Administrivia

HW 3 is due today
- you get 2 more late days for the quarter

HW 4 is out later today

Midterm Prep
- Attend section tomorrow, solutions have already been posted
- Review TONIGHT in PAA 102 4-5:50, panopto will be available
- Erik filmed a review session, check it out on the website
A message about grades...
Midterm Logistics

50 minutes

8.5 x 11 in note page, front and back

Math identities sheet provided (see posted on website)

We will be scanning your exams to grade...
- Do not write on the back of pages
- Try not to cram answers into margins or corners
Asymptotic Analysis
Asymptotic Analysis

**asymptotic analysis** – the process of mathematically representing runtime of a algorithm in relation to the number of inputs and how that relationship changes as the number of inputs grow

**Two step process**

1. **Model** – the process of mathematically representing how many operations a piece of code will run in relation to the number of inputs \( n \)

2. **Analyze** – compare runtime/input relationship across multiple algorithms
   1. Graph the model of your code where \( x = \) number of inputs and \( y = \) runtime
   2. For which inputs will one perform better than the other?
Code Modeling

**code modeling** – the process of mathematically representing how many operations a piece of code will run in relation to the number of inputs $n$

Examples:
- Sequential search $f(n) = n$
- Binary search $f(n) = \log_2 n$

What counts as an “operation”?

**Basic operations**
- Adding ints or doubles
- Variable assignment
- Variable update
- Return statement
- Accessing array index or object field

**Consecutive statements**
- Sum time of each statement

**Function calls**
- Count runtime of function body

**Conditionals**
- Time of test + worst case scenario branch

**Loops**
- Number of iterations of loop body x runtime of loop body

Assume all operations run in equivalent time
Code Modeling

```java
public int mystery(int n) {
    int result = 0;
    for (int i = 0; i < n/2; i++) {
        result++;
    }
    for (int i = 0; i < n/2; i+=2) {
        result++;
    }
    result * 10;
    return result;
}
```

\[ f(n) = 3 + \frac{3}{4}n = C_1 + \frac{3}{4}n \]
public String mystery (int n) {
    ChainedHashDictionary<Integer, Character> alphabet =
        new ChainedHashDictionary<Integer, Character>();
    for (int i = 0; i < 26; i++) {
        char c = 'a' + (char)i;
        alphabet.put(i, c);
    }
    DoubleLinkedList<Character> result = new DoubleLinkedList<Character>();
    for (int i = 0; i < n; i += 2) {
        char c = alphabet.get(i);+26
        result.add(c); +1
    }
    String final = "";
    for (int i = 0; i < result.size(); i++) {
        final += result.remove(); +1
    }
    return final;
}
Imagine you have three possible algorithms to choose between. Each has already been reduced to its mathematical model:

- \( f(n) = n \)
- \( g(n) = 4n \)
- \( h(n) = n^2 \)

The growth rate for \( f(n) \) and \( g(n) \) looks very different for small numbers of input. 

...but since both are linear eventually look similar at large input sizes. 

whereas \( h(n) \) has a distinctly different growth rate. 

But for very small input values \( h(n) \) actually has a slower growth rate than either \( f(n) \) or \( g(n) \). 

Definitions

\(O(f(n))\) is the “family” or “set” of all functions that are dominated by \(f(n)\)
- \(f(n) \in O(g(n))\) when \(f(n) \leq g(n)\)
- The upper bound of an algorithm’s function

\(\Omega(f(n))\) is the family of all functions that dominate \(f(n)\)
- \(f(n) \in \Omega(g(n))\) when \(f(n) \geq g(n)\)
- The lower bound of an algorithm’s function

\(\Theta(f(n))\) is the family of functions that are equivalent to \(f(n)\)
- We say \(f(n) \in \Theta(g(n))\) when both
- \(f(n) \in O(g(n))\) and \(f(n) \in \Omega(g(n))\) are true
- A direct fit of an algorithm’s function

\[ f(n) \in O(g(n)) \text{ if there exist positive constants } c, n_0 \text{ such that for all } n \geq n_0, \]
\[ f(n) \leq c \cdot g(n) \]

\[ f(n) \in \Omega(g(n)) \text{ if there exist positive constants } c, n_0 \text{ such that for all } n \geq n_0, \]
\[ f(n) \geq c \cdot g(n) \]

\[ f(n) \in \Theta(g(n)) \text{ if } f(n) \in O(g(n)) \text{ and } f(n) \in \Omega(g(n)) \text{ is true.} \]
Proving Domination

\[ f(n) = 5(n + 2) \]

\[ g(n) = 2n^2 \]

