Algorithmics and Programming II: Introduction

Jordi Cortadella and Jordi Petit
Department of Computer Science
• Lecturers:
  – Jordi Cortadella (jordi.cortadella@upc.edu)
  – Emma Rollón (erollon@cs.upc.edu)
  – Jordi Petit (jordi.petit-silvestre@cs.upc.edu)

• Sessions:
  – Theory & (Jordi C.)
  – Lab (Emma & Jordi P.)
Material

• Slides, exercises:

https://www.cs.upc.edu/~jordicf/Teaching/AP2

• Jutge (for lab sessions):

https://jutge.org

• Lliçons (by J. Petit and S. Roura):

https://lliçons.jutge.org
The lectures of the course are combined with practical programming sessions using a virtual learning environment for computer programming (Jutge.org).

<table>
<thead>
<tr>
<th>Lectures</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
<td></td>
</tr>
<tr>
<td>2. Abstract Data Types</td>
<td></td>
</tr>
<tr>
<td>3. Algorithm Analysis</td>
<td></td>
</tr>
<tr>
<td>4. Divide and Conquer</td>
<td></td>
</tr>
<tr>
<td>5. Memory Management</td>
<td></td>
</tr>
<tr>
<td>6. Containers: Stack</td>
<td></td>
</tr>
<tr>
<td>7. Containers: Queue and List</td>
<td></td>
</tr>
<tr>
<td>8. Containers: Priority Queue</td>
<td></td>
</tr>
<tr>
<td>9. Graphs: Connectivity</td>
<td></td>
</tr>
<tr>
<td>11. Trees</td>
<td></td>
</tr>
<tr>
<td>12. Containers: Set and Dictionary</td>
<td></td>
</tr>
<tr>
<td>13. Hashing</td>
<td></td>
</tr>
<tr>
<td>14. Fast Fourier Transform</td>
<td></td>
</tr>
<tr>
<td>15. Cryptography</td>
<td></td>
</tr>
</tbody>
</table>
Evaluation

• Evaluation items:
  – Projects (Proj), Parcial Lab (PLab), Final Theory (FTh), Final (FLab).

• Grading:
  – $N_1 = 0.2 \text{ Proj} + 0.25 \text{ PLab} + 0.25 \text{ FLab} + 0.3 \text{ FTh}$
  – $N_2 = 0.2 \text{ Proj} + 0.4 \text{ FLab} + 0.4 \text{ FTh}$
  – $N = \max(N_1, N_2)$
First project: Containers

• Design a class to manage containers.
• Language: Python.
Peer and self assessment

• The project will be evaluated by the students themselves.

• Each project will be evaluated by three students. The grade will be calculated as the average grade given by the students.

• The evaluation will be completely blind.

• Biased evaluations will be detected and penalized.

• Each student will have the right to request the evaluation by the professor (who can upgrade or downgrade the evaluation given by the students).
“Programming is an art of telling another human what one wants the computer to do.”

“An algorithm must be seen to be believed.”

“The real problem is that programmers have spent far too much time worrying about efficiency in the wrong places and at the wrong times; premature optimization is the root of all evil (or at least most of it) in programming.”
Second Project: GPS

Telegram

Language: Python
Objective of the course

Confronting large and difficult problems. How?

- Skills for abstraction and algorithmic reasoning.
- Design and use of complex data structures.
- Techniques for complexity analysis.
- Methodologies for modular programming.
- High-quality code.
Compute the convex hull of $n$ given points in the plane.
The Closest-Points problem

• **Input:** A list of $n$ points in the plane
  \[\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}\]
  
• **Output:** The pair of closest points

• **Simple approach:** check all pairs $\rightarrow O(n^2)$

• We want an $O(n \log n)$ solution!
Navigation: find the shortest path

Graph showing the shortest paths between cities in Germany, with distances in kilometers:

- Frankfurt to Mannheim: 85 km
- Frankfurt to Würzburg: 217 km
- Würzburg to Stuttgart: 173 km
- Stuttgart to Kassel: 183 km
- Kassel to Nürnberg: 502 km
- Nürnberg to Augsburg: 167 km
- Augsburg to München: 84 km
- Mannheim to Erfurt: 80 km
- Erfurt to Karlsruhe: 186 km
- Karlsruhe to Stuttgart: 103 km
- Karlsruhe to München: 250 km
How to encrypt messages?

This file is written in plaintext

Encryption Algorithm

2q9733i10
79c4
ffsfsf24yu3

Decryption Algorithm

This file is written in plaintext
The secret: training, training, training, training ...
... up to the finish line
Abstract Data Types (I) (and Object-Oriented Programming)

Jordi Cortadella and Jordi Petit
Department of Computer Science
How many horses can you distinguish?
I forget how I wanted to begin this story. That's probably because my mind, just like everyone else's, can only remember a few things at a time. Researchers have often debated the maximum amount of items we can store in our conscious mind, in what's called our working memory, and a new study puts the limit at three or four.

Working memory is a more active version of short-term memory, which refers to the temporary storage of information. Working memory relates to the information we can pay attention to and manipulate.
Two examples

// Main loop of binary search
while (left <= right) {
    int i = (left + right)/2;
    if (x < A[i]) right = i - 1;
    else if (x > A[i]) left = i + 1;
    else return i;
}

// Main loop of insertion sort
for (int i = 1; i < A.size(); ++i) {
    int x = A[i];
    int j = i;
    while (j > 0 and A[j - 1] > x) {
        A[j] = A[j - 1];
        --j;
    }
    A[j] = x;
}
Hiding details: abstractions
Different types of abstractions
Concept maps are hierarchical: why?

Each level has few items
The computer systems stack

Image Credit: Christopher Batten, Cornell University
The computer systems stack

- Application
- Algorithm
- Programming Language
- Operating System
- Instruction Set Architecture
- Microarchitecture
- Register-Transfer Level
- Gate Level
- Circuits
- Devices
- Technology

How data flows through system

- Boolean logic gates and functions

Combining devices to do useful work

- Transistors and wires

- Silicon process technology

Image Credit: Christopher Batten, Cornell University
The computer systems stack

- Application
- Algorithm
- Programming Language
- Operating System
- Instruction Set Architecture
- Microarchitecture
- Register-Transfer Level
- Gate Level
- Circuits
- Devices
- Technology

Mac OS X, Windows, Linux
Handles low-level hardware management

MIPS32 Instruction Set
Instructions that machine executes

```
blez $a2, done
move $a7, $zero
li $t4, 99
move $a4, $a1
move $v1, $zero
li $a3, 99
lw $a5, 0($a4)
addiu $a4, $a4, 4
slt $a6, $a5, $a3
movn $v0, $v1, $a6
addiu $v1, $v1, 1
movn $a3, $a5, $a6
```

Image Credit: Christopher Batten, Cornell University
The computer systems stack

Sort an array of numbers
2,6,3,8,4,5 -> 2,3,4,5,6,8

Insertion sort algorithm
1. Find minimum number in input array
2. Move minimum number into output array
3. Repeat steps 1 and 2 until finished

C implementation of insertion sort
```c
void isort( int b[], int a[], int n ) {
    for ( int idx, k = 0; k < n; k++ ) {
        int min = 100;
        for ( int i = 0; i < n; i++ ) {
            if ( a[i] < min ) {
                min = a[i];
                idx = i;
            }
        }
        b[k] = min;
        a[idx] = 100;
    }
}
```

Image Credit: Christopher Batten,
Cornell University
Our challenge

• We need to design large systems and reason about complex algorithms.

• Our working memory can only manipulate 4 things at once.

• We need to interact with computers using programming languages.

• Solution: abstraction
  – Abstract reasoning.
  – Programming languages that support abstraction.

• We already use a certain level of abstraction: functions. But it is not sufficient. We need much more.
Data types

• Programming languages have a set of primitive data types (e.g., int, bool, double, char, ...).

• Each data type has a set of associated operations:
  – We can add two integers.
  – We can concatenate two strings.
  – We can divide two doubles.
  – But we cannot divide two strings!

• Programmers can add new operations to the primitive data types:
  – gcd(a,b), match(string1, string2), ...

• The programming languages provide primitives to group data items and create structured collections of data:
  – C++: array, struct.
  – python: list, tuple, dictionary.
Abstract Data Types (ADTs)

A set of objects and a set of operations to manipulate them

Operations:
- Number of vertices
- Number of edges
- Shortest path
- Connected components

Data type: Graph
Abstract Data Types (ADTs)

A set of objects and a set of operations to manipulate them:

\[ P(x) = x^3 - 4x^2 + 5 \]

Data type: Polynomial

Operations:
- \( P + Q \)
- \( P \times Q \)
- \( P/Q \)
- \( \gcd(P, Q) \)
- \( P(x) \)
- \( \text{degree}(P) \)
Abstract Data Types (ADTs)

• Separate the notions of specification and implementation:
  – Specification: “what does an operation do?”
  – Implementation: “how is it done?”

• Benefits:
  – Simplicity: code is easier to understand
  – Encapsulation: details are hidden
  – Modularity: an ADT can be changed without modifying the programs that use it
  – Reuse: it can be used by other programs
Abstract Data Types (ADTs)

• An ADT has two parts:
  – Public or external: abstract view of the data and operations (methods) that the user can use.
  – Private or internal: the actual implementation of the data structures and operations.

• Operations:
  – Creation/Destruction
  – Access
  – Modification
Abstract Data Types (ADTs)

- Internal Data Representation
- Private Operations
  - Create
  - Destruct
  - Read
  - Write
  - Modify

User Interface (API)

API: Application Programming Interface
A point can be represented by two coordinates \((x, y)\).

Several operations can be envisioned:

- Get the \(x\) and \(y\) coordinates.
- Calculate distance between two points.
- Calculate polar coordinates.
- Move the point by \((\Delta x, \Delta y)\).
Example: a Point

// Things that we can do with points

Point p1(5.0, -3.2); // Create a point (a variable)
Point p2(2.8, 0); // Create another point

// We now calculate the distance between p1 and p2
double dist12 = p1.distance(p2);

// Distance to the origin
double r = p1.distance();

// Create another point by adding coordinates
Point p3 = p1 + p2;

// We get the coordinates of the new point
double x = p3.getX(); // x = 7.8
double y = p3.getY(); // y = -3.2
ADTs and Object-Oriented Programming

• OOP is a programming paradigm: a program is a set of objects that interact with each other.

• An object has:
  – fields (or attributes) that contain data
  – functions (or methods) that contain code

• Objects (variables) are instances of classes (types). A class is a template for all objects of a certain type.

• In OOP, a class is the natural way of implementing an ADT.
Classes and Objects

class

car

objects

polo

mini

beetle
// The declaration of the class Point
class Point {

public:
  // Constructor
  Point(double x_coord, double y_coord);
  // Constructor for (0,0)
  Point();
  // Gets the x coordinate
  double getX() const;
  // Gets the y coordinate
  double getY() const;
  // Returns the distance to point p
  double distance(const Point& p) const;
  // Returns the distance to the origin
  double distance() const;
  // Returns the angle of the polar coordinate
  double angle() const;
  // Creates a new point by adding the coordinates of two points
  Point operator + (const Point& p) const;

private:
  double x, y; // Coordinates of the point
};
Implementation of the class Point

// The constructor: different implementations
Point::Point(double x_coord, double y_coord) {
    x = x_coord; y = y_coord;
}

// or also
Point::Point(double x_coord, double y_coord) : x(x_coord), y(y_coord) {}

// or also
Point::Point(double x, double y) : x(x), y(y) {}

All of them are equivalent, but only one of them should be chosen. We can have different constructors with different signatures.

// The other constructor
Point::Point() : x(0), y(0) {}
Implementation of the class Point

double Point::getX() const {
    return x;
}

double Point::getY() const {
    return y;
}

double Point::distance(const Point& p) const {
    double dx = getX() - p.getX(); // Better getX() than x
    double dy = getY() - p.getY();
    return sqrt(dx*dx + dy*dy);
}

double Point::distance() const {
    return sqrt(getX()*getX() + getY()*getY());
}

Note: compilers are smart. Small functions are expanded inline.
## Implementation of the class Point

```cpp
double Point::angle() const {
    if (getX() == 0 and getY() == 0) return 0;
    return atan(getY()/getX());
}

Point Point::operator + (const Point& p) const {
    return Point(getX() + p.getX(), getY() + p.getY());
}
```

- **operator overloading**
- **p1 + p2**
- “this” object
- parameter of operator +
File organization: one file

Point.hh

```cpp
#ifndef __POINT_H__
#define __POINT_H__

class Point {
  public:
    // Constructor
    Point(double x, double y) : x(x), y(y) {}

    // Gets the x coordinate
    double getX() const {
        return x;
    }

  private:
    double x, y; // Coordinates of the point
};

#endif // __POINT_H__
```

Only one header file (\.hh) that contains the specification and the implementation.

Advantages:
- Easy distribution.
- Useful to implement templates.

Disadvantages:
- More compile effort.
- The implementation is revealed.
A header file (.hh) containing the specification and a C++ file (.cc) containing the implementation.

Advantages:
- Less compile effort.
- Hidden implementation.

Disadvantages:
- Need to distribute a library.
- Data representation still visible.
Conclusions

• The human brain has limitations: 4 things at once.

• Modularity and abstraction are for designing large maintainable systems.
Abstract Data Types (II)
(and Object-Oriented Programming)

Jordi Cortadella and Jordi Petit
Department of Computer Science
Public or private?

• What should be public?
  – Only the methods that need to interact with the external world. Hide as much as possible. Make a method public only if necessary.

• What should be private?
  – All the attributes.
  – The internal methods of the class.

• Can we have public attributes?
  – Theoretically yes (C++ and python allow it).
Class Point: a new implementation

• Let us assume that we need to represent the point with polar coordinates for efficiency reasons (e.g., we need to use them very often).

• We can modify the private section of the class without modifying the specification of the public methods.

• The private and public methods may need to be rewritten, but not the programs using the public interface.
Let us design the new type for Point

// The declaration of the class Point
class Point {

public:
    // Constructor
    Point(double x_coord, double y_coord);

    // Constructor for (0,0)
    Point();

    // Gets the x coordinate
    double getX() const;

    // Gets the y coordinate
    double getY() const;

    // Returns the distance to point p
    double distance(const Point& p) const;

    // Returns the distance to the origin
    double distance() const;

    // Returns the angle of the polar coordinate
    double angle() const;

    // Creates a new point by adding the coordinates of two points
    Point operator + (const Point& p) const;

private:
    double _radius, _angle; // Polar coordinates
};
Point::Point(double x, double y) :
    _radius(sqrt(x*x + y*y)),
    _angle(x == 0 and y == 0 ? 0 : atan(y/x))
{};

double Point::getX() const {
    return _radius*cos(_angle);
}

double Point::getY() const {
    return _radius*sin(_angle);
}

double Point::distance(const Point& p) const {
    double dx = getX() - p.getX();
    double dy = getY() - p.getY();
    return sqrt(dx*dx + dy*dy);
}

double Point::distance() const {
    return _radius;
}
double Point::angle() const {
    return _angle;
}

// Notice that no changes are required for the + operator
// with regard to the initial implementation of the class
Point Point::operator + (const Point& p) const {
    return Point(getX() + p.getX(), getY() + p.getY());
}

Discussion:
• How about having x and y (or _radius and _angle) as public attributes?
• Programs using p.x and p.y would not be valid for the new implementation.
• Programs using p.getX() and p.getY() would still be valid.

Recommendation (reminder):
• All attributes should be private.
Public/private: let’s summarize

class Z {
    public:
        ...
        void f(const Z& x) {
            ... a ... // “this” attribute
            ... x.a ... // x’s attribute
            g(); // Ok
            x.g(); // Ok
        }
    ...
    private:
        T a; // Attribute
        ...
        void g(...) {...}
        ...
};

int main () {
    Z v1, v2;
    ...
    v1.f(v2); // Ok
    ...
    v1.g(...); // Wrong! (private)
    ...
    v1.a ... // Wrong! (private)
    v1 = v2; // Ok (copy)
}
A new class: Rectangle

• We will only consider orthogonal rectangles (axis-aligned).

• An orthogonal rectangle can be represented in different ways:

  Two points (extremes of diagonal)

  One point, width and height
Rectangle: abstract view

Create

Rectangle(8,5)

(0,0)

Scale

scale(0.5)

Rotate

rotate(false)

Move

move(10,2)

(1,8)

(11,10)

Flip (horizontally/vertically)

Intersection

Point inside?
class Rectangle {
public:
    // Constructor (LL at the origin)
    Rectangle(double width, double height);

    // Returns the area of the rectangle
    double area() const;

    // Scales the rectangle with a factor s > 0
    void scale(double s);

    // Returns the intersection with another rectangle
    Rectangle operator * (const Rectangle& R) const;

...;

};
Rectangle: using the ADT

Rectangle R1(4,5); // Creates a rectangle 4x5
Rectangle R2(8,4); // Creates a rectangle 8x4

R1.move(2,3); // Moves the rectangle
R1.scale(1.2); // Scales the rectangle
double Area1 = R1.Area(); // Calculates the area

Rectangle R3 = R1 * R2;

if (R3.empty()) ...
class Rectangle {
public:

private:
    Point ll;       // Lower-left corner of the rectangle
    double w, h;   // width/height of the rect.

    Other private data and methods (if necessary)

};
// LL at the origin
Rectangle::Rectangle(double w, double h) :
   ll(Point(0,0)), w(w), h(h) {} 

// LL specified at the constructor
Rectangle::Rectangle(const Point& p, double w, double h) :
   ll(p), w(w), h(h) {} 

// LL and UR specified at the constructor
Rectangle::Rectangle(const Point& ll, const Point& ur) :
   ll(ll), w(ur.getX() - ll.getX()), h(ur.getY() - ll.getY()) {} 

// Empty rectangle (using another constructor)
Rectangle::Rectangle() : Rectangle(0, 0) {}
Rectangle: overloaded operators

```cpp
Rectangle& Rectangle::operator *= (const Rectangle& R) {
    // Calculate the ll and ur coordinates
    Point Rll = R.getLL();
    ll = Point(max(ll.getX(), Rll.getX()),
               max(ll.getY(), Rll.getY()));

    Point ur = getUR();
    Point Rur = R.getUR();
    double urx = min(ur.getX(), Rur.getX());
    double ury = min(ur.getY(), Rur.getY());

    // Calculate width and height (might be negative → empty)
    w = urx - ll.getX();
    h = ury - ll.getY();

    return *this;
}

// Use *= to implement *
Rectangle Rectangle::operator * (const Rectangle& R) const {
    Rectangle result = *this;  // Make a copy of itself
    result *= R;
    return result;
}
```

Rectangle: other public methods

```cpp
Point Rectangle::getLL() const {
    return ll;
}

Point Rectangle::getUR() const {
    return ll + Point(w, h);
}

double Rectangle::getWidth() const {
    return w;
}

double Rectangle::getHeight() const {
    return h;
}

double Rectangle::area() const {
    return w*h;
}

// Notice: not a const method
void Rectangle::scale(double s) {
    w *= s;
    h *= s;
}

bool Rectangle::empty() const {
    return w <= 0 or h <= 0;
}
```
What is *this?

- *this* is a pointer (memory reference) to the object (pointers will be explained later)

- *this* is the object itself
Let us work with rectangles

Rectangle R1(Point(2,3), Point(6,8));
double areaR1 = R1.area(); // areaR1 = 20

Rectangle R2(Point(3,5), 2, 4); // LL=(3,5) UR=(5,9)

// Check whether the point (4,7) is inside the
// intersection of R1 and R2.
bool in = (R1*R2).isPointInside(Point(4,7));
// The object R1*R2 is “destroyed” after the assignment.

R2.rotate(false); // R2 is rotated counterclockwise
R2 *= R1; // Intersection with R1

Exercise: draw a picture of R1 and R2 after the execution of the previous code.
Yet another class: Rational

What we would like to do:

Rational R1(3); // R1 = 3
Rational R2(5, 4); // R2 = 5/4
Rational R3(8, -10); // R3 = -4/5

R3 += R1 + Rational(-1, 5); // R3 = 2

if (R3 >= Rational(2)) {
    R3 = -R1*R2; // R3 = -15/4
}

cout << R3.to_str() << endl;
class Rational {
private:
    int num, den; // Invariant: den > 0 and gcd(num,den)=1

    // Simplifies the fraction (used after each operation)
    void simplify();

public:
    // Constructor (if some parameter is missing, the default value is taken)
    Rational(int num = 0, int den = 1);

    // Returns the numerator of the fraction
    int getNum() const {
        return num;
    }

    // Returns the denominator of the fraction
    int getDen() const {
        return den;
    }

    // Returns true if the number is integer and false otherwise.
    bool isInteger() const {
        return den == 1;
    }

    ...}
};
class Rational {

public:

    ...

    // Arithmetic operators
    Rational& operator += (const Rational& r);
    Rational operator + (const Rational& r) const;
    // Similarly for -, *, and /

    // Unary operator
    Rational operator - () const {
        return Rational(-getNum(), getDen());
    }

    // Equality comparison
    bool operator == (const Rational& r);

    // Returns a string representing the rational
    string to_str() const;
};
Rational::Rational(int num, int den) : num(num), den(den) {
    simplify();
}

void Rational::simplify() {
    assert(den != 0); // We will discuss assertions later
    if (den < 0) {
        num = -num;
        den = -den;
    }

    // Divide by the gcd of num and den
    int d = gcd(abs(num), den);
    num /= d;
    den /= d;
}
Rational: arithmetic and relational operators

Rational& Rational::operator += (const Rational& r) {
    num = getNum() * r.getDen() + getDen() * r.getNum();
    den = getDen() * r.getDen();
    simplify();
    return *this;
}

Rational Rational::operator + (const Rational& r) {
    Rational result = *this;  // A copy of this object
    result += r;
    return result;
}

bool Rational::operator == (const Rational& r) {
    return getNum() == r.getNum() and getDen() == r.getDen();
}

bool Rational::operator != (const Rational& r) {
    return not operator ==(r);
}

string Rational::to_str() const {
    string s(to_string(getNum()));
    if (not isInteger()) s += “/” + to_string(getDen());
    return s;
}
Classes and Objects in Python
A Python session with rational numbers

```python
>>> from rational import Rational # from file rational.py
>>> a = Rational(4, -6)           # construct with num and den
>>> print(a)
-2/3
>>> b = Rational(4)               # integer value
>>> print(b)
4
>>> print((a+b).num(), (a+b).den())
10 3
>>> c = Rational()                # c = 0
>>> if a < c:
...   print(a, "is negative")
...
-2/3 is negative
>>> print(a*b)                    # uses the __str__ method (see later)
-8/3
>>> a/b                           # uses the __repr__ method (see later)
Rational(-1/6)
```
The Rational class in Python

```python
class Rational:
    def __init__(self, num=0, den=1):
        if not isinstance(num, int):
            raise TypeError("The numerator is not an integer")
        if not isinstance(den, int):
            raise TypeError("The denominator is not an integer")
        if den == 0:
            raise ZeroDivisionError("The denominator is zero")
        self._num = num
        self._den = den
        self._simplify()
```

Classes only have one constructor (**`__init__`**). Multiple constructors can be “simulated” by checking the number and type of arguments.

All class methods and attributes are public.

Naming convention: use underscore prefix for "private" attributes and methods (e.g., `_num`, `_den`, `_simplify`)

**self** in Python is similar to **this** in C++. All methods must be declared with **self** as first argument. Exception: static methods (not discussed here).
class Rational:
    :

def num(self):
    # Public method
    return self._num

def den(self):
    # Public method
    return self._den

def _simplify(self):
    # Private method
    if self._den < 0:
        self._num *= -1
        self._den *= -1
d = math.gcd(abs(self._num), self._den)
    self._num //= d
    self._den //= d
```python
class Rational:
    
    def __str__(self):
        """Returns a user-friendly string with information about the value of the object. It is invoked by str(x) or print(x)."""
        if self._den == 1:
            return str(self._num)
        return str(self._num) + "/" + str(self._den)

    def __repr__(self):
        """Returns a string with information about the representation of the class. It is invoked by repr(x) or simply 'x'."""
        return "Rational(" + str(self) + ")"
```
class Rational:

    
    def __neg__(self):
        """Returns -self.""
        return Rational(-self._num, self._den)

    def __add__(self, rhs):
        """Returns self + rhs.""
        num = self._num*rhs._den + self._den*rhs._num
        den = self._den*rhs._den
        return Rational(num, den)

    # Similarly for __sub__, __mul__, __truediv__
Relational operators

class Rational:
    :
    def __eq__(self, rhs):
        """Checks whether self == rhs.""
        return self._num == rhs._num and self._den == rhs._den
    def __ne__(self, rhs):
        """Checks whether self != rhs.""
        return not self == rhs
    def __lt__(self, rhs):
        """Checks whether self < rhs.""
        return self._num*rhs._den < self._den*rhs._num
    def __le__(self, rhs):
        """Checks whether self <= rhs.""
        return not rhs < self
    # Similarly for __gt__ and __ge__
>>> from rational import Rational
>>> help(Rational.__add__)
Help on function __add__ in module rational:

__add__(self, rhs)
    Returns self + rhs.

>>> help(Rational)
class Rational(builtins.object)
    Rational(num=0, den=1)

    Class to manipulate rational numbers.

    The class includes the basic arithmetic and relational operators.

    Methods defined here:

    __add__(self, rhs)
        Returns self + rhs.

    __eq__(self, rhs)
        Checks whether self == rhs.
Python documentation: *docstrings*

- The first line after a module, class or function can be used to insert a string that documents the component.

- Triple quotes ("""") are very convenient to insert multi-line strings.

- The *docstrings* are stored in a special attribute of the component named __doc__.

- Different ways of print the *docstrings* associated to a component:
  - print(Rational.num.__doc__)
  - help(Rational.num)
# geometry.py

"""geometry.py
Provides two classes for representing Polygons and Circles."""

# author: Euclid of Alexandria

from math import pi, sin, cos

class Polygon:
    """Represents polygons and provides methods to calculate area, intersection, convex hull, etc."""

    def __init__(self, list_vertices):
        """Creates a polygon from a list of vertices."""
        ...

class Circle:
    ...

Documentation of the module

Documentation of the class

Documentation of the method
Using a module: example

```python
import geometry
p = geometry.Poligon(...)
c = geometry.Circle(...)
```

Imports the module. Now all classes can be used with the prefix of the module.

```python
import geometry as geo
p = geo.Poligon(...)
c = geo.Circle(...)
```

Imports and renames the module.

```python
from geometry import *
p = Poligon(...)
c = Circle(...)
```

Imports all classes in the module. No need to add the prefix of the module.

```python
from geometry import Poligon as plg, Circle as cir
p = plg(...)
c = cir(...)
```

Imports and renames the classes in the module.
Conclusions

• Finding the appropriate hierarchy is a fundamental step towards the design of a complex system.

• User-friendly documentation is indispensable.
EXERCISES
Implement the following methods for the class Rectangle:

```cpp
// Rotate the rectangle 90 degrees clockwise or counterclockwise, depending on the value of the parameter. The rotation should be done around the lower-left corner of the rectangle.
void rotate(bool clockwise);

// Flip horizontally (around the left edge) or vertically (around the bottom edge), depending on the value of the parameter.
void flip(bool horizontally);

// Check whether point p is inside the rectangle
bool isPointInside(const Point& p) const;
```
Re-implement a class

Re-implement the class Rectangle using an internal representation with two Points:

- Lower-Left (LL)
- Upper-Right (UR)
Algorithm Analysis (I)

Jordi Cortadella and Jordi Petit
Department of Computer Science
What do we expect from an algorithm?

• Correct
• Easy to understand
• Easy to implement
• Efficient:
  – Every algorithm requires a set of resources
    • Memory
    • CPU time
    • Energy
// Pre: n ≥ 0
// Returns the Fibonacci number of order n.

int fib(int n) { // Recursive solution
    if (n <= 1) return n;
    return fib(n - 1) + fib(n - 2);
}
Fibonacci

How many recursive calls?
Fibonacci: runtime

\[ T(0) = 1 \]
\[ T(1) = 1 \]
\[ T(n) = T(n - 1) + T(n - 2) \]

Let us assume that \( T(n) = a^n \) for some constant \( a \). Then,

\[ a^n = a^{n-1} + a^{n-2} \quad \Rightarrow \quad a^2 = a + 1 \]

\[ a = \frac{1 + \sqrt{5}}{2} = \varphi \approx 1.618 \quad \text{(golden ratio)} \]

Therefore, \( T(n) \approx 1.6^n \).

If \( T(0) = 1 \) ns, then \( T(100) \approx 2.6 \cdot 10^{20} \) ns > 8000 yrs.

With the age of Universe (14 · 10^9 yrs), we could compute up to \text{fib}(128).\)
Fibonacci numbers: iterative version

// Pre: n ≥ 0
// Returns the Fibonacci number of order n.
int fib(int n) { // iterative solution
    int f_i = 0;
    int f_i1 = 1;
    // Inv: f_i is the Fibonacci number of order i.
    //      f_i1 is the Fibonacci number of order i+1.
    for (int i = 0; i < n; ++i) {
        int f = f_i + f_i1;
        f_i = f_i1;
        f_i1 = f;
    }
    return f_i;
}
Fibonacci numbers

Algebraic solution: find matrix $A$ such that

\[
\begin{bmatrix}
F_{n+2} \\
F_{n+1}
\end{bmatrix} = \begin{bmatrix}
? & ? \\
? & ?
\end{bmatrix} \cdot \begin{bmatrix}
F_{n+1} \\
F_n
\end{bmatrix}
\]

\[
\begin{bmatrix}
F_{n+2} \\
F_{n+1}
\end{bmatrix} = \begin{bmatrix}
1 & 1 \\
1 & 0
\end{bmatrix} \cdot \begin{bmatrix}
F_{n+1} \\
F_n
\end{bmatrix}
\]

\[
\begin{bmatrix}
F_{n+1} \\
F_n
\end{bmatrix} = A^n \cdot \begin{bmatrix}
1 \\
0
\end{bmatrix}
\]
Fibonacci numbers

\[ A^1 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \quad A^2 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \]

\[ A^4 = \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix} \quad A^8 = \begin{bmatrix} 34 & 21 \\ 21 & 13 \end{bmatrix} \]

\[ A^{16} = \begin{bmatrix} 1597 & 987 \\ 987 & 610 \end{bmatrix} \quad \cdots \quad A^n = \begin{bmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{bmatrix} \]

Runtime \( \approx \log_2 n \) \ 2x2 matrix multiplications
Given an algorithm that reads inputs from a domain $D$, we want to define a cost function $C$:

$$C : D \rightarrow \mathbb{R}^+$$

where $C(x)$ represents the cost of using some resource (CPU time, memory, energy, ...). Analyzing $C(x)$ for every possible $x$ is impractical.
Algorithm analysis: simplifications

- Analysis based on the size of the input: $|x| = n$

- Only the best/average/worst cases are analyzed:

  $$C_{\text{worst}}(n) = \max\{C(x) : x \in D, |x| = n\}$$

  $$C_{\text{best}}(n) = \min\{C(x) : x \in D, |x| = n\}$$

  $$C_{\text{avg}}(n) = \sum_{x \in D, |x| = n} p(x) \cdot C(x)$$

  $p(x)$: probability of selecting input $x$ among all the inputs of size $n$. 
Properties:

\[ \forall n \geq 0 : \quad C_{\text{best}}(n) \leq C_{\text{avg}}(n) \leq C_{\text{worst}}(n) \]

\[ \forall x \in D : \quad C_{\text{best}}(|x|) \leq C(x) \leq C_{\text{worst}}(|x|) \]

- We want a notation that characterizes the cost of algorithms independently from the technology (CPU speed, programming language, efficiency of the compiler, etc.).

