B16 Structured Programming Sheet 1

Instructor Andrea Vedaldi (andrea.vedaldi@eng.ox.ac.uk)
Web Page http://www.robots.ox.ac.uk/~vedaldi/teach.html
Revision HT 2019 (typeset on August 16, 2018)

Instructions: In a number of the questions you are asked to write actual C or MATLAB code. Try to compile and execute the code for each exercise on your computer (see the second problem to get started). If you cannot get the code to work, make sure to get to the end of the exercise using pen and paper. Note that using the online (PDF) version of this problem sheet could be more convenient for cutting and pasting.

Problem 1 Software lifecycle

Write notes contrasting the waterfall model of the software development lifecycle with extreme programming.

Solution to problem 1

Bookwork -- really, just for interest's sake (and to understand the importance of safety-critical software).

Problem 2 Hello world!

The goal of this exercise is to edit, compile, and execute a simple C program. You will use the same procedure later for more complex exercises as well. There are a few options:

- **Use an on-line C compiler.** On-line compilers are suitable for trying out simple programs such as the exercises in this tutorial. However, they are *not* very representative of a real-world programming environment. There are several such on-line environments (if you are given the option, make sure that you select C or C/C++ for this tutorial). We recommend using *C++ Shell*:

  - [http://cpp.sh](http://cpp.sh)

  Note also that such tools *many not save your code*. Therefore, make sure to save a copy on your computer (e.g. by cutting and pasting from a text editor to your browser and viceversa).

- **Install a C compiler on your machine.** This is more complicated, but much closer to a real work environment. Depending on your platform, you can get started here:

  - **Windows.** Install a copy of the Microsoft Visual C integrated development environment (IDE). As Oxford students, you have the right to install a copy of this software for free.
  
  - **Mac OS X.** Install *Xcode* from the Mac App Store for free. Xcode is Apple's IDE, similar to Visual C.
- **Linux.** Install **GCC** using your Linux distribution package manager. GCC is not an IDE; in this case you will need also a text editor and to use the terminal to compile and run your program (do not worry, it is easy!).

Appendix B contains some instructions on how to get started with these environment; however, for any technical question *Google & Stack Overflow are your friends!*.

Once you have decided which editor to use, cut and paste the following program

```c
#include <stdio.h>

int main(int argc, const char * argv[]) {
    printf("Hello, World!\n");
    return 0;
}
```

Make sure you can compile and execute the program, getting the correct output (which one?).

**Solution to problem 2**

The program should print the string **Hello, World!** in the console/terminal.

**Problem 3  Function encapsulation and side-effects**

1. Explain what are function **side-effects** and why should they be avoided.

2. Explain what is meant by **encapsulation** in software design and why it is important.

3. Determine what the code below does, indicate its design flaws, and improve its design and presentation.

```c
#include <stdio.h>
int num,res; void r(int s) {
   while (num>=0) { num = num-s; } num=num+s; res=num;}
int main() { num = 10;
r(4); printf ("%d\n", res ); }
```

4. Compile and run the code above and your improved version. Make sure that they produce the same result.

**Solution to problem 3**

1. Function side-effects are changes to the program state made in the execution of a function that are not explicit in the function interface. Examples include reading a global variable, since the behaviour of the function may vary depending on the value of this variable. Likewise, the setting of a global variable, or other hidden'' state changes are side-effects. They should be avoided because they violate the encapsulation of the function as providing black-box functionality, making the function harder to understand and difficult to use.
2. Encapsulation is the process of separating explicitly the interface of a software component from its implementation. The interface defines the ways in which the software component can interact with other components. If the interface is clearly specified, then the component can be re-used with confidence, even if its internal implementation details change. Another benefit of good encapsulation is to make the behaviour of the function explicit, reducing the risk of unexpected side-effects.

In the context of procedural programming, the software components of primary interest are functions/procedures. The operation and interface of a well designed function are transparent from the function header (prototype) which gives the set of inputs (parameters) and outputs (function return values or in the case of C++, perhaps non-constant reference parameters). This encapsulates the function's meaning, without revealing its internal implementation details. The C and MATLAB standard libraries are good examples of this level of encapsulation. A programmer can use these libraries without needing to know the internal implementation details, and (unless otherwise documented and in limited and well defined cases) can be confident that the operation is neither dependent on, nor affects the values of, global variables.

3. First, reformattting the code should make it much easier to find out that the program defines a function \( r() \) which computes the remainder in the integer division \( \text{num}/\text{s} \) using repeated subtraction.

```c
#include <stdio.h>

int num, res;

void r(int s)
{
    while (num >= 0) num = num - s;
    num = num + s;
    res = num;
}

int main()
{
    num = 10;
    r(4);
    printf("%d\n", res);
}
```

The variable \( \text{num} \) is an implicit input parameter to the function \( r() \) and \( \text{res} \) is being used as an implicit output. In this simple example such side-effects can be readily identified, because the implementation of \( r() \) is so tiny. However in more complex code this becomes a real problem. To remove the side-effects we make these relationships explicit in the definition of \( r() \), so that \( \text{num} \) is passed in (by value) to the function, and the result passed out as a return value. When we do this, the local variable \( g \) becomes redundant because \( \text{num} \) is passed in by value.

```c
#include <stdio.h>
```
```c
int r(int num, int s)
{
    while (num >= 0) num = num - s;
    num = num + s;
    return num;
}

int main()
{
    printf("%d\n", r(10, 4));
}
```

We could argue that the code is better but still unsatisfactory. What happens if \( s \) is not a positive number?

Problem 4  Parameters, recursion and the stack

The Chebyshev polynomials of the first kind \( T_n(x) \) are defined recursively by

\[
\begin{align*}
T_0(x) &= 1, \\
T_1(x) &= x, \\
T_{n+1}(x) &= 2xT_n(x) - T_{n-1}(x).
\end{align*}
\]

The following MATLAB code evaluates \( T_n(x) \) at point \( x \) for a given polynomial order \( n \):

```matlab
function ChebyshevExample()
% plot the Chebyshev polynomial of order n in the interval -10,10
n = 4;
t = [-10:0.01:10];
plot(t,chebyshev(n,t));
y = chebyshev(3, [-2:1:2]);

% Chebyshev polynomials evaluation
function val = chebyshev(n,x)
if n == 0
    val = ones(size(x)) ;
else
    if n == 1
        val = x ;
    else
        val = 2 .* x .* chebyshev(n-1,x) - chebyshev(n-2,x) ;
    end
end
end
```

1. With reference to the code, define and identify the formal and actual parameters of the function \texttt{chebyshev} and the function call.

