Acknowledgement: The course slides are adapted from the slides prepared by R. Sedgewick and K. Wayne of Princeton University.

Cast of characters

Programmer needs to develop a working solution.

Client wants to solve problem efficiently.

Student might play any or all of these roles someday.

Theoretician wants to understand.

Basic blocking and tackling is sometimes necessary. [this lecture]

Running time

"As soon as an Analytic Engine exists, it will necessarily guide the future course of the science. Whenever any result is sought by its aid, the question will arise—By what course of calculation can these results be arrived at by the machine in the shortest time?" — Charles Babbage (1864)
Reasons to analyse algorithms

- Predict performance.
- Compare algorithms.
- Provide guarantees.
- Understand theoretical basis.

Primary practical reason: avoid performance bugs.

Some algorithmic successes

N-body simulation.
- Simulate gravitational interactions among $N$ bodies.
- Brute force: $N^2$ steps.
- Barnes-Hut algorithm: $N \log N$ steps, enables new research.

Discrete Fourier transform.
- Break down waveform of $N$ samples into periodic components.
- Applications: DVD, JPEG, MRI, astrophysics, ....
- Brute force: $N^2$ steps.
- FFT algorithm: $N \log N$ steps, enables new technology.

- sFFT: Sparse Fast Fourier Transform algorithm (Hassanieh et al., 2012)
  - A faster Fourier Transform: $k \log N$ steps (with $k$ sparse coefficients)

The challenge

Q. Will my program be able to solve a large practical input?

Key insight. [Knuth 1970s] Use scientific method to understand performance.
**Scientific method applied to analysis of algorithms**

A framework for predicting performance and comparing algorithms.

**Scientific method.**
- **Observe** some feature of the natural world.
- **Hypothesize** a model that is consistent with the observations.
- **Predict** events using the hypothesis.
- **Verify** the predictions by making further observations.
- **Validate** by repeating until the hypothesis and observations agree.

**Principles.**
- Experiments must be reproducible.
- Hypotheses must be falsifiable.

Feature of the natural world = computer itself.

---

**Example: 3-sum**

3-sum. Given $N$ distinct integers, how many triples sum to exactly zero?

```java
public class ThreeSum {
    public static int count(int[] a) {
        int N = a.length;
        int count = 0;
        for (int i = 0; i < N; i++)
            for (int j = i+1; j < N; j++)
                for (int k = j+1; k < N; k++)
                    if (a[i] + a[j] + a[k] == 0)
                        count++;
        return count;
    }

    public static void main(String[] args) {
        int[] a = In.readInts(args[0]);
        StdOut.println(count(a));
    }
}
```

---

**Context.** Deeply related to problems in computational geometry.
Q. How to time a program?
A. Manual.

Measuring the running time

Measuring the running time

Q. How to time a program?
A. Automatic.

public class Stopwatch
{
    private final long start = System.currentTimeMillis();
    public double elapsedTime()
    {
        long now = System.currentTimeMillis();
        return (now - start) / 1000.0;
    }
}

Implementation (part of stdlib.jar)

public static void main(String[] args)
{
    int[] a = In.readInts(args[0]);
    Stopwatch stopwatch = new Stopwatch();
    StdOut.println(ThreeSum.count(a));
    double time = stopwatch.elapsedTime();
}

Empirical analysis

Run the program for various input sizes and measure running time.

<table>
<thead>
<tr>
<th>N</th>
<th>time (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>0</td>
</tr>
<tr>
<td>500</td>
<td>0</td>
</tr>
<tr>
<td>1000</td>
<td>0.1</td>
</tr>
<tr>
<td>2000</td>
<td>0.8</td>
</tr>
<tr>
<td>4000</td>
<td>6.4</td>
</tr>
<tr>
<td>8000</td>
<td>51.1</td>
</tr>
<tr>
<td>16000</td>
<td>?</td>
</tr>
</tbody>
</table>
Data analysis

**Standard plot.** Plot running time $T(N)$ vs. input size $N$.

<table>
<thead>
<tr>
<th>Problem Size $N$</th>
<th>Running Time $T(N)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1K</td>
<td>10</td>
</tr>
<tr>
<td>2K</td>
<td>20</td>
</tr>
<tr>
<td>4K</td>
<td>40</td>
</tr>
<tr>
<td>8K</td>
<td>80</td>
</tr>
</tbody>
</table>

Log-log plot. Plot running time $T(N)$ vs. input size $N$ using log-log scale.

