

Examination

January 29th, 2013

Duration of examination: 180 minutes.

Total points: 180

Problem 1 Sparse matrix

In this problem we use the sparse matrix

```
A = delsq(numgrid('C',n));
```

(1a) [10 points] Write a Matlab script

```
function fillin()
```

which measures the operator complexity (sum of the nonzeros of all factors of a given factorization divided by the number of nonzeros of A) of different decompositions, and plots them for $n = 2^1, \dots, 2^8$. Measure the operator complexity of the following things:

- Cholesky decomposition of A
- Incomplete Cholesky decomposition of A (see `ichol`)
- propose and implement an alternative way to reduce the fill in of the Cholesky decomposition.

Is it possible to measure an operator complexity number below 1? Explain why.

HINT: You may use the Matlab function `nnz`, and `ichol`.

Solution: Yes, a operator complexity number below 1 is possible, since the matrix **A** is symmetric with a lower and an upper triangular part. The Cholesky factor is only the lower triangular part. The operator complexity of the Cholesky factorization may therefore be as small as 0.5. This happens if the matrix is dense.

You may use reordering. The ones presented in the lecture are `symrcm`, `amd`, `symamd` and `colamd`. (e.g. `ilu` is not a good solution)

($n = 2^2, \dots, 2^8$ is OK as well)

Listing 1 and Figure 1 (The solution is with `cholinc`, as `ichol` did not work with my matlab version.)

Listing 1: `fillin()`

```
1 function fillin()
2 clear all; close all;
3 k = 2.^[1:8];
4
5 fill = inf*ones(length(k), 3);
6 n = inf*ones(length(k), 1);
7
8 for l = 1:length(k)
9
10     A = delsq(numgrid('C',k(l)));
11     n(l) = size(A,1);
```

```

12
13 % Fill in measurements
14 fill(1,1) = nnz(A);
15 fill(1,2) = nnz(chol(A));
16 fill(1,3) = nnz(cholinc(A,10^-1));
17 p = amd(A);
18 fill(1,4) = nnz(chol(A(p,p)));
19 end
20
21 loglog(n, fill(:,1:4))
22 legend('A', 'chol', 'cholinc', 'amd')
23
24 print -deps2c 'fillin.eps'

```

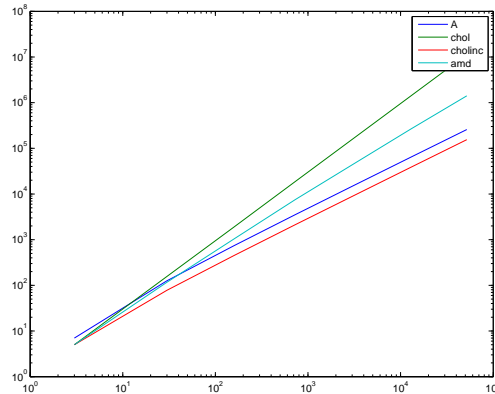


Figure 1: Convergence plot

(1b) [15 points] Plot the runtime (time to compute the solution \mathbf{x} , e.g. factorization and backward/forward substitution) and the measured error (norm of the residual) for $n = 2^1, \dots, 2^8$ of the three different approaches (of (1a)) to solve $\mathbf{Ax} = \mathbf{b}$. Use a random vector \mathbf{b} . Implement this functionality in the Matlab script

```
function conf()
```

Comment on the obtained results.

HINT: You may use `tic`, `toc`, `rand`, `norm` and `sort`.

HINT: Read (1c) before starting this sub problem

HINT: As you measure the runtime call `maxNumCompThreads(1)` to insure that Matlab is running in single core mode.

Solution: Together with (1c)

(1c) [10 points] Propose and implement (in (1b)) an alternative method which takes advantage of the sparsity of \mathbf{A} , and solves $\mathbf{Ax} = \mathbf{b}$ not exactly but with a residual norm of 10^{-3} .

Comment on the obtained results (may be together with (1b)).

Solution: You should use CG or preconditioned CG.

The runtime of `amd` is the best option, with short runtime and small error. The error for `ichol` is large as expected since `C` in this case is only an approximate Cholesky factor of \mathbf{A} and hence not useful. The runtime of `chol` seems to grow stronger than `pgc`.

