DELFT UNIVERSITY OF TECHNOLOGY

FACULTY OF ELECTRICAL ENGINEERING, MATHEMATICS AND COMPUTER SCIENCE

ANSWERS OF THE TEST SCIENTIFIC COMPUTING (wi4201)
Wednesday January 22 2025, 13:30-16:30

This document contains short answers, which indicate how the exercises can be answered. In most of the cases more details are needed to give a sufficiently clear answer.

1. (a) Yes
   Note that the following inequality is valid:

$$\|\mathbf{u}\|_2 = \sqrt{\sum_{i=1}^{n}(u_i)^2} \leq \sqrt{n \max_{1\leq i\leq n}|u_i|^2} = \sqrt{n}\|\mathbf{u}\|_\infty.$$

   (b) Yes. Note that D is SPD. Check the properties of the innerproduct.

   - Symmetry

$$(\mathbf{x},\mathbf{y})_D = \mathbf{x}^T D\mathbf{y} = (D^T\mathbf{x})^T\mathbf{y} = \mathbf{y}^T(D^T\mathbf{x}) = \mathbf{y}^T D\mathbf{x} = (\mathbf{y},\mathbf{x})_D$$

   - Linear

$$(a\mathbf{x}+b\mathbf{y},\mathbf{z})_D = a\mathbf{x}^T D\mathbf{z} + b\mathbf{y}^T D\mathbf{z} = a(\mathbf{x},\mathbf{z})_D + b(\mathbf{y},\mathbf{z})_D$$

   - positivity

$$(\mathbf{x},\mathbf{x})_D = \mathbf{x}^T D\mathbf{x} = \sum_{i=1}^{n}\frac{i}{n}x_i^2$$

     This implies that $(\mathbf{x},\mathbf{x})_D > 0$ if $\mathbf{x}$ is not the zero vector and if $(\mathbf{x},\mathbf{x})_D = 0$ $\mathbf{x}$ is the zero vector.

   (c) Yes. Since Q is an orthogonal matrix we know that $QQ^T = Q^TQ = I$. Suppose $\lambda \in \sigma(A)$.

$$A\mathbf{v} = \lambda\mathbf{v}$$
$$Q^TA\mathbf{v} = \lambda Q^T\mathbf{v}$$
$$Q^TAQQ^T\mathbf{v} = \lambda Q^T\mathbf{v}$$

   suppose $\mathbf{w} = Q^T\mathbf{v}$

$$Q^TAQ\mathbf{w} = \lambda\mathbf{w}$$

   so $\lambda \in \sigma(Q^TAQ)$.

(d) Yes. If $\rho(A) < 1$, then $I - A$ has eigenvalues bounded away from zero, and is therefore non-singular. We furthermore have the equality

$$(I - A^{k+1}) = (I - A)(I + A + A^2 + \ldots + A^k)\,, \tag{1}$$

or equivalently

$$(I - A)^{-1}(I - A^{k+1}) = (I + A + A^2 + \ldots + A^k)\,. \tag{2}$$

Taking the limit as $k \to \infty$ and taking into account that since $\rho(A) < 1$ it follows that $\lim_{k\to\infty} \|A^k\|_2 = 0$ yields the desired result.

(e) Yes, Bi-CGSTAB is a short recurrcence method. So the number of vector updates and inner products per iteration are independent of the iteration number.

2. No point deductions for treatment of the RHS function as it said (x, y) on the exam rather than $g(x, y)$ being some function of only the coordinates.

The ellipticity is analyzed by considering the coefficients of the highest order derivatives. Specifically, we use:

$$\mathcal{L}(u) = a_{11}\frac{\partial^2 u}{\partial x^2} + 2a_{12}\frac{\partial^2 u}{\partial x \partial y} + a_{22}\frac{\partial^2 u}{\partial y^2} + b_1\frac{\partial u}{\partial x} + b_2\frac{\partial u}{\partial y} + cu(x, y).$$

We classify these equation based on the sign of the determinant

$$D = \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}^2.$$

The coefficients in our equation are $a_{11} = -1, a_{12} = 0, a_{22} = -1$. The determinant is given by:

$$(-1)(-1) - 0^2 = 1 > 0$$

Hence, the equation is classified as elliptic.