Regression. Fit straight line through data points: $a N^b$.

Hypothesis. The running time is about $1.006 \times 10^{-10} \times N^{2.999}$ seconds.

Prediction and validation

**Hypothesis.** The running time is about $1.006 \times 10^{-10} \times N^{2.999}$ seconds.

**Predictions.**
- 51.0 seconds for $N = 8,000$.
- 408.1 seconds for $N = 16,000$.

**Observations.**

<table>
<thead>
<tr>
<th>$N$</th>
<th>Time (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8,000</td>
<td>51.1</td>
</tr>
<tr>
<td>8,000</td>
<td>51</td>
</tr>
<tr>
<td>8,000</td>
<td>51.1</td>
</tr>
<tr>
<td>16,000</td>
<td>410.8</td>
</tr>
</tbody>
</table>

Doubling hypothesis

**Doubling hypothesis.** Quick way to estimate $b$ in a power-law relationship.

Run program, **doubling** the size of the input.

<table>
<thead>
<tr>
<th>$N$</th>
<th>time (seconds)</th>
<th>ratio</th>
<th>lg ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>500</td>
<td>0</td>
<td>4.8</td>
<td>2.3</td>
</tr>
<tr>
<td>1,000</td>
<td>0.1</td>
<td>6.9</td>
<td>2.8</td>
</tr>
<tr>
<td>2,000</td>
<td>0.8</td>
<td>7.7</td>
<td>2.9</td>
</tr>
<tr>
<td>4,000</td>
<td>6.4</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>8,000</td>
<td>51.1</td>
<td>8</td>
<td>3</td>
</tr>
</tbody>
</table>

Hypothesis. Running time is about $a N^b$ with $b = \lg$ ratio.

**Caveat.** Cannot identify logarithmic factors with doubling hypothesis.
Doubling hypothesis

Doubling hypothesis. Quick way to estimate $b$ in a power-law hypothesis.

Q. How to estimate $a$ (assuming we know $b$)?
A. Run the program (for a sufficient large value of $N$) and solve for $a$.

$$51.1 = a \times 8000^3$$

$$a = 0.998 \times 10^{-10}$$

Hypothesis. Running time is about $0.998 \times 10^{-10} \times N^3$ seconds.

almost identical hypothesis to one obtained via linear regression

In practice, constant factors matter too!

Q. How long does this program take as a function of $N$?

```java
String s = StdIn.readString();
int N = s.length();
for (int i = 0; i < N; i++)
for (int j = 0; j < N; j++)
distance[i][j] = ...
```

<table>
<thead>
<tr>
<th>$N$</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>0.11</td>
</tr>
<tr>
<td>2.000</td>
<td>0.35</td>
</tr>
<tr>
<td>4.000</td>
<td>1.6</td>
</tr>
<tr>
<td>8.000</td>
<td>6.5</td>
</tr>
</tbody>
</table>

Jenny ~ $c_1 N^2$ seconds
Kenny ~ $c_2 N$ seconds

Experimental algorithmics

System independent effects.
- Algorithm.
- Input data.
  - determines exponent $b$ in power law
  - determines constant $a$ in power law

System dependent effects.
- Hardware: CPU, memory, cache, ...
- Software: compiler, interpreter, garbage collector, ...
- System: operating system, network, other applications, ...

Bad news. Difficult to get precise measurements.
Good news. Much easier and cheaper than other sciences.
  - e.g., can run huge number of experiments

Analysis of Algorithms

- Observations
- Mathematical models
- Order-of-growth classifications
- Dependencies on inputs
- Memory
Mathematical models for running time

Total running time: sum of cost $\times$ frequency for all operations.