- Runtime is usually the most important resource to analyze.
Asymptotic notation

Let us consider all functions $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$

Definitions:

$$O(f(n)) = \{g(n) : \exists k > 0, \exists n_0, \forall n \geq n_0 : g(n) \leq k \cdot f(n)\}$$

$$\Omega(f(n)) = \{g(n) : \exists k > 0, \exists n_0, \forall n \geq n_0 : g(n) \geq k \cdot f(n)\}$$

$$\Theta(f(n)) = O(f(n)) \cap \Omega(f(n))$$
Asymptotic notation

\[ g(n) \in O(f_1(n)) \]
\[ g(n) \in \Omega(f_2(n)) \]
Asymptotic notation

\[ g(n) \in \Theta(f(n)) \]
Asymptotic notation: example

\[ O(n^2) = \{3n^2 - n + 20, \]
\[ 0.5n^2 - 3, \]
\[ 4n \log_2 n, \]
\[ 2n - 5, \]
\[ n < 20 \ ? \ 2^n : 2n^2 + 1000, \]
\[ \ldots \]
\[ \} \]

\[ \Omega(n^2) = \{3n^2 - n + 20, \]
\[ 0.5n^2 - 3, \]
\[ 3n^3 \log_2 n, \]
\[ 2^n - 4n^3, \]
\[ n < 20 \ ? \ 2^n : 2n^2 + 1000, \]
\[ \ldots \]
\[ \} \]

\[ \Theta(n^2) = O(n^2) \cap \Omega(n^2) = \{3n^2 - n + 20, \]
\[ 0.5n^2 - 3, \]
\[ n < 20 \ ? \ 2^n : 2n^2 + 1000, \]
\[ \ldots \]
\[ \} \]
### Examples

<table>
<thead>
<tr>
<th>Big-O</th>
<th>Big-Ω</th>
</tr>
</thead>
<tbody>
<tr>
<td>$13n^3 - 4n + 8 \in O(n^3)$</td>
<td>$13n^3 - 4n + 8 \in \Omega(n^3)$</td>
</tr>
<tr>
<td>$2n - 5 \in O(n)$</td>
<td>$n^2 \in \Omega(n)$</td>
</tr>
<tr>
<td>$n^2 \notin O(n)$</td>
<td>$n^2 \notin \Omega(n^3)$</td>
</tr>
<tr>
<td>$2^n \in O(n!)$</td>
<td>$n! \in \Omega(2^n)$</td>
</tr>
<tr>
<td>$3^n \notin O(2^n)$</td>
<td>$3^n \notin \Omega(2^n)$</td>
</tr>
<tr>
<td>$3 \log_2 n \in O(\log n)$</td>
<td>$3 \log_2 n \in \Omega(\log n)$</td>
</tr>
<tr>
<td>$n \log_2 n \in O(n^2)$</td>
<td>$n \log_2 n \in \Omega(n)$</td>
</tr>
<tr>
<td>$O(n^2) \subseteq O(n^3)$</td>
<td>$\Omega(n^3) \subseteq \Omega(n^2)$</td>
</tr>
</tbody>
</table>
Complexity ranking

<table>
<thead>
<tr>
<th>Function</th>
<th>Common name</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n!$</td>
<td>factorial</td>
</tr>
<tr>
<td>$2^n$</td>
<td>exponential</td>
</tr>
<tr>
<td>$n^d$, $d &gt; 3$</td>
<td>polynomial</td>
</tr>
<tr>
<td>$n^3$</td>
<td>cubic</td>
</tr>
<tr>
<td>$n^2$</td>
<td>quadratic</td>
</tr>
<tr>
<td>$n \sqrt{n}$</td>
<td></td>
</tr>
<tr>
<td>$n \log n$</td>
<td>quasi-linear</td>
</tr>
<tr>
<td>$n$</td>
<td>linear</td>
</tr>
<tr>
<td>$\sqrt{n}$</td>
<td>root - $n$</td>
</tr>
<tr>
<td>$\log n$</td>
<td>logarithmic</td>
</tr>
<tr>
<td>$1$</td>
<td>constant</td>
</tr>
</tbody>
</table>
The limit rule

Let us assume that $L$ exists (may be $\infty$) such that:

$$L = \lim_{n \to \infty} \frac{f(n)}{g(n)}$$

\[
\begin{align*}
&\text{if } L = 0 & \text{then } f \in O(g) \\
&\text{if } 0 < L < \infty & \text{then } f \in \Theta(g) \\
&\text{if } L = \infty & \text{then } f \in \Omega(g)
\end{align*}
\]

Note: If both limits are $\infty$ or 0, use L’Hôpital rule:

$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = \lim_{x \to \infty} \frac{f'(x)}{g'(x)}$$
Properties

- $f \in \text{O}(f)$
- $\forall c > 0, \quad \text{O}(f) = \text{O}(c \cdot f)$
- $f \in \text{O}(g) \land g \in \text{O}(h) \Rightarrow f \in \text{O}(h)$
- $f_1 \in \text{O}(g_1) \land f_2 \in \text{O}(g_2) \Rightarrow f_1 + f_2 \in \text{O}(g_1 + g_2) = \text{O}(\max\{g_1, g_2\})$
- $f \in \text{O}(g) \Rightarrow f + g \in \text{O}(g)$
- $f_1 \in \text{O}(g_1) \land f_2 \in \text{O}(g_2) \Rightarrow f_1 \cdot f_2 \in \text{O}(g_1 \cdot g_2)$
- $f \in \text{O}(g) \iff g \in \Omega(f)$
Asymptotic complexity (small values)
Asymptotic complexity (larger values)

\[ 2^n, n^3, n^2, n \log n, n \]

10 20 30 40 50 60 70 80 90

Algorithm Analysis

© Dept. CS, UPC
Execution time: example

Let us consider that every operation can be executed in 1 ns \((10^{-9} \text{ s})\).

<table>
<thead>
<tr>
<th>Function</th>
<th>(n = 1,000)</th>
<th>(n = 10,000)</th>
<th>(n = 100,000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\log_2 n)</td>
<td>10 ns</td>
<td>13.3 ns</td>
<td>16.6 ns</td>
</tr>
<tr>
<td>(\sqrt{n})</td>
<td>31.6 ns</td>
<td>100 ns</td>
<td>316 ns</td>
</tr>
<tr>
<td>(n)</td>
<td>1 (\mu)s</td>
<td>10 (\mu)s</td>
<td>100 (\mu)s</td>
</tr>
<tr>
<td>(n \log_2 n)</td>
<td>10 (\mu)s</td>
<td>133 (\mu)s</td>
<td>1.7 ms</td>
</tr>
<tr>
<td>(n^2)</td>
<td>1 ms</td>
<td>100 ms</td>
<td>10 s</td>
</tr>
<tr>
<td>(n^3)</td>
<td>1 s</td>
<td>16.7 min</td>
<td>11.6 days</td>
</tr>
<tr>
<td>(n^4)</td>
<td>16.7 min</td>
<td>116 days</td>
<td>3171 yr</td>
</tr>
<tr>
<td>(2^n)</td>
<td>(3.4 \times 10^{284}) yr</td>
<td>(6.3 \times 10^{2993}) yr</td>
<td>(3.2 \times 10^{30086}) yr</td>
</tr>
</tbody>
</table>
How about “big data”?

**Source:** Jon Kleinberg and Éva Tardos, Algorithm Design, Addison Wesley 2006.

**Table 2.1** The running times (rounded up) of different algorithms on inputs of increasing size, for a processor performing a million high-level instructions per second. In cases where the running time exceeds $10^{25}$ years, we simply record the algorithm as taking a very long time.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$n$</th>
<th>$n \log_2 n$</th>
<th>$n^2$</th>
<th>$n^3$</th>
<th>$1.5^n$</th>
<th>$2^n$</th>
<th>$n!$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 10$</td>
<td>&lt; 1 sec</td>
<td>&lt; 1 sec</td>
<td>&lt; 1 sec</td>
<td>&lt; 1 sec</td>
<td>&lt; 1 sec</td>
<td>&lt; 1 sec</td>
<td>4 sec</td>
</tr>
<tr>
<td>$n = 30$</td>
<td>&lt; 1 sec</td>
<td>&lt; 1 sec</td>
<td>&lt; 1 sec</td>
<td>&lt; 1 sec</td>
<td>&lt; 1 sec</td>
<td>18 min</td>
<td>$10^{25}$ years</td>
</tr>
<tr>
<td>$n = 50$</td>
<td>&lt; 1 sec</td>
<td>&lt; 1 sec</td>
<td>&lt; 1 sec</td>
<td>&lt; 1 sec</td>
<td>11 min</td>
<td>36 years</td>
<td>very long</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>&lt; 1 sec</td>
<td>&lt; 1 sec</td>
<td>&lt; 1 sec</td>
<td>1 sec</td>
<td>12,892 years</td>
<td>$10^{17}$ years</td>
<td>very long</td>
</tr>
<tr>
<td>$n = 1,000$</td>
<td>&lt; 1 sec</td>
<td>&lt; 1 sec</td>
<td>1 sec</td>
<td>18 min</td>
<td>very long</td>
<td>very long</td>
<td>very long</td>
</tr>
<tr>
<td>$n = 10,000$</td>
<td>&lt; 1 sec</td>
<td>&lt; 1 sec</td>
<td>2 min</td>
<td>12 days</td>
<td>very long</td>
<td>very long</td>
<td>very long</td>
</tr>
<tr>
<td>$n = 100,000$</td>
<td>&lt; 1 sec</td>
<td>2 sec</td>
<td>3 hours</td>
<td>32 years</td>
<td>very long</td>
<td>very long</td>
<td>very long</td>
</tr>
<tr>
<td>$n = 1,000,000$</td>
<td>1 sec</td>
<td>20 sec</td>
<td>12 days</td>
<td>31,710 years</td>
<td>very long</td>
<td>very long</td>
<td>very long</td>
</tr>
</tbody>
</table>

This is often the practical limit for big data.
A robot stands in front of a wall that is infinitely long to the right and left side. The wall has a door somewhere and the robot has to find it to reach the other side. Unfortunately, the robot can only see the part of the wall in front of it.

The robot does not know neither how far away the door is nor what direction to take to find it. It can only execute moves to the left or right by a certain number of steps.

Let us assume that the door is at a distance $d$. How to find the door in a minimum number of steps?
Algorithm 1:

• Pick one direction and move until the door is found.

Complexity:

• If the direction is correct $\rightarrow O(d)$.
• If incorrect $\rightarrow$ the algorithm does not terminate.
Algorithm 2:

- 1 step to the left,
- 2 steps to the right,
- 3 steps to the left, ...
- ... increasing by one step in the opposite direction.

Complexity:

\[ T(d) = 3d + \sum_{i=1}^{d-1} 4i = 3d + 4 \frac{d(d - 1)}{2} = 2d^2 + d = O(d^2) \]
The robot and the door in an infinite wall

Algorithm 3:
• 1 step to the left and return to origin,
• 2 steps to the right and return to origin,
• 3 steps to the left and return to origin,...
• ... increasing by one step in the opposite direction.

Complexity:

\[
T(d) = d + \sum_{i=1}^{d} 2i = d + 2 \frac{d(d + 1)}{2} = d^2 + 2d = O(d^2)
\]
Algorithm 4:
- 1 step to the left and return to origin,
- 2 steps to the right and return to origin,
- 4 steps to the left and return to origin,
- ... doubling the number of steps in the opposite direction.

**Complexity** (assume that $d = 2^n$):

$$T(d) = d + 2 \sum_{i=0}^{n} 2^i = d + 2(2^{n+1} - 1) = 5d - 2 = O(d)$$
Runtime analysis rules

• Variable declarations cost no time.

• *Elementary operations* are those that can be executed with a *small number of basic computer steps* (an assignment, a multiplication, a comparison between two numbers, etc.).

• Vector sorting or matrix multiplication are not elementary operations.

• We consider that the cost of elementary operations is $O(1)$. 
Runtime analysis rules

- Consecutive statements:
  - If $S_1$ is $O(f)$ and $S_2$ is $O(g)$,
    then $S_1;S_2$ is $O(\max\{f, g\})$

- Conditional statements:
  - If $S_1$ is $O(f)$, $S_2$ is $O(g)$ and $B$ is $O(h)$,
    then if $(B)$ $S_1$; else $S_2$; is $O(\max\{f + h, g + h\})$,
    or also $O(\max\{f, g, h\})$. 
Runtime analysis rules

• For/While loops:
  – Running time is at most the running time of the statements inside the loop times the number of iterations

• Nested loops:
  – Analyze inside out: running time of the statements inside the loops multiplied by the product of the sizes of the loops
Nested loops: examples

\[
\text{for (int } i = 0; i < n; ++i) \\
\quad \text{for (int } j = 0; j < n; ++j) \\
\qquad \text{DoSomething(); // } O(1)
\]
\[\Rightarrow O(n^2)\]

\[
\text{for (int } i = 0; i < n; ++i) \\
\quad \text{for (int } j = i; j < n; ++j) \\
\qquad \text{DoSomething(); // } O(1)
\]
\[\Rightarrow O(n^2)\]

\[
\text{for (int } i = 0; i < n; ++i) \\
\quad \text{for (int } j = 0; j < m; ++j) \\
\quad \text{for (int } k = 0; k < p; ++k) \\
\qquad \text{DoSomething(); // } O(1)
\]
\[\Rightarrow O(n \cdot m \cdot p)\]
Linear time: $O(n)$

Running time proportional to input size

// Compute the maximum of a vector
// with n numbers

```c
int m = a[0];
for (int i = 1; i < a.size(); ++i) {
    if (a[i] > m) m = a[i];
}
```
Linear time: $O(n)$

Other examples:

– Reversing a vector

– Merging two sorted vectors

– Finding the largest null segment of a sorted vector: a linear-time algorithm exists (a null segment is a compact sub-vector in which the sum of all the elements is zero)
Logarithmic time: $O(\log n)$

• Logarithmic time is usually related to divide-and-conquer algorithms

• Examples:
  – Binary search
  – Calculating $x^n$
  – Calculating the $n$-th Fibonacci number
Example: recursive $x^y$

// Pre: $x \neq 0, y \geq 0$
// Returns $x^y$

```c
int power(int x, int y) {
    if (y == 0) return 1;
    if (y%2 == 0) return power(x*x, y/2);
    return x*power(x*x, y/2);
}
```

// Assumption: each */% takes O(1)

$$T(x^y) \leq 4 + T((x^2)^{y/2}) \leq 4 + 4 + T((x^4)^{y/4}) \leq \ldots$$

$$T(x^y) \leq 4 + 4 + \ldots + 4$$

$$\log_2 y \text{ times}$$

$$\Rightarrow O(\log y)$$
Linearithmic time: $O(n \log n)$

- **Sorting**: Merge sort and heap sort can be executed in $O(n \log n)$.

- **Largest empty interval**: Given $n$ time-stamps $x_1, \cdots, x_n$ on which copies of a file arrive at a server, what is largest interval when no copies of file arrive?
  - $O(n \log n)$ solution. Sort the time-stamps. Scan the sorted list in order, identifying the maximum gap between successive time-stamps.
Algorithm Analysis (II)

Jordi Cortadella and Jordi Petit
Department of Computer Science
Examples

• Selection sort

• Insertion sort

• The Maximum Subsequence Sum Problem

• Convex Hull
Selection sort uses this invariant:

- This is sorted and contains the \( i-1 \) smallest elements.
- This may not be sorted, but all elements here are larger than or equal to the elements in the sorted part.
Selection Sort

```cpp
void selection_sort(vector<elem>& v) {
    int last = v.size() - 1; // v.size() = n
    for (int i = 0; i < last; ++i) { // 0..n-2
        int k = i;
        for (int j = i + 1; j <= last; ++j) { // i+1..n-1
            if (v[j] < v[k]) k = j;
        }
        swap(v[k], v[i]);
    }
}
```

**Algorithm Analysis**

Observation: notice that \( T(n) \in \Omega(n^2) \), also. Therefore, \( T(n) \in \Theta(n^2) \).
Insertion Sort

• Let us use inductive reasoning:
  – If we know how to sort arrays of size $n-1$,
  – do we know how to sort arrays of size $n$?
Insertion Sort

```cpp
void insertion_sort(vector<elem>& v) {
    for (int i = 1; i < v.size(); ++i) { // n-1 times
        elem x = v[i];
        int j = i;
        while (j > 0 and v[j - 1] > x) { // 0..i times
            v[j] = v[j - 1];
            --j;
        }
        v[j] = x;
    }
}
```

\[
T(n) = \Omega(n)
\]
\[
T(n) = O(n^2)
\]

\[
T_{\text{worst}}(n) = \sum_{i=1}^{n-1} i \cdot O(1) = O(n^2)
\]  \Rightarrow \text{sorted in reverse order}

\[
T_{\text{best}}(n) = \sum_{i=1}^{n-1} O(1) = O(n)
\]  \Rightarrow \text{already sorted}

Algorithm Analysis © Dept. CS, UPC
The Maximum Subsequence Sum Problem

• Given (possibly negative) integers $A_1, A_2, \ldots, A_n$, find the maximum value of $\sum_{k=i}^{j} A_k$. (the max subsequence sum is 0 if all integers are negative).

• Example:
  – Input: -2, 11, -4, 13, -5, -2
  – Answer: 20 (subsequence 11, -4, 13)

The Maximum Subsequence Sum Problem

```cpp
int maxSubSum(const vector<int>& a) {
    int maxSum = 0;
    // try all possible subsequences
    for (int i = 0; i < a.size(); ++i)
        for (int j = i; j < a.size(); ++j) {
            int thisSum = 0;
            for (int k = i; k <= j; ++k) thisSum += a[k];
            if (thisSum > maxSum) maxSum = thisSum;
        }
    return maxSum;
}
```

\[
T(n) = \sum_{i=0}^{n-1} \sum_{j=i}^{n-1} \sum_{k=i}^{j} 1
\]
The Maximum Subsequence Sum Problem

\[ T(n) = \sum_{i=0}^{n-1} \sum_{j=i}^{n-1} \sum_{k=i}^{j} 1 \]

\[ = \sum_{i=0}^{n-1} \sum_{j=i}^{n-1} (j - i + 1) \]

\[ = \sum_{i=0}^{n-1} \frac{(n - i + 1)(n - i)}{2} = \ldots \]

\[ = \frac{n^3 + 3n^2 + 2n}{6} = O(n^3) \]
The Maximum Subsequence Sum Problem

```cpp
int maxSubSum(const vector<int>& a) {
    int maxSum = 0;
    // try all possible subsequences
    for (int i = 0; i < a.size(); ++i) {
        int thisSum = 0;
        for (int j = i; j < a.size(); ++j) {
            thisSum += a[j]; // reuse computation
            if (thisSum > maxSum)  maxSum = thisSum;
        }
    }
    return maxSum;
}
```

Algorithm Analysis

\[ T(n) = \sum_{i=0}^{n-1} \sum_{j=i}^{n-1} 1 = O(n^2) \]
Max Subsequence Sum: Divide & Conquer

<table>
<thead>
<tr>
<th>First half</th>
<th>Second half</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>-3</td>
</tr>
</tbody>
</table>

The max sum can be in one of three places:

- 1\textsuperscript{st} half
- 2\textsuperscript{nd} half
- Spanning both halves and crossing the middle

In the 3\textsuperscript{rd} case, two max subsequences must be found starting from the center of the vector (one to the left and the other to the right)
int maxSumRec(const vector<int>& a, int left, int right) {
    // base cases
    if (left == right)
        if (a[left] > 0) return a[left];
        else return 0;

    // Recursive cases: left and right halves
    int center = (left + right)/2;
    int maxLeft = maxSumRec(a, left, center);
    int maxRight = maxSumRec(a, center + 1, right);

    // 
    
}
Max Subsequence Sum: Divide&Conquer

: 

```c
int maxRCenter = 0, rightSum = 0;
for (int i = center + 1; i <= right; ++i) {
    rightSum += a[i];
    if (rightSum > maxRCenter) maxRCenter = rightSum;
}

int maxLCenter = 0, leftSum = 0;
for (int i = center; i >= left; --i) {
    leftSum += a[i];
    if (leftSum > maxLCenter) maxLCenter = leftSum;
}

return max(maxLeft, maxRight, maxRCenter + maxLCenter);
```

Algorithm Analysis

© Dept. CS, UPC

13
Max Subsequence Sum: Divide & Conquer

\[ T(1) = 1 \]
\[ T(n) = 2T(n/2) + O(n) \]

We will see how to solve this equation formally in the next lesson (Master Theorem). Informally:

\[ T(n) = 2T(n/2) + n = 2(2(T(n/4) + n/2)) + n \]
\[ = 4T(n/4) + n + n = 8T(n/8) + n + n + n = \cdots \]
\[ = 2^k T(n/2^k) + n + n + \cdots + n \]

when \( n = 2^k \) we have that \( k = \log_2 n \)

\[ T(n) = 2^k T(1) + kn = n + n \log_2 n = O(n \log n) \]

But, can we still do it faster?
The Maximum Subsequence Sum Problem

• Observations:
  – If \( a[i] \) is negative, it cannot be the start of the optimal subsequence.
  – Any negative subsequence cannot be the prefix of the optimal subsequence.

• Let us consider the inner loop of the \( O(n^2) \) algorithm and assume that all prefixes of \( a[i..j-1] \) are positive and \( a[i..j] \) is negative:

  – If \( p \) is an index between \( i+1 \) and \( j \), then any subsequence from \( a[p] \) is not larger than any subsequence from \( a[i] \) and including \( a[p-1] \).
  – If \( a[j] \) makes the current subsequence negative, we can advance \( i \) to \( j+1 \).
The Maximum Subsequence Sum Problem

```cpp
int maxSubSum(const vector<int>& a) {
    int maxSum = 0, thisSum = 0;
    for (int i = 0; i < a.size(); ++i) {
        int thisSum += a[i];
        if (thisSum > maxSum) maxSum = thisSum;
        else if (thisSum < 0) thisSum = 0;
    }
    return maxSum;
}
```

\[ T(n) = O(n) \]

<table>
<thead>
<tr>
<th>a:</th>
<th>4</th>
<th>-3</th>
<th>5</th>
<th>-4</th>
<th>-3</th>
<th>-1</th>
<th>5</th>
<th>-2</th>
<th>6</th>
<th>-3</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>thisSum:</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>9</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>maxSum:</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>
A polygon can be represented by a sequence of vertices.

Two consecutive vertices represent an edge of the polygon.

The last edge is represented by the first and last vertices of the sequence.

Vertices: (1,3) (4,1) (7,3) (5,4) (6,7) (2,6)

Edges: (1,3)-(4,1)-(7,3)-(5,4)-(6,7)-(2,6)-(1,3)

// A polygon (an ordered set of vertices)
using Polygon = vector<Point>;;
Create a polygon from a set of points

Given a set of $n$ points in the plane, connect them in a simple closed path.
Simple polygon

- **Input**: \( p_1, p_2, \ldots, p_n \) (points in the plane).
- **Output**: \( P \) (a polygon whose vertices are \( p_1, p_2, \ldots, p_n \) in some order).

- Select a point \( z \) with the largest \( x \) coordinate (and smallest \( y \) in case of a tie in the \( x \) coordinate). Assume \( z = p_1 \).
- For each \( p_i \in \{ p_2, \ldots, p_n \} \), calculate the angle \( \alpha_i \) between the lines \( z - p_i \) and the \( x \) axis.
- Sort the points \( \{ p_2, \ldots, p_n \} \) according to their angles. In case of a tie, use distance to \( z \).
Simple polygon
Simple polygon

Implementation details:

- There is no need to calculate angles (requires \( \arctan \)). It is enough to calculate slopes (\( \Delta y / \Delta x \)).
- There is not need to calculate distances. It is enough to calculate the square of distances (no sqrt required).

**Complexity:** \( O(n \log n) \). The runtime is dominated by the sorting algorithm.
Compute the convex hull of $n$ given points in the plane.
Clockwise and counter-clockwise

How to calculate whether three consecutive vertices are in a **clockwise** or **counter-clockwise** turn.

**counter-clockwise**

\[ p_3 \text{ at the left of } \overrightarrow{p_1p_2} \]

\[ \alpha < \beta \]

**clockwise**

\[ p_3 \text{ at the right of } \overrightarrow{p_1p_2} \]

\[ \alpha > \beta \]

// Returns true if \( p_3 \) is at the left of \( \overrightarrow{p_1p_2} \)
bool leftof(p1, p2, p3) {
    return (p2.x - p1.x) * (p3.y - p1.y) > (p2.y - p1.y) * (p3.x - p1.x);
}

© Dept. CS, UPC
Convex hull: gift wrapping algorithm

Convex hull: gift wrapping algorithm

- **Input:** \(p_1, p_2, \ldots, p_n\) (points in the plane).
- **Output:** P (the convex hull of \(p_1, p_2, \ldots, p_n\)).

- **Initial points:**
  \(p_0\) with the smallest \(x\) coordinate.

- **Iteration:** Assume that a partial path with \(k\) points has been built (\(p_k\) is the last point). Pick some arbitrary \(p_{k+1} \neq p_k\). Visit the remaining points. If some point \(q\) is at the left of \(p_k p_{k+1}\) redefine \(p_{k+1} = q\).

- Stop when P is complete (back to point \(p_0\)).

**Complexity:** At each iteration, we calculate \(n\) angles. \(T(n) = O(hn)\), where \(h\) is the number of points in the convex hull. In the worst case, \(T(n) = O(n^2)\).
Polygon convexHull(const vector<Point>& points) {
    int n = points.size();
    Polygon hull;

    // Pick the leftmost point
    int left = 0;
    for (int i = 1; i < n; i++)
        if (points[i].x < points[left].x) left = i;

    int p = left;
    do {
        hull.push_back(points[p]); // Add point to the convex hull

        int q = (p + 1)%n; // Pick a point different from p
        for (int i = 0; i < n; i++)
            if (leftof(points[p], points[q], points[i])) q = i;

        p = q; // Leftmost point for the convex hull
    } while (p != left); // While not closing polygon

    return hull;
}
Convex hull: Graham scan

**Input:** \(p_1, p_2, \ldots, p_n\) (points in the plane).

**Output:** \(q_1, q_2, \ldots, q_m\) (the convex hull).

**Initially:**
Create a simple polygon \(P\) (complexity \(O(n \log n)\)).
Assume the order of the points is \(p_1, p_2, \ldots, p_n\).

// \(Q = (q_1, q_2, \ldots)\) is a vector where the points // of the convex hull will be stored.
\[
q_1 = p_1; \ q_2 = p_2; \ q_3 = p_3; \ m = 3;
\]
for \(k = 4\) to \(n\):
\[
\text{while leftof}(q_{m-1}, q_m, p_k): \ m = m - 1;
\]
\[
m = m + 1;
\]
\[
q_m = p_k;
\]

**Observation:** each point \(p_k\) can be included in \(Q\) and deleted at most once.
The main loop of Graham scan has linear cost.

**Complexity:** dominated by the creation of the simple polygon \(\rightarrow O(n \log n)\).
EXERCISES
Summations

Prove the following equalities:

\[ \sum_{i=1}^{n} i = \frac{n(n + 1)}{2} \]

\[ \sum_{i=1}^{n} i^2 = \frac{n(n + 1)(2n + 1)}{6} \]

\[ \sum_{i=0}^{n} 2^i = 2^{n+1} - 1 \]
For loops: analyze the cost of each code

// Code 1
int s = 0;
for (int i = 0; i < n; ++i) ++s;

// Code 2
int s = 0;
for (int i = 0; i < n; i += 2) ++s;

// Code 3
int s = 0;
for (int i = 0; i < n; ++i) ++s;
for (int j = 0; j < n; ++j) ++s;

// Code 4
int s = 0;
for (int i = 0; i < n; ++i) {
    for (int j = 0; j < n; ++j) ++s;
}

// Code 5
int s = 0;
for (int i = 0; i < n; ++i) {
    for (int j = 0; j < i; ++j) ++s;
}
For loops: analyze the cost of each code

// Code 6
int s = 0;
for (int i = 0; i < n; ++i) {
    for (int j = i; j < n; ++j) ++s;
}

// Code 7
int s = 0;
for (int i = 0; i < n; ++i) {
    for (int j = 0; j < n; ++j) {
        for (int k = 0; k < n; ++k) ++s;
    }
}

// Code 8
int s = 0;
for (int i = 0; i < n; ++i) {
    for (int j = 0; j < i; ++j) {
        for (int k = 0; k < j; ++k) ++s;
    }
}
For loops: analyze the cost of each code

// Code 9
int s = 0;
for (int i = 1; i <= n; i *= 2) ++s;

// Code 10
int s = 0;
for (int i = 0; i < n; ++i) {
    for (int j = 0; j < i*i; ++j) {
        for (int k = 0; k < n; ++k) ++s;
    }
}

// Code 11
int s = 0;
for (int i = 0; i < n; ++i) {
    for (int j = 0; j < i*i; ++j) {
        if (j%i == 0) {
            for (int k = 0; k < n; ++k) ++s;
        }
    }
}
The following statements refer to the *insertion sort* algorithm and the $X$’s hide an occurrence of $O$, $\Omega$ or $\Theta$. For each statement, find which options for $X \in \{O, \Omega, \Theta\}$ make the statement true or false. Justify your answers.

1. The worst case is $X(n^2)$
2. The worst case is $X(n)$
3. The best case is $X(n^2)$
4. The best case is $X(n)$
5. For every probability distribution, the average case is $X(n^2)$
6. For every probability distribution, the average case is $X(n)$
7. For some probability distribution, the average case is $X(n \log n)$
The following algorithms try to determine whether \( n \geq 0 \) is prime. Find which ones are correct and analyze their cost as a function of \( n \).

```c
bool isPrime1(int n) {
    if (n <= 1) return false;
    for (int i = 2; i < n; ++i) if (n%i == 0) return false;
    return true;
}

bool isPrime2(int n) {
    if (n <= 1) return false;
    for (int i = 2; i*i < n; ++i) if (n%i == 0) return false;
    return true;
}

bool isPrime3(int n) {
    if (n <= 1) return false;
    for (int i = 2; i*i <= n; ++i) if (n%i == 0) return false;
    return true;
}

bool isPrime4(int n) {
    if (n <= 1) return false;
    if (n == 2) return true;
    if (n%2 == 0) return false;
    for (int i = 3; i*i <= n; i += 2) if (n%i == 0) return false;
    return true;
}
```
The Sieve of Eratosthenes

The following program is a version of the Sieve of Eratosthenes. Analyze its complexity.

```cpp
vector<bool> Primes(int n) {
    vector<bool> p(n + 1, true);
p[0] = p[1] = false;
    for (int i = 2; i*i <= n; ++i) {
        if (p[i]) {
            for (int j = i*i; j <= n; j += i) p[j] = false;
        }
    }
    return p;
}
```

You can use the following equality, where $p \leq x$ refers to all primes $p \leq x$:

$$
\sum_{p \leq x} \frac{1}{p} = \log \log x + O(1)
$$
The Cell Phone Dropping Problem

• You work for a cell phone company which has just invented a new cell phone protector and wants to advertise that it can be dropped from the $f^{th}$ floor without breaking.

• If you are given 1 or 2 phones and an $n$ story building, propose an algorithm that minimizes the worst-case number of trial drops to know the highest floor it won’t break.

• Assumption: a broken cell phone cannot be used for further trials.

• How about if you have $p$ cell phones?

