Teknisk-vetenskapliga beräkningar
Exam

April 12th, 2012

Instructions: This exam consists of four major problems. The maximum score is 100 points or 25 points per problem. You are allowed to use anything which is either printed or written on paper prior to the exam. This includes lecture notes, your own notes, your mandatory projects and any textbook that you might care to reference. Moreover, you may use a programmable calculator. While the class was taught in English you may write your answers in Swedish or English.

Problem 1 This problem concerns the computation of $s^{1/3}$ where s is a real number.

1. (5pt) Explain why it is enough to solve the equation $f(x) = 0$ where

$$f(x) = s - x^3$$

and write down Newton's iteration for this problem.

Solution: We have

$$f(x) = 0 \Leftrightarrow x = \sqrt[3]{s}$$

where the bi-implication holds because the function $x \to x^3$ is monotone, hence has an inverse function $x \to \sqrt[3]{x}$. Given an initial guess x_0, Newton's iteration takes the form

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{s - x_n^3}{-3x_n^2} = x_n + \frac{s - x_n^3}{3x_n^2}$$

2. (10pt) Let $s \in [1, 8]$. Show that the initial guess

$$x_0 = x_0(s) = as + b, \quad a = \frac{1}{7}, \quad b = \frac{3}{7} + \frac{1}{3}\sqrt{\frac{7}{3}}$$

satisfies

$$|s^{1/3} - x_0| \leq \frac{1}{3}\sqrt{\frac{7}{3}} - \frac{3}{7} \approx 0.0806.$$

Solution: We define a function $\phi : [1, 8] \to \mathbb{R}$ as

$$\phi(x) = \sqrt[3]{s} - (as + b)$$

We have

$$\phi(1) = 1 - a - b = 1 - \frac{1}{7} - \frac{3}{7} - \frac{1}{3}\sqrt{\frac{7}{3}} = \frac{3}{7} - \frac{1}{3}\sqrt{\frac{7}{3}}$$

and

$$\phi(8) = 2 - 8a - b = 2 - \frac{8}{7} - \frac{3}{7} + \frac{1}{3}\sqrt{\frac{7}{3}} = \frac{3}{7} - \frac{1}{3}\sqrt{\frac{7}{3}}$$

Finally, we see that

$$\phi'(s) = \frac{1}{3}s^{-\frac{2}{3}} - a = 0 \Leftrightarrow s = s_0 = (3a)^{-\frac{3}{2}}$$

and at this particular point

$$\phi(s_0) = (3a)^{-\frac{1}{2}} - a(3a)^{-\frac{3}{2}} - b = -\left(\frac{3}{7} - \frac{1}{3}\sqrt{\frac{7}{3}}\right)$$

It follows that

$$|\phi(x)| \leq \left(\frac{3}{7} - \frac{1}{3}\sqrt{\frac{7}{3}}\right)$$

3. (10pt) Compute $(12)^{1/3}$ with a relative error of at most 10^{-6}.

   Solution: Incomplete! We have $12 = 8 \times \frac{3}{2}$. Therefore $\sqrt[3]{12} = 2 \times \sqrt[3]{\frac{3}{2}}$. You should pick $x_0 = x_0(\frac{3}{2})$ and then execute a few Newton iterations. The error can be judge by using x_n to generate an interval (a_n, b_n) which can be show to contain the root. Pick $a_n = x_n - \Delta$ and $b_n = x_n + \Delta$ and verify that $f(a_n)$ and $f(b_n)$ have different sign.

Problem 2 This problem concerns the fundamental differences between real arithmetic and floating point arithmetic. You will be investigating the function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} \frac{\sinh(x)-x}{x^3}, & x \neq 0, \\ \frac{1}{6}, & x = 0. \end{cases}$$

where $\sinh(x)$ is the hyperbolic sinus function given by

$$\sinh(x) = \frac{\exp(x) - \exp(-x)}{2}.$$

1. (5pt) Show that f is continuous for all $x \in \mathbb{R}$, including the special case of $x = 0$.