- Need to analyze program to determine set of operations.
- Cost depends on machine, compiler.
- Frequency depends on algorithm, input data.

In principle, accurate mathematical models are available.

Cost of basic operations

<table>
<thead>
<tr>
<th>operation</th>
<th>example</th>
<th>nanoseconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>integer add</td>
<td>$a + b$</td>
<td>2.1</td>
</tr>
<tr>
<td>integer multiply</td>
<td>$a \times b$</td>
<td>2.4</td>
</tr>
<tr>
<td>integer divide</td>
<td>$a / b$</td>
<td>5.4</td>
</tr>
<tr>
<td>floating-point add</td>
<td>$a + b$</td>
<td>4.6</td>
</tr>
<tr>
<td>floating-point multiply</td>
<td>$a \times b$</td>
<td>4.2</td>
</tr>
<tr>
<td>floating-point divide</td>
<td>$a / b$</td>
<td>13.5</td>
</tr>
<tr>
<td>sine</td>
<td>Math.sin(theta)</td>
<td>91.3</td>
</tr>
<tr>
<td>arctangent</td>
<td>Math.atan2(y, x)</td>
<td>129</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

† Running OS X on Macbook Pro 2.2GHz with 2GB RAM

Example: 1-sum

Q. How many instructions as a function of input size $N$?

```
int count = 0;
for (int i = 0; i < N; i++)
  if (a[i] == 0)
    count++;
```

Cost of instructions

<table>
<thead>
<tr>
<th>operation</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>variable declaration</td>
<td>2</td>
</tr>
<tr>
<td>assignment statement</td>
<td>2</td>
</tr>
<tr>
<td>less than compare</td>
<td>$N + 1$</td>
</tr>
<tr>
<td>equal to compare</td>
<td>$N$</td>
</tr>
<tr>
<td>array access</td>
<td>$N$</td>
</tr>
<tr>
<td>increment</td>
<td>$N$ to $2N$</td>
</tr>
</tbody>
</table>
Example: 2-sum

Q. How many instructions as a function of input size \( N \)?

```c
int count = 0;
for (int i = 0; i < N; i++)
for (int j = i+1; j < N; j++)
  if (a[i] + a[j] == 0)
    count++;
```

\[
0 + 1 + 2 + \ldots + (N - 1) = \frac{N(N - 1)}{2}
\]

<table>
<thead>
<tr>
<th>Operation</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>variable declaration</td>
<td>( N + 2 )</td>
</tr>
<tr>
<td>assignment statement</td>
<td>( N + 2 )</td>
</tr>
<tr>
<td>less than compare</td>
<td>( \frac{1}{2} ) ( N(N + 2) )</td>
</tr>
<tr>
<td>equal to compare</td>
<td>( \frac{1}{2} N ) ( N - 1 )</td>
</tr>
<tr>
<td>array access</td>
<td>( N(N - 1) )</td>
</tr>
<tr>
<td>increment</td>
<td>( \frac{1}{2} N ) ( N - 1 ) to ( N(N - 1) )</td>
</tr>
</tbody>
</table>

Simplification 1: cost model

Cost model. Use some basic operation as a proxy for running time.

```c
int count = 0;
for (int i = 0; i < N; i++)
for (int j = i+1; j < N; j++)
  if (a[i] + a[j] == 0)
    count++;
```

\[
0 + 1 + 2 + \ldots + (N - 1) = \frac{N(N - 1)}{2}
\]

Simplification 2: tilde notation

- Estimate running time (or memory) as a function of input size \( N \).
- Ignore lower order terms.
  - when \( N \) is large, terms are negligible
  - when \( N \) is small, we don’t care

Ex 1. \( \frac{1}{6} N^3 + 20 N + 16 \sim \frac{1}{6} N^3 \)
Ex 2. \( \frac{1}{6} N^3 + 100 \frac{1}{2} N^{4/3} + 56 \sim \frac{1}{6} N^3 \)
Ex 3. \( \frac{1}{6} N^3 - \frac{1}{2} N^2 + \frac{1}{6} \sim \frac{1}{6} N^3 \)

Leading-term approximation
e (e.g., \( N = 1000 \): 500 thousand vs. 166 million)

Technical definition. \( f(N) \approx g(N) \) means

\[
limit_{N \to \infty} \frac{f(N)}{g(N)} = 1
\]
Estimate running time (or memory) as a function of input size $N$.

