Algorithms

ROBERT SEDGEWICK | KEVIN WAYNE

http://algs4.cs.princeton.edu

2.1 ELEMENTARY SORTS

- *rules of the game*
- *selection sort*
- *insertion sort*
- *shellsort*
- *shuffling*
- *convex hull*

Algorithms

ROBERT SEDGEWICK | KEVIN WAYNE

http://algs4.cs.princeton.edu

2.1 ELEMENTARY SORTS

- *rules of the game*
- *selection sort*
- *insertion sort*
- *shellsort*
- *shuffling*
- *convex hull*

Sorting problem

Ex. Twitter Dossier.

Name	YOB	Tweets	Followers	Most Recent Tweet Containing
Justin Bieber	1994	23,532	44,865,050	u saved my life ... @justinbieber
Kevin Shields	1963	21	6,165	Primal Scream (& Kevin Shields)
Kevin Barnes	1974	1,211	43,470	kevin barnes drank too much cough syrup
Kevin Mitnick	1963	9,380	103,952	Kevin Mitnick gives solution for NSA spying
Lil B	1989	115,746	678,610	GUYS LIL B COMMENTED ON MY PICTURE BYE
Josh Hug	1981	48	30	I need a josh hug
Carly Rae Jepsen	1981	5,040	8,874,941	carly rae jepsen stop trying already

item → Kevin Mitnick

key → Carly Rae Jepsen

Sort. Rearrange array of N items into ascending order.

Name	YOB	Tweets	Followers	Most Recent Tweet Containing
Carly Rae Jepsen	1981	5,040	8,874,941	carly rae jepsen stop trying already
Josh Hug	1981	48	30	I need a josh hug
Justin Bieber	1994	23,532	44,865,050	u saved my life ... @justinbieber
Kevin Barnes	1974	1,211	43,470	kevin barnes drank too much cough syrup
Kevin Mitnick	1963	9,380	103,952	Kevin Mitnick gives solution for NSA spying
Kevin Shields	1963	21	6,165	Primal Scream (& Kevin Shields)
Lil B	1989	115,746	678,610	GUYS LIL B COMMENTED ON MY PICTURE BYE

Inversions (a.k.a. Yodaness)

Def. An inversion is a pair of keys in an array that are out of order.

A E E L M O T R X P S

T-R T-P T-S R-P X-P X-S

(6 inversions out of 55 max)

Goal.

- Given an array with N inversions.
- Perform a sequence of operations that reduces inversions to 0.

Introduction à l'analyse des lignes courbes algébriques, Gabriel Cramer, 1750

Sample sort client 1

Goal. Sort any type of data.

Ex 1. Sort random real numbers in ascending order.

seems artificial, but stay tuned for an application

```
public class Experiment
{
   public static void main(String[] args)
   {
      int N = Integer.parseInt(args[0]);
      Double[] a = new Double[N];
      for (int i = 0; i < N; i++)
         a[i] = StdRandom.uniform();
      Insertion.sort(a);
      for (int i = 0; i < N; i++)
         StdOut.println(a[i]);
   }
}
```

```
% java Experiment 10
0.08614716385210452
0.09054270895414829
0.10708746304898642
0.21166190071646818
0.363292849257276
0.460954145685913
0.5340026311350087
0.7216129793703496
0.9003500354411443
0.9293994908845686
```


Sample sort client 2

Goal. Sort any type of data.

Ex 2. Sort strings from file in alphabetical order.

```
public class StringSorter
{
   public static void main(String[] args)
   {
      String[] a = In.readStrings(args[0]);
      Insertion.sort(a);
      for (int i = 0; i < a.length; i++)
         StdOut.println(a[i]);
   }
}
```

```
% more words3.txt
it's friday, friday gotta get down on friday

% java StringSorter words3.txt
down friday friday friday, get gotta it's on
```


Sample sort client 3

Goal. Sort any type of data.

Ex 3. Sort the files in a given directory by filename.

```
import java.io.File;
public class FileSorter
{
   public static void main(String[] args)
   {
      File directory = new File(args[0]);
      File[] files = directory.listFiles();
      Insertion.sort(files);
      for (int i = 0; i < files.length; i++)
         StdOut.println(files[i].getName());
   }
}
```

```
% java FileSorter .
Insertion.class
Insertion.java
InsertionX.class
InsertionX.java
Selection.class
Selection.java
Shell.class
Shell.java
ShellX.class
ShellX.java
```

```
public class File {
 private String path;
 private int prefixLength;
 ...
}
```


Callbacks

Goal. Sort any type of data.

Q. How can `sort()` know how to compare data of type `Double`, `String`, and `java.io.File` without any information about the type of an item's key?

Callback = reference to executable code.

- Client passes array of objects to `sort()` function.
- The `sort()` function calls back object's `compareTo()` method as needed.

Implementing callbacks.

- Java: interfaces.

Callbacks: roadmap

client

```
import java.io.File;
public class FileSorter
{
   public static void main(String[] args)
   {
      File directory = new File(args[0]);
      File[] files = directory.listFiles();
      Insertion.sort(files);
      for (int i = 0; i < files.length; i++)
         StdOut.println(files[i].getName());
   }
}
```

object implementation

```
public class File
implements Comparable<File>
{
   ...
   public int compareTo(File b)
   {
      ...
      return -1;
      ...
      return +1;
      ...
      return 0;
   }
}
```

Comparable interface (built in to Java)

```
public interface Comparable<Item>
{
   public int compareTo(Item that);
}
```

Insertion sort implementation

```
public static void sort(Comparable[] a)
{
   int N = a.length;
   for (int i = 0; i < N; i++)
      for (int j = i; j > 0; j--)
         if (a[j].compareTo(a[j-1]) < 0)
            exch(a, j, j-1);
         else break;
}
```

key point: no dependence on `File` data type

Callbacks: roadmap

client

```
import java.io.File;
public class FileSorter
{
   public static void main(String[] args)
   {
      File directory = new File(args[0]);
      File[] files = directory.listFiles();
      Insertion.sort(files);
      for (int i = 0; i < files.length; i++)
         StdOut.println(files[i].getName());
   }
}
```

object implementation

```
public class File
implements Comparable<File>
{
   ...
   public int compareTo(File b)
   {
      ...
      return -1;
      ...
      return +1;
      ...
      return 0;
   }
}
```

pollEv.com/jhug text to **37607**

Q: If we omit "compareTo()", which file will fail to compile?

A. FileSorter.java [778751]
B. File.java [778752]
C. InsertionSort.java [778753]

Insertion sort implementation

```
public static void sort(Comparable[] a)
{
   int N = a.length;
   for (int i = 0; i < N; i++)
      for (int j = i; j > 0; j--)
         if (a[j].compareTo(a[j-1]) < 0)
            exch(a, j, j-1);
         else break;
}
```


Callbacks: roadmap

client

```
import java.io.File;
public class FileSorter
{
   public static void main(String[] args)
   {
      File directory = new File(args[0]);
      File[] files = directory.listFiles();
      Insertion.sort(files);
      for (int i = 0; i < files.length; i++)
         StdOut.println(files[i].getName());
   }
}
```

pollEv.com/jhug text to **37607**

Q: If we omit "`implements Comparable`", which file will fail to compile?

