Computational Complexity

Solving Computational Problems
- First, get it right.
- Then, make it faster.
- Avoid optimizing prematurely.

Topics
- Algorithm analysis
- Fast algorithm synthesis
- Empirical measurement of complexity

"Complexity" in the algorithm-analysis context
- means the cost of a program’s execution
  - (in running time, memory, ...)
  - rather than
- the cost of creating the program
  - (in # of statements, development time, ...)
- In this context, less-complex programs may require more development time.

Functions associated with a program
- Consider a program with one natural number as input (for simplicity).
- Two functions that can be associated with the program are:
  - \( f(n) \), the function computed by the program
  - \( T(n) \), the running time of the program

Running time \( T(n) \)

It is common to measure \( T \) based on the size of the input, rather than the input value itself.
Possible size measures n for T(n)

- Total number of bits used to encode the input.
- Number of data values in the input (e.g. size of an array)

The second is viewed as an approximation to the first.

Primitive Operations

- These are operations which we don’t further decompose in our analysis.
- They are considered the fundamental building blocks of the algorithm, e.g. + * - / if()
- Typically, the time they take is assumed to be constant.
- Caution: This assumption is not always valid, and may need to be revisited.

Is multiplication really “primitive”?

- In doing arithmetic on arbitrarily-large numbers, the size of the numerals representing those numbers may have a definite effect on the time required.
  - 2 x 2
  - vs. 26378491562329846 x 786431258901237

Size of Numerals

- For a single number (n) input, size of the corresponding numeral is typically on the order of log(n)
- e.g. decimal numeral encoding:
  - size(n) = #digits of n = \[\lceil \log_{10} n \rceil \]
  - \[\lceil x \rceil = \text{smallest integer} \geq x\]
  - (called the “ceiling of” \(x\))

Does Numeral Radix Matter?

- Not much, as long as it’s > 1.
- r-ary numeral encoding:
  - size(n) = #digits of n
    - \(= \lceil \log_r n \rceil \)
  - \(\log_r n\) vs. \(\log_{10} n\)?
- All logs differ by only a constant multiple:
  - \(\log_r n = \log_{10}(2) \log_{10}(n) = 0.301 \log_{10}(n)\)
  - \(\log_{10}(n) = 3.32 \log_{10}(n)\)

\(\log_2\) is the norm
Asymptotic Analysis

- Typically in measuring complexity, we look at the growth-rate of the time function as size increases without bound, rather than the value itself.
- There is therefore a tendency to pay less attention to constant factors in the run-time function.
- This is called asymptotic analysis.
- It is only one part of the story; sometimes constants are important.

Step Counting

- Using exact running time to measure an algorithm requires calibration based on the type of machine, clock rate, etc.
- Instead, we usually just count steps taken in the algorithm.
- Often we will assume primitives take one step each.
- This is usually enough to give us an accurate view of the growth rate of running time.

Straight-Line Code

\[
\begin{align*}
\text{x} &= \text{x} + 1; \\
\text{v} &= \text{x} / \text{z}; \\
\text{w} &= \text{x} + \text{v};
\end{align*}
\]

3 operations, therefore 3 steps
(Not counting assignment as an operation here)

Loop Code

\[
\begin{align*}
\text{for} ( \text{int} \ i = 0; \ i < \text{n}; \ i++) \\
&\{ \\
&\quad \text{x} = \text{x} + 1; \\
&\quad \text{v} = \text{x} / \text{z}; \\
&\quad \text{w} = \text{x} + \text{v}; \\
&\}\n\end{align*}
\]

(These could count as steps too.)
\[\text{n iterations x 5 steps + 2} = 5\text{n+2 steps}\]

Non-Numeric Loop Control

\[
\text{for( OpenList L = A; !L.isEmpty(); L = L.rest(); )}
\{ \\
\quad \text{... loop body ...}
\}
\]

# of iterations = A.length()

Recursive Code

\[
\text{fac(n) = n} \text{ ==} 0 ? 1 : \text{n*fac(n-1)}
\]

3n +1 steps, if we count multiply, -, and comparison as steps.
Steps are involved in the overhead for function calls too. The number of such steps would still be proportional to n in this case.
Recurrence Formulas represent time for recursive expressions

\[
\text{fac}(n) = n \begin{cases} 0 \text{ if } n = 0, & 1 \text{ if } n > 0 \end{cases} \quad n \cdot \text{fac}(n-1)
\]

\[
T(0) \Rightarrow 1;
T(n) \Rightarrow T(n-1) + 3;
\]

Cost of computing \(\text{fac}(n)\) Incremental cost (\(*, -, ==\) comparison)

By McCarthy's Transformation, recurrence formulas can be used for arbitrary imperative computations as well.

