CS/Math 3414 Assignment 1

Solution Sketches

---

1. There are several ways to show that a finite binary representation corresponds to a finite decimal number. The most elegant answer is to notice that every binary representation corresponds to $\pm m/2^n$, for some m, n. But the same quantity can be rewritten as $\pm m\, 5^n/10^n$, which is a finite representation in the decimal system (because of the 10 in the denominator)!

2. Three things can happen when x^2 is being computed. If it is too large, overflow will happen. If it is too small, underflow occurs. If it is within the range of representation, then an error bounded by $\epsilon/2$ can be expected. The difference between the two ways of computing x^2 can be summarized as:

$$\begin{aligned} fl(fl(x)fl(x)) &= (x(1+\delta_1)x(1+\delta_2))(1+\delta_3) \\ &\leq x^2(1+\delta)^3 \\ &\approx x^2(1+3\delta) \\ fl(xx) &= x^2(1+\delta) \end{aligned}$$

3. We can write $f(x)$ as $f(x) = g(x) - h(x)$, where

$$g(x) = \frac{\sqrt{1+x^2}-1}{x^2}$$

and

$$h(x) = \frac{x^2 sin\, x}{x - tan\, x}$$

When $x \approx 0$, there is possibility of cancelation in both $g(x)$ (in its numerator) and $h(x)$ (in the denominator). We can rewrite $g(x)$ as:

$$\begin{aligned} g(x) &= \frac{(\sqrt{1+x^2}-1)}{x^2}\frac{(\sqrt{1+x^2}+1)}{(\sqrt{1+x^2}+1)} \\ &= \frac{1}{\sqrt{1+x^2}+1} \end{aligned}$$

Similarly, by expanding the terms in $h(x)$ in a Taylor series, we get:

$$\begin{aligned} h(x) &= x^2\left(\frac{x - x^3/3! + x^5/5! - \cdots}{x - x - x^3/3 - 2x^5/15 - \cdots}\right) \\ &\approx -\left(\frac{x - x^3/6}{x/3 + 2x^3/15}\right) \\ &= -3 + \frac{17x^2}{10} \end{aligned}$$

Notice that the last step has to be done carefully. If you just omit the x^3 terms in the denominator, you don't get a very good approximation (you will get $-3 + x^2/2$ which has a big error in the coefficient of the second order term). Rather, you should carry out the division completely *before* approximating to get the right coefficient (in this case, 17/10). We can now compute $f(x)$ at $x = 0$ as $0.5 - (-3) = 3.5$.

4. It is trivial to show that $a(x) = b(x)$ (just multiply both numerator and denominator of $a(x)$ by $(1 + cos\, x)$ and cancel out $sin\, x$). To show that $c(x)$ approximates $a(x)$ in a neighborhood around zero, we can expand both numerator and denominator of $a(x)$ in a Taylor series around $x = 0$. We know that

$$\begin{aligned} cos\, x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots \\ sin\, x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots \\ \frac{1 - cos\, x}{sin\, x} &= \frac{\frac{x^2}{2!} - \frac{x^4}{4!} + \cdots}{x - \frac{x^3}{3!} + \cdots} \\ &\approx \frac{x}{2} + \frac{x^3}{24} \end{aligned}$$

Once again, notice that in the last step we just didn't blindly remove the x^3 term from the denominator. If you had done this, you will get $\frac{x}{2} - \frac{x^3}{24}$ as the answer instead of $\frac{x}{2} + \frac{x^3}{24}$!

5. Recall that $tan\, x$ has period π, so we can keep subtracting π from 10^{100} till we get a number in the range $[-\pi/2, \pi/2]$, and compute the tan of this quantity instead. If you think about what this involves, we will perform a calculation such as:

$$\theta = 10^{100} - n\,\pi$$

where n is the number of times π has to be subtracted. For the subtraction to be accurate, we will need at least 100 digits in π, since all corresponding 100 digits in 10^{100} have to cancel out!