









f( <i>n</i> )	Number	of data ite	ems proces	sed per:
	1 minute	1 day	1 year	1 century
n	10	14,400	5.26.106	5.26·10 <sup>8</sup>
n log n	10	3,997	883,895	6.72·10 <sup>7</sup>
n <sup>1.5</sup>	10	1,275	65,128	1.40.106
<i>n</i> <sup>2</sup>	10	379	7,252	72,522
<i>n</i> <sup>3</sup>	10	112	807	3,746
2 <sup>n</sup>	10	20	29	35

# Beware exponential complexity

- ☺ If a linear, O(*n*), algorithm processes 10 items per minute, then it can process 14,400 items per day, 5,260,000 items per year, and 526,000,000 items per century.
- If an exponential, O(2<sup>n</sup>), algorithm processes 10 items per minute, then it can process only 20 items per day and 35 items per century...

## Ascending order of complexity $1 \leftarrow \log \log n \leftarrow \log n \leftarrow n \leftarrow n \log n$ $\leftarrow n^{\alpha}; 1 < \alpha < 2 \leftarrow n^2 \leftarrow n^3 \leftarrow n^m; m > 3 \leftarrow 2^n \dots$ *Questions:* - Where is the place of $n^3 \log n$ ? - Where is the place of $n^{2.79}$ ?

Answer  $1 \leftarrow \log \log n \leftarrow \log n \leftarrow n \leftarrow n \log n$  $\leftarrow n^{\alpha} : 1 < \alpha < 2 \leftarrow n^2 \leftarrow \mathbf{n}^{2.79} \leftarrow n^3 \leftarrow$  $n^{3}\log n \leftarrow n^{m}; m > 3 \leftarrow 2^{n} \dots$ 

- The exact running time function is not important, since it can be multiplied by any arbitrary positive constant.
- Two functions are compared only *asymtotically*, for large *n*, and not near the origin
  - If the constants involved are very large, then the asymptotical behaviour is of no practical interest!

11

## **Example** • Let algorithms **A** and **B** have running times $T_A(n) = 20n \text{ ms}$ and $T_B(n) = 0.1n \log_2 n \text{ ms}$ • In the "Big-Oh"sense, **A** is better than **B**... • But: on which data volume can **A** outperform **B**? $T_A(n) < T_B(n)$ if $20n < 0.1n \log_2 n$ , or $\log_2 n > 200$ , that is, when $n > 2^{200} \approx 10^{60}$ !

 Thus, in all practical cases B is better than A...

#### Example

Let algorithms **A** and **B** have running times  $T_A(n) = 20n \text{ ms}$  and  $T_B(n) = 0.1n^2 \text{ ms}$ 

- In the "Big-Oh"sense, A is better than B...
- But: on which data volumes A outperforms B?

 $T_{\rm A}(n) < T_{\rm B}(n)$  if  $20n < 0.1n^2$ , or n > 200

• Thus **A** is better than **B** in most practical cases except for *n* < 200 when **B** becomes faster...

13

#### Examples

Running time T( <i>n</i> )	Complexity O(n)
$n^2 + 100 n + 1$	$O(n^2)$
$0.001n^3 + n^2 + 1$	$O(n^3)$
23 n	O( <i>n</i> )
$2^{3n}$	$O(8^n)$ as $2^{3n} \equiv (2^3)^n$
$2^{3+n}$	$O(2^n)$ as $2^{3+n} \equiv 2^3 \cdot 2^n$
$2 \cdot 3^n$	O(3 <sup><i>n</i></sup> )

Running time T(n)	Complexity O(n)
$0.0001 \ n + 10000$	O(n)
100000 n + 10000	O(n)
$0.0001 n^2 + 10000 n$	$O(n^2)$
$100000 n^2 + 10000 n$	$O(n^2)$
$30 \log_{20}(23n)$	$O(\log n)$ as
actually NOT that	$\log_{c}(ab) = \log_{c}a + \log_{c}b$
hard	

#### Example

Assume that we have an algorithm with a running time

 $T(n) = \sqrt{n}$ 

that can process x input items on your computer. What input size it can process if the computer is 100 times faster?