Ignore lower order terms.

- when $N$ is large, terms are negligible
- when $N$ is small, we don’t care

Simplification 2: tilde notation

<table>
<thead>
<tr>
<th>operation</th>
<th>frequency</th>
<th>tilde notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>variable declaration</td>
<td>$N + 2$</td>
<td>$\sim N$</td>
</tr>
<tr>
<td>assignment statement</td>
<td>$N + 2$</td>
<td>$\sim N$</td>
</tr>
<tr>
<td>less than compare</td>
<td>$\frac{1}{2} (N + 1) (N + 2)$</td>
<td>$\sim \frac{1}{2} N^2$</td>
</tr>
<tr>
<td>equal to compare</td>
<td>$\frac{1}{2} N (N - 1)$</td>
<td>$\sim \frac{1}{2} N^2$</td>
</tr>
<tr>
<td>array access</td>
<td>$N (N - 1)$</td>
<td>$\sim N^2$</td>
</tr>
<tr>
<td>increment</td>
<td>$\frac{1}{2} N (N - 1)$ to $N (N - 1)$</td>
<td>$\sim \frac{1}{2} N^2$ to $\sim N^2$</td>
</tr>
</tbody>
</table>

Example: 2-sum

Q. Approximately how many array accesses as a function of input size $N$?

A. $\sim N^2$ array accesses.

Bottom line. Use cost model and tilde notation to simplify frequency counts.

Example: 3-sum

Q. Approximately how many array accesses as a function of input size $N$?

A. $\sim \frac{1}{2} N^3$ array accesses.

Bottom line. Use cost model and tilde notation to simplify frequency counts.

Example: 2-sum

```java
int count = 0;
for (int i = 0; i < N; i++)
    for (int j = i+1; j < N; j++)
        if (a[i] + a[j] == 0)
            count++;
```

Q. How to estimate a discrete sum?

A1. Take discrete mathematics course.

A2. Replace the sum with an integral, and use calculus!

Ex 1. $1 + 2 + \ldots + N$. $\sum_{i=1}^{N} i \sim \int_{x=1}^{N} x \, dx = \frac{1}{2} N^2$

Ex 2. $1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{N}$. $\sum_{i=1}^{N} \frac{1}{i} \sim \int_{x=1}^{N} \frac{1}{x} \, dx = \ln N$

Ex 3. 3-sum triple loop. $\sum_{i=1}^{N} \sum_{j=i+1}^{N} \sum_{k=j+1}^{N} 1 \sim \int_{x=1}^{N} \int_{y=x}^{N} \int_{z=y}^{N} dz \, dy \, dx \sim \frac{1}{6} N^3$
Mathematical models for running time

In principle, accurate mathematical models are available.

In practice,
• Formulas can be complicated.
• Advanced mathematics might be required.
• Exact models best left for experts.

Bottom line. We use approximate models in this course: \( T(N) \sim cN^3 \).

Common order-of-growth classifications

Good news. The small set of functions
1. \( \log N \), \( N \), \( N \log N \), \( N^2 \), \( N^3 \), and \( 2^N \)
suffices to describe order-of-growth of typical algorithms.