(b) For the discretization, we use the Finite Difference Method on the given grid. Using nearest neighbors, we have for the internal nodes that

$$\frac{\partial^2 u}{\partial x^2}(x_{ij}, y_{ij}) = \frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{h^2} + O(h^2) \quad \text{for } 2 \le i, j \le m$$

(and similar for the y-derivative). Using Taylor polynomials, the claim that the error is $O(h^2)$ should be shown. The approximation to the partial differential equation discretized on internal points of the grid can be written as

$$-\frac{u_{i,j-1} - 2u_{i,j} + u_{i,j+1}}{h^2} - \frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{h^2} + ku_{i,j} = kg(x_i, y_j) \quad \text{for } 2 \le i, j \le m.$$

(c) The stencil on the internal nodes is given by:

$$\frac{1}{h^2}\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4+kh^2 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

The stencil in the lower left corner is:

$$\frac{1}{h^2}\begin{bmatrix} 0 & -1 & 0 \\ 0 & 4+kh^2 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

(d) The matrix A has the following structure:

$$\frac{1}{h^2}\begin{bmatrix} T_h & -I & 0 & \cdots & 0 \\ -I & T_h & -I & \cdots & 0 \\ 0 & -I & T_h & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & T_h \end{bmatrix}$$

where T_h is given by

$$\begin{bmatrix} 4+kh^2 & -1 & 0 & \cdots & 0 \\ -1 & 4+kh^2 & -1 & \cdots & 0 \\ 0 & -1 & 4+kh^2 & \cdots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & -1 & \cdots & 4+kh^2 \end{bmatrix}$$

Finally, the bandwidth is equal to m.

(e) Gershgorin's theorem is given by:

If $\lambda \in \sigma(A)$, then λ is located in one of the n closed disks in the complex plane that has center a_{ii} and radius

$$\rho_i = \sum_{j=1, j\neq i}^{n} |a_{ij}|,$$

i.e.,

$$\lambda \in \sigma(A) \implies |a_{ii} - \lambda| \leq \rho_i.$$

Now with our uniform grid and zero Dirichlet conditions we get

$$-\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} + 2u = 2g(x, y)$$

Substituting the second order central difference approximations, we get at grid point (i, j):

$$\frac{-u_{i+1,j} - u_{i-1,j} - u_{i,j+1} - u_{i,j-1} + \left(\frac{4}{h^2} + 2\right) u_{i,j}}{h^2} = kg(x_i, y_j).$$

The matrix A representing this system has elements:

$$a_{ii} = \frac{4}{h^2} + 2, \quad a_{i,i\pm1} = a_{i,i\pm N} = -\frac{1}{h^2} \text{ (for interior points)}.$$

Applying Gershgorin using $a_{ii} = \frac{4}{h^2} + 2$ gives:

- Interior Nodes: Each eigenvalue λ of A is within:

$$\text{Radius } \rho_i = \frac{4}{h^2}, \quad \text{Interval: } \left[2, 2 + \frac{8}{h^2}\right].$$

- Edge Nodes (not corners): They have three neighbors contributing, thus:

$$\text{Radius } \rho_i = \frac{3}{h^2}, \quad \text{Interval: } \left[2 + \frac{1}{h^2}, 2 + \frac{7}{h^2}\right].$$

- Corner Nodes: With two neighbors, their bounds are:

$$\text{Radius } \rho_i = \frac{2}{h^2}, \quad \text{Interval: } \left[2 + \frac{2}{h^2}, 2 + \frac{6}{h^2}\right].$$

The eigenvalues lie in the union of these circles in the complex plane, hence we get:

$$2 \leq \lambda \leq \frac{8}{h^2} + 2$$

(f) Similarly, we now get for center $a_{ii} = \frac{4}{h^2} - 2$

- Interior Nodes:

$$\text{Radius } \rho_i = \frac{4}{h^2}, \quad \text{Interval: } \left[-2, \frac{8}{h^2} - 2\right].$$