($n = 2^2, \dots, 2^8$ is OK as well)

Listing 2, Figure 3 and Figure 2

```
Listing 2: conf()
```

```

1 function conf()
2 clear all; close all;
3 %%%%%%%% Prevent Matlab from using multiple cores %%%%%%%%%
4 maxNumCompThreads(1)
5
6 MyTol = 10-3;
7 k = 2.^[2:7];
8 t = inf*ones(length(k), 5);
9 err = inf*ones(length(k), 5);
10 n = inf*ones(length(k), 1);
11
12 for l = 1:length(k)
13   A = delsq(numgrid('C',k(l)));
14   n(l) = size(A,1);
15   b = rand(size(A,1),1);
16   MaxIt = min(1000,size(A,1));
17
18   for j=1:3
19     % time + error for cg:
20     tic;
21     x = pcg(A,b,MyTol,MaxIt);
22     t(l, 1) = min(t(l, 1), toc);
23     err(l, 1) = norm(A*x-b);
24
25     % time + error for pcg:
26     tic;
27     M = cholinc(A,10-1);
28     x = pcg(A,b,MyTol,MaxIt,M,M);
29     t(l, 2) = min(t(l, 2), toc);
30     err(l, 2) = norm(A*x-b);
31
32     % time + error for chol:
33     tic;
34     C = chol(A);
35     x = C\(C'\b);
36     t(l, 3) = min(t(l, 3), toc);
37     err(l, 3) = norm(A*x-b);
38
39     % time + error for cholinc:
40     tic;
41     C = cholinc(A,10-1);
42     x = C\(C'\b);
43     t(l, 4) = min(t(l, 3), toc);
44     err(l, 4) = norm(A*x-b);
45
46     % time + error for amd:
47     tic;
48     p = amd(A);
49     C = chol(A(p,p));
50     x = C\(C'\b(p));
51     [temp1, p1] = sort(p);
52     x = x(p1);
53     t(l, 5) = min(t(l, 3), toc);
54     err(l, 5) = norm(A*x-b);
55
56   end
57 end
58
59 loglog(n, t(:,1:5), '-*', n, 1e-6*n, 'k-', n, 1e-8*n.^2, 'k—')
60 legend('cg', 'pcg', 'chol', 'cholinc', 'amd')
61
62 print -deps2c 'time.eps'

```

```

63 figure
64
65 loglog(n, err(:,1:5), '-*')
66 legend('cg', 'pcg', 'chol', 'cholinc', 'amd')
67 title('error')
68
69
70 print -deps2c 'conf.eps'

```

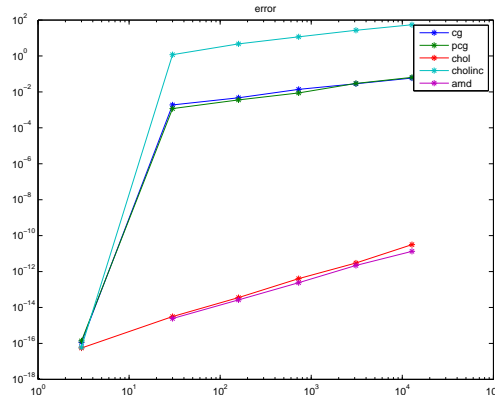


Figure 2: Error plot

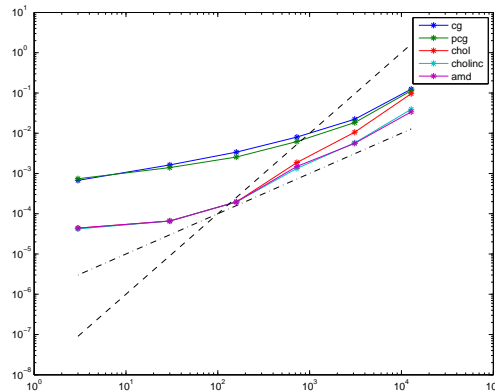


Figure 3: Convergence plot

Problem 2 Linear least squares

Let two vectors $\mathbf{z}, \mathbf{c} \in \mathbb{R}^n$, $n \in \mathbb{N}$ of measured data be given. The two numbers α^* and β^* are defined as

$$(\alpha^*, \beta^*) = \operatorname{argmin}_{\alpha, \beta \in \mathbb{R}} \|\mathbf{T}_{\alpha, \beta} \mathbf{z} - \mathbf{c}\|_2, \quad (1)$$

with the tridiagonal matrix

$$\mathbf{T}_{\alpha, \beta} = \begin{pmatrix} \alpha & \beta & 0 & \dots & 0 \\ \beta & \alpha & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \alpha & \beta \\ 0 & \dots & 0 & \beta & \alpha \end{pmatrix} \in \mathbb{R}^{n, n}.$$

