Scientific Computation Comprehensive Examination

BY FALL 2016

Answer 5 questions of your choice explaining all steps that lead to a solution. Partial credit will be awarded for presenting a viable solution strategy. No credit will be given to computations presented without motivation.

I. Find a solution of the system

$$\begin{cases} \sin(x)+\cos(y)+\exp(x\,y)=1.5, \\ \arctan(x+y)-x\,y=0 \end{cases} \tag{1}$$

to two significant digits of accuracy.

Solution. This is a nonlinear system of form $F(X)=0$. Given an initial approximation $X_0=(x_0,y_0)^T$ close to the solution, Newton's method

$$F'(X_n)(X_{n+1}-X_n)=-F(X_n),$$

$$X_n=\begin{pmatrix} x_n \\ y_n \end{pmatrix}, F(X)=\begin{pmatrix} f(x,y) \\ g(x,y) \end{pmatrix}=\begin{pmatrix} \sin(x)+\cos(y)+\exp(x\,y)-1.5 \\ \arctan(x+y)-x\,y \end{pmatrix},$$

$$F'(X)=\begin{pmatrix} \cos(x)+y\exp(x\,y) & -\sin(y)+x\exp(x\,y) \\ \frac{1}{1+(x+y)^2}-y & \frac{1}{1+(x+y)^2}-x \end{pmatrix},$$

converges to the solution quadratically, $e_{n+1}\leqslant Ce_n^2$, with $e_n=\|X_n-X\|$, $C\sim\|F'\|/\|F''\|$, X system solution. Since X is unknown, use fact that $\mathbb{R}^2$ is complete to state quadratic convergence as $\varepsilon_{n+1}\leqslant C\varepsilon_n^2$, $\varepsilon_n=\|X_{n+1}-X_n\|$. The imposed accuracy of two significant digits would be obtained upon a Newton iteration with initial error $\varepsilon_0\cong 0.1/\sqrt{C}$. Hence the main task is to find an initial approximation of the desired accuracy, and estimate C. Additionally, evaluation of the transcendental functions f, g requires construction of readily hand-computable approximants. The simplest to evaluate initial approximation is $X_0=(0,0)^T$, which leads to

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}X_1=-\begin{pmatrix} 0.5 \\ 0 \end{pmatrix}\Rightarrow X_1=\begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix}\Rightarrow \varepsilon_1=\sqrt{2}, C\sim\mathcal{O}(1),$$

implying $\varepsilon_2>0.01$, hence a better initial approximation is needed.

(*Note: statement of the above algorithm, analysis, recognition of need for transcendental function approximation, likely excessive computation from starting point (0,0) would result in 8/10 score for this problem, remaining 2/10 for obtaining a better initial approximant)*

To find a good initial approximation, note that solution of system (1) corresponds to intersection of curves (α,β) defined implicitly by

$$\begin{array}{ll} \alpha\colon & f(x,y)=\sin(x)+\cos(y)+\exp(x\,y)-1.5=0 \\ \beta\colon & g(x,y)=\arctan(x+y)-x\,y=0 \end{array}.$$

Use the implicit function theorem to find the slopes of the α, β curves as

$$\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)_\alpha = -\left(\frac{\partial f}{\partial x}\right)/\left(\frac{\partial f}{\partial y}\right) = \frac{\cos(x)+y\exp(xy)}{\sin(y)-x\exp(xy)} = s(x,y), \tag{2}$$

$$\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)_\beta = -\left(\frac{\partial g}{\partial x}\right)/\left(\frac{\partial g}{\partial y}\right) = -\frac{1-y-y(x+y)^2}{1-x-x(x+y)^2} = t(x,y). \tag{3}$$

Note that $g(0,0)=0$, hence the solution to (1) is a point on the curve defined by the initial value problem (IVP)

$$\beta: \begin{cases} y' = t(x,y) \\ y(x=0) = 0 \end{cases}, \tag{4}$$

with slope at $(x,y)=(0,0)$ given by $y'_\beta(0)=t(0,0)=-1$. The corresponding (IVP) for the α curve is

$$\alpha: \begin{cases} y' = s(x,y) \\ y(x=0) = \eta \end{cases},$$

with η the solution of $f(0,\eta)=\cos\eta-0.5=0 \Rightarrow \eta=\pi/3\cong 1.047$. The slope of the α curve at $(x,y)=(0,0)$ is $y'_\alpha(0)=s(0,\pi/3)=(1+\pi/3)/\left(\sqrt{3}/2\right)\cong 2.364$. These linear approximants

$$p_1(x)=2.364x+1.047,\ q_1(x)=-x,$$

of the integral curves (α,β) intersect at $(x_0,y_0)=(-0.31,0.31)$. The qualitative behavior of the curves is shown in Fig. 1.