- Edge Nodes (not corners):

$$\text{Radius } \rho_i = \frac{3}{h^2}, \quad \text{Interval: } \left[\frac{1}{h^2} - 2, \frac{7}{h^2} - 2\right].$$

- Corner Nodes:

$$\text{Radius } \rho_i = \frac{2}{h^2}, \quad \text{Interval: } \left[\frac{2}{h^2} - 2, \frac{6}{h^2} - 2\right].$$

The eigenvalues lie in the union of these circles in the complex plane, hence we get:

$$-2 \leq \lambda \leq \frac{8}{h^2} - 2$$

(g) The analysis shows that for $k = -2$ all eigenvalues are real, yet some are non-positive, indicating that the matrix is indefinite. This affects the choice of numerical methods for solving the linear system, with methods like Conjugate Gradient likely being inappropriate due to the presence of negative and potentially zero eigenvalues. Both indefiniteness and possible singularity require a method like GMRES.

3. (a) We have

$$A_1 = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 5 \\ 4 & 6 & 8 \end{bmatrix},$$

To eliminate the first column below the diagonal, we calculate the Gauss vector $\alpha^{(1)}$ as:

$$\alpha^{(1)} = \begin{bmatrix} 0 \\ \frac{a_{21}}{a_{11}} \\ \frac{a_{31}}{a_{11}} \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{2}{1} \\ \frac{4}{1} \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix}.$$

Using this Gauss vector, the matrix after one step of elimination becomes:

$$L_1 A_1 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 5 \\ 4 & 6 & 8 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & 2 & 4 \end{bmatrix}.$$

To eliminate the second column below the diagonal, we calculate the next Gauss vector $\alpha^{(2)}$ as:

$$\alpha^{(2)} = \begin{bmatrix} 0 \\ 0 \\ \frac{a_{32}^{(1)}}{a_{22}^{(1)}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \frac{2}{1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}.$$

Using this Gauss vector, the matrix after the second elimination step becomes:

$$L_2(L_1 A_1) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & -2 \end{bmatrix}.$$

From the elimination process, the upper triangular matrix is:

$$U = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & -2 \end{bmatrix}.$$

The lower triangular matrix L is obtained by combining the Gauss vectors:

$$L = \begin{bmatrix} 1 & 0 & 0 \\ \alpha_2^{(1)} & 1 & 0 \\ \alpha_3^{(1)} & \alpha_3^{(2)} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 4 & 2 & 1 \end{bmatrix}.$$

Finally, $A_1 = LU$:

$$A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 4 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 5 \\ 4 & 6 & 8 \end{bmatrix}.$$

(b) In this case we have

$$\hat{L}\hat{U} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2+\epsilon & 5 \\ 4 & 6 & 4 \end{bmatrix} \neq A_2$$

If we use $\hat{L}$ and $\hat{U}$ to solve for x using $Ly = b$ and $Ux = y$, then

$$\hat{x} = \begin{bmatrix} \frac{11}{2} - \frac{2}{3}\epsilon \\ -2 \\ \frac{2}{3}\epsilon - \frac{2}{3} \end{bmatrix} \approx \begin{bmatrix} \frac{11}{2} \\ -2 \\ -\frac{2}{3} \end{bmatrix}$$

Thus we get

$$\|x - \hat{x}\| \approx 4.069$$

(c) We now have

$$A_3 = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 5 \\ 4 & 6 & 8 \end{bmatrix}.$$

We use partial pivoting. The largest element in the first column is 4 (in row 3). Thus, we swap row 1 with row 3. The updated matrix is:

$$PA_3 = \begin{bmatrix} 4 & 6 & 8 \\ 2 & 2 & 5 \\ 1 & 1 & 1 \end{bmatrix}.$$

The permutation matrix P is:

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

To eliminate the first column below the diagonal, we compute the first Gauss vector $\alpha^{(1)}$:

$$\alpha^{(1)} = \begin{bmatrix} 0 \\ \frac{a_{21}}{a_{11}} \\ \frac{a_{31}}{a_{11}} \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{2}{4} \\ \frac{1}{4} \end{bmatrix} = \begin{bmatrix} 0 \\ 0.5 \\ 0.25 \end{bmatrix}.$$

