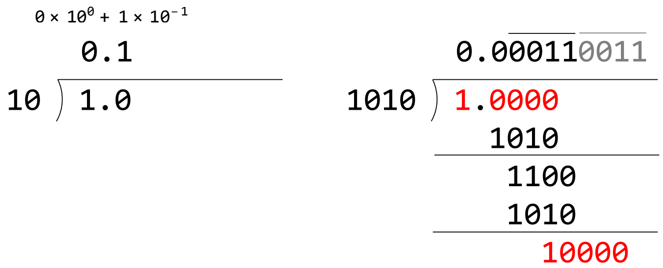
SI211 Homework 1

Prof. Boris Houska

Deadline March 8

1. *IEEE 754.* The figure below illustrates how to calculate $1/10$ using decimal or binary numbers. We use x_{D} to denote decimal numbers, and x_{B} to denote binary numbers.



   (a) Expand $0.0\overline{0011}_{\mathrm{B}}$ by writing this number in the form of an infinite geometric series.

   (b) In IEEE 754 standard, 32 bits (s`:1 bit` | c`:8 bits` | f`:23 bits`) will be used to represent a single precision floating point number $x = (-1)^s 2^{c-127}(1+f)$.

      i. What are c and f for $x_1 = 0.1_{\mathrm{D}}$?

      ii. What are c and f for $x_2 = 9.75_{\mathrm{D}}$?

   (c) Work out the IEEE 754 single precision bit-wise representation of x_1 and x_2. You may use `bitstring(Float32(x))` in `Julia`, or `num2bin(quantizer('single'), x)` in `MATLAB` to verify your result.

2. *Numerical difference.* Let f be a smooth function.

   (a) Consider the 4-point central difference formula

$$f'(x) \approx \frac{f(x-3h) - 27f(x-h) + 27f(x+h) - f(x+3h)}{\alpha h} \, .$$

   For which value of α can this approximation be expected to be accurate? Work out the mathematical approximation error and explain how to choose a sutiable h that minimizes the sum of the round-off error and the mathematical approximation error.

   (b) Next, consider the 5-point central difference formula

$$f''(x) \approx \frac{f(x-3h) - 3f(x-h) + 4f(x) - 3f(x+h) + f(x+3h)}{\beta h} .$$

For which value of β is this approximation accurate? Work out the mathematical approximation error and explain how to choose a suitable h.

(c) Consider the function

$$f(x) = \sinh(x)\sin(x)$$

i. Compute $f'$and f'' by hand.

ii. Plot the total error of the approximations (a) and (b) at $x = 1.0$ for different choices of $h \in [10^{-15}, 10^{-1}]$. Use logarithmic scales on both axes.

3. *Algorithmic differentiation.* Let $f\colon \mathbb{R}^3 \mapsto \mathbb{R}$ be given by

$$f(x) = \frac{\mathrm{e}^{x_2}}{\mathrm{e}^{x_1} + \mathrm{e}^{x_2} + \mathrm{e}^{x_3}}$$

Implement your own version of algorithmic differentiation to evaluate $\nabla_x f$ at $x = (5, 6, 7)$. Compare the program's output with your hand-derived result. Attach your code.