A. FileSorter.java [778757]
B. File.java [778758]
C. InsertionSort.java [778759]

object implementation

```
public class File
implements Comparable<File>
{
   ...
   public int compareTo(File b)
   {
      ...
      return -1;
      ...
      return +1;
      ...
      return 0;
   }
}
```

Insertion sort implementation

```
public static void sort(Comparable[] a)
{
   int N = a.length;
   for (int i = 0; i < N; i++)
      for (int j = i; j > 0; j--)
         if (a[j].compareTo(a[j-1]) < 0)
            exch(a, j, j-1);
         else break;
}
```


Comparable API

Implement `compareTo()` so that `v.compareTo(w)`

- Is a total order.
- Returns a negative integer, zero, or positive integer if v is less than, equal to, or greater than w, respectively.
- Throws an exception if incompatible types (or either is `null`).

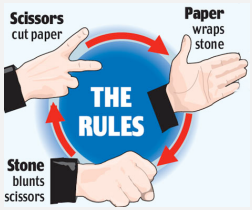
Total order

A total order is a binary relation $\leq$ that satisfies:

- Antisymmetry: if $v \leq w$ and $w \leq v$, then $v = w$.
- Transitivity: if $v \leq w$ and $w \leq x$, then $v \leq x$.
- Totality: either $v \leq w$ or $w \leq v$ or both.

Ex.

- Standard order for natural and real numbers.
- Chronological order for dates or times.
- Alphabetical order for strings.
- ...



an intransitive relation

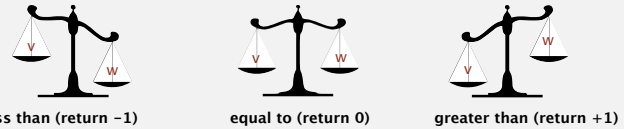
violates totality: (Double.NaN <= Double.NaN) is false

Surprising but true. The <= operator for double is not a total order. (!)

Comparable API

Implement compareTo() so that v.compareTo(w)

- Is a total order.
- Returns a negative integer, zero, or positive integer if v is less than, equal to, or greater than w, respectively.
- Throws an exception if incompatible types (or either is null).



less than (return -1) equal to (return 0) greater than (return +1)

Built-in comparable types. Integer, Double, String, Date, File, ...

User-defined comparable types. Implement the Comparable interface.

Implementing the Comparable interface

Date data type. Simplified version of java.util.Date.

```
public class Date implements Comparable<Date>
{
   private final int month, day, year;

   public Date(int m, int d, int y)
   {
      month = m;
      day   = d;
      year  = y;
   }

   public int compareTo(Date that)
   {
      if (this.year  < that.year ) return -1;
      if (this.year  > that.year ) return +1;
      if (this.month < that.month) return -1;
      if (this.month > that.month) return +1;
      if (this.day   < that.day  ) return -1;
      if (this.day   > that.day  ) return +1;
      return 0;
   }
}
```

only compare dates to other dates

Callbacks

Goal. Sort any type of data.

Q. How can sort() know how to compare data of type Double, String, and java.io.File without any information about the type of an item's key?

Callback = reference to executable code.

- Client passes array of objects to sort() function.
- The sort() function calls back object's compareTo() method as needed.

Implementing callbacks.

- Java: interfaces.
- C: function pointers.
- C++: class-type functors.
- C#: delegates.
- Python, Perl, ML, Javascript: first-class functions.

Summary.

Generic Sorting.

- Generic sorting algorithm expects array of Comparables
- Comparable: Class implements .compareTo() method
  - Must contain compareTo() method to compile
  - compareTo() should obey certain rules to guarantee function

Today's Sorting Algorithms.

- Will only interact with the `Comparable` array via helper functions
  - `exch(i ,j)`: swaps items at position `i` and `j`
  - `less(v, w)`: returns `true` if `v.compareTo(w) < 0`

Two useful sorting abstractions

Helper functions. Refer to data through compares and exchanges.

Less. Is item `v` less than `w` ?

```
private static boolean less(Comparable v, Comparable w)
{  return v.compareTo(w) < 0;  }
```

Exchange. Swap item in array `a[]` at index `i` with the one at index `j`.

- Why exchange?

```
private static void exch(Comparable[] a, int i, int j)
{
   Comparable swap = a[i];
   a[i] = a[j];
   a[j] = swap;
}
```


Testing

Goal. Test if an array is sorted.

```
private static boolean isSorted(Comparable[] a)
{
   for (int i = 1; i < a.length; i++)
      if (less(a[i], a[i-1])) return false;
   return true;
}
```

a[] ⟶ `TroubleSort.sort(a)`

a[] ⟶ `isSorted(a)` ⟶ true

pollEv.com/jhug text to **37607**

Q. If the sorting algorithm passes the test, did it correctly sort the array?

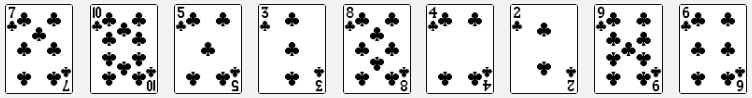
A. Yes [778645]
B. No [778646]

2.1 Elementary Sorts

- *rules of the game*
- *selection sort*
- *insertion sort*
- *shellsort*
- *shuffling*
- *convex hull*

Selection sort demo

- In iteration `i`, find index `min` of smallest remaining entry.
- Swap `a[i]` and `a[min]`.



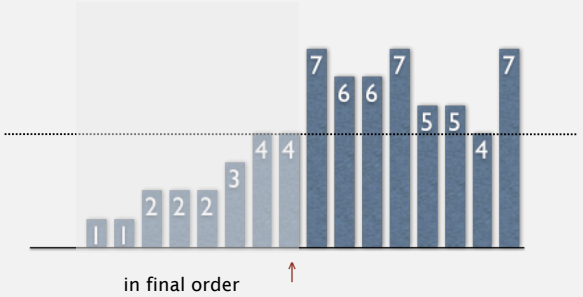
initial

Selection sort

Algorithm. ↑ scans from left to right.

Invariants.