Solving Recurrence Formulas

One Method: Repeated Substitution

\[
T(0) \Rightarrow 1;
T(n) \Rightarrow T(n-1) + 3;
\]

\[
T(n) = T(n-1) + 3
= T(n-1) + 3 + 3
= T(n-1-1) + 3 + 3
= \ldots
= T(0) + n \times 3
= 3n + 1
\]

\(T(n): 3n+1\) is a "closed form" solution to the recurrence.

Try These

\[
T(0) \Rightarrow 0;
T(n) \Rightarrow n + T(n-1);
\]

\[
S(1) \Rightarrow 0;
S(n) \Rightarrow 1 + S(\lfloor n/2 \rfloor);
\]

Approximate Solutions

- Rather than solve a recurrence exactly, it is often simpler, yet serves the same purpose, to get an approximate solution.

- More specifically, we'd like a "tight upper bound" on the solution, that is, an approximation that differs from actuality by at most a constant multiple.

Gauss' 3rd Grade Technique

Compute \(1 + 2 + 3 + \ldots + 1000\):

\[1+1000 + 2+999 + 3+998 + \ldots + 500+501 = 500 \times 1001 = 500500\]

In general, \(1+2+3+\ldots+n = \frac{n(n+1)}{2}\) ("arithmetic" series vs. "geometric" series)
**"O" Notation**

- "O" is letter "Oh" (for "order")
- Used to express upper bounds on running time (number of steps)
- \( T \in O(f) \) means that 
  \[
  \forall c \in (\mathbb{R}_0^+ \cup \infty) \forall n_0 \in \mathbb{N} \quad T(n) < c f(n)
  \]
- Think of \( O(f) \) as the set of functions that "grow no faster than" a constant times the value of \( f \).
- This is "big-Oh"; "little-oh" has a different meaning.

**Multiplicative constants are irrelevant in "O" comparisons:**
- If \( g \) is a function, then any function \( n \in \mathbb{N} \) \( d \cdot g(n) \) (using anonymous function notation, a la \( \text{la}\) \( \text{rex} \)), where \( d \) is a constant, is \( \in O(g) \).
- Examples:
  - \( n \in 2.5n^2 \in O(n^2) \)
  - \( n \in 100000000n^2 \in O(n^2) \)

**Notation Abuse**

- It is customary and usual to drop the "n \( \in \)" and just use the body expression of the anonymous function.
- Examples:
  - \( O(n^2) \) instead of \( O(n \cdot n^2) \)
  - \( O(n) \) instead of \( O(n \cdot n) \)
  - \( O(1) \) instead of \( O(n \cdot 1) \)

**Words for Functions**

- A function \( f \) is called
  - constant if \( f \in O(1) \)
  - logarithmic if \( f \in O(\log(n)) \)
  - linear if \( f \in O(n) \)
  - quadratic if \( f \in O(n^2) \)
  - cubic if \( f \in O(n^3) \)
  - polynomial if \( f \in O(n^k) \), for some constant \( k \)
  - exponential if \( f \in O(2^{p(n)}) \), for some polynomial \( p(n) \)
  - factorial if \( f \in O(n!) \)

**Example algorithms**

- \( O(n^2) \):
- \( O(n^3) \):
- \( O(n) \):
- \( O(\log(n)) \):
- \( O(1) \):
- \( O(2^n) \):
- \( O(n!) \):
Why Asymptotic Complexity Matters

Running Time as a function of Complexity

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>log^2n</td>
<td>10.361</td>
<td>44.140</td>
<td>99.317</td>
<td>176.54</td>
<td>273.85</td>
<td>397.24</td>
</tr>
<tr>
<td>sqrt n</td>
<td>3.162</td>
<td>10</td>
<td>31.622</td>
<td>100</td>
<td>316.22</td>
<td>1000</td>
</tr>
<tr>
<td>n</td>
<td>10</td>
<td>100</td>
<td>1000</td>
<td>10000</td>
<td>100000</td>
<td>1000000</td>
</tr>
<tr>
<td>n log n</td>
<td>33.219</td>
<td>664.38</td>
<td>9965.8</td>
<td>132877</td>
<td>16610E5</td>
<td>199658</td>
</tr>
<tr>
<td>n^2</td>
<td>3.16</td>
<td>101</td>
<td>103</td>
<td>106</td>
<td>109</td>
<td>1012</td>
</tr>
<tr>
<td>n^3</td>
<td>1000</td>
<td>10^6</td>
<td>10^9</td>
<td>10^12</td>
<td>10^15</td>
<td>10^18</td>
</tr>
<tr>
<td>n!</td>
<td>3628800</td>
<td>3.6288E6</td>
<td>3.6288E7</td>
<td>3.6288E8</td>
<td>3.6288E9</td>
<td>3.6288E10</td>
</tr>
</tbody>
</table>

Values of various functions vs. values of argument n.

Even a computer trillions of times faster won't help with such functions.

