Homework #1

2. Write $x = \pm 0.d_1 d_2 \dots d_k d_{k+1} \dots \cdot 10^a$ with $d_1 \neq 0$. If $d_{k+1} < 5$, then $fl(x) = \pm 0.d_1 d_2 \dots d_k \cdot 10^a$. So

$$\frac{|x - fl(x)|}{|x|} = \frac{0.d_{k+1}d_{k+2}\dots \cdot 10^{-k}}{0.d_1 d_2 \dots} \leq \frac{0.499\dots \cdot 10^{-k}}{0.1} = 5 \cdot 10^{-k}.$$

If $d_{k+1} \geq 5$, then $fl(x) = \pm(0.d_1 d_2 \dots d_k + 10^{-k}) \cdot 10^a$. So

$$\frac{|x - fl(x)|}{|x|} = \frac{(1 - 0.d_{k+1}d_{k+2}\dots) \cdot 10^{-k}}{0.d_1 d_2 \dots} \leq \frac{0.5\dots \cdot 10^{-k}}{0.1} = 5 \cdot 10^{-k}.$$

4. We are interested in $fl(fl(x)+fl(y))$, the machine computation of $x+y$. For any real z, $fl(z) = z(1+r)$, for some r satisfying

$$|r| = \frac{|fl(z) - z|}{|z|} = \text{relative error of } fl(z) \leq u.$$

So $fl(x) = x(1+r_1)$ and $fl(y) = y(1+r_2)$ and $fl(fl(x)+fl(y)) = (fl(x)+fl(y))(1+r_3)$, for some $|r_i| \leq u$, with the last equation using $z = fl(x) + fl(y)$. Hence

$$\begin{aligned}
fl(fl(x) + fl(y)) &= (x(1+r_1) + y(1+r_2))(1+r_3) \\
&= x(1 + r_1 + r_3 + r_1 r_3) + y(1 + r_2 + r_3 + r_2 r_3) \\
&= (x+y)\left(1 + \frac{(r_1 + r_3 + r_1 r_3)x}{x+y} + \frac{(r_2 + r_3 + r_2 r_3)y}{x+y}\right) \\
&= (x+y)(1+R),
\end{aligned}$$

where

$$R = \frac{(r_1 + r_3 + r_1 r_3)x}{x+y} + \frac{(r_2 + r_3 + r_2 r_3)y}{x+y}.$$

Note

$$|R| = \frac{|fl(fl(x) + fl(y)) - (x+y)|}{|x+y|} = \text{relative error of } fl(fl(x) + fl(y))$$

$$\begin{aligned}
|R| &\leq (|r_1| + |r_3| + |r_1||r_3|)\frac{|x|}{|x|+|y|} + (|r_2| + |r_3| + |r_2||r_3|)\frac{|y|}{|x|+|y|} \\
&\leq (2u + u^2)\frac{|x|+|y|}{|x+y|} = (2u + u^2) = 2u + \mathcal{O}(u^2).
\end{aligned}$$

7. (a) There exists a root, call it p, in $[-3, 2]$ by the Intermediate Value Theorem. Bisection method error formula says

$$|p_n - p| \leq \frac{b-a}{2^n},$$

for starting interval $[a, b]$. So, in this case,

$$|p_{15} - p| \leq \frac{5}{2^{15}} = 1.52587890625\dots \cdot 10^{-4}.$$

(b) Bisection method error formula in our case says

$$|p_n - p| \le \frac{5}{2^n}.$$

Enforcing a bound on the error bound, $5/2^n \le 10^{-12}$, forces $|p_n - p| \le 10^{-12}$ as well. This is achieved when $5 \cdot 10^{12} \le 2^n$, or $\log(5 \cdot 10^{12})/\log(2) \le n$, which simplifies to $n \ge 42.1850652335357\ldots$, or simply $n \ge 42$.

8. (Matlab)

   (a) See "hw1afn.m".

   (b) See "hw1bfn.m".

   (c) $p_1 = 0.785398163397448$, $p_5 = 0.736310778185108$, $p_{10} = 0.737844758972993$, $p_{20} = 0.739086624278811$.

9. (Math 274)

   (a) For any real z, $fl(z) = z(1+r)$, for some r satisfying

$$|r| = \frac{|fl(z) - z|}{|z|} = \text{relative error of } fl(z) \le u.$$

By induction, for $n = 1$, $\prod_{x=1}^{n} x_i = x_1$, and its machine computation's relative error, taking $z = x_1$, is $\le u = (2n-1)u$. Now suppose the machine computation of $\prod_{x=1}^{k} x_i$, call it A, has relative error $(2k-1)u + \mathcal{O}(u^2)$. So $A = \left(\prod_{x=1}^{k} x_i\right)(1 + R_1)$, where $|R_1| = (2k-1)u + \mathcal{O}(u^2)$. The machine computation of $\prod_{x=1}^{k+1} x_i$ is $fl(A \cdot fl(x_{k+1}))$, noting A is already a machine number, and

$$\begin{aligned} fl(A \cdot fl(x_{k+1})) &= (Ax_{k+1}(1+r_1))(1+r_2) \\ &= \left(\prod_{x=1}^{k} x_i\right) x_{k+1}(1+R_1)(1+r_1+r_2+r_1r_2) \\ &= \left(\prod_{x=1}^{k+1} x_i\right)(1+R_2), \end{aligned}$$

where $R_2 = R_1 + r_1 + r_2 + r_1r_2 + R_1r_1 + R_1r_2 + R_1r_1r_2$, and $|R_2|$ is the relative error of $fl(A \cdot fl(x_{k+1}))$. So

$$\begin{aligned} |R_2| &\le |R_1| + |r_1| + |r_2| + |r_1||r_2| + |R_1||r_1| + |R_1||r_2| + |R_1||r_1||r_2| \\ &\le (2k-1)u + u + u + u^2 + (2k-1)u^2 + (2k-1)u^2 + (2k-1)u^3 + \mathcal{O}(u^2) \\ &= (2k+1)u + \mathcal{O}(u^2). \end{aligned}$$

So induction gives us our result.