

## Finite Numbers

Choose three of the following exercises.

1. Use the Matlab functions `eps`, `realmax`, `realmin` to compute the Floating Point Systems parameters  $t, L, U$  assuming that  $\beta = 2$ . Fill the gap.

```
-----  
MATLAB FLOATING POINT SYSTEM  
F(2,t,L,U)  
-----  
UFL   =  
OFL   =  
Precisione Macchina =  
t     =  
L     =  
U     =  
-----
```

### Tips:

Given the Floating Point system  $\mathcal{F}(\beta, t, L, U)$ , you know that:

$$\epsilon_{mach} = \beta^{1-t}, UFL = \beta^L, \quad OFL = (\beta - \beta^{-t+1})\beta^U.$$

Can you get  $t$  from the expression of  $\epsilon_{mach}$ ?

2. Compute the machine precision  $\epsilon$  for single and double precision, using its alternative definition:

$$fl(1 + \epsilon) > 1.$$

### Tips:

Use a `while` structure. To set a single precision use the function `single`.

3. Truncation error in  $\pi$  approximation. In mathematics, the Leibnitz formula for  $\pi$ , states that:

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \dots$$

It is obtained using Taylor formula for the  $\tan^{-1}$  function:

$$\arctan(x) = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + o(x^9),$$

setting  $x = 1$  since  $\arctan(1) = \frac{\pi}{4}$ . Fix  $n \in \mathbb{N}$  and compute the approximation of  $\pi$  using Leibnitz formula truncated at  $n$ -th term. Compare the result with the true value of  $\pi$ .

4. Let's consider the sequence  $a_n = (1 + \frac{1}{n})^n$ . It is well known that:

$$\lim_{n \rightarrow +\infty} a_n = e,$$

where  $e$  is the Euler constant. Choose different values for  $n$ , compute  $a_n$  and compare it to the real value of the Euler constant. What happens if you choose a large value of  $n$ ?