Allowable Problem Size as a Function of Available Time

<table>
<thead>
<tr>
<th>Time Multiple</th>
<th>10</th>
<th>100</th>
<th>1000</th>
<th>10000</th>
<th>100000</th>
<th>1000000</th>
</tr>
</thead>
<tbody>
<tr>
<td>log n</td>
<td>1024</td>
<td>10^10</td>
<td>10^20</td>
<td>10^30</td>
<td>10^40</td>
<td>10^50</td>
</tr>
<tr>
<td>log^2 n</td>
<td>8</td>
<td>10^2</td>
<td>10^6</td>
<td>10^10</td>
<td>10^15</td>
<td>10^20</td>
</tr>
<tr>
<td>sqrt n</td>
<td>4.5</td>
<td>22</td>
<td>140</td>
<td>1000</td>
<td>7.710^5</td>
<td>6.210^6</td>
</tr>
<tr>
<td>n^2</td>
<td>3</td>
<td>10</td>
<td>100</td>
<td>1000</td>
<td>320</td>
<td>10000</td>
</tr>
<tr>
<td>n^3</td>
<td>2</td>
<td>4</td>
<td>10</td>
<td>21</td>
<td>46</td>
<td>100</td>
</tr>
<tr>
<td>2n</td>
<td>3</td>
<td>6</td>
<td>9</td>
<td>13</td>
<td>16</td>
<td>21</td>
</tr>
<tr>
<td>3n</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td>21</td>
<td>30</td>
<td>42</td>
</tr>
</tbody>
</table>

Increase in size of problem that can be run based on increase in allowed time, assuming algorithm runs problem size 1 in time 1.

Rules for "O"

Additive Rule

- f+g \( O(\max(f, g)) \)
- Here \( f+g \) means \( f(n) + g(n) \)
- \( \max(f, g) \) means \( \max(f(n), g(n)) \)
- Corollary: If \( g \subseteq O(f) \), then \( f+g \subseteq O(f) \)
- Example: For polynomials, the highest order term dominates, e.g. \( n^5 - 1000000n^3 + 100000n^2 \subseteq O(n^5) \)

Transitivity Rule

- If \( f \subseteq O(g) \) and \( g \subseteq O(h) \), then \( f \subseteq O(h) \).

Multiplicative Rule

- If \( f \subseteq O(g) \), then \( h*f \subseteq O(h*g) \)
- where \( h*f \) means \( h(n)*f(n) \)
- For example, \( n*log(n) \subseteq O(n^2) \)
- since \( \log(n) \subseteq O(n) \)

Derivative Rule

- If \( f' \subseteq O(g) \),
- where \( ' \) denotes the derivative,
- then \( f \subseteq O(g) \).
- Example: \( \log(n) \subseteq O(n) \)
- This follows from the derivative rule because \( 1/n \subseteq O(1) \).
**Limit Rule**

- If \( \lim_{n \to \cdot} f(n)/g(n) = k \)
  - then
    - If \( k > 0 \), \( f \in O(g) \), and \( g \in O(f) \).
    - If \( k = 0 \), \( f \notin O(g) \), but not conversely.

**Tight Bounds**

- A bound \( f \in O(g) \) is **tight** if \( g \in O(f) \) also.

**Example: \( \log(n) \in O(n^{1/2}) \).**

- This holds provided \( 1/n \in O(n^{-1/2}) \), according to the derivative rule.
- Apply the limit rule:
  \[
  \lim ((1/n) / n^{-1/2}) = \lim (1/ n^{1/2}) = 0
  \]
- This bound is not tight, since the limit is 0.
- Therefore \( n^{1/2} \notin O(1/n) \).

**Black-box vs. White-box Complexity**

- **Black-box**: We have a copy of the program, with no code. We can run it for different sizes of input.
- **White-box** (aka "Clear box"): We have the code. We can analyze it.
- These terms are from the area of software testing.

**Doubling Input Size**

<table>
<thead>
<tr>
<th>Complexity</th>
<th>Doubling the input causes execution time to</th>
</tr>
</thead>
<tbody>
<tr>
<td>( O(1) )</td>
<td>stay the same</td>
</tr>
<tr>
<td>( O(\log n) )</td>
<td>increase by an additive constant</td>
</tr>
<tr>
<td>( O(n^{1/2}) )</td>
<td>increase by a factor of ( \sqrt{n} )</td>
</tr>
<tr>
<td>( O(n) )</td>
<td>double</td>
</tr>
<tr>
<td>( O(n \log n) )</td>
<td>double, plus increase by a constant factor times ( n )</td>
</tr>
<tr>
<td>( O(n^2) )</td>
<td>increase by a factor of 4</td>
</tr>
<tr>
<td>( O(n^k) )</td>
<td>increase by a factor of ( 2^k )</td>
</tr>
<tr>
<td>( O(n^{2k}) )</td>
<td>square</td>
</tr>
</tbody>
</table>
### Black-Box Complexity

- Run on sizes 32, 64, 128, 512, ...
- For each n, get time T(n).
- How can we estimate the order of run-time (e.g. O(n^2), O(n^3), etc.)?

