Otto-von-Guericke-University Magdeburg
Max Planck Institute for Dynamics of Complex Technical Systems
Computational Methods in Systems and Control Theory

Dr. Jens Saak, Dipl.-Math. Martin Köhler
Website: http://www.mpi-magdeburg.mpg.de/csc/teaching/18ws/sc1/

Winterterm 2018/2019

Scientific Computing 1
5th Homework

Handout: 8th Nov. 2018 **Return:** 16th Nov. 2018

Exercise 1: (4 Points)

Write two C functions which determine the machine epsilon in single and double precision. Check if the `-ffast-math` compiler option influences your implementation. If your implementation gets influenced by this, try to find a second implementation which is not affected by the compiler switch.

Exercise 2: (4 Points)

Write a C program which finds the smallest double precision floating point number $1 < x < 2$ such that $x \cdot \frac{1}{x}$ does not yield 1 exactly. Use your code to determine the machine epsilon from the previous exercise. Does changing the rounding behavior to `FE_UPWARD`, `FE_DOWNWARD` or `FE_TOWARDZERO` influence the result? Details about changing the rounding mode can be found in the manpage of `fenv`.

Exercise 3: (4 Points)

a.) Convert $(1011.101)_2$ and $(0.011111\ldots)_2$ to the decimal system.

b.) Convert $(1CBA)_{16}$ and $(C2D2.E3)_{16}$ into the binary and the decimal system.

c.) Convert $(131)_{10}$ and $(0.3)_{10}$ into the hexadecimal system.

d.) Convert $(763)_{10}$ and $(101101)_2$ into the octal system (base 8).

Exercise 4: (2 Points)

Proof that the grouping of 4 digits in the binary-system to one digit in the hexadecimal-system is correct.

Exercise 5: (5 Points)

Write a C program which prints all numbers that are contained in a given $\mathbb{M}(p, t, e_{min}, e_{max})$. Use the output to plot the members of $\mathbb{M}(2, 4, -2, 4)$ and $\mathbb{M}(3, 2, -1, 2)$ on the number ray. (This can be done with an arbitrary tool, e.g., MATLAB®, gnuplot, TikZ, or by hand.)

Exercise 6: (6 Points)

a.) Draw all positive numbers that can be expressed using $\mathbb{M}(2, 3, -1, 3)$ on a linearly scaled number ray and on a logarithmically scaled number ray. What difference between the linearly and the logarithmically scaled plot do you recognize?

b.) Given is an arbitrary machine number set

$$\mathbb{M} := \mathbb{M}(p, t, e_{\min}, e_{\max}) := \{ \pm 0.\alpha_1\alpha_2 \ldots \alpha_t \cdot p^b \quad | \, \alpha_i \in \{0, \ldots, p-1\}, \alpha_1 \neq 0, \\ e_{\min} \leq b \leq e_{\max}\} \cup \{0\} \, .$$

Proof that for two neighboring powers p^b and p^{b+1} ($e_{min} \leq b < b+1 \leq e_{max}$) the number of representable elements in $\mathbb{M} \cap [p^b, p^{b+1})$ is the same independent of the choice of b. What can you say about the relative length of two neighboring intervals of this type?

Overall Points: 25