Winterterm 2012/2013

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

Dr. Jens Saak, Dipl.-Math. Martin Köhler
Website: http://www.mpi-magdeburg.mpg.de/mpcsc/lehre/2012_WS_SC/

Scientific Computing 1
6th Homework

Handout: 11/08/2012 **Return:** 11/15/2012

Exercise 1: **(5 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 last homework. 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 2: **(7 Points)**

a.) Consider the following Integral:

$$y_n := \int_0^1 x^n \frac{1}{10+x} \mathrm{d}x \tag{1}$$

Compute y_0 and proof that

$$y_n = -10 y_{n-1} + \frac{1}{n} \tag{2}$$

holds $\forall n > 0$.

Hint: Use integration by parts and:

$$\int_0^1 (n-1)x^{n-1} \mathrm{d}x = \frac{n-1}{n}, \qquad \int_0^1 (n-1)x^{n-2} \mathrm{d}x = 1.$$

b.) Implement the recursion from (2) as a C function. Additionally implement the backward-recursion from y_n to y_{n-1} with the initial value $y_{30} = 0$ as a second function. Compute y_i for $0 \leq i \leq 30$ using those two functions. What do you recognize? Figure out possible reasons for this behavior.
Hint: Use $y_{20} \approx 4.34703 \cdot 10^{-3}$ to compare your results.

Exercise 3: **(4 Points)**

Floating point exception can be checked using the `fpclassify` mechanism in the C libary. Read the man page of `fpclassify` and use the described functions to check the results of the following computations:

- 1^{∞}, 2^{∞}
- e^{∞}, $e^{-\infty}$
- NaN^0, 1^{NaN}

- 2^{100}, 2^{800}, 2^{2000}
- $\log(\infty)$, $\log(\infty)$

Hint: C99 defines the two constants `INFINITY` and `NAN` in `math.h`.

Exercise 4: **(6 Points)**

Consider a generic polynomial

$$P_n(x) := \sum_{i=0}^{n} a_i x^i$$

with $a_i \in \mathbb{R}^n$.

a.) Write a C function which takes the degree n, the coefficients a_i as an array, and x as inputs and evaluates $P_n(x)$ naively.

b.) Write a second C function with the same arguments which evaluates $P_n(x)$ using the Horner-Scheme: The polynomial

$$P_n(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0$$

can be rewritten as

$$P_n(x) = (((a_n x + a_{n-1})x + a_{n-2})x + \ldots)x + a_0.$$

Use these two functions to evaluate

$$P(x) = x^6 - 998x^5 - 998x^4 - 998x^3 - 998x^2 - 998x - 998$$

at different points. For example $x = 999$. Compare and discuss the results of both functions.

Exercise 5: **(3 Points)**

Did other floating point formats than IEEE-754 play a role in computer science? Search for historic computers and calculators in the Internet and write down the corresponding $\mathbb{M}(p, t, e_{min}, e_{max})$. Why are such formats no longer in use?

Overall Points: 25

In the Tutorial

The following exercises are not part of the homework and will be solved in the tutorial. However, if you solve them successfully during the homework, you can earn some extra points.

Exercise 6:

Let x be the exact and $\hat{x}$ the computed solution of a problem. The classical definition of the relative error is $E_{rel}(\hat{x}) = |x - \hat{x}|/|x|$. In practice $\tilde{E}_{rel}(\hat{x}) = |x - \hat{x}|/|\hat{x}|$ is often used as replacement. Find inequalities to estimate $\tilde{E}_{rel}(\hat{x})$ with respect to $E_{rel}(\hat{x})$. Is the use of $\tilde{E}$ instead of E justifiable?

Exercise 7:

Reformulate the following expressions to avoid cancellation:

a.) $\sqrt{1+x} - 1, \quad x \approx 0$

b.) $\frac{1-\cos x}{\sin x}, \quad x \approx 0$

c.) $\frac{1}{1+2x} - \frac{1-x}{1+x}, \quad x \approx 0$