

1

| Big O: Upper bound <br> $O(g(n))$ is the set of functions: $O(n))=\left\{\begin{array}{ll} f(n): & \begin{array}{l} \text { there exists positive constants } c \text { and } n_{0} \text { such that } \\ 0 \leq f(n) \leq C g(n) \text { for all } n \geq n_{0} \end{array} \end{array}\right\}$ |
| :---: |

3

Proving bounds: find constants that satisfy inequalities

Show that $5 n^{2}-15 n+100$ is $\Theta\left(n^{2}\right)$
Step 1: Prove $O\left(n^{2}\right)$ - Find constants $c$ and $n_{0}$ such that $5 n^{2}-15 n+100 \leq c n^{2}$ for all $n>n_{0}$

$$
\begin{aligned}
c n^{2} & \geq 5 n^{2}-15 n+100 \\
c & \geq 5-15 / n+100 / n^{2}
\end{aligned}
$$

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

## Proving bounds

Step 2: Prove $\Omega\left(n^{2}\right)-$ Find constants $c$ and $n_{0}$ such that $5 n^{2}$ $-15 n+100 \geq c n^{2}$ for all $n>n_{0}$

$$
\begin{aligned}
c n^{2} & \leq 5 n^{2}-15 n+100 \\
c & \leq 5-15 / n+100 / n^{2}
\end{aligned}
$$

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 $100 / \mathrm{n}^{2}$ since it is always between 0 and 100.

## Bounds

Is $5 n^{2} O(n)$ ? No
How would we prove it?


5


7

## 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

9

## MergeSort

```
Merge-Sort(A)
    if length[A]==1
    return A
    else
            q\leftarrow[length }[A]/2
            create arrays }L[1..q] and R[q+1.. length[A]
            copy }A[1..q] to L
            copy }A[q+1..length[A]] to 
            LS}\leftarrow\operatorname{Merge-Sort(L)
            RS}\leftarrowMERge-Sort(R
            return Merge(LS, RS)
```


## MergeSort: Merge

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

```
Merge(L,R)
    create array B of length length[L]+length[R]
    i\leftarrow1
    j\leftarrow1
    for }k\leftarrow1\mathrm{ to length[B]
            if j>length[R] or (i\leqlength[L] and L[i] \leqR[j])
                B[k]\leftarrowL[i]
            i\leftarrowi+1
            else
            B[k]\leftarrowR[j]
            j\leftarrowj+1
    return B
```

Ierge(L,R)
Ierge(L,R)
create array B of length length[L]+ length[R]
create array B of length length[L]+ length[R]
2 i\leftarrow1
2 i\leftarrow1
for }k\leftarrow1\mathrm{ to length[B]
for }k\leftarrow1\mathrm{ to length[B]
if j>length[R] or (i\leqlength[L] and L[i]\leqR[j])
if j>length[R] or (i\leqlength[L] and L[i]\leqR[j])
B[k]\leftarrowL[i]
B[k]\leftarrowL[i]
B[k]\leftarrowL[l]
B[k]\leftarrowL[l]
else
else
\ B[k]\leftarrowR[j]
\ B[k]\leftarrowR[j]
return B
return B

## Merge

14

## Merge

L: 1358 R: 2467
B:
Merge(L.R)
Merge(L.R)
1 create array B of length length [L] + length [R]
1 create array B of length length [L] + length [R]
2
2
3 j}
3 j}
for }k\leftarrow1\mathrm{ to length[B]
for }k\leftarrow1\mathrm{ to length[B]
B[l]\leftarrowL[l]
B[l]\leftarrowL[l]
else
else
N[k]\leftarrowR[j]
N[k]\leftarrowR[j]
15


16


17


18


19


20


21


22


23


24


25


26


27


28


29


30


31
L: $1358^{\downarrow^{i}} \quad$ R: $2467^{\downarrow^{j}}$
B: 12345678
Merge(L,R)
Merge(L,R)
create array B of length length[L]+ length [R]
create array B of length length[L]+ length [R]
2 i
2 i
for }k\leftarrow1\mathrm{ to length[B]
for }k\leftarrow1\mathrm{ to length[B]
5 if j> length[R] or (i\leqlength[L] and L[i]\leqR[j])
5 if j> length[R] or (i\leqlength[L] and L[i]\leqR[j])
else
else
return B
return B

32


36

## Merge

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

```
Merge(L,R)
create array B of length length[L]+length[R]
2 i
3 j}\leftarrow
for }k\leftarrow1\mathrm{ to length[B]
if j> length[R] or ( }i\leq\mathrm{ length [L] and L[i] }\leqR[j]
                B[k]}\leftarrowL[i
            i\leftarrowi+1
    else
        B[k]\leftarrowR[j]
        j\leftarrowj+1
    return B
```


## MergeSort

Running time?