Find a \( c \) and \( n_0 \) that show that \( f(n) \in O(g(n)) \).

\[ f(n) = 5(n + 2) = 5n + 10 \]

\[ 5n \leq c \cdot 2n^2 \text{ for } c = 3 \text{ when } n \geq 1 \]

\[ 10 \leq c \cdot 2n^2 \text{ for } c = 5 \text{ when } n \geq 1 \]

\[ 5n + 10 \leq 3(2n^2) + 5(2n^2) \text{ when } n \geq 1 \]

\[ 5n + 10 \leq 8(2n^2) \text{ when } n \geq 1 \]

\[ f(n) \leq c \cdot g(n) \text{ when } c = 8 \text{ and } n_0 = 1 \]
**O, Ω, Θ Examples**

For the following functions give the simplest tight $O$ bound

- $a(n) = 10 \log n + 5 \quad O(\log n)$
- $b(n) = 3^n - 4n \quad O(3^n)$
- $c(n) = \frac{n}{2} \quad O(n)$

For the above functions indicate whether the following are true or false

- $a(n) \in O(b(n)) \quad$ TRUE
- $a(n) \in O(c(n)) \quad$ TRUE
- $a(n) \in \Omega(b(n)) \quad$ FALSE
- $a(n) \in \Omega(c(n)) \quad$ FALSE
- $a(n) \in \Theta(b(n)) \quad$ FALSE
- $a(n) \in \Theta(c(n)) \quad$ TRUE
- $b(n) \in O(a(n)) \quad$ FALSE
- $b(n) \in O(c(n)) \quad$ FALSE
- $b(n) \in \Omega(a(n)) \quad$ TRUE
- $b(n) \in \Omega(c(n)) \quad$ TRUE
- $b(n) \in \Theta(a(n)) \quad$ FALSE
- $b(n) \in \Theta(c(n)) \quad$ TRUE
- $c(n) \in O(b(n)) \quad$ TRUE
- $c(n) \in O(a(n)) \quad$ FALSE
- $c(n) \in \Omega(b(n)) \quad$ FALSE
- $c(n) \in \Omega(a(n)) \quad$ TRUE
- $c(n) \in \Theta(b(n)) \quad$ FALSE
- $c(n) \in \Theta(a(n)) \quad$ FALSE
- $c(n) \in \Theta(c(n)) \quad$ TRUE
**Review: Complexity Classes**

**complexity class** – a category of algorithm efficiency based on the algorithm’s relationship to the input size $N$

<table>
<thead>
<tr>
<th>Class</th>
<th>Big O</th>
<th>If you double $N$...</th>
<th>Example algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>$O(1)$</td>
<td>unchanged</td>
<td>Add to front of linked list</td>
</tr>
<tr>
<td>logarithmic</td>
<td>$O(\log_2 n)$</td>
<td>Increases slightly</td>
<td>Binary search</td>
</tr>
<tr>
<td>linear</td>
<td>$O(n)$</td>
<td>doubles</td>
<td>Sequential search</td>
</tr>
<tr>
<td>log-linear</td>
<td>$O(n \log_2 n)$</td>
<td>Slightly more than doubles</td>
<td>Merge sort</td>
</tr>
<tr>
<td>quadratic</td>
<td>$O(n^2)$</td>
<td>quadruples</td>
<td>Nested loops traversing a 2D array</td>
</tr>
<tr>
<td>cubic</td>
<td>$O(n^3)$</td>
<td>Multiplies by 8</td>
<td>Triple nested loop</td>
</tr>
<tr>
<td>polynomial</td>
<td>$O(n^c)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>exponential</td>
<td>$O(c^n)$</td>
<td>Multiplies drastically</td>
<td></td>
</tr>
</tbody>
</table>

http://bigocheatsheet.com/
Modeling Complex Loops

```java
for (int i = 0; i < n; i++) {
    for (int j = 0; j < i; j++) {
        System.out.println("Hello!");
    }
}
```