2. For the second function call, show the evolution of the stack and the final value of the vector returned.
Solution to problem 4

1. In the particular case here, the formal parameters are \(n\) and \(x\) (and the return parameter is \(\text{val}\)), appearing in the function header `function \(\text{val} = \text{chebyshev}(n, x)\)` in the script. The actual parameters appear in the two calls to `chebyshev()` in the script. In the first call the params are \(a = 4\) (mapping to formal parameter \(n\)) and \(t = [-10:0.01:10]\) (mapping to formal parameter \(x\)) while in the second they are the value 3 (mapping to \(n\)) and \(t = [-2:1:2]\), mapping to formal parameter \(x\).

2. The first call `chebyshev(3, [-2:1:2])` pushes the following onto the stack:

\[
\begin{array}{c}
\text{% chebyshev(3, [-2:1:2])} \\
3 \\
[-2 -1 0 1 2]
\end{array}
\]

and calls \(2 \times x \times \text{chebyshev}(2, [-2:1:2]) + \text{chebyshev}(1, [-2:1:2])\). The first of these calls pushes to give a stack:

\[
\begin{array}{c}
\text{% chebyshev(3, [-2:1:2])} \\
3 \\
[-2 -1 0 1 2] \\
\text{% chebyshev(2, [-2:1:2])} \\
2 \\
[-2 -1 0 1 2]
\end{array}
\]

and runs `chebyshev()`), yielding two more calls: \(2 \times x \times \text{chebyshev}(1, [-2:1:2]) + \text{chebyshev}(0, [-2:1:2])\). The first of these pushes so the stack is now:

\[
\begin{array}{c}
\text{% chebyshev(3, [-2:1:2])} \\
3 \\
[-2 -1 0 1 2] \\
\text{% chebyshev(2, [-2:1:2])} \\
2 \\
[-2 -1 0 1 2] \\
\text{% chebyshev(1, [-2:1:2])} \\
1 \\
[-2 -1 0 1 2]
\end{array}
\]

and this call returns \([-2 -1 0 1 2]\) because \(n==1\), popping the bottom two parameters in the process. Hence the stack is now:

\[
\begin{array}{c}
\text{% chebyshev(3, [-2:1:2])} \\
3 \\
[-2 -1 0 1 2] \\
\text{% chebyshev(2, [-2:1:2])} \\
2 \\
[-2 -1 0 1 2]
\end{array}
\]

The second call `chebyshev(0, [-2:1:2])` pushes so the stack is:

\[
\begin{array}{c}
\text{% chebyshev(3, [-2:1:2])} \\
3
\end{array}
\]
and returns \([1 \ 1 \ 1 \ 1 \ 1]\) because \(n=0\) (popping the two bottom parameters in the process). \(\text{chebyshev}(3,[-2:1:2])\) can now be evaluated as \(2.*[-2 \ -1 \ 0 \ 1 \ 2].*[-2 \ -1 \ 0 \ 1 \ 2] - [1 \ 1 \ 1 \ 1 \ 1] = [7 \ 1 \ -1 \ 1 \ 7]\) and its parameters popped from the stack, so the stack is now back to

\[
\begin{array}{l}
\% \text{chebyshev}(3,[-2:1:2])
\end{array}
\]

\[
3
\]

\[
[-2 \ -1 \ 0 \ 1 \ 2]
\]

We now have \(2.*x.*[7 \ 1 \ -1 \ 1 \ 7] - \text{chebyshev}(1,[-2:1:2])\). The second call pushes so the stack is

\[
\begin{array}{l}
\% \text{chebyshev}(3,[-2:1:2])
\end{array}
\]

\[
3
\]

\[
[-2 \ -1 \ 0 \ 1 \ 2]
\]

\[
\% \text{chebyshev}(1,[-2:1:2])
\]

\[
1
\]

\[
[-2 \ -1 \ 0 \ 1 \ 2]
\]

and returns \([-2 \ -1 \ 0 \ 1 \ 2]\) since \(n=1\), popping the parameters as it returns.

Finally we can evaluate \(2.*x.*[7 \ 1 \ -1 \ 1 \ 7] - [-2 \ -1 \ 0 \ 1 \ 2] = [-26 \ -1 \ 0 \ 1 \ 26]\) and pop the initial parameters from the stack.

**Problem 5  Parameters and recursion**

1. Write a recursive MATLAB function to compute the \(n^{th}\) Fibonacci number, where the first two Fibonacci numbers are defined to be 1 and 1, and each subsequent number is the sum of the previous two.

2. Write a non-recursive function to do the same.

3. If your answer to part (i) involves binary recursion (i.e. the function calls itself twice), consider the cost of this, and devise a recursive algorithm that uses linear recursion instead (i.e. the function call itself only once).

4. Compare the amount of work required by binary recursion and linear recursion by sketching a trace of the function calls. Is binary recursion doing redundant work?

**Solution to problem 5**
1. The following function computes the Fibonacci numbers using binary recursion:

```plaintext
function res = fib2(n)
if n==1 || n==2
    res = 1;
else
    res = fib2(n-1) + fib2(n-2);
end
```

2. The following function computes the Fibonacci numbers using a simple iteration:

```plaintext
function res = fib_nonrecursive(n)
if n==1 || n==2
    res = 1;
else
    a = 1;
    b = 1;
    for i=3:n
        temp = b;
        b = a+b;
        a = temp;
    end
    res = b;
end
```

3. The following function computes the Fibonacci number using linear recursion:

```plaintext
function [res,prev] = fib1(n)
if n==1
    res = 1;
    prev = 0;
elseif n==2
    res = 1;
    prev = 1;
else
    [f1,f2] = fib1(n-1);
    res = f1+f2;
    prev = f1;
end
```

4. The difference between the two recursive implementations can be appreciated by inspecting the call trace:
Note that binary recursion computes the same values over and over again, with exponential complexity (this could be avoided by memoisation).