- Entries the left of ↑ (including ↑) fixed and in ascending order.
- No entry to right of ↑ is smaller than any entry to the left of ↑.



in final order

Selection sort inner loop

To maintain algorithm invariants:

- Move the pointer to the right.

```
i++;
```

- Identify index of minimum entry on right.

```
int min = i;
for (int j = i+1; j < N; j++)
   if (less(a[j], a[min]))
      min = j;
```

- Exchange into position.

```
exch(a, i, min);
```

Selection sort: Java implementation

```
public class Selection
{
   public static void sort(Comparable[] a)
   {
      int N = a.length;
      for (int i = 0; i < N; i++)
      {
         int min = i;
         for (int j = i+1; j < N; j++)
            if (less(a[j], a[min]))
               min = j;
         exch(a, i, min);
      }
   }

   private static boolean less(Comparable v, Comparable w)
   {  /* as before */  }

   private static void exch(Comparable[] a, int i, int j)
   {  /* as before */  }
}
```

Selection sort: mathematical analysis

Proposition. Selection sort uses $(N-1)+(N-2)+\ \ldots\ +1+0\ \sim\ N^2/2$ compares and N exchanges.

i	min	a[] 0	1	2	3	4	5	6	7	8	9	10
		S	O	R	T	E	X	A	M	P	L	E
0	6	S	O	R	T	E	X	A	M	P	L	E
1	4	A	O	R	T	E	X	S	M	P	L	E
2	10	A	E	R	T	O	X	S	M	P	L	E
3	9	A	E	E	T	O	X	S	M	P	L	R
4	7	A	E	E	L	O	X	S	M	P	T	R
5	7	A	E	E	L	M	X	S	O	P	T	R
6	8	A	E	E	L	M	O	S	X	P	T	R
7	10	A	E	E	L	M	O	P	X	S	T	R
8	8	A	E	E	L	M	O	P	R	S	T	X
9	9	A	E	E	L	M	O	P	R	S	T	X
10	10	A	E	E	L	M	O	P	R	S	T	X
		A	E	E	L	M	O	P	R	S	T	X

entries in black are examined to find the minimum

entries in red are a[min]

entries in gray are in final position

Trace of selection sort (array contents just after each exchange)

Running time insensitive to input. Quadratic time, even if input is sorted.

Data movement is minimal. Linear number of exchanges.

Selection sort: mathematical analysis

Proposition. Selection sort uses $(N-1)+(N-2)+\ \ldots\ +1+0\ \sim\ N^2/2$ compares and N exchanges.

i	min	a[] 0	1	2	3	4	5	6	7	8	9	10
		S	O	R	T	E	X	A	M	P	L	E
0	6	S	O	R	T	E	X	A	M	P	L	E
1	4	A	O	R	T	E	X	S	M	P	L	E
2	10	A	E	R	T	O	X	S	M	P	L	E
3	9	A	E	E	T	O	X	S	M	P	L	R
4	7	A	E	E	L	O	X	S	M	P	T	R
5	7	A	E	E	L	M	X	S	O	P	T	R
6	8	A	E	E	L	M	O	S	X	P	T	R
7	10	A	E	E	L	M	O	P	X	S	T	R
8	8	A	E	E	L	M	O	P	R	S	T	X
9	9	A	E	E	L	M	O	P	R	S	T	X
10	10	A	E	E	L	M	O	P	R	S	T	X
		A	E	E	L	M	O	P	R	S	T	X

entries in black are examined to find the minimum

entries in red are a[min]

entries in gray are in final position

Selection Sort (Compares).

Best Average Worst

~N²/2 ~N²/2 ~N²/2

Selection Sort (Exchanges).

Best Average Worst

N N N

Selection sort: mathematical analysis

pollEv.com/jhug

text to 37607

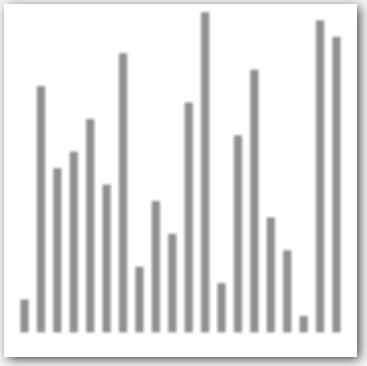
Q. Let R be the number of un-fixed elements before a call to exch. By how much is the inversion count reduced on that call to exch?

A. 1 [779846]
B. N [779847]
C. R [779848]
D. Between 0 and R [779849]
E. Between 0 and N [779850]

```
exch(a, i, min);
```

Selection sort: animations

20 random items



algorithm position

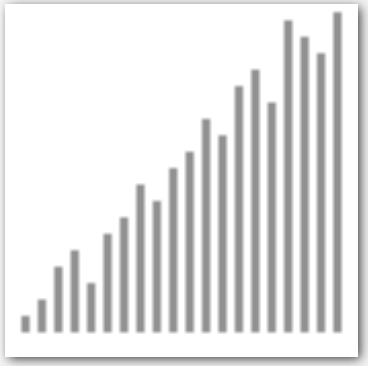
in final order

not in final order

http://www.sorting-algorithms.com/selection-sort

Selection sort: animations

20 partially-sorted items



algorithm position
in final order
not in final order

http://www.sorting-algorithms.com/selection-sort

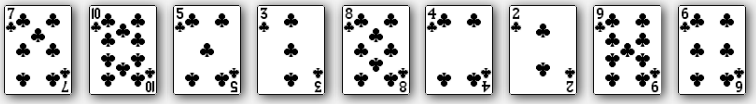
Robert Sedgewick | Kevin Wayne
http://algs4.cs.princeton.edu

2.1 Elementary Sorts

- *rules of the game*
- *selection sort*
- *insertion sort*
- *shellsort*
- *shuffling*
- *convex hull*

Insertion sort demo

- In iteration `i`, swap `a[i]` with each larger entry to its left.

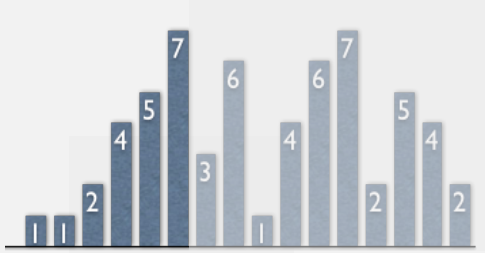


Insertion sort

Algorithm. ↑ scans from left to right.

Invariants.

- Entries to the left of ↑ (including ↑) are in ascending order.
- Entries to the right of ↑ have not yet been seen.



Insertion sort inner loop

To maintain algorithm invariants:

- Move the pointer to the right.

```
i++;
```

- Moving from right to left, exchange `a[i]` with each larger entry to its left.

```
for (int j = i; j > 0; j--)
   if (less(a[j], a[j-1]))
        exch(a, j, j-1);
   else break;
```

Insertion sort: Java implementation

```
public class Insertion
{
   public static void sort(Comparable[] a)
   {
      int N = a.length;
      for (int i = 0; i < N; i++)
         for (int j = i; j > 0; j--)
            if (less(a[j], a[j-1]))
               exch(a, j, j-1);
            else break;
   }
 ...
}
```

pollEv.com/jhug text to **37607**

Q. What is the worst case number of compares to complete an entire insertion sort?

A. ~N [780115] D. ~N² [780118]
B. ~N/2 [780116] E. ~N²/2 [780119]
C. ~N/4 [780117] F. ~N²/4 [780120]

Insertion sort: Java implementation

```
public class Insertion
{
   public static void sort(Comparable[] a)
   {
      int N = a.length;
      for (int i = 0; i < N; i++)
         for (int j = i; j > 0; j--)
            if (less(a[j], a[j-1]))
               exch(a, j, j-1);
            else break;
   }
 ...
}
```

Q. What is the best case number of compares to complete an entire insertion sort? Exchanges?

Insertion sort: Java implementation

```
public class Insertion
{
   public static void sort(Comparable[] a)
   {
      int N = a.length;
      for (int i = 0; i < N; i++)
         for (int j = i; j > 0; j--)
            if (less(a[j], a[j-1]))
               exch(a, j, j-1);
            else break;
   }
 ...
}
```

Q. What is the best case number of compares to complete an entire insertion sort? Exchanges?

Insertion Sort (Compares).