(Source: Wood & Yasskin, Texas A&M University)
Divide & Conquer (I)

Jordi Cortadella and Jordi Petit
Department of Computer Science
Divide-and-conquer algorithms

• Strategy:
  – Divide the problem into smaller subproblems of the same type of problem
  – Solve the subproblems recursively
  – Combine the answers to solve the original problem

• The work is done in three places:
  – In partitioning the problem into subproblems
  – In solving the basic cases at the tail of the recursion
  – In merging the answers of the subproblems to obtain the solution of the original problem
Conventional product of polynomials

Example:

\[ P(x) = 2x^3 + x^2 - 4 \]

\[ Q(x) = x^2 - 2x + 3 \]

\[ (P \cdot Q)(x) = 2x^5 + (-4 + 1)x^4 + (6 - 2)x^3 + 8x - 12 \]

\[ (P \cdot Q)(x) = 2x^5 - 3x^4 + 4x^3 + 8x - 12 \]
Conventional product of polynomials

```plaintext
function PolynomialProduct(P, Q)
    // P and Q are vectors of coefficients.
    // Returns R = P × Q.
    // degree(P) = size(P)-1, degree(Q) = size(Q)-1.
    // degree(R) = degree(P)+degree(Q).

    R = vector with size(P)+size(Q)-1 zeros;

    for each P_i
        for each Q_j
            R_{i+j} = R_{i+j} + P_i \cdot Q_j

    return R
```

Complexity analysis:
- Multiplication of polynomials of degree n: \( O(n^2) \)
- Addition of polynomials of degree n: \( O(n) \)
Product of polynomials: Divide & Conquer

Assume that we have two polynomials with \( n \) coefficients (degree \( n - 1 \))

\[
\begin{array}{c|ccc|c}
& \text{n-1} & \text{n/2} & \text{0} \\
\hline
\text{P} : & \text{P}_L & \text{P}_R \\
\text{Q} : & \text{Q}_L & \text{Q}_R \\
\end{array}
\]

\[
P(x) \cdot Q(x) = P_L(x) \cdot Q_L(x) \cdot x^n + \\
( P_R(x) \cdot Q_L(x) + P_L(x) \cdot Q_R(x) ) \cdot x^{n/2} + \\
P_R(x) \cdot Q_R(x)
\]

\[
T(n) = 4 \cdot T(n/2) + O(n) = O(n^2) \quad \text{← Shown later}
\]
Product of complex numbers

• The product of two complex numbers requires four multiplications:

\[(a + bi)(c + di) = ac - bd + (bc + ad)i\]

• Carl Friedrich Gauss (1777-1855) noticed that it can be done with just three: \(ac, bd\) and \((a + b)(c + d)\)

\[bc + ad = (a + b)(c + d) - ac - bd\]

• A similar observation applies for polynomial multiplication.
Product of polynomials with Gauss’s trick

\[ R_1 = P_L Q_L \]
\[ R_2 = P_R Q_R \]
\[ R_3 = (P_L + P_R)(Q_L + Q_R) \]

\[ PQ = \underbrace{P_L Q_L \, x^n}_{R_1} + \underbrace{(P_R Q_L + P_L Q_R) \, x^{n/2}}_{R_3-R_1-R_2} + \underbrace{P_R Q_R}_{R_2} \]

\[ T(n) = 3T(n/2) + O(n) \]
Polynomial multiplication: recursive step

\[ P \times Q = P_L Q_R + P_R Q_L \]

\[ P = \begin{array}{cccccc} 1 & -2 & 3 & 2 & 0 & -1 \\ Q = \begin{array}{cccc} 2 & 1 & 0 & -1 \\ & 3 & 0 \end{array} \]

\[ P_L = \begin{array}{cccc} 1 & -2 & 3 \\ 2 & -3 & 4 & 3 & 0 \end{array} \]
\[ Q_L = \begin{array}{cccc} 2 & 1 & 0 \\ 2 & -3 & 4 & 3 & 0 \end{array} \]
\[ Q_R = \begin{array}{cccc} -1 & 3 & 0 \\ -2 & 6 & 1 & -3 & 0 \end{array} \]

\[ P_L + P_R = \begin{array}{cccc} 3 & -2 & 2 \\ 3 & 10 & -6 & 8 \end{array} \]
\[ Q_L + Q_R = \begin{array}{cccc} 1 & 4 & 0 \\ 3 & 10 & -6 & 8 \end{array} \]
\[ P_L Q_R + P_R Q_L = \begin{array}{cccc} 3 & 7 & -11 & 8 \\ 2 & -3 & 4 & 3 & 0 \end{array} \]
Pattern of recursive calls

- **Divide & Conquer**
- 

![Diagram](image)

- **Branching factor**: 3

- **Equation**: $n/2$

- **Levels**: $\log_2 n$

- **Values**:
  - 2
  - 1 1 1

© Dept. CS, UPC
Useful reminders

• Sum of geometric series with ratio $r$:

$$S = a + ar + ar^2 + ar^3 + \cdots + ar^{n-1}$$

$$S = a \left( \frac{1 - r^n}{1 - r} \right) = \frac{a}{1 - r} + \frac{r}{r - 1} ar^{n-1}$$

• Logarithms:

$$\log_b n = \log_b a \cdot \log_a n$$

$$a^{\log_b n} = a^{(\log_a n)(\log_b a)} = n^{\log_b a}$$
Complexity analysis

• The time spent at level $k$ is

$$3^k \cdot O\left(\frac{n}{2^k}\right) = \left(\frac{3}{2}\right)^k \cdot O(n)$$

• For $k = 0$, runtime is $O(n)$.

• For $k = \log_2 n$, runtime is $O(3^{\log_2 n})$, which is equal to $O(n^{\log_2 3})$.

• The runtime per level increases geometrically by a factor of $3/2$ per level. The sum of any increasing geometric series is, within a constant factor, simply the last term of the series.

• Therefore, the complexity is $O(n^{1.59})$. 
A popular recursion tree

Branching factor: 2

Example: efficient sorting algorithms.

\[ T(n) = 2 \cdot T\left(\frac{n}{2}\right) + O(n) \]

Algorithms may differ on the amount of work done at each level: \( O(n^c) \)
## Examples

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Branch</th>
<th>c</th>
<th>Runtime equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power ($x^y$)</td>
<td>1</td>
<td>0</td>
<td>$T(y) = T(y/2) + O(1)$</td>
</tr>
<tr>
<td>Binary search</td>
<td>1</td>
<td>0</td>
<td>$T(n) = T(n/2) + O(1)$</td>
</tr>
<tr>
<td>Merge sort</td>
<td>2</td>
<td>1</td>
<td>$T(n) = 2 \cdot T(n/2) + O(n)$</td>
</tr>
<tr>
<td>Polynomial product</td>
<td>4</td>
<td>1</td>
<td>$T(n) = 4 \cdot T(n/2) + O(n)$</td>
</tr>
<tr>
<td>Polynomial product (Gauss)</td>
<td>3</td>
<td>1</td>
<td>$T(n) = 3 \cdot T(n/2) + O(n)$</td>
</tr>
</tbody>
</table>
Master theorem

• Typical pattern for Divide&Conquer algorithms:
  – Split the problem into $a$ subproblems of size $n/b$
  – Solve each subproblem recursively
  – Combine the answers in $O(n^c)$ time

• Running time:  \[ T(n) = a \cdot T(n/b) + O(n^c) \]

• Master theorem:
  \[
  T(n) = \begin{cases} 
    O(n^c) & \text{if } c > \log_b a \quad (a < b^c) \\
    O(n^c \log n) & \text{if } c = \log_b a \quad (a = b^c) \\
    O(n^{\log_b a}) & \text{if } c < \log_b a \quad (a > b^c) 
  \end{cases}
  \]
Master theorem: recursion tree

Size $n$

Size $n/b$

Size $n/b^2$

Branching factor $a$

Depth $\log_b n$

Width $a^{\log_b n} = n^{\log_b a}$
Master theorem: proof

• For simplicity, assume \( n \) is a power of \( b \).
• The base case is reached after \( \log_b n \) levels.
• The \( k \)th level of the tree has \( a^k \) subproblems of size \( n/b^k \).
• The total work done at level \( k \) is:

\[
a^k \times O\left(\frac{n}{b^k}\right)^c = O(n^c) \times \left(\frac{a}{b^c}\right)^k
\]

• As \( k \) goes from 0 (the root) to \( \log_b n \) (the leaves), these numbers form a geometric series with ratio \( a/b^c \). We need to find the sum of such a series.

\[
T(n) = O(n^c) \cdot \left(1 + \frac{a}{b^c} + \frac{a^2}{b^{2c}} + \frac{a^3}{b^{3c}} + \cdots + \frac{a^{\log_b n}}{b^{(\log_b n)c}}\right)
\]

\( \log_b n \) terms
Master theorem: proof

- Case $a/b^c < 1$. Decreasing series. The sum is dominated by the first term ($k = 0$): $O(n^c)$.

- Case $a/b^c > 1$. Increasing series. The sum is dominated by the last term ($k = \log_b n$):

$$n^c \left( \frac{a}{b^c} \right)^{\log_b n} = n^c \left( \frac{a^{\log_b n}}{(b^{\log_b n})^c} \right) = a^{\log_b n} = n^{\log_b a}$$

- Case $a/b^c = 1$. We have $O(\log n)$ terms all equal to $O(n^c)$. 

Divide & Conquer © Dept. CS, UPC
Master theorem: visual proof

\[ a > b^c \]

\[ a^{\log_b n} \]

\[ n^{c} \]

\[ n^{\log_b a} \]

\[ n^{c} \]

\[ n^{c} \log n \]

\[ \log_b n \text{ levels} \]

\[ a < b^c \]

\[ n^{c} \]

\[ a = b^c \]

\[ n^{c} \]

\[ n^{c} \]

\[ \log_b n \text{ levels} \]
Master theorem: examples

Running time: \( T(n) = a \cdot T(n/b) + O(n^c) \)

\[
T(n) = \begin{cases} 
O(n^c) & \text{if } a < b^c \\
O(n^c \log n) & \text{if } a = b^c \\
O(n^{\log_b a}) & \text{if } a > b^c
\end{cases}
\]

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>a</th>
<th>c</th>
<th>Runtime equation</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power ((x^y))</td>
<td>1</td>
<td>0</td>
<td>(T(y) = T(y/2) + O(1))</td>
<td>(O(\log y))</td>
</tr>
<tr>
<td>Binary search</td>
<td>1</td>
<td>0</td>
<td>(T(n) = T(n/2) + O(1))</td>
<td>(O(\log n))</td>
</tr>
<tr>
<td>Merge sort</td>
<td>2</td>
<td>1</td>
<td>(T(n) = 2 \cdot T(n/2) + O(n))</td>
<td>(O(n \log n))</td>
</tr>
<tr>
<td>Polynomial product</td>
<td>4</td>
<td>1</td>
<td>(T(n) = 4 \cdot T(n/2) + O(n))</td>
<td>(O(n^2))</td>
</tr>
<tr>
<td>Polynomial product (Gauss)</td>
<td>3</td>
<td>1</td>
<td>(T(n) = 3 \cdot T(n/2) + O(n))</td>
<td>(O(n^{\log_2 3}))</td>
</tr>
</tbody>
</table>

\(b = 2\) for all the examples
Divide & Conquer (II)

Jordi Cortadella and Jordi Petit
Department of Computer Science
Examples

- Quick sort
- The selection problem
- The closest-points problem
Quick sort (Tony Hoare, 1959)

• Suppose that we know a number $x$ such that one-half of the elements of a vector are greater than or equal to $x$ and one-half of the elements are smaller than $x$.
  – Partition the vector into two equal parts ($n-1$ comparisons)
  – Sort each part recursively

• Problem: we do not know $x$.

• The algorithm also works no matter which $x$ we pick for the partition. We call this number the pivot.

• Observation: the partition may be unbalanced.
Quick sort with Hungarian, folk dance
The key step of quick sort is the partitioning algorithm.

**Question:** how to find a good pivot?
Quick sort

https://en.wikipedia.org/wiki/Quicksort
Quick sort: partition

```plaintext
function Partition(A, left, right)
    // A[left..right]: segment to be sorted
    // Returns the middle of the partition with
    //   A[middle] = pivot
    //   A[left..middle-1] ≤ pivot
    //   A[middle+1..right] > pivot
    x = A[left];    // the pivot
    i = left; j = right;

    while i < j do
        while i ≤ right and A[i] ≤ x do i = i+1;
        while j ≥ left and A[j] > x do j = j-1;
        if i < j then swap(A[i], A[j]);

    swap(A[left], A[j]);
    return j;
```

Divide & Conquer © Dept. CS, UPC
Quick sort partition: example

pivot
→
i             ← j

6 2 8 5 10 9 12 1 15 7 3 13 4 11 16 14
6 2 4 5 10 9 12 1 15 7 3 13 8 11 16 14
6 2 4 5 3 9 12 1 15 7 10 13 8 11 16 14
6 2 4 5 3 1 12 9 15 7 10 13 8 11 16 14
1 2 4 5 3 6 12 9 15 7 10 13 8 11 16 14
1 2 4 5 3 6 12 9 15 7 10 13 8 11 16 14

middle

Divide & Conquer
© Dept. CS, UPC
function Qsort(A, left, right)

// A[left..right]: segment to be sorted

if left < right then
    mid = Partition(A, left, right);
    Qsort(A, left, mid-1);
    Qsort(A, mid+1, right);
function HoarePartition(A, left, right)

// A[left..right]: segment to be sorted.
// Output: The left part has elements \leq \text{ than the pivot.}
// The right part has elements \geq \text{ than the pivot.}
// Returns the index of the last element of the left part.

x = A[left]; // the pivot
i = left-1; j = right+1;

while true do
    do i = i+1; while A[i] < x;
    do j = j-1; while A[j] > x;

    if i \geq j then return j;

    swap(A[i], A[j]);

Admire a unique piece of art by Hoare:
The first swap creates two sentinels.
After that, the algorithm flies ...
Quick sort partition: example

First swap: 4 is a sentinel for R; 6 is a sentinel for L → no need to check for boundaries

pivot

6 2 8 5 10 9 12 1 15 7 3 13 4 11 16 14

4 2 8 5 10 9 12 1 15 7 3 13 6 11 16 14

i → j

4 2 3 5 10 9 12 1 15 7 8 13 6 11 16 14

i

4 2 3 5 1 9 12 10 15 7 8 13 6 11 16 14

j

4 2 3 5 1 9 12 10 15 7 8 13 6 11 16 14

j (middle)
Quick sort with Hoare’s partition

function Qsort(A, left, right)

// A[left..right]: segment to be sorted

if left < right then
    mid = HoarePartition(A, left, right);
    Qsort(A, left, mid);
    Qsort(A, mid+1, right);
function Qsort(A, left, right)
    // A[left..right]: segment to be sorted.
    // K is a break-even size in which insertion sort is
    // more efficient than quick sort.
    if right - left ≥ K then
        mid = HoarePartition(A, left, right);
        Qsort(A, left, mid);
        Qsort(A, mid+1, right);

function Sort(A):
    Qsort(A, 0, A.size()-1);
    InsertionSort(A);
Quick sort: complexity analysis

• The partition algorithm is $O(n)$. 

• Assume that the partition is balanced:
  
  $$T(n) = 2 \cdot T(n/2) + O(n) = O(n \log n)$$

• Worst case runtime: the pivot is always the smallest element in the vector $\Rightarrow O(n^2)$

• Selecting a good pivot is essential. There are different strategies, e.g.,
  – Take the median of the first, last and middle elements
  – Take the pivot at random
Quick sort: complexity analysis

• Let us assume that $x_i$ is the $i$-th smallest element in the vector.

• Let us assume that each element has the same probability of being selected as pivot.

• The runtime if $x_i$ is selected as pivot is:

$$T(n) = n - 1 + T(i - 1) + T(n - i)$$
Quick sort: complexity analysis

\[ T(n) = n - 1 + \frac{1}{n} \sum_{i=1}^{n} (T(i - 1) + T(n - i)) \]

\[ T(n) = n - 1 + \frac{1}{n} \sum_{i=1}^{n} T(i - 1) + \frac{1}{n} \sum_{i=1}^{n} T(n - i) \]

\[ T(n) = n - 1 + \frac{2}{n} \sum_{i=0}^{n-1} T(i) \leq 2(n + 1)(H(n + 1) - 1.5) \]

\[ H(n) = 1 + 1/2 + 1/3 + \cdots + 1/n \] is the Harmonic series, that has a simple approximation: \( H(n) = \ln n + \gamma + O(1/n) \).
\[ \gamma = 0.577 \ldots \] is Euler’s constant.

\[ T(n) \leq 2(n + 1)(\ln n + \gamma - 1.5) + O(1) \in O(n \log n) \]
Quick sort: complexity analysis summary

• Runtime of quicksort:

\[ T(n) = O(n^2) \]
\[ T(n) = \Omega(n \log n) \]
\[ T_{\text{avg}}(n) = O(n \log n) \]

• Be careful: some malicious patterns may increase the probability of the worst case runtime, e.g., when the vector is sorted or almost sorted.

• Possible solution: use random pivots.
The selection problem

- Given a collection of $N$ elements, find the $k$-th smallest element.

- Options:
  - Sort a vector and select the $k$-th location: $O(N \log N)$
  - Read $k$ elements into a vector and sort them. The remaining elements are processed one by one and placed in the correct location (similar to insertion sort). Only $k$ elements are maintained in the vector. Complexity: $O(kN)$. Why?
Quick sort with Hoare’s partition

function Qsort(A, left, right)

// A[left..right]: segment to be sorted

if left < right then
    mid = HoarePartition(A, left, right);
    Qsort(A, left, mid);
    Qsort(A, mid+1, right);
Quick select with Hoare’s partition

// Returns the element at location \( k \) assuming
// \( A[left..right] \) would be sorted in ascending order.
// Pre: \( left \leq k \leq right \).
// Post: The elements of \( A \) have changed their locations.

function Qselect(A, left, right, k)
    if left == right then return A[left];

    mid = HoarePartition(A, left, right);
    // We only need to sort one half of \( A \)
    if k \leq mid then return Qselect(A, left, mid, k);
    else return Qselect(A, mid+1, right, k);
Quick Select: complexity

• Assume that the partition is balanced:
  – Quick sort: $T(n) = 2T(n/2) + O(n) = O(n \log n)$
  – Quick select: $T(n) = T(n/2) + O(n) = O(n)$

• The average linear time complexity can be achieved by choosing good pivots (similar strategy and complexity computation to qsort).
The Closest-Points problem

- **Input:** A list of \( n \) points in the plane 
  \[ \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \]
- **Output:** The pair of closest points
- **Simple approach:** check all pairs \( \rightarrow O(n^2) \)
- We want an \( O(n \log n) \) solution!
The Closest-Points problem

• We can assume that the points are sorted by the $x$-coordinate. Sorting the points is free from the complexity standpoint ($O(n \log n)$).

• Split the list into two halves. The closest points can be both at the left, both at the right or one at the left and the other at the right (center).

• The left and right pairs are easy to find (recursively). How about the pairs in the center?
The Closest-Points problem

• Let $\delta = \min(\delta_L, \delta_R)$. We only need to compute $\delta_C$ if it improves $\delta$.

• We can define a strip around the center with distance $\delta$ at the left and right. If $\delta_C$ improves $\delta$, then the points must be within the strip.

• In the worst case, all points can still reside in the strip.

• But how many points do we really have to consider?
The Closest-Points problem

Let us take all points in the strip and sort them by the \( y \)-coordinate. We only need to consider pairs of points with distance smaller than \( \delta \).

Once we find a pair \((p_i, p_j)\) with \( y \)-coordinates that differ by more than \( \delta \), we can move to the next \( p_i \).

\[
\text{for } (i=0; i < \text{NumPointsInStrip}; ++i) \\
\quad \text{for } (j=i+1; j < \text{NumPointsInStrip}; ++j) \\
\quad \quad \text{if } (p_i \text{ and } p_j \text{'s } y \text{-coordinate differ by more than } \delta) \text{ break; } // \text{ Go to next } p_i \\
\quad \quad \text{if } (\text{dist}(p_i, p_j) < \delta) \delta = \text{dist}(p_i, p_j) \;
\]

But, how many pairs \((p_i, p_j)\) do we need to consider?
The Closest-Points problem

• For every point $p_i$ at one side of the strip, we only need to consider points from $p_{i+1}$.

• The relevant points only reside in the $2\delta \times \delta$ rectangle below point $p_i$. There can only be 8 points at most in this rectangle (4 at the left and 4 at the right). Some points may have the same coordinates.
The Closest-Points problem: algorithm

• Sort the points according to their $x$-coordinates.

• Divide the set into two equal-sized parts.

• Compute the min distance at each part (recursively). Let $\delta$ be the minimal of the two minimal distances.

• Eliminate points that are farther than $\delta$ from the separation line.

• Sort the remaining points according to their $y$-coordinates.

• Scan the remaining points in the $y$ order and compute the distances of each point to its 7 neighbors.
The Closest-Points problem: complexity

• Initial sort using $x$-coordinates: $O(n \log n)$. It comes for free.

• Divide and conquer:
  – Solve for each part recursively: $2T(n/2)$
  – Eliminate points farther than $\delta$: $O(n)$
  – Sort remaining points using $y$-coordinates: $O(n \log n)$
  – Scan the remaining points in $y$ order: $O(n)$

$$T(n) = 2T(n/2) + O(n) + O(n \log n) = O(n \log^2 n)$$

• Can we do it in $O(n \log n)$? Yes, we need to sort by $y$ in a smart way.
The Closest-Points problem: complexity

- Let $Y$ a vector with the points sorted by the $y$-coordinates. This can be done initially for free.

- Each time we partition the set of points by the $x$-coordinate, we also partition $Y$ into two sorted vectors (using an “unmerging” procedure with linear complexity)

\[
Y_L = Y_R = \emptyset \quad // \text{Initial lists of points}
\]

```cpp
foreach \(p_i \in Y\) in ascending order of \(y\) do
  if \(p_i\) is at the left part then \(Y_L\).push_back\(p_i\)
  else \(Y_R\).push_back\(p_i\)
```

- Now, sorting the points by the $y$-coordinate at each iteration can be done in linear time, and the problem can be solved in $O(n \log n)$
Sometimes we may find recurrences with the following structure:

$$T(n) = a \cdot T(n - b) + O(n^c)$$

Examples:

Hanoi(n) = 2 \cdot Hanoi(n - 1) + O(1)

Sort(n) = Sort(n - 1) + O(n)

Muster theorem:

$$T(n) = \begin{cases} 
0(n^c) & \text{if } a < 1 \quad \text{(never occurs)} \\
0(n^{c+1}) & \text{if } a = 1 \\
0(n^c a^{n/b}) & \text{if } a > 1
\end{cases}$$
Muster theorem: recursion tree

Size $n$

Size $n - b$

Size $n - 2b$

Branching factor $a$

Depth $n/b$

Width $a^{n/b}$

Size 1
Muster theorem: examples

• **Hanoi:** \( T(n) = 2T(n - 1) + O(1) \)

We have \( a = 2 \) and \( c = 0 \), thus \( T(n) = O(2^n) \).

• **Selection sort** (recursive version):
  – Select the min element and move it to the first location
  – Sort the remaining elements

\[
T(n) = T(n - 1) + O(n) \quad (a = c = 1)
\]

Thus, \( T(n) = O(n^2) \)
Muster theorem: examples

**Fibonacci:** \( T(n) = T(n - 1) + T(n - 2) + O(1) \)

We can compute bounds:

\[
2T(n - 2) + O(1) \leq T(n) \leq 2T(n - 1) + O(1)
\]

Thus,

\[
O\left(2^{n/2}\right) \leq T(n) \leq O(2^n)
\]
EXERCICES
The skyline problem

Given the exact locations and shapes of several rectangular buildings in a city, draw the skyline (in two dimensions) of these buildings, eliminating hidden lines (source: Udi Manber, *Introduction to Algorithms*, Addison-Wesley, 1989).

**Input:**

\[(1,11,5) \ (2,6,7) \ (3,13,9) \ (12,7,16) \ (14,3,25) \]
\[(19,18,22) \ (23,13,29) \ (24,4,28) \]

**Output:**

\[(1,11,3,13,9,0,12,7,16,3,19,18,22,3,23,13,29,0) \]

(numbers in boldface represent heights)

Describe (in natural language) two different algorithms to solve the skyline problem:

- By induction: assume that you know how to solve it for \( n - 1 \) buildings.
- Using Divide&Conquer: solve the problem for \( n/2 \) buildings and combine.

Analyze the cost of each solution.
A, B or C?

Suppose you are choosing between the following three algorithms:

• Algorithm A solves problems by dividing them into five subproblems of half the size, recursively solving each subproblem, and then combining the solutions in linear time.

• Algorithm B solves problems of size $n$ by recursively solving two subproblems of size $n - 1$ and then combining the solutions in constant time.

• Algorithm C solves problems of size $n$ by dividing them into nine subproblems of size $n/3$, recursively solving each subproblem, and then combining the solutions in $O(n^2)$ time.

What are the running times of each of these algorithms (in big-O notation), and which one would you choose?

Crazy sorting

Let $T[i..j]$ be a vector with $n = j - i + 1$ elements. Consider the following sorting algorithm:

a) If $n \leq 2$ the vector is easily sorted (constant time).

b) If $n \geq 3$, divide the vector into three intervals $T[i..k - 1]$, $T[k..l]$ and $T[l + 1..j]$, where $k = i + \lfloor n/3 \rfloor$ and $l = j - \lfloor n/3 \rfloor$. The algorithm recursively sorts $T[i..l]$, then it sorts $T[k..j]$, and finally sorts $T[i..l]$.

• Proof the correctness of the algorithm.
• Analyze the asymptotic complexity of the algorithm (give a recurrence of the runtime and solve it).
A majority element in a vector, $A$, of size $n$ is an element that appears more than $n/2$ times (thus, there is at most one). For example, the vector $[3, 3, 4, 2, 4, 2, 4, 4]$ has a majority element ($4$), whereas the vector $[3, 3, 4, 2, 4, 4, 2, 2]$ does not. If there is no majority element, your program should indicate this. Here is a sketch of an algorithm to solve the problem:

First, a candidate majority element is found (this is the hardest part). This candidate is the only element that could possibly be the majority element. The second step determines if this candidate is actually the majority. This is just a sequential search through the vector. To find a candidate in the vector, $A$, form a second vector, $B$. Then compare $A_0$ and $A_1$. If they are equal, add one of these to $B$; otherwise do nothing. Then compare $A_2$ and $A_3$. Again if they are equal, add one of these to $B$; otherwise do nothing. Continue in this fashion until the entire vector is read. The recursively find a candidate for $B$; this is the candidate for $A$ (why?).

• How does the recursion terminate?
• What is the running time of the algorithm?
• How can we avoid using an extra array, $B$?
• Prove the correctness of the algorithm (hint: prove it for $n$ even)
• How is the case where $n$ is odd handled?

Breaking into pieces

Let us assume that \( f \) is \( \Theta(1) \) and \( g \) has a runtime proportional to the size of the vector it has to process, i.e., \( \Theta(j - i + 1) \). What is the asymptotic cost of \( A \) and \( B \) as a function of \( n \)? \( n \) is the size of the vector.

If both functions do the same, which one would you choose?

double A(vector<double>& v, int i, int j) {
    if (i < j) {
        int x = f(v, i, j);
        int m = (i+j)/2;
        return A(v, i, m-1) + A(v, m, j) + A(v, i+1, m) + x;
    } else {
        return v[i];
    }
}

double B(vector<double>& v, int i, int j) {
    if (i < j) {
        int x = g(v, i, j);
        int m1 = i + (j-i+1)/3;
        int m2 = i + (j-i+1)*2/3;
        return B(v, i, m1-1) + B(v, m1, m2-1) + B(v, m2, j) + x;
    } else {
        return v[i];
    }
}
APPENDIX
## Logarithmic identities

<table>
<thead>
<tr>
<th>Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>( b^{\log_b a} = \log_b b^a = a )</td>
</tr>
<tr>
<td>( \log_b(xy) = \log_b x + \log_b y )</td>
</tr>
<tr>
<td>( \log_b \frac{x}{y} = \log_b x - \log_b y )</td>
</tr>
<tr>
<td>( \log_b x^c = c \log_b x )</td>
</tr>
<tr>
<td>( \log_b x = \frac{\log_c x}{\log_c b} )</td>
</tr>
<tr>
<td>( x^{\log_b y} = y^{\log_b x} )</td>
</tr>
</tbody>
</table>

\[ H(n) = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots + \frac{1}{n} \]

\[ \gamma = \lim_{n \to \infty} \left( -\ln n + \sum_{k=1}^{n} \frac{1}{k} \right) \]

\[ \gamma = 0.5772 \ldots \text{ (Euler-Mascheroni constant)} \]

\[ \sum_{k=1}^{n} \frac{1}{k} \in \Theta(\log n) \text{ (Harmonic series)} \]
A recurrence that depends on all the previous values of the function.

\[ T(n) = n - 1 + \sum_{i=0}^{n-1} \frac{2}{n} T(i) \]

\[ nT(n) = n(n - 1) + 2 \sum_{i=0}^{n-1} T(i), \quad (n + 1)T(n + 1) = (n + 1)n + 2 \sum_{i=0}^{n} T(i) \]

\[ (n + 1)T(n + 1) - nT(n) = (n + 1)n - n(n - 1) + 2T(n) = 2n + 2T(n) \]

\[ T(n + 1) = \frac{n + 2}{n + 1} T(n) + \frac{2n}{n + 1} \leq \frac{n + 2}{n + 1} T(n) + 2 \]

\[ T(n) \leq 2 + \frac{n + 1}{n} \left( 2 + \frac{n}{n - 1} \left( 2 + \frac{n - 1}{n - 2} \left( \cdots \frac{4}{3} \right) \right) \right) \]

\[ T(n) \leq 2 \left( 1 + \frac{n + 1}{n} + \frac{n + 1}{n - 1} + \frac{n + 1}{n - 2} + \cdots + \frac{n + 1}{n - 1} \frac{n}{n - 2} \cdots \frac{4}{3} \right) \]

\[ T(n) \leq 2 \left( 1 + \frac{n + 1}{n} + \frac{n + 1}{n - 1} + \frac{n + 1}{n - 2} + \cdots + \frac{n + 1}{3} \right) = 2(n + 1) \left( \frac{1}{n + 1} + \frac{1}{n} + \frac{1}{n - 1} + \cdots + \frac{1}{3} \right) \]

\[ T(n) \leq 2(n + 1)(H(n + 1) - 1.5) \in \Theta(n \log n) \]
Muster theorem: proof

- Expanding the recursion (assume that $f(n)$ is $O(n^c)$)

$$T(n) = aT(n - b) + f(n)$$
$$= a(aT(n - 2b) + f(n - b)) + f(n)$$
$$= a^2T(n - 2b) + af(n - b) + f(n)$$
$$= a^3T(n - 3b) + a^2f(n - 2b) + af(n - b) + f(n)$$

- Hence:

$$T(n) = \sum_{i=0}^{\frac{n}{b}} a^i \cdot f(n - ib)$$

- Since $f(n - ib)$ is in $O((n - ib)^c)$, which is in $O(n^c)$, then

$$T(n) = O\left(n^c \sum_{i=0}^{\frac{n}{b}} a^i\right)$$

- The proof is completed by this property:

$$\sum_{i=0}^{\frac{n}{b}} a^i = \begin{cases} 
0(1), & \text{if } a < 1 \\
0(n), & \text{if } a = 1 \\
0(a^{n/b}), & \text{if } a > 1 
\end{cases}$$
Memory management

Jordi Cortadella and Jordi Petit
Department of Computer Science
The memory

Address

1036 0 0 1 0 0 1 0 0 1 1 1 0 0 0 1 0 1 0 1 0 0 1 0 0 1 0 1 0 0 0 1 0
1040
1044 0 0 1 0 0 1 0 0 0 1 1 1 0 0 0 0 0 0 1 1 0 1 0 0 1 0 1 0 1 0 0 0 0 0
1048 1 0 0 1 0 0 1 1
1052 0 1 0 0 1 0 1 0 1 0 0 0 0 1 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 1 0 0
1056 0 0 1 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 0 1 0 0 1 0 0 1 0 0 0 0 0 1
1060 H e l l
1064 o r l l
1068
1072 0 1 1 1 0 0 1 1 0 1 0 0 1 0 0 0 1 0 0 0 1 0 1 0 1 0 1 0 0 0 0 0 0 0 0
1076
1080 0 0 1 0 1 1 0 1 0 0 0 0 0 0 1 0 0 0 0 1 1 1 1 1 0 0 1 0 0 0 1 0 0
1084

int

byte

Hello world

Memory management

© Dept. CS, UPC
• Given a type $T$, $T^*$ is another type called “pointer to $T$”. It holds the memory address of an object of type $T$.

