1.4 Analysis of Algorithms

- introduction
- observations
- mathematical models
- order-of-growth classifications
- theory of algorithms
- memory (read your textbook)
1.4 Analysis of Algorithms

- introduction
- observations
- mathematical models
- order-of-growth classifications
- theory of algorithms
- memory
Cast of characters

**Programmer** needs to develop a working solution.

**Client** wants to solve problem efficiently.

**Theoretician** seeks to understand.

**Student (you)** might play any or all of these roles someday.
“As soon as an Analytical Engine exists, it will necessarily guide the future course of the science. Whenever any result is sought by its aid, the question will then arise—By what course of calculation can these results be arrived at by the machine in the shortest time? ” — Charles Babbage (1864)
Reasons to analyze algorithms

Predict performance.

Compare algorithms. this course
(CIS 121)

Provide guarantees. theory of algorithms
(CIS 320)

Understand theoretical basis.

Primary practical reason: avoid performance bugs.

client gets poor performance because programmer
did not understand performance characteristics
Another algorithmic success story

Discrete Fourier transform.
- Express signal as weighted sum of sines and cosines.
- Applications: DVD, JPEG, MRI, astrophysics, ...
- Brute force: \( N^2 \) steps.
- FFT algorithm: \( N \log N \) steps, enables new technology.

![Graph showing time and size with different growth rates: linear, linearithmic, and quadratic limits.](image)
The challenge

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

Why is my program so slow?  Why does it run out of memory?

Insight. [Knuth 1970s] Use scientific method to understand performance.
Scientific method applied to the 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.
1.4 Analysis of Algorithms

- introduction
- observations
- mathematical models
- order-of-growth classifications
- theory of algorithms
- memory
Running example: 3-SUM

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

% more 8ints.txt
8
30 -40 -20 -10 40 0 10 5
% java ThreeSum 8ints.txt
4

<table>
<thead>
<tr>
<th></th>
<th>a[i]</th>
<th>a[j]</th>
<th>a[k]</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>-40</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>-20</td>
<td>-10</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>-40</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>-10</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

Context. Deeply related to problems in computational geometry.
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)
    {
        In in = new In(args[0]);
        int[] a = in.readAllInts();
        StdOut.println(count(a));
    }
}
Measuring the running time

**Q.** How to time a program?

**A.** Manual.
Measuring the running time

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

```java
public class Stopwatch {
    // create a new stopwatch
    public Stopwatch()
    {
        // time since creation (in seconds)
        double elapsedTime()
    }
}
```

```java
public static void main(String[] args)
{
    In in = new In(args[0]);
    int[] a = in.readInts();
    Stopwatch stopwatch = new Stopwatch();
    StdOut.println(ThreeSum.count(a));
    double time = stopwatch.elapsedTime();
    StdOut.println("elapsed time = " + time);
}
```
Empirical analysis

Run the program for various input sizes and measure running time.
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>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>
<tr>
<td>16,000</td>
<td>?</td>
</tr>
</tbody>
</table>

† on a 2.8GHz Intel PU-226 with 64GB DDR E3 memory and 32MB L3 cache; running Oracle Java 1.7.0_45-b18 on Springdale Linux v. 6.5
Data analysis

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

![Diagram showing the relationship between problem size $N$ and running time $T(N)$](image-url)
**Data analysis**

**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) $\dagger$</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>

"order of growth" of running time is about $N^3$ [stay tuned] validates hypothesis!
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) ( \dagger )</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>

\[
\frac{T(N)}{T(N/2)} = \frac{aN^b}{a(N/2)^b} = 2^b
\]

\( \lg (6.4 / 0.8) = 3.0 \)

seems to converge to a constant \( b \approx 3 \)

**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 relationship.

**Q.** How to estimate $a$ (assuming we know $b$)?

**A.** Run the program (for a sufficient large value of $N$) and solve for $a$.

<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>
</tbody>
</table>

$51.1 = a \times 8000^3$

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

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

almost identical hypothesis to one obtained via regression
Experimental algorithmics

System independent effects.
- Algorithm.
- Input data.

\[ \text{determines exponent } b \]
\[ \text{in power law } a N^b \]

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

\[ \text{determines constant } a \]
\[ \text{in power law } a N^b \]

Bad news. Sometimes difficult to get precise measurements.

Good news. Much easier and cheaper than other sciences.
Good news for computer science

Algorithmic experiments are virtually free by comparison with other sciences.

