DELFT UNIVERSITY OF TECHNOLOGY
FACULTY OF ELECTRICAL ENGINEERING, MATHEMATICS AND COMPUTER SCIENCE

The final grade of the test: (total number of points)/5

Examiner responsible: C. Vuik

Examination reviewer: V. Dwarka

TEST SCIENTIFIC COMPUTING (wi4201)
Wednesday January 22 2026, 13:30-16:30

1. Below 5 statements are given. If the statement is true give a short proof. If the statement is wrong give a counter example or an explanation.

   (a) $I \in \mathbb{R}^{n\times n}$ is the identity matrix. $\Rightarrow K_2(I) = 2$. (2 pt.)

   (b) For $A \in \mathbb{R}^{n\times n}$ the matrix norm $\|A\|_{max}$ is defined as $\|A\|_{max} = \max_{1\le i\le n, 1\le j\le n} |a_{i,j}|$. This norm has the multiplicative property. (2 pt.)

   (c) $A = \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -2 & 2 \end{pmatrix}$ The spectral radius of A is larger than 6. (2 pt.)

   (d) $A \in \mathbb{R}^{n\times n}$, and $A = A^T \Rightarrow \|A\|_2 = \max\limits_{1\le i\le n} \lambda_i$ where λ_i is an eigenvalue of A. (2 pt.)

   (e) $A \in \mathbb{R}^{n\times n}$ is a lower triangular matrix with only zero elements on the main diagonal $\Rightarrow A^n = 0$. (2 pt.)

2. Consider the boundary value problem on the unit square

$$-\nabla \cdot (\alpha \nabla u)(x,y) + \sigma u(x,y) = f(x,y) \text{ for } (x,y) \in \Omega = (0,1)\times(0,1)$$

   with homogeneous Dirichlet boundary conditions

$$u(x,y) = 0 \text{ for } (x,y) \in \partial\Omega.$$

   Here, $\alpha > 0$ and $\sigma \in \mathbb{R}$ are constants. For the discretization we use a uniform grid with n intervals in each coordinate direction and elimination of the boundary conditions.

   (a) Classify the PDE (elliptic, parabolic, hyperbolic). (1 pt.)

   (b) Give a discretization such that the truncation error is $O(h^2)$. (2 pt.)

(c) Take $n = 3$. Using a lexicographic ordering, give the full matrix A and vector f such that $Au = f$. Motivate why A is symmetric. (2 pt.)

We now continue by considering a general n and h.

(d) Assume $\alpha = 1$. Take $\sigma = 1$ and use Gershgorin's theorem to formulate a bound (as an interval in terms of h) that contains all the eigenvalues of A. (2 pt.)

(e) Now repeat the previous question for $\sigma = -1$. (1 pt.)

(f) Using the bounds from the previous parts, characterize all σ for which you can guarantee A is nonsingular, including any nontrivial cases. (2 pt.)

3. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric positive definite matrix and consider solving

$$Ax = f$$

(a) Consider a perturbation of the right hand side $f \mapsto f + \Delta f$ with $\|\Delta f\| \leq \delta \|f\|$. Prove a bound of the form

$$\frac{\|\Delta x\|}{\|x\|} \leq \delta\, \kappa(A),$$

for any submultiplicative matrix norm. (2 pt.)

(b) Prove that if A is symmetric positive definite, then

$$\kappa_2(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}.$$

Use the fact that we can diagonalize A with an orthogonal basis of eigenvectors. (3 pt.)

(c) Assume the computed solution $\hat{x}$ returned by a Cholesky based solver satisfies the backward error relation

$$(A + E)\hat{x} = f, \qquad \|E\|_2 \leq \varepsilon \|A\|_2,$$

for some $\varepsilon > 0$. Using only norm inequalities, show that if

$$\varepsilon\, \kappa_2(A) < 1,$$

then

$$\frac{\|\hat{x} - x\|_2}{\|x\|_2} \leq \frac{\varepsilon\, \kappa_2(A)}{1 - \varepsilon\, \kappa_2(A)}.$$

You may use that if $\|B\|_2 < 1$, then $(I + B)$ is invertible and

$$\|(I + B)^{-1}\|_2 \leq \frac{1}{1 - \|B\|_2}.$$

(3 pt.)

(d) Now suppose A is the matrix obtained by the standard five point finite difference discretization of $-\Delta u = f$ on $\Omega = (0,1) \times (0,1)$ with Dirichlet boundary conditions, using a uniform grid with mesh size h. Assume you know the eigenvalue bounds

$$c_1 \leq \lambda_{\min}(A) \leq \lambda_{\max}(A) \leq \frac{c_2}{h^2},$$

with constants $c_1, c_2 > 0$ independent of h. Prove that

$$\kappa_2(A) = O(h^{-2}),$$

and combine this with (a) to explain why the upper bound deteriorates as $h \to 0$ if machine precision is fixed. (2 pt.)