• Examples:

```cpp
int* pi;
myClass* pC;
```

• Pointers are often used to manage the memory space of dynamic data structures.

• In some cases, it might be convenient to obtain the address of a variable during runtime.
Pointers

• Address of a variable (reference operator &):

```c
int i;
int* pi = &i;
// &i means: “the address of i”
```

• Access to the variable (dereference operator *)

```c
int j = *pi;
// j gets the value of the variable pointed by pi
```

• Null pointer (points to nowhere; useful for initialization)

```c
int* pi = nullptr;
```

Memory management
Dynamic Object Creation/Destruction

• The `new` operator returns a pointer to a newly created object:

```cpp
myClass* c = new myClass( );
myClass* c = new myClass{ }; // C++11
myClass* c = new myClass;
```

• The `delete` operator destroys an object through a pointer (deallocate the memory space associated to the object):

```cpp
delete c;  // c must be a pointer
```
Access to members through a pointer

The members of a class can be accessed through a pointer to the class via the -> operator:

```cpp
vector<int>* vp = new vector<int> (100);
...
vp->push_back(5);
...
int n = vp->size();
```
new and delete can also create/destroy arrays of objects

// c is a pointer to an array of objects
myClass* c = new myClass[10];

// c[i] refers to one of the elements of the array
c[i].some_method_of_myClass();

// deallocating the array
delete [] c;
References

• A reference defines a new name for an existing value (a synonym).

• References are not pointers, although they are usually implemented as pointers (memory references).

• Typical uses:
  – Avoiding copies (e.g., parameter passing)
  – Aliasing long names
  – Range for loops
auto & r = x[getIndex(a, b)].val;
...
r.defineValue(n); // avoids a long name for r
...

// avoids a copy of a large data structure
bigVector & V = myMatrix.getRow(i);
...

// An alternative for the following loop:
// for (int i =0; i < a.size(); ++i) ++a[i];

for (auto x: arr) ++x; // does not work (why?)

for (auto & x: arr) ++x; // it works!
Pointers vs. references

• A *pointer* holds the memory address of a variable. A *reference* is an alias (another name) for an existing variable.

• In practice, references are implemented as pointers.

• A pointer can be re-assigned any number of times, whereas a reference can only be assigned at initialization.

• Pointers can be NULL. References always refer to an object.

• You can have pointers to pointers. You cannot have references to references.

• You can use pointer arithmetic (e.g., &object+3). Reference arithmetic does not exist.
The Vector class

(an approximation to the STL vector class)
The Vector class

• The natural replacement of C-style arrays.

• Main advantages:
  – It can dynamically grow and shrink.
  – It is a *template* class (can handle any type T)
  – No need to take care of the allocated memory.
  – Data is allocated in a contiguous memory area.

• We will implement a **Vector** class, a simplified version of STL’s vector.

• Based on Weiss’ book *(4th edition)*, see Chapter 3.4.
The Vector class

A generic abstract class that can handle various datatypes. The template parameters determine the genericity of the class.

Example of declarations:
- Vector<int> V;
- Vector<polygon> Vp;
- Vector<Vector<double>> M;

```cpp
template<typename Object>
class Vector {
    public:
        ...
    private:
        ...
};
```
A Vector may allocate more memory than needed (size vs. capacity).

The memory must be reallocated when there is no enough capacity in the storage area.

The pointer stores the base memory address (location of objects[0]). Pointers can be used to allocate/free memory blocks via new/delete operators.
The Vector class

theSize: 3
theCapacity: 5

theSize: 4
theCapacity: 5

a.push_back(...)
Data structures usually have a descriptor (fixed size) and a storage area (variable size). Memory blocks cannot always be resized. They need to be reallocated and initialized with the old block. After that, the old block can be freed.

Programmers do not have to take care of memory allocation, but it is convenient to know the basics of memory management.

Beware of memory leaks, fragmentation, ...

After resizing B’s storage (e.g., B.push_back(...))
public:
    // Returns the size of the vector
    int size() const
    { return theSize; }

    // Is the vector empty?
    bool empty() const
    { return size() == 0; }

    // Adds an element to the back of the vector
    void push_back(const Object & x) {
        if (theSize == theCapacity) reserve(2*theCapacity + 1);
        objects[theSize++] = x;
    }

    // Removes the last element of the array
    void pop_back()
    { theSize--; }
The Vector class

```cpp
public:
    // Returns a const ref to the last element
    const Object& back() const
    { return objects[theSize - 1]; }

    // Returns a ref to the i-th element
    Object& operator[](int i)
    { return objects[i]; }

    // Returns a const ref to the i-th element
    const Object& operator[](int i) const
    { return objects[i]; }

    // Modifies the size of the vector (destroying
    // elements in case of shrinking)
    void resize(int newSize) {
        if (newSize > theCapacity) reserve(newSize);
        theSize = newSize;
    }
```
// Reserves space (to increase capacity)
void reserve(int newCapacity) {
    if (newCapacity <= theCapacity) return;

    // Allocate the new memory block
    Object* newArray = new Object[newCapacity];

    // Copy the old memory block
    for (int k = 0; k < theSize; ++k)
        newArray[k] = objects[k];

    theCapacity = newCapacity;
    // Swap pointers and free old memory block
    std::swap(objects, newArray);
    delete [] newArray;
}
Constructors, copies, assignments

MyClass a, b; // Constructor is used

MyClass c = a; // Copy constructor is used

b = c; // Assignment operator is used

// Copy constructor is used when passing
// the argument to a function (or returning
// the value from a function)

void foo(Myclass x);

... foo(c); // Copy constructor is used
The Vector class

// Default constructor with initial size
Vector(int initSize = 0) : theSize{initSize},
    theCapacity{initSize + SPARE_CAPACITY}
    { objects = new Object[theCapacity]; }

// Copy constructor
Vector(const Vector& rhs) : theSize{rhs.theSize},
    theCapacity{rhs.theCapacity}, objects{nullptr} {
    objects = new Object[theCapacity];
    for (int k = 0; k < theSize; ++k) objects[k] = rhs.object[k];
}

// Assignment operator
Vector& operator=(const Vector& rhs) {
    if (this != &rhs) { // Avoid unnecessary copy
        theSize = rhs.theSize;
        theCapacity = rhs.theCapacity;
        delete [] objects;
        objects = new Object[theCapacity];
        for (int k = 0; k < theSize; ++k) objects[k] = rhs.object[k];
    }
    return *this;
}

// Destructor
Vector~() { delete [] objects; }
Memory layout of a program

<table>
<thead>
<tr>
<th>Region</th>
<th>Type of data</th>
<th>Lifetime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static</td>
<td>Global data</td>
<td>Lifetime of the program</td>
</tr>
<tr>
<td>Stack</td>
<td>Local variables of a function</td>
<td>Lifetime of the function</td>
</tr>
<tr>
<td>Heap</td>
<td>Dynamic data</td>
<td>Since created (new) until destroyed (delete)</td>
</tr>
</tbody>
</table>
int g(vector<int>& X) {
    int i;
    vector<double> Y;
    ...
}

void f() {
    int z;
    vector<int> A;
    ...
    z = g(A);
    ...
}

int main() {
    double r;
    ...
    f();
    ...
}
Memory management models

- **Programmer-controlled management:**
  - The programmer decides when to allocate (new) and deallocate (delete) blocks of memory.
  - Example: C++.
  - Pros: efficiency, memory management can be optimized.
  - Cons: error-prone (dangling references and memory leaks)

- **Automatic management:**
  - The program periodically launches a garbage collector that frees all non-referenced blocks of memory.
  - Examples: Java, Python, R.
  - Pros: the programmer does not need to worry about memory management.
  - Cons: cannot optimize memory management, less control over runtime.
Dangling references and memory leaks

```c++
myClass* A = new myClass;
myClass* B = new myClass;
// We have allocated space for two objects

A = B;
// A and B point at the same object!
// Possible memory leak (unreferenced memory space)

delete A;
// Now B is a dangling reference
// (points at free space)

delete B; // Error
```
Pointers and memory leaks

```c
for (i = 0; i < 1000000; ++i) {
    myClass* A = new Myclass; // 1Kbyte
    ...
    do_something(A, i);
    ...
    // forgets to delete A
}
// This loop produces a 1-Gbyte memory leak!
```

Recommendations:
- Do not use pointers, unless you are very desperate.
- If you have to use pointers, hide their usage inside a class and define consistent constructors/destructors.
- Use valgrind (valgrind.org) to detect memory leaks.
- Remember: no pointers $\Rightarrow$ no memory leaks.
Pointers and references to dynamic data

```cpp
vector<myClass> a;
...
// do some push_back's to a
...
const myClass& x = a[0]; // ref to a[0]
// x contains a memory address pointing at a[0]

a.push_back(something);
// the address of a[0] might have changed!
// x might be pointing at garbage data
```

**Recommendation:** don’t trust on pointers/references to dynamic data.

**Possible solution:** use indices/keys to access data in containers.
How much memory space does an object take?

- The methods (code) are shared by all the objects of the same class.
- Each object has its own copy of attributes (data).
About C (the predecessor of C++)

• Developed by Dennis Ritchie at Bell Labs (1972) and used to re-implement Unix.

• It was designed to be easily mappable to machine instructions and provide low-level memory access.

• Today, it is still necessary to use some C libraries, designed by skilled experts, that have not been rewritten in other languages (the same happens with some FORTRAN libraries).

• Some aspects that must be known to interface with C libraries:
  – No references (only pointers).
  – No object-oriented support (no STL, no vectors, no maps, etc).
  – Vectors must be implemented as arrays.
// a is received by value and b by reference
// (using a pointer)
int f(int a, int* b) {
    *b = *b + a;
    return a*a;
}

int main() {
    int x, y, z;
    ...
    // pointers used to pass by reference
    z = f(x, &y);
    ...
}
C: arrays

```c
int a[100];  // an array of 100 int's
double m[30][40]; // a matrix of size 30x40
int c[]; // also int* c: an array of unknown size
int* c; // an array of unknown size
char* s; // a string (array of unknown size)
...
// memory allocation for n integers
int* c = malloc(n*sizeof(int));

// memory allocation for m chars
char* s = malloc(m*sizeof(char));
...
// Anything that is allocated must be freed
// (or memory leaks will occur)
free(c);
free(s);
```
C: arrays

// Equivalent declaration: double sum(double* v, int n)
// The size of the array must be indicated explicitly.
double sum(double v[], int n) {
    double s = 0;
    for (int i = 0; i < n; ++i) s += v[i];
    return s;
}

// Or we can use a sentinel as a terminator of an array.
// Example: strings are usually terminated by a 0.
int length(char* s) {
    int len;
    for (len = 0; s[len] != 0; ++len);
    return len;
}

int main() {
    char hello[] = "Hello, world!";
    // arrays are implicitly passed by reference
    int l = length(hello);
    ...
}
double sum(double v[], int n) {
    double s = 0;
    for (int i = 0; i < n; ++i) s += v[i];
    return s;
}

// Any C++ vector is an object that contains
// a private array. This array can be accessed
// using the ‘data()’ method (C++11).
int main() {
    vector<double> a;
    ...
    double s = sum(a.data(), a.size());
}
Assertions vs. error handling

• Assertions:
  – Runtime checks of properties (e.g., invariants, pre-/post-conditions).
  – Useful to detect internal errors.
  – They should always hold in a bug-free program.
  – They should give meaningful error messages to the programmers.

• Error handling:
  – Detect improper use of the program (e.g., incorrect input data).
  – The program should not halt unexpectedly.
  – They should give meaningful error messages to the users.
assert (x >= 0);
double y = sqrt(x);

assert (i >= 0 and i < A.size());
int v = A[i];

assert (p != nullptr);
int n = p->number;

string id;
cin >> id;
if (isdigit(name[0])) {
    cerr << "Invalid identifier" << endl;
    return false;
}

Containers: Stacks
© Dept. CS, UPC
Summary

• Memory is a valuable resource in computing devices. It must be used efficiently.

• Languages are designed to make memory management transparent to the user, but a lot of inefficiencies may arise unintentionally (e.g., copy by value).

• Pointers imply the use of the heap and all the problems associated to memory management (memory leaks, fragmentation).

• Recommendation: do not use pointers unless you have no other choice. Not using pointers will save a lot of debugging time.

• In case of using pointers, try to hide the pointer manipulation and memory management (new/delete) inside the class in such a way that the user of the class does not need to “see” the pointers.
EXERCISES
Constructors

Consider two versions of the program below, each one using a different definition of the class Point. Comment on the behavior of the program at compile time and runtime.

```cpp
class Point {
    int x, y;
public:
    Point(const Point &p) {
        x = p.x; y = p.y;
    }
    int getX() { return x; }
    int getY() { return y; }
};

class Point {
    int x, y;
public:
    Point(int i=0, int j=0) {
        x = i; y = j;
    }
    int getX() { return x; }
    int getY() { return y; }
};

int main() {
    Point p1(10);
    Point p2 = p1;
    cout << "x = " << p2.getX() << endl;
    cout << "y = " << p2.getY() << endl;
}
```

Consider two versions of the program below, each one using a different definition of the class Point. Comment on the behavior of the program at compile time and runtime.
```cpp
int c = 0; // Global variable

class A {
    int id;
public:
    A() : id(++c) { cout << "N"; }
    A(const A& x) {
        id = x.id; cout << "C";
    }
    A& operator=(const A& x) {
        id = x.id; cout << "A";
    }
    ~A() { cout << id; }
};

void f(const A& x, A y) {
    A z = x;
    A w;
}

int main() {
    A v1, *v2, v3;
    A v4 = v3;
    v2 = new A();
    v1 = *v2;
    f(v1, v4);
    delete v2;
    A v5;
}
```

What is the output of this program? Explain why.
Consider the following definition of a list of students organized as shown in the picture.

You can assume that the vector of marks is never empty.

The last student in the list points at “null” (nullptr).

Design the function `BestStudent` with the following specification:

L points at the first student of the list. `BestStudent` returns the name of the student with the best average mark. In case no student has an average mark greater than or equal to 5, the function must return the string “Bad Teacher”.

```cpp
struct Student {
    string name;
    vector<double> marks;
    Student* next;
};

string BestStudent(Student* L);
```
Containers: Stack

Jordi Cortadella and Jordi Petit
Department of Computer Science
The Stack ADT

- A stack is a list of objects in which insertions and deletions can only be performed at the top of the list.
- Also known as LIFO (Last In, First Out)

Containers: Stacks

**push** (insert an element at the top)

**pop** (delete the most recently inserted element)

**top** (the most recently inserted element)

**empty** (is there any element?)

*Note:* pop and top generate an error on an empty stack
The definition can handle generic stacks of any type T.

The default constructor does not need to do anything: a zero-sized vector is constructed by default.
template <typename T>
class Stack {
    public:
        ...
        bool empty() const {
            return data.size() == 0;
        }

        const T& top() const { // Returns a const reference
            assert (not empty());
            return data.back();
        }

        T& top() { // Returns a reference
            assert (not empty());
            return data.back();
        }

        void pop() {
            assert (not empty());
            data.pop_back();
        }

        void push(const T& x) {
            data.push_back(x);
        }
    }
}
Balancing symbols

• **Balancing symbols**: check for syntax errors when expressions have opening/closing symbols, e.g., () [] {}  

  Correct: [ ( ) { ( ) [ ( ) ] } ( ) ]  
  Incorrect: [ ( ) { ( ) ]...

• **Algorithm** (linear): read all chars until end of file. For each char, do the following:
  – If the char is opening, push it onto the stack.
  – If the char is closing and stack is empty → error, otherwise pop a symbol from the stack and check they match. If not → error.
  – At the end of the file, check the stack is empty.

• **Exercise**: implement and try the above examples.
Evaluation of postfix expressions

• This is an infix expression. What’s his value? 42 or 144?

\[ 8 \times 3 + 10 + 2 \times 4 \]

• It depends on the operator precedence. For scientific calculators, \(*\) has precedence over \(+\).

• Postfix (reverse Polish notation) has no ambiguity:

\[ 8 \, 3 \times \, 10 \, + \, 2 \, 4 \times \, + \]

• Postfix expressions can be evaluated using a stack:
  – each time an operand is read, it is pushed on the stack
  – each time an operator is read, the two top values are popped and operated. The result is push onto the stack
Evaluation of postfix expressions: example

\[ 6 \ 5 \ 2 \ 3 \ + \ 8 \ * \ + \ 3 \ + \ * \]

push(6)
push(5)
push(2)
push(3)  +  push(8)  *  +  push(3)  +  *

push(6)  5  2  3  +  8  *  +  3  +  *

push(8)  40  5  6
push(3)  45  6
push(2)  5  5  6
push(5)  48  6
push(6)  288
From infix to postfix

Algorithm:
- When an operand is read, write it to the output.
- If we read a right parenthesis, pop the stack writing symbols until we encounter the left parenthesis.
- For any other symbol (+, *, (), pop entries and write them until we find an entry with lower priority. After popping, push the symbol onto the stack. Exception: (‘ can only be removed when finding a ’).
- When the end of the input is reached, all symbols in the stack are popped and written onto the output.
## From Infix to Postfix

### Expression
\[ a + b \times c + ( d \times e + f ) \times g \]

### Priority

<table>
<thead>
<tr>
<th>Operation</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>*</td>
<td>a</td>
</tr>
<tr>
<td>+</td>
<td>a</td>
</tr>
<tr>
<td>(</td>
<td>a</td>
</tr>
<tr>
<td>+</td>
<td>a</td>
</tr>
<tr>
<td>+</td>
<td>a b</td>
</tr>
</tbody>
</table>
From infix to postfix

<table>
<thead>
<tr>
<th>Priority</th>
<th>Expression</th>
<th>Postfix</th>
</tr>
</thead>
<tbody>
<tr>
<td>*</td>
<td>a + b * c + ( d * e + f ) * g</td>
<td>a b c * +</td>
</tr>
<tr>
<td>+</td>
<td>a + b * c + ( d * e + f ) * g</td>
<td>a b c * +</td>
</tr>
<tr>
<td>(</td>
<td>a + b * c + ( d * e + f ) * g</td>
<td>a b c * +</td>
</tr>
<tr>
<td>+</td>
<td>a + b * c + ( d * e + f ) * g</td>
<td>a b c * +</td>
</tr>
</tbody>
</table>

Containers: Stacks

© Dept. CS, UPC
From infix to postfix

Priority

*  
+  
(  

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>*</td>
<td>+</td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>*</td>
<td>+</td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>*</td>
<td>+</td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>*</td>
<td>+</td>
</tr>
</tbody>
</table>
From infix to postfix

Priority

<table>
<thead>
<tr>
<th>*</th>
<th>+</th>
<th>(</th>
<th>+</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
<td>+</td>
</tr>
<tr>
<td>d</td>
<td>e</td>
<td>f</td>
<td></td>
</tr>
<tr>
<td>*</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
<td>*</td>
</tr>
<tr>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
<td>*</td>
</tr>
<tr>
<td>d</td>
<td>e</td>
<td>f</td>
<td></td>
</tr>
<tr>
<td>*</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
<td>+</td>
</tr>
<tr>
<td>d</td>
<td>e</td>
<td>f</td>
<td></td>
</tr>
<tr>
<td>*</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
<td>*</td>
</tr>
<tr>
<td>d</td>
<td>e</td>
<td>f</td>
<td></td>
</tr>
<tr>
<td>*</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>b</td>
<td>c</td>
<td>+</td>
</tr>
<tr>
<td>d</td>
<td>e</td>
<td>f</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
From infix to postfix

Priority

\[ \text{a + b * c + ( d * e + f ) * g} \]

Complexity: \( O(n) \)

Suggested exercise:
- Add substraction (same priority as addition) and division (same priority as multiplication).
EXERCISES
Interleaved push/pop operations

Suppose that an intermixed sequence of push and pop operations are performed. The pushes push the integers 0 through 9 in order; the pops print out the return value. Which of the following sequences could not occur?

a) 4 3 2 1 0 9 8 7 6 5
b) 4 6 8 7 5 3 2 9 0 1
c) 2 5 6 7 4 8 9 3 1 0
d) 4 3 2 1 0 5 6 7 8 9

Source: Robert Sedgewick, Computer Science 126, Princeton University.
Design the class **MidStack** implementing a stack with the following operations:

- Push/pop: the usual operations on a stack.
- FindMiddle: returns the value of the element in the middle.
- DeleteMiddle: deletes the element in the middle.

All the operations must be executed in $O(1)$ time.

Suggestion: use some container of the STL to implement it.

**Note:** if the stack has $n$ elements at locations $0..n-1$, where $0$ is the location at the bottom, the middle element is the one at location $\lfloor (n - 1)/2 \rfloor$. 
Containers: Queue and List

Jordi Cortadella and Jordi Petit
Department of Computer Science
Queue

• A container in which insertion is done at one end (the tail) and deletion is done at the other end (the head).

• Also called FIFO (First-In, First-Out)
Queue usage

Queue<int> Q; // Constructor
Q.push(5); // Inserting few elements
Q.push(8);
Q.push(6);

int n = Q.size(); // n = 3

while (not Q.empty()) {
    int elem = Q.front(); // Get the first element
    cout << elem << endl;
    Q.pop(); // Delete the element
}
template<typename T>
class Queue {
public:

Queue(); // Constructor

~Queue(); // Destructor

Queue(const T& Q); // Copy constructor

Queue& operator=(const Queue& Q); // Assignment operator

void push(const& T x); // Enqueues an element

void pop(); // Dequeues the first element

T front() const; // Returns the first element

int size() const; // Number of elements in the queue

bool empty() const; // Is the queue empty?
};
Implementation with linked lists

```
template<typename T>
class Queue {

private:
    struct Node {
        T elem;
        Node* next;
    };

    Node *first; // Pointer to the first element
    Node *last;  // Pointer to the last element
    int n;       // Number of elements
```

Containers © Dept. CS, UPC
Implementation with linked lists

Q.push(5)
Q.push(8)
Q.push(6)
Q.pop()
Queue: some methods

```cpp
/** Returns the number of elements. */
int size() const {
    return n;
}

/** Checks whether the queue is empty. */
bool empty() const {
    return size() == 0;
}

/** Inserts a new element at the end of the queue. */
void push(const T& x) {
    Node* p = new Node {x, nullptr};
    if (n++ == 0) first = last = p;
    else last = last->next = p;
}

/** Removes the first element. Pre: the queue is not empty. */
void pop() {
    assert(not empty());
    Node* old = first;
    first = first->next;
    delete old;
    if (--n == 0) last = nullptr;
}

/** Returns the first element. Pre: the queue is not empty. */
T front() const {
    assert(not empty());
    return first->elem;
}
```
/** Default constructor: an empty queue. */
Queue() : first(nullptr), last(nullptr), n(0) { }

/** Copy constructor. */
Queue(const Queue& Q) {
    copy(Q);
}

/** Assignment operator. */
Queue& operator= (const Queue& Q) {
    if (&Q != this) {
        free();
        copy(Q);
    }
    return *this;
}

/** Destructor. */
~Queue() {
    free();
}

private:
/** Frees the linked list of nodes in the queue. */
void free() {
    Node* p = first;
    while (p) {
        Node* old = p;
        p = p->next;
        delete old;
    }
}
/*** Copies a queue. */
void copy(const Queue& Q) {
    n = Q.n;
    if (n == 0) {
        first = last = nullptr;
    } else {
        Node* p1 = Q.first;
        Node* p2 = first = new Node {p1->elem};
        while (p1->next) {
            p1 = p1->next;
            p2 = p2->next = new Node {p1->elem};
        }
        p2->next = nullptr;
        last = p2;
    }
}
Implementation with circular buffer

• A queue can also be implemented with an array (vector) of elements.

• It is a more efficient representation if the maximum number of elements in the queue is known in advance.
Implementation with circular buffer

Containers

© Dept. CS, UPC
Implementation with circular buffer

after `Q.push(e)`
Implementation with circular buffer

after $Q.pop()$
Implementation with circular buffer
Implementation with circular buffer

after \texttt{Q.push(d)}
Queue: Complexity

• All operations in queues can run in constant time, except for:
  – Copy: linear in the size of the list.
  – Delete: linear in the size of the list.

• Queues do not allow to access/insert/delete elements in the middle of the queue.
List

• List: a container with sequential access.

• It allows to insert/erase elements in the middle of the list in constant time.

• A list can be considered as a sequence of elements with one or several cursors (iterators) pointing at internal elements.

• For simplicity, we will only consider lists with one iterator.

• Check the STL list: it can be visited by any number of iterators.
List: graphical representation

first

5 8 3 0 4 3 9 1 4 8

L.insert(7)

5 8 3 0 4 7 3 9 1 4 8

L.move_right()

5 8 3 0 4 7 3 9 1 4 8

L.erase()

5 8 3 0 4 7 3 1 4 8
List implementation: doubly linked nodes

Sentinel

cursor

cursor: pointer at the first node after the cursor

Empty list:

© Dept. CS, UPC
The class List: private representation

template <typename T>
class List {

  /** Doubly linked node of the list. */
  struct Node {
    Node* prev;  /** Pointer to the previous node. */
    T elem;      /** The element of the list. */
    Node* next;  /** Pointer to the next element. */
  };

  Node* sentinel;  /** Sentinel of the list. */
  Node* cursor;    /** Node after the cursor. */
  int n;           /** Number of elements (without sentinel). */
The class List: public methods

public:
/** Constructor of an empty list. */
List() : sentinel(new Node), cursor(sentinel), n(0) {
  sentinel->next = sentinel->prev = sentinel;
}

/** Destructor. */
~List() {
  free();
}

/** Copy constructor. */
List(const List& L) {
  copy(L);
}

/** Assignment operator. */
List& operator=(const List& L) {
  if (&L != this) {
    free();
    copy(L);
  }
  return *this;
}

/** Returns the number of elements in the list. */
int size() const {
  return n;
}

/** Checks whether the list is empty. */
bool empty() const {
  return size() == 0;
}
public:
/** Checks whether the cursor is at the beginning of the list. */
bool is_at_front() const {
    return cursor == sentinel->next;
}

/** Checks whether the cursor is at the end of the list. */
bool is_at_end() const {
    return cursor == sentinel;
}

/** Moves the cursor one position backward.
   Pre: the cursor is not at the beginning of the list. */
void move_backward() {
    assert(!is_at_front());
    cursor = cursor->prev;
}

/** Moves the cursor one position forward.
   Pre: the cursor is not at the end of the list. */
void move_forward() {
    assert(!is_at_end());
    cursor = cursor->next;
}
The class List: public methods

public:

/** Moves the cursor to the beginning of the list. */
void move_to_front() {
    cursor = sentinel->next;
}

/** Moves the cursor to the end of the list. */
void move_to_end() {
    cursor = sentinel;
}

/** Inserts an element x before the cursor. */
void insert(const T& x) {
    Node* p = new Node {cursor->prev, x, cursor};
    cursor->prev = cursor->prev->next = p;
    ++n;
}
The class List: public methods

public:

/** Erases the element after the cursor.  
   Pre: cursor is not at the end. */
void erase() {
    assert(not is_at_end());
    Node* p = cursor;
    p->next->prev = p->prev;
    cursor = p->prev->next = p->next;
    delete p;
    --n;
}

/** Returns the element after the cursor.  
   Pre: the cursor is not at the end. */
T front() const {
    assert(not is_at_end());
    return cursor->elem;
}

Exercises: implement the private methods copy() and free().
Higher-order functions

• A higher-order function is a function that can receive other functions as parameters or return a function as a result.

• Most languages support higher-order functions (C++, python, R, Haskell, Java, JavaScript, ...).

• The have different applications:
  – `sort` in STL is a higher-order function (the compare function is a parameter).
  – functions to visit the elements of containers (lists, trees, etc.) can be passed as parameters.
  – Mathematics: functions for composition and integration receive a function as parameter.
  – etc...
Higher-order functions: example

template <typename T>
class List {

...  
/** Transforms every element of the list using f.  
   It returns a reference to the list. */  
List<T>& transform(void f(T&));

/** Returns a list with the elements for which f is true */  
List<T> filter(bool f(const T&)) const;

/** Applies f sequentially to the list and returns a single value. For the list \([x_1, x_2, x_3, ..., x_n]\) it returns \(f(\ldots f(f(\text{init}, x_1), x_2)\ldots, x_n)\).  
If the list is empty, it returns init. */  
T reduce(T f(const T&, const T&), T init) const;
}
Higher-order functions: example

/** Checks whether a number is prime */
bool isPrime(int n) {...}

/** Adds two numbers */
int add(int x, int y) {
    return x + y;
}

/** Substitutes a number by its square */
void square(int& x) {
    x = x*x;
}

/** The following code computes:
\[ \sum_{x \in L, \text{x is prime}} x^2 \] */

int n = L.filter(isPrime).transform(square).reduce(add, 0);
Higher-order functions: example

List<T>& transform(void f(T&)) {
    Node* p = sentinel->next;
    while (p != sentinel) { // Visit all elements and apply f to each one
        f(p->elem);
        p = p->next;
    }
    return *this;
}

List<T> filter(bool f(const T&)) const {
    List<T> L;
    Node* p = sentinel->next;
    while (p != sentinel) { // Pick elements only if f is asserted
        if (f(p->elem)) L.insert(p->elem);
        p = p->next;
    }
    return L;
}

T reduce(T f(const T&, const T&), T init) const {
    T x = init; // Initial value
    Node* p = sentinel->next; // First element (if any)
    while (p != sentinel) {
        x = f(x, p->elem); // Composition with next element
        p = p->next;
    }
    return x;
}
Queues implemented as circular buffers

• Design the class queue implemented with a circular buffer (using a vector):
  – The push/pop/front operations should run in constant time.
  – The copy and delete operations should run in linear time.
  – The class should have a constructor with a parameter $n$ that should indicate the maximum number of elements in the queue.

• Consider the design of a variable-size queue using a circular buffer. Discuss how the implementation should be modified.
Reverse and Josephus

• Design the method `reverse()` that reverses the contents of the list:
  – No auxiliary lists should be used.
  – No copies of the elements should be performed.

• Solve the Josephus problem, for $n$ people and executing every $k$-th person, using a circular list:

• Design the method `merge(const List& L)` that merges the list with another list L, assuming that both lists are sorted. Assume that a pair of elements can be compared with the operator `<.

• Design the method `sort()` that sorts the list according to the `<` operator. Consider merge sort and quick sort as possible algorithms.
Graphs: Connectivity

Jordi Cortadella and Jordi Petit
Department of Computer Science
The network graph formed by Wikipedia editors (edges) contributing to different Wikipedia language versions (vertices) during one month in summer 2013
Transportation systems
Social networks
World Wide Web
Disease transmission network

Transmission of renewable energy

Topology of regional transmission grid model of continental Europe in 2020
What would we like to solve on graphs?

• Finding paths: which is the shortest route from home to my workplace?

• Flow problems: what is the maximum amount of people that can be transported in Barcelona at rush hours?

• Constraints: how can we schedule the use of the operating room in a hospital to minimize the length of the waiting list?