Insertion Sort (Exchanges).

Insertion sort: mathematical analysis

Proposition. To sort a randomly-ordered array with distinct keys, insertion sort uses $\sim \frac{1}{4} N^2$ compares and $\sim \frac{1}{4} N^2$ exchanges on average.

Pf. Expect each entry to move halfway back.

		a[]										
i	j	0	1	2	3	4	5	6	7	8	9	10
		S	O	R	T	E	X	A	M	P	L	E
1	0	O	S	R	T	E	X	A	M	P	L	E
2	1	O	R	S	T	E	X	A	M	P	L	E
3	3	O	R	S	T	E	X	A	M	P	L	E
4	0	E	O	R	S	T	X	A	M	P	L	E
5	5	E	O	R	S	T	X	A	M	P	L	E
6	0	A	E	O	R	S	T	X	M	P	L	E
7	2	A	E	M	O	R	S	T	X	P	L	E
8	4	A	E	M	O	P	R	S	T	X	L	E
9	2	A	E	L	M	O	P	R	S	T	X	E
10	2	A	E	E	L	M	O	P	R	S	T	X
		A	E	E	L	M	O	P	R	S	T	X

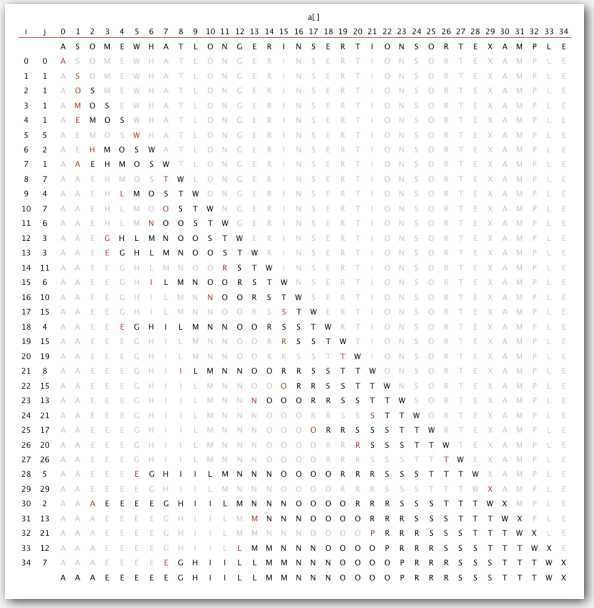
entries in gray do not move

entry in red is a[j]

entries in black moved one position right for insertion

Trace of insertion sort (array contents just after each insertion)

Insertion sort: trace



Insertion sort vs. selection sort

Selection Sort (Compares).

Best	Average	Worst
~N²/2	~N²/2	~N²/2

Selection Sort (Exchanges).

Best	Average	Worst
N	N	N

Insertion Sort (Compares).

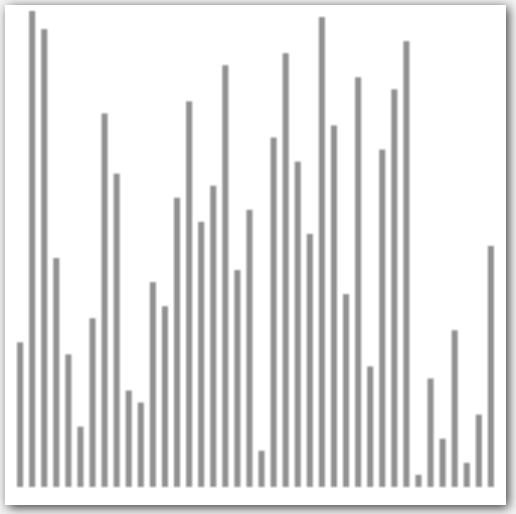
Best	Average	Worst
N−1	~N²/4	~N²/2

Insertion Sort (Exchanges).

Best	Average	Worst
0	~N²/4	~N²/2

Insertion sort: animation

40 random items

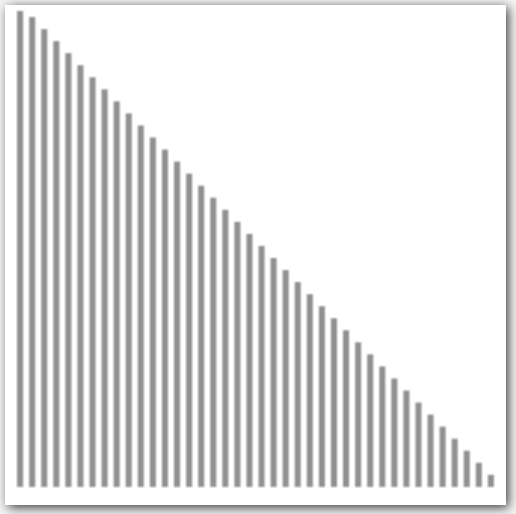


algorithm position
in order
not yet seen

http://www.sorting-algorithms.com/insertion-sort

Insertion sort: animation

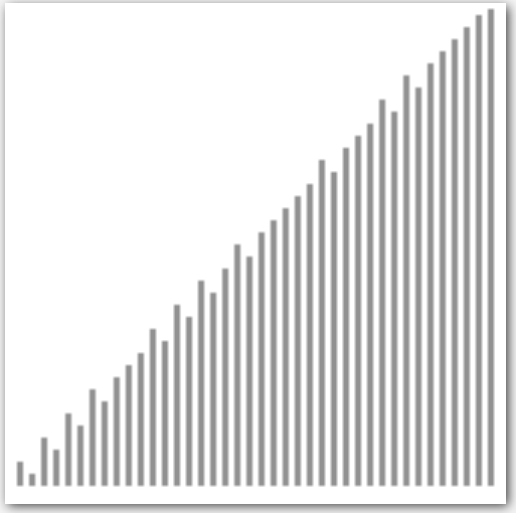
40 reverse-sorted items



http://www.sorting-algorithms.com/insertion-sort

Insertion sort: animation

40 partially-sorted items



http://www.sorting-algorithms.com/insertion-sort

Insertion sort: Java implementation

```
public class Insertion
{
   public static void sort(Comparable[] a)
   {
      int N = a.length;
      for (int i = 0; i < N; i++)
         for (int j = i; j > 0; j--)
            if (less(a[j], a[j-1]))
               exch(a, j, j-1);
            else break;
   }


}
```

Q1. What happens to the inversion on each call to exch?

Q2. Given an array with inversion count C, how many calls to exch will be made total before sorting is complete?

Insertion sort: partially-sorted arrays

Def. An inversion is a pair of keys that are out of order.

A E E L M O T R X P S

T-R T-P T-S R-P X-P X-S

(6 inversions)

Def. An array is partially sorted if the number of inversions is $\leq c\,N$.

- Ex 1. A subarray of size 10 appended to a sorted subarray of size N.
- Ex 2. An array of size N with only 10 entries out of place.

