1. Compute the roots of the following quadratic equation

\[ x^2 + 111.11x + 1.1111 = 0 \]

in the floating-point number system \( F(10,5,-50,50) \) with rounded arithmetic.

Sol. Theoretically, the roots of the above equations are:

\[ x_1 = \frac{-111.11 + \sqrt{(111.11)^2 - 4(1.1111)}}{2} \quad (1.1) \]

\[ x_2 = \frac{-111.11 - \sqrt{(111.11)^2 - 4(1.1111)}}{2} \quad (1.2) \]

However, since there is a cancellation in (1.1), we should compute \( x_1 \) as follows:

\[ x_1 = \frac{-2(1.1111)}{111.11 + \sqrt{(111.11)^2 - 4(1.1111)}} \quad (1.3) \]

By carrying out the all the operations of (1.3) and (1.2) in \( F(10,5,-50,50) \) with rounded arithmetic, we have

\[ x_1 = \frac{-2.2222}{111.11 + \sqrt{12345 - 4.4444}} \]

\[ = \frac{-2.2222}{111.11 + 12341} \]

\[ = \frac{-2.2222}{111.11 + 111.09} \]

\[ = \frac{-2.2222}{222.2} \]

\[ = -0.010001 \]

and

\[ x_2 = \frac{-111.11 - 111.09}{2} \]
Note that if $x_1$ is calculated using (1.1), we would get

$$x_1 = -0.01$$

2. Generally, when a list of floating-point numbers is added, less round-off error will occur if the numbers are added in the order of increasing magnitude. Give an example of your own to illustrate this principle.

Sol. Consider $x_1 + x_2 + x_3$ in $F(10,3,-50,50)$ with chopped arithmetic, where

$$x_1 = 10 \quad x_2 = 0.15 \quad x_3 = 0.05$$

If we add the numbers in the order of decreasing magnitude, i.e., from $x_1$ to $x_3$, we have

$$fl(x_1 + x_2 + x_3) = fl(fl(x_1 + x_2) + x_3)$$
$$= fl(fl(10 + 0.15) + 0.15)$$
$$= fl(10.1 + 0.05)$$
$$= 10.1$$

If we add the numbers in the order of increasing magnitude, i.e., from $x_3$ to $x_1$, we have

$$fl(x_1 + x_2 + x_3) = fl(x_1 + fl(x_2 + x_3))$$
$$= fl(10 + fl(0.15 + 0.15))$$
$$= fl(10 + 0.2)$$
$$= 10.2$$

Obviously 10.2 is a better solution because it is the exact solution.

3. [Forward Elimination without Pivoting]

Convert the following system of linear equations into an equivalent, upper triangular system of linear equations without pivoting, in the floating-point number system $F(10,5,-128,127)$ with chopped arithmetic.

$$\begin{align*}
2x_1 - 3x_2 + 4x_3 &= 8 \quad \text{......... (1)} \\
5x_1 + 2x_2 - x_3 &= 6 \quad \text{......... (2)} \\
4x_1 + x_2 + 2x_3 &= 12 \quad \text{......... (3)}
\end{align*}$$

Sol. First, eliminate entries in the first column that are below the diagonal line:
(2) - (1)*\(\frac{5}{2}\)
\[
\begin{align*}
5x_1 + 2.0x_2 - x_3 &= 6 \\
-5x_1 + 7.5x_2 + 10x_3 &= 20 \\
\end{align*}
\]
\[
9.5x_2 - 11x_3 = -14
\]

(3) - (1)*\(\frac{4}{2}\)
\[
\begin{align*}
4x_1 + x_2 + 2x_3 &= 12 \\
-4x_1 - 6x_2 + 8x_3 &= 16 \\
\end{align*}
\]
\[
7x_2 - 6x_3 = -4
\]

We get
\[
\begin{align*}
2x_1 - 3x_2 + 4x_3 &= 8 \quad \ldots \ldots \quad (1) \\
9.5x_2 - 11x_3 &= -14 \quad \ldots \ldots \quad (2) \\
7.0x_2 - 6x_3 &= -4 \quad \ldots \ldots \quad (3)
\end{align*}
\]

Then we eliminate the entry in the second column that are below the main diagonal:

(3) - (2)*\(\frac{7}{9.5}\)
\[
0.73684
\]
\[
\begin{align*}
7x_2 - 6.0000x_3 &= -4.000 \\
-7x_2 + 8.1052x_3 &= -10.315 \\
\end{align*}
\]
\[
2.1052x_3 = 6.315
\]

Hence, the upper triangular system is:
\[
\begin{align*}
2x_1 - 3x_2 + 4x_3 &= 8 \quad \ldots \ldots \quad (1) \\
9.5x_2 - 11x_3 &= -14 \quad \ldots \ldots \quad (2) \\
-2.1052x_3 &= 6.315 \quad \ldots \ldots \quad (3)
\end{align*}
\]

4. [Forward Elimination with Pivoting]

Convert the above system of linear equations into an equivalent, upper triangular system of linear equations with pivoting, in the floating-point number system \(F(10,5,-128,127)\) with chopped arithmetic. (Note that you have to physically move the rows of the system in some cases)

Sol. First, we switch Eq.(1) and Eq.(2) to make Eq.(2) the pivot equation for column one:
Next, eliminate entries in the first column that are below the diagonal line:

(2) - (1)\*\frac{2}{5} = 0.4

\[2x_1 - 3.0x_2 + 4.0x_3 = 8.0\]
\[2x_1 + 0.8x_2 - 0.4x_3 = 2.4\]

\[\text{----------------------------------------}\]
\[-3.8x_2 + 4.4x_3 = 5.6\]

(3) - (1)\*\frac{4}{5} = 0.8

\[4x_1 + x_2 + 2.0x_3 = 12.0\]
\[4x_1 + 1.6x_2 - 0.8x_3 = 4.8\]

\[\text{----------------------------------------}\]
\[-0.6x_2 + 2.8x_3 = 7.2\]

We get

\[\begin{cases}
5x_1 + 2x_2 - x_3 = 6 \\
-3.8x_2 + 4.4x_3 = 5.6 \\
-0.6x_2 + 2.8x_3 = 7.2
\end{cases}\]  \hspace{1cm} (1)  \hspace{1cm} (2)  \hspace{1cm} (3)

For stage 2, no row switching is needed because the coefficient of \(x_2\) in the second equation is already the largest in the second column. Hence, we only need to perform the following operation:

(3) - (2)\*\frac{0.6}{3.8} = 0.15789

\[\text{----------------------------------------}\]
\[2.1052x_3 = 6.3158\]

Hence, the upper triangular system is:

\[\begin{cases}
5x_1 + 2x_2 - x_3 = 6 \\
-3.8x_2 + 4.4x_3 = 5.6 \\
2.1052x_3 = 6.3158
\end{cases}\]  \hspace{1cm} (1)  \hspace{1cm} (2)  \hspace{1cm} (3)