(2a) [8 points] Reformulate (1) as a linear least squares problem in the usual form

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^k} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2$$

with suitable $\mathbf{A} \in \mathbb{R}^{m,k}$, $\mathbf{b} \in \mathbb{R}^m$, $m, k \in \mathbb{N}$.

Solution: The vector whose norm has to be minimized can be written as

$$\mathbf{T}_{\alpha,\beta}\mathbf{z} - \mathbf{c} \stackrel{(1)}{=} \begin{pmatrix} \alpha z_1 + \beta z_2 - c_1 \\ \alpha z_2 + \beta(z_1 + z_3) - c_2 \\ \dots \\ \alpha z_{n-1} + \beta(z_{n-2} + z_n) - c_{n-1} \\ \alpha z_n + \beta z_{n-1} - c_n \end{pmatrix} \stackrel{(2)}{=} \begin{pmatrix} z_1 & z_2 \\ z_2 & z_1 + z_3 \\ \vdots & \vdots \\ z_{n-1} & z_{n-2} + z_n \\ z_n & z_{n-1} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - \mathbf{c} \stackrel{(3)}{=} \mathbf{Ax} - \mathbf{b}.$$

In this way, we have moved the unknown α and β into the vector \mathbf{x} , while the data \mathbf{z} and \mathbf{c} are moved into the matrix \mathbf{A} and the hand side vector $\mathbf{b} = \mathbf{c}$ (that is not affected by the transformation). The number of unknown k is equal to 2 and the number of equations m is equal to n .

(2b) [13 points] Write a Matlab function

```
[alpha, beta] = lsqest(z,c)
```

that computes the values of the optimal parameter α^* and β^* according to (1) from the data vectors \mathbf{z} and \mathbf{c} (i.e., \mathbf{z} and \mathbf{c}). Use the QR-decomposition to solve the linear least squares problem.

HINT: For $\mathbf{z} = (1, 2, \dots, 10)^T$ and $\mathbf{c} = (10, 9, \dots, 1)^T$ you should get $\alpha^* \approx -0.4211$ and $\beta^* \approx 0.5789$.

Solution:

Listing 3: Linear least squares for the problem (1).

```
1 function lsqest
2 z = (1:10)';
3 c = (10:-1:1)';
4 [alpha, beta] = lsqest(z,c)
5
6 % z and c are column vectors
7 function [alpha, beta] = lsqest(z,c)
8 A = [z, [z(2:end);0] + [0;z(1:end-1)]];
9 [Q,R] = qr(A);
10 c = Q'*c;
11 v = R\c;
12 alpha = v(1);
13 beta = v(2);
```

Problem 3 Bézier semi-circle

(3a) [12 points] Write a Matlab function

```
function plot_bezcurv(d)
```

which draws the Bézier curve, the control points \mathbf{d} , and the convex hull defined by the control points \mathbf{d} ($2 \times n$ matrix).

HINT: You may use `convhull` and `fill`

Solution: Listing 4

Listing 4: `plot_bezcurv()`

```
1 function plot_bezcurv(d)
2 n = size(d,2)-1; %number of nodes
3 k = convhull(d(1,:),d(2,:)); %k = indices of re-ordered vertices of the
   convex hull
4 fill(d(1,k),d(2,k),[0.8 1 0.8]); %draw the c.h. using reordered vertices
5 hold on;
6 plot(d(1,:),d(2,:),'b-*'); %draw the nodes
7
```

```

8 x = 0:0.001:1;
9 V = bernstein(n,x);
10 bc = d*V; %compute the Bezier curve
11 plot(bc(1,:),bc(2,:), 'r-', 'LineWidth', 2); %plot the Bezier curve
12
13 xlabel('x');
14 ylabel('y');
15 legend('Convex Hull', 'Control points', 'Bezier curve');
16 hold off;
17 end
18
19 function B = bernstein(n,t)
20 B = [ones(1,length(t)); zeros(n,length(t))];
21 for j=1:n
22     for k=j+1:-1:2
23         B(k,:) = t.*B(k-1,:) + (1-t).*B(k,:);
24     end
25     B(1,:) = (1-t).*B(1,:);
26 end
27 end