### Use Empirical Analysis to Predict

- Suppose we are trying to bolster a hypothesis T(n) \(\in\) O(f(n))
- From the definition, we know this means there is a constant c such that for all n, T(n) \(\leq\) c f(n)
- If our hypothesis is correct, we therefore expect for all n, T(n) / f(n) \(\leq\) c
- We can simply examine this ratio.

### Examples: Sorting Programs

- See turing:/cs/cs60/examples/sorting
- Use unix command:
  ```
  run <type>
  ```
  which will try random dataset sizes in the ranges 64, 128, 256, ..., 65536
- type can be one of:
  - quicksort, heapsort, minsort, radixsort, bucketsort

### Black-Box Complexity

- If we see T(n) / f(n) \(\leq\) c
- then the hypothesis T(n) \(\in\) O(f(n)) is supported.
- If T(n) / f(n) is approximately a constant as n increases, then the bound appears to be tight.
- If T(n) / f(n) decreases as n increase, the bound is loose.
- If T(n) / f(n) increases, we don’t have a bound.

### Sorting Programs:
The “fruit flies” of algorithm analysis
### Result Tabulation from Turing runs

(times in ms.)

<table>
<thead>
<tr>
<th>size</th>
<th>heapsort</th>
<th>quicksort</th>
<th>bucketsort</th>
<th>radixsort</th>
<th>msort</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>1</td>
<td>1</td>
<td>51</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>128</td>
<td>3</td>
<td>2</td>
<td>38</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>256</td>
<td>5</td>
<td>3</td>
<td>45</td>
<td>11</td>
<td>25</td>
</tr>
<tr>
<td>512</td>
<td>9</td>
<td>6</td>
<td>45</td>
<td>20</td>
<td>96</td>
</tr>
<tr>
<td>1024</td>
<td>23</td>
<td>12</td>
<td>54</td>
<td>40</td>
<td>380</td>
</tr>
<tr>
<td>2048</td>
<td>49</td>
<td>30</td>
<td>45</td>
<td>79</td>
<td>1507</td>
</tr>
<tr>
<td>4096</td>
<td>100</td>
<td>53</td>
<td>39</td>
<td>176</td>
<td>6119</td>
</tr>
<tr>
<td>8192</td>
<td>230</td>
<td>111</td>
<td>57</td>
<td>321</td>
<td>24657</td>
</tr>
<tr>
<td>16384</td>
<td>482</td>
<td>246</td>
<td>104</td>
<td>537</td>
<td>98681</td>
</tr>
<tr>
<td>32768</td>
<td>1058</td>
<td>535</td>
<td>104</td>
<td>1290</td>
<td>397230</td>
</tr>
<tr>
<td>65536</td>
<td>2333</td>
<td>1183</td>
<td>165</td>
<td>2758</td>
<td>1590497</td>
</tr>
</tbody>
</table>

Example: msort

We predict that \( T(n) \) is about 0.00037 \( n^2 \) milliseconds for large \( n \).

How long will it take to sort 100,000 items? One million items?

<table>
<thead>
<tr>
<th>( n )</th>
<th>( T(n) ) ms</th>
<th>( T(n) ) in days</th>
</tr>
</thead>
<tbody>
<tr>
<td>10000</td>
<td>37000</td>
<td>0.1027778</td>
</tr>
<tr>
<td>100000</td>
<td>37000000</td>
<td>10.277778</td>
</tr>
<tr>
<td>1000000</td>
<td>3700000000</td>
<td>1027.7778</td>
</tr>
</tbody>
</table>

White-Box Complexity

- Here we examine the code (loop structure, etc.)

```
for( int j = 0; j < n; j++ )
{
    O(1)
}
```

Means some constant-time computation.

Assume it does not modify loop index.

Complexity: \( O(n) \)

Example: msort

Suppose we hypothesize \( T(n) \in O(n^2) \) and compute \( T(n)/n^2 \).

\[
\begin{array}{cccccc}
380 & 1507 & 819 & 24657 & 98681 & 397230 & 1590497 \\
0.0003624 & 0.0003593 & 0.00036472 & 0.00036742 & 0.00036762 & 0.00036995 & 0.00037032 \\
\end{array}
\]

The hypothesis is supported. Moreover, we can predict that \( T(n) \) is about 0.00037 \( n^2 \) milliseconds for large \( n \).
Analysis of Typical Loops

for( int j = 0; j < n; j++ )
for( int k = 0; k < n; k++ )
{
  O(1)
}

Analysis of Typical Loops

for( int j = 1; j <= n; j = 2*j )
{
  O(1)  // j doubles each time.
  The loop stops by iteration k, where \( 2^k \geq n \), i.e. by ceiling(log(n)) iterations.
}

Complexity: \( O(\log(n)) \)  This bound is tight.