Problem 6  Sorting, loop invariants and passing functions as parameters

The following pseudo-code defines an algorithm to sort an array A of numbers into ascending order:

```plaintext
% ROUTINE: BubbleSort
% INPUT: A is array of N elements
% OUTPUT: A is sorted array of elements
for i=1 to N-1
    for j=i+1 to N
    end
end
return A
```

1. Determine the order $O(f(N))$ of the algorithm.
2. Determine a loop invariant for each loop and hence satisfy yourself that the function will have the desired effect.
3. Explain how the sorting function can be made generic by replacing the comparison $A[i] > A[j]$ with a call to a function GreaterThan, which takes two items and returns true if $x$ is less than $y$, and false otherwise, and where GreaterThan is a formal parameter to BubbleSort. Write down the likely function prototype for BubbleSort.
4. Write such a comparison function that compares the lengths of two vectors under the 2-norm and show how it can be used with the generic sorting function to sort the rows of a matrix.
5. Write two more comparison functions that use (i) the 1-norm and (ii) the infinity norm.
6. Comment on the advantages of this programming construct in this context.
Solution to problem 6

1. There are two nested loops. The outer loop iterates over \( i \) for values \( 1, \ldots, N - 1 \), and, for each value of \( i \), the inner loop iterates \( N - i \) times. The inner loop runs therefore

\[
\sum_{i=1}^{N-1} (N - i) = N(N - 1) - \frac{N(N - 1)}{2} = \frac{N(N - 1)}{2}
\]

times. Hence the cost of the algorithm is \( O(N^2) \).

2. Intuitively, the algorithm starts with the array \( A \) and an empty array \( B \). It then gradually finds and removes the smallest element from \( A \) and appends it to \( B \). Rather than working with two arrays explicitly, however, the code divides \( A \) in two parts and use swaps pairs of elements to move the data between them.

The intuition above can be formalised by means of loop invariants, which can be proved by mathematical induction by inspection of the program. There are two invariants \( P_1(i) \) and \( P_2(i, j) \), one for each loop:

```plaintext
% P1(0) is true here
for i=1 to N-1
  % P1(i-1) is true here
  % P2(i,i) is true here
  for j=i+1 to N
    % P2(i,j-1) is true here
    % P2(i,j) is true here
  end
% P1(i) is true here
end
return A
```

The invariants are as follows:

(a) The outer loop invariant \( P_1(i) \) is a proposition which states that the first \( i \) elements \( A[1], \ldots, A[i] \) of the array \( A \) are sorted. It also says that \( A[i] \) not larger than the tail \( A[i + 1], \ldots, A[N] \) of the array. As special cases, \( P_1(0) \) (valid at the beginning of the algorithm) says nothing, and \( P_1(N) \) (valid at the end of the algorithm) says that the whole array is sorted. Formally, this can now be written as:

\[
\begin{align*}
  &\text{for } 1 \leq i \leq N - 1 \quad P_1(i) : \quad A[1] \leq \cdots \leq A[i] \quad \text{and} \quad A[i] \leq A[i + 1], \ldots, A[N]; \\
  &\text{for } i = N \quad P_1(N) : \quad A[1] \leq \cdots \leq A[N]; \\
  &\text{for } i = 0 \quad P_1(0) : \quad \text{empty statement (always true).}
\end{align*}
\]

(b) The inner loop invariant \( P_2(i, j) \) is a proposition which states that \( A[i] \) is the smallest of the subarray \( A[i], \ldots, A[j] \); it also says that the outer invariant \( P_1(i - 1) \) is maintained (i.e. it remains true). Formally:

\[
\begin{align*}
  &\text{for } 1 \leq i \leq N - 1, \ i \leq j \leq N \quad P_2(i, j) : \quad P_1(i - 1) \quad \text{and} \quad A[i] \leq A[i + 1], \ldots, A[j].
\end{align*}
\]
Inner loop correctness. We now prove that the inner loop is correct by showing that, assuming that $P_2(i, i)$ is true at line 3, then $P_2(i, j)$ must be true at line 8 for each iteration $j = i + 1, i + 2, \ldots, N$ of the inner loop. This is proved by induction:

- At the first iteration of the inner loop, $j = i+1$, the proposition $P_2(i, j-1) = P_2(i, i)$ is true at line 6 because it is assumed to be true at line 3 and the code has yet to change the array. At iterations $j = i + 2, \ldots$, the proposition $P_2(i, j-1)$ at line 6 is equivalent to the proposition $P_2(i, j)$ at line 8 at the end of the previous iteration of the loop. Thus it is true if the latter is, which we show next.

- Now we show that at any iteration of the inner loop, if $P_2(i, j-1)$ is true the beginning of the loop body (line 6), then $P_2(i, j)$ is true at the end of the loop body (line 8). Formally, we need to show that:

$$P_2(i, j-1): P_1(i-1) \text{ and } [A[i] \leq A[i+1], \ldots, A[j-1]] \quad \text{(at line 6)}$$

implies that

$$P_2(i, j): P_1(i-1) \text{ and } [A[i] \leq A[i+1], \ldots, A[j-1], A[j]] \quad \text{(at line 8)}$$

This fact can be proved by studying the effect of executing line 7. First, notice that the line 7 potentially swaps $A[i]$ and $A[j]$ and examine the effect of such a swap on the first part of the proposition $P_2(i, j)$, namely $P_1(i-1)$. Recall that this proposition is:

$$P_1(i-1): A[1] \leq \cdots \leq A[i-1] \text{ and } A[i-1] \leq A[i], \ldots, A[j], \ldots, A[N]$$

Suppose the values are swapped, and denote with $\tilde{A}[i] \leftarrow A[j]$ and $\tilde{A}[j] \leftarrow A[i]$ the updated array entries. After the swapped, the proposition thus becomes:

$$P_1(i-1): A[1] \leq \cdots \leq A[i-1] \text{ and } A[i-1] \leq \tilde{A}[i], \ldots, \tilde{A}[j], \ldots, A[N]$$

Is the proposition still true? Yes, because both $A[i]$ and $A[j]$, and hence $\tilde{A}[j]$ and $\tilde{A}[i]$, are not smaller than $A[i-1]$ to start with.