Common order-of-growth classifications

<table>
<thead>
<tr>
<th>order of growth</th>
<th>name</th>
<th>typical code framework</th>
<th>description</th>
<th>example</th>
<th>( T(2N) / T(N) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 1 )</td>
<td>constant</td>
<td>( a = b + c; )</td>
<td>statement</td>
<td>add two numbers</td>
<td>1</td>
</tr>
<tr>
<td>( \log N )</td>
<td>logarithmic</td>
<td>while ( (N &gt; 1) ) { ( N = N / 2 );… }</td>
<td>divide in half</td>
<td>binary search</td>
<td>-1</td>
</tr>
<tr>
<td>( N )</td>
<td>linear</td>
<td>for ( { \text{int } i = 0; i &lt; N; i++ } )</td>
<td>loop</td>
<td>find the maximum</td>
<td>2</td>
</tr>
<tr>
<td>( N \log N )</td>
<td>linearithmic</td>
<td>[see mergesort lecture]</td>
<td>divide and conquer</td>
<td>mergesort</td>
<td>-2</td>
</tr>
<tr>
<td>( N^2 )</td>
<td>quadratic</td>
<td>for ( { \text{int } i = 0; i &lt; N; i++ } )</td>
<td>double loop</td>
<td>check all pairs</td>
<td>4</td>
</tr>
<tr>
<td>( N^3 )</td>
<td>cubic</td>
<td>for ( { \text{int } i = 0; i &lt; N; i++ } )</td>
<td>triple loop</td>
<td>check all triples</td>
<td>8</td>
</tr>
<tr>
<td>( 2^N )</td>
<td>exponential</td>
<td>[see combinatorial search lecture]</td>
<td>exhaustive search</td>
<td>check all subsets</td>
<td>( T(N) )</td>
</tr>
</tbody>
</table>
### Practical implications of order-of-growth

<table>
<thead>
<tr>
<th>growth rate</th>
<th>problem size solvable in minutes</th>
<th>time to process millions of inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>any</td>
<td>any</td>
</tr>
<tr>
<td>log N</td>
<td>any</td>
<td>any</td>
</tr>
<tr>
<td>N</td>
<td>millions</td>
<td>tens of millions</td>
</tr>
<tr>
<td>N log N</td>
<td>hundreds of thousands</td>
<td>millions</td>
</tr>
<tr>
<td>N²</td>
<td>hundreds thousand</td>
<td>thousands</td>
</tr>
<tr>
<td>N³</td>
<td>hundred hundreds thousand</td>
<td>thousand</td>
</tr>
<tr>
<td>2ⁿ</td>
<td>20</td>
<td>20s</td>
</tr>
</tbody>
</table>

### Practical implications of order-of-growth

<table>
<thead>
<tr>
<th>growth rate</th>
<th>name</th>
<th>description</th>
<th>effect on a program that runs for a few seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>time for 100x more data</td>
<td>size for 100x faster computer</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>constant</td>
<td>independent of input size</td>
<td>–</td>
</tr>
<tr>
<td>log N</td>
<td>logarithmic</td>
<td>nearly independent of input size</td>
<td>–</td>
</tr>
<tr>
<td>N</td>
<td>linear</td>
<td>optimal for N inputs</td>
<td>a few minutes</td>
</tr>
<tr>
<td>N log N</td>
<td>linearithmic</td>
<td>nearly optimal for N inputs</td>
<td>a few minutes</td>
</tr>
<tr>
<td>N²</td>
<td>quadratic</td>
<td>not practical for large problems</td>
<td>several hours</td>
</tr>
<tr>
<td>N³</td>
<td>cubic</td>
<td>not practical for medium problems</td>
<td>several weeks</td>
</tr>
<tr>
<td>2ⁿ</td>
<td>exponential</td>
<td>useful only for tiny problems</td>
<td>forever</td>
</tr>
</tbody>
</table>

### Binary search

**Goal.** Given a sorted array and a key, find index of the key in the array?

**Binary search.** Compare key against middle entry.
- Too small, go left.
- Too big, go right.
- Equal, found.

<table>
<thead>
<tr>
<th>i</th>
<th>j</th>
<th>mid</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
</tbody>
</table>

↑ | ↑ | ↑

lo | mid | hi
**Binary search demo**

*Goal.* Given a sorted array and a key, find index of the key in the array!