 $\sqrt{n_{new}} = 100 \sqrt{x}$  $n_{new} = 10^{4} x$ 

#### Worst-Case Performance

- · Upper bounds: simple to obtain
- Lower bounds: a difficult matter...
- Worst case data may be unlikely to be met in practice
- Unknown "Big-Oh" constants c and  $n_0$  may not be small
- Inputs in practice lead to much lower running times
- **Example:** the most popular fast sorting algorithm, **QuickSort**, has O(n<sup>2</sup>) running time in the worst case but in practice the time is O(n log n)

17

## Average-Case Performance

- Estimate average time for each operation
- Estimate frequencies of operations
- May be a difficult challenge... (take COMPSCI.320 for details)
- May be no natural "average" input at all
- May be hard to estimate (average time for each operation depends on data)

18

#### **Recurrence Relations**

#### Divide-and-conquer principle:

- to divide a large problem into smaller ones and recursively solve each subproblem, then
- to combine solutions of the subproblems to solve the original problem
- **Running time**: by a *recurrence relation* combining the size and number of the subproblems and the cost of dividing the problem into the subproblems

19

#### "Telescoping" a Recurrence

• Recurrence relation and its base condition (i.e., the difference equation and initial condition):

 $T(n) = 2 \cdot T(n-1) + 1; T(0) = 0$ 

20

• Closed (explicit) form for T(*n*) by "telescoping":



"Telescoping" = Substitution  

$$T(n) = 2T(n-1)+1$$
  
 $2T(n-1) = 2^{2}T(n-2)+2$   
 $2^{2}T(n-2) = 2^{3}T(n-3)+2^{2}$   
...  
 $2^{n-1}T(2) = 2^{n}T(0)+2^{n-1}$   
 $T(n) = 1+2+2^{2}+...+2^{n-1} = 2^{n}-1$ 

Basic Recurrence: 1  

$$T(n) = T(n-1) + n \iff T(n) = \frac{n(n+1)}{2}$$

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

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

$$T(1) = 1$$
23

1: Explicit Expression for T(n)  

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

$$= \overline{T(n-2) + (n-1)} + n$$

$$= \overline{T(n-3) + (n-2)} + (n-1) + n$$

$$= \overline{T(2) + 3} + \dots + (n-2) + (n-1) + n$$

$$= \overline{T(1) + 2} + \dots + (n-2) + (n-1) + n$$

$$= \overline{1} + 2 + \dots + (n-2) + (n-1) + n = \frac{n(n+1)}{2}$$

#### Guessing to Solve a Recurrence

- Guess a hypothetic solution T(n); n ≥ 0 from a sequence of numbers T(0), T(1), T(2), ..., obtained from the recurrence relation
- Prove T(n) by math induction: <u>Base condition</u>: T holds for n = n<sub>base</sub>, e.g. T(0) or T(1) <u>Induction hypothesis to verify</u>: for every n > n<sub>base</sub>,

if T holds for n - 1, then T holds for n **Strong induction**: if T holds for  $n_{\text{base}}, ..., n - 1$ , then...

25

Explicit Expression for T(n)

- T(1) = 1; T(2) = 1 + 2 = 3; T(3) = 3 + 3 = 6; T(4) = 6 + 4 = 10 ⇒ Hypothesis:
- Base condition holds: T(1) = 1.2 / 2 = 1
- If the hypothetic closed-form relationship T(n) holds for n 1 then it holds also for n:
- Thus, the expression for T(*n*) holds for all *n* > 1



2: Explicit Expression for T(*n*)  

$$T(2^{m}) = T(2^{m-1}) + 1$$

$$= T(2^{m-2}) + 1 + 1$$

$$\dots = T(2^{1}) + 1 + 1 + \dots + 1$$

$$= T(2^{0}) + 1 + 1 + \dots + 1 + 1$$

$$T(2^{m}) = m \implies T(n) = \log_{2} n$$

Basic Recurrence: 3  
• Scan and halve the input:  

$$T(n) = T(n/2) + n \iff T(n) \cong 2n$$
  
• "Telescoping" (for  $n = 2^m$ ):  
 $T(2^m) = T(2^{m-1}) + 2^m$   $T(2^2) = T(2^1) + 2^2$   
 $T(2^{m-1}) = T(2^{m-2}) + 2^{m-1}$   $T(2^1) = T(2^0) + 2^1$   
...  $T(2^0) = 1$ 

3: Explicit Expression for T(*n*)  

$$T(2^{m}) = T(2^{m-1}) + 2^{m}$$

$$= \overline{T(2^{m-2}) + 2^{m-1}} + 2^{m}$$

$$= \overline{T(2^{1}) + 2^{2}} + \dots + 2^{m-1} + 2^{m}$$

$$= \overline{T(2^{0}) + 2^{1}} + 2^{2} + \dots + 2^{m-1} + 2^{m}$$

$$T(2^{m}) = 2^{m+1} - 1 \implies T(n) \cong 2n$$



4: Explicit Expression for T(n)  

$$T(2^{m})/2^{m} = T(2^{m-1})/2^{m-1} + 1$$

$$= T(2^{m-2})/2^{m-2} + 1 + 1$$

$$= T(2^{1})/2^{1} + 1 + ... + 1 + 1$$

$$= T(2^{0})/2^{0} + 1 + 1 + ... + 1 + 1$$

$$= 0 + 1 + ... + 1 = m$$

$$T(2^{m}) = m \cdot 2^{m} \implies T(n) = n \log_{2} n$$





### Capabilities and Limitations

- Rough complexity analysis cannot result immediately in an efficient practical program but it helps in predicting of empirical running time of the program
- "Big-Oh" analysis is unsuitable for small input and hides the constants c and  $n_0$  crucial for a practical task
- "Big-Oh" analysis is unsuitable if costs of access to input data items vary and if there is lack of sufficient memory
- But complexity analysis provides ideas how to develop new efficient methods