Let us now look at the second part of the proposition $P_2(i, j)$. For $P_2(i, j)$ to be true after line 7 is executed, we must show that:

$$A[i] \leq A[i+1], \ldots, A[j-1], A[j]. \quad (1)$$

There are two possibilities. The first one is that $A[i] \leq A[j]$, in this case the if condition at line 7 is false and no exchange takes place. Since we already know from $P_2(i, j-1)$ that $A[i] \leq A[i+1], \ldots, A[j-1]$, the fact that $A[i] \leq A[j]$ implies that eq. (1) is true at line 8.

The second possibility is that $A[i] > A[j]$. In this case the if condition is true and values $\tilde{A}[i] \leftarrow A[j]$ and $\tilde{A}[j] \leftarrow A[i]$ are exchanged, so that the second part of $P_2(i, j)$ can be written as:

$$\tilde{A}[i] \leq A[i+1], \ldots, A[j-1], \tilde{A}[j]. \quad (2)$$

Clearly the swap restores the last inequality $\tilde{A}[i] \leq \tilde{A}[j]$, but are all other inequalities $\tilde{A}[i] \leq A[i+1], \ldots, A[j-1]$ still satisfied?

The answer is yes (and this is really the key to the algorithm) because the value $A[j]$ that is now in front of the sequence is smaller than $A[i]$ (because we just found that $A[j] < A[i]$), which in turn was known to be smaller than the other elements from the previous iteration of the loop (i.e., $A[i] \leq A[i+1], \ldots, A[j-1]$). Hence:

**Outer loop correctness.** Now that we have proved the correctness of the inner loop, we can prove the correctness of the outer loop in a similar manner:

- At the first iteration $i = 1$, the proposition $P_1(i - 1) = P_1(0)$ at the beginning of the outer loop body (line 3) is true because it is true at line 1 and nothing changed in the array yet. At iterations $i = 2, 3, \ldots$, proposition $P_1(i - 1)$ is true provided that $P_1(i)$ is true at the end (line 10) of the previous iteration of the loop, which we show next.
- We must show that $P_1(i)$ is true at line 10 given that $P_1(i - 1)$ is true at line 3. In order to do so, we make use of the fact that the inner loop causes the predicate $P_2(i, N)$ to be true at line 8, as we proved above. In formulas, given that

\[
P_2(i, N) : \quad P_1(i - 1) \quad \text{and} \quad A[i] \leq A[i + 1], \ldots, A[N];
\]

\[
P_1(i - 1) : \quad A[1] \leq \cdots \leq A[i - 1] \quad \text{and} \quad A[i - 1] \leq A[i], \ldots, A[N]
\]

we need to show that

\[
P_1(i) : \quad A[1] \leq \cdots \leq A[i - 1] \leq A[i] \quad \text{and} \quad A[i] \leq A[i + 1], \ldots, A[N].
\]

The second part of proposition $P_1(i)$ is true due to the second part of proposition $P_2(i, N)$. The inequalities $A[1] \leq \cdots \leq A[i - 1]$ are true due to the first part of proposition $P_1(i - 1)$. The remaining inequality $A[i - 1] \leq A[i]$ is true due to the second part of $P_1(i - 1)$.

This conclude the proof of correctness.

3. The solution is to pass in a function parameter `GT` that compares two elements (`GT` is assumed to be a function handle). For example, in MATLAB this might look like:

```matlab
function A = BubbleSort(A,GT)
% The function sorts the columns of the arra A. N is the number
% of columns. As a special case, the function can sort a vector
% of number.

N = size(A,2) ;

for i=1:N-1
    for j=i+1:N
        if (feval(GT,A(:,i),A(:,j)))
            temp = A(:,j) ;
            A(:,j) = A(:,i) ;
            A(:,i) = temp ;
        end
    end
end
```

The following function demonstrates how to use `BubbleSort` to sort a list of scalars or a list of vectors. These are coded as columns of a matrix.

```matlab
function BubbleSortExample()

% Sort a row vector using the standard definiton
```
Problem 7  Data structures and interfaces

In this exercise we explore the design and implementation of a list data structure in the C language. The code comprises three files: list.h, containing the list declaration, list.c, containing the list implementation, and main.c, containing a program using the list. The first file, list.h, is as follows:

```c
#define __LIST_H__

typedef double DataType;

typedef struct ListElement {
    DataType x;
    ListElement *next;
} ListElement;

typedef ListElement *List;

typedef struct ListElement _ListElement {
    DataType x;
    ListElement *next;
} ListElement;

#define __LIST_H__
```

4. % of the greater-than operator
5. A = [1 16 20 10 -10];
6. B = BubbleSort(A, gt); disp('Before sorting:'); disp(A);
7. disp('After sorting:'); disp(B);
8. A = [1 2 3 5 0 ; 2 3 4 5 0 ; 5 6 7 8 0 ; 1 2 8 6 0];
9. B = BubbleSort(A, GreaterThan);
10. disp('Before sorting:'); disp(A);
11. disp('After sorting:'); disp(B);
12. function res = GreaterThan(x, y)
13.     res = norm(x) > norm(y);
14. ```
The second file, `main.c`, is as follows:

```c
#include "list.h"

int main(int argc, char **argv)
{
    List* list;
    ListElement *element;
    list = ListCreate();
    ListAppend(list, 1);
    ListAppend(list, 2);
    ListAppend(list, 3);
    ListAppend(list, 4);
    ListDelete(list);
    return 0;
}
```

The content of the third file, `main.c`, is given in Appendix A.