*Successful search.* Binary search for 33.

<table>
<thead>
<tr>
<th>6</th>
<th>13</th>
<th>14</th>
<th>25</th>
<th>33</th>
<th>43</th>
<th>51</th>
<th>53</th>
<th>64</th>
<th>72</th>
<th>84</th>
<th>93</th>
<th>95</th>
<th>96</th>
<th>97</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
</tbody>
</table>

↑

lo

↑

mid

↑

hi

**Goal.* Given a sorted array and a key, find index of the key in the array!

*Successful search.* Binary search for 33.

<table>
<thead>
<tr>
<th>6</th>
<th>13</th>
<th>14</th>
<th>25</th>
<th>33</th>
<th>43</th>
<th>51</th>
<th>53</th>
<th>64</th>
<th>72</th>
<th>84</th>
<th>93</th>
<th>95</th>
<th>96</th>
<th>97</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>14</td>
</tr>
</tbody>
</table>

↑

lo

↑

mid

↑

hi

**Goal.* Given a sorted array and a key, find index of the key in the array!

*Successful search.* Binary search for 33.

<table>
<thead>
<tr>
<th>6</th>
<th>13</th>
<th>14</th>
<th>25</th>
<th>33</th>
<th>43</th>
<th>51</th>
<th>53</th>
<th>64</th>
<th>72</th>
<th>84</th>
<th>93</th>
<th>95</th>
<th>96</th>
<th>97</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
</tbody>
</table>

↑

mid

return 4
**Binary search demo**

**Goal.** Given a sorted array and a key, find index of the key in the array!

**Unsuccessful search.** Binary search for 34.

```
<table>
<thead>
<tr>
<th>6</th>
<th>13</th>
<th>14</th>
<th>25</th>
<th>33</th>
<th>43</th>
<th>51</th>
<th>53</th>
<th>64</th>
<th>72</th>
<th>84</th>
<th>93</th>
<th>95</th>
<th>96</th>
<th>97</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
</tbody>
</table>
```

lo = hi

return -1
Binary search: Java implementation

Trivial to implement?

• First binary search published in 1946; first bug-free one published in 1962.
• Bug in Java’s Arrays.binarySearch() discovered in 2006.

```
public static int binarySearch(int[] a, int key)
{
    int lo = 0, hi = a.length-1;
    while (lo <= hi)
    {
        int mid = lo + (hi - lo) / 2;
        if      (key < a[mid]) hi = mid - 1;
        else if (key > a[mid]) lo = mid + 1;
        else return mid;
    }
    return -1;
}
```

Invariant. If key appears in the array a[], then a[lo] \leq key \leq a[hi].

Binary search: mathematical analysis

Proposition. Binary search uses at most 1 + \lg N compares to search in a sorted array of size N.

Def. \( T(N) = \# \) compares to binary search in a sorted subarray of size at most N.

Binary search recurrence. \( T(N) \leq T(N/2) + 1 \) for \( N > 1 \), with \( T(1) = 1 \).

Pf sketch.

Algorithm.

• Sort the N (distinct) numbers.
• For each pair of numbers a[i] and a[j], binary search for -(a[i] + a[j]).

Analysis. Order of growth is \( N^2 \log N \).

• Step 1: \( N^2 \) with insertion sort.
• Step 2: \( N^2 \log N \) with binary search.

An \( N^2 \) log N algorithm for 3-sum

Input

```
30 -40 -20 -10 40  0 10  5
```

Sort

```
-40 -20 -10  0  5 10 30 40
```

Binary search

```
(-40, -20) 60
(-40, -10) 50
(-40,  0)  40
(-40,  5)  30
(-40, 10)  20
(-40, 40)  10
(-10,  0)  10
(-20,  10)  9
```

Analysis. Order of growth is \( N^2 \log N \).

• Step 1: \( N^2 \) with insertion sort.
• Step 2: \( N^2 \log N \) with binary search.

