Homework 0

Theory:
1. Give examples of floating-point numbers $x$, $y$, and $z$ for which addition is not associative. Do the same for multiplication. (Assume a 64 bit floating point representation, with 10 bits allocated to the power, 53 to the mantissa, and 1 to the sign.)

2. (Challenge 1.6) Suppose $x$ and $y$ are true (non-zero) values and $\tilde{x}$ and $\tilde{y}$ are our approximations to them. Express the errors as

$$\tilde{x} = x(1 - r)$$
$$\tilde{y} = y(1 - s).$$

Show that the relative error in $\tilde{x}$ is $|r|$ and the relative error in $\tilde{y}$ is $|s|$. Also show that the relative error in $\tilde{x}\tilde{y}$ (as an approximation to $xy$) is bounded by $|r| + |s| + |rs|.$

Computation:
1. Write a Matlab program that determines the length of the mantissa on your machine. (Hint: use a simple while loop.)

2. (Challenge 1.8) Write a Matlab function that computes the two roots of a quadratic polynomial with good relative precision.