Summation

\[ 1 + 2 + 3 + 4 + \ldots + n = \sum_{i=1}^{n} i \]

Definition: Summation

\[ \sum_{i=a}^{b} f(i) = f(a) + f(a+1) + f(a+2) + \ldots + f(b-2) + f(b-1) + f(b) \]

\[ T(n) = \sum_{i=0}^{n-1} \sum_{j=0}^{i-1} c \]
Function Modeling: Recursion

public int factorial(int n) {
    if (n == 0 || n == 1) {
        return 1;
    } else {
        return n * factorial(n - 1);
    }
}

\[ T(n) = \begin{cases} 
    C_1 & \text{when } n = 0 \text{ or } 1 \\
    C_2 + T(n-1) & \text{otherwise} 
\end{cases} \]

Definition: Recurrence

Mathematical equivalent of an if/else statement

\[ f(n) = \begin{cases} 
    \text{runtime of base case when conditional} & \text{otherwise} 
\end{cases} \]
Tree Method Formulas

How much work is done by recursive levels (branch nodes)?
1. How many recursive calls are on the i-th level of the tree?
   - i = 0 is overall root level
   \[ \text{numberOfNodesPerLevel}(i) = 2^i \]
2. At each level i, how many inputs does a single node process?
   \[ \text{inputsPerRecursiveCall}(i) = \frac{n}{2^i} \]
3. How many recursive levels are there?
   - Based on the pattern of how we get down to base case
   \[ \text{branchCount} = \log_2 n - 1 \]
\[ \text{Recursive work} = \sum_{i=0}^{\log_2 n - 1} \text{branchNum}(i) \text{branchWork}(i) \]
\[ T(n > 1) = \sum_{i=0}^{\log_2 n - 1} 2^i \left( \frac{n}{2^i} \right) \]

How much work is done by the base case level (leaf nodes)?
1. How much work is done by a single leaf node?
   \[ \text{leafWork} = 1 \]
2. How many leaf nodes are there?
   \[ \text{leafCount} = 2^{\log_2 n} = n \]
\[ \text{NonRecursive work} = \text{leafWork} \times \text{leafCount} = \text{leafWork} \times \text{branchNum}^{\text{numLevels}} \]
\[ T(n \leq 1) = 1 \left( 2^{\log_2 n} \right) = n \]
\[ \text{total work} = \text{recursive work} + \text{nonrecursive work} = T(n) = \sum_{i=0}^{\log_2 n - 1} 2^i \left( \frac{n}{2^i} \right) + n = n \log_2 n + n \]
Tree Method Example

\[ T(n) = \begin{cases} 
3 & \text{when } n = 1 \\
3T\left(\frac{n}{3}\right) + n & \text{otherwise}
\end{cases} \]

Size of input at level \( i \)?: \( \frac{n}{3^i} \)

Number of nodes at level \( i \)?: \( 3^i \)

How many levels of the tree?: \( \log_3(n) \)

How many nodes are on the bottom level?: \( 3^{\log_3(n)} = n \)

How much work done in base case?: \( 3n \)

Total recursive work:

\[
\sum_{i=0}^{\log_3 n - 1} \frac{n}{3^i} 3^i = n \log_3(n)
\]

\[ T(n) = n \log_3(n) + 3n \]
Reflecting on Master Theorem

Given a recurrence of the form:

\[
T(n) = \begin{cases} 
  d & \text{when } n = 1 \\
  aT\left(\frac{n}{b}\right) + n^c & \text{otherwise}
\end{cases}
\]

- If \( \log_b a < c \) then \( T(n) \in \Theta(n^c) \)
- If \( \log_b a = c \) then \( T(n) \in \Theta(n^c \log_2 n) \)
- If \( \log_b a > c \) then \( T(n) \in \Theta(n^{\log_b a}) \)

\[ \text{height} \approx \log_b a \]
\[ \text{branchWork} \approx n^c \log_b a \]
\[ \text{leafWork} \approx d(n^{\log_b a}) \]

The \( \log_b a < c \) case
- Recursive case conquers work more quickly than it divides work
- Most work happens near “top” of tree
- Non recursive work in recursive case dominates growth, \( n^c \) term

The \( \log_b a = c \) case
- Work is equally distributed across call stack (throughout the “tree”)
- Overall work is approximately work at top level x height