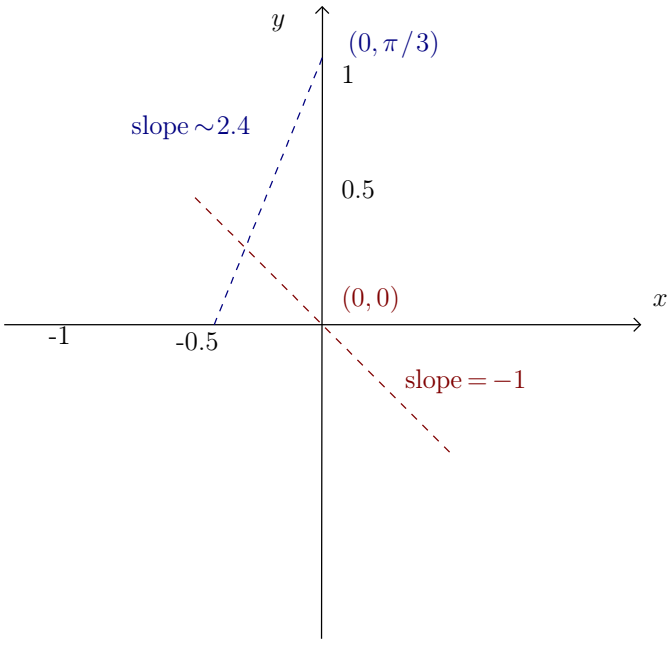

Figure 1.

Use this as an initial approximation for Newton's method

$$F'(X_0)(X_1 - X_0) = -F(X_0)$$

and evaluate

$$F(X_0) = \begin{pmatrix} -S + C + \exp(P) - 1.5 \\ -P \end{pmatrix}, F'(X) = \begin{pmatrix} C + y_0 \exp(P) & -S + x_0 \exp(P) \\ 1 - y_0 & 1 - x_0 \end{pmatrix}$$

through series expansions

$$\begin{aligned} P &= -0.31^2 = -0.0961 \\ \exp(P) &\cong 1 + P + P^2/2 \cong 0.909 \\ S &= \sin(0.31) \cong 0.31 - \frac{0.31^3}{3!} \cong 0.305 \\ C &= \cos(0.31) \cong 1 - \frac{0.31^2}{2!} \cong 0.952 \end{aligned}$$

to give

$$\begin{pmatrix} 1.234 & -0.587 \\ 0.69 & 1.31 \end{pmatrix} \begin{pmatrix} x_1 + 0.31 \\ y_1 - 0.31 \end{pmatrix} = \begin{pmatrix} 0.055 \\ 0.0961 \end{pmatrix}.$$

Solving gives $(x_1, y_1) = (-0.318, 0.387)$, hence $\varepsilon_0 \cong 0.08$, with $C \sim \mathcal{O}(1)$, and one more Newton iterate (not computed due to time constraint) would give desired precision.

II. Find the best approximation in the inf-norm on interval [-1,1] of $\cosh(x) = \frac{1}{2}(e^x + e^{-x})$ by a quadratic polynomial.

Solution. Since $\cosh(x)$ is an even function, the problem can be stated as

$$\min_{a,b} \max_{0 \leqslant x \leqslant 1} |\varepsilon(x)|, \varepsilon(x) = a + bx^2 - \cosh(x).$$

Extrema of $\varepsilon: [0,1] \to \mathbb{R}$ are either endpoint values $\varepsilon(0)$,$\varepsilon(1)$ (since $[0,1]$ is compact), or local maxima/minima $\varepsilon(t)$, with t solutions within (0,1) of

$$\varepsilon'(x) = 2bx - \sinh(x) = 0.$$

The possible extrema are therefore

$$y_1(a) = \varepsilon(0) = a - 1, y_2(a,b) = \varepsilon(1) = a + b - \cosh(1),$$

$$y_3(a,b) = \varepsilon(t) = a + bt^2 - \cosh t, 2bt - \sinh(t) = 0.$$

From the Chebyshev alternating theorem (equioscillation theorem)

$$y_1(a) = -y_3(a,b) = y_2(a,b),$$

and the best inf-norm approximant is $b = \cosh(1) - 1$, and a from solution of system

$$\begin{cases} 2a = 1 - bt^2 + \cosh t \\ 2bt = \sinh(t) \\ b = \cosh(1) - 1 \end{cases}.$$