```

(3b) [12 points] Write a Matlab function

```
len = function bezLength(d)
```

which approximately computes the length of the Bézier curve by some numerical quadrature.

Solution: Listing 5

Listing 5: bezLength()

```

1 function d = bezLength(d)
2 n = size(d,2)-1; %number of nodes
3 x = 0:0.001:1;
4 V = bernstein(n,x);
5 bc = d*V; %compute the Bezier curve
6
7 % Add up the segments of a bezier curve. Estimate the length of each
8 % segment with the Pythagorean theorem.
9 d = sqrt(sum((bc(:,1:end-1) - bc(:,2:end)).^2));
10 d = sum(d);
11 end
12
13 function B = bernstein(n,t)
14 B = [ones(1,length(t)); zeros(n,length(t))];
15 for j=1:n
16     for k=j+1:-1:2
17         B(k,:) = t.*B(k-1,:) + (1-t).*B(k,:);
18     end
19     B(1,:) = (1-t).*B(1,:);
20 end
21 end

```

(3c) [16 points] We want to approximate a unit half circle by (one segment of) a Bézier curve with 5 Bézier control points. The Bézier control points shall be such that

- the Bézier curve is symmetric.
- the Bézier curve passes through the points $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$.
- the tangent at the two end points $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ corresponds to the tangent of the semi circle.

- the length of the Bézier curve corresponds to the length of a semi circle (π).

Write a Matlab function

```
function halfcircle(d)
```

where you compute the Bézier control points and plot the resulting Bézier curve.

HINT: You may use the following 5 Bézier control points: $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} -1 \\ \alpha \end{pmatrix}$, $\begin{pmatrix} 0 \\ \beta \end{pmatrix}$, $\begin{pmatrix} 1 \\ \alpha \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

HINT: Find a linear (analytical) relation between α and β such that the Bézier curve passes through the point $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Then only one independent parameter is left. Then use a **simple** numerical method from the lecture to determine a good value for this parameter such that the Bézier curve has length π

HINT: If you are unable to complete task (3a) and/or (3b) use the functions `plot_bezcurv_p(d)` and/or `len = bezLength_p(d)` instead.

Solution: The Bezier curve is given by

$$P(t) = \sum_{i=0}^4 P_i B_{i,4}(t)$$

with

$$B_{i,4}(t) = \binom{4}{i} (1-t)^{4-i} t^i.$$

With the 5 Bezier points the Bezier curve yields

$$P(t) = \binom{-1}{0} (1-t)^4 + 4 \binom{-1}{\alpha} (1-t)^3 t + 6 \binom{0}{\beta} (1-t)^2 t^2 + 4 \binom{1}{\alpha} (1-t) t^3 + \binom{1}{0} t^4.$$

To determine α and β we use that $P(t)$ for $t = 0.5$ should lie on the half unit circle. $P(.5)$ is in the middle of the half circle, hence

$$P(.5) = \begin{pmatrix} 0 \\ (8\alpha + 6\beta)0.5^4 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

leading to $\beta = \frac{8}{3} - \frac{4}{3}\alpha$.

