Institutionen för datavetenskap
Umeå universitet

14 december 2009
Numeriska metoder för civilingenjörer

Review exercises for quiz and final exam

1 Theme 1

1. What is the definition of machine epsilon ϵ_M?
2. What is the largest error that can occur when rounding a real number x to a floating point number $fl(x)$ (assuming that over- or underflow does not occur)? Estimate the error in the floating point representation of the number π when $\epsilon_M = 2^{-52}$ (IEEE 754 double precision).
3. As part of an algorithm, there is a need to test whether $f(x) = a$, where $f(x)$ and a are real numbers. The values of $f(x)$ and a are stored as floating point numbers `f` and `a` in a Matlab implementation of the algorithm. How should the test be implemented?
4. Which error, the rounding error or the discretization error, usually dominates in scientific computations?
5. Name two cases in which rounding errors may have a large impact on the result of a computation that uses floating point numbers.
6. What is meant by underflow in floating point calculations? Give example of a calculation that yields `NaN` in the IEEE 754 standard.
7. Give an example of a calculation that may cause cancellation of significant digits.
8. Assume that we wish to approximate the derivative with the so-called central difference
$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h}.$$
What kind of errors dominate for different sizes of h?
9. Why does the divergent series $\sum_{k=1}^{\infty} 1/k$ has a finite sum in floating point arithmetic?

2 Theme 2

1. *Without* the use of matrix inverse (`inv`), write a Matlab expression that computes the vector
$$x = B^{-1}(2A + I)(C^{-1} + A)b,$$
where A, B, and C are given $n \times n$ matrices, with B and C nonsingular, and b is a given n vector.
2. Assume that the LU factorization of a matrix A of size 5000-by-5000 takes 11 s. Estimate how long time the backward and forward substitutions will take. (You may ignore delays due to startup and memory operations.)
3. When solving linear systems of equations $Ax = b$ in Matlab, the common recommendation is to write `x=A\b` and not `x=inv(A)*b`. Why?
4. What is the recommended strategy when solving a linear system with multiple right-hand sides?
5. Assume that the LU factorization of a matrix is known, that is, upper and lower triangular matrices L and U that satisfy $A = LU$ are known. How can this information be used in order to solve the linear system $A^T x = b$?

6. Does the conditioning of a linear system depend on the computer algorithm that is used to solve the problem?

7. Compute matrix norms $\|A\|_\infty$, $\|A\|_1$ for

$$A = \begin{pmatrix} -1.1 & 2 & 0.5 & 1 \\ 2 & 3 & -3 & 0.5 \\ 1.5 & 0.9 & 0 & 0.1 \\ 0.5 & 0.6 & -0.6 & 0.7 \end{pmatrix}.$$

   Explain how the matrix norm $\|A\|_2$ could be computed.

8. Classify following matrices as *well conditioned* (a low condition number) or *ill conditioned* (a high condition number):

$$\begin{pmatrix} 10^{10} & 0 \\ 0 & 10^{-10} \end{pmatrix}, \quad \begin{pmatrix} 10^{10} & 0 \\ 0 & 10^{10} \end{pmatrix}, \quad \begin{pmatrix} 10^{-10} & 0 \\ 0 & 10^{-10} \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}.$$

9. (a) Show that the matrix

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

   cannot be LU-factorized; that is, there are no matrices L and U (with the structure we have discussed in the course) such that $A = LU$.

   (b) What action does software for solution of linear system (such as Matlab) take in order to be able to factor matrices like A?

10. (a) Perform LU factorization, without pivoting, of the matrix

$$A = \begin{pmatrix} 1 & 0.5 & 1.5 & -1 \\ 2 & 3 & 2 & -2 \\ 0 & 2 & 1 & 0 \\ 0 & 4 & 2 & 2 \end{pmatrix}.$$

   (b) What problems can occur in a computer implementation of LU factorization if row pivoting is not performed?

11. (a) LU factorize the matrix

$$A = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 10 \end{pmatrix}$$

   without row pivoting. Use the L and U factors to solve the equation system $Ax = b$ with $b = (1,1,1)^T$.

   (b) Now assume that vector b above is not exactly known but rounded to 3 correct decimals (that is, the magnitude of the rounding error in each component of b is not larger than 5×10^{-4}). Moreover, it holds that $\|A^{-1}\|_\infty = 7$. How large will then the error in the computed solution be in the worst case?

3 Theme 3

1. What is meant by *linear convergence rate* for a method to solve nonlinear systems of equations?

2. Assume that the errors in an iterative method develop as follows:

   (a) 10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}

   (b) 10^{-3}, 10^{-5}, 10^{-7}, 10^{-9}

   What are corresponding convergence rates?

3. How many iterations does Newton's method require to solve a *linear* system of equations $\mathbf{f}(\mathbf{x}) = \mathbf{0}$? Motivate.