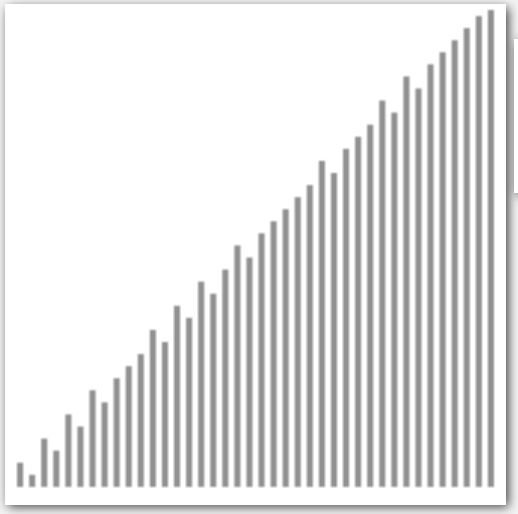
Proposition. For partially-sorted arrays, insertion sort runs in linear time.

Pf. Number of exchanges equals the number of inversions.

number of compares = exchanges + (N - 1)

Insertion sort: animation

40 partially-sorted items



Insertion Sort (Exchanges).

Best	PS	Average	Worst
0	Θ(N)	~N²/4	~N²/2

PS: Partially sorted

▲ algorithm position
in order
not yet seen

http://www.sorting-algorithms.com/insertion-sort

Algorithms

Robert Sedgewick | Kevin Wayne

http://algs4.cs.princeton.edu

2.1 Elementary Sorts

- *rules of the game*
- *selection sort*
- *insertion sort*
- *shellsort [see lecture/slides online]*
- *shuffling*
- *convex hull*

Shellsort overview

Idea. Move entries more than one position at a time by *h*-sorting the array.

an h-sorted array is h interleaved sorted subsequences

h = 4

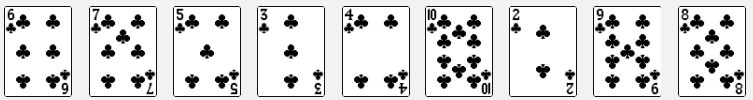
L	E	E	A	M	H	L	E	P	S	O	L	T	S	X	R
L				M				P				T			
	E				H				S				S		
		E				L				O				X	
			A				E				L				R

Shellsort. [Shell 1959] *h*-sort array for decreasing sequence of values of *h*.

input	S	H	E	L	L	S	O	R	T	E	X	A	M	P	L	E
13-sort	P	H	E	L	L	S	O	R	T	E	X	A	M	S	L	E
4-sort	L	E	E	A	M	H	L	E	P	S	O	L	T	S	X	R
1-sort	A	E	E	E	H	L	L	L	M	O	P	R	S	S	T	X

h-sorting demo

In iteration `i`, swap `a[i]` with each larger entry `h` positions to its left.



h-sorting

How to h-sort an array? Insertion sort, with stride length h.

3-sorting an array

M	O	L	E	E	X	A	S	P	R	T
E	O	L	M	E	X	A	S	P	R	T
E	E	L	M	O	X	A	S	P	R	T
E	E	L	M	O	X	A	S	P	R	T
A	E	L	E	O	X	M	S	P	R	T
A	E	L	E	O	X	M	S	P	R	T
A	E	L	E	O	P	M	S	X	R	T
A	E	L	E	O	P	M	S	X	R	T
A	E	L	E	O	P	M	S	X	R	T
A	E	L	E	O	P	M	S	X	R	T

Why insertion sort?

- Big increments ⇒ small subarray.
- Small increments ⇒ nearly in order. [stay tuned]

Shellsort example: increments 7, 3, 1

input

S	O	R	T	E	X	A	M	P	L	E

7-sort

S	O	R	T	E	X	A	M	P	L	E
M	O	R	T	E	X	A	S	P	L	E
M	O	R	T	E	X	A	S	P	L	E
M	O	L	T	E	X	A	S	P	R	E
M	O	L	E	E	X	A	S	P	R	T

3-sort

M	O	L	E	E	X	A	S	P	R	T
E	O	L	M	E	X	A	S	P	R	T
E	E	L	M	O	X	A	S	P	R	T
E	E	L	M	O	X	A	S	P	R	T
A	E	L	E	O	X	M	S	P	R	T
A	E	L	E	O	X	M	S	P	R	T
A	E	L	E	O	P	M	S	X	R	T
A	E	L	E	O	P	M	S	X	R	T
A	E	L	E	O	P	M	S	X	R	T

1-sort

A	E	L	E	O	P	M	S	X	R	T
A	E	L	E	O	P	M	S	X	R	T
A	E	L	E	O	P	M	S	X	R	T
A	E	E	L	O	P	M	S	X	R	T
A	E	E	L	O	P	M	S	X	R	T
A	E	E	L	M	O	P	S	X	R	T
A	E	E	L	M	O	P	S	X	R	T
A	E	E	L	M	O	P	S	X	R	T
A	E	E	L	M	O	P	R	S	X	T
A	E	E	L	M	O	P	R	S	T	X

result

A	E	E	L	M	O	P	R	S	T	X

Shellsort: Java implementation

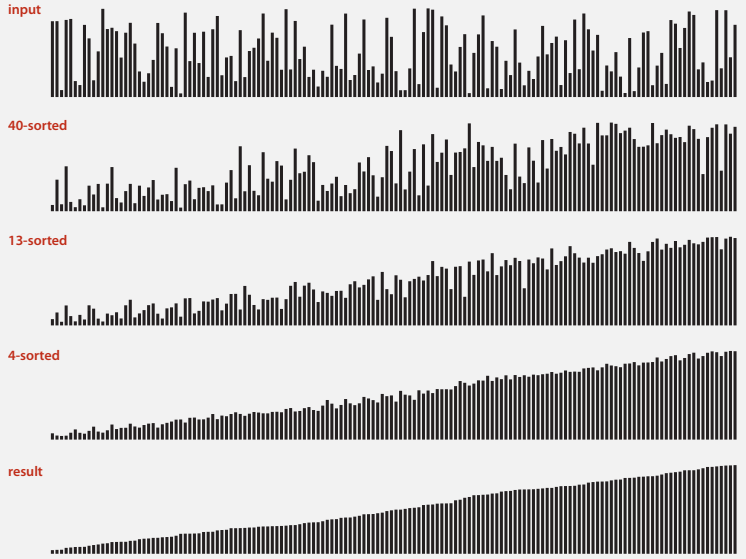
```
public class Shell
{
   public static void sort(Comparable[] a)
   {
      int N = a.length;

      int h = 1;
      while (h < N/3) h = 3*h + 1; // 1, 4, 13, 40, 121, 364, ...

      while (h >= 1)
      {  // h-sort the array.
         for (int i = h; i < N; i++)
         {
            for (int j = i; j >= h && less(a[j], a[j-h]); j -= h)
               exch(a, j, j-h);
         }

         h = h/3;
      }
   }

   private static boolean less(Comparable v, Comparable w)
   { /* as before */ }
   private static boolean exch(Comparable[] a, int i, int j)
   { /* as before */ }
}
```