Using $\alpha^{(1)}$, we perform the first elimination step:

$$L_1 P A_3 = \begin{bmatrix} 4 & 6 & 8 \\ 0 & -1 & 1 \\ 0 & -0.5 & -1 \end{bmatrix}.$$

To eliminate the second column below the diagonal, we compute the second Gauss vector $\alpha^{(2)}$:

$$\alpha^{(2)} = \begin{bmatrix} 0 \\ 0 \\ \frac{a_{32}^{(1)}}{a_{22}^{(1)}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \frac{-0.5}{-1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0.5 \end{bmatrix}.$$

Using $\alpha^{(2)}$, we perform the second elimination step:

$$L_2(L_1 P A_3) = \begin{bmatrix} 4 & 6 & 8 \\ 0 & -1 & 1 \\ 0 & 0 & -1.5 \end{bmatrix}.$$

The matrices L, U, and P from the factorization are:

$$U = \begin{bmatrix} 4 & 6 & 8 \\ 0 & -1 & 1 \\ 0 & 0 & -1.5 \end{bmatrix}.$$

The entries in L are constructed using the Gauss vectors:

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ 0.25 & 0.5 & 1 \end{bmatrix}.$$

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Finally, $PA_3 = LU$.

$$PA_3 = \begin{bmatrix} 4 & 6 & 8 \\ 2 & 2 & 5 \\ 1 & 1 & 1 \end{bmatrix},$$

and

$$LU = \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ 0.25 & 0.5 & 1 \end{bmatrix} \begin{bmatrix} 4 & 6 & 8 \\ 0 & -1 & 1 \\ 0 & 0 & -1.5 \end{bmatrix} = \begin{bmatrix} 4 & 6 & 8 \\ 2 & 2 & 5 \\ 1 & 1 & 1 \end{bmatrix}.$$

4. (a) If A is SPD show that $||\mathbf{u}||_A = \sqrt{\mathbf{u}^T A \mathbf{u}}$ is a norm.

i.) $||\mathbf{u}||_A \geq 0$ with equality only for $\mathbf{u} = \mathbf{0}$

$$\begin{aligned} ||\mathbf{u}||_A &= \sqrt{\mathbf{u}^T A \mathbf{u}} \\ &\geq 0 \quad \text{(this follows from } A \text{ is SPD)} \end{aligned}$$

Since A is SPD, equality can only occur if $\mathbf{u} = \mathbf{0}$

ii.) $||c\mathbf{u}|| = |c|||\mathbf{u}||$

$$\begin{aligned} ||c\mathbf{u}|| &= \sqrt{c\mathbf{u}^T A c\mathbf{u}} \\ &= \sqrt{c^2 \mathbf{u}^T A \mathbf{u}} \\ &= |c|\sqrt{\mathbf{u}^T A \mathbf{u}} \\ &= |c|||\mathbf{u}||_A \end{aligned}$$

iii.) $||\mathbf{u} + \mathbf{v}||_A \leq ||\mathbf{u}||_A + ||\mathbf{v}||_A$

$$\begin{aligned} ||\mathbf{u} + \mathbf{v}||_A &= \sqrt{(\mathbf{u} + \mathbf{v})^T A (\mathbf{u} + \mathbf{v})} \\ &= \sqrt{\mathbf{u}^T A \mathbf{u} + 2\mathbf{u}^T A \mathbf{v} + \mathbf{v}^T A \mathbf{v}} \\ &= \sqrt{||\mathbf{u}||_A^2 + 2\mathbf{u}^T A \mathbf{v} + ||\mathbf{v}||_A^2} \\ &\leq \sqrt{||\mathbf{u}||_A^2 + 2||\mathbf{u}||_A||\mathbf{v}||_A + ||\mathbf{v}||_A^2} \\ &= \sqrt{(||\mathbf{u}||_A + ||\mathbf{v}||_A)^2} \\ &= ||\mathbf{u}||_A + ||\mathbf{v}||_A \end{aligned}$$