• Clustering: can we identify groups of friends by analyzing their activity in twitter?
A significant part of the material used in this chapter has been inspired by the book:


(several examples, figures and exercises are taken from the book)
A graph is specified by a set of vertices (or nodes) $V$ and a set of edges $E$.

$$V = \{1, 2, 3, 4, 5\}$$

$$E = \{(1, 2), (1, 3), (2, 4), (3, 4), (4, 5), (5, 2), (5, 5)\}$$

Graphs can be directed or undirected. Undirected graphs have a symmetric relation.
A graph with $n = |V|$ vertices, $v_1, \ldots, v_n$, can be represented by an $n \times n$ matrix with:

$$a_{i,j} = \begin{cases} 1 & \text{if there is an edge from } v_i \text{ to } v_j \\ 0 & \text{otherwise} \end{cases}$$

$$a = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Space: $O(n^2)$

For undirected graphs, the matrix is symmetric.
A graph can be represented by $|V|$ lists, one per vertex. The list for vertex $u$ holds the vertices connected to the outgoing edges from $u$.

The lists can be implemented in different ways (vectors, linked lists, ...)

Space: $O(|E|)$

**Undirected graphs**: use bi-directional edges
Dense and sparse graphs

• A graph with $|V|$ vertices could potentially have up to $|V|^2$ edges (all possible edges are possible).

• We say that a graph is **dense** when $|E|$ is close to $|V|^2$. We say that a graph is **sparse** when $|E|$ is close to $|V|$.

• How big can a graph be?

Dense graph

Sparse graph
• December 2017: 50 billion web pages \((50 \times 10^9)\).
• Size of adjacency matrix: \(25 \times 10^{20}\) elements.
  (not enough computer memory in the world to store it).
• Good news: The web is very sparse. Each web page has about half a dozen hyperlinks to other web pages.

[www.worldwidewebsize.com](http://www.worldwidewebsize.com)
Adjacency matrix vs. adjacency list

• Space:
  – Adjacency matrix is $O(|V|^2)$
  – Adjacency list is $O(|E|)$

• Checking the presence of a particular edge $(u, v)$:
  – Adjacency matrix: constant time
  – Adjacency list: traverse $u$’s adjacency list

• Which one to use?
  – For dense graphs $\rightarrow$ adjacency matrix
  – For sparse graphs $\rightarrow$ adjacency list

• For many algorithms, traversing the adjacency list is not a problem, since they require to iterate through all neighbors of each vertex. For sparse graphs, the adjacency lists are usually short (can be traversed in constant time)
// Declaration of a graph that stores
// a string (name) for each vertex
Graph<string> G;

// Create the vertices
int a = G.addVertex("a");
int b = G.addVertex("b");
int c = G.addVertex("c");

// Create the edges
G.addEdge(a,a);
G.addEdge(a,b);
G.addEdge(b,c);
G.addEdge(c,b);

// Print all edges of the graph
for (int src = 0; src < G.numVertices(); ++src) {
    // all vertices
    for (auto dst: G.succ(src)) {
        // all successors of src
        cout << G.info(src) << " -> " << G.info(dst) << endl;
    }
}
template<
typename vertexType>
class Graph {
private:
    struct Vertex {
        vertexType info;  // Information of the vertex
        vector<int> succ; // List of successors
        vector<int> pred; // List of predecessors
    };

    vector<Vertex> vertices; // List of vertices

public:
    /** Constructor */
    Graph() {};

    /** Adds a vertex with information associated to the vertex.
    Returns the index of the vertex */
    int addVertex(const vertexType& info) {
        vertices.push_back(Vertex{info});
        return vertices.size() - 1;
    }
/** Adds an edge src \(\rightarrow\) dst */
void addEdge(int src, int dst) {
    vertices[src].succ.push_back(dst);
    vertices[dst].pred.push_back(src);
}

/** Returns the number of vertices of the graph */
int numVertices() const {
    return vertices.size();
}

/** Returns the information associated to vertex v */
const vertexType& info(int v) const {
    return vertices[v].info;
}

/** Returns the list of successors of vertex v */
const vector<int>& succ(int v) const {
    return vertices[v].succ;
}

/** Returns the list of predecessors of vertex v */
const vector<int>& pred(int v) const {
    return vertices[v].pred;
}
Reachability: exploring a maze

Which vertices of the graph are reachable from a given vertex?
To explore a labyrinth we need a ball of string and a piece of chalk:

- The chalk prevents looping, by marking the visited junctions.
- The string allows you to go back to the starting place and visit routes that were not previously explored.
How to simulate the string and the chalk with an algorithm?

- Chalk: a boolean variable for each vertex (visited).
- String: a stack
  - push vertex to unwind at each junction
  - pop to rewind and return to the previous junction

Note: the stack can be simulated with recursion.
function explore(G, v):
// Input: G = (V,E) is a graph
// Output: visited(u) is true for all the nodes reachable from v

visited(v) = true
previsit(v)
for each edge (v,u) ∈ E:
    if not visited(u): explore(G,u)
postvisit(v)

Notes:
• Initially, visited(v) is assumed to be false for every v ∈ V.
• pre/postvisit functions are not required now.
Finding the nodes reachable from another node

```plaintext
function explore(G, v):
    visited(v) = true
    for each edge (v, u) ∈ E:
        if not visited(u): explore(G, u)
```

- All visited nodes are reachable because the algorithm only moves to neighbors and cannot jump to an unreachable region.

- Does it miss any reachable vertex? No. Proof by contradiction.
  - Assume that a vertex \( u \) is missed.
  - Take any path from \( v \) to \( u \) and identify the last vertex that was visited on that path (\( z \)). Let \( w \) be the following node on the same path. Contradiction: \( w \) should have also been visited.
Finding the nodes reachable from another node

```python
function explore(G, v):
    visited(v) = true
    for each edge (v, u) ∈ E:
        if not visited(u): explore(G, u)
```

Dotted edges are ignored (*back edges*): they lead to previously visited vertices.
The solid edges (*tree edges*) form a tree.
Depth-first search

```python
function DFS(G):
    for all \( v \in V \):
        visited(v) = false
    for all \( v \in V \):
        if not visited(v): explore(G, v)
```

DFS traverses the entire graph.

**Complexity:**
- Each vertex is visited only once (thanks to the chalk marks)
- For each vertex:
  - A fixed amount of work (pre/postvisit)
  - All adjacent edges are scanned

**Running time** is \( O(|V| + |E|) \).
Difficult to improve: reading a graph already takes \( O(|V| + |E|) \).
**DFS example**

The outer loop of DFS calls *explore* three times (for A, C and F)
- Three trees are generated. They constitute a *forest*.

---

**function DFS(G):**

```plaintext
for all v ∈ V:
   visited(v) = false
for all v ∈ V:
   if not visited(v): explore(G, v)
```
Connectivity

• An undirected graph is connected if there is a path between any pair of vertices.

• A disconnected graph has disjoint *connected components*.

• Example: this graph has 3 connected components:

\[ \{A, B, E, I, J\} \quad \{C, D, G, H, K, L\} \quad \{F\} \]
function explore(G, v, cc):
    // Input: G = (V, E) is a graph, cc is a CC number
    // Output: ccnum[u] = cc for each vertex u in the same CC as v
    ccnum[v] = cc
    for each edge (v, u) ∈ E:
        if ccnum[u] == 0: explore(G, u, cc)

function ConnComp(G):
    // Input: G = (V, E) is a graph
    // Output: Every vertex v has a CC number in ccnum[v]
    for all v ∈ V: ccnum[v] = 0; // Clean cc numbers
    cc = 1; // Identifier of the first CC
    for all v ∈ V:
        if ccnum[v] = 0: // A new CC starts
            explore(G, v, cc); cc = cc + 1;

• Performs a DFS traversal assigning a CC number to each vertex.
• The outer loop of ConnComp determines the number of CC’s.
• The variable ccnum[v] also plays the role of visited[v].
Revisiting the explore function

function explore(G, v):
    visited(v) = true
    previsit(v)
    for each edge (v, u) ∈ E:
        if not visited(u):
            explore(G, u)
    postvisit(v)

function previsit(v):
    pre[v] = clock
    clock = clock + 1

function postvisit(v):
    post[v] = clock
    clock = clock + 1

Let us consider a global variable clock that can determine the occurrence times of previsit and postvisit.

Every node v will have an interval (pre[v], post[v]) that will indicate the time the node was first visited (pre) and the time of departure from the exploration (post).

Property: Given two nodes u and v, the intervals (pre[u], post[u]) and (pre[v], post[v]) are either disjoint or one is contained within the other.

The pre/post interval of u is the lifetime of explore(u) in the stack (LIFO).
Example of pre/postvisit orderings

Graphs: Connectivity

Recursion depth

© Dept. CS, UPC
DFS in directed graphs: types of edges

- **Tree edges**: those in the DFS forest.
- **Forward edges**: lead to a nonchild descendant in the DFS tree.
- **Back edges**: lead to an ancestor in the DFS tree.
- **Cross edges**: lead to neither descendant nor ancestor.
DFS in directed graphs: types of edges

- **Tree edges**: those in the DFS forest.
- **Forward edges**: lead to a nonchild descendant in the DFS tree.
- **Back edges**: lead to an ancestor in the DFS tree.
- **Cross edges**: lead to neither descendant nor ancestor.
A cycle is a circular path:

\[ v_0 \rightarrow v_1 \rightarrow v_2 \rightarrow \cdots \rightarrow v_k \rightarrow v_0. \]

Examples:

- B → E → F → B
- C → D → A → C

**Property:** A directed graph has a cycle iff its DFS reveals a back edge.

**Proof:**

\[ \Leftarrow \]

If \((u, v)\) is a back edge, there is a cycle with \((u, v)\) and the path from \(v\) to \(u\) in the search tree.

\[ \Rightarrow \]

Let us consider a cycle \(v_0 \rightarrow v_1 \rightarrow v_2 \rightarrow \cdots \rightarrow v_k \rightarrow v_0\). Let us assume that \(v_i\) is the first discovered vertex (lowest pre number). All the other \(v_j\) on the cycle are reachable from \(v_i\) and will be its descendants in the DFS tree. The edge \(v_{i-1} \rightarrow v_i\) leads from a vertex to its ancestor and is thus a back edge.
A list of tasks that must be executed in a certain order (cannot be executed if it has cycles).

**Legal task linearizations (or topological sorts):**
Directed Acyclic Graphs (DAGs)

A DAG is a directed graph without cycles.

DAGs are often used to represent causalities or temporal dependencies, e.g., task A must be completed before task C.

• Cyclic graphs cannot be linearized.

• All DAGs can be linearized. How?
  – Decreasing order of the post numbers.
  – The only edges \((u, v)\) with \(\text{post}[u] < \text{post}[v]\) are back edges (do not exist in DAGs).

• Property: In a DAG, every edge leads to a vertex with a lower post number.

• Property: Every DAG has at least one source and at least one sink.
  (source: highest post number, sink: lowest post number).
Topological sort

```plaintext
function explore(G, v):
    visited(v) = true
    previsit(v)
    for each edge (v,u) ∈ E:
        if not visited(u):
            explore(G,u)
    postvisit(v)

Initially: TSort = ∅

function postvisit(v):
    TSort.push_front(v)

// After DFS, TSort contains
// a topological sort
```

Another algorithm:
- Find a source vertex, write it, and delete it (mark) from the graph.
- Repeat until the graph is empty.

It can be executed in linear time. How?
This graph is connected (undirected view), but there is no path between any pair of nodes.

For example, there is no path $K \to \cdots \to C$ or $E \to \cdots \to A$.

The graph is not *strongly connected*. Two nodes $u$ and $v$ of a directed graph are connected if there is a path from $u$ to $v$ and a path from $v$ to $u$.

The *connected* relation is an equivalence relation and partitions $V$ into disjoint sets of *strongly connected components*.  

**Strongly Connected Components**

\[
\begin{align*}
\{A\} \\
\{B, E\} \\
\{C, F\} \\
\{D\} \\
\{G, H, I, J, K, L\}
\end{align*}
\]
Every directed graph can be represented by a *meta-graph*, where each meta-node represents a strongly connected component.

**Property:** every directed graph is a DAG of its strongly connected components. (Exercise: prove it!)

A directed graph can be seen as a 2-level structure. At the top we have a DAG of SCCs. At the bottom we have the details of the SCCs.
Properties of DFS and SCCs

• **Property:** If the *explore* function starts at \( u \), it will terminate when all vertices reachable from \( u \) have been visited.
  
  – If we start from a vertex in a sink SCC, it will retrieve exactly that component.
  
  – If we start from a non-sink SCC, it will retrieve the vertices of several components.

• **Examples:**
  
  – If we start at \( K \) it will retrieve the component \( \{G, H, I, J, K, L\} \).
  
  – If we start at \( B \) it will retrieve all vertices except \( A \).
Properties of DFS and SCCs

- **Intuition for the algorithm:**
  - Find a vertex located in a sink SCC
  - Extract the SCC

- **To be solved:**
  - How to find a vertex in a sink SCC?
  - What to do after extracting the SCC?

- **Property:** If $C$ and $C'$ are SCCs and there is an edge $C \rightarrow C'$, then the highest post number in $C$ is bigger than the highest post number in $C'$.

- **Property:** The vertex with the highest DFS post number lies in a source SCC.
Properties of DFS and SCCs

But we would like executing DFS starting from the sink nodes!

How can we do that?

highest post-visit numbers
Reverse graph \((G^R)\)
SCC algorithm

function SCC(G):
// Input: G(V,E) a directed graph
// Output: each vertex v has an SCC number in ccnum[v]
G^R = reverse(G)
DFS(G^R) // calculates post numbers
sort V // decreasing order of post number
ConnComp(G)

Runtime complexity:
• DFS and ConnComp run in linear time O(|V| + |E|).
• Can we reverse G in linear time?
• Can we sort V by post number in linear time?
Reversing $G$ in linear time

function SCC($G$):

// Input: $G(V,E)$ a directed graph
// Output: each vertex $v$ has an SCC number in ccnum[$v$]

$G^R$ = reverse($G$)

DFS($G^R$) // calculates post numbers

sort $V$ // decreasing order of post number

ConnComp($G$)

function reverse($G$)

// Input: $G(V,E)$ graph represented by an adjacency list
//        edges[$v$] for each vertex $v$.
// Output: $G(V,E^R)$ the reversed graph of $G$, with the
//         adjacency list edgesR[$v$].

for each $u \in V$:
    for each $v \in \text{edges}[u]$:
        edgesR[$v$].insert($u$)

return $(V,\text{edgesR})$
Sorting $V$ in linear time

```plaintext
function SCC(G):
    // Input: $G(V,E)$ a directed graph
    // Output: each vertex $v$ has an SCC number in ccnum[$v$]
    $G^R = \text{reverse}(G)$
    DFS($G^R$) // calculates post numbers
    sort $V$ // decreasing order of post number
    ConnComp($G$)
```

Use the explore function for topological sort:
- Each time a vertex is post-visited, it is inserted at the top of the list.
- The list is ordered by decreasing order of post number.
- It is executed in linear time.
Sorting $V$ in linear time

Assume the initial order:
$F, A, B, C, D, E, J, G, H, I, K, L$

$G^R$

DFS tree

<table>
<thead>
<tr>
<th>Vertex:</th>
<th>J</th>
<th>L</th>
<th>K</th>
<th>H</th>
<th>G</th>
<th>I</th>
<th>D</th>
<th>F</th>
<th>C</th>
<th>B</th>
<th>E</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post:</td>
<td>24</td>
<td>23</td>
<td>22</td>
<td>21</td>
<td>17</td>
<td>16</td>
<td>12</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>
Crawling the Web

• Crawling the Web is done using depth-first search strategies.

• The graph is unknown and no recursion is used. A stack is used instead containing the nodes that have already been visited.

• The stack is not exactly a LIFO. Only the most “interesting” nodes are kept (e.g., page rank).

• Crawling is done in parallel (many computers at the same time) but using a central stack.

• How do we know that a page has already been visited? Hashing.
Summary

• Big data is often organized in big graphs (objects and relations between objects)

• Big graphs are usually sparse. Adjacency lists is the most common data structure to represent graphs.

• Connectivity can be analyzed in linear time using depth-first search.
EXERCISES
DFS: stack overflow

• DFS can be implemented with an elegant recursive algorithm, but it may experiment stack overflow problems. Explain why.

• Design an iterative version of DFS.

• Challenge: do not to search in stack overflow
Perform DFS on the two graphs. Whenever there is a choice of vertices, pick the one that is alphabetically first. Classify each edge as a tree edge, forward edge, back edge or cross edge, and give the pre and post number of each vertex.
Run the DFS-based topological ordering algorithm on the graph. Whenever there is a choice of vertices to explore, always pick the one that is alphabetically first.

1. Indicate the pre and post numbers of the nodes.
2. What are the sources and sinks of the graph?
3. What topological order is found by the algorithm?
4. How many topological orderings does this graph have?
Run the SCC algorithm on the two graphs. When doing DFS of $G^R$: whenever there is a choice of vertices to explore, always pick the one that is alphabetically first. For each graph, answer the following questions:
1. In what order are the SCCs found?
2. Which are source SCCs and which are sink SCCs?
3. Draw the meta-graph (each meta-node is an SCC of $G$).
4. What is the minimum number of edges you must add to the graph to make it strongly connected?
The police department in the city of Computopia has made all streets one-way. The mayor contends that there is still a way to drive legally from any intersection in the city to any other intersection, but the opposition is not convinced. A computer program is needed to determine whether the mayor is right. However the city elections are coming up soon, and there is just enough time to run a \textit{linear-time} algorithm.

a) Formulate this problem graph-theoretically, and explain why it can indeed be solved in linear time.

b) Suppose it now turns out that the mayor’s original claim is false. She next claims something weaker: if you start driving from town hall, navigating one-way streets, then no matter where you reach, there is always a way to drive legally back to the town hall. Formulate this weaker property as a graph-theoretic problem, and carefully show how it too can be checked in linear time.
We have three containers whose sizes are 10 pints, 7 pints and 4 pints, respectively. The 7-pint and 4-pint containers start out full of water, but the 10-pint container is initially empty. We are allowed one type of operation: pouring the contents of one container into another, stopping only when the source container is empty or the destination container is full. We want to know if there is a sequence of pourings that leaves exactly 2 pints in the 4-pint container.

a) Model this as a graph problem: give a precise definition of the graph involved and state the specific question about this graph that needs to be answered.

b) What algorithm should be applied to solve the problem?

c) Give a sequence of pourings, if it exists, or prove that it does not exist any sequence.

**Hint:** A vertex of the graph can be represented by a triple of integers.
Containers: Priority Queues

Jordi Cortadella and Jordi Petit
Department of Computer Science
A priority queue

• A priority queue is a queue in which each element has a priority.

• Elements with higher priority are served before elements with lower priority.

• It can be implemented as a vector or a linked list. For a queue with \( n \) elements:
  – Insertion is \( O(n) \).
  – Extraction is \( O(1) \).

• A more efficient implementation can be proposed in which insertion and extraction are \( O(\log n) \): binary heap.
Binary Heap

- Complete binary tree (except at the bottom level).
- Height $h$: between $2^h$ and $2^{h+1} - 1$ nodes.
- For $N$ nodes, the height is $O(\log N)$.
- It can be represented in a vector.
Binary Heap

Heap-Order Property: the key of the parent of X is smaller than (or equal to) the key in X.

Locations in the vector:

\[ i \]

\[ 2i \]

\[ 2i + 1 \]

\[ \left\lfloor \frac{i}{2} \right\rfloor \]

\[ i \text{ (even)} \]

\[ i + 1 \text{ (odd)} \]
Two main operations on a binary heap:
• Insert a new element
• Remove the min element

Both operations must preserve the properties of the binary heap:
• Completeness
• Heap-Order property
Binary Heap: insert 14

Insert in the last location

... and bubble up ...

done!
Binary Heap: remove min

Extract the min element and move the last one to the root of the heap

... and bubble down ...
Binary Heap: remove min

done!
Binary Heap: complexity

• Bubble up/down operations do at most $h$ swaps, where $h$ is the height of the tree and

$$h = \lfloor \log_2 N \rfloor$$

• Therefore:
  – Getting the min element is $O(1)$
  – Inserting a new element is $O(\log N)$
  – Removing the min element is $O(\log N)$
Binary Heap: other operations

- Let us assume that we have a method to know the location of every key in the heap.

- Increase/decrease key:
  - Modify the value of one element in the middle of the heap.
  - If decreased $\rightarrow$ bubble up.
  - If increased $\rightarrow$ bubble down.

- Remove one element:
  - Set value to $-\infty$, bubble up and remove min element.
Heaps are sometimes constructed from an initial collection of $N$ elements. How much does it cost to create the heap?

- Obvious method: do $N$ insert operations.
- Complexity: $O(N \log N)$

Can it be done more efficiently?
Building a heap from a set of elements

Containers: Priority Queues

© Dept. CS, UPC
// Constructor from a collection of items
BinaryHeap(const vector<Elem>& items) {
    v.push_back(Elem()); // v is the vector holding the elements
    for (auto& e: items) v.push_back(e);
    for (int i = size()/2; i > 0; --i) bubble_down(i);
}

Sum of the heights of all nodes:
- 1 node with height $h$
- 2 nodes with height $h - 1$
- 4 nodes with height $h - 2$
- $2^i$ nodes with height $h - i$

$$S = h + 2(h - 1) + 4(h - 2) + 8(h - 3) + 16(h - 4) + \cdots + 2^{h-1}(1)$$
$$2S = 2h + 4(h - 1) + 8(h - 2) + 16(h - 3) + \cdots + 2^h(1)$$
Subtract the two equations:
$$S = -h + 2 + 4 + 8 + \cdots + 2^{h-1} + 2^h = (2^{h+1} - 1) - (h + 1) = O(N)$$

A heap can be built from a collection of items in linear time.
Heap sort

```cpp
template <typename T>
void HeapSort(vector<T>& v) {
    BinaryHeap<T> heap(v);
    for (T& e: v) e = heap.remove_min();
}
```

- Complexity: $O(n \log n)$
  - Building the heap: $O(n)$
  - Each removal is $O(\log n)$, executed $n$ times.
Exercise: insert/remove element

Given the binary heap implemented in the following vector, draw the tree represented by the vector.

```
6  7  9  10 11 12 13 15 19 14 21 17 16
```

Execute the following sequence of operations

```
insert(8); remove_min(); insert(6); insert(18); remove_min();
```

and draw the tree after the execution of each operation.
Exercise: guess $a$ and $b$

Consider the binary heap of integer keys implemented by the following vector:

| 3 | 7 | $a$ | 10 | 15 | 18 | $b$ | 25 | 13 | 20 | 17 | 22 | 19 |

After executing the operations `insert(8)` and `remove_min()` the contents of the binary heap is:

| 7 | 10 | 8 | $b$ | 15 | 18 | $a$ | 25 | 13 | 20 | 17 | 22 | 19 |

Discuss about the possible values of $a$ and $b$. Assume there can never be two identical keys in the heap.
Exercise: the $k$-th element

The $k$-th element of $n$ sorted vectors.

Let us consider $n$ vectors sorted in ascending order.

Design an algorithm with cost $\Theta(k \log n + n)$ that finds the $k$-th global smallest element.
Consider the following declaration for a Binary Heap:

```cpp
template <typename T>  // T must be a comparable type
class BinaryHeap {
private:
    vector<Elem> v;  // Table for the heap (location 0 not used)

    // Bubbles up the element at location i
    void bubble_up(int i);

    // Bubbles down the element at location i
    void bubble_down(int i);
};
```

Give an implementation for the methods `bubble_up` and `bubble_down`. 
Graphs: Shortest paths

Jordi Cortadella and Jordi Petit
Department of Computer Science
Distance in a graph

Depth-first search finds vertices reachable from another given vertex. The paths are not the shortest ones.

Distance between two nodes: length of the shortest path between them.
Breadth-first search

Similar to a wave propagation
Breadth-first search
Breadth-first search
BFS algorithm

• BFS visits vertices layer by layer: 0, 1, 2, ..., $d$.

• Once the vertices at layer $d$ have been visited, start visiting vertices at layer $d + 1$.

• Algorithm with two active layers:
  – Vertices at layer $d$ (currently being visited).
  – Vertices at layer $d + 1$ (to be visited next).

• Central data structure: a queue.
BFS algorithm: simulation

Graphs: Shortest paths

© Dept. CS, UPC
BFS queue

Pop elements with distance $d$

Push elements with distance $d + 1$
BFS algorithm

function BFS(G, s)
// Input: Graph G(V,E), source vertex s.
// Output: For each vertex u, dist[u] is
//        the distance from s to u.

for all u ∈ V: dist[u]=∞

dist[s] = 0
Q = {s} // Queue containing just s
while not Q.empty():
    u = Q.pop_front()
    for all (u,v) ∈ E:
        if dist[v] = ∞:
            dist[v] = dist[u] + 1
        Q.push_back(v)

Runtime O(|V| + |E|): Each vertex is visited once, each edge is visited once
(for directed graphs) or twice (for undirected graphs).
Reachability: BFS vs. DFS

**Input:** A graph $G$ and a source node $s$.

**Output:** $\forall u \in V$: reached[$u$] $\iff$ $u$ is reachable from $s$.

---

**function** BFS($G$, $s$)

```
for all $u \in V$:
    reached[$u$] = false

Q = $\mathbb{Q}$ // Empty queue
Q.push_back($s$)
reached[$s$] = true
while not Q.empty():
    $u = Q$.pop_front()
    for all $(u, v) \in E$:
        if not reached[$v$]:
            reached[$v$] = true
            Q.push_back($v$)
```

**function** DFS($G$, $s$)

```
for all $u \in V$:
    reached[$u$] = false

S = $\mathbb{S}$ // Empty stack
S.push($s$)
reached[$s$] = true
while not S.empty():
    $u = S$.pop()
    for all $(u, v) \in E$:
        if not reached[$v$]:
            reached[$v$] = true
            S.push($v$)
```
Reachability: BFS vs. DFS

DFS order: A B C E F G H D
BFS order: A B D C F E G H
Distance: 0 1 1 2 2 3 3 3 3
Weights on edges

Image credits: https://thegadgetflow.com/blog/google Maps-vs-google-earth/
Inefficient: many cycles without any interesting progress. How about real numbers?
Dijkstra’s algorithm: invariant

Shortest paths already computed (completed vertices)

Frontier

Data structure:
The set of non-completed vertices with their shortest distance from S using only the completed vertices.

Graphs: Shortest paths
Example

Graphs: Shortest paths

$t = 0$

**Done** | **Queue**
---|---
A: 0 | B: ∞
B: ∞ | E: ∞
D: ∞ | C: ∞
C: ∞ | F: ∞
F: ∞ | G: ∞
Example

\[
\begin{align*}
\text{Done} & \quad \text{Queue} \\
A:0 & \quad B:2 \\
D:1 & \quad E:3 \\
C:3 & \\
G:5 & \\
F:9 & \\
\end{align*}
\]

\[
\begin{align*}
\text{Done} & \quad \text{Queue} \\
A:0 & \quad E:3 \\
D:1 & \quad C:3 \\
B:2 & \quad G:5 \\
F:9 & \\
\end{align*}
\]
Example

Graphs: Shortest paths

© Dept. CS, UPC
Example

$t = 5$

<table>
<thead>
<tr>
<th>Done</th>
<th>Queue</th>
</tr>
</thead>
<tbody>
<tr>
<td>A:0</td>
<td></td>
</tr>
<tr>
<td>D:1</td>
<td></td>
</tr>
<tr>
<td>B:2</td>
<td></td>
</tr>
<tr>
<td>E:3</td>
<td></td>
</tr>
<tr>
<td>C:3</td>
<td></td>
</tr>
<tr>
<td>G:5</td>
<td></td>
</tr>
</tbody>
</table>

$t = 6$

<table>
<thead>
<tr>
<th>Done</th>
<th>Queue</th>
</tr>
</thead>
<tbody>
<tr>
<td>A:0</td>
<td></td>
</tr>
<tr>
<td>D:1</td>
<td></td>
</tr>
<tr>
<td>B:2</td>
<td></td>
</tr>
<tr>
<td>E:3</td>
<td></td>
</tr>
<tr>
<td>C:3</td>
<td></td>
</tr>
<tr>
<td>G:5</td>
<td></td>
</tr>
<tr>
<td>F:6</td>
<td></td>
</tr>
</tbody>
</table>
We need to:

- keep a list non-completed vertices and their expected distances.
- select the non-completed vertex with shortest distance.
- update the distances of the neighbouring vertices.
Dijkstra’s algorithm for shortest paths

```
function ShortestPaths(G, l, s)

// Input: Graph G(V,E), source vertex s,
//        positive edge lengths {l_e:e ∈ E}
// Output: dist[u] has the distance from s,
//        prev[u] has the predecessor in the tree

    for all u ∈ V:
        dist[u] = ∞
        prev[u] = nil

    dist[s] = 0
    Q = makequeue(V)  // using dist as keys

    while not Q.empty():
        u = Q.deleteMin()
        for all (u,v) ∈ E:
            if dist[v] > dist[u] + l(u,v):
                dist[v] = dist[u] + l(u,v)
                prev[v] = u
                Q.decreaseKey(v)

```
Dijkstra’s algorithm: complexity

\[ Q = \text{makequeue}(V) \]

\[ \textbf{while not } \ Q.\text{empty}(): \]
\[ u = Q.\text{deletemin}() \]
\[ \textbf{for all } (u, v) \in E:\]
\[ \quad \textbf{if } \text{dist}[v] > \text{dist}[u] + l(u, v): \]
\[ \quad \quad \text{dist}[v] = \text{dist}[u] + l(u, v) \]
\[ \quad \quad \text{prev}[v] = u \]
\[ \quad Q.\text{decreasekey}(v) \]

- The skeleton of Dijkstra’s algorithm is based on BFS, which is \( O(|V| + |E|) \)
- We need to account for the cost of:
  - makequeue: insert \(|V|\) vertices to a list.
  - deletemin: find the vertex with min dist in the list (\(|V|\) times)
  - decreasekey: update dist for a vertex (\(|E|\) times)

- Let us consider two implementations for the list: \textbf{vector} and \textbf{binary heap}
## Dijkstra’s algorithm: complexity

<table>
<thead>
<tr>
<th>Implementation</th>
<th>deletemin</th>
<th>insert/decreasekey</th>
<th>Dijkstra’s complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vector</td>
<td>$O(</td>
<td>V</td>
<td>)$</td>
</tr>
<tr>
<td>Binary heap</td>
<td>$O(\log</td>
<td>V</td>
<td>)$</td>
</tr>
</tbody>
</table>

### Binary heap:
- The elements are stored in a complete (balanced) binary tree.
- **Insertion**: place element at the bottom and let it *bubble up* swapping the location with the parent (at most $\log_2 |V|$ levels).
- **Deletemin**: Remove element from the root, take the last node in the tree, place it at the root and let it *bubble down* (at most $\log_2 |V|$ levels).
- **Decreasekey**: decrease the key in the tree and let it *bubble up* (same as insertion). A data structure might be required to known the location of each vertex in the heap (table of pointers).

For connected graphs: $O((|V| + |E|) \log |V|) = O(|E| \log |V|)$
Why Dijkstra’s works

- A tree of open paths with distances is maintained at each iteration.
- The shortest paths for the internal nodes have already been calculated.
- The node in the frontier with shortest distance is “frozen” and expanded. Why? Because no other shorter path can reach the node.

Disclaimer: this is only true if the distances are non-negative!
Graphs with negative edges

- Dijkstra’s algorithm does not work:

  - Problem: shortest paths are consolidated too early.
  - Possible solution: add a constant weight to all edges, make them positive, and apply Dijkstra.
    - It does not work, prove it!

  **Dijkstra**

  Let **S** → **A** have length = 3.

  \[
  \text{dist}(v) = \min\{\text{dist}(v), \text{dist}(u) + l(u, v)\}
  \]

- Dijkstra is based on a safe update each time an edge \((u, v)\) is treated:
Graphs with negative edges

- The shortest path from \(s\) to \(t\) can have at most \(|V| - 1\) edges:

\[
\begin{align*}
    s & \rightarrow u_1 \rightarrow u_2 \rightarrow u_3 \rightarrow \cdots \rightarrow u_k \rightarrow t
\end{align*}
\]

- If the sequence of updates includes

\[(s, u_1), (u_1, u_2), (u_2, u_3), \ldots, (u_k, t),\]

in that order, the shortest distance from \(s\) to \(t\) will be computed correctly (updates are always safe). Note that the sequence of updates does not need to be consecutive.