III. For a function $f(x), x \in [0,1]$, consider the composite midpoint rule for computing

$$I(f) = \int_0^1 f(x)\, d\, x \approx h \sum_{i=1}^{n} f\left((i - \frac{1}{2})\, h\right) = Q(f, n)$$

where $h = \frac{1}{n}$,
(1) Suppose that $f \in C^{\infty}[0,1]$, prove that

$$I(f) - Q(f, n) = a_1\, h^2 + O(h^3). \tag{5}$$

Solution. Write

$$I(f) = \sum_{i=1}^{n} \int_{(i-1)h}^{ih} f(x)\, d\, x,$$

Taylor series expand in each subinterval,

$$I(f) = \sum_{i=1}^{n} \int_{(i-1)h}^{ih} \left\{ f_{i-1/2} + f'_{i-1/2} \cdot \left[x - (i - \frac{1}{2})\, h \right] + \frac{1}{2} f''_{i-1/2} \cdot \left[x - (i - \frac{1}{2})\, h \right]^2 + ... \right\} d\, x,$$

compute the individual integrals of odd/even functions

$$\int_{(i-1)h}^{ih} \left[x - (i - \frac{1}{2})\, h \right]^{2k+1} \mathrm{d}x = 0, \int_{(i-1)h}^{ih} \left[x - (i - \frac{1}{2})\, h \right]^{2k} \mathrm{d}x = \mathcal{O}(h^{2k+1}),$$

to obtain

$$I(f) = \sum_{i=1}^{n} \left\{ h\, f_{i-1/2} + \mathcal{O}(h^3) \right\} \cong Q(f, n) + n\mathcal{O}(h^3) + n\mathcal{O}(h^5),$$

with $n = \mathcal{O}(1/h)$.

This verifies (5), which could also be written as the tighter bounds

$$I(f) - Q(f, n) = a_1 h^2 + a_2 h^4 + ... = a_1 h^2 + \mathcal{O}(h^4) = a_1 h^2 + o(h^3).$$

(2) Consider $f(x) = \frac{1}{x^{\alpha}}$, with $0 < \alpha < 1$, notice that there is a singularity at $x = 0$. Could you find β in the formula

$$I(f) - Q(f, n) = c\, h^{\beta}\, ?$$

Solution. Previous estimate holds except for subinterval $[0, h]$ on which the exact integral is

$$I_0 = \int_0^h f(x)\, \mathrm{d}x = \int_0^h \frac{\mathrm{d}x}{x^{\alpha}} = \frac{h^{1+\alpha}}{1+\alpha},$$

approximated by

$$Q_0 = h\, f_{1/2} = \frac{h}{(h/2)^\alpha} = 2^\alpha h^{1-\alpha},$$

leading to error

$$e_0 = I_0 - Q_0 = \frac{h^{1+\alpha}}{1+\alpha} - 2^\alpha h^{1-\alpha} = \mathcal{O}(h^{1-\alpha}),$$

hence $\beta = 1-\alpha$.

(3) ("midpoint rule" with end-point corrections) Now consider $f(x) = \frac{1}{\sqrt{x}}\, g(x)$, describe how to modify the midpoint rule and get higher order accuracy by adding a "local correction", i.e., by changing the weight for the function value $f(\frac{h}{2}) = \frac{g(\frac{h}{2})}{\sqrt{\frac{h}{2}}}$.

Solution. Again, initial estimates holds (assuming $g \in C^\infty[0,\,1]$) except for subinterval $[0, h]$. Apply mean-value theorem

$$I_0 = \int_0^h \frac{g(x)}{\sqrt{x}}\, \mathrm{d}x = 2\sqrt{h}\; g(\xi),$$

and the local approximant is

$$Q_0 = \sqrt{2h}\, w g(\frac{h}{2}).$$

Since

$$g(\xi) = g\bigg(\frac{h}{2}\bigg) + g'\bigg(\frac{h}{2}\bigg)\frac{h}{2} + \mathcal{O}(h^2),$$

the error becomes

$$I_0 - Q_0 = 2\sqrt{h}\bigg[\, g\bigg(\frac{h}{2}\bigg) + g'\bigg(\frac{h}{2}\bigg)\frac{h}{2} + \mathcal{O}(h^2)\,\bigg] - \sqrt{2h}\, w g(\frac{h}{2})$$

$$I_0 - Q_0 = g\bigg(\frac{h}{2}\bigg)\big[\,2\sqrt{h} - \sqrt{2h}\, w\,\big] + \mathcal{O}(h^{5/2}),$$

and choosing $w = \sqrt{2}$ increases accuracy to $\mathcal{O}(h^{5/2})$ from $\mathcal{O}(h^{1/2})$.