Only count if a[i] < a[j] < a[k] to avoid double counting.
Comparing programs

Hypothesis. The $N^2 \log N$ three-sum algorithm is significantly faster in practice than the brute-force $N^3$ algorithm.

<table>
<thead>
<tr>
<th>N</th>
<th>time (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>0.1</td>
</tr>
<tr>
<td>2.000</td>
<td>0.8</td>
</tr>
<tr>
<td>4.000</td>
<td>6.4</td>
</tr>
<tr>
<td>8.000</td>
<td>51.1</td>
</tr>
</tbody>
</table>

ThreeSum.java

Guiding principle. Typically, better order of growth $\Rightarrow$ faster in practice.

Types of analyses

**Best case.** Lower bound on cost.
- Determined by “easiest” input.
- Provides a goal for all inputs.

**Worst case.** Upper bound on cost.
- Determined by “most difficult” input.
- Provides a guarantee for all inputs.

**Average case.** Expected cost for random input.
- Need a model for “random” input.
- Provides a way to predict performance.

Ex 1. Array accesses for brute-force 3 sum.

- Best: $\sim \frac{1}{2} N^3$
- Average: $\sim \frac{1}{2} N^3$
- Worst: $\sim \frac{1}{2} N^3$

Ex 2. Compares for binary search.

- Best: $\sim 1$
- Average: $\sim \log N$
- Worst: $\sim \log N$

Analysis of Algorithms

- Observations
- Mathematical models
- Order-of-growth classifications
- Dependencies on inputs
- Memory

Types of analyses

**Best case.** Lower bound on cost.

**Worst case.** Upper bound on cost.

**Average case.** “Expected” cost.

Actual data might not match input model?
- Need to understand input to effectively process it.
- Approach 1: design for the worst case.
- Approach 2: randomize, depend on probabilistic guarantee.
**Theory of Algorithms**

**Goals.**
- Establish “difficulty” of a problem.
- Develop “optimal” algorithms.

**Approach.**
- Suppress details in analysis: analyze “to within a constant factor”.
- Eliminate variability in input model by focusing on the worst case.

**Optimal algorithm.**
- Performance guarantee (to within a constant factor) for any input.
- No algorithm can provide a better performance guarantee.

---

**Commonly-used notations**

<table>
<thead>
<tr>
<th>notation</th>
<th>provides</th>
<th>example</th>
<th>shorthand for</th>
<th>used to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tilde</td>
<td>leading term</td>
<td>$\tilde{10 N^2}$</td>
<td>$10 N^2 + 22 N \log N$</td>
<td>provide approximate model</td>
</tr>
<tr>
<td>Big Theta</td>
<td>asymptotic growth rate</td>
<td>$\Theta(N^p)$</td>
<td>$\frac{1}{2} N^p \quad 5 N^p + 22 N \log N + 3N$</td>
<td>classify algorithms</td>
</tr>
<tr>
<td>Big Oh</td>
<td>$\Theta(N^p)$ and smaller</td>
<td>$O(N^p)$</td>
<td>$10 N^p \quad 100 N \quad 22 N \log N + 3N$</td>
<td>develop upper bounds</td>
</tr>
<tr>
<td>Big Omega</td>
<td>$\Theta(N^p)$ and larger</td>
<td>$\Omega(N^p)$</td>
<td>$\frac{1}{2} N^p \quad N^p \quad N^p + 22 N \log N + 3N$</td>
<td>develop lower bounds</td>
</tr>
</tbody>
</table>

**Common mistake.** Interpreting big-Oh as an approximate model.

---

**Tilde notation vs. big-Oh notation**

We use tilde notation whenever possible.
- Big-Oh notation suppresses leading constant.
- Big-Oh notation only provides upper bound (not lower bound).