The \( \log_b a > c \) case
- Recursive case divides work faster than it conquers work
- Most work happens near “bottom” of tree
- Leaf work dominates branch work
Master Theorem Example

Given a recurrence of the form:

\[ T(n) = \begin{cases} 
  d & \text{when } n = 1 \\
  aT\left(\frac{n}{b}\right) + n^c & \text{otherwise}
\end{cases} \]

If \( \log_b a < c \) then \( T(n) \in \Theta(n^c) \)

If \( \log_b a = c \) then \( T(n) \in \Theta(n^c \log n) \)

If \( \log_b a > c \) then \( T(n) \in \Theta(n^{\log_b a}) \)

\[ a = 3 \\
 b = 3 \\
 c = 1 \\
 \log_3 3 = 1 \\
 T(n) \text{ is in } \Theta(n \log n) \]
BST & AVL Trees
A **binary search tree** is a binary tree that contains comparable items such that for every node, all children to the left contain smaller data and all children to the right contain larger data.
Meet AVL Trees

**AVL Trees** must satisfy the following properties:

- **binary trees**: all nodes must have between 0 and 2 children
- **binary search tree**: for all nodes, all keys in the left subtree must be smaller and all keys in the right subtree must be larger than the root node
- **balanced**: for all nodes, there can be no more than a difference of 1 in the height of the left subtree from the right. Math.abs(height(left subtree) – height(right subtree)) ≤ 1

AVL stands for Adelson-Velsky and Landis (the inventors of the data structure)
Two AVL Cases

**Line Case**
Solve with 1 rotation

**Kink Case**
Solve with 2 rotations

**Rotate Right**
Parent’s left becomes child’s right
Child’s right becomes its parent

**Rotate Left**
Parent’s right becomes child’s left
Child’s left becomes its parent

**Right Kink Resolution**
Rotate subtree left
Rotate root tree right

**Left Kink Resolution**
Rotate subtree right
Rotate root tree left
Hashing
public V get(int key) {
    int newKey = getKey(key);
    this.ensureIndexNotNull(key);
    return this.data[key].value;
}

public void put(int key, int value) {
    this.array[getKey(key)] = value;
}

public void remove(int key) {
    int newKey = getKey(key);
    this.ensureIndexNotNull(key);
    this.data[key] = null;
}

public int getKey(int value) {
    return value % this.data.length;
}
First Hash Function: % table size

<table>
<thead>
<tr>
<th>indices</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>element</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“poo”</td>
<td>“biz”</td>
</tr>
<tr>
<td>s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“bar”</td>
<td>“bop”</td>
</tr>
</tbody>
</table>

put(0, “foo”); 0 % 10 = 0
put(5, “bar”); 5 % 10 = 5
put(11, “biz”); 11 % 10 = 1
put(18, “bop”); 18 % 10 = 8
put(20, “poo”); 20 % 10 = 0

Collision!
Handling Collisions

Solution 1: Chaining

Each space holds a “bucket” that can store multiple values. Bucket is often implemented with a LinkedList.

<table>
<thead>
<tr>
<th>Operation</th>
<th>Array w/ indices as keys</th>
</tr>
</thead>
<tbody>
<tr>
<td>put(key,value)</td>
<td>best: O(1)</td>
</tr>
<tr>
<td></td>
<td>average: O(1 + λ)</td>
</tr>
<tr>
<td></td>
<td>worst: O(n)</td>
</tr>
<tr>
<td>get(key)</td>
<td>best: O(1)</td>
</tr>
<tr>
<td></td>
<td>average: O(1 + λ)</td>
</tr>
<tr>
<td></td>
<td>worst: O(n)</td>
</tr>
<tr>
<td>remove(key)</td>
<td>best: O(1)</td>
</tr>
<tr>
<td></td>
<td>average: O(1 + λ)</td>
</tr>
<tr>
<td></td>
<td>worst: O(n)</td>
</tr>
</tbody>
</table>

**Average Case:**
Depends on average number of elements per chain

Load Factor λ
If n is the total number of key-value pairs
Let c be the capacity of array
Load Factor λ = \( \frac{n}{c} \)
Handling Collisions

Solution 2: Open Addressing

Resolves collisions by choosing a different location to store a value if natural choice is already full.