Bottom line. No excuse for not running experiments to understand costs.
1.4 **Analysis of Algorithms**

- introduction
- observations
- *mathematical models*
- order-of-growth classifications
- theory of algorithms
- memory
Mathematical models for running time

**Total running time:** sum of cost × 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.
Example: 1-SUM

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

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

<table>
<thead>
<tr>
<th>operation</th>
<th>cost (ns) †</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>variable declaration</td>
<td>2/5</td>
<td>2</td>
</tr>
<tr>
<td>assignment statement</td>
<td>1/5</td>
<td>2</td>
</tr>
<tr>
<td>less than compare</td>
<td>1/5</td>
<td>$N + 1$</td>
</tr>
<tr>
<td>equal to compare</td>
<td>1/10</td>
<td>$N$</td>
</tr>
<tr>
<td>array access</td>
<td>1/10</td>
<td>$N$</td>
</tr>
<tr>
<td>increment</td>
<td>1/10</td>
<td>$N$ to $2N$</td>
</tr>
</tbody>
</table>

† representative estimates (with some poetic license)
Example: 2-SUM

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

```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++;
```

Pf. [Gauss]

\[
\begin{align*}
T(N) & = 0 + 1 + \ldots + (N-2) + (N-1) \\
+ T(N) & = (N-1) + (N-2) + \ldots + 1 + 0 \\
2T(N) & = (N-1) + (N-1) + \ldots + (N-1) + (N-1) \\
\Rightarrow T(N) & = \frac{N(N-1)}{2}
\end{align*}
\]

Karl Friedrich Gauss, when he was 10 year's old and a schoolteacher made him add $1+2+3+\ldots+100$

https://www.youtube.com/watch?v=tpkzn2e5mtl
Example: 2-SUM

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

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++;

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

<table>
<thead>
<tr>
<th>operation</th>
<th>cost (ns)</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>variable declaration</td>
<td>2/5</td>
<td>$N + 2$</td>
</tr>
<tr>
<td>assignment statement</td>
<td>1/5</td>
<td>$N + 2$</td>
</tr>
<tr>
<td>less than compare</td>
<td>1/5</td>
<td>$\frac{1}{2} (N + 1) (N + 2)$</td>
</tr>
<tr>
<td>equal to compare</td>
<td>1/10</td>
<td>$\frac{1}{2} N (N - 1)$</td>
</tr>
<tr>
<td>array access</td>
<td>1/10</td>
<td>$N (N - 1)$</td>
</tr>
<tr>
<td>increment</td>
<td>1/10</td>
<td>$\frac{1}{2} N (N + 1)$ to $N^2$</td>
</tr>
</tbody>
</table>

$1/4 N^2 + 13/20 N + 13/10$ ns to $3/10 N^2 + 3/5 N + 13/10$ ns (tedious to count exactly)
Simplifying the calculations

"It is convenient to have a measure of the amount of work involved in a computing process, even though it be a very crude one. We may count up the number of times that various elementary operations are applied in the whole process and then given them various weights. We might, for instance, count the number of additions, subtractions, multiplications, divisions, recording of numbers, and extractions of figures from tables. In the case of computing with matrices most of the work consists of multiplications and writing down numbers, and we shall therefore only attempt to count the number of multiplications and recordings. " — Alan Turing

ROUNDING-OFF ERRORS IN MATRIX PROCESSES

By A. M. TURING

(National Physical Laboratory, Teddington, Middlesex)

[Received 4 November 1947]

SUMMARY

A number of methods of solving sets of linear equations and inverting matrices are discussed. The theory of the rounding-off errors involved is investigated for some of the methods. In all cases examined, including the well-known ‘Gauss elimination process’, it is found that the errors are normally quite moderate: no exponential build-up need occur.
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++;
```

<table>
<thead>
<tr>
<th>operation</th>
<th>cost (ns)</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>variable declaration</td>
<td>2/5</td>
<td>$N + 2$</td>
</tr>
<tr>
<td>assignment statement</td>
<td>1/5</td>
<td>$N + 2$</td>
</tr>
<tr>
<td>less than compare</td>
<td>1/5</td>
<td>$\frac{1}{2} (N + 1) (N + 2)$</td>
</tr>
<tr>
<td>equal to compare</td>
<td>1/10</td>
<td>$\frac{1}{2} N (N - 1)$</td>
</tr>
<tr>
<td>array access</td>
<td>1/10</td>
<td>$N (N - 1)$</td>
</tr>
<tr>
<td>increment</td>
<td>1/10</td>
<td>$\frac{1}{2} N (N + 1)$ to $N^2$</td>
</tr>
</tbody>
</table>

$0 + 1 + 2 + \ldots + (N - 1) = \frac{1}{2} N (N - 1) = \binom{N}{2}$

(cost model = array accesses)

(we assume compiler/JVM do not optimize any array accesses away!)
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. $\frac{1}{6} N^3 + 100 N^{4/3} + 56 \sim \frac{1}{6} N^3$

Ex 3. $\frac{1}{6} N^3 - \frac{1}{2} N^2 + \frac{1}{3} N \sim \frac{1}{6} N^3$

discard lower-order terms
(e.g., $N = 1000$: 166.67 million vs. 166.17 million)

Technical definition. $f(N) \sim g(N)$ means $\lim_{N \to \infty} \frac{f(N)}{g(N)} = 1$
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

<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 $N^2$</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 $N$?