- Solution: update all edges \(|V| - 1\) times!

- Complexity: \(O(|V| \cdot |E|)\).
function ShortestPaths(G, l, s)

// Input: Graph G(V,E), source vertex s,
//        edge lengths \{l_e:\in E\}, no negative cycles.
// Output: dist[u] has the distance from s,
//         prev[u] has the predecessor in the tree

for all u \in V:
    dist[u] = \infty
    prev[u] = nil

dist[s] = 0
repeat |V| - 1 times:
    for all (u, v) \in E:
        if dist[v] > dist[u] + l(u, v):
            dist[v] = dist[u] + l(u, v)
            prev[v] = u
Bellman-Ford: example

Graphs: Shortest paths

<table>
<thead>
<tr>
<th>Node</th>
<th>Iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>∞</td>
</tr>
<tr>
<td>B</td>
<td>∞</td>
</tr>
<tr>
<td>C</td>
<td>∞</td>
</tr>
<tr>
<td>D</td>
<td>∞</td>
</tr>
<tr>
<td>E</td>
<td>∞</td>
</tr>
<tr>
<td>F</td>
<td>∞</td>
</tr>
<tr>
<td>G</td>
<td>∞</td>
</tr>
</tbody>
</table>
Negative cycles

- What is the shortest distance between S and A?

- A negative cycle produces $-\infty$ distances by endlessly applying rounds to the cycle.

- How to detect negative cycles?
  - Apply Bellman-Ford (update edges $|V| - 1$ times)
  - Perform an extra round and check whether some distance decreases.

Bellman-Ford does not work as it assumes that the shortest path will not have more than $|V| - 1$ edges.
Shortest paths in DAGs

• DAG’s property:

\textit{In any path of a DAG, the vertices appear in increasing topological order.}

• Any sequence of updates that preserves the topological order will compute distances correctly.

• Only one round visiting the edges in topological order is sufficient: $O(|V| + |E|)$.

• How to calculate the longest paths?
  – Negate the edge lengths and compute the shortest paths.
  – Alternative: update with max (instead of min).
function DagShortestPaths(G, l, s)

// Input: DAG G(V,E), source vertex s,
//        edge lengths \{l_e : \in E\}.
// Output: dist[u] has the distance from s,
//         prev[u] has the predecessor in the tree

for all \( u \in V \):
    dist[u] = \infty
    prev[u] = nil

dist[s] = 0
Linearize G
for all \( u \in V \) in linearized order:
    for all \( (u, v) \in E \):
        if \( \text{dist}[v] > \text{dist}[u] + l(u, v) \):
            \( \text{dist}[v] = \text{dist}[u] + l(u, v) \)
            \( \text{prev}[v] = u \)
DAG shortest/longest paths: example

Linearization: S A B C D E F G H

Shortest paths

Longest paths
Shortest paths: summary

**Single-source shortest paths**

<table>
<thead>
<tr>
<th>Graph</th>
<th>Algorithm</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-negative edges</td>
<td>Dijkstra</td>
<td>$O((</td>
</tr>
<tr>
<td>Negative edges</td>
<td>Bellman-Ford</td>
<td>$O(</td>
</tr>
<tr>
<td>DAG</td>
<td>Topological sort</td>
<td>$O(</td>
</tr>
</tbody>
</table>

A related problem: All-pairs shortest paths

- Floyd-Warshall algorithm ($O(|V|^3)$), based on dynamic programming.
- Other algorithms exist.
EXERCISES
Run Dijkstra’s algorithm starting at node A:

- Draw a table showing the intermediate distance values of all the nodes at each iteration
- Show the final shortest-path tree
Run Bellman-Ford algorithm starting at node S:

- Draw a table showing the intermediate distance values of all the nodes at each iteration
- Show the final shortest-path tree
There is a network of roads $G = (V, E)$ connecting a set of cities $V$. Each road in $E$ has an associated length $l_e$. There is a proposal to add one new road to this network, and there is a list $E'$ of pairs of cities between which the new road can be built. Each such potential road $e' \in E'$ has an associated length. As a designer for the public works department you are asked to determine the road $e' \in E'$ whose addition to the existing network $G$ would result in the maximum decrease in the driving distance between two fixed cities $s$ and $t$ in the network. Give an efficient algorithm for solving this problem.
Nesting boxes

A $d$-dimensional box with dimensions $(x_1, x_2, \ldots, x_d)$ nests within another box with dimensions $(y_1, y_2, \ldots, y_d)$ if there exists a permutation $\pi$ on \{1,2, ..., $d$\} such that:

\[ x_{\pi(1)} < y_1, x_{\pi(2)} < y_2, \ldots, x_{\pi(d)} < y_d. \]

a. Argue that the nesting relation is transitive.

b. Describe an efficient method to determine whether or not one $d$-dimensional box nests inside another.

c. Suppose that you are given a set of $n$ $d$-dimensional boxes $\{B_1, B_2, \ldots, B_n\}$. Describe an efficient algorithm to determine the longest sequence $\langle B_{i_1}, B_{i_2}, \ldots, B_{i_k} \rangle$ of boxes such that $B_{i_j}$ nests within $B_{i_{j+1}}$ for $j = 1, 2, \ldots, k - 1$. Express the running time of your algorithm in terms of $n$ and $d$.

Graphs: Minimum Spanning Trees and Maximum Flows

Jordi Cortadella and Jordi Petit
Department of Computer Science
Laying a communication network

Source: https://www.javatpoint.com/applications-of-minimum-spanning-tree
Laying a communication network

Source: https://www.javatpoint.com/applications-of-minimum-spanning-tree
Minimum Spanning Trees

- Nodes are computers
- Edges are links
- Weights are maintenance cost
- Goal: pick a subset of edges such that
  - the nodes are connected
  - the maintenance cost is minimum

The solution is not unique. Find another one!

Property:
An optimal solution cannot contain a cycle.
Minimum Spanning Tree

• Given an undirected graph $G = (V, E)$ with edge weights $w_e$, find a tree $T = (V, E')$, with $E' \subseteq E$, that minimizes

$$\text{weight}(T) = \sum_{e \in E'} w_e .$$

• Greedy algorithm: repeatedly add the next lightest edge that does not produce a cycle.

Note: We will now see that this strategy guarantees an MST.
Properties of trees

- **Definition:** A tree is an undirected graph that is connected and acyclic.

- **Property:** Any connected, undirected graph \( G = (V, E) \) has \( |E| \geq |V| - 1 \) edges.

- **Property:** A tree on \( n \) nodes has \( n - 1 \) edges.
  - Start from an empty graph. Add one edge at a time making sure that it connects two disconnected components. After having added \( n - 1 \) edges, a tree has been formed.

- **Property:** Any connected, undirected graph \( G = (V, E) \) with \( |E| = |V| - 1 \) is a tree.
  - It is sufficient to prove that \( G \) is acyclic. If not, we can always remove edges from cycles until the graph becomes acyclic.

- **Property:** Any undirected graph is a tree iff there is a unique path between any pair of nodes.
  - If there would be two paths between two nodes, the union of the paths would contain a cycle.
The cut property

Suppose edges $X$ are part of an MST of $G = (V, E)$. Pick any subset of nodes $S$ for which $X$ does not cross between $S$ and $V - S$, and let $e$ be the lightest edge across this partition. Then $X \cup \{e\}$ is part of some MST.

Proof (sketch): Let $T$ be an MST and assume $e$ is not in $T$. If we add $e$ to $T$, a cycle will be created with another edge $e'$ across the cut $(S, V - S)$. We can now remove $e'$ and obtain another tree $T'$ with weight$(T') \leq \text{weight}(T)$. Since $T$ is an MST, then the weights must be equal.
The cut property: example

Graphs: MST and Max Flow

© Dept. CS, UPC
Minimum Spanning Tree

Any scheme like this works (because of the properties of trees):

\[ X = \{\} \quad // \text{The set of edges of the MST} \]

repeat \(|V| - 1\) times:

- pick a set \(S \subset V\) for which \(X\) has no edges between \(S\) and \(V - S\)
- let \(e \in E\) be the minimum-weight edge between \(S\) and \(V - S\)
- \(X = X \cup \{e\}\)
MST: two strategies

**Invariant:**
- A set of nodes \( (S) \) is in the tree.

**Progress:**
- The lightest edge with exactly one endpoint in \( S \) is added.

---

**Prim’s algorithm**

---

**Invariant:**
- A set of trees (forest) has been constructed.

**Progress:**
- The lightest edge between two trees is added.
Prim’s algorithm

function Prim(G, w)
// Input: A connected undirected Graph G(V,E)
// with edge weights w(e).
// Output: An MST defined by the vector prev.
for all u ∈ V:
    visited(u) = false
    prev(u) = nil
pick any initial node u₀
visited(u₀) = true
n = 1

// Q: priority queue of edges using w(e) as priority
Q = makequeue()
for each (u₀, v) ∈ E: Q.insert(u₀, v)

while n < |V|:
    (u, v) = deletemin(Q) // Edge with smallest weight
    if not visited(v):
        visited(v) = true
        prev(v) = u
        n = n + 1
    for each (v, x) ∈ E:
        if not visited(x): Q.insert(v, x)

Complexity: O(|E| log |V|)
Prim’s algorithm

A
6
B
2
C
5
D
1
E
2
F
4

Q:

(AD,4) (AB,5) (AC,6)
(DB,2) (DC,2) (DF,4) (AB,5) (AC,6)
(BC,1) (DC,2) (DF,4) (AB,5) (AC,6)
(DC,2) (CF,3) (DF,4) (AB,5) (CE,5) (AC,6)
(CF,3) (DF,4) (AB,5) (CE,5) (AC,6)
(DF,4) (FE,4) (AB,5) (CE,5) (AC,6)
(FE,4) (AB,5) (CE,5) (AC,6)
Kruskal’s algorithm

Informal algorithm:
• Sort edges by weight.
• Visit edges in ascending order of weight and add them as long as they do not create a cycle.

How do we know whether a new edge will create a cycle?

```python
function Kruskal(G, w)

// Input: A connected undirected Graph G(V,E)
//        with edge weights w_e.

// Output: An MST defined by the edges in X.

X = {}  
sort the edges in E by weight
for all (u,v) ∈ E, in ascending order of weight:
    if (X has no path connecting u and v):
        X = X ∪ {(u,v)}
```
Disjoint sets

- A data structure to store a collection of disjoint sets.

- Operations:
  - makeset\( (x) \): creates a singleton set containing just \( x \).
  - find\( (x) \): returns the identifier of the set containing \( x \).
  - union\( (x, y) \): merges the sets containing \( x \) and \( y \).

- Kruskal’s algorithm uses disjoint sets and calls
  - makeset: \(|V|\) times
  - find: \(2 \cdot |E|\) times
  - union: \(|V| - 1\) times
Kruskal’s algorithm

```plaintext
function Kruskal(G, w)

// Input: A connected undirected Graph G(V,E)
//        with edge weights w_e.

// Output: An MST defined by the edges in X.

for all u ∈ V: makeset(u)

X = {}
sort the edges in E by weight
for all (u,v) ∈ E, in ascending order of weight:
    if (find(u) ≠ find(v)):
        X = X ∪ {(u,v)}
        union(u,v)
```

Graphs: MST and Max Flow © Dept. CS, UPC
Disjoint sets

• The nodes are organized as a set of trees. Each tree represents a set.

• Each node has two attributes:
  – parent ($\pi$): ancestor in the tree
  – rank: height of the subtree

• The root element is the representative for the set: its parent pointer is itself (self-loop).

• The efficiency of the operations depends on the height of the trees.

function makeset($x$):
  $\pi(x) = x$
  rank($x$) = 0

function find($x$):
  while $x \neq \pi(x)$: $x = \pi(x)$
  return $x$
function union(x, y):
    \[ r_x = \text{find}(x) \]
    \[ r_y = \text{find}(y) \]
    if \( r_x = r_y \): return

    if rank(\( r_x \)) > rank(\( r_y \)):
        \[ \pi(r_y) = r_x \]
    else:
        \[ \pi(r_x) = r_y \]

    if rank(\( r_x \)) = rank(\( r_y \)):
        rank(\( r_y \)) = rank(\( r_y \)) + 1

function makeset(x):
    \[ \pi(x) = x \]
    rank(x) = 0

function find(x):
    while \( x \neq \pi(x) \): \( x = \pi(x) \)
    return \( x \)
Disjoint sets

After `makeset(A),...,makeset(G)`: 

![Diagram showing initial disjoint set configuration with nodes A, B, C, D, E, F, G.

After `union(A,D), union(B,E), union(C,F)`:

![Diagram showing merged disjoint set configurations with nodes A, B, C, D, E, F, G.

After `union(C,G), union(E,A)`:

![Diagram showing final disjoint set configuration with nodes A, B, C, D, E, F, G.

After `union(B,G)`:

![Diagram showing final disjoint set configuration with nodes A, B, C, D, E, F, G.

**Property:** Any root node of rank $k$ has at least $2^k$ nodes in its tree.

**Property:** If there are $n$ elements overall, there can be at most $n/2^k$ nodes of rank $k$. Therefore, all trees have height $\leq \log n$. 
**Disjoint sets**

**Property 1:** Any root node of rank $k$ has at least $2^k$ nodes in its tree.

**Property 2:** If there are $n$ elements overall, there can be at most $n/2^k$ nodes of rank $k$. Therefore, all trees have height $\leq \log n$.

---

**Property 1: proof by induction**

![Diagram](image)

$\text{rank } k + 1$

$\geq 2^k$

$\text{rank } k$

$\geq 2^k$

$\geq 2^{k+1}$

**Property 2:**

For $n$ nodes, the tallest possible tree could have rank $k$, such that:

$$n \geq 2^k$$

$$k \leq \log_2 n$$

Therefore, find($x$) is $O(\log n)$
Disjoint sets: path compression

• Complexity of Kruskal’s algorithm: \( O(|E| \log |V|) \).
  – Sorting edges: \( O(|E| \log |E|) = O(|E| \log |V|) \).
  – Find + union (2 \cdot |E| times): \( O(|E| \log |V|) \).

• How about if the edges are already sorted or sorting can be done in linear time (weights are integer and small)?

• Path compression:
Disjoint sets: path compression

function find(x):
    if x ≠ π(x): π(x) = find(π(x))
    return π(x)

Amortized cost of find: \(O(1)\)
Kruskal’s cost: \(O(|E|)\)
(if sorting has linear cost)
Max-flow/min-cut problems

How much water can you pump from source to target?

What is the fewest number of green tubes that need to be cut so that no water will be able to flow from the hydrant to the bucket?

https://brilliant.org/wiki/max-flow-min-cut-algorithm/
Max-flow/min-cut problems: applications

• Networks that carry data, water, oil, electricity, cars, etc.
  – How to maximize usage?
  – How to minimize cost?
  – How to maximize reliability?

• Multiple application domains:
  – Computer networks
  – Image processing
  – Computational biology
  – Airline scheduling
  – Data mining
  – Distributed computing
  – ...
Max-flow problem

Model:
- A directed graph $G = (V, E)$.
- Two special nodes $s, t \in V$.
- Capacities $c_e > 0$ on the edges.

Goal: assign a flow $f_e$ to each edge $e$ of the network satisfying:
- $0 \leq f_e \leq c_e$ for all $e \in E$ (edge capacity not exceeded)
- For all nodes $u$ (except $s$ and $t$), the flow entering the node is equal to the flow exiting the node:

$$ \sum_{(w,u) \in E} f_{wu} = \sum_{(u,z) \in E} f_{uz}. $$

Size of a flow: total quantity sent from $s$ to $t$ (equal to the quantity leaving $s$):

$$ \text{size}(f) = \sum_{(s,u) \in E} f_{su}. $$
Max-flow problem: intuition

Find an augmenting path

An augmenting path may reverse some of the flow previously assigned

After augmenting flow by $f$

$c - f$

$s$

$t$

Graphs: MST and Max Flow

© Dept. CS, UPC
Augmenting paths

Given a flow, an **augmenting path** represents a feasible additional flow from $s$ to $t$.

**Augmenting path**

**New flow**
Augmenting paths can have forward and backward edges.

Augmenting path

New flow
Given a flow $f$, an augmenting path is a directed path from $s$ to $t$, which consists of edges from $E$, but not necessarily in the same direction. Each of these edges $e$ satisfies exactly one of the following two conditions:

- $e$ is in the same direction as in $E$ (forward) and $f_e < c_e$. The difference $c_e - f_e$ is called the *slack* of the edge.

- $e$ is in the opposite direction (backward) and $f_e > 0$. It represents the fact that some flow can be borrowed from the current flow.
Residual graph

Graph

Flow

Residual graph

Graphs: MST and Max Flow

© Dept. CS, UPC
Ford-Fulkerson algorithm: example

Flow

Residual

Graphs: MST and Max Flow

© Dept. CS, UPC
Ford-Fulkerson algorithm: example
**Ford-Fulkerson algorithm**

```plaintext
function Ford-Fulkerson(G, s, t)

// Input: A directed Graph G(V,E) with edge capacities c_e.
// s and t and the source and target of the flow.
// Output: A flow f that maximizes the size of the flow.
// For each (u,v) ∈ E, f(v,u) represents its flow.

for all (u,v) ∈ E:
  f(u,v) = c(u,v)   // Forward edges
  f(v,u) = 0       // Backward edges

while there exists a path p = s ⇝ t in the residual graph:
  f(p) = min{f(u,v): (u,v) ∈ p}
  for all (u,v) ∈ p:
    f(u,v) = f(u,v) − f(p)
    f(v,u) = f(v,u) + f(p)
```

© Dept. CS, UPC
Ford-Fulkerson algorithm: complexity

- Finding a path in the residual graph requires $O(|E|)$ time (using BFS or DFS).

- How many iterations (augmenting paths) are required?
  - The worst case is really bad: $O(C \cdot |E|)$, with $C$ being the largest capacity of an edge (if only integral values are used).
  - By selecting the path with fewest edges (using BFS) the maximum number of iterations is $O(|V| \cdot |E|)$.
  - By carefully selecting fat augmenting paths (using some variant of Dijkstra’s algorithm), the number of iterations can be reduced.

- Ford-Fulkerson algorithm is $O(|V| \cdot |E|^2)$ if BFS is used to select the path with fewest edges (Edmonds-Karp algorithm).
Max-flow problem

**Cut:** An \((s, t)\)-cut partitions the nodes into two disjoint groups, \(L\) and \(R\), such that \(s \in L\) and \(t \in R\).

For any flow \(f\) and any \((s, t)\)-cut \((L, R)\):

\[
\text{size}(f) \leq \text{capacity}(L, R).
\]

**The max-flow min-cut theorem:**

The size of the maximum flow equals the capacity of the smallest \((s, t)\)-cut.

**The augmenting-path theorem:**

A flow is maximum iff it admits no augmenting path.
Finding a cut with minimum capacity:

1. Solve the max-flow problem with Ford-Fulkerson.
2. Compute $L$ as the set of nodes reachable from $s$ in the residual graph.
3. Define $R = V - L$.
4. The cut $(L, R)$ is a min-cut.
Bipartite matching

There is an edge between a boy and a girl if they like each other.

Can we pick couples so that everyone has exactly one partner that he/she likes?

Bad matching: if we pick (Aleix, Anna) and (Bernat, Cristina), then we cannot find couples for Berta, Duna, Carles and David.

A **perfect matching** would be: (Aleix, Berta), (Bernat, Duna), (Carles, Anna) and (David, Cristina).
Reduced to a max-flow problem with $c_e = 1$.

**Question:** can we always guarantee an integer-valued flow?

**Property:** if all edge capacities are integer, then the optimal flow found by Ford-Fulkerson’s algorithm is integral. It is easy to see that the flow of the augmenting path found at each iteration is integral.
Extensions of Max-Flow

• **Max-Flow with Edge Demands**
  – Each edge $e$ has a demand $d(e)$. The flow $f$ must satisfy $d(e) \leq f(e) \leq c(e)$.

• **Node Supplies and Demands**
  – An extra flow $x(v)$ can be injected (positive) or extracted (negative) at every vertex $v$. The flow must satisfy:

  $$\sum_{u \in V} f(u \rightarrow v) - \sum_{w \in V} f(v \rightarrow w) = x(v).$$

• **Min-cost Max-Flow**
  – Each edge $e$ has a weight $w_e$. Compute a max-flow of minimum cost:

  $$\text{cost}(f) = \sum_{e \in E} w_e \cdot f(e)$$

• **Max-Weight Bipartite Matching**
  – Each edge $e$ has a weight $w_e$. Find a maximum cardinality matching with maximum total weight.
EXERCISES
Minimum Spanning Trees

- Calculate the shortest path tree from node A using Dijkstra’s algorithm.

- Calculate the MST using Prim’s algorithm. Indicate the sequence of edges added to the tree and the evolution of the priority queue.

- Calculate the MST using Kruskal’s algorithm. Indicate the sequence of edges added to the tree and the evolution of the disjoint sets. In case of a tie between two edges, try to select the one that is not in Prim’s tree.
We want to classify a set of points into $k$ clusters. We define the distance between two points as the Euclidean distance. We define the *spacing* of the clustering as the minimum distance between any pair of points in different clusters.

Describe an algorithm such that, given an integer $k$, finds a $k$-clustering such that *spacing* is maximized. Argue about the complexity of the algorithm.

Note: $k$-clustering of maximum spacing is the basis for the construction of dendograms.
• Find the maximum flow from S to T. Give a sequence of augmenting paths that lead to the maximum flow.
• Draw the residual graph after finding the maximum flow.
• Find a minimum cut between S and T.
Contagious disease

The island of Sodor is home to a large number of towns and villages, connected by an extensive rail network. Recently, several cases of a deadly contagious disease (Covid 19) have been reported in the village of Ffarquhar. The controller of the Sodor railway plans to close down certain railway stations to prevent the disease from spreading to Tidmouth, his home town. No trains can pass through a closed station. To minimize expense (and public notice), he wants to close down as few stations as possible. However, he cannot close the Ffarquhar station, because that would expose him to the disease, and he cannot close the Tidmouth station, because then he couldn’t visit his favorite pub.

Describe and analyze an algorithm to find the minimum number of stations that must be closed to block all rail travel from Ffarquhar to Tidmouth. The Sodor rail network is represented by an undirected graph, with a vertex for each station and an edge for each rail connection between two stations. Two special vertices $F$ and $T$ represent the stations in Ffarquhar and Tidmouth.

For example, given the following input graph, your algorithm should return the number 2.

Source: Jeff Erickson, Algorithms, UIUC, 2015.
Enthusiastic celebration of a sunny day at a prominent northeastern university has resulted in the arrival at the university's medical clinic of 169 students in need of emergency treatment. Each of the 169 students requires a transfusion of one unit of whole blood. The clinic has supplies of 170 units of whole blood. The number of units of blood available in each of the four major blood groups and the distribution of patients among the groups is summarized below.

<table>
<thead>
<tr>
<th>Blood type</th>
<th>A</th>
<th>B</th>
<th>O</th>
<th>AB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supply</td>
<td>46</td>
<td>34</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>Demand</td>
<td>39</td>
<td>38</td>
<td>42</td>
<td>50</td>
</tr>
</tbody>
</table>

Type A patients can only receive type A or O; type B patients can receive only type B or O; type O patients can receive only type O; and type AB patients can receive any of the four types.

Give a maxflow formulation that determines a distribution that satisfies the demands of a maximum number of patients.

Can we have enough blood units for all the students?

Given a digraph $G = (V, E)$ and vertices $s, t \in V$, describe an algorithm that finds the maximum number of edge-disjoint paths from $s$ to $t$.

Note: two paths are edge-disjoint if they do not share any edge.
Trees

Jordi Cortadella and Jordi Petit
Department of Computer Science
Trees

Data are often organized hierarchically

source: https://en.wikipedia.org/wiki/Tree_structure
Filesystems

Trees
Company structure
Genealogical trees
Probability trees
Parse trees

```
S
  /   \
 NP   VP
  / \
 Det Nom
  |   |
  the little bear

V
  |
saw

VP
  /\ \
 NP P
  | |
 Det in Det
  | |  |
  the in the

Adj N
  | |
  fine fat trout

Nom N
  |
  brook
```
Image representation (quad-trees)
Decision trees

Tree: definition

• Graph theory: a tree is an undirected graph in which any two vertices are connected by exactly one path.

• Recursive definition (CS). A non-empty tree $T$ consists of:
  – a root node $r$
  – a list of trees $T_1, T_2, \ldots, T_n$ that hierarchically depend on $r$. 

Diagram:

- A root node $r$ with branches leading to $T_1$, $T_2$, ..., $T_n$. 
- Each $T_i$ is a subtree with its own root and branches to further subtrees.
A is the root node.
Nodes with no children are leaves (e.g., B and P).
Nodes with the same parent are siblings (e.g., K, L and M).
The depth of a node is the length of the path from the root to the node. Examples: depth(A)=0, depth(L)=2, depth(Q)=3.
Tree: representation with linked lists

```cpp
struct TreeNode {
    Type element;
    list<TreeNode> children; // Linked list of children
};
```
struct TreeNode {
    Type element;
    vector<TreeNode> children;  // Vector of children
};
Print a tree

```cpp
struct Tree {
    string name;
    vector<Tree> children;
};

print(const Tree& T, int depth=0);
```
/** Prints a tree indented according to depth.  
 * Pre: The tree is not empty. */
void print(const Tree& T, int depth) {

    // Print the root indented by 2*depth
    cout << string(2*depth, ' ') << T.name << endl;

    // Print the children with depth + 1
    for (const Tree& child: T.children)
        print(child, depth + 1);
}

This function executes a **preorder** traversal of the tree: each node is processed **before** the children.
Print a tree (postorder traversal)

Postorder traversal: each node is processed after the children.
/** Prints a tree (in postorder) indented according to depth. * Pre: The tree is not empty. */
void printPostOrder(const Tree& T, int depth) {

    // Print the children with depth + 1
    for (const Tree& child: T.children)
        printPostOrder(child, depth + 1);

    // Print the root indented by 2*depth
    cout << string(2*depth, ' ') << T.name << endl;
}

This function executes a postorder traversal of the tree: each node is processed after the children.
Binary trees

Nodes with at most two children.

```
struct BinTree {
    Type element;
    BinTree* left;
    BinTree* right;
};
```
Example: expression trees

Expression tree for: \( a + b \times c + (d \times e + f) \times g \)

Postfix representation: \( a \ b \ c \ + \ d \ e \ + \ f \ g \ + \) 

How can the postfix representation be obtained?
Expressions are represented by strings in postfix notation in which the characters ‘a’…‘z’ represent operands and the characters ‘+’ and ‘*’ represent operators.

```cpp
struct ExprTree {
    char op;  // operand or operator
    ExprTree* left;
    ExprTree* right;
};

using Expr = ExprTree*;

/** Builds an expression tree from a correct * expression represented in postfix notation. */
Expr buildExpr(const string& expr);

/** Generates a string with the expression in * infix notation. */
string infixExpr(const Expr T);

/** Evaluates an expression taking V as the value of the * variables (e.g., V[‘a’] contains the value of a). */
int evalExpr(const Expr T, const map<char,int>& V);
```
How to build an expression tree

$$a \ b \ c \ast \ + \ d\ e \ast\ f \ + \ g \ast \ast$$

Trees © Dept. CS, UPC
How to build an expression tree

\[ a \ b \ c \ * \ + \ d \ e \ * \ f \ + \ g \ * \ + \]

Stack

a
How to build an expression tree

Stack

a b c * + d e * f + g * +
How to build an expression tree

$ab\cdot c + de\cdot f + g$
How to build an expression tree

Stack

a b c * + d e * f + g * *
How to build an expression tree

```
  a b c * + d e * f + g * +
```

Stack

```
+   *
|   |
a b c
```

Trees © Dept. CS, UPC
How to build an expression tree

\[ a \cdot b \cdot c + d \cdot e \cdot f + g \cdot \]
How to build an expression tree

```
ab c * + de f * g +
```

Stack

Tree representation:
How to build an expression tree

a b c * + d e * f + g * +

Stack

Trees
How to build an expression tree

\[ a \cdot b + c \cdot d + e \cdot f + g \]
How to build an expression tree

```
   a   b   c
    *   +
     d   e
      *   f
      +
    g
```

Stack
How to build an expression tree

a b c * + d e * f + g ++

Stack
How to build an expression tree

a b c * + d e * f + g * +

Stack
How to build an expression tree
Example: expression trees

```cpp
Expr buildExpr(const string& expr) {
    stack<Expr> S;
    // Visit the chars of the string sequentially
    for (char c: expr) {
        if (c >= 'a' and c <= 'z') {
            // We have an operand in {'a'…'z'}. Create a leaf node.
            S.push(new ExprTree{c, nullptr, nullptr});
        } else {
            // c is an operator ('+' or '*')
            Expr right = S.top();
            S.pop();
            Expr left = S.top();
            S.pop();
            S.push(new ExprTree{c, left, right});
        }
    }
    // The stack has only one element and is freed after return
    return S.top();
}
```

Remember: **using Expr = ExprTree**;
Example: expression trees

```cpp
/** Returns a string with an infix representation of T. */
string infixExpr(const Expr T) {

    // Let us first check the base case (an operand)
    if (T->left == nullptr) return string(1, T->op);

    // We have an operator. Return ( T->left T->op T->right )
    return "(" +
        infixExpr(T->left) +
        T->op +
        infixExpr(T->right) +
        ")";
}

Inorder traversal: node is visited between the left and right children.
```
/** Evaluates an expression taking V as the value of the * variables (e.g., V[‘a’] contains the value of a). */
int evalExpr(const Expr T, const map<char, int>& V) {
    if (T->left == nullptr) return V.at(T->op);
    int l = evalExpr(T->left, V);
    int r = evalExpr(T->right, V);
    return T->op == ‘+’ ? l+r : l*r;
}

/** Example of usage of ExprTree. */
int main() {
    Expr T = buildExpr("abc*+de*f+g++");
    cout << infixExpr(T) << endl;
    cout << “Eval = ” << evalExpr(T, {{‘a’,3}, {‘b’,1}, {‘c’,0}, {‘d’,5},
    {‘e’,2}, {‘f’,1}, {‘g’,6}}) << endl;
    freeExpr(T); // Not implemented yet
}
Exercises

• Design the function `freeExpr`.

• Modify `infixExpr` for a nicer printing:
  – Minimize number of parenthesis.
  – Add spaces around + (but not around *)

• Extend the functions to support other operands, including the unary – (e.g., –a/b).
Tree traversals

**Traversal**: algorithm to visit the nodes of a tree in some specific order.

The actions performed when visiting each node can be a parameter of the traversal algorithm.

```cpp
struct TreeNode {
    Tinfo info;
    TreeNode* left;
    TreeNode* right;
};

using Tree = TreeNode*;

using visitor = void (int &);

// This function matches the type visitor
void print(int& i) {
    cout << i << endl;
}

void traversal(Tree T, visitor v);
```
void preorder(Tree T, visitor v) {
    if (T != nullptr) {
        v(T->elem);
        preorder(T->left, v);
        preorder(T->right, v);
    }
}
Tree traversals

Preorder: A B D G H E I C F J K
Postorder: G H D I E B J K F C A

void preorder(Tree T, visitor v) {
  if (T != nullptr) {
    v(T->elem);
    preorder(T->left, v);
    preorder(T->right, v);
  }
}

void postorder(Tree T, visitor v) {
  if (T != nullptr) {
    postorder(T->left, v);
    postorder(T->right, v);
    v(T->elem);
  }
}
Tree traversals

Preorder:   A B D G H E I C F J K
Postorder:  G H D I E B J K F C A
Inorder:    G D H B E I A J F K C

```c
void inorder(Tree T, visitor v) {
    if (T != nullptr) {
        inorder(T->left, v);
        v(T->elem);
        inorder(T->right, v);
    }
}
```
Tree traversals

Preorder:  A B D G H E I C F J K
Postorder: G H D I E B J K F C A
Inorder:   G D H B E I A J F K C
By levels: A B C D E F G H I J K

```cpp
void byLevels(Tree T, visitor v) {
    queue<Tree> Q; Q.push(T);
    while (not Q.empty()) {
        T = Q.front(); Q.pop();
        if (T != nullptr) {
            v(T->elem);
            Q.push(T->left); Q.push(T->right);
        }
    }
}
```
EXERCISES
Traversals: Full Binary Trees

• A Full Binary Tree is a binary tree where each node has 0 or 2 children.