Note that `main.c` includes the `list.h` file. Neither `list.h` nor `main.c` define the structure `ListElement` (`list.h` only declares `ListElement` to be a structure, but does not provide any detail). In other words, from the viewpoint of `main.c`, both `ListElement` and `List` are opaque data types.

1. How can the program in `main.c` make use of opaque data types? Are opaque data types a good idea? Why?

2. Compile and execute the program. To do so, note that you need to combine two implementation files, `main.c` and `list.c`. This can be done by running the C compiler individually on `main.c` and `list.c` to produce the intermediate object files `main.o` and `list.o` and then by running the linker to combine the object files into a single executable file. In practice:

   - If you compile the code online using Coding Grounds, simply create the three files `list.h`, `list.c`, and `main.c` and compile and execute as usual.
   - If you use instead an IDE such as Visual C or Xcode, create the three files in the interface. Then, the IDE should be smart enough to correctly understand what needs to be done automatically.
   - If you use the command line, such as GCC in Linux, simply specify the two implementation files; the compiler is smart enough to do all the steps (including linking) automatically:

     ```bash
     > gcc -o list list.c main.c
     ```

3. Study carefully the implementation of the list (in `list.c`) then look at the `main()` function again. Which ones of the interface function allocate memory on the heap and for which purpose? Which ones free memory? Does the program leak any memory?
4. **ListCreate()** is used in the main() function to return a pointer to a new list. Intuitively, this corresponds to creating a new instance of a list object. Which part of the program logically owns this object instance? Should the object instance be explicitly disposed of? As an author of the main() function, is it your responsibility to dispose of the object instance? If so, how should this be done?

5. Which part of the program is logically responsible for handling the memory required to hold the list elements?

6. The header file list.h contains enough information for the compiler to understand how to use the list interface. However, a user might at best guess the meaning of the interface functions and how they should be used. A clear understanding can only be achieved by studying the implementation file, which defies the purpose of having an interface in the first place (i.e. encapsulation).

The correct solution is to document the interface properly. For example, the documentation for ListCreate() might read:

```c
/** ListCreate - Create a new list.
 ** return - pointer to the new list.
 **
 ** The function ListCreate() allocates and initializes a new empty
 ** list and returns a pointer to it. The ownership of the new object is
 ** transferred to the caller. The object must be disposed of by
 ** calling ListDelete().
 **
 ** Errors: if there is not sufficient memory to allocate the new list, the
 ** function returns a NULL pointer.
 ***/
```

The documentation of the ListAppend() function might read:

```c
/** ListAppend - Append a new element to a list.
 ** param "0060list"0060 - a pointer to the list.
 ** param "0060x"0060 - the value of the new list element.
 **
 ** The function ListAppend() creates a new element at the end
 ** of the list "0060list"0060 and initializes its value to "0060x"0060.
 ** Note that
 ** the list retains the ownership of the new element.
 **
 ** Errors: if there is not sufficient memory to allocate the
 ** new list element, an exception is raised terminating the program.
 ***/
```

Write a similar documentation for the ListInsertAfterElement() and ListFindElement() functions.

**Solution to problem 7**
1. The compiler only requires the function declaration to be able to generate code that calls them. Furthermore, all opaque data types are pointers such that the compiler does not need to know the detail of the pointed objects to manipulate them (for example to pass them as function arguments).

2. Practical work.

3. ListCreate() allocates memory to store the list object and ListAppend allocates memory to store a new list element. ListDelete() deletes the memory used to store the list object; furthermore, by inspecting the code in main.c, note that this function first calls ListClear() which removes all elements in the list, including freeing the corresponding memory. Hence, ListDelete() deletes all the memory associated to the list.

4. ListCreate() allocates memory to hold the list object; however, it then transfers the ownership of this memory to the caller function (in this case main()). Note that this is purely a design choice, or a software pattern -- the compiler is completely oblivious to this concept of ownership; instead, it is a concept used by the programmer to understand how the interface is meant to be used. In this case, the main() function should be thought of as an user of the interface. The fact that main() becomes the owner of the object means that the it is responsible for deleting the object when the latter is no longer needed, by calling ListDelete().

5. The list elements are allocated by the List object, for example when the ListAppend() function is called (ListInsertAfterElement() has a similar effect). Note that the ListAppend() function does not transfer the ownership of the new list element to main() (which is clear because it does not return a pointer to the new element). Instead, the list object retains ownership of all the list elements. This is why ListDelete() must delete any list element in the list before deleting the list itself.

6. For example:

```c
/** ListInsertAfterElement - Insert a new element after another.
 ** param "0060element"0060 - a pointer to the reference list element.
 ** param "0060x"0060 - the value of the new list element.
 **
 ** The function ListInsertAfterElement() creates a new list element
 ** with value "0060x"0060 and inserts it after "0060element"0060. Note
 ** that the list
 ** retains the ownership of the new element.
 **
 ** Errors: if there is not sufficient memory to allocate the
 ** new list element, an exception is raised terminating the program.
 **/

/** ListFindElement - Find an element by value.
 ** param "0060list"0060 - a pointer to the list.
 ** param "0060x"0060 - the element value to search for.
 **
 ** The function ListFindElement() searches the list "0060list"0060 for the
```
Problem 8  Simple algorithms on dynamic structures

With reference to the list program developed in the previous question

1. Add to the file main.c a function void print(List* list) that prints the value of the elements in the list list. Note that the DataType of the values is double, such that the correct code for printf is "%g" (i.e. printf("%g", x) prints x correctly provided that x is a variable of type double).