```
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++;
```

A. \( \sim N^2 \) array accesses.

Bottom line. Use cost model and tilde notation to simplify 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 counts.
Estimating a discrete sum

Q. How to estimate a discrete sum?
Estimating a discrete sum

Q. How to estimate a discrete sum?
A2. Replace the sum with an integral, and use calculus!

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

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.
\[
\sum_{i=1}^{N} \sum_{j=i}^{N} \sum_{k=j}^{N} 1 \sim \int_{x=1}^{N} \int_{y=x}^{N} \int_{z=y}^{N} dz \, dy \, dx \sim \frac{1}{6} N^3
\]

Ex 4. \(1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots\)
\[
\int_{x=0}^{\infty} \left(\frac{1}{2}\right)^x \, dx = \frac{1}{\ln 2} \approx 1.4427
\]
\[
\sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^i = 2
\]

integral trick doesn't always work!
Estimating a discrete sum

Q. How to estimate a discrete sum?
A3. Use Wolfram Alpha.

\[
\sum_{i=1}^{N} \left( \sum_{j=i+1}^{N} \left( \sum_{k=j+1}^{N} 1 \right) \right) = \frac{1}{6} N \left( N^2 - 3N + 2 \right)
\]

wolframalpha.com
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.

\[ T_N = c_1 A + c_2 B + c_3 C + c_4 D + c_5 E \]

- \(A\) = array access
- \(B\) = integer add
- \(C\) = integer compare
- \(D\) = increment
- \(E\) = variable assignment

Costs (depend on machine, compiler)

Frequencies (depend on algorithm, input)

Bottom line. We use approximate models in this course: \(T(N) \sim c N^3\).
1.4 Analysis of Algorithms

- introduction
- observations
- mathematical models
- order-of-growth classifications
- theory of algorithms
- memory
Common order-of-growth classifications

**Definition.** If \( f(N) \sim c \ g(N) \) for some constant \( c > 0 \), then the order of growth of \( f(N) \) is \( g(N) \).
- Ignores leading coefficient.
- Ignores lower-order terms.

**Ex.** The order of growth of the running time of this code is \( N^3 \).

```c
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++;
```

**Typical usage.** Mathematical analysis of running times.
Common order-of-growth classifications

**Good news.** The set of functions

\[ 1, \log N, \ N, \ N \log N, \ N^2, \ N^3, \text{ and } 2^N \]
suffices to describe the order of growth of most common 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>
</tbody>
</table>
| $\log N$        | logarithmic   | while $(N > 1)$
  
  { $N = N/2; \ ... \ }$ | divide in half | binary search | $\sim 1$                    |
| $N$             | linear        | for (int $i = 0; i < N; i++$)
  
  { $\ ... \ }$ | single loop | find the maximum | 2                           |
| $N \log N$      | linearithmic  | see mergesort lecture                         | divide and conquer | mergesort | $\sim 2$ |
| $N^2$           | quadratic     | for (int $i = 0; i < N; i++$)
  
  for (int $j = 0; j < N; j++$)
  
  { $\ ... \ }$ | double loop | check all pairs | 4                           |
| $N^3$           | cubic         | for (int $i = 0; i < N; i++$)
  
  for (int $j = 0; j < N; j++$)
  
  for (int $k = 0; k < N; k++$)
  
  { $\ ... \ }$ | triple loop | check all triples | 8                           |
| $2^N$           | exponential   | see combinatorial search lecture               | exhaustive search | check all subsets | $2^N$ |
# 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></td>
<td></td>
<td>time for 100x more data</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>N^2</td>
<td>quadratic</td>
<td>not practical for large problems</td>
<td>several hours</td>
</tr>
<tr>
<td>N^3</td>
<td>cubic</td>
<td>not practical for medium problems</td>
<td>several weeks</td>
</tr>
<tr>
<td>2^N</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.
**Binary search demo**

**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.

---

**successful 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
Binary search demo

**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.

successful search for 33

<table>
<thead>
<tr>
<th></th>
<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>lo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>mid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Binary search demo

**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.

---

**successful 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
Binary search demo

**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.

---

**successful search for 33**