3x+1 increment sequence

insertion sort

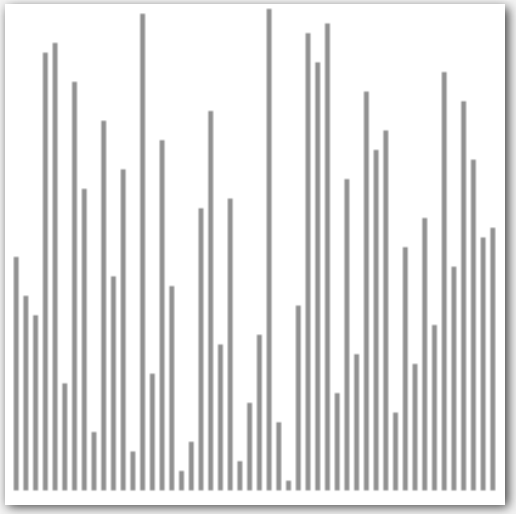
move to next increment

Shellsort: visual trace



Shellsort: animation

50 random items

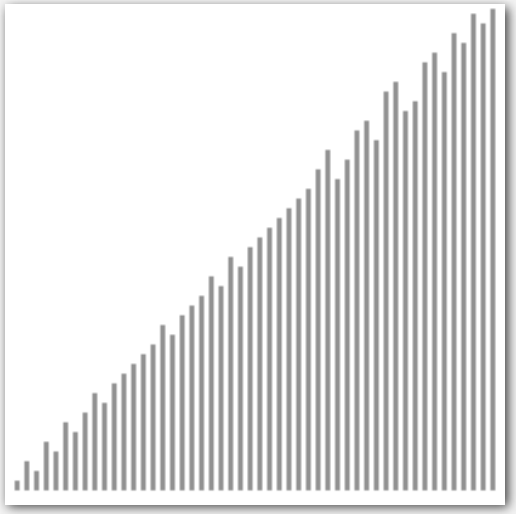


algorithm position
h-sorted
current subsequence
other elements

http://www.sorting-algorithms.com/shell-sort

Shellsort: animation

50 partially-sorted items



algorithm position
h-sorted
current subsequence
other elements

http://www.sorting-algorithms.com/shell-sort

Shellsort: which increment sequence to use?

Powers of two. 1, 2, 4, 8, 16, 32, ...
No.

Powers of two minus one. 1, 3, 7, 15, 31, 63, ...
Maybe.

→ 3x + 1. 1, 4, 13, 40, 121, 364, ...
OK. Easy to compute.

Sedgewick. 1, 5, 19, 41, 109, 209, 505, 929, 2161, 3905, ...
Good. Tough to beat in empirical studies.

merging of $(9 \times 4^i) - (9 \times 2^i) + 1$ and $4^i - (3 \times 2^i) + 1$

Shellsort: intuition

Proposition. A g-sorted array remains g-sorted after h-sorting it.

7-sort

S	O	R	T	E	X	A	M	P	L	E
M	O	R	T	E	X	A	S	P	L	E
M	O	R	T	E	X	A	S	P	L	E
M	O	L	T	E	X	A	S	P	R	E
M	O	L	E	E	X	A	S	P	R	T

3-sort

M	O	L	E	E	X	A	S	P	R	T
E	O	L	M	E	X	A	S	P	R	T
E	E	L	M	O	X	A	S	P	R	T
E	E	L	M	O	X	A	S	P	R	T
A	E	L	E	O	X	M	S	P	R	T
A	E	L	E	O	X	M	S	P	R	T
A	E	L	E	O	P	M	S	X	R	T
A	E	L	E	O	P	M	S	X	R	T
A	E	L	E	O	P	M	S	X	R	T
A	E	L	E	O	P	M	S	X	R	T

still 7-sorted

Challenge. Prove this fact—it's more subtle than you'd think!

Shellsort: analysis

Proposition. The worst-case number of compares used by shellsort with the 3x+1 increments is $O(N^{3/2})$.

Property. Number of compares used by shellsort with the 3x+1 increments is at most by a small multiple of N times the # of increments used.

N	compares	$N^{1.289}$	2.5 N lg N
5,000	93	58	106
10,000	209	143	230
20,000	467	349	495
40,000	1022	855	1059
80,000	2266	2089	2257

measured in thousands

Remark. Accurate model has not yet been discovered (!)

Why are we interested in shellsort?

Example of simple idea leading to substantial performance gains.

Useful in practice.

bzip2, /linux/kernel/groups.c

- Fast unless array size is huge (used for small subarrays).
- Tiny, fixed footprint for code (used in some embedded systems).
- Hardware sort prototype.

uClibc

Simple algorithm, nontrivial performance, interesting questions.

- Asymptotic growth rate?
- Best sequence of increments? ← open problem: find a better increment sequence
- Average-case performance?

Lesson. Some good algorithms are still waiting discovery.

Summary.

Sorting Techniques.

- Today's sorts:
  - Selection Sort: Order of growth: N^2.
  - Insertion Sort: N^2.
  - Shell Sort: $N^{3/2}$.
- Next week: $N \lg N$ sorts.
  - Merge sort.
  - Quick sort.
  - Heap sort.
- Novelty sorts:
  - Bogo sort: $N\,N!$ (average case). Never completes (worst case).
  - Gnome sort: N^2.

Robert Sedgewick | Kevin Wayne

http://algs4.cs.princeton.edu

2.1 Elementary Sorts

- *rules of the game*
- *selection sort*
- *insertion sort*
- *shellsort*
- *shuffling*
- *convex hull*

Robert Sedgewick | Kevin Wayne
http://algs4.cs.princeton.edu

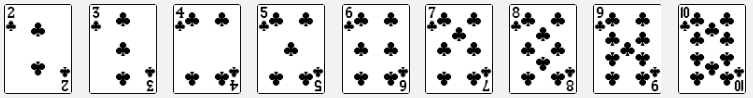
2.1 Elementary Sorts

- *rules of the game*
- *selection sort*
- *insertion sort*
- *shellsort*
- *shuffling*
- *convex hull*

How to shuffle an array

Goal. Rearrange array so that result is a uniformly random permutation.

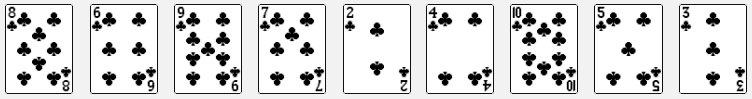
Uniformly Random Permutation. All permutations equally likely.



How to shuffle an array

Goal. Rearrange array so that result is a uniformly random permutation.

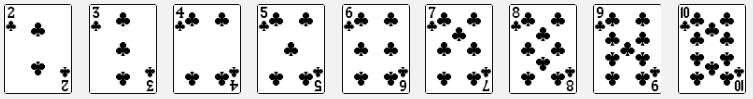
Uniformly Random Permutation. All permutations equally likely.