Analysis of Typical Loops

for( int j = 1; j < n; j++ )
for( int k = j; k > 0; k = k/2 )
{
  O(1)
}

Analysis of Typical Loops

for( int j = 1; j < n; j++ )
for( int k = j; k > 0; k = k/2 )
{
  O(1)
}

\[ \log(1) \cdot \log(2) \cdot \ldots \cdot \log(n-2) \cdot \log(n-1) = O(n \log(n)) \]
\[ n/2 \text{ terms, each } \geq \log(n/2) \Rightarrow \text{bound is tight} \]
Analysis of Typical Loops

for( int j = n; j > 0; j = j/2 )
    for( int k = 1; k <= j; k++ )
        { O(1) }

Complexity: O(n log(n)), but this is not tight.

Analysis of Actual Programs

minsort:
    sorting by selecting minima

class minsort
    { private double array[]; // The array being sorted

    // Calling minsort constructor on array of doubles sorts
    // the array elements 0..(N-1).
    minsort(double array[], int N)
        { this.array = array;
            for( int i = 0; i < N; i++ )
                { swap(i, findMin(i, N));
                }
    }

    // swap(i, j) interchanges the values
    // in array[i] and array[j]
    void swap(int i, int j)
        { double temp = array[i];
            array[i] = array[j];
            array[j] = temp;
        }

minsort
minsort

// findMin(M, N) finds the index of the minimum among
// array[M], array[M+1], ...., array[N-1].
int findMin(int minSoFar, int N) {
    // by default, the element at minSoFar is the minimum
    for(int j = minSoFar + 1; j < N; j++)
        if(array[j] < array[minSoFar])
            minSoFar = j; // a smaller value is found
    return minSoFar;
}

Analysis: \( \text{N} \) steps

minsort

// swap(i, j) interchanges the values
// in array[i] and array[j];
void swap(int i, int j) {
    double temp = array[i];
    array[i] = array[j];
    array[j] = temp;
}

Analysis: 3 steps

minsort

class minsort {
    private double array[]; // The array being sorted
    // Calling minsort constructor on array of doubles sorts
    // the array elements 0..(N-1).
    minsort(double array[], int N) {
        this.array = array;
        for(int i = 0; i < N; i++)
            swap(i, findMin(i, N));
    }
}

Analysis: \( \text{N} \) \(-\text{i}-2 \) times

\( 3 \) steps

\( \text{N} \) \(-\text{i}-2 \) steps

i ranges from 0 to \( \text{N} \)-1
\( (\text{N}+2) \) \(+ (\text{N}-1) + \ldots + 3 \) steps

O(\( \text{N}^2 \)) steps

Similar analysis, with the same
result, is obtained for:

- bubble sort
- simple insertion sort

(For a live demo, see:
http://www.cs.oswego.edu/~mohammad/classes/osc241/samples/Sort2-E.html)

insertion sort in rex

isort([]) => [];
isort([A | X]) => insert(A, isort(X));
// insert inserts the first item into a list
// in the proper place, assuming the list
// is in order
insert(A, []) => [A];
insert(A, [B | X]) =>
recurrence for isort

\[
isort(\{\}) \Rightarrow \{\};
\]

\[
isort([A \mid X]) \Rightarrow \text{insert}(A, isort(X));
\]

argument below is the length of the list

\[
T_{\text{isort}}(0) \Rightarrow 0;
\]

\[
T_{\text{isort}}(N) \Rightarrow T_{\text{isort}}(N-1) + T_{\text{isort}}(N-1);
\]

returning to
recurrence for isort

\[
T_{\text{isort}}(0) \Rightarrow 0;
\]

\[
T_{\text{isort}}(N) \Rightarrow T_{\text{isort}}(N-1) + T_{\text{isort}}(N-1);
\]

\[
\leq T_{\text{isort}}(N-1) + cN
\]

Solving

\[
T_{\text{isort}}(N) = c(1+2+ \ldots + N)
\]

\[O(n^2)\] steps

recurrence for insert

\[
isort(A, \{\}) \Rightarrow \{A\};
\]

\[
isort(A, [B \mid X]) \Rightarrow \begin{cases} A < B \? [A, B \mid X] \mid \text{insert}(A, X) \end{cases};
\]

\[
T_{\text{isort}}(0) = 1
\]

\[
T_{\text{isort}}(N) \leq 1 + T_{\text{isort}}(N-1);
\]

Solving:

\[
T_{\text{isort}}(N) \in O(N).
\]

Is \(O(n^2)\) the best we can do for sorting?