Type 1: Linear Probing

If there is a collision, keep checking the next element until we find an open spot.

```
public int hashFunction(String s) {
    int naturalHash = this.getHash(s);
    if (natural hash in use) {
        int i = 1;
        while (index in use) {
            try (naturalHash + i);
            i++;
        }
    }
}
```

Type 2: Quadratic Probing

If we collide instead try the next $i^2$ space

```
public int hashFunction(String s) {
    int naturalHash = this.getHash(s);
    if (natural hash in use) {
        int i = 1;
        while (index in use) {
            try (naturalHash + i * i);
            i++;
        }
    }
}
```
Linear Probing

Insert the following values into the Hash Table using a hashFunction of % table size and linear probing to resolve collisions:
1, 5, 11, 7, 12, 17, 6, 25
Quadratic Probing

Insert the following values into the Hash Table using a hashFunction of % table size and quadratic probing to resolve collisions
89, 18, 49, 58, 79

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>58</td>
<td>79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
<td></td>
<td>49</td>
</tr>
</tbody>
</table>

(49 % 10 + 0 * 0) % 10 = 9
(49 % 10 + 1 * 1) % 10 = 0
(58 % 10 + 0 * 0) % 10 = 8
(58 % 10 + 1 * 1) % 10 = 9
(58 % 10 + 2 * 2) % 10 = 2
(79 % 10 + 0 * 0) % 10 = 9
(79 % 10 + 1 * 1) % 10 = 0
(79 % 10 + 2 * 2) % 10 = 3

Problems:
If λ ≥ ½ we might never find an empty spot
Infinite loop!
Can still get clusters
Handling Collisions

Solution 3: Double Hashing

If the natural hash location is taken, apply a second and separate hash function to find a new location. $h'(k, i) = (h(k) + i \cdot g(k)) \% T$

```java
public int hashFunction(String s)
    int naturalHash = this.getHash(s);
    if(naturalHash in use) {
        int i = 1;
        while (index in use) {
            try (naturalHash + i * jump_Hash(key));
            i++;
        }
    }
```
### ArrayDictionary<\(K, V\)>

<table>
<thead>
<tr>
<th>Function</th>
<th>Best case</th>
<th>Worst case</th>
</tr>
</thead>
<tbody>
<tr>
<td>get(K key)</td>
<td>(O(1)) Key is first item looked at</td>
<td>(O(n)) Key is not found</td>
</tr>
<tr>
<td>put(K key, V value)</td>
<td>(O(1)) Key is first item looked at</td>
<td>(O(n)) N search, N resizing</td>
</tr>
<tr>
<td>remove(K key)</td>
<td>(O(1)) Key is first item looked at</td>
<td>(O(n)) Key is not found</td>
</tr>
<tr>
<td>containsKey(K key)</td>
<td>(O(1)) Key is first item looked at</td>
<td>(O(n)) N search, C swapping</td>
</tr>
<tr>
<td>size()</td>
<td>(O(1)) Return field</td>
<td>(O(1)) Return field</td>
</tr>
</tbody>
</table>

### DoubleLinkedList<T>

<table>
<thead>
<tr>
<th>Function</th>
<th>Best case</th>
<th>Worst case</th>
</tr>
</thead>
<tbody>
<tr>
<td>get(int index)</td>
<td>(O(1)) Index is 0 or size</td>
<td>(n/2 \rightarrow O(n)) Index is size/2</td>
</tr>
<tr>
<td>add(T item)</td>
<td>(O(1)) Item added to back</td>
<td>(O(1)) Item added to back</td>
</tr>
<tr>
<td>remove()</td>
<td>(O(1)) Item removed from back</td>
<td>(O(1)) Item removed from back</td>
</tr>
<tr>
<td>delete(int index)</td>
<td>(O(1)) Index is 0 or size</td>
<td>(n/2 \rightarrow O(n)) Index is size/2</td>
</tr>
<tr>
<td>set(int index, T item)</td>
<td>(O(1)) Index is 0 or size</td>
<td>(n/2 \rightarrow O(n)) Index is size/2</td>
</tr>
<tr>
<td>insert(int index, T item)</td>
<td>(O(1)) Index is 0 or size</td>
<td>(n/2 \rightarrow O(n)) Index is size/2</td>
</tr>
</tbody>
</table>