University of Toronto at Scarborough
CSCC37—Numerical Algorithms for Computational Mathematics, Fall 2022

Assignment #1: FP Systems / Matrix&Vector Norms

Due: October 12, 2022 at 11:45 p.m.
This assignment is worth 10% of your final grade.

Warning: Your electronic submission on *MarkUs* affirms that this assignment is your own work and no one else's, and is in accordance with the University of Toronto Code of Behaviour on Academic Matters, the Code of Student Conduct, and the guidelines for avoiding plagiarism in CSCC37.

This assignment is due by 11:45 p.m. October 12. If you haven't finished by then, you may hand in your assignment late with a penalty as specified in the course information sheet.

[10] 1. The (fictional) space probe *Venus 22* sent back the following picture of a Venusian rock-carving, assumed to constitute an addition. What is the base of the number system used by the Venusians?

Warning: Be sure to justify your answer, and the justification is not as simple as "Three symbols means base 3."

$$\boxed{\begin{array}{ccc} & \# & * \\ & \# & * \\ \hline \# & \Diamond & \# \end{array}}$$

[10] 2. Design a computer that is able to store $(0.1)_{10}$ exactly. A floating point number on your computer must be represented internally in a base less than 10, and must have a mantissa with a finite number of digits. (**Hint:** Is this possible?)

[10] 3. Prove that machine precision ϵ, as defined in lecture, can be used as a bound for relative round-off. In other words, prove that

$$\epsilon = \begin{cases} b^{1-t} & \text{chopping} \\ \frac{1}{2}b^{1-t} & \text{rounding} \end{cases}$$

where b is the base of the computer's floating point number system and t is the mantissa length.

[10] 4. For $x \in \mathbb{R}^n$, prove that

$$\|x\|_\infty = \lim_{p\to\infty} \|x\|_p$$

where

$$\|x\|_p = (|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}} \quad p \in \mathbb{N},\ p \geq 1$$

[15] 5. When both A and B are $n \times n$ upper-triangular matrices, the entries of $C = AB$ are defined as follows:

$$c_{ij} = \begin{cases} \sum_{k=i}^{j} a_{ik} b_{kj} & 1 \le i \le j \le n \\ 0 & 1 \le j < i \le n \end{cases} .$$

For $n = 2$ show that $fl(AB) = \hat{A}\hat{B}$ where $\hat{A} = A + E_A$ and $\hat{B} = B + E_B$. Derive bounds for $\|E_A\|$ and $\|E_B\|$ showing that they are small relative to $\|A\|$ and $\|B\|$, respectively. In other words, show that the computed product is the exact product of slightly perturbed A and B.

[total: 55 marks]