   Solution: For $x \neq 0$ there are no problems as f is built from functions which are known to be continuous using a finite set of basic arithmetic operations. In the case of $x = 0$ we proceed using l'Hospital's rule. With

   $$T(x) = \sinh(x) - x$$

   we have

   $$T'(x) = \cosh(x) - 1, T''(x) = \sinh(x), T^{(3)}(x) = \cosh(x)$$

   and

   $$N(x) = x^3 \Rightarrow N^{(3)}(x) = 6.$$

   Therefore

   $$\frac{T^{(3)}(x)}{N^{(3)}(x)} \to \frac{1}{6}, \quad x \to 0, \quad x \neq 0$$

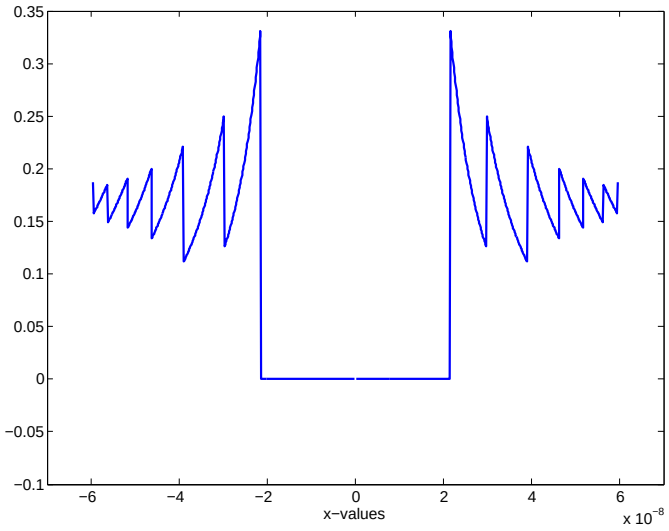
   Using l'Hospitals rule three times in a recursive manner, we conclude that

   $$f(x) = \frac{T(x)}{N(x)} \to \frac{1}{6} = f(0), \quad x \to 0, x \neq 0$$

   It follows that f is also continuous at the point $x = 0$.

2. (10pt) The following plot was generated with the MATLAB command

   ```
   >> k=-24; x=linspace(-2^-k,2^-k,1025); f=(sinh(x)-x)./x.^3; plot(x,f)
   ```



Explain the presence of the violent oscillations as well as the strange plateau in the interval $[-2, 2] \times 10^{-8}$.

Solution: The expression $d(x) = \sinh(x) - x$ will cancel catastropically for $x \approx 0$. Therefore, f is not computed reliably for small values of x. Eventually, when x is small enough the machine approximates $x \to \sinh(x)$ with the simple function $x \to x$, which is why we have a plateau around $x = 0$.

3. (10pt) Show that the correct double precision representation of f satisfies

$$\text{fl}(f(x)) = \text{fl}\left(\frac{1}{6}\right)$$

for all $[-1, 1] \times 10^{-8}$.

Solution: The Taylor series for sinh at the point $x_0 = 0$ is

$$g(x) = \sinh(x) = x + \frac{1}{3!}x^3 + \frac{1}{5!}x^5 \dots .$$

It follows that

$$f(x) = \frac{1}{3!} + \frac{g^{(5)}(\xi)}{5!}x^2$$

for some ξ between 0 and x. Now, let $0 < x < 10^{-8}$. Since $g(x) = \sinh(x)$ we have $g^{(5)}(\xi) = \cosh(\xi) \approx 1$. Finally,

$$3!\frac{g^{(5)}(\xi)}{5!}x^2 \leq 6 \times \frac{1}{5!}(10^{-8})^2 = \frac{1}{20} \times 10^{-16} < 2^{-53}$$

which show that the term $\frac{g^{(5)}(\xi)}{5!}x^2$ is so small, that

$$\mathrm{fl}(f(x)) = \mathrm{fl}\left(\frac{1}{3!} + \frac{g^{(5)}(\xi)}{5!}x^2\right) = \mathrm{fl}\left(\frac{1}{6}\right).$$

Problem 3 This problem concerns the numerical computation of the integral

$$I = \int_0^1 \exp(-x^2)dx.$$

In the table below the column `Sh` contains the Simpson sum S_h corresponding to the stepsize $h = \frac{1}{2N}$ and the column `fraction` contains the values of the usual fraction ν given by