35



- Ordering relation: places each pair α, β of *countable* items in a fixed order denoted as (α,β) or <α,β>
- **<u>Order notation</u>**:  $\alpha \leq \beta$  (less than or equal to)
- Countable item: labelled by a specific integer key
- <u>Comparable objects in Java</u>: if an object can be less than, equal to, or greater than another object: object1.compareTo( object2 ) <0, =0, >0







Insertion Sort

· Sequentially contracts the unordered part,

n-i unordered and i ordered

ordered part unordered part

 $a_i, \ldots, a_{n-1}$ 

39

· Splits an array into a unordered and

 $a_0, \ldots, a_{i-1}$ 

ordered parts

one element per stage:

• Stage *i* = 1, ..., *n*-1:

elements





- i + 1 positions to place a next item:  $0 \ 1 \ 2 \ \dots \ i$ -1 i
- i-j+1 comparisons and i-j moves for each position j = i, i-1, ..., 1
- *i* comparisons and *i* moves for position j = 0

$$E_i = \frac{1+2+\ldots+i+i}{i+1} = \frac{i}{2} + \frac{i}{i+1}$$

• 
$$n-1$$
 stages for  $n$  input items: the total average number of comparisons:  

$$E = E_1 + E_2 + \ldots + E_{n-1} = \frac{n^2}{4} + \frac{3n}{4} - H_n$$

$$\left(\frac{1}{2} + \frac{1}{2}\right) + \left(\frac{2}{2} + \frac{2}{3}\right) + \ldots + \left(\frac{n-1}{2} + \frac{n-1}{n}\right)$$

$$= \frac{1}{2}(1+2+\ldots+(n-1)) + \left(\frac{1}{2} + \frac{2}{3} + \ldots + \frac{n-1}{n}\right)$$

$$= \frac{(n-1)n}{4} + n - \left(\frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}\right)$$
•  $H_n \cong \ln n + 0.577$  when  $n \to \infty$  is the  $n$ -th harmonic number

	A	nalysi	s of In	versio	ns	
A 0 >	n <b>inversi</b> ordered pa ⊳ <i>a<sub>j</sub></i> : e.g.,	<b>on</b> in an a air of posit [, 2,,	array $\mathbf{A} =$ ions $(i, j)$ 1] or [10	$[a_1, a_2,$ such that 0,, 35,	, <i>a<sub>n</sub></i> ] is an <i>i</i> ≤ <i>j</i> but <i>a</i> …]	ıy a <sub>i</sub>
	Α	# invers.	A <sub>reverse</sub>	# invers.	Total #	]
	3,2,5	1	5,2,3	2	3	
	3,2,5,1	4	1,5,2,3	2	6	
	1,2,3,4,7	0	7,4,3,2,1	10	10	
						45









Example of ShellSort: step 1											
gap	<i>i</i> :C:		Data to be sorted								
	М	25	8	2	91	70	50	20	31	15	65
5	5:1:0	25					50				
	6:1:0		8					20			
	7:1:0			2					31		
	8:1:1				15					91	
	9:1:1					65					70
		25	8	2	15	65	50	20	31	91	70
											50

	Example of ShellSort: step 2										
gap	<i>i</i> :C: M	25	8	2	15	65	50	20	31	91	70
2	2:1:1	2		25							
	3:1:0		8		15						
	4:1:0			25		65					
	5:1:0				15		50				
	6:3:2	2		20		25		65			
	7:2:1				15		31		50		
	8:1:0							65		91	
	9:1:0								50		70
											51

## Example of ShellSort: step 3

yap	i:C:M	2	8	20	15	25	31	65	50	91	70
1	1:1:0	2	8								
	2:1:0		8	20							
[	3:2:1		8	15	20						
	4:1:0				20	25					
Γ	5:1:0					25	31				
	6:1:0						31	65			
ſ	7:2:1						31	50	65		
	8:1:0								65	91	
	9:2:1								65	70	91

## Time complexity of ShellSort

- Heavily depends on gap sequences
- Shell's sequence: <sup>*n*</sup>/<sub>2</sub>, <sup>*n*</sup>/<sub>4</sub>, ..., 1:
- $O(n^2)$  worst;  $O(n^{1.5})$  average • "Odd gaps only" (if even: gap/2 + 1):
  - $O(n^{1.5})$  worst;  $O(n^{1.25})$  average
- Heuristic sequence:  $g^{ap}/_{2.2}$ : better than  $O(n^{1.25})$
- A very simple algorithm with an extremely complex analysis