```
   6   13   14   25   33   43   51   53   64   72   84   93   95   96   97
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14

lo = hi

mid
return 4
```
Binary search demo

**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.

unsuccessful search for 34
Binary search demo

**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.

**unsuccessful 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>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
Binary search demo

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.

unsuccessful 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>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
Binary search demo

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.

unsuccessful 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>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 = hi

mid

return -1
Binary search: implementation

Trivial to implement?

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

---

Extra, Extra - Read All About It: Nearly All Binary Searches and Mergesorts are Broken

Posted by Joshua Bloch, Software Engineer

I remember vividly Jon Bentley’s first Algorithms lecture at CMU, where he asked all of us incoming Ph.D. students to write a binary search, and then dissected one of our implementations in front of the class. Of course it was broken, as were most of our implementations. This made a real impression on me, as did the treatment of this material in his wonderful *Programming Pearls* (Addison-Wesley, 1986; Second Edition, 2000). The key lesson was to carefully consider the invariants in your programs.

http://googleresearch.blogspot.com/2006/06/extra-extra-read-all-about-it-nearly.html
Binary search: Java implementation

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

```java
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;
}
```

why not mid = (lo + hi) / 2 ?

one "3-way compare"
Proposition. Binary search uses at most $1 + \lg N$ key compares to search in a sorted array of size $N$.

Def. $T(N) = \#$ key compares to binary search a sorted subarray of size $\leq N$.

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

Pf sketch. [assume $N$ is a power of 2]

\[
T(N) \leq T(N/2) + 1 \quad \text{[ given ]}
\]
\[
\leq T(N/4) + 1 + 1 \quad \text{[ apply recurrence to first term ]}
\]
\[
\leq T(N/8) + 1 + 1 + 1 \quad \text{[ apply recurrence to first term ]}
\]
\[
\vdots
\]
\[
\leq T(N/N) + 1 + 1 + \ldots + 1 \quad \text{[ stop applying, $T(1) = 1$ ]}
\]
\[
= 1 + \lg N
\]
Comparing programs

**Hypothesis.** The sorting-based $N^2 \log N$ algorithm for 3-SUM 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*

<table>
<thead>
<tr>
<th>N</th>
<th>time (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000</td>
<td>0.14</td>
</tr>
<tr>
<td>2,000</td>
<td>0.18</td>
</tr>
<tr>
<td>4,000</td>
<td>0.34</td>
</tr>
<tr>
<td>8,000</td>
<td>0.96</td>
</tr>
<tr>
<td>16,000</td>
<td>3.67</td>
</tr>
<tr>
<td>32,000</td>
<td>14.88</td>
</tr>
<tr>
<td>64,000</td>
<td>59.16</td>
</tr>
</tbody>
</table>

*ThreeSumDeluxe.java*

**Guiding principle.** Typically, better order of growth $\Rightarrow$ faster in practice.
1.4 Analysis of Algorithms

- introduction
- observations
- mathematical models
- order-of-growth classifications
- theory of algorithms
- memory
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:_ \(~ \frac{1}{2} N^3\)

_average:_ \(~ \frac{1}{2} N^3\)

_worst:_ \(~ \frac{1}{2} N^3\)

**Ex 2.** Compares for binary search.

_best:_ \(~ 1\)

_average:_ \(~ \log N\)

_worst:_ \(~ \log N\)
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: focus on the worst case.

Upper bound. Performance guarantee of algorithm for any input.
Lower bound. Proof that no algorithm can do better.
Optimal algorithm. Lower bound = upper bound (to within a constant factor).
## Commonly-used notations in the theory of algorithms

<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><strong>Big Theta</strong></td>
<td>asymptotic order of growth</td>
<td>$\Theta(N^2)$</td>
<td>$\frac{1}{2}N^2$</td>
<td>classify algorithms</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$10N^2$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$5N^2 + 22N\log N + 3N$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\vdots$</td>
<td></td>
</tr>
<tr>
<td><strong>Big O</strong></td>
<td>$\Theta(N^2)$ and smaller</td>
<td>$O(N^2)$</td>
<td>$10N^2$</td>