4. (a) Given the linear system $A\,\mathbf{u} = \mathbf{f}$ with an n-by-n real-valued coefficient matrix A. Assume a splitting of this coefficient matrix of the form $A = M - N$ where M is non-singular and assume that a basic iterative solution method for the linear system is derived from this splitting. Derive a recursion formula for the iterates $\mathbf{u}^k$. Derive a recursion formula for the residual vector $\mathbf{r}^k$.

(2 pt.)

(b) Give the relation between the error $\mathbf{e}^k$ and the residual vector $\mathbf{r}^k$. Use this relation to derive the defect-correction scheme that use the approximation $\widehat{A}$ to A;

(2 pt.)

(c) Consider a 2D Poisson equation with Dirichlet boundary conditions on a square domain, discretized by a 5-point stencil. Give the stencil notation of the Jacobi iteration matrix for an interior grid point. (2 pt.)

(d) Derive the damped Jacobi method and give the damped Jacobi iteration matrix. (2 pt.)

(e) Do 1 iteration with the backward Gauss-Seidel method to the following linear system, where we start with the zero vector.

$$\begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

(2 pt.)

5. In this exercise we have to solve a linear system $A\mathbf{u} = \mathbf{f}$, where A is an $n \times n$ non-singular matrix.

(a) Take $\mathbf{u}_1 = \alpha \mathbf{f}$. Derive an expression for α such that $\|\mathbf{u} - \mathbf{u}_1\|_{A^T A}$ is minimal. (2 pt.)

(b) The CGNR method is the CG method applied to $A^TA\mathbf{u} = A^T\mathbf{f}$. Show that the 2-norm of the residuals is monotone decreasing. (2 pt.)

(c) Give a 3×3 non-diagonal matrix such that CGNR converges in one iteration for every right-hand side vector $\mathbf{f}$. (2 pt.)

(d) Given the algorithm

Bi-CGSTAB method

$\mathbf{u}^0$ is an initial guess; $\mathbf{r}^0 = \mathbf{f} - A\mathbf{u}^0$;
$\bar{\mathbf{r}}^0$ is an arbitrary vector, such that $(\bar{\mathbf{r}}^0)^T\mathbf{r}^0 \neq 0$, e.g., $\bar{\mathbf{r}}^0 = \mathbf{r}^0$;
$\rho_{-1} = \alpha_{-1} = \omega_{-1} = 1$;
$\mathbf{v}^{-1} = \mathbf{p}^{-1} = 0$;
for $i = 0, 1, 2, \ldots$ do
  $\rho_i = (\bar{\mathbf{r}}^0)^T\mathbf{r}^i$; $\beta_{i-1} = (\rho_i/\rho_{i-1})(\alpha_{i-1}/\omega_{i-1})$;
  $\mathbf{p}^i = \mathbf{r}^i + \beta_{i-1}(\mathbf{p}^{i-1} - \omega_{i-1}\mathbf{v}^{i-1})$;
  $\hat{\mathbf{p}} = M^{-1}\mathbf{p}^i$;
  $\mathbf{v}^i = A\hat{\mathbf{p}}$;
  $\alpha_i = \rho_i/(\bar{\mathbf{r}}^0)^T\mathbf{v}^i$;
  $\mathbf{s} = \mathbf{r}^i - \alpha_i\mathbf{v}^i$;
  if $\|\mathbf{s}\|$ small enough then
    $\mathbf{u}^{i+1} = \mathbf{u}^i + \alpha_i\hat{\mathbf{p}}$; quit;
  $\mathbf{z} = M^{-1}\mathbf{s}$;
  $\mathbf{t} = A\mathbf{z}$;
  $\omega_i = \mathbf{t}^T\mathbf{s}/\mathbf{t}^T\mathbf{t}$;
  $\mathbf{u}^{i+1} = \mathbf{u}^i + \alpha_i\hat{\mathbf{p}} + \omega_i\mathbf{z}$;
  if $\mathbf{u}^{i+1}$ is accurate enough then quit;
  $\mathbf{r}^{i+1} = \mathbf{s} - \omega_i\mathbf{t}$;
end for

The matrix M in this scheme represents the preconditioning matrix based on ILU(0). Determine the minimal amount of memory and flops per iteration, where we assume that A is based on a 5-point stencil. (2 pt.)

(e) Give a comparison of the mathematical properties of the CGNR and Bi-CGSTAB method (both without preconditioning). (2 pt.)