tree can help
  - Calculate exactly (like we did with MergeSort)
  - Use it to get a good guess, then prove with substitution method.

Master method: Provides solutions for recurrences of the form:

$$T(n) = aT(n/b) + f(n)$$

Substitution method

Guess the form of the solution
Then prove it’s correct by induction

$$T(n) = T(n/2) + d$$

Halves the input then a constant amount of work

Guesses?

Substitution method

Guess the form of the solution
Then prove it’s correct by induction

$$T(n) = T(n/2) + d$$

Halves the input then a constant amount of work

Similar to binary search:
  - Guess: $O(\log n)$
Proof?

Proof by induction:
- Assume it’s true for smaller $T(k)$, i.e. $k < n$
- Prove that it’s then true for current $T(n)$

$T(n) = T(n/2) + d = O(\log n)$

Key question: does a constant exist such that:
$T(n) \leq c \log n - c \log 2 + d$ residual

How do we now prove $T(n) = O(\log n)$?

To prove that $T(n) = O(\log n)$ identify the appropriate constants:

$O(g(n)) = \left\{ f(n) : \right.$ there exists positive constants $c$ and $n$ such that $\left. 0 \leq f(n) \leq cg(n) \text{ for all } n \geq n_0 \right\}$

i.e. some constant $c'$ such that $T(n) \leq c' \log n$

$T(n) = T(n/2) + d$

$\leq c \log \left( \frac{n}{2} \right) + d$ from our inductive hypothesis

$\leq c \log n - c \log 2 + d$

$\leq c \log n - c + d$ residual

Assume $T(k) = O(\log k)$ for all $k < n$
Show that $T(n) = O(\log n)$

From our assumption, $T(n/2) = O(\log n/2)$:

From the definition of big-O: $T(n/2) \leq c \log(n/2)$

How do we now prove $T(n) = O(\log n)$?

Ideas?
To prove that \( T(n) = O(\log n) \) identify the appropriate constants:

\[
O(g(n)) = \begin{cases} 
\text{there exists positive constants } c \text{ and } n \text{ such that} \\
0 \leq f(n) \leq cg(n) \text{ for all } n \geq n_0
\end{cases}
\]

i.e. some constant \( c' \) such that \( T(n) \leq c' \log n \)

Key question: does a constant exist such that:

\[
T(n) \leq c' \log n
\]

if \( c' \geq d \), then, just let \( c' = c \)

\[
T(n) \leq c \log n - c + d \leq c \log n
\]

if \( c' < d \), let \( c' = d+1 \) and

\[
T(n) \leq c \log n - c + d \leq d \log n + \log n
\]
Substitution method

Guess the form of the solution
Then prove it's correct by induction

\[ T(n) = T(n/2) + d \]

Halves the input then constant amount of work
Similar to binary search:

Guess: \( O(\log_2 n) \)
\[ T(n) = 2T(n/2) + n \]

What if we guess wrong, e.g. \( O(n) \)?

Assume \( T(k) = O(k) \) for all \( k < n \)
- again, this implies that \( T(n/2) \leq c(n/2) \)
Show that \( T(n) = O(n) \)

\[ T(n) = 2T(n/2) + n \]
\[ \leq 2cn/2 + n \]
\[ = cn + n \]
\[ \leq cn \]

factor of \( n \) so we can just roll it in?

\[ T(n) = 2T(n/2) + n \]
\[ \leq 2cn/2 + n \]
\[ = cn + n \]
\[ \leq cn \]

factor of \( n \) so we can just roll it in?

\[ T(n) = 2T(n/2) + n \]
\[ \leq 2cn/2 + n \]
\[ = cn + n \]
\[ \leq cn \]

factor of \( n \) so we can just roll it in?

\[ T(n) = 2T(n/2) + n \]
\[ \leq 2cn/2 + n \]
\[ = cn + n \]
\[ \leq cn \]

factor of \( n \) so we can just roll it in?