Listing 6

Listing 6: `halfcircle()`

```

1 function halfcircle
2 clear all; close all;
3 bezcurv([-1 -1 0 1 1; 0 0.873 1.5027 0.873 0]);
4
5 %bisection
6 bigAlpha = 1;
7 smallAlpha = .5;
8 beta = 8/3-4/3*bigAlpha;
9 alpha = bigAlpha
10 d = bezLength([-1 -1 0 1 1; 0 alpha beta alpha 0]);
11 i = 0;
12 while abs(d-pi) > 10^-10
13 alpha = (bigAlpha+smallAlpha)/2;
14 beta = 8/3-4/3*alpha;
15 d = bezLength([-1 -1 0 1 1; 0 alpha beta alpha 0]);
16
17 if d < pi
18     smallAlpha = alpha;
19 else
20     bigAlpha = alpha;
21 end
22 i=i+1
23

```

```
24 end
25 beta
26
27 end
```

Problem 4 Best rank-1 approximation

Given $\mathbf{A} \in \mathbb{R}^{n,n}$ with positive diagonal we consider the minimization problem

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{A} - \mathbf{x}\mathbf{x}^\top\|_F^2. \quad (2)$$

(4a) [6 points] Reformulate (2) as a standard non-linear least squares problem $\frac{1}{2} \|F(\mathbf{x})\|_2^2 \rightarrow \min$ for a suitable function F .

Solution:

$$F_{i+(k-1)n} = a_{ik} - x_i x_k \quad i, k = 1, \dots, n$$

(4b) [15 points] Derive the Jacobian and implement it in the Matlab function

```
function df = DF(x).
```

Solution:

$$J(\mathbf{x}) = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \dots & \frac{\partial F_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_{n^2}}{\partial x_1} & \dots & \frac{\partial F_{n^2}}{\partial x_n} \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial F_1}{\partial x_1} \\ \vdots \\ \frac{\partial F_{n^2}}{\partial x_1} \end{bmatrix} = \begin{bmatrix} -2x_1 \\ -x_2 \\ \vdots \\ -x_n \\ -x_2 \\ 0 \\ \vdots \\ 0 \\ -x_3 \\ 0 \\ \vdots \\ 0 \\ -x_4 \\ \vdots \end{bmatrix} \quad \begin{bmatrix} \frac{\partial F_1}{\partial x_i} \\ \vdots \\ \frac{\partial F_{n^2}}{\partial x_i} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ -x_1 \\ 0 \\ \vdots \\ -x_{i-1} \\ 0 \\ \vdots \\ -x_1 \\ -x_2 \\ \vdots \\ -x_{i-1} \\ -2x_i \\ x_{i+1} \\ \vdots \\ x_n \\ 0 \\ \vdots \\ 0 \\ -x_{i+1} \\ 0 \\ \vdots \end{bmatrix}$$

Listing 7

Listing 7: DF()

```

1 function df = DF( x )
2 % construct DF
3
4 n = length(x);
5 df = zeros(n*n,n);
6 for i = 1:n
7     first = 1 + (i-1)*n;
8     last = n + (i-1)*n;
9     df(first:last,i) = -x;
10    ind = [1:n:n^2] + i-1;
11    df(ind,i) = df(ind,i) -x;
12 end
13
14 end

```

(4c) [11 points] Write a Matlab code

```
function x = rankoneapprox(A)
```

that computes the solution of (2) by means of the Gauss-Newton iteration with initial guess $x_i^{(0)} = \sqrt{a_{ii}}, i = 1, \dots, n$ with a tolerance of 10^{-3} .

HINT: You may use `reshape`. HINT: If you are unable to complete task 2 use the function `df = DF_p(x)` instead.

HINT: Use `test4.m` to test your function.

Solution: Listing 8

Listing 8: `rankoneapprox()`

```
1 function x = rankoneapprox( A )
2
3 % config
4 tol = 1e-6;
5 n = size(A,1);
6 x = sqrt(diag(A));
7
8 % construct F
9 F = @(x) reshape(A,n*n,1) - reshape(x*x',n*n,1);
10
11 % Newton correction
12 % s = DF(x) \ F(x);
13
14 % Newton iterations
15 while norm(s) > tol * norm(x)
16     s = DF(x) \ F(x);
17     x = x - s;
18 end
19
20 end
```

(4d) [8 points] Explain, why the MATLAB built in function `eig` can be used to solve (2) provided that \mathbf{A} is *symmetric* (hermitian).

Solution:

$$\|\mathbf{A} - \mathbf{x}\mathbf{x}^\top\|_F = \sqrt{\sum_{i=1}^n \sigma_i^2(\mathbf{A} - \mathbf{x}\mathbf{x}^\top)} \quad (3)$$

Choose \mathbf{x} as the eigenvalue of the largest eigenvector of \mathbf{A} times the square root of the largest eigenvalue. This minimizes the right hand side of (3).