$$\nu = \frac{S_{2h} - S_{4h}}{S_h - S_{2h}}$$

Here is the table of computed values

N	Sh	(Sh-S2h)/15	fraction
524288	0.7468241328124283	5.92e-17	-1.00000000
262144	0.7468241328124274	-5.92e-17	-1.25000000
131072	0.7468241328124283	7.40e-17	-0.50000000
65536	0.7468241328124272	-3.70e-17	-0.80000000
32768	0.7468241328124278	2.96e-17	-2.00000000
16384	0.7468241328124273	-5.92e-17	-0.62500000
8192	0.7468241328124282	3.70e-17	-0.20000000
4096	0.7468241328124277	-7.40e-18	-7.00000000
2048	0.7468241328124278	5.18e-17	-10.42857143
1024	0.7468241328124270	-5.40e-16	13.71232877
512	0.7468241328124351	-7.41e-15	16.05494505
256	0.7468241328125462	-1.19e-13	16.00161782
128	0.7468241328143305	-1.90e-12	15.99929616
64	0.7468241328428812	-3.05e-11	15.99704551
32	0.7468241332996726	-4.87e-10	15.98792233
16	0.7468241406069851	-7.79e-09	15.94720317
8	0.7468242574357303	-1.24e-07	15.70468214
4	0.7468261205274666	-1.95e-06	11.10927205
2	0.7468553797909874	-2.17e-05	0.00000000
1	0.7471804289095103	0.00e+00	0.00000000

1. (5pt) Explain why the fraction ν misbehaves for large values of N.

   Solution: In exact arithmetic we have $S_h \to I$ as $h \to 0$. Therefore the real numbers S_h, S_{2h} and S_{4h} are all very close. As a result we experiences catastropic cancellation in both the nominator and the denominater when we attempt to calculate the fraction. Therefore, it is extremely unlikely that the fraction will be close to 16, when h is suitably small.

2. (8pt) Determine the range of values of N for which you can trust the error estimate, but remember to justify your choice.

   Solution: In exact arithmetic the fractions will converge monotonically to 16. This behavior is seen from $k = 4$ to $k = 128$. The value at $k = 256$ has jumped to the other side of 16 indicating that this is the point where the rounding errors are starting to become important. I would trust the sign, the magnitude and the first couple of digits of the error estimates for $k = 8$ to $k = 128$ were the fractions are not only close to 16 but converging monotonically to 16.

3. (12pt) Compute the integral I with a relative error which is less than 10^{-12}.

   Solution: We choose the value corresponding to $k = 128$ and compute

$$A_h = S_h + \frac{S_h - S_{2h}}{15} = 0.746824132812427$$

   a value with carries an error which is $O(h^6)$ because we killed the $O(h^4)$ term. Similarly we have

$$A_{2h} = S_{2h} + \frac{S_{2h} - S_{4h}}{15} = 0.746824132812428$$

   Our new error estimate is $\frac{A_h - A_{2h}}{2^6 - 1} \approx -2.1 \times 10^{-17}$. It is easy to see that A_h approximates the integral with a relative error which is much less than the desired tolerance.

Problem 4 Let A be the matrix given by

$$A = \begin{bmatrix} 4 & 2 & 0 & 0 \\ 2 & 5 & 2 & 0 \\ 0 & 2 & 5 & 2 \\ 0 & 0 & 2 & 5 \end{bmatrix}$$

1. (8pt) Compute an LU factorization of A.

   Solution: We have

   $$A = LU \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & 1 & 0 & 0 \\ 0 & \frac{1}{2} & 1 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} 4 & 2 & 0 & 0 \\ 0 & 4 & 2 & 0 \\ 0 & 0 & 4 & 2 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

   but a student would need to do Gaussian elimination by hand.

2. (7pt) Given that A^{-1} satisfies

   $$A^{-1} = \frac{1}{256} \begin{bmatrix} 85 & -42 & 20 & -8 \\ -42 & 84 & -40 & 16 \\ 20 & -40 & 80 & -32 \\ -8 & 16 & -32 & 64 \end{bmatrix}$$

   compute the condition number of A with respect to the infinity norm.

   Solution: We have $\|A\|_\infty = 9$ and $\|A^{-1}\|_\infty = \frac{182}{256}$. The condition number is

   $$\kappa_\infty(A) = \|A\|_\infty \|A^{-1}\|_\infty = \frac{1638}{256} = \frac{819}{128}$$

3. (10pt) Find the floating point representation of the solution of the linear system $Ax = f$, where $f = (1, 1, 1, 1)^T$.

   Solution: We are literally handed the inverse matrix, so the exact solution is given by

   $$x = A^{-1} f = \frac{1}{256} \begin{bmatrix} 55 \\ 18 \\ 28 \\ 40 \end{bmatrix}$$

   which is vector of machine numbers in both single and double precision.