Algorithm Speedup Techniques

- Divide and Conquer
- The "digital" principle: Use data values to do selection or direct accessing
- Try trees instead of linear arrangement
- "Dynamic" programming (later)

Technique #1: Divide-and-Conquer

- Rearrange the array to be sorted:
  - Low elements on the bottom
  - High elements on the top
  - Use some element as the "pivot" value
- Sort the low and high portions recursively
- Called "quicksort"
- Invented by C.A.R. (Tony) Hoare
Tony Hoare

Professor C.A.R. Hoare, FRS

James Martin Professor of Computing
Oxford University Computing Laboratory,
Wolfson Building, Parks Road, Oxford OX1 3QD, England.

Quicksort Illustrated

Using the rule that we split using the element at the middle of the sub-array.
Splitting sends elements < to the left and > to the right (= can go to either.)

Quicksort Analysis

- Overall steps = cn x number of levels.
- O(log(n)) levels in optimistic case.
- O(n) levels in pessimistic case.
- Overall O(n²).
- The average case can be shown to be O(n log(n)) based on a probabilistic argument, assuming the data are initially randomly distributed.

Another Divide-and-Conquer Sort
Mergesort

- Sort by successively merging longer and longer sorted sequences.
- Useful with linked lists, or large files.
- More difficult to program for arrays.
- Different versions exists:
  - Top-Down: Split unordered sequence, vs.
  - Bottom-Up: Start with sequences of length 1 and create increasingly longer ones.

Bottom-Up Mergesort

- Start with sequences of length 1; these are sorted by default.
- Merge pairs of sorted sequences to form single sorted sequences.
- until there is only one sequence left.
Bottom-Up Mergesort Example

\[
\{7, 4, 5, 2, 1, 5, 0, 6\} \\
\{4, 7, 2, 3, 1, 5, 0, 6\} \\
\{2, 3, 4, 7, 0, 1, 5, 6\} \\
\{0, 1, 2, 3, 4, 5, 6, 7\}
\]

Bottom-Up Mergesort Analysis

- Merging two sequences of length \(n/2\) each can be done in \(O(n)\) if linked lists are used.
- In merge sort of \(n\) elements, we merge:
  - \(n/2\) pairs of sequences of length 1
  - \(n/4\) pairs of sequences of length 2
  - \(n/8\) pairs of sequences of length 4
  - ... 
  - 1 pair of sequences of length \(n/2\)

Bottom-Up Mergesort Analysis

- At each "level" \(O(n)\) steps are used.
- There are \(\log(n)\) levels.
- Therefore mergesort is \(O(n \log(n))\) worst case.

Bottom-Up Mergesort in rex (1)

```plaintext
// first the initial list is transformed to a list of 1-element lists, then those lists are merged repeatedly
merge_sort(List) = merge_repeatedly(map((X) => [X], List));

// merge_repeatedly merges pairs in a list of lists until there is only one list left.
merge_repeatedly([A]) => A; // only one list
merge_repeatedly(Lists) => merge_repeatedly(merge_pairs(Lists));
```

Bottom-Up Mergesort Analysis

- Let \(T(j)\) = steps at levels \(\leq j\)
  - Then
    - \(T(1) = c \cdot n\), \(c\) some constant
    - \(T(j+1) = T(j) + c \cdot n\)
- \(T(\log(n))\) is the time to sort \(n\) elements
  - \(= c \cdot n \log(n)\)

Bottom-Up Mergesort in rex (2)

```plaintext
// merge_pairs merges pairs of lists in a list until none is left.
merge_pairs([]) => []; // no more lists
merge_pairs([A]) => [A]; // only one list
merge_pairs([A, B | L]) => [merge(A, B) | merge_pairs(L)];

// merge creates a single ordered list from two ordered lists
merge([], []) => []; // no lists
merge([A], []) => [A]; // only one list
merge([], [B]) => [B]; // only one list
merge([A, B | L], [M]) => [merge(A, B) | merge_pairs(L, M)];
```

A => B ? [A | merge(L, [B | M])] : [B | merge([A | L], M)];
Top-Down Mergesort Example

\{7, 4, 3, 2, 1, 5, 0, 6\}

split into two

\{7, 4, 3, 2\} \{1, 5, 0, 6\}

recursively mergesort each half

\{2, 3, 4, 7\} \{0, 1, 5, 6\}

merge the sorted halves

\{0, 1, 2, 3, 4, 5, 6, 7\}

Alternate ways to split

\{7, 4, 3, 2, 1, 5, 0, 6\}

split into two

\{7, 5, 1\} \{4, 2, 5, 6\}

Top-Down Mergesort Analysis

- $T_{merge}(n) = cn$ Time to merge two lists of length $n/2$
- $T_{split}(n) = dn$ Time to split a list of length $n$
- $T_{sort}(1) = 1$;
- $T_{sort}(n) = T_{split}(n) + 2 T_{sort}(n/2) + T_{merge}(n)$
  
  \[= 2 T_{sort}(n/2) + en\]

  where $e = c + d$

Top-Down Mergesort Analysis

- $T_{sort}(1) = 1$;
- $T_{sort}(n) = 2 T_{sort}(n/2) + en$
- Substituting
  
  $T(n) = 2 T(n/2) + en$
  
  \[= 2 (T(n/4) + en/2) + en\]
  
  \[= 2 (2(T(n/8) + en/4) + en/2) + en\]
  
  \[= \ldots\]
  
  \[= 2^{\log(n)} + \log(n) \cdot en\]
  
  \[= n + en \log(n)\]