(4e) [5 points] Write a MATLAB code

```
function x = symrankoneapprox(A)
```

that computes the solution of (2) for *symmetric* (hermitian) \mathbf{A} using MATLAB's `eig`.

HINT: Use `test4.m` to test your function.

Solution:

Listing 9

Listing 9: `symrankoneapprox()`

```
1 function x = symrankoneapprox( A )
2
3 [V,D] = eig(A);
4 D = diag(D);
5 [d,ind] = max(D);
6 x = V(:,ind)*sqrt(d);
7
8 end
```

Problem 5 ODE / Runge-Kutta method

We want to solve the initial value problem $\dot{y} = f(t, y)$, $y(t_0) = y_0$ by the Runge-Kutta method characterized by the following Butcher table:

$$\begin{array}{c|ccc} 0 & 0 & & \\ \frac{2}{3} & \frac{2}{3} & 0 & \\ 0 & -1 & 1 & 0 \\ \hline & 0 & \frac{3}{4} & \frac{1}{4} \end{array} \quad (4)$$

(5a) [11 points]

Implement two Matlab functions

$$y1 = \text{RK_step}(\text{odefun}, t, y0, h),$$

and

$$[t, y] = \text{RK}(\text{odefun}, \text{tspan}, y0, N).$$

$\text{RK_step}(\text{odefun}, t, y0, h)$ implements one Runge-Kutta step defined by (4), from t to $t+h$ with the starting value $y0$. $\text{odefun}(t, y)$ defines the right hand side of the initial value problem $\dot{y} = f(t, y)$, $y(t_0) = y_0$.

$\text{RK}(\text{odefun}, \text{tspan}, y0, N)$ uses RK_step to solve the ODE over the whole interval specified by tspan using $N \in \mathbb{N}$ uniform timesteps.

Solution: Listing 10 and Listing 11

Listing 10: $\text{RK_step}()$

```
1 function y1 = RK_step(odefun, t, y0, h)
2
3 K1 = odefun(t, y0);
4 K2 = odefun(t + h*2/3, y0 + h*K1*2/3);
5 K3 = odefun(t, y0 + h*(K2 - K1));
6
7 y1 = y0 + h*K2*3/4 + h*K3/4;
```

Listing 11: $\text{RK}()$

```
1 function [t, y] = RK(odefun, tspan, y0, N)
2
3 h = diff(tspan)/N;
4 t = linspace(tspan(1), tspan(2), N+1);
5 y = [y0, zeros(length(y0), N)];
6
7 for k = 1:N
8     y(:, k+1) = RK_step_p(odefun, t(k), y(:, k), h);
9 end
10
11 y = y.'; % change to Matlab convention
```

(5b) [5 points] Bring the following ODE into a suitable form to solve it with $\text{RK_step}(\text{odefun}, t, y0, h)$:

$$x''(t) + x(t) = \sin(t), \quad x(0) = 100, \quad x'(0) = 5. \quad (5)$$

Implement this suitable form in the Matlab function

$$\text{out} = \text{odefun}(t, y).$$

Solution: Listing 12

Listing 12: odefun()

```

1 function out = odefun(t, y)
2
3 out = [y(2); sin(t)-y(1)];

```

(5c) [12 points] Implement a Matlab function

exRK(),

to graphically (with a plot) determine the order of the chosen Runge-Kutta method (4) for the given ODE (5). Use the Matlab function ode45 with relative tolerance $100 * \text{eps}$ and absolute tolerance eps to determine a reference solution.

HINT: The error of a method could be computed by $\|x(T) - \hat{x}(T)\|_{L_2} + \|x'(T) - \hat{x}'(T)\|_{L_2}$, where x is the reference solution and \hat{x} the approximate one. T is the stopping time.

HINT: You may use norm.

HINT: If you are unable to complete task (5a) and/or (5b) use the functions $[t, y] = \text{RK}_p(\text{odefun}, \text{tspan}, y_0, N)$ and/or $\text{out} = \text{odefun}_p(t, y)$ instead.