Shuffle sort

- Generate a random number for each array entry.
- Sort the array.

useful for shuffling columns in a spreadsheet

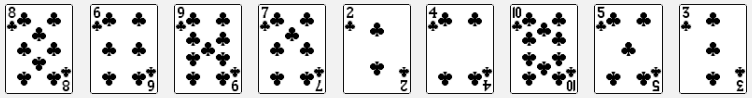


0.8003 0.9706 0.9157 0.9649 0.1576 0.4854 0.1419 0.4218 0.9572

Shuffle sort

- Generate a random number for each array entry.
- Sort the array.

useful for shuffling columns in a spreadsheet

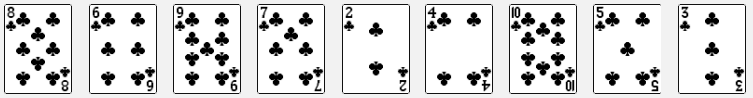


0.1419 0.1576 0.4218 0.4854 0.8003 0.9157 0.9572 0.9649 0.9706

Shuffle sort

- Generate a random number for each array entry.
- Sort the array.

useful for shuffling columns in a spreadsheet



0.1419 0.1576 0.4218 0.4854 0.8003 0.9157 0.9572 0.9649 0.9706

Proposition. Shuffle sort produces a uniformly random permutation of the input array, provided no duplicate numbers.

assuming numbers uniformly at random

Problem with Duplicates. Identical items aren't randomly distributed.

War story (Microsoft)

Microsoft antitrust probe by EU. Microsoft agreed to provide a randomized ballot screen for users to select browser in Windows 7.

http://www.browserchoice.eu

appeared last 50% of the time

War story (Microsoft)

Microsoft antitrust probe by EU. Microsoft agreed to provide a randomized ballot screen for users to select browser in Windows 7.

Solution? Implement shuffle sort by making comparator always return a random answer.

```
public int compareTo(Browser that)
{
   double r = Math.random();
   if (r < 0.5) return -1;
   if (r > 0.5) return +1;
   return 0;
}
```

browser comparator (should implement a total order)

War story (Microsoft)

Microsoft antitrust probe by EU. Microsoft agreed to provide a randomized ballot screen for users to select browser in Windows 7.

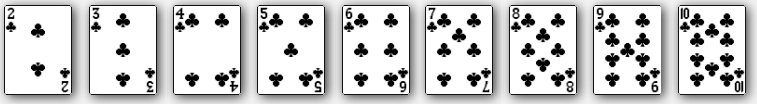
Programming Error... or was it?

http://www.browserchoice.eu

appeared last 50% of the time

Knuth shuffle demo

- In iteration `i`, pick integer `r` between `0` and `i` uniformly at random.
- Swap `a[i]` and `a[r]`.



Proposition. [Fisher-Yates 1938] Knuth shuffling produces a uniformly random permutation of the input array in linear time.

Knuth shuffle

- In iteration `i`, pick integer `r` between `0` and `i` uniformly at random.
- Swap `a[i]` and `a[r]`.

common bug: between 0 and N – 1
correct variant: between i and N – 1

```
public class StdRandom
{
   ...
   public static void shuffle(Object[] a)
   {
      int N = a.length;
      for (int i = 0; i < N; i++)
      {
         int r = StdRandom.uniform(i + 1);
         exch(a, i, r);
      }
   }
}
```

between 0 and i

War story (online poker)

Texas hold'em poker. Software must shuffle electronic cards.

How We Learned to Cheat at Online Poker: A Study in Software Security
http://www.datamation.com/entdev/article.php/616221

War story (online poker)

Shuffling algorithm in FAQ at www.planetpoker.com

```
for i := 1 to 52 do begin
   r := random(51) + 1;
   swap := card[r];
   card[r] := card[i];
   card[i] := swap;
end;
```

← between 1 and 51

Bug 1. Random number `r` never 52 ⇒ 52nd card can't end up in 52nd place.

Bug 2. Shuffle not uniform (should be between 1 and i).

Bug 3. `random()` uses 32-bit seed ⇒ 2^{32} possible shuffles.

Bug 4. Seed = milliseconds since midnight ⇒ 86.4 million shuffles.

> *" The generation of random numbers is too important to be left to chance. "*
>
> *— Robert R. Coveyou*

War story (online poker)

Best practices for shuffling (if your business depends on it).

- Use a hardware random-number generator that has passed both the FIPS 140-2 and the NIST statistical test suites.
- Continuously monitor statistic properties: hardware random-number generators are fragile and fail silently.
- Use an unbiased shuffling algorithm.

Bottom line. Shuffling a deck of cards is hard!

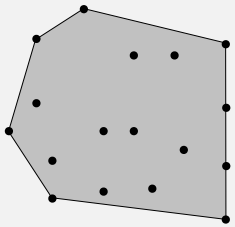
Robert Sedgewick | Kevin Wayne
http://algs4.cs.princeton.edu

2.1 Elementary Sorts

- *rules of the game*
- *selection sort*
- *insertion sort*
- *shellsort*
- *shuffling*
- *convex hull*

Convex hull

The convex hull of a set of N points is the smallest perimeter fence enclosing the points.

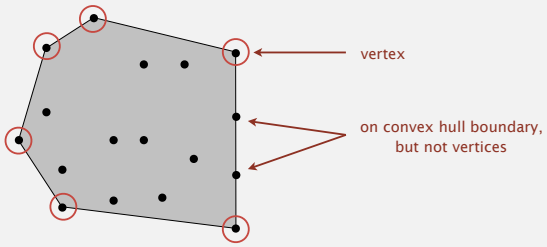


Equivalent definitions.

- Smallest convex set containing all the points.
- Smallest area convex polygon enclosing the points.
- Convex polygon enclosing the points, whose vertices are points in set.

Convex hull

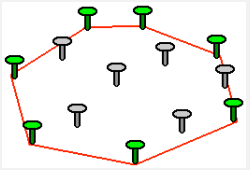
The convex hull of a set of N points is the smallest perimeter fence enclosing the points.



Convex hull output. Sequence of vertices in counterclockwise order.

Convex hull: mechanical algorithm

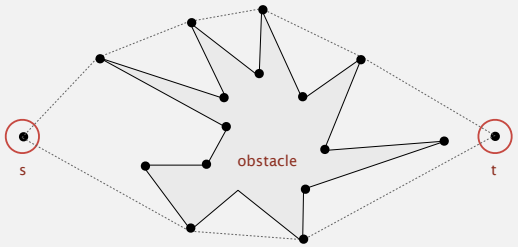
Mechanical algorithm. Hammer nails perpendicular to plane; stretch elastic rubber band around points.



http://www.idlcoyote.com/math_tips/convexhull.html

Convex hull application: motion planning

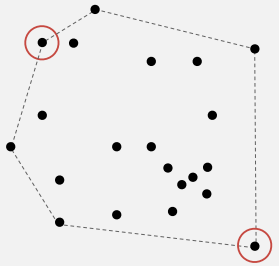
Robot motion planning. Find shortest path in the plane from s to t that avoids a polygonal obstacle.



Fact. Shortest path is either straight line from s to t or it is one of two polygonal chains of convex hull.

Convex hull application: farthest pair

Farthest pair problem. Given N points in the plane, find a pair of points with the largest Euclidean distance between them.