2. Using the new print() function, compile and run the following version of the main() function:

```c
int main(int argc, char **argv)
{
    List *list;
    ListElement *element;

    list = ListCreate();

    ListAppend(list, 1);
    ListAppend(list, 2);
    ListAppend(list, 3);
    ListAppend(list, 4);
    print(list);

    element = ListFindElement(list, 3);
    ListInsertAfterElement(element, 10);
    print(list);

    ListRemoveAfterElement(ListFindElement(list, 2));
    print(list);

    ListClear(list);
    print(list);

    ListAppend(list, 3);
    print(list);

    ListDelete(list);
    return 0;
}
```

Comment on the generated output.
3. Now write a `merge(list3, list1, list2)` function merging two lists `list1` and `list2` into a third list `list3` in such a way that, if `list1` and `list2` are in non-decreasing order, so is `list3` (this is the \textit{merge} step of merge sort). Test your function using the following program:

```c
int main(int argc, char **argv)
{
    List *list1, *list2, *list3;

    list1 = ListCreate();
    list2 = ListCreate();
    list3 = ListCreate();

    ListAppend(list1, 1);
    ListAppend(list1, 3);
    ListAppend(list1, 4);
    ListAppend(list1, 7);

    ListAppend(list2, 2);
    ListAppend(list2, 5);
    ListAppend(list2, 6);

    merge(list3, list1, list2);

    print(list1);
    print(list2);
    print(list3);

    ListDelete(list3);
    ListDelete(list2);
    ListDelete(list1);
    return 0;
}
```

\textbf{Solution to problem 8}

1. An implementation is:

```c
#include <stdio.h>

void print(List *list)
{
    ListElement* element = ListGetFirstElement(list);
    printf("(");
    while (element) {
        printf("%d", ListElementGetValue(element));
        element = ListElementGetSuccessor(element);
    }
}
```
Note that, by using suitable functions of the interface, there is no need to know the implementation details of both List and ListElement, which remain opaque objects.

2. For example:

```c
void merge(List *list3, List *list1, List *list2)
{
    ListElement *element1 = ListGetFirstElement(list1);
    ListElement *element2 = ListGetFirstElement(list2);

    while (element1 || element2) {
        int consume;
        DataType x1 = element1 ? ListElementGetValue(element1) : 0;
        DataType x2 = element2 ? ListElementGetValue(element2) : 0;
        if (element1 && element2) {
            consume = (x1 <= x2) ? 1 : 2;
        } else {
            consume = (element1) ? 1 : 2;
        }
        if (consume == 1) {
            ListAppend(list3, x1);
            element1 = ListElementGetSuccessor(element1);
        } else {
            ListAppend(list3, x2);
            element2 = ListElementGetSuccessor(element2);
        }
    }
}
```

Problem 9 Algorithms and loop invariants

The method of bisection for finding the roots of a function $f(x)$ is as follows: Starting with two points $x_1$ and $x_2$ such that $x_1 \leq x_2$ in the domain of $f$:

- Compute $f(r)$ where $r = (x_1 + x_2)/2$.
- If $x_2$ is sufficiently close to $x_1$ return $r$ as a root.
- If not, then examine the signs of $f(r)$, $f(x_1)$ and $f(x_2)$. If the signs of $f(r)$ and $f(x_1)$ differ then look for a root in $[x_1, r)$, otherwise look in $(r, x_2]$.

1. Write a MATLAB function to implement it.
2. What are the preconditions and loop invariant required to implement this algorithm as a loop?
3. The loop invariant is insufficient to prove the correctness of this algorithm. One must also show that the algorithm terminates and that it does so in a useful state. What is an upper bound on the number of iterations that the algorithm executes? What can you say about the precision of the estimated root at the end of the algorithm?

4. If the preconditions are not satisfied, what does your implementation do?

Solution to problem 9

1. The following MATLAB function implements `Bisection()`:

```matlab
function r = Bisection(func, x1, x2)

fx1 = feval(func, x1);
fx2 = feval(func, x2);

% Check the validity of the precondition
assert(x1 <= x2);
assert(fx1*fx2 <= 0);

eps = 1.0e-6;
r = (x1+x2)/2;
fr = feval(func,r);

while (x2 - x1 > eps)
    fx1 = feval(func,x1);
    if (fx1*fr <= 0)
        x2 = r;
    else
        x1 = r;
    end
    r = (x1+x2)/2;
    fr = feval(func,r);
end
```

The following MATLAB function demonstrates how to use `Bisection()`:

```matlab
function BisectionExample()

% find a root of cos(x)
x = Bisection(cos, 0, pi); fprintf('x =
    Bisection((x) x.^2 - 3*x + 2, 0, 1.5);
    fprintf("0027%.2f is a root of x.^2 - 3*x + 2\n"0027, x);
```

2. The preconditions are the requirements that the input must satisfy for the algorithm to execute correctly. In this case, the requirements are that \( f \) is a continuous function, \( x_1 \leq x_2 \), and \( f(x_1) \) and \( f(x_2) \) have opposite signs or are null. Taken together, these conditions
guarantee that there is a root in the interval \([x_1, x_2]\). The loop invariant \(P\) enforces this condition throughout the execution of the algorithm:
\[
P: \quad x_1 \leq r \leq x_2 \quad \text{and} \quad f(x_1)f(x_2) \leq 0.
\]

Proving the validity of the invariant requires showing that \(P\) is true at the end of the loop (line 21) given that the invariant \(P'\) from the previous iteration is true at the beginning of the loop (line 12). Note that, in this case, \(P\) and \(P'\) formally coincide; however, \(P'\) is a statement about the variables before the body of the loop is executed (i.e. at the end of the previous iteration), while \(P\) after that.

To show that \(P\) is true, examine the if statement at lines 16–20: if \(f(x_1)f(r) \leq 0\), then setting \(x_2 \leftarrow r\) retains the invariant \(f(x_1)f(x_2) \leq 0\); otherwise, use the fact that \(f(x_1)f(x_2) \leq 0\); since \(f(x_1) * f(r) > 0\), we can multiply this quantity to both sides of the inequality, obtaining \(f(x_1)^2f(r)f(x_2) \leq 0\). Since \(f(x_1)f(r) > 0\), \(f(x_1)\) is not zero and \(f(x_1)^2\) is strictly positive. Hence this quantity can be simplified, resulting in \(f(r)f(x_2) \leq 0\). Thus setting \(x_1 \leftarrow r\) retains the invariant \(f(x_1)f(x_2) \leq 0\). Finally, line 21 \(r \leftarrow (x_1 + x_2) < 2\) restores the property \(x_1 \leq r \leq x_2\).