(b) We assume that $\mathbf{u}^1 = \alpha_0 \mathbf{r}^0$. Determine α_0 such that $\|\mathbf{u} - \mathbf{u}^1\|_A$ is minimal.

$$\begin{aligned} \|\mathbf{u} - \mathbf{u}^1\|_A^2 &= (\mathbf{u} - \mathbf{u}^1)^T A (\mathbf{u} - \mathbf{u}^1) \\ &= \|\mathbf{u}\|_A - 2\alpha_0 (\mathbf{r}^0)^T A \mathbf{u} + \alpha_0^2 \|\mathbf{r}^0\|_A \end{aligned}$$

$$\frac{d}{d\alpha_0} \|\mathbf{u} - \mathbf{u}^1\|_A^2 = -2(\mathbf{r}^0)^T A \mathbf{u} + 2\alpha_0 \|\mathbf{r}^0\|_A \tag{3}$$

Then we impose that (3) be equal zero to obtain:

$$\alpha_0 = \frac{(\mathbf{r}^0)^T A \mathbf{u}}{\|\mathbf{r}^0\|_A} = \frac{(\mathbf{r}^0)^T \mathbf{f}}{\|\mathbf{r}^0\|_A}$$

(c) The matrix A corresponds to a shifted discretized Poisson operator. The eigenvalues are given by

$$\lambda_{k,l} = 5 - 2\cos\frac{\pi k}{61} - 2\cos\frac{\pi l}{61}, \quad 1 \leq k, l \leq 60.$$

Determine the linear rate of convergence for the Conjugate Gradient method. The matrix A is SPD, hence we can use:

$$\kappa_2(A) = \frac{\lambda_{max}(A)}{\lambda_{min}(A)}$$

to obtain:

$$\kappa_2(A) = \frac{\lambda_{60,60}(A)}{\lambda_{1,1}(A)} = \frac{8.994}{1.005} = 8.9492$$

So in terms of Theorem 7.1.3 of the lecture notes, the linear rate of convergence is:

$$\frac{\sqrt{\kappa_2(A)} - 1}{\sqrt{\kappa_2(A)} + 1} = 0.4989$$

(d) Based on the first paragraph of Section 7.2 of the lecture notes, a preconditioner matrix M should satisfy the following:

- M is SPD,
- the eigenvalues of $M^{-1}A$ are clustered around 1,
- $M^{-1}y$ is obtainable at low cost.

The PCG method is obtained, given a suitable preconditioner $M = PP^T$, by applying the CG method to a preconditioned linear system $\tilde{A}\tilde{\mathbf{u}} = \tilde{\mathbf{y}}$, where $\tilde{A} = P^{-1}AP^{-T}$, $\mathbf{u} = P^{-T}\tilde{\mathbf{u}}$ and $\tilde{\mathbf{y}} = P^{-1}\mathbf{y}$, and P is a nonsingular matrix. This can also be rewritten such that CG is applied to the system $M^{-1}A\mathbf{u} = M^{-1}\mathbf{f}$.

(e) The eigenvalues of the matrix A are:

$$\lambda_1 = 1,\ \lambda_2 = 80,\ \lambda_3 = 82$$

A is symmetric and all its eigenvalues are positive, then A is SPD.

$$\kappa_2(A) = \frac{\lambda_{max}(A)}{\lambda_{min}(A)} = 82$$

$$\frac{\sqrt{\kappa_2(A)} - 1}{\sqrt{\kappa_2(A)} + 1} = 0.8011$$

We give an estimate of convergence in terms of the number of iterations k needed to obtain:

$$\frac{\|\mathbf{u} - \mathbf{u}^k\|_A}{\|\mathbf{u} - \mathbf{u}^0\|_A} \leq 2(0.8011)^k = 10^{-12} \tag{4}$$

thus k is at most:

$$k = \frac{\log\left(\frac{10^{-12}}{2}\right)}{\log(0.8011)} = 127.71 \tag{5}$$

We consider a preconditioner where P is a diagonal matrix whose diagonal elements $p_{i,i} = \sqrt{a_{i,i}}$:

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{pmatrix}$$

in order to estimate the convergence of the CG method we look at the eigenvalues of the matrix:

$$\tilde{A} = P^{-1}AP^{-T} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -\frac{1}{81} \\ 0 & -\frac{1}{81} & 1 \end{pmatrix} \tag{6}$$

these eigenvalues are $\lambda_1 = 1$, $\lambda_2 = \frac{80}{81}$, $\lambda_3 = \frac{82}{81}$

$$\kappa_2(\tilde{A}) = \frac{\lambda_{max}(\tilde{A})}{\lambda_{min}(\tilde{A})} = \frac{82}{80}$$

$$\frac{\sqrt{\kappa_2(\tilde{A})} - 1}{\sqrt{\kappa_2(\tilde{A})} + 1} = 0.006$$

$$\frac{\|\mathbf{u} - \mathbf{u}^k\|_{\tilde{A}}}{\|\mathbf{u} - \mathbf{u}^0\|_{\tilde{A}}} \leq 2(0.006)^k = 10^{-12}$$

thus k is at most:

$$k = \frac{\log\left(\frac{10^{-12}}{2}\right)}{\log(0.006)} = 5.33$$

5. (a) We assume that $\mathbf{q}_{k-1} = \mathbf{v}_1 + \mathbf{w}$ with $\|\mathbf{w}\|_2 = O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)$. From the algorithm we know that $\lambda^{(k)} = \mathbf{q}_{k-1}^T \mathbf{z}_k$, which is equal to $\lambda^{(k)} = \bar{\mathbf{q}}_{k-1}^T A \mathbf{q}_{k-1}^T = (\mathbf{v}_1 + \mathbf{w})^T(\mathbf{v}_1 + \mathbf{w}) = \lambda_1 \mathbf{v}_1^T \mathbf{v}_1 + O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)$ In order to prove the result we have to show that $\mathbf{v}_1^T \mathbf{v}_1$ is close to 1. This can be shown as follows: $\mathbf{v}_1^T \mathbf{v}_1 = (\mathbf{q}_{k-1} - \mathbf{w})^T(\mathbf{q}_{k-1} - \mathbf{w}) = 1 + O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)$ This proves the result.

   (b) For the shifted power method we apply the power method to the matrix $A - cI$. To obtain the original eigenvalue the result of this power method approximation should be shifted back by adding the value c. We know that for the shifted power method the convergence is determined by the ratio $\left|\frac{\lambda_2 - c}{\lambda_1 - c}\right|$ if we assume that $|\lambda_1 - c| > |\lambda_2 - c| \geq |\lambda_n - c|$. We obtain fast convergence if the ratio $\left|\frac{\lambda_2 - c}{\lambda_1 - c}\right|$ is as small as possible. This implies that $|\lambda_2 - c| = |\lambda_n - c|$. This leads to $c = \frac{\lambda_2 + \lambda_n}{2}$.

(c) Two options are possible or based on the linear converging result, or based on the residual. For the first stopping criterion we can use:

$$\text{estimate} \quad r \quad \text{from} \quad \tilde{r} = \frac{|\lambda^{(k+1)} - \lambda^{(k)}|}{|\lambda^{(k)} - \lambda^{(k-1)}|} \, ,$$

and stop if $\frac{\tilde{r}}{1-\tilde{r}} \frac{|\lambda^{(k+1)} - \lambda^{(k)}|}{|\lambda^{(k+1)}|} \leq \varepsilon$. Or the residual is small

$$\frac{\|\lambda^{(k)} \mathbf{q}_k - A\mathbf{q}_k\|_2}{|\lambda^{(k)}|} < \varepsilon$$

(d) To approximate the smallest eigenvalue where $\lambda_{n-1} = 1.1$ and $\lambda_n = 1$ the inverse power method is the method of choice. This means that the power method is applied to A^{-1}. If the shifted power method is used the convergence will be very slow $\frac{1001}{1001.1} = 0.9999$, whereas if the inverse power method is used the convergence is given by $\frac{1}{1.1} = 0.9091$. This is much faster.