IV. Construct the third order explicit Runge-Kutta formula that approximates solutions of the ordinary differential equation $y'(t) = f(t,\, y)$ and uses evaluations of f at intermediate steps $t, t + h/2, t + 3h/4$.

Solution. The formula is of form

$$\begin{aligned}
&y_{n+1} = y_n + h(w_1 K_1 + w_2 K_2 + w_3 K_3),\\
&K_1 = f(t_n,\, y_n),\\
&K_2 = f(t_n + h/2,\, y_n + \alpha_{21} h K_1),\\
&K_3 = f(t_n + 3h/4,\, y_n + \alpha_{31} h K_1 + \alpha_{32} h K_2).
\end{aligned}$$

The unknown weights w_1, w_2, w_3 and coefficients α_{11}, α_{21}, α_{22} are determined by matching Taylor series expansion up to 3rd order.

$$
\begin{aligned}
&y_{n+1} = y_n + y_n' h + \tfrac{1}{2} y_n'' h^2 + \tfrac{1}{6} y_n''' h^3 + \mathcal{O}(h^4) \\
&y_n' = f \\
&y_n'' = f_t + f_y f \\
&y_n''' = f_{tt} + f_{ty} f + f_y (f_t + f_y f) + (f_{yt} + f_{yy} f) f \\
&= f_{tt} + 2 f f_{ty} + f_y f_t + f_y^2 f + f_{yy} f^2 \\
&K_1 = f \\
&K_2 = f + \tfrac{h}{2} f_t + \alpha_{21} h K_1 f_y + \tfrac{h^2}{8} f_{tt} + \tfrac{1}{2} \alpha_{21} h^2 K_1 f_{ty} + \tfrac{1}{2} (\alpha_{21} h K_1)^2 f_{yy} + \mathcal{O}(h^3) \\
&K_3 = f + \tfrac{3h}{4} f_t + (\alpha_{31} K_1 + \alpha_{32} K_2) h f_y + \tfrac{9h^2}{32} f_{tt} + \tfrac{3h^2}{4} (\alpha_{31} K_1 + \alpha_{32} K_2) f_{ty} + \\
&\tfrac{h^2}{2} (\alpha_{31} K_1 + \alpha_{22} K_2)^2 f_{yy} + \mathcal{O}(h^3)
\end{aligned}
$$

with $f \equiv f(t_n, y_n)$, $f_t = \partial_t f(t_n, y_n)$, ... Identification of coefficients of h^p, $p = 0, 1, 2, 3$ and various combinations of f, f_t, f_y, ... yields

$$
\begin{array}{lllll}
p=0: & y_n: & & 1 = 1 \\
p=1: & f: & \text{(eq1)} & 1 = w_1 + w_2 + w_3 \\
p=2: & f_t: & \text{(eq2)} & \frac{1}{2} = \frac{1}{2} w_2 + \frac{3}{4} w_3 \\
 & f f_y: & \text{(eq3)} & \frac{1}{2} = \alpha_{21} w_2 + (\alpha_{31} + \alpha_{32}) w_3 \\
p=3: & f_{tt}: & \text{(eq4)} & \frac{1}{6} = \frac{1}{8} w_2 + \frac{9}{32} w_3 \\
 & f f_{ty}: & \text{(eq5)} & \frac{1}{3} = \frac{1}{2} \alpha_{21} w_2 + \frac{3}{4} (\alpha_{31} + \alpha_{32}) w_3 \\
 & f_t f_y: & \text{(eq6)} & \frac{1}{6} = \frac{1}{2} \alpha_{32} w_3 \\
 & f f_y^2: & \text{(eq7)} & \frac{1}{6} = \alpha_{21} \alpha_{32} w_3 \\
 & f^2 f_{yy}: & \text{(eq8)} & \frac{1}{6} = \frac{1}{2} \alpha_{21}^2 w_2 + \frac{1}{2} (\alpha_{31} + \alpha_{22})^2 w_3
\end{array}
$$

