Homework 2

due 2/2-2015

Task 1: Machine Epsilon

The following code can be used in MATLAB to determine the so-called machine accuracy ε.

```
numprec=double(1.0); % Define 1.0 with double precision
numprec=single(1.0); % Define 1.0 with single precision
while(1 < 1 + numprec)
    numprec=numprec*0.5;
end
numprec=numprec*2
```

a) Determine ε using the above program, both for single and double precision.

   Note: The implementation of single/double precision arithmetics differs between versions of MATLAB. If runs with both single and double precision give the same answer, please try another computer/version of MATLAB if possible. Otherwise, write down your MATLAB version and move on. The above code is working properly on release 2009a to 2012b on Linux, for instance.

b) Give a definition of the machine accuracy based on the code above. Try to use words and not mathematical expressions. What does ε mean?

Task 2: Round-off Error

In this exercise, the errors involved in numerically calculating derivatives are examined. For example, the derivative of a function f can be approximated with central differences:

$$f'_{num}(x) = \frac{f(x + \Delta x) - f(x - \Delta x)}{2\Delta x} \tag{1}$$

a) Determine the relative error ξ of the derivative of the function $f(x) = \dfrac{1}{1+x} + x$ when using the central difference approximation defined above:

$$\xi = \frac{|f'(x) - f'_{num}(x)|}{|f'(x)|}$$

   at the location $x = 2$. In the calculation use the stepsizes $\Delta x = 10^{-20} \ldots 10^{0}$. Use both single and double precision for the calculation, and present the results in a double logarithmic plot (ξ vs. Δx). In MATLAB double logarithmic plots are obtained by the function `loglog()`. Remember that all variables used here should be defined as double or single precision as in Task 1.

b) The general formula for the propagation error, for a function $h(X_j)$ with n variables X_j is given by:

$$\xi_h = \sum_{j=1}^{n} \left| \frac{X_j}{h} \frac{\partial h}{\partial X_j} \right| \varepsilon_{X_j} \,,$$

where ε_{X_j} is the accuracy on the quantity X_j. Based on this, show that the propagation error ξ_h, when adding two numbers X_1 and X_2, is given by:

$$\xi_h = \frac{|X_1|}{|X_1 + X_2|}\varepsilon_{X_1} + \frac{|X_2|}{|X_1 + X_2|}\varepsilon_{X_2}$$

where ε_{X_1} and ε_{X_2} are the corresponding accuracies of each number.

c) Show that the relative discretisation error of using equation (1) is given by:

$$\xi_d = \frac{\Delta x^2 |f'''(x)|}{6|f'(x)|}$$

(Hint: Taylor expansion)
and the propagation error is given by (round-off error):

$$\xi_r = \frac{\varepsilon \cdot |f(x)|}{\Delta x |f'(x)|}$$

(Hint: Use equation from part b)
with the machine accuracy ε from Task 1. Find the value of Δx that minimises the total error:

$$\xi_g = \xi_r + \xi_d$$

Plot the results for ξ_r, ξ_d, ξ_g together with the results from part a).

Task 3: Integration of Ordinary Differential Equation (ODE)

In this problem the stability and convergence order of some simple integration methods is examined. The following first order, ordinary, linear differential equation with constant coefficient is considered (Dahlquist equation)

$$\frac{\mathrm{d}u}{\mathrm{d}t} = A(u) = \lambda u \ , \qquad u(0) = 1$$

where $0 \leq t \leq T$ and $\lambda = \text{const.} \in \mathbb{C}$. The time interval $[0, T]$ is split into N parts with the same length Δt. The following integration methods should be used:

- explicit Euler method
$$u^{n+1} - u^n = \Delta t A(u^n)$$
- implicit Euler method
$$u^{n+1} - u^n = \Delta t A(u^{n+1})$$
- Crank-Nicolson method
$$u^{n+1} - u^n = \frac{1}{2}\Delta t(A(u^{n+1}) + A(u^n))$$

where $n = 0, ..., N$. Calculate until $T = 10$ and use the discretisation with $N = 20, 40, 50, 100, 200$ steps.

a) Derive the analytical solution u_{ex}.

b) For $\lambda = -\sqrt{3}/2 + \pi i$, calculate the numerical solution with the given discretisations and integration methods. Plot the real part of the analytical solution and the three numerical solutions for each value of N.

c) Discuss the usefulness and accuracy of the methods.

d) As $\lambda \to -\infty$ the problem becomes more and more stiff. Derive, for the three considered numerical schemes, the expression of the amplification factor, $G(z)$, where $z = \lambda \Delta t$. Which of the schemes provides a better approximation of the exact (analytical) amplification for one time step? Plot $G(z)$ for the three schemes and the analytical solution in the neighborhood of zero.

e) For $\lambda = -\sqrt{3}/2 + i$, do as in b) and calculate the numerical and analytical solutions. Show also the error $|u_{ex} - u_{num}|$ at time $t = 3$ as a function of N in a double logarithmic plot. Explain the differences between the methods.

Note (for all tasks): Together with your solutions, hand in the MATLAB-codes that you have written yourself.