3. The algorithm terminates when when \(x_2 - x_1 \leq \epsilon\). Since there is always a root in the interval \([x_1, x_2]\), it means that the point \(r\) returned by the algorithm is within a distance \(\epsilon/2\) from the root. Note also that the algorithm halves the length of the interval at each iteration. Hence it stops when: \((x_2 - x_1)/2^t \leq \epsilon\), i.e. \(t = \lceil \log_2 \frac{x_2 - x_1}{\epsilon} \rceil\).

4. If precondition that \(f\) continuous not satisfied then behaviour unpredictable. Other precondition is easily tested before the loop and an error can be flagged.

**Problem 10  Advanced algorithms**

An undirected graph can be represented in MATLAB using a single matrix in which the \((i,j)\)-th entry in the matrix is the weight of the edge between nodes \(i\) and \(j\), with a weight of zero meaning no edge exists (we assume that edges must have strictly positive weight).

A spanning tree of a graph is a tree (i.e. every vertex can be reached from every other vertex via one and only one path) containing all the nodes of the graph. A minimal spanning tree of a graph is the panning tree whose sum of edge weights is minimal.

1. The MATLAB code below finds the minimal spanning tree of the input argument graph using Prim's algorithm

   \[
   \text{http://en.wikipedia.org/wiki/Prim's_algorithm}
   \]

   Read and understand the MATLAB code, and add comments where indicated that describe the meaning of each part of the code.

2. Write down the matrix \(G\) that represents the graph shown in the figure below.

3. Show an execution trace of the function running on this graph, clearly stating the values of the variables at each starred location in the code.
%% Compute minimal spanning tree of graph G
function T = MinST(G)

%% Add comment here describing V1 and V2
V1 = [1];
V2 = 2: length(G);

%% T is the result, set to have no edges
T = zeros(size(G));

%% max is just a big number bigger than any edge weight in G
max = 100000;

while (~isempty(V2))
    % *** Execution trace: V1, V2, T
    % Add a comment here: what do the following loops do?
    min = max;
    for i=1:length(V1)
        for j=1:length(V2)
            if (G(V1(i),V2(j))>0 && G(V1(i),V2(j))<min)
                % add a comment here: what is the condition
                % and what do these lines do?
                min = G(V1(i),V2(j));
                u = V1(i);
                v = V2(j);
            end
        end
    end
    % *** Execution trace: u, v, min

%% What does the following line do and do the invaraints
%% of the loops above ensure it is the right thing?
T(u,v) = min;

%% explain the following two lines
V1 = [V1 v];
Solution to problem 10

```matlab
% Compute minimal spanning tree of graph G
function T = MinST(G)

% V1 is the set of vertices that have been added to the minimal
% spanning tree. V2 is the set of vertices remaining. These two are
% disjoint sets, and at each step we try to pick the edge between the
% two sets that in minimal (a greedy choice)
V1 = [1];
V2 = 2:length(G);

% T is the result, set to have no edges
T = zeros(size(G));

% max is just a big number bigger than any edge weight in G
max = 100000;

while (~isempty(V2))
    % print trace
    V1
    V2
    T

    % min keeps track of the lowest weight edge so far. We start it
    % out as a "big" number; i.e. bigger than any potential edge weight.
    % the two loops consider each pair of vertices with one from V1 and one
    % from V2 and choose the one with the least weight.
    min = max;
    for i=1:length(V1)
        for j=1:length(V2)
            if (G(V1(i),V2(j))>0 && G(V1(i),V2(j))<min)
                % the condition checks that there is an edge (>0) and
                % if its weight is smaller than the best so far. If it beats the
                % current best then we store the weight and save which two vertices
                % are involved in u and v
                min = G(V1(i),V2(j));
                u = V1(i);
                v = V2(j);
            end
        end
    end
end
```
% T is the matrix representing the graph of the minimal spanning tree
% Here we have decided that u and v are joined in the MST so we add that
% to T, with its weight min (which equals G(u,v)). The invariant of the
% loops above is that min, u, v contains the lowest weight edge between the
% V1(1)...V1(i) and V2(1)...V2(j). The terminating condition is clearly
% that we have the lowest overall because i is past the end of V1 and
% likewise j is past the end of V2.
T(u,v) = min;

V1 = [V1 v]; % Add v to V1 set of vertices in the MST
V2(V2==v)=[ ]; % Remove v from V2, the set of vertices not yet visited
end

1.

G = [0 1.5 3.9 4.8;
    1.5 0 1.8 0;
    3.9 1.8 0 3.2;
    4.8 0 3.2 0]

2.

V1 = 1
V2 =
    2 3 4
T =
    0 0 0 0
    0 0 0 0
    0 0 0 0
    0 0 0 0
u = 1
v = 2
min = 1.5000
V1 =
    1 2
V2 =
    3 4
T =
\[
\begin{array}{cccc}
0.00000 & 1.50000 & 0.00000 & 0.00000 \\
0.00000 & 0.00000 & 0.00000 & 0.00000 \\
0.00000 & 0.00000 & 0.00000 & 0.00000 \\
0.00000 & 0.00000 & 0.00000 & 0.00000 \\
\end{array}
\]

\[
u = 2
\]
\[
v = 3
\]
\[
\text{\texttt{min}} = 1.8000
\]
\[
V1 = \\
1 
2 
3
\]
\[
V2 = 4
\]
\[
T = \\
0.00000 & 1.50000 & 0.00000 & 0.00000 \\
0.00000 & 0.00000 & 1.80000 & 0.00000 \\
0.00000 & 0.00000 & 0.00000 & 0.00000 \\
0.00000 & 0.00000 & 0.00000 & 0.00000 \\
\]