## $\operatorname{Merge-Sort}(A)$

if length $[A]==1$
return A

## else

$q \leftarrow\lfloor$ length $[A] / 2\rfloor$
create arrays $L[1 . . q]$ and $R[q+1$.. length $[A]]$
copy $A[1 . . q]$ to $L$
copy $A[q+1$.. length $[A]]$ to $R$
$L S \leftarrow \operatorname{Merge-Sort}(\mathrm{~L})$
$R S \leftarrow \operatorname{Merge-Sort}(\mathrm{R})$
return $\operatorname{MERGE}(L S, R S)$

37

## Merge-Sort

Running time?

$$
T(n)=\left\{\begin{array}{cc}
c & \text { if } n \text { is small } \\
2 T(n / 2)+D(n)+C(n) & \text { otherwise }
\end{array}\right.
$$

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


41


43


44


45


46


47

## Merge-Sort

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

$$
\begin{aligned}
\frac{n}{2^{d}} & =1 \\
2^{d} & =n \\
\log 2^{d} & =\log n \\
d \log 2 & =\log n \\
d & =\log _{2} n
\end{aligned}
$$

48


49

| Recurrence |  |
| :--- | :--- |
| A function that is defined with respect to itself on |  |
| smaller inputs |  |
| $T(n)=2 T(n / 2)+n$ |  |
| $T(n)=16 T(n / 4)+n$ |  |
| $T(n)=2 T(n-1)+n^{2}$ |  |
|  |  |

51

## Why are we interested in recurrences?

Computational cost of divide and conquer algorithms

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

- 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. $n, n^{2}, \ldots$

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

53

## 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)


57

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

Assume $T(k)=O(\log k)$ for all $k<n$
Show that $T(n)=O(\log n)$
From our assumption, $T(n / 2)=O(\log n / 2)$ :

$$
O(g(n))=\left\{\begin{array}{ll}
f(n): & \begin{array}{l}
\text { there exists positive constants } c \text { and } n \text { such that } \\
0 \leq f(n) \leq \operatorname{cg}(n) \text { for all } n \geq n_{0}
\end{array}
\end{array}\right\}
$$

From the definition of big- $O: T(n / 2) \leq c \log (n / 2)$
How do we now prove $T(n)=O(\log n)$ ?

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

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

$$
O(g(n))=\left\{\begin{array}{ll}
f(n): & \begin{array}{l}
\text { there exists positive constants } c \text { and } n \text { such that } \\
0 \leq f(n) \leq \operatorname{cg}(n) \text { for all } n \geq n_{0}
\end{array}
\end{array}\right\}
$$

i.e. some constant $c^{\prime}$ such that $T(n) \leq c^{\prime} \log n$

$$
\begin{aligned}
\mathrm{T}(n) & =T(n / 2)+d \\
& \leq c \log \left(\frac{n}{2}\right)+d \quad \text { from our inductive hypothesis } \\
& \leq c \log n-c \log 2+d \\
& \leq c \log n-c+d \text { residual }
\end{aligned}
$$

Key question: does a constant exist such that: $T(n) \leq c^{\prime} \log n$

60
if $c \geq d$ ，then，just let $\mathrm{c}^{\prime}=\mathrm{c}$

$$
\mathrm{T}(\mathrm{n}) \leq c \log n-c+d \leq c \log n
$$

61

## Base case？

For an inductive proof we need to show two things：
－Show that it holds for some base case
－Assuming it＇s true for $k<n$ show it＇s true for $n$

What is the base case in our situation？

$$
T(n)=\left\{\begin{array}{cc}
\Theta(1) & \text { if } n \text { is small } \\
T(n / 2)+d & \text { otherwise }
\end{array}\right.
$$

63

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

Guess the solution？
At each iteration，does a linear amount of work（i．e． iterate over the data）and reduces the size by one at each step
$O\left(n^{2}\right)$

Assume $T(k)=O\left(k^{2}\right)$ for all $k<n$
－again，this implies that $T(n-1) \leq c(n-1)^{2}$
Show that $T(n)=O\left(n^{2}\right)$ ，i．e．$T(n) \leq c^{\prime} n^{2}$

64

$$
\begin{aligned}
T(n) & =T(n-1)+n \\
& \leq c(n-1)^{2}+n \quad \text { from our inductive hypothesis } \\
& =c\left(n^{2}-2 n+1\right)+n \\
& =c n^{2}-2 c n+c+n \text { residual } \\
& \text { if } \quad-2 c n+c+n \leq 0
\end{aligned}
$$

then let $\mathrm{c}^{\prime}=\mathrm{c}$ and there exists a constant such that $T(n) \leq c^{\prime} n^{2}$

## Substitution method

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

Guess the solution?
Recurses into 2 sub-problems that are half the size and performs some operation on all the elements $O(n \log n)$

What if we guess wrong, e.g. $\mathrm{O}\left(n^{2}\right)$ ?

Assume $T(k)=O\left(k^{2}\right)$ for all $k<n$

- again, this implies that $T(n / 2) \leq c(n / 2)^{2}$

Show that $T(n)=O\left(n^{2}\right)$


69

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

What if we guess wrong, e.g. $\mathrm{O}(n)$ ?

Assume $T(k)=O(k)$ for all $k<n$

- again, this implies that $T(n / 2) \leq c(n / 2)$ Show that $T(n)=O(n)$


71

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

Prove $T(n)=O\left(n \log _{2} n\right)$
Assume $T(k)=O\left(k \log _{2} k\right)$ for all $k<n$

- again, this implies that $T(k)=c k \log _{2} k$

Show that $T(n)=O\left(n \log _{2} n\right)$

$$
\begin{aligned}
T(n) & =2 T(n / 2)+n \\
\leq & 2 c n / 2 \log (n / 2)+n \\
\leq & c n\left(\log _{2} n-\log _{2} 2\right)+n \\
\leq & c n \log _{2} n-c n+n \quad \text { residual } \\
\leq & c n \log _{2} n \\
& \quad \text { if } c n \geq n, c>1
\end{aligned}
$$

72