There are 8 nonlinear equations for 6 unknowns w_1, w_2, w_3, α_{21}, α_{31}, α_{32}. From (eq6,eq7), $\alpha_{21} = 1/2$. Solving the linear subsystem (eq1,2,4) gives

$$
w_1 = \frac{2}{9}, w_2 = \frac{3}{9}, w_3 = \frac{4}{9}.
$$

Replacing in above gives

$$
\begin{aligned}
\alpha_{31} + \alpha_{32} &= \frac{3}{4}, \\
\alpha_{31} + \alpha_{32} &= \frac{3}{4}, \\
(\alpha_{31} + \alpha_{32})^2 &= \frac{9}{16},
\end{aligned}
$$

showing a 1-parameter family of 3rd-order Runge-Kutta methods that use evaluations at t, $t + h/2, t + 3h/4$. One efficient choice is $\alpha_{31} = 0$, $\alpha_{32} = 3/4$.

V. Suppose the $n \times n$ non-singular matrix A is factored as $A = L\ H$ where L is lower triangular with ones on its diagonal and H is upper Hessenberg (i.e., all elements $h_{i,j} = 0$ if $j < i-1$). Design an efficient algorithm to compute the LU decomposition of A from L and H. Discuss the number of operations and memory usage of your algorithm. You may assume pivoting is not necessary.

Solution. Rewrite the factorization as $A = LG_{n-1}^{-1}...G_2^{-1}G_1^{-1}G_1G_2...G_{n-1}H$ with G_i the Gauss multiplier matrix that eliminates element $h_{j+1,j}$ of the Hessenberg form

$$G_i = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & 0 & & & \text{row } i \\ & & -g_i & 1 & & & \text{row } i+1 \\ & & & & \ddots & & \\ & & & & & 1 & \end{pmatrix}, g_i = \frac{h_{i+1,i}}{h_{i,i}}, G_i^{-1} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & 0 & & & \\ & & g_i & 1 & & & \\ & & & & \ddots & & \\ & & & & & 1 & \end{pmatrix}.$$

The desired LU decomposition is $A = TU$, $T = LG_{n-1}^{-1}...G_1^{-1}$, $U = G_1...G_{n-1}H$. No additional memory is required (overwriting L, H) and the number of operations is $4 \times (1+2+...+n-1) = \mathcal{O}(2n^2)$ (1 flop = 1 $\times$ or 1 +).

Algorithm

for $i = 1$ to $n-1$
  $g = h_{i+1,i}/h_{i,i}$
  for $j = i$ to n
    $h_{i+1,j} = h_{i+1,j} - gh_{i,j}$
    $l_{j,i} = l_{j,i} + gl_{j,i+1}$

VI. Using the singular value decomposition, one can determine the numerical rank of a matrix by studying the singular values and can also approximate the original matrix by a lower rank matrix to a prescribed accuracy requirement. Describe an algorithm based on the modified Gram-Schmidt (MGS) scheme and proper pivoting technique (permutation matrix), so that the modified QR algorithm is also rank revealing (i.e., the diagonal entries of the R matrix play the same role as the singular values).

Solution. The modified Gram-Schmidt algorithm (below) without permutations for $A = (a_1...a_n) \in \mathbb{R}^{m \times n}$ fails if $r_{ii} = 0$ indicating $a_i \in \text{span}\{a_1, ..., a_{i-1}\}$.

Algorithm

for $i = 1{:}\, n$
  $r_{ii} = \|a_i\|;\ q_i = a_i/r_{ii}$
  for $j = 2{:}\, n$
    $r_{ij} = a_i^T q_j;\ a_j = a_j - r_{ij}q_i$

To obtain behavior similar to the SVD, the columns of A should be permuted such that at each stage r_{ii} is maximal.

Algorithm

for $i = 1{:}\, n$
  for $l = i + 1{:}\, n$
    if $\|a_l\| > \|a_i\|$: swap(a_l, a_i)
  $r_{ii} = \|a_i\|$
  if $|r_{ii}| \leqslant \epsilon$: 'Numerical rank=',$i - 1$; exit
  $q_i = a_i / r_{ii}$
  for $j = 2{:}\, n$
    $r_{ij} = a_i^T q_j$; $a_j = a_j - r_{ij} q_i$