Solution: Listing 13 and Figure 4

Listing 13: exRK()

```

1 function exRK
2 clear all; close all; tic;
3
4 % Interval and initial condition:
5 tspan = [0 1];
6 y0 = [100; 5];
7
8 % Reference solution with ode45:
9 opts = odeset('RelTol', 100*eps, 'AbsTol', eps);
10 [t, y_ex] = ode45(@odefun, tspan, y0, opts);
11 ex = y_ex(end, :); % extract final value to compute error
12
13 h = 2.^(-(4:10)); % meshwidth
14 n = diff(tspan) ./ h; % # steps
15
16 for l = 1:length(n)
17 [t, y_rk] = RK(@odefun, tspan, y0, n(l));
18 err_rk(l) = norm(ex - y_rk(end, :));
19 a = y_rk(end, :);
20 err_rk(l) = norm(a(1)-ex(1)) + norm(a(2)-ex(2));
21 end
22
23 figure(2);
24 loglog(h, err_rk, 'r', h, 1e3*h.^3, 'k--');
25 axis tight; xlabel('meshwidth\h'); ylabel('error');
26 legend('3-stage RK', 'h^3', 'location', 'nw');
27 set(0, 'DefaultLineWidth', 'default');
28 print -depsc2 'ExRKconvergence.eps';
29 toc

```

(5d) [13 points] Analytically determine the convergence order and the stability interval of the Runge-Kutta method in (4)

Solution: Ansatz: $y'(t) = \lambda y$, $y(0) = 1$ for negative λ the solution is known: $y(t) = e^{\lambda t}$.

For the given RK table we have:

$$k_1 = f(t, y) = \lambda y_j$$

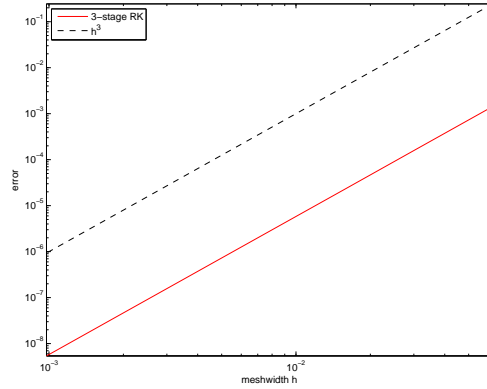


Figure 4: Convergence plot

$$k_2 = f\left(t + \frac{2}{3}h, y_j + \frac{2}{3}hk_1\right) = \lambda_i\left(y_j + \frac{2}{3}hk_1\right) = \left(\lambda_i + \frac{2}{3}h\lambda_i^2\right)y_j$$

$$k_3 = f\left(t + 0, y_j + h(-k_1 + k_2)\right) = \lambda_i\left(y_j + h(-k_1 + k_2)\right) = \left(\lambda_i - h\lambda_i^2 + h\lambda_i^2 + \frac{2}{3}h^2\lambda_i^3\right)y_j = \left(\lambda_i + \frac{2}{3}h^2\lambda_i^3\right)y_j$$

$$y_{j+1} = y_j + h\left(\frac{3}{4}k_2 + \frac{1}{4}k_3\right) = \left(1 + \frac{3}{4}h\lambda_i + \frac{3}{4}\frac{2}{3}h^2\lambda_i^2 + \frac{1}{4}h\lambda_i + \frac{1}{4}\frac{2}{3}h^3\lambda_i^3\right)y_j = \left(1 + h\lambda_i + \frac{1}{2}h^2\lambda_i^2 + \frac{1}{6}h^3\lambda_i^3\right)y_j$$

This are 4 terms of the Taylor series of $e^{h\lambda}$, hence the RK method has order 3.

$$F(h\lambda) = 1 + h\lambda + \frac{1}{2}h^2\lambda^2 + \frac{1}{6}h^3\lambda^3$$

The stability interval is given by $B = \{\mu \in \mathbb{R} \mid |F(\mu)| < 1\}$. By solving for $F(\mu) = 1$ is easy and leads to $\mu = 0$. Solving for $F(\mu) = -1$ is harder and leads to $\mu \approx -2.51$ (alternatively this result is available in the slides). This leads to the stability interval $(-2.51, 0)$.