---

**Theory of algorithms: example 1**

**Goals.**
- Establish “difficulty” of a problem and develop “optimal” algorithms.
- Ex. 1-SUM = “Is there a 0 in the array? ”

**Upper bound.** A specific algorithm.
- Running time of the optimal algorithm for 1-SUM is $O(N)$.

**Lower bound.** Proof that no algorithm can do better.
- Ex. Have to examine all $N$ entries (any unexamined one might be 0).
- Running time of the optimal algorithm for 1-SUM is $\Omega(N)$.

**Optimal algorithm.**
- Lower bound equals upper bound (to within a constant factor).
- Ex. Brute-force algorithm for 1-SUM is optimal: its running time is $\Theta(N)$. 
Theory of algorithms: example 2

**Goals.**
- Establish “difficulty” of a problem and develop “optimal” algorithms.
- Ex. 3-SUM

**Upper bound.** A specific algorithm.
- Ex. Brute-force algorithm for 3-SUM
- Running time of the optimal algorithm for 3-SUM is $O(N^3)$.

**Open problems.**
- Optimal algorithm for 3-SUM?
- Subquadratic algorithm for 3-SUM?
- Quadratic lower bound for 3-SUM?

Algorithm design approach

**Start.**
- Develop an algorithm.
- Prove a lower bound.

**Gap!**
- Lower the upper bound (discover a new algorithm).
- Raise the lower bound (more difficult).

**Golden Age of Algorithm Design.**
- 1970s-.
- Steadily decreasing upper bounds for many important problems.
- Many known optimal algorithms.

**Caveats.**
- Overly pessimistic to focus on worst case?
- Need better than “to within a constant factor” to predict performance.

Analysis of Algorithms

- Observations
- Mathematical models
- Order-of-growth classifications
- Dependencies on inputs
- Memory
Basics

Bit. 0 or 1.
Byte. 8 bits.
Megabyte (MB). 1 million or $2^{20}$ bytes.
Gigabyte (GB). 1 billion or $2^{30}$ bytes.

Old machine. We used to assume a 32-bit machine with 4 byte pointers.

Modern machine. We now assume a 64-bit machine with 8 byte pointers.
• Can address more memory.
• Pointers use more space.

Object overhead. 16 bytes.
Reference. 8 bytes.
Padding. Each object uses a multiple of 8 bytes.

Ex 1. A Date object uses 32 bytes of memory.

Ex 2. A String of length $N$ uses $\sim 2^N$ bytes of memory.
Typical memory usage summary

Total memory usage for a data type value:
- Primitive type: 4 bytes for int, 8 bytes for double, …
- Object reference: 8 bytes.
- Array: 24 bytes + memory for each array entry.
- Object: 16 bytes + memory for each instance variable + 8 if inner class.

Shallow memory usage: Don’t count referenced objects.

Deep memory usage: If array entry or instance variable is a reference, add memory (recursively) for referenced object.

Turning the crank: summary

Empirical analysis.
- Execute program to perform experiments.
- Assume power law and formulate a hypothesis for running time.
- Model enables us to make predictions.

Mathematical analysis.
- Analyze algorithm to count frequency of operations.
- Use tilde notation to simplify analysis.
- Model enables us to explain behaviour.

Scientific method.
- Mathematical model is independent of a particular system; applies to machines not yet built.
- Empirical analysis is necessary to validate mathematical models and to make predictions.

Memory profiler

Classmexer library. Measure memory usage of a Java object by querying JVM.
http://www.javamex.com/classmexer

import com.javamex.classmexer.MemoryUtil;
public class Memory {
  public static void main(String[] args) {
    Date date = new Date(12, 31, 1999);
    StdOut.println(MemoryUtil.memoryUsageOf(date));
    String s = "Hello, World";
    StdOut.println(MemoryUtil.memoryUsageOf(s));
    StdOut.println(MemoryUtil.deepMemoryUsageOf(s));
  }
}

% javac -cp ..:classmexer.jar Memory.java
% java -cp ..:classmexer.jar -javaagent:classmexer.jar Memory
32
40
88
2N + 64

use -XX:-UseCompressedOops on OS X to match our model