  $O(n \log(n))$

Technique #2: Using data values to do selection

- Under this category, we have
  - bucket sort
  - radix sort
- We make non-general assumptions about the data:
  - The size of keys to be sorted is limited to integers with a fixed upper bound.

Bucket Sort

- Related to hashing
  - Both use indexing, which is $O(1)$, to find “bucket”
- Suppose the set of keys to be sorted is known to be limited to a relatively small integer range, say $\{0, 1, \ldots, R-1\}$.
- Create an array of size $R$ of lists, each entry corresponding to a key value.
- Go through the data once, putting each element in the corresponding list.
- Concatenate the resulting lists.
Analysis of Bucket Sort

- Assume that the number of data elements is comparable to, or larger than, the number of buckets.
- Go through the data once, putting each element in the corresponding list. This is O(n).
- Concatenate the resulting lists. This is also O(n).
- Therefore we have O(n) overall.
- Remember that bucket sorting makes special assumptions about the data.

Radix Sort

- Like bucket sort, but with smaller arrays.
- Trades array size for multiple passes.
- Represent the range as radix b integers.
- Make $P = \log_b(R)$ passes.
- Sort on the least significant digit first, progressing toward the most.
- Re-collect the data after each pass and redistributed.

Analysis of Radix Sort

- Assuming a bounded range, the number of passes $P$ is a fixed constant.
- Each pass uses O(n).
- Therefore we have O(n) overall.

Demo of Radix Sort

Technique #3: Use a Tree instead of Linear List

- For the same amount of data, a sufficiently balanced tree uses only $O(\log n)$ to traverse a chain rather than O(n) (as in naive bubble, selection, or insertion sort).
- Heapsort is a form of sorting that uses trees.

Heapsort

- Loosely based on the "Peter Principle" (Dr. Laurence J. Peter)
- books by Laurence Peter
Laurence J. Peter

The Peter Principle

- "In a hierarchy, everyone rises to his/her level of incompetence".
- [A person having reached this level is said to have the "final placement syndrome"]

A Hierarchy

The Peter Principle
applied to Sorting

- A heap is a binary tree structure in which, for each sub-tree, the root is > all elements in that sub-tree.
- It suffices for each root to be > its children.
- [This is one definition of "heap" used in CS. The other definition is the area in memory from which storage is allocated dynamically, e.g. when new is called.]

A Heap

A Heap?

Example of final placement syndrome
**The Two Phases in Heapsort**

- Phase I: Form the data into a heap
- Phase II: Transform the heap into a linear sequence

**Phase I: Forming a Heap**

- This is the actual Peter Principle.
- Start with the data distributed randomly in the tree.
- Form "sub-heaps" beginning at the leaves and working toward the root.
- Successively combine two sub-heaps with a common root into a single heap.

**Combination as a Tournament**

![Combination as a Tournament Diagram]

**Analysis of forming a heap from two sub-heaps**

- 3-way rounds are played from the parent downward to the leaves
- Only one path toward the leaves is followed, since other sub-trees along the path don't change
- A 3-way round uses O(1) steps (2! comparisons, possible exchange)
- The time to form a heap at level k from the leaves is therefore O(k).
- In the worst case, this is O(log(n)).

**Iterating sub-heap formation from leaves to root**

- The leaves are already heaps by themselves.
- We need to play the tournament (O(k) rounds at level k) n/2 times, corresponding to the non-leaves.
- Coarsely, Phase I is O(n log(n)) overall.
- Is this bound tight?
Illustrating Phase I

After Level 1 Heap Formation

Level 2 Heap Formation

After Level 2 Heap Formation

Level 3 Heap Formation

Heaps

Red lines indicate where exchanges will occur

Green lines indicate where exchanges occurred
After Level 3 Heap Formation

Heapsort Phase II

Now that we have a heap, what do we do?

Heapsort Phase II

One antidote for final-placement syndrome is that people retire.

In our heap, we "retire" the maximum, which is guaranteed to be at the root.

This leaves a hole that needs to be filled.

Retiring the maximum

Filing the hole

The way this happens is different from in a corporation.

We pick the rightmost leaf, and tentatively place it in the hole at the root.