• Draw the full binary trees corresponding to the following tree traversals:
  – Preorder: 2 7 3 6 1 4 5; Postorder: 3 6 7 4 5 1 2
  – Preorder: 3 1 7 4 9 5 2 6 8; Postorder: 1 9 5 4 6 8 2 7 3

• Given the pre- and post-order traversals of a binary tree (not necessarily full), can we uniquely determine the tree?
  – If yes, prove it.
  – If not, show a counterexample.
Traversals: Binary Trees

• Draw the binary trees corresponding the following traversals:
  – Preorder: 3 6 1 8 5 2 4 7 9; Inorder: 1 6 3 5 2 8 7 4 9
  – Level-order: 4 8 3 1 2 7 5 6 9; Inorder: 1 8 5 2 4 6 7 9 3
  – Postorder: 4 3 2 5 9 6 8 7 1; Inorder: 4 3 9 2 5 1 7 8 6

• Describe an algorithm that builds a binary tree from the preorder and inorder traversals.
We want to draw the skeleton of a binary tree as it is shown in the figure. For that, we need to assign \((x, y)\) coordinates to each tree node. The layout must fit in a predefined bounding box of size \(W \times H\), with the origin located in the top-left corner.

Design the function

```c
void draw(Tree T, double W, double H)
```

to assign values to the attributes \(x\) and \(y\) of all nodes of the tree in such a way that the lines that connect the nodes do not cross.

Suggestion: calculate the coordinates in two steps. First assign \((x, y)\) coordinates using some arbitrary unit. Next, shift/scale the coordinates to exactly fit in the bounding box.
Containers:
Set and Dictionary

Jordi Cortadella and Jordi Petit
Department of Computer Science
Sets and Dictionaries

• A set: a collection of items. The typical operations are:
  – Add/remove one element
  – Does it contain an element?
  – Size?, Is it empty?
  – Visit all items

• A dictionary (map): a collection of key-value pairs. The typical operations are:
  – Put a new key-value pair
  – Remove a key-value pair with a specific key
  – Get the value associated to a key
  – Does it contain a key?
  – Visit all key-value pairs
Sets and Dictionaries

• A dictionary can be treated as a set of keys, each key having an associated value.

• We will focus on the implementation of sets.

Source: Natural Language Processing with Python, by Steven Bird, Ewan Klein and Edward Loper
## Possible implementations of a set

### Unsorted list or vector

<table>
<thead>
<tr>
<th>Operation</th>
<th>Time Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insertion</td>
<td>$O(n)$, if checking for duplicate keys, $O(1)$ otherwise.</td>
</tr>
<tr>
<td>Deletion</td>
<td>$O(n)$ since it has to find the item along the list.</td>
</tr>
<tr>
<td>Lookup</td>
<td>$O(n)$ since the list must be scanned.</td>
</tr>
<tr>
<td>Good for</td>
<td>Small sets.</td>
</tr>
</tbody>
</table>

### Sorted vector

<table>
<thead>
<tr>
<th>Operation</th>
<th>Time Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insertion</td>
<td>$O(n)$ in the worst case (similar to insertion sort)</td>
</tr>
<tr>
<td>Deletion</td>
<td>$O(n)$ since it has to sift the elements after deletion.</td>
</tr>
<tr>
<td>Lookup</td>
<td>$O(\log n)$ with binary search.</td>
</tr>
<tr>
<td>Good for</td>
<td>Read-only collections (only lookups) or very few updates.</td>
</tr>
</tbody>
</table>

Can we have a data structure with efficient insertion/deletion/lookup operations?

Note: $n$ is the number of items in the set.
Binary Search Trees

**BST property**: for every node in the tree with value V:
- All values in the left subtree are smaller than V.
- All values in the right subtree are larger than V.

This is a binary search tree

This is *not* a binary search tree
Sets and dictionaries

Set

Dictionary (Key=Name, Value=Age)

Requirement: keys must be comparable
template<typename T>
class Set {
    public:
        // Constructors, assignment and destructor
        Set();
        Set(const Set& S);
        Set& operator=(const Set& S);
        ~Set();

        // Finding elements
        const T& findMin() const;
        const T& findMax() const;
        bool contains(const T& x) const;
        int size() const;
        bool isEmpty() const;

        // Insert/remove methods
        void insert(const T& x);
        void remove(const T& x);
};
Binary Search Trees: find min/max

Find min: Go to the leftmost element.

Find max: Go to the rightmost element.
Binary Search Trees: contains

Contains: 
  • Move to left/right depending on the value.
  • Stop when:
    ➢ The value is found (contained)
    ➢ No more elements exist (not contained)
Binary Search Trees: insert

Insert 5

Insert:
- Move to left/right depending on the value.
- Stop when the element is found (nothing to do) or a null is found.
- If not found, substitute null by the new element.
remove: simple case (no children)

Before:

```
  6
 /   \
2     8
|     |
1  4   |
|       |
3       
```

After:

```
  6
 /   \
2     8
|     |
1     4
|     |
3     |
```
remove: simple case (one child)

Before:

```
   6
  / \
 2   8
 / \  \
1   4  \
    /  \
   3   
```

After:

```
   6
  /  \
 2   8
   \  
    \ 
     3
```

`remove(4)`
remove: complex case (two children)

1. Find the element.

2. Find the min value of the right subtree.

3. Copy the min value onto the element to be removed.

remove(2)
remove: complex case (two children)

1. Find the element.
2. Find the min value of the right subtree.
3. Copy the min value onto the element to be removed.
4. Remove the min value in the right subtree (simple case).

remove(2)
remove: complex case (two children)

1. Find the element.

2. Find the min value of the right subtree.

3. Copy the min value onto the element to be removed.

4. Remove the min value in the right subtree (simple case).

remove(2)
Visiting the items in ascending order

Question:
How can we visit the items of a BST in ascending order?

Answer:
Using an **in-order** traversal
• Copying and deleting the full tree takes $O(n)$.

• We are mostly interested in the runtime of the insert/remove/contains methods.
  – The complexity is $O(d)$, where $d$ is the depth of the node containing the required element.

• But, how large is $d$?
BST: runtime analysis

• Internal path length (IPL): The sum of the depths of all nodes in a tree. Let us calculate the average IPL considering all possible insertion sequences.

\[
\begin{align*}
\text{ILP} & = 0 \times 1 + 1 \times 2 + 2 \times 3 + 3 \times 5 = 23 \\
\text{Avg. IPL} & = \frac{23}{11} \approx 2.09
\end{align*}
\]
Internal path length (IPL): The sum of the depths of all nodes in a tree. Let us calculate the average IPL considering all possible insertion sequences.

\[ D(n) \] is the IPL of a tree with \( n \) nodes. \( D(1) = 0 \). The left subtree has \( i \) nodes and the right subtree has \( n - i - 1 \) nodes. Thus,

\[
D(n) = D(i) + D(n - i - 1) + (n - 1)
\]

If all subtree sizes are equally likely, then the average value for \( D(i) \) and \( D(n - i - 1) \) is

\[
\frac{1}{n} \sum_{j=0}^{n-1} D(j)
\]
BST: runtime analysis

• Therefore,

\[ D(n) = \frac{2}{n} \left( \sum_{j=0}^{n-1} D(j) \right) + n - 1 \]

• The previous recurrence gives: \( D(n) = O(n \log n) \)

• The average height of nodes after \( n \) random insertions is \( O(\log n) \).

• However, the \( O(\log n) \) average height is not preserved when doing deletions.
Random BST

Source: Fig 4.29 of Weiss textbook
Random BST after $n^2$ insert/removes

Reason: the deletion algorithm is asymmetric (deletes elements from the right subtree)

Source: Fig 4.30 of Weiss textbook
Worst-case runtime: $O(n)$
Balanced trees

• The worst-case complexity for insert, remove and search operations in a BST is $O(n)$, where $n$ is the number of elements.

• Various representations have been proposed to keep the height of the tree as $O(\log n)$:
  – AVL trees
  – Red-Black trees
  – Splay trees
  – B-trees
AVL trees

• Named after Adelson-Velsky and Landis (1962).

• Main idea: invest some additional time to balance the tree each time a new element is inserted or deleted.

• Properties:
  – The height of the tree is always $\Theta(\log n)$.
  – The time devoted to balancing is $O(\log n)$. 
AVL tree: definition

- An AVL tree is a BST such that, for every node, the difference between the heights of the left and right subtrees is at most 1.
AVL tree in action

https://en.wikipedia.org/wiki/AVL_tree
AVL trees

Smallest AVL tree with $h = 9$. 
AVL trees

The smallest AVL tree with $h = 6$.

Smallest AVL tree with $h = 6$.

The important question: what is the size of an AVL tree with height $h$?
• Theorem: The height of an AVL tree with $n$ nodes is $\Theta(\log n)$.

• Proof in two steps:
  – The height is $\Omega(\log n)$.
  – The height is $O(\log n)$. 
The height is $\Omega(\log n)$

- The size $n$ of a tree with height $h$ is:

$$n \leq 1 + 2 + 4 + \cdots + 2^h = 2^{h+1} - 1$$

- Therefore,

$$\log_2(n + 1) - 1 \leq h$$

and $h = \Omega(\log n)$. 
The height is $O(\log n)$

- Let $S(h)$ be the min number of nodes of an AVL tree with height $h$.

- One of the children (e.g., left) must have height $h - 1$. The other child must have height $h - 2$ (because the AVL has min size).

- Therefore,
  \[ S(h) = S(h - 1) + S(h - 2) + 1. \]

- Thus,
  \[ S(h) \geq 2 \cdot S(h - 2). \]

- Given that $S(0) = 1$ and $S(1) = 2$, it can be easily proven, by induction, that:
  \[ S(h) \geq 2^{h/2} \]

- Since $n \geq S(h)$ and $S(h) \geq 2^{h/2}$, then $\log_2 n \geq h/2$:
  \[ h = O(\log n). \]
Height of an AVL tree

• The recurrence

\[ S(h) = S(h - 1) + S(h - 2) + 1 \]

resembles the one of the Fibonacci numbers. A tighter bound can be obtained.

• Theorem: the height of an AVL tree with \( n \) internal nodes satisfies:

\[ h < 1.44 \log_2 (n + 2) - 1.328 \]
Unbalanced insertion: 4 cases

Any newly inserted item may fall into any of the four subtrees (LL, LR, RL or RR).

A new insertion may violate the balancing property. Re-balancing might be required.
Single rotation: the left-left case

Insertion

Rotation
Single rotation: the right-right case

Sets & Dictionaries

© Dept. CS, UPC
Double rotation: the left-right case

Insertion

Single rotation does not work

Double rotation
Double rotation: the right-left case

Insertion

Double rotation
Example: insertions

1. Insertion of 1
2. Insertion of 4
3. Insertion of 5
4. Insertion of 9

RR rotation: 4 and 5 are swapped.
Example: insertions

Initial tree:

```
  2
  |
  5
  |
  4
  |
  9
  |
  3
```

Insert 3:

```
  2
  |
  5
  |
  4
  |
  9
  |
  3
  |
  6
```

RL rotation on 2:

```
  4
  |
  2
  |
  5
  |
  9
  |
  3
  |
  6
```

Insert 6:

```
  4
  |
  2
  |
  5
  |
  9
  |
  3
  |
  6
  |
  1
  |
  3
```

RL rotation on 5:

```
  4
  |
  2
  |
  6
  |
  1
  |
  3
  |
  5
  |
  9
```

Final tree:

```
  4
  |
  2
  |
  6
  |
  1
  |
  3
  |
  5
  |
  9
```
Example: insertions

1. **Insert 7**
   - Initial tree: 4
   - After inserting 7: 4
     - 2
       - 6
         - 1 3
         - 5
         - 9
         - 7
     - 1 3 5 9 7
   - 4

2. **Insert 16**
   - After inserting 16: 4
     - 2
       - 6
         - 1 3
         - 5
         - 9
         - 7
         - 16
     - 1 3 5 9 7 16
   - 4

3. **Insert 0**
   - After inserting 0: 4
     - 2
       - 6
         - 1 3
         - 5
         - 9
         - 7
         - 0
         - 16
     - 1 3 5 9 7 16
   - 4

Sets & Dictionaries
Example: deletion

Remove 5

Apply LL rotation on 3
Example: deletion

RR rotation

Sets & Dictionaries © Dept. CS, UPC
Implementation details

• The height must be stored at each node. Only the unbalancing factor (\{-1,0,1\}) is strictly required.

• The insertion/deletion operations are implemented similarly as in BSTs (recursively).

• The re-balancing of the tree is done when the recursive calls return to the ancestors (check heights and rotate if necessary).
Complexity

- Single and double rotations only need the manipulation of few pointers and the height of the nodes ($O(1)$).

- Insertion: the height of the subtree after a rotation is the same as the height before the insertion. Therefore, at most only one rotation must be applied for each insertion.

- Deletion: more complicated. More than one rotation might be required.

- Worst case for deletion: $O(\log n)$ rotations (a chain effect from leaves to root).
EXERCISES
• Starting from an empty BST, depict the BST after inserting the values 32, 15, 47, 67, 78, 39, 63, 21, 12, 27.

• Depict the previous BST after removing the values 63, 21, 15 and 32.
Merging BSTs

• Describe an algorithm to generate a sorted list from a BST. What is its cost?

• Describe an algorithm to create a balanced BST from a sorted list. What is its cost?

• Describe an algorithm to create a balanced BST that contains the union of the elements of two BSTs. What is its cost?
Depict the three AVL trees after sequentially inserting the values 31, 32 and 33 in the following AVL tree:
• Build an AVL tree by inserting the following values: 15, 21, 23, 11, 13, 8, 32, 33, 27. Show the tree before and after applying each rotation.

• Depict the AVL tree after removing the elements 23 and 21 (in this order). When removing an element, move up the largest element of the left subtree.
Hashing

Jordi Cortadella and Jordi Petit
Department of Computer Science
We want to keep a database of the cars inside a parking lot. The database is automatically updated each time the cameras at the entry and exit points of the parking read the plate of a car.

Each plate is represented by a free-format short string of alphanumeric characters (each country has a different system).

The following operations are needed:
- Add a plate to the database (when a car enters).
- Remove a plate from the database (when a car exits).
- Check whether a car is in the parking.

Constraint: we want the previous operations to be very efficient, i.e., executed in constant time.
(This constraint is overly artificial, since the activity in a parking lot is extremely slow compared to the speed of a computer.)
Naïve implementation options

• Lists, vectors or binary search trees are not valid options, since the operations take too long:
  – Unsorted lists: adding takes $O(1)$. Removing/checking takes $O(n)$.
  – Sorted vector: adding/removing takes $O(n)$. Checking takes $O(\log n)$.
  – AVL trees: adding/removing/checking takes $O(\log n)$.

• A (Boolean) vector with one location for each possible plate:
  – The operations could be done in constant time!, but ...
  – The vector would be extremely large (e.g., only the Spanish system can have 80,000,000 different plates).
  – We may not even know the size of the domain (all plates in the world).
  – Most of the vector locations would be “empty” (e.g. assume that the parking has 1,000 places).

• Can we use a data structure with size $O(n)$, where $n$ is the size of the parking?
A hash function maps data of arbitrary size to a table of fixed size. Important questions:

- How to design a good hash function?
- The hash function is not injective. How to handle collisions?
Hash function

• We can calculate the location for item $x$ as

$$h(x) \mod m$$

where $h$ is the hash function and $m$ is the size of the hash table.

• A good hash function must scatter items *randomly* and *uniformly* (to minimize the impact of collisions).

• A hash function must also be *consistent*, i.e., give the same result each time it is applied to the same item.
Hashing the plates: some attempts

• Add the last three characters (e.g., ASCII codes) of plate:

\[ h(x) = x_{n-1} + x_{n-2} + x_{n-3} \]

Bad choice: For the Spanish system, this would concentrate the values between 198 (BBB) and 270 (ZZZ).

• Multiply the last three characters:

\[ h(x) = x_{n-1} \cdot x_{n-2} \cdot x_{n-3} \]

The values are distributed between 287,496 and 729,000. However the distribution is not uniform. The last three characters denote the age of the car. The population of new cars is larger than the one of old cars (e.g., about 15% of the cars are less than 1-year old).

Moreover: consecutive plates would fall into the same slot. Some companies (e.g., car renting) have cars with consecutive plates and they could be located in the neighbourhood of the parking lot.
Hashing the plates: some attempts

• Multiply all characters of the plate:

\[ h(x) = x_0 \cdot x_1 \cdots x_{n-1} \]

Better choice, but not fully random and uniform. Two plates with permutations of characters would fall into the same slot, e.g., 3812 DXF and 8321 FDX.

• The perfect hash function does not exist, but using prime numbers is a good option since most data have no structure related to prime numbers.

• Where can we use prime numbers?
  – In the size of the hash table
  – In the coefficients of the hash function
Example of hash function for strings

- A usual hash function for a string with size $n$ is as follows:

$$h(x) = \sum_{i=0}^{n-1} x_i \cdot p^i$$

where $p$ is a prime number and $x_i$ is the character at location $i$. This function can be efficiently implemented using Horner’s rule for the evaluation of a polynomial.

- Here is a slightly different implementation (reversed string):

```c++
/** Hash function for strings */
unsigned int hash(const string& key, int tableSize) {
    unsigned int hval = 0;
    for (char c: key) hval = 37*hval + c;
    return hval%tableSize;
}
```
Handling collisions

• A collision is produced when

\[ h(x_1) \equiv h(x_2) \mod m \]

• There are two main strategies to handle collisions:
  – Using lists of items with the same hash value (separate chaining)
  – Using alternative cells in the same hash table (linear probing, double hashing, ...)
Handling collisions: separate chaining

Each slot is a list of the items that have the same hash value.

Load factor: $\lambda = \frac{\text{number of items}}{\text{table size}}$

$\lambda$ is the average length of a list.

A successful search takes about $\lambda/2$ links to be traversed, on average.

Table size: make it similar to the number of expected items.
Common strategy: when $\lambda > 1$, do rehashing.
Handling collisions: using the same hash table

• If the slot is occupied, find alternative cells in the same table. To avoid long trips finding empty slots, the load factor should be below $\lambda = 0.5$.

• Deletions must be “lazy” (slots must be invalidated but not deleted, thus avoiding truncated searches).

• **Linear probing:** if the slot is occupied, use the next empty slot in the table.

• **Double hashing:** if the slot is occupied using the first hash function $h_1$, use a second hash function $h_2$. The sequence of slots that is visited is $h_1(x)$, $h_1(x) + h_2(x)$, $h_1(x) + 2h_2(x)$, etc.
An example

Insertion of the elements 54, 26, 93, 17, 77, 31, 44, 55, 20.
Hash function: $h(x) = x \mod 11$.

Separate chaining:

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Linear probing:

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>77</td>
<td>44</td>
<td>55</td>
<td>20</td>
<td>26</td>
<td>93</td>
<td>17</td>
<td></td>
<td></td>
<td>31</td>
<td>54</td>
</tr>
</tbody>
</table>

What if we remove 55? Use lazy deletion!
Rehashing

• When the table gets too full, the probability of collision increases (and the cost of each operation).

• Rehashing requires building another table with a larger size and rehash all the elements to the new table. Running time: $O(n)$.

• New size: $2n$ (or a prime number close to it). Rehashing occurs very infrequently and the cost is amortized by all the insertions. The average cost remains constant.
The hash table occupies $O(M + n)$ space.
Each slot has $n/M$ items, on average.
The runtime to find an item is $O(n/M)$, on average.

<table>
<thead>
<tr>
<th>Cases</th>
<th>Space: $O(n + M)$</th>
<th>Time: $O(n/M)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M \gg n$</td>
<td>$O(M)$</td>
<td>$O(1)$</td>
</tr>
<tr>
<td>$n \gg M$</td>
<td>$O(n)$</td>
<td>$O(n)$</td>
</tr>
<tr>
<td>$M = O(n)$</td>
<td>$O(n)$</td>
<td>$O(1)$</td>
</tr>
</tbody>
</table>

The best strategy is to have $M = O(n)$ that allows to maintain a constant-time access without wasting too much memory.

Rehashing should be applied to maintain $M = O(n)$. 
## Binary Search Trees vs. Hash Tables

Not a clear winner

<table>
<thead>
<tr>
<th>Operation</th>
<th>Binary Search Tree</th>
<th>Hash Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insertion/Deletion/Lookup</td>
<td>$O(\log n)$</td>
<td>$O(1)$</td>
</tr>
<tr>
<td>Sorted Iteration</td>
<td>In-order traversal: $O(n)$</td>
<td>Needs an extra sorted vector: $O(n \log n)$</td>
</tr>
<tr>
<td>Hash function</td>
<td>Not required</td>
<td>Required</td>
</tr>
<tr>
<td>Total order</td>
<td>Required</td>
<td>Not required</td>
</tr>
<tr>
<td>Range search</td>
<td>$O(\log n)$</td>
<td>$O(n)$</td>
</tr>
</tbody>
</table>

Hashing  
© Dept. CS, UPC  
15
Hash functions are used to guarantee the integrity of data (files, messages, etc) when distributed between different locations.

Different hashing algorithms exist: MD5, SHA1, SHA255, ...

The probability of collision is extremely low.
Security is based on the fact that hashing functions are cryptographic (not reversible).

Be careful: there are databases of hash values for “popular” passwords (e.g., 123456, qwerty, password, barcelona, samsung,…).
EXERCISES
Given the values \{2341, 4234, 2839, 430, 22, 397, 3920\}, a hash table of size 7, and hash function \( h(x) = x \mod 7 \), show the resulting tables after inserting the values in the given order with each of these collision strategies:

- Separate chaining
- Linear probing
Let us assume that we have a list with $n$ elements. Design an algorithm that can check that all elements are different. Analyze the complexity of the algorithm considering different data structures:

- Checking the elements without any additional data structure, i.e., using the same list.
- Using AVLs.
- Using hash tables.
Cryptography

Jordi Cortadella and Jordi Petit
Department of Computer Science
Where do we need cryptography?

- Communication (e.g., sending private emails).
- Digital signatures, i.e., guarantee that digital documents are authentic.
- Network services over unsecure networks (e.g., secure shell (ssh) for remote login, file transfers, remote command execution, etc.).
- HyperText Transfer Protocol Secure (HTTPS): secure communication on Internet.
- Cryptocurrencies (e.g., bitcoin)
• How can we avoid an eavesdropper (Eve) to overhear a message sent from Alice to Bob?

• Solution: encrypt the message!
• The encryption function $e(x)$ must be invertible. The inverse is $d(\cdot)$.

• Two schemes:
  – The traditional: Secret-key protocols (symmetric).
  – The modern: Public-key protocols (asymmetric).
• Alice and Bob have to meet privately and chose a secret key.

• They can use the secret key to mutually exchange messages.

• There are many secret-key protocols. We will explain two of them:
  – XOR encoding.
  – Advanced Encryption Standard (AES).
Secret-key protocol: XOR encoding

- A secret key $r$ is chosen (a binary string).

- The encoding and decoding functions are identical: $e_r(x) = d_r(x) = x \oplus r$.

- Example: $r = 11011100$.

- It is convenient that the bits of $r$ are randomly generated.

- Still, this is not a very robust scheme since Eve can figure out important information by listening several messages.
Secret-key protocol: XOR encoding

• If the key is too short, it needs to be applied many times (once for each block). Messages often show similarities and repeated patterns (same header, same tail, long sequences of zeros, ...).

• If we send two messages, $e_r(x)$ and $e_r(z)$, then $e_r(x) \oplus e_r(z)$ may reveal important information. It is convenient to change the key at every message (one-time pad).

$$r = 0101$$

$$x = 1001\,0000\,0000\,0000\,0000\,0000\,1011\,1010$$

$$z = 1001\,1001\,0000\,0000\,0000\,0000\,0000\,0111$$

$$e_r(x) = 1100\,0101\,0101\,0101\,0101\,1110\,1111$$

$$e_r(z) = 1100\,1100\,0101\,0101\,0101\,0101\,0010$$

$$e_r(x) \oplus e_r(z) = 0000\,1001\,0000\,0000\,0000\,0000\,1011\,1101$$

Same header.
Same sender?

The key !!!

Long identical sequence.
Maybe zeros?
Secret-key protocol: AES

• AES: Advanced Encryption Standard.

• Established as a standard by the U.S. National Institute of Standards and Technology (NIST) in 2001.

• Very robust and used worldwide.

• A family of ciphers with different key and block sizes (key sizes: 128, 196 and 256 bits).
AES scheme

Plain text (128 bits)  +  Round Key (0)
                     +  Round Key (i)
                     +  Round Key

SubBytes
ShiftRows
MixColumn

Cipher text (128 bits)

Cipher Key

Key Expansion

N-1 Rounds

Final Round

Cryptography
© Dept. CS, UPC
AES steps

1. **SubBytes**
   - Transform each byte of the state with the AES S-box.

2. **ShiftRows**
   - Shift the rows of the state matrix to the left by a certain number of positions.
   - No change for first row, shift left by 1 for second row, shift left by 2 for third row, shift left by 3 for fourth row.

3. **MixColumns**
   - Perform a matrix multiplication with the AES MixColumns matrix over GF(2^8).
   - Multiply each column of the state matrix by the MixColumns matrix and reduce the result modulo the irreducible polynomial.

4. **AddRoundKey**
   - XOR the state with the round key.

The process is repeated for each round until the final round, where **SubBytes** is applied without **ShiftRows**.
Secret-key protocols: problems

Every channel requires a different key

The key cannot be transmitted through the communication channel!
Public-key protocols
Public-key protocols
Public-key protocols

Cryptography
Public-key protocols

• Each participant generates a public key \( P \) and a (private) secret key \( S \). Public keys are revealed to everybody.

• The public/secret keys are a matched pair, i.e.,

\[
M = S(P(M)) = P(S(M))
\]

• If Alice has the pair \( (P_A, S_A) \), anybody can compute \( P_A(X) \), but only Alice can compute \( S_A(X) \).

• If Bob wants to send a secret message \( M \) to Alice, Bob will compute \( X = P_A(M) \) and send it to Alice. Only Alice will be able to decipher the message: \( M = S_A(X) \).
Public-key protocols

But, how to create a cryptosystem like this?
Using number theory.
RSA cryptosystem

• Public-key cryptosystem (Rivest-Shamir-Adleman, 1977).

• Based upon number theory: modular arithmetic and prime numbers.

• Security: based on the fact that factoring a large number (product of two large primes) is hard.
Bézout’s identity

• **Lemma:** If $d$ divides both $a$ and $b$, and $d = ax + by$ for some integers $x$ and $y$, then necessarily $d = \gcd(a, b)$.

• **Proof:**
  – Clearly, $d \leq \gcd(a, b)$, since $d$ is a divisor of $a$ and $b$.
  – Since $\gcd(a, b)$ is a divisor of $a$ and $b$, it must also be a divisor of $ax + by = d$. This implies that $\gcd(a, b) \leq d$.
  – Therefore, $d = \gcd(a, b)$. 
Extended Euclid’s algorithm

```plaintext
function extendedGcd(a, b)
  // Input: two positive integers a and b, with a ≥ b ≥ 0
  // Output: Integers x, y, d such that d = gcd(a, b) and ax + by = d
  if b = 0: return (1, 0, a)
  (x′, y′, d) = extendedGcd(b, a mod b)
  return (y′, x′ − [a/b]y′, d)
```

Proof by induction using the identity

\[ a \mod b = a - \lceil a/b \rceil b \]

<table>
<thead>
<tr>
<th>extgcd(25,11)</th>
<th>(4,-9,1)</th>
<th>4 \cdot 25 - 9 \cdot 11 = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>extgcd(11,3)</td>
<td>(-1,4,1)</td>
<td>-1 \cdot 11 + 4 \cdot 3 = 1</td>
</tr>
<tr>
<td>extgcd(3,2)</td>
<td>(1,-1,1)</td>
<td>1 \cdot 3 - 1 \cdot 2 = 1</td>
</tr>
<tr>
<td>extgcd(2,1)</td>
<td>(0,1,1)</td>
<td>0 \cdot 2 + 1 \cdot 1 = 1</td>
</tr>
<tr>
<td>extgcd(1,0)</td>
<td>(1,0,1)</td>
<td>1 \cdot 1 + 0 \cdot 0 = 1</td>
</tr>
</tbody>
</table>
Modular arithmetic: properties

• $x \mod N$ is the remainder when $x$ is divided by $N$.

\[
x \equiv y \pmod{N} \iff N \text{ divides } (x - y).
\]

• Addition and multiplication. Let us assume $a_1 \equiv b_1 \pmod{N}$ and $a_2 \equiv b_2 \pmod{N}$, then

\[
\begin{align*}
    a_1 + a_2 & \equiv b_1 + b_2 \pmod{N} \\
    a_1 \cdot a_2 & \equiv b_1 \cdot b_2 \pmod{N}
\end{align*}
\]

• Exponentiation. Let us assume $a \equiv b \pmod{N}$, then

\[
a^k \equiv b^k \pmod{N}
\]
Modular arithmetic: properties

• Given $a$ and $N$, we say that $x$ is the multiplicative inverse of $a$ (mod $N$) if

$$ax \equiv 1 \pmod{N}.$$  

Example. The multiplicative inverse of 4 (mod 7) is 2:

$$4 \cdot 2 \equiv 1 \pmod{7}$$

• When $\gcd(a, N) = 1$, the multiplicative inverse of $a$ always exists and can be calculated by the extended Euclid’s algorithm:

$$ax + Ny = 1 \implies x \text{ is } a \text{’s inverse } (\text{mod } N)$$
Let $p$ and $q$ be any two primes and $N = pq$. 

$\phi(N) = (p - 1)(q - 1)$ is the totient of $N$, i.e., the number of positive integers smaller than $N$ which are co-prime to $N$.

For any $e$ co-prime to $\phi(N)$:

1. The mapping $x \mapsto x^e \mod N$ is a bijection on \{0,1,...,N − 1\}.

2. The inverse mapping can be obtained as follows. Let $d$ be the inverse of $e$ modulo $\phi(N)$. Then for all $x \in \{0, ..., N - 1\}$,

\[(x^e)^d \equiv x \mod N.\]
Bob chooses public and secret keys:
- Bob picks two large random primes, \( p \) and \( q \).
- The public key is \( (N, e) \), where \( N = pq \) and \( e \) is a small number co-prime to \((p - 1)(q - 1)\).
- The secret key is \( d \), the inverse of \( e \) modulo \((p - 1)(q - 1)\), computed using the extended Euclid’s algorithm.

Alice sends a message \( x \) to Bob:
- Alice takes Bob’s public key \( (N, e) \) and sends \( y = (x^e \mod N) \).
- Bob decodes the message by computing \( y^d \mod N \).
The RSA cryptosystem: example

• Let \( p = 5 \) and \( q = 17 \), thus \( N = 85 \) and \( \phi(N) = 64 \).

• Let \( e = 3 \). It satisfies: \( \gcd(e, \phi(N)) = \gcd(3, 64) = 1 \).

