Solutions to Homework 2

1. a) Let $X$ be the largest floating-point number in the system. In this case the addition $X+X$ will produce an overflow since $2X>X$.

b) Using the notation in a), and setting $Y=-X$, then $X-Y$ will produce an overflow.

c) Using the notation in a) then $X*X$ will produce an overflow.

d) Using the notation in a), and letting $Z$ be any floating-point number between 0 and 1, then $X/Y$ will produce an overflow since $X/Y > X$.

2. a) Using the expression for relative error in class we have:

$$-\varepsilon/2 \leq (x_f-x)/x \leq \varepsilon/2$$

By considering each inequality separately, we obtain the result.

b) From part a), writing $x_f = x(1+\delta)$ and $y_f = y(1+\delta)$, we see that

$$(x_f y_f) = x_f y_f (1+\delta) = x(1+\delta) y(1+\delta) (1+\delta)$$

The result follows once you expand $(1+\delta)^3$ and note that $\delta^2$ and $\delta^3$ terms may be ignored because of the small value of $\delta$.

3. a) Given a function $g(x)$ then its third degree Taylor polynomial about the point $x=0$ is

$$g(0) + xg'(0) + \frac{x^2}{2}g''(0) + \frac{x^3}{6}g'''(0)/6$$

Letting $g(x) = \ln(1-x)$ then $g' = -(1-x)^{-1}$, $g'' = -(1-x)^{-2}$, $g''' = -2(1-x)^{-3}$. Inserting these into the above Taylor polynomial gives:

$$\ln(1-x) \approx -x - \frac{x^2}{2} - \frac{x^3}{3}$$

The subsequent approximation we obtain for $f(x)$ is

$$\ln(1-x)/x \approx -1 - x$$

From this it follows that we should assign the value –1 to $f(0)$.

b) Using the MATLAB command `fplot('log(1-x)/x', [-5.0e-16, 5.0e-16])` produces the appropriate plot. The plot shows that MATLAB assigns the value zero to $f(x)$ near $x=0$. 

c) Using the MATLAB command `fplot('log(1-x)', [-1.0e-15, 1.0e-15])` produces the appropriate plot. To explain the first step in the graph after \(x=0\) it is best to investigate what values the steps have. This is done using the following MATLAB commands.

```matlab
>> format long e
>> log(1-1.0e-16)
ans =
-1.110223024625157e-016
```

```matlab
>> log (1-2.0e-16)
ans =
-2.220446049250314e-016
```

So the first non-zero value is -1.110223024625157e-016 which equals \(\varepsilon/2\), which can be checked in MATLAB as follows:

```matlab
>> eps/2
ans =
1.110223024625157e-016
```

The stairs appear in the graph because of the limited set of values the computer is able to assign to \(h(x) = 1-x\) when \(x\) is very near zero. This happens because \(h(0)=1\) and the floating-point numbers just to the left of 1 are \(1-\varepsilon/2, 1-\varepsilon, 1-3\varepsilon/2,...\) etc. So, as \(x\) increases from \(x=0\) the values of \(\ln(1-x)\) that the computer has to work with are \(\ln(1-\varepsilon/2), \ln(1-\varepsilon), \ln(1-3\varepsilon/2),...\) These are exactly the values for the first three steps to the left of \(x=0\) in the plot.

The steps explain the oscillations in part b) as the peaks seen in b) correspond to the edges of the steps in the plot for part c). During the interval in which \(\ln(1-x)\) is constant, MATLAB is essentially plotting \(a/x\) where \(a\) is a constant for a particular interval, but whose value is different in different intervals. So the oscillations to the right of zero correspond to a piecewise function \(a/x (a<0)\) and the oscillations to the left of zero correspond to a piecewise function \(a/x (a>0)\).

d) The oscillations are not symmetric because the floating-point numbers are not distributed symmetrically near \(x=1\). In particular the floating-point numbers to the left and right of \(h(0)\) are \(...1-3\varepsilon/2, 1-\varepsilon, 1-\varepsilon/2, 1, 1+\varepsilon, 1+2\varepsilon, 1+3\varepsilon,...\) Consequently the values for the function \(1-x\) when \(x>0\) are \(1-\varepsilon/2, 1-\varepsilon, 1-3\varepsilon/2,...\) and when \(x<0\) the values are \(1+\varepsilon, 1+2\varepsilon, 1+3\varepsilon,...\). This also explains why there are more oscillations to the right.
4. a) \[ A = \begin{bmatrix} 80 & 0 & 30 & 10 \\ 0 & 80 & 10 & 10 \\ 16 & 20 & 60 & 72 \\ 4 & 0 & 0 & 8 \end{bmatrix} \]

\[ \mathbf{b} = \begin{bmatrix} 40 \\ 27 \\ 31 \\ 2 \end{bmatrix} \]

b) \[ A=[80 \ 0 \ 30 \ 10; \ 0 \ 80 \ 10 \ 10; \ 16 \ 20 \ 60 \ 72; \ 4 \ 0 \ 0 \ 8]; \]

\[ \mathbf{b}=[40;27;31;2]; \]

\[ \mathbf{x}=A\backslash\mathbf{b} \]

\[ \mathbf{x} = \begin{bmatrix} 0.4000 \\ 0.3000 \\ 0.2500 \\ 0.0500 \end{bmatrix} \]