Then we play the tournament from the root so that the new value reaches its proper level.

Filling the hole

18 (Hey, the view from the top is great!)
Filling the hole

25

The tournament path

20

9

17

21

13

6

10

5

4

The tournament path

25

18

Result of the tournament

22

20

9

18

17

21

10

6

19

4

5

1

13

1

13

9

10

17

6

19

22

20

A happy heap once again.

The cost of keeping the heap happy

- The tournament path could be as long as the path from the root to a leaf.
- Along the path, a number of $O(1)$ rounds were played.
- The cost for one post-retirement adjustment is therefore $O(\log(n))$.
- Retiring all $n$ elements in sequence gives us the sorted order.
- Overall then, we have $O(n \log(n))$ for Phase II.

The Rest of Phase II illustrated

20

9

18

17

21

4

5

1

13

10

6

19

22

25

20

Phase II, step 2, continued

20

9

18

17

21

4

5

1

13

10

6

19

22

25

Phase II, step 3

20

9

18

17

21

4

5

1

13

10

6

19

22

25
Where to put the retirees?

- We can put the retirees back in the tree, as long as we know not to play any more tournaments with them.
- There is an easy way to keep track of this, as we shall see.
Phase II with Retiree Placement

Phase II with Retiree Placement

Phase II with Retiree Placement

Phase II with Retiree Placement

Phase II with Retiree Placement

Phase II with Retiree Placement
Phase II with Retiree Placement

Phase II with Retiree Placement

Phase II with Retiree Placement

Phase II with Retiree Placement

Phase II with Retiree Placement

Phase II with Retiree Placement
Phase II with Retiree Placement

Node Numbering

- Note that the sorted sequence is readable from the nodes in breadth-first order.
- Further, we can maintain the entire tree as an array (with no explicit links), as shown next.

Array Node Numbering
Permits Read-out of Sorted Data

Use as a Tree:
Finding your parents/children

Heapsort Code (1)

```c
heapsort(float array[], int N) {
    this.array = array;
    int Last = N-1;
    // phase 1: form heap
    for( int Top = Last/2; Top >= 0; Top-- )
        adjust(Top, Last);
    // phase 2: use heap to sort
    while( Last > 0 )
    {
        swap(0, Last);
        adjust(0, --Last);
    }
}
```
Heapsort Code (2)

```c
void adjust(int Top, int Last) {
    float TopVal = array[Top]; // Set aside top of heap
    int Parent, Child;
    for (Parent = Top; Parent < Last; Parent = Child) {
        Child = 2*Parent+1; // Child means left child
        if (Child > Last) break; // Left child non-existent
        if (Child+1 <= Last && array[Child] < array[Child+1]) { // Right child exists and right child is larger
            Child++; // Child is the larger child
        }
        if (TopVal >= array[Child]) break; // Location for TopVal found
        array[Parent] = array[Child]; // Move larger child up in tree
    }
    array[Parent] = TopVal; // Install TopVal in place
}
```

Heapsort Summary

- $O(n \log(n))$ steps
- In-place array implementation

Priority Queue

- A priority queue is a data repository that has two methods:
  - insert
  - removeMax (or removeMin, depending on the version)
- A heap is a natural implementation of a priority queue:
  - Both insertions and deletions are $O(\log n)$

Priority Queue Applications

- A priority queue (with min removal) is often found in simulation applications.
  - Events are time-stamped and put in a priority queue, which orders them by smallest time first.
- On a typical simulation cycle:
  - The event with the next timestamp is removed.
  - The event may cause the insertion of new events with later timestamps.

Lower-Bound for Sorting

- Sorting based on pairwise comparisons (which excludes radix and bucket sorts) requires a minimum of $c \cdot n \log_2(n)$ steps.
- Sorting methods that achieve this lower bound are called "optimal".
- Heapsort and mergesort are the two optimal sorts we've studied.

Derivation of the Sorting Lower-Bound

- A sorting algorithm effectively establishes which of $n!$ permutations the data are originally in and through a series of comparisons and exchanges permutes the data to a fixed order.
- This can be viewed as a tree with $n!$ nodes, constructed with binary internal nodes for comparisons.
Derivation of the Sorting Lower-Bound

- The worst-case run time of the sorting algorithm is the height of a tree having $n!$ nodes.

- The minimum height of such a tree is $\log_2(n!)$. 

- By Stirling’s formula, $n! \sim (2\pi n)^{1/2} (n/e)^n$, so $\log_2(n!)$ is $O(n \log(n))$.

Sorting Summary

- Minsort, insertion sort, bubble sort: easy to code, slow, avoid for large data sets
- Quicksort: fast on average, slow worst case
- Bucket sort / Radix sort: fastest asymptotic performance, special assumptions about data
- Heap sort: optimal performance, sorts arrays in-place
- Merge sort: optimal performance, sorts lists, more difficult for arrays