• We calculate \( d = 3^{-1} \mod 64 = 43 \) using extended Euclid’s algorithm:

\[
43 \cdot 3 - 2 \cdot 64 = 1
\]

Note: the algorithm gives \( 1 \cdot 64 - 21 \cdot 3 = 1 \), but \( -21 = 43 \) (mod 64)

• Let us consider the message \( x = 12 \).
  – The sender must encrypt \( x \) as \( y = 12^3 \mod 85 = 28 \).
  – The receiver must decrypt \( y \) by computing \( x = 28^{43} \mod 85 = 12 \).

• Remember: \( x^k \) can be efficiently computed with \( \log_2 k \) multiplications.
  Note: Multiplication and division of “long” numbers is required (similar to multiplication of polynomials).
Why is RSA secure?

• Typical sizes for $p$ and $q$ are 1024-bit numbers with values larger than $2^{1023.5} \approx 1.8 \times 10^{308}$.

• Eve knows the public key $(N, e)$ and the message $y$. How can she guess $x$? There are two options:

1. Try all possible values of $x$ and check whether $y = x^e \mod N$. But $x$ is a large $n$-bit number and checking all values would take exponential time (impractical).

2. Try to guess $d$ and calculate $x^d \mod N$. This would require to calculate the inverse of $e$ modulo $(p - 1)(q - 1)$. But $p$ and $q$ are not known unless the factors of $N$ are calculated. Factoring is still a hard problem.
Hybrid cryptosystems

- Public-key cryptosystems (e.g., RSA) are convenient (no need to share keys) but computationally expensive. Secret-key (symmetric) cryptosystems (e.g. AES) are more efficient. Both can be combined.

- Bob wants to send an encrypted message to Alice:
  - Bob generates a new symmetric key $k$ and encrypts the data with this key (using AES).
  - Bob encrypts $k$ using Alice’s public key (using RSA).
  - Bob sends both encryptions to Alice.

- Alice wants to decrypt Bob’s message:
  - Alice uses her private key to decrypt the encrypted symmetric key $k$.
  - Alice uses the symmetric key $k$ to decrypt the data (using AES).
A CHF maps data of arbitrary size to a fixed-size bit string.
Cryptographic hash function (CHF)

• Properties:
  – Easy to compute.
  – Pre-image resistance: if \( y = h(x) \), it is difficult to find \( x \) from \( y \).
  – Collision resistance: It is difficult to find two inputs, \( x_1 \) and \( x_2 \), such that \( h(x_1) = h(x_2) \).

• Popular CHFs:
  – Message Digest: MD2, MD4, MD5 and MD6. It is a 128-bit hash function.
  – Secure Hash Function: SHA-0, SHA-1, SHA-2, SHA-3. They produce hash values with 160 bits (SHA-1) or 256 bits (SHA-2).
  – And some others ...
One step of SHA-1

The result is “accumulated” to the result of previous steps.
Example: SHA-1

**Input**

- **Fox**
  - cryptographic hash function
  - Hash value: DFCD 3454 BBEA 788A 751A 696C 24D9 7009 CA99 2D17

- The red fox jumps over the blue dog
  - cryptographic hash function
  - Hash value: 0086 46BB FB7D CBE2 823C ACC7 6CD1 90B1 EE6E 3ABC

- The red fox jumps over the blue dog
  - cryptographic hash function
  - Hash value: 8FD8 7558 7851 4F32 D1C6 76B1 79A9 0DA4 AEFE 4819

- The red fox jumps over the blue dog
  - cryptographic hash function
  - Hash value: FCD3 7FDB 5AF2 C6FF 915F D401 C0A9 7D9A 46AF FB45

- The red fox jumps over the blue dog
  - cryptographic hash function
  - Hash value: 8ACA D682 D588 4C75 4BF4 1799 7D88 BCF8 92B9 6A6C
Digital signatures

• A scheme to guarantee that a message is authentic.

• Consider the following case:
  – Alice sends a document (possibly unencrypted) to Bob and wants Bob to electronically sign the document.
  – Bob “signs” the document and sends it back to Alice.

• Questions:
  – How does Alice know that the document has not been altered? \(\rightarrow\) integrity.
  – How does Alice know that Bob has signed the document (and not somebody else)? \(\rightarrow\) authentication.
Digital signatures

**Alice**

SHA-2

\[ P_B \left( S_B \left( \begin{array}{c} h \end{array} \right) \right) \]

**Bob** \((P_B, S_B)\)

SHA-2 (hash)

\[ h \]

**Signature**

\[ S_B \left( \begin{array}{c} h \end{array} \right) \]

Yes: the document has not been altered (same h) and has been signed by Bob (encrypted with \( S_B \))

No: the document has been altered (different h) or has not been signed by Bob (encrypted with \( S_X \))

Cryptography © Dept. CS, UPC
The pending challenge

How to generate large prime numbers?

(not explained in this lecture)
Internet + Privacy → Cryptography
EXERCISES
Simple cryptographic hash

We want to use the XOR operator $\oplus$ for cryptographic hashing as follows. We split every message $M$ into blocks $B_i$ of 5 bits, e.g., $M = 11101 \cdot 00011 \cdot 10100 \cdot 110$. In case the length is not a multiple of 5, additional zeroes are added at the end of the message.

For a message $M$ with $k$ blocks, we define the cryptographic hash $h$ as follows:

$$h(M) = B_1 \oplus B_2 \oplus \cdots \oplus B_k.$$ 

where $\oplus$ means the bitwise application of XOR. For example, $01110 \oplus 11010 = 10100$.

- What would be the output $h(M)$ for the previous message $M$?

- If we change one bit of a message, does the output change a lot?

- Assume that we know $h(M)$ and the length of $M$. Is it easy to find another $M'$ with the same length such that $h(M) = h(M')$? Justify your answer.
Simple RSA

Assume you have $p = 5$ and $q = 7$.

- Which is the smallest value for $e$?
- What is the corresponding value for $d$?
- Encrypt the message $M = 3$.
- Find all possible pairs $(e, d)$ valid for this cryptosystem.
Implement an RSA cryptosystem

• Given two primes, \( p \) and \( q \), design an RSA cryptosystem (in C++ or python) as follows:
  - Let \( N = p \cdot q \). Find the smallest \( e \geq 3 \), such that \((N, e)\) can be used as public key. Use the extended gcd algorithm.
  - Find \( d \) that can be used for secret key.
  - Implement the function \( \text{encode}(x, e, N) \) that computes \( x^e \mod N \). This function must be efficient. Note: assume that \( N^2 \) can be represented as an int.
  - Implement a function to double check, for \( 0 \leq x < N \), that \( \text{encode}(\text{encode}(x, e, N), d, N) = x \).

• Example: \( p = 79 \), \( q = 491 \).

Public key: \((38789, 11)\), Secret key: \(31271\).

\( e(2) = 2048 \), \( e(19) = 23855 \), \( e(32757) = 4 \),
\( e(38788) = 38788 \), \( e(10) = 18550 \).
Fast Fourier Transform

Jordi Cortadella and Jordi Petit
Department of Computer Science
Why Fourier Transform?
Fourier series

• Periodic function $f(t)$ of period 1:

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(2\pi nt) + \sum_{n=1}^{\infty} b_n \sin(2\pi nt)$$

• Fourier coefficients:

$$a_n = 2 \int_{0}^{T} f(t) \cos(2\pi nt) \, dt, \quad b_n = 2 \int_{0}^{T} f(x) \sin(2\pi nt) \, dt$$

• Fourier series is fundamental for signal analysis (to move from time domain to frequency domain, and vice versa)
Discrete-time signals

\[ x[t] \]

© Dept. CS, UPC
Polynomial representation

\[ P(x) = x^3 - 2x^2 - 3x + 1 \]

\[ P(x) = (1, -2, -3, 1) \]

Coefficient representation

Evaluation

Interpolation

Point-value representation

\[ P(x) = \{(-1, 1), (0, 1), (1, -3), (2, -5)\} \]
Polynomials: coefficient representation

- A polynomial is represented as a vector of coefficients \((a_0, a_1, \ldots, a_{n-1})\):

  \[
  A(x) = 2x^4 + x^2 - 4x + 3
  \]

  \[
  A = (3, -4, 1, 0, 2)
  \]

- Addition: \(O(n)\)

  \[
  A(x) + B(x) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_{n-1} + b_{n-1})x^{n-1}
  \]

- Evaluation: \(O(n)\) using Horner’s method

  \[
  A(x) = a_0 + (x(a_1 + x(a_2 + \cdots + x(a_{n-2} + x(a_{n-1}))) \cdots))
  \]

- Multiplication: \(O(n^2)\) using brute force

  \[
  A(x) \cdot B(x) = \sum_{i=0}^{2n-2} c_i x^i, \quad \text{where} \quad c_i = \sum_{j=0}^{i} a_j b_{i-j}
  \]
Polynomials: point-value representation

- **Fundamental Theorem (Gauss):** A degree-$n$ polynomial with complex coefficients has exactly $n$ complex roots.

- **Corollary:** A degree-$n$ polynomial $A(x)$ is uniquely identified by its evaluation at $n + 1$ distinct values of $x$. 
Polynomials: point-value representation

- A polynomial is represented as a set of pairs \((x_i, y_i)\):

  \[
  A(x) = \{(x_0, y_0), \ldots, (x_{n-1}, y_{n-1})\}
  \]

  \[
  B(x) = \{(x_0, z_0), \ldots, (x_{n-1}, z_{n-1})\}
  \]

- Addition: \(O(n)\)

  \[
  A(x) + B(x) = \{(x_0, y_0 + z_0), \ldots, (x_{n-1}, y_{n-1} + z_{n-1})\}
  \]

- Multiplication: \(O(n)\), but with \(2n - 1\) points

  \[
  A(x) \cdot B(x) = \{(x_0, y_0 \cdot z_0), \ldots, (x_{n-1}, y_{n-1} \cdot z_{n-1})\}
  \]

- Interpolation: \(O(n^2)\) using Lagrange’s formula

  \[
  A(x) = \sum_{k=0}^{n-1} y_k \frac{\prod_{j \neq k} (x - x_j)}{\prod_{j \neq k} (x_k - x_j)}
  \]
Interpolation: Lagrange polynomials

\[ A(x) = \sum_{k=0}^{n-1} y_k \frac{\prod_{j \neq k} (x - x_j)}{\prod_{j \neq k} (x_k - x_j)} \]

\[ A(x) = \{(−1,−1), (0,−2), (2,2)\} \]

\[ A(x) = -1 \frac{(x - 0)(x - 2)}{(-1 - 0)(-1 - 2)} - 2 \frac{(x + 1)(x - 2)}{(0 + 1)(0 - 2)} + 2 \frac{(x + 1)(x - 0)}{(2 + 1)(2 - 0)} \]

\[ A(x) = -\frac{x(x - 2)}{3} + (x + 1)(x - 2) + \frac{(x + 1)x}{3} \]

\[ A(x) = x^2 - 2 \]
Interpolation: Lagrange polynomials

\[ A(x) = \{(-1, -1), (0, -2), (2, 2)\} \]

\[ A(x) = x^2 - 2 \]
Could we have an efficient algorithm to move from coefficient to point-value representation and vice versa?
From coefficients to point-values

Given a polynomial \( a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1} \), evaluate it at \( n \) different points \( x_0, \ldots, x_{n-1} \):

\[
\begin{bmatrix}
  y_0 \\
  y_1 \\
  y_2 \\
  \vdots \\
  y_{n-1}
\end{bmatrix} =
\begin{bmatrix}
  1 & x_0 & x_0^2 & \cdots & x_0^{n-1} \\
  1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\
  1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  1 & x_{n-1} & x_{n-1}^2 & \cdots & x_{n-1}^{n-1}
\end{bmatrix}
\begin{bmatrix}
  a_0 \\
  a_1 \\
  a_2 \\
  \vdots \\
  a_{n-1}
\end{bmatrix}
\]

**Runtime:** \( O(n^2) \) matrix-vector multiplication (apply Horner \( n \) times).

**Horner’s rule:**

\[
p(x) = a_0 + x (a_1 + x (a_2 + x( a_3 + \cdots + x (a_{n-1} + x a_n) \cdots )))
\]
Evaluation by divide-and-conquer


- We want to evaluate $A(x)$ at $n$ different points. Let us choose them to be positive-negative pairs: $\pm x_0, \pm x_1, \ldots, \pm x_{n/2-1}$

- The computations for $A(x_i)$ and $A(-x_i)$ overlap a lot.

- Split the polynomial into odd and even powers

  $3 + 4x + 6x^2 + 2x^3 + x^4 + 10x^5 = (3 + 6x^2 + x^4) + x(4 + 2x^2 + 10x^4)$

- The terms in parenthesis are polynomials in $x^2$:

  $$A(x) = A_e(x^2) + xA_o(x^2)$$
Evaluation by divide-and-conquer

• The calculations needed for $A(x_i)$ can be reused for computing $A(-x_i)$.

$$A(x_i) = A_e(x_i^2) + x_iA_o(x_i^2)$$
$$A(-x_i) = A_e(x_i^2) - x_iA_o(x_i^2)$$

• Evaluating $A(x)$ at $n$ paired points

$$\pm x_0, \pm x_1, \ldots, \pm x_{n/2-1}$$

reduces to evaluating $A_e(x)$ and $A_o(x)$ at just $n/2$ points: $x_0^2, \ldots, x_{n/2-1}^2$
Evaluation by divide-and-conquer

If we could recurse, we would get a running time:

\[ T(n) = 2 \cdot T(n/2) + O(n) = O(n \log n) \]

But can we recurse?
Evaluation by divide-and-conquer

Evaluate: $A(x)$
degree $\leq n - 1$

Evaluate:
$A_e(x^2)$ and $A_o(x^2)$
degree $\leq n/2 - 1$

The problem:
We need $x_0^2$ and $x_1^2$ to be a plus-minus pair.
But a square cannot be negative!

Not if we use real numbers. How about complex numbers?
Selection of the evaluation points

Note:
\[
\sqrt{i} = \pm \frac{1}{\sqrt{2}} (1 + i)
\]
\[
\sqrt{-i} = \pm \frac{1}{\sqrt{2}} (1 - i)
\]
Selection of the evaluation points

- $\sqrt{-1}$
- $-\sqrt{-1}$
- $+\sqrt{i}$
- $-\sqrt{i}$
- $+\sqrt{-i}$
- $-i$
- $+i$
- $-1$
- $+1$
Complex numbers: review

\[ z = a + bi \]

\[ z = r(\cos \theta + i \sin \theta) = re^{i\theta} \]

Polar coordinates: \((r, \theta)\)

Length: \(r = \sqrt{a^2 + b^2}\)

Angle \(\theta \in [0, 2\pi)\): \(\cos \theta = \frac{a}{r}, \sin \theta = \frac{b}{r}\)

\(\theta\) can always be reduced modulo \(2\pi\)

Some examples:

<table>
<thead>
<tr>
<th>Number</th>
<th>(-1)</th>
<th>(i)</th>
<th>(5 + 5i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polar coords</td>
<td>((1, \pi))</td>
<td>((1, \pi/2))</td>
<td>((5\sqrt{2}, \pi/4))</td>
</tr>
</tbody>
</table>
Complex numbers: multiplication

For any \( z = (r, \theta) \):

\[-z = (r, \theta + \pi), \text{ since } -1 = (1, \pi)\]

If \( z \) is on the unit circle, then \( z^n = (1, n\theta) \)
Complex numbers: the $n$th roots of unity

Solutions to the equation $z^n = 1$

($(n = 16)$

$\frac{4\pi}{n}$

Angle $2\pi/n$

$\frac{2\pi}{n} + \pi$

Solutions are $z = (1, \theta)$, for $\theta$ a multiple of $2\pi/n$

All roots are plus-minus paired:

$-(1, \theta) = (1, \theta + \pi)$
Roots of unity for $n = 8$
Recursive divide-and-conquer

\[ x_0^2 + x_1^2 - x_0 - x_1 + x_2 - x_2 + x_3 - x_3 \]

\[ x_0^4 + 1 \]

\[ x_2^4 - 1 \]
Divide-and-conquer step

Evaluate $A(x)$ at $n$th roots of unity

Evaluate $A_e(x^2)$ and $A_o(x^2)$ at $(n/2)$nd roots of unity
Divide-and-conquer steps
Divide-and-conquer steps

\[ A(x) \]

\[ A_e(x) \]

\[ A_0(x) \]
function FFT(A, ω)

Inputs: \( A = (a_0, a_1, \ldots, a_{n-1}) \), for \( n \) a power of 2
\( \omega \): A primitive \( n \)th root of unity

Output: \((A(1), A(\omega), A(\omega^2), \ldots, A(\omega^{n-1}))\)

if \( \omega = 1 \): return \( A \) // Only 1 coef. (constant)

\((A_e(\omega^0), A_e(\omega^2), \ldots, A_e(\omega^{n-2}))\) = FFT\((A_e, \omega^2)\)
\((A_o(\omega^0), A_o(\omega^2), \ldots, A_o(\omega^{n-2}))\) = FFT\((A_o, \omega^2)\)

for \( k = 0 \) to \( n - 1 \):
\( A(\omega^k) = A_e(\omega^{2k}) + \omega^k A_o(\omega^{2k}) \)

return \((A(1), A(\omega), A(\omega^2), \ldots, A(\omega^{n-1}))\)
FFT algorithm

\[
\text{for } k = 0 \text{ to } n - 1: \quad A(\omega^k) = A_e(\omega^{2k}) + \omega^k A_o(\omega^{2k})
\]

Example \((n = 8)\):

\[
\begin{align*}
A(\omega^0) &= A_e(\omega^0) + \omega^0 A_o(\omega^0) \\
A(\omega^1) &= A_e(\omega^2) + \omega^1 A_o(\omega^2) \\
A(\omega^2) &= \boxed{A_e(\omega^4)} + \omega^2 A_o(\omega^4) \\
A(\omega^3) &= A_e(\omega^6) + \omega^3 A_o(\omega^6) \\
A(\omega^4) &= A_e(\omega^8) + \omega^4 A_o(\omega^8) \\
A(\omega^5) &= A_e(\omega^{10}) + \omega^5 A_o(\omega^{10}) \\
A(\omega^6) &= \boxed{A_e(\omega^{12})} + \omega^6 A_o(\omega^{12}) \\
A(\omega^7) &= A_e(\omega^{14}) + \omega^7 A_o(\omega^{14})
\end{align*}
\]

\(\omega^4 = \omega^{12}\) \hspace{1cm} \(\omega^2 = -\omega^6\)
FFT algorithm

for \( k = 0 \) to \( n - 1 \): 
\[ A(\omega^k) = A_e(\omega^{2k}) + \omega^k A_o(\omega^{2k}) \]

Example \((n = 8)\):

\[
\begin{align*}
A(\omega^0) &= A_e(\omega^0) + \omega^0 A_o(\omega^0) \\
A(\omega^1) &= A_e(\omega^2) + \omega^1 A_o(\omega^2) \\
A(\omega^2) &= A_e(\omega^4) + \omega^2 A_o(\omega^4) \\
A(\omega^3) &= A_e(\omega^6) + \omega^3 A_o(\omega^6) \\
A(\omega^4) &= A_e(\omega^8) + \omega^4 A_o(\omega^8) \\
A(\omega^5) &= A_e(\omega^{10}) + \omega^5 A_o(\omega^{10}) \\
A(\omega^6) &= A_e(\omega^{12}) + \omega^6 A_o(\omega^{12}) \\
A(\omega^7) &= A_e(\omega^{14}) + \omega^7 A_o(\omega^{14}) \\
\end{align*}
\]

\( \omega^4 = \omega^{12} \) \quad \omega^2 = -\omega^6
FFT algorithm

for \( k = 0 \) to \( n - 1 \): \( A(\omega^k) = A_e(\omega^{2k}) + \omega^k A_o(\omega^{2k}) \)

Example \((n = 8)\):

\[
\begin{align*}
A(\omega^0) &= A_e(\omega^0) + \omega^0 A_o(\omega^0) \\
A(\omega^4) &= A_e(\omega^0) - \omega^0 A_o(\omega^0) \\
A(\omega^1) &= A_e(\omega^2) + \omega^1 A_o(\omega^2) \\
A(\omega^5) &= A_e(\omega^2) - \omega^1 A_o(\omega^2) \\
A(\omega^2) &= A_e(\omega^4) + \omega^2 A_o(\omega^4) \\
A(\omega^6) &= A_e(\omega^4) - \omega^2 A_o(\omega^4) \\
A(\omega^3) &= A_e(\omega^6) + \omega^3 A_o(\omega^6) \\
A(\omega^7) &= A_e(\omega^6) - \omega^3 A_o(\omega^6)
\end{align*}
\]
FFT algorithm

for $k = 0 \text{ to } n - 1$: \[ A(\omega^k) = A_e(\omega^{2k}) + \omega^k A_o(\omega^{2k}) \]

Example ($n = 8$):

\[
\begin{align*}
A(\omega^0) &= A_e(\omega^0) + \omega^0 A_o(\omega^0) \\
A(\omega^4) &= A_e(\omega^0) - \omega^0 A_o(\omega^0) \\
A(\omega^1) &= A_e(\omega^2) + \omega^1 A_o(\omega^2) \\
A(\omega^5) &= A_e(\omega^2) - \omega^1 A_o(\omega^2) \\
A(\omega^2) &= A_e(\omega^4) + \omega^2 A_o(\omega^4) \\
A(\omega^6) &= A_e(\omega^4) - \omega^2 A_o(\omega^4) \\
A(\omega^3) &= A_e(\omega^6) + \omega^3 A_o(\omega^6) \\
A(\omega^7) &= A_e(\omega^6) - \omega^3 A_o(\omega^6)
\end{align*}
\]

for $k = 0 \text{ to } n/2 - 1$:

\[
\begin{align*}
A(\omega^k) &= A_e(\omega^{2k}) + \omega^k A_o(\omega^{2k}) \\
A(\omega^{k+n/2}) &= A_e(\omega^{2k}) - \omega^k A_o(\omega^{2k})
\end{align*}
\]
FFT algorithm

function FFT($a, \omega$)

Inputs: $a = (a_0, a_1, ..., a_{n-1})$, for $n$ a power of 2 
$\omega$: A primitive $n$th root of unity 
Output: $(a(1), a(\omega), a(\omega^2), ..., a(\omega^{n-1}))$

if $\omega = 1$: return $a$ // $n = 1$, $a$ has only one element

($s_0, s_1, ..., s_{n/2-1}$) = FFT(((a_0, a_2, ..., a_{n-2}), \omega^2)
($s'_0, s'_1, ..., s'_{n/2-1}$) = FFT((a_1, a_3, ..., a_{n-1}), \omega^2)

for $k = 0$ to $n/2 - 1$: // FFT shuffling
    $r_k = s_k + \omega^k s'_k$
    $r_{k+n/2} = s_k - \omega^k s'_k$

return ($r_0, r_1, ..., r_{n-1}$)
• The runtime of the FFT can be expressed as:

\[ T(n) = 2 \cdot T(n/2) + O(n) \]

• Using the *Master Theorem* we conclude:

Runtime FFT(n) = O(n \log n)
Unfolding the FFT

\[ \begin{align*}
F \tau_T n/2 = a_0 + \sum_{k=1}^{n/2} a_k \omega^k + \sum_{k=1}^{n/2} \bar{a}_k \bar{\omega}^k r_k \\
&= a_0 + \sum_{k=1}^{n/2} a_k \omega^k + \sum_{k=1}^{n/2} \bar{a}_k \bar{\omega}^k r_k + n/2
\end{align*} \]
Unfolding the FFT (butterfly diagram)
Why is it called a butterfly diagram?
Conversion between both representations

<table>
<thead>
<tr>
<th>representation</th>
<th>addition</th>
<th>multiplication</th>
<th>evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>coefficient</td>
<td>$O(n)$</td>
<td>$O(n^2)$</td>
<td>$O(n)$</td>
</tr>
<tr>
<td>point-value</td>
<td>$O(n)$</td>
<td>$O(n)$</td>
<td>$O(n^2)$</td>
</tr>
</tbody>
</table>

\[
\langle \text{values} \rangle = \text{FFT}(\langle \text{coefficients} \rangle, \omega)
\]

$\{a_0, a_1, \cdots, a_{n-1}\}$

Coefficient representation

$O(n \log n)$

evaluation

$\{x_0, y_0\}, \cdots, (x_{n-1}, y_{n-1}\}$

Point-value representation

Interpolation

$\uparrow$

?
From point-values to coefficients

The Fast Fourier Transform computes:

\[
\begin{bmatrix}
  y_0 \\
y_1 \\
y_2 \\
\vdots \\
y_{n-1}
\end{bmatrix}
= \begin{bmatrix}
  1 & 1 & 1 & 1 & \cdots & 1 \\
  1 & \omega & \omega^2 & \omega^3 & \cdots & \omega^{n-1} \\
  1 & \omega^2 & \omega^4 & \omega^6 & \cdots & \omega^{2(n-1)} \\
  \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
  1 & \omega^{n-1} & \omega^{2(n-1)} & \omega^{3(n-1)} & \cdots & \omega^{(n-1)^2}
\end{bmatrix}
\begin{bmatrix}
a_0 \\
a_1 \\
a_2 \\
\vdots \\
a_{n-1}
\end{bmatrix}
\]

where \( \omega = e^{2\pi i/n} = (1, 2\pi/n) \).

Let us call \( F_n(\omega) \) the Fourier matrix. Thus,

\[ y = F_n(\omega) \cdot a \]

How about if we know \( y \) and we want to obtain \( a \)?
From point-values to coefficients

\[ y = F_n(\omega) \cdot a \]
\[ \Downarrow \]
\[ [F_n(\omega)]^{-1} \cdot y = a \]

\( F_n(\omega) \) is a unitary matrix and has the following property:

\[ [F_n(\omega)]^{-1} = \frac{1}{n} \cdot F_n(\omega^{-1}) \]

and also

If \( \omega \) is a primitive \( n \)th root of unit,
then \( 1/\omega \) is also a primitive \( n \)th root of unit.

Note: The inverse of unitary matrix is its conjugate transpose.
Conversion between both representations

<table>
<thead>
<tr>
<th>representation</th>
<th>addition</th>
<th>multiplication</th>
<th>evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>coefficient</td>
<td>$O(n)$</td>
<td>$O(n^2)$</td>
<td>$O(n)$</td>
</tr>
<tr>
<td>point-value</td>
<td>$O(n)$</td>
<td>$O(n)$</td>
<td>$O(n^2)$</td>
</tr>
</tbody>
</table>

$\langle \text{values} \rangle = \text{FFT}(\langle \text{coefficients} \rangle, \omega)$

$\langle \text{coefficients} \rangle = \frac{1}{n} \text{FFT}(\langle \text{values} \rangle, \omega^{-1})$
Polynomial multiplication

**Input:** Coefficients of two polynomials $A(x)$ and $B(x)$, of degree $d_A$ and $d_B$, respectively. Let $d = d_A + d_B$.

**Output:** The product $C = A \cdot B$.

1. **Selection:**
   - Pick $\omega = (1, 2\pi/n)$, such that $n \geq d + 1$ and $n$ is a power of two.

2. **Evaluation (FFT):**
   - Compute $A(1), A(\omega), A(\omega^2), \ldots, A(\omega^{n-1})$.
   - Compute $B(1), B(\omega), B(\omega^2), \ldots, B(\omega^{n-1})$.

3. **Multiplication:**
   - Compute $C(\omega^k) = A(\omega^k) \cdot B(\omega^k)$, for all $k = 0, \ldots, n - 1$.

4. **Interpolation (inverse FFT):**
   - Recover $C(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_d x^d$. 
Example: from values to coefficients

Let us consider a polynomial:

\[ P(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 \]

We have \( n = 4 \) and \( \omega = i \). Let us assume that the evaluation at four different points is:

\[
\begin{align*}
P(1) &= 2 \\
P(i) &= 1 - i \\
P(-1) &= 4 \\
P(-i) &= 1 + i
\end{align*}
\]

We want to calculate the coefficients \((a_0, a_1, a_2, a_3)\) using the inverse FFT, i.e.,

\[
[a_0, a_1, a_2, a_3] = \frac{1}{4} \text{FFT}([2, 1 - i, 4, 1 + i], \omega^{-1})
\]
Example: from values to coefficients

\[
\text{FFT}([2,1 - i, 4,1 + i], \omega = -i)
\]

\[
k = 0 \quad \begin{cases} 
  r_0 = 6 + 2 = 8 \\
  r_2 = 6 - 2 = 4 \\
  r_1 = -2 + (-i)(-2i) = -4 \\
  r_3 = -2 - (-i)(-2i) = 0
\end{cases}
\]

\[
k = 1
\]

\[
\text{FFT}([2,4], \omega = -1)
\]

\[
\begin{array}{l}
  r_0 = 2 + 4 = 6 \\
  r_1 = 2 - 4 = -2
\end{array}
\]

\[
\text{FFT}([2], \omega = 1) \quad \text{FFT}([4], \omega = 1)
\]

\[
2 \quad 4
\]

\[
\text{FFT}([1 - i], \omega = 1) \quad \text{FFT}([1 + i], \omega = 1)
\]

\[
1 - i \quad 1 + i
\]

\[
[a_0, a_1, a_2, a_3] = \frac{1}{4} [8, -4,4,0]
\]

\[
P(x) = x^2 - x + 2
\]
Example: from coefficients to values

\[ P(x) = x^2 - x + 2 \]

\[
\text{FFT}([2, -1, 1], \omega = i)
\]

\[
\begin{align*}
k = 0 & : \begin{cases} 
    r_0 = 3 - 1 = 2 \\
    r_2 = 3 + 1 = 4
\end{cases} \\
n = 1 & : \begin{cases} 
    r_1 = 1 + i \cdot (-1) = 1 - i \\
    r_3 = 1 - i \cdot (-1) = 1 + i
\end{cases}
\]

\[
\begin{align*}
\text{FFT}([2,1], \omega = -1) & : \\
r_0 = 2 + 1 = 3 & : \\
r_1 = 2 - 1 = 1 & :
\end{align*}
\]

\[
\begin{align*}
\text{FFT}([2], \omega = 1) & : 2 \\
\text{FFT}([1], \omega = 1) & : 1
\end{align*}
\]

\[
\begin{align*}
\text{FFT}([-1,0], \omega = -1) & : \\
r_0 = -1 + 0 = -1 & : \\
r_1 = -1 - 0 = -1 & :
\end{align*}
\]

\[
\begin{align*}
\text{FFT}([-1], \omega = 1) & : -1 \\
\text{FFT}([0], \omega = 1) & : 0
\end{align*}
\]

\[
[y_0, y_1, y_2, y_3] = \text{FFT}([a_0, a_1, a_2, a_3], \omega = i)
\]

\[
[y_0, y_1, y_2, y_3] = [2, 1 - i, 4, 1 + i]
\]
Conclusions

• Gilbert Strang (MIT, 1994): “the most important numerical algorithm of our lifetime”.

EXERCISES
Consider the polynomials \( 1 + x - 2x^2 + x^3 \) and \(-1 + x^2\):

– Choose an appropriate power of two to execute the FFT for the polynomial multiplication. Find the value of \(\omega\).

– Give the result of the FFT for \(x^2 - 1\) using the value of \(\omega\) required for the multiplication (no need to execute the FFT).
Consider the FFT of the polynomial $x^2 + 2x + 1$:

– Find the value of $\omega$ to execute the FFT.
– In which points the polynomial must be evaluated?
– Execute the FFT and give the point-value representation of the polynomial.
Multiplication using FFT

Consider the polynomials $-1 + 2x + x^2$ and $1 + 2x$:

- Choose an appropriate power of two to execute the FFT. Find the value of $\omega$.
- Calculate their point-value representation using the FFT (execute the FFT algorithm manually).
- Calculate the product of the point-value representations.
- Execute the inverse FFT to obtain the coefficients of the product.