\[
u = 3
\]
\[
v = 4
\]
\[
\text{\texttt{min}} = 3.2000
\]
\[
\text{\texttt{ans}} = \\
0.00000 & 1.50000 & 0.00000 & 0.00000 \\
0.00000 & 0.00000 & 1.80000 & 0.00000 \\
0.00000 & 0.00000 & 0.00000 & 3.20000 \\
0.00000 & 0.00000 & 0.00000 & 0.00000 \\
\]
A  Linked list implementation

```c
#include "list.h"
#include <stdio.h>
#include <stdlib.h>
#include <assert.h>

/* -------------------------------------------------------------- */
/* ListaElement implementation */
/* -------------------------------------------------------------- */

typedef struct ListaElement_
{
    struct ListaElement_ *next ;
    DataType data ;
} ListaElement ;

ListaElement * ListaElementGetSuccessor(ListaElement *element)
{
    if (element) {
        return element->next ;
    } else {
        return NULL ;
    }
}

DataType ListaElementGetValue(ListaElement *element)
{
    assert(element) ;
    return element->data ;
}

/* -------------------------------------------------------------- */
/* Lista implementation */
/* -------------------------------------------------------------- */

Lista * ListaCreate()   
{  
    Lista* list = malloc(sizeof(Lista)) ;
    if (list) {
        list->next = NULL ;
    }
    return list ;
}
```
```c
void ListDelete(List *list) {
    assert(list);
    ListClear(list);
    free(list);
}

void ListClear(List *list) {
    assert(list);
    while (!ListIsEmpty(list)) {
        ListRemoveAfterElement(list);
    }
}

int ListIsEmpty(List *list) {
    return (list->next == NULL);
}

ListElement * ListGetFirstElement(List *list) {
    assert(list);
    return list->next;
}

ListElement * ListGetLastElement(List *list) {
    assert(list);
    ListElement* last = list;
    while (last->next) {
        last = last->next;
    }
    return last;
}

ListElement * ListFindElement(List *list, DataType x) {
    assert(list);
    ListElement* current = ListGetFirstElement(list);
    while (current && current->data != x) {
        current = current->next;
    }
    return current;
}

void ListInsertAfterElement(ListElement *element, DataType x)
```
```c
{  
    assert(element);  
    ListElement *newElement = malloc(sizeof(ListElement));  
    assert(newElement);  
    newElement->next = element->next;  
    newElement->data = x;  
    element->next = newElement;  
}

void ListRemoveAfterElement(ListElement *element)  
{  
    assert(element);  
    ListElement *elementToRemove = element->next;  
    if (elementToRemove) {  
        element->next = elementToRemove->next;  
        free(elementToRemove);  
    }  
}

void ListAppend(List *list, DataType x)  
{  
    assert(list);  
    ListInsertAfterElement(ListGetLastElement(list), x);  
}
```

### B  Getting started with compiling and running programs

These instructions should help you getting started with entering, compiling, and running your C programs in Windows, Mac OS X, or Linux. Remember to use Google if you are in trouble, or switch to one of the on-line C compilers as suggested in Problem 2.

#### B.1  Windows and Visual C


- Download and install the Visual C IDE. Use Microsoft *Dreamspark* to download a free student copy using your Oxford credentials ([https://www.dreamspark.com](https://www.dreamspark.com)).

- In the Visual C++ project types pane, click *Win32*, and then click **Win32 Console Application**.

- Type a name for the project (e.g. *b16-tutorial*). By default, the solution that contains the project has the same name as the project, but you can type a different name. You can also type a different location for the project.

- Click **OK** to create the project.

- In the *Win32 Application Wizard*, click **Next**, select **Empty Project**, and then click **Finish**.
• If Solution Explorer is not displayed, on the View menu, click Solution Explorer.

• Add a new source file to the project, as follows.
  
  – In Solution Explorer, right-click the Source Files folder, point to Add, and then click New Item.
  
  – In the Code node, click C File (.c), type a name for the file, and then click Add. The .c file appears in the Source Files folder in Solution Explorer, and the file is opened in the Visual Studio editor.
  
  – In the file in the editor, write the code of your program.
  
  – Save the file.

• On the Build menu, click Build Solution. The Output window displays information about the compilation progress, for example, the location of the build log and a message that indicates the build status.

• On the Debug menu, click Start without Debugging. A command window is displayed with the output of your program.

B.2 Mac OS X and Xcode

• Install Xcode from the Mac App Store.

• Open Xcode.

• Select the menu File→New→Project…. This will show a window similar to the following:
As a template, choose **OS X→Application→Command Line Tool**.

- Choose a name for your product" (project) and make sure you choose C as the Language. Do not include spaces in the product name.

![Screenshot of Xcode project creation window](image)

- This will start Xcode with the following window. On the left, there is a navigator to find and access the project files. On the right information about the selected item (in this case the product" or executable program). In the middle you have the editor, which you can use to change the item.
• Choose the `main.c` file in the navigator. In the editor, paste your program. Check your spelling and then press the play'' symbol (triangle) in the top left. This will compile your program and execute it.

• Provided that there are no errors, Xcode will show the output of your program in the bottom right. If there are errors, select them in the navigator, fix them, and try again.
B.3 Linux and GCC

- The most common Linux C/C++ compiler is GCC. Download and install the GCC compiler package for your platform (the exact instructions depend on the specific Linux distribution; from the command line `apt-get install gcc` might work). Make also sure you install a text editor (e.g. `emacs` or `vi` and learn how to use it).

- Open the terminal application and create a new directory for your project. For example

  ```bash
  > mkdir b16-tutorial
  > cd b16-tutorial
  ```

- Create a file for your program, for example `exercise1.c`:

  ```bash
  > edit exercise1.c
  ```

  This will invoke the default editor. Instead of `edit`, you can of course use the editor of your preference.

- Save your program and go back to the terminal. Make sure that the `exercise1.c` file contains your text. To do so, you can for example use the `cat` command:

  ```bash
  > cat exercise1.c
  ```

- Run GCC to compile your program

  ```bash
  > gcc -o exercise1 exercise1.c
  ```

  The `-o exercise1` option tells GCC that you want the executable program to be called `exercise1` (without extension).

- If there are errors, find them in your program using the editor, fix them, save, and try again.
• Once the compilation completes, you can execute your program. Go back to the terminal and type

`> ./exercise1`

Note the `./` at the beginning: this tells Linux to search for the executable in the current directory. Now your program should run.