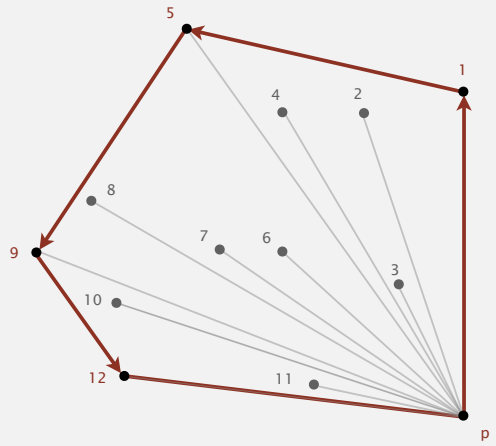


Fact. Farthest pair of points are extreme points on convex hull.

Convex hull: geometric properties

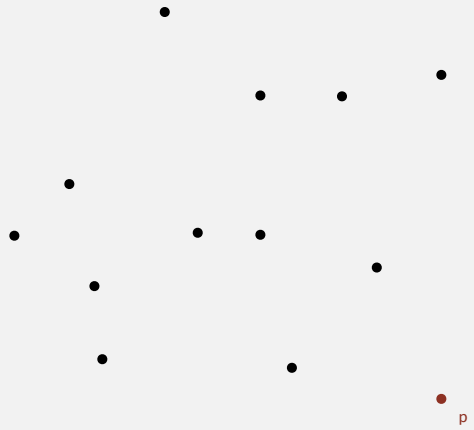
Fact. Can traverse the convex hull by making only counterclockwise turns.

Fact. The vertices of convex hull appear in increasing order of polar angle with respect to point p with lowest y-coordinate.



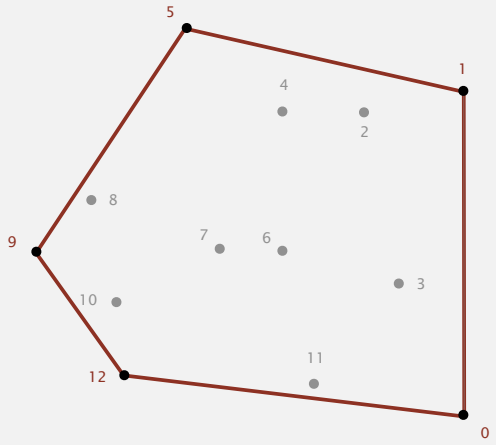
Graham scan demo

- Choose point p with smallest y-coordinate.
- Sort points by polar angle with p.
- In sorted order: Add point, then discard old points until all turns are ccw.



Graham scan demo

- Choose point p with smallest y-coordinate.
- Sort points by polar angle with p.
- In sorted order: Add point, then discard old points until all turns are ccw.



Graham scan: implementation challenges

Q. How to find point p with smallest y-coordinate?
A. Define a total order, comparing by y-coordinate. [next lecture]

Q. How to sort points by polar angle with respect to p ?
A. Define a total order for each point p. [next lecture]

Q. How to determine whether $p_1 \rightarrow p_2 \rightarrow p_3$ is a counterclockwise turn?
A. Computational geometry. [next two slides]

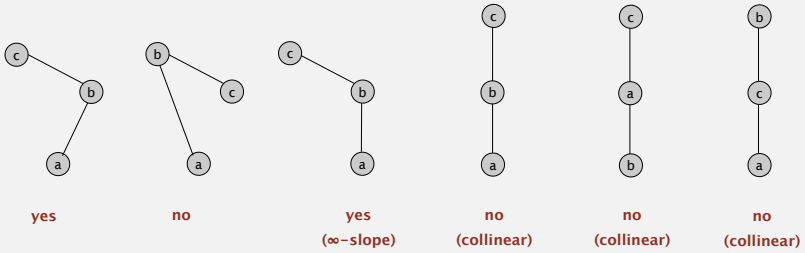
Q. How to sort efficiently?
A. Mergesort sorts in $N \log N$ time. [next lecture]

Q. How to handle degeneracies (three or more points on a line)?
A. Requires some care, but not hard. [see booksite]

Implementing ccw

CCW. Given three points a, b, and c, is $a \rightarrow b \rightarrow c$ a counterclockwise turn?

is c to the left of the ray a→b



Lesson. Geometric primitives are tricky to implement.

- Dealing with degenerate cases.
- Coping with floating-point precision.

Implementing ccw

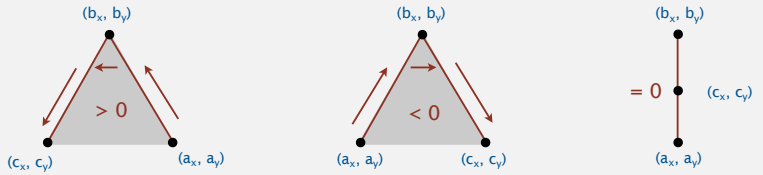
CCW. Given three points a, b, and c, is $a \rightarrow b \rightarrow c$ a counterclockwise turn?

- Determinant of special matrix gives 2x signed area of planar triangle.

$$2 \times Area(a,b,c) = \begin{vmatrix} a_x & a_y & 1 \\ b_x & b_y & 1 \\ c_x & c_y & 1 \end{vmatrix} = (b_x - a_x)(c_y - a_y) - (b_y - a_y)(c_x - a_x)$$

|(b – a) × (c – a)|

- If signed area > 0, then $a \rightarrow b \rightarrow c$ is counterclockwise.
- If signed area < 0, then $a \rightarrow b \rightarrow c$ is clockwise.
- If signed area $= 0$, then $a \rightarrow b \rightarrow c$ are collinear.



Immutable point data type

```
public class Point2D
{
   private final double x;
   private final double y;

   public Point2D(double x, double y)
   {
      this.x = x;
      this.y = y;
   }

   ...

   public static int ccw(Point2D a, Point2D b, Point2D c)
   {
      double area2 = (b.x-a.x)*(c.y-a.y) - (b.y-a.y)*(c.x-a.x);
      if      (area2 < 0) return -1;  // clockwise
      else if (area2 > 0) return +1;  // counter-clockwise
      else                return  0;  // collinear
   }
}
```

danger of floating-point roundoff error

Graham scan: implementation

Simplifying assumptions. No three points on a line; at least 3 points.

```
Stack<Point2D> hull = new Stack<Point>();

Arrays.sort(p, Point2D.Y_ORDER);
Arrays.sort(p, p[0].BY_POLAR_ORDER);

hull.push(p[0]);
hull.push(p[1]);

for (int i = 2; i < N; i++) {
   Point2D top = hull.pop();
   while (Point2D.ccw(hull.peek(), top, p[i]) <= 0)
      top = hull.pop();
   hull.push(top);
   hull.push(p[i]);
}
```

p[0] is now point with lowest y-coordinate (can do more efficiently without sorting)

sort by polar angle with respect to p[0]

definitely on hull

discard points that would create clockwise turn

add p[i] to putative hull

Running time. $N \log N$ for sorting and linear for rest.

Pf. $N \log N$ for sorting; each point pushed and popped at most once.

In closing

Sorting.

- Useful on its own.
- Can be used as a stepping stone to solving other problems.
  - Shuffling.
  - Convex hull.
  - Finding duplicates in an array.
  - Finding similarities between arrays.
- COS226: Solving diverse problems using standard algorithmic tools.