<td>develop upper bounds</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$100N$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$22N\log N + 3N$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\vdots$</td>
<td></td>
</tr>
<tr>
<td><strong>Big Omega</strong></td>
<td>$\Theta(N^2)$ and larger</td>
<td>$\Omega(N^2)$</td>
<td>$\frac{1}{2}N^2$</td>
<td>develop lower bounds</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$N^5$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$N^3 + 22N\log N + 3N$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\vdots$</td>
<td></td>
</tr>
</tbody>
</table>
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 $O(N^3)$. 
Theory of algorithms: example 2

Goals.

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

Upper bound. A specific algorithm.

- Ex. Improved algorithm for 3-SUM.
- Running time of the optimal algorithm for 3-SUM is $O(N^2 \log N)$.

Lower bound. Proof that no algorithm can do better.

- Ex. Have to examine all $N$ entries to solve 3-SUM.
- Running time of the optimal algorithm for solving 3-SUM is $\Omega(N)$.

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.
### Commonly-used notations in the theory of algorithms

<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><strong>Tilde</strong></td>
<td>leading term</td>
<td>( \sim 10N^2 )</td>
<td>( 10N^2 ) ( 10N^2 + 22N\log N ) ( 10N^2 + 2N + 37 )</td>
<td>provide approximate model</td>
</tr>
<tr>
<td><strong>Big Theta</strong></td>
<td>asymptotic order of growth</td>
<td>( \Theta(N^2) )</td>
<td>( \frac{1}{2}N^2 ) ( 10N^2 ) ( 5N^2 + 22N\log N + 3N )</td>
<td>classify algorithms</td>
</tr>
<tr>
<td><strong>Big Oh</strong></td>
<td>( \Theta(N^2) ) and smaller</td>
<td>( O(N^2) )</td>
<td>( 10N^2 ) ( 100N ) ( 22N\log N + 3N )</td>
<td>develop upper bounds</td>
</tr>
<tr>
<td><strong>Big Omega</strong></td>
<td>( \Theta(N^2) ) and larger</td>
<td>( \Omega(N^2) )</td>
<td>( \frac{1}{2}N^2 ) ( N^5 ) ( N^3 + 22N\log N + 3N )</td>
<td>develop lower bounds</td>
</tr>
</tbody>
</table>

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

**This course.** Focus on approximate models: use Tilde-notation
1.4 Analysis of Algorithms

- introduction
- observations
- mathematical models
- order-of-growth classifications
- theory of algorithms
- 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.

64-bit machine. We assume a 64-bit machine with 8-byte pointers.

NIST most computer scientists

some JVMs "compress" ordinary object pointers to 4 bytes to avoid this cost
## Typical memory usage for primitive types and arrays

### Primitive types

<table>
<thead>
<tr>
<th>type</th>
<th>bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>boolean</td>
<td>1</td>
</tr>
<tr>
<td>byte</td>
<td>1</td>
</tr>
<tr>
<td>char</td>
<td>2</td>
</tr>
<tr>
<td>int</td>
<td>4</td>
</tr>
<tr>
<td>float</td>
<td>4</td>
</tr>
<tr>
<td>long</td>
<td>8</td>
</tr>
<tr>
<td>double</td>
<td>8</td>
</tr>
</tbody>
</table>

### One-dimensional arrays

<table>
<thead>
<tr>
<th>type</th>
<th>bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>char[]</td>
<td>$2N + 24$</td>
</tr>
<tr>
<td>int[]</td>
<td>$4N + 24$</td>
</tr>
<tr>
<td>double[]</td>
<td>$8N + 24$</td>
</tr>
</tbody>
</table>

### Two-dimensional arrays

<table>
<thead>
<tr>
<th>type</th>
<th>bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>char[][]</td>
<td>$\sim 2 M N$</td>
</tr>
<tr>
<td>int[][]</td>
<td>$\sim 4 M N$</td>
</tr>
<tr>
<td>double[][]</td>
<td>$\sim 8 M N$</td>
</tr>
</tbody>
</table>
Typical memory usage for objects in Java

**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.

```java
public class Date {
    private int day;
    private int month;
    private int year;
    ...
}
```

<table>
<thead>
<tr>
<th>Field</th>
<th>Size (bytes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>day</td>
<td>4</td>
</tr>
<tr>
<td>month</td>
<td>4</td>
</tr>
<tr>
<td>year</td>
<td>4</td>
</tr>
<tr>
<td>padding</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>32 bytes</strong></td>
</tr>
</tbody>
</table>

16 bytes (object overhead)
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.
- **Padding:** round up to multiple of 8 bytes.

+ 8 extra bytes per inner class object (for reference to enclosing class)

**Note.** Depending on application, we may want to count memory for any referenced objects (recursively).
Turning the crank: summary

Empirical analysis.
- Execute program to perform experiments.
- Assume power law.
- 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 behavior.

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.