4. List one advantage of fixed-point iterations compared to Newton's method for solving nonlinear equations. List one advantage with Newton's method.

5. Write up Newton's method for the solution of the minimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}),$$

   where $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth function of n variables.

4 Theme 4

1. True or false: if a solution y to an initial value problem for an ODE satisfies $y(t) \to +\infty$ as $t \to +\infty$, then the ODE is necessarily unstable.

2. Classify the following ODEs as stable, asymptotically stable, or unstable with respect to initial values.

   (a) $y' - y = -t^2$, (b) $y' - y = t^2$, (c) $y' + y = t^2$; (d) $y' = t^2$.

3. Write following equations in the standard form for initial value problems.

   (a) Duffing's equation:
$$y'' + \delta y' + \sigma(y^3 - y) = \gamma \sin \omega t$$

   (b) Blasius' equation:
$$f''' + \frac{1}{2} f f'' = 0$$

   (c) Van der Pol's equation:
$$y'' - y'(1 - y^2) - y = 0.$$

4. Shortly explain what is meant by the following terminology in connection with the numerical solution of initial-value problems for ordinary differential equations:

   (a) *implicit* and *explicit* methods;

   (b) *truncation error*;

   (c) *order of accuracy*.

5. True or false: can a numerical method be unstable when it is applied to an ODE which is stable with respect to initial conditions?

6. Implicit numerical methods for the solution of initial-value problems for ordinary differential equations typically have a much larger range of time steps for which the method is stable compared to explicit methods. Why then are implicit methods not always used?

7. The initial value problem for the ordinary differential equation $y' = f(t, y)$ can be solved numerically using the so-called α scheme:

$$y_{k+1} = y_k + \Delta t\big[\alpha f(t_{k+1}, y_{k+1}) + (1-\alpha) f(t_k, y_k)\big],$$

where $0 \le \alpha \le 1$.

   (a) What methods are obtained when selecting $\alpha = 0$, 1, and 1/2, respectively?

   (b) Determine the order of accuracy for the scheme for all $0 \le \alpha \le 1$.

   (c) For all $0 \le \alpha \le 1$, determine the stability condition on the negative real axis, that is, the condition of stability for the scheme when applied to the equation $y' = \lambda y$ for $\lambda < 0$.

   (d) What changes have to be done to the α-scheme in order to obtain Heun's method?

8. The initial value problem for the ordinary differential equation $y' = f(t, y)$ can be numerically solved with the scheme

$$y_{k+1} = y_k + \frac{\Delta t}{2}\big[3 f(t_k, y_k) - f(t_{k-1}, y_{k-1})\big].$$

   (a) Is the method explicit or implicit?

   (b) Derive the order of accuracy for the method.

9. What is meant by a *stiff* system of ODEs? Why are implicit methods often recommended for stiff system of ODEs?

5 Theme 5

1. Let $(x_0, y_0), \ldots, (x_n, y_n)$ be a set of pair of numbers.

   (a) Give a condition under which the point sets can be interpolated by polynomials.

   (b) What problem can occur when interpolating arbitrary point sets with polynomials?

2. You have been given the task of providing a drawing of a church vault (*valv*) by interpolating a set of 25 measured x- and y-coordinates. Suggest a suitable computational technique to interpolate the points in order to plot the shape of the vault.

3. A function f is defined on the interval $[0, 1]$, but the function values are known only at the points in the table. The table suggests that the underlying function has a maximum within the interval. It is likely that the maximum is not exactly at one of the given points.

   (a) When interpolating a suitable subset of the points with a polynomial in order to estimate the exact location of the maximum, which points and which order of the polynomial is appropriate to use?

   (b) Carry out the interpolation and estimate the location of the maximum.

x	$f(x)$
0.0	0.1206
0.1	0.2506
0.2	0.4329
0.3	0.6532
0.4	0.8748
0.5	1.0441
0.6	1.1136
0.7	1.0656
0.8	0.9213
0.9	0.7295
1.0	0.5427

4. Derive a formula for numerical computation of the double integral

$$\int_0^1 \int_0^1 f(x,y)\,\mathrm{d}x\mathrm{d}y$$

through repeated use of the trapezoidal method with equidistant intervals first in one coordinate direction and then in the other.

5. A flight is planned between Stockholm and New York (distance 7800 km) using an airplane with a nominal cruising speed of 700 km/h. The effective speed (the speed relative to ground) varies depending on the wind direction. Using meteorological data, the effective speed has been estimated according to the table.

   The flight time for a flight from x_1 to x_2 is given by the integral

$$T = \int_{x_1}^{x_2} \frac{1}{v(x)}\,\mathrm{d}x$$

x (km)	v (km/h)
0	750
1300	680
2600	630
3900	640
5200	690
6500	760
7800	830

   Estimate the flight time by numerically computing the above integral.