Solution 9 – MATH-250 Advanced Numerical Analysis I

The exercise sheet is divided into two sections: quiz and exercises. The quiz will be discussed in the beginning of the lecture on Thursday, May 8. The exercises marked with ($\star$) are graded homework. The exercises marked with **(Python)** are implementation based and can be solved in the Jupyter notebooks which are available on Moodle/Noto. **The deadline for submitting your solutions to the homework is Friday, May 9 at 10h15.**

Quiz

(a) For *any* invertible matrix A, right-hand side $\mathbf{b}$, and starting vector $\mathbf{x}_0$, there is a choice of α such that the Richardson method converges.

☐ True ■ False

(b) Consider a family of linear systems

$$A_n \mathbf{x} = \mathbf{b}_n, \quad A_n \in \mathbb{R}^{n \times n},$$

such that

- A_n is symmetric positive definite;
- $\kappa_2(A_n) = \|A_n\|_2 \|A_n^{-1}\|_2 = O(n^2)$ for $n \to \infty$;
- $\|\mathbf{x}\|_2 = 1$.

Consider fixed accuracy $\varepsilon > 0$. Let k_n denote the minimal number of iterations of the Richardson method (with optimal α, zero starting vector, no preconditioner) needed to attain $\|\mathbf{x}_{k_n} - \mathbf{x}\|_2 \le \varepsilon$. Then for $n \to \infty$ it holds that

☐ $k_n = O(1)$ ☐ $k_n = O(n)$

☐ $k_n = O(\log n)$ ■ $k_n = O(n^2)$

(c) Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable on $\mathbb{R}^n$. If $\mathbf{x}$ is a minimum of f then $\nabla f(\mathbf{x}) = 0$.

■ True ☐ False

(d) Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable on $\mathbb{R}^n$ and $\mathbf{x}$ such that $\nabla f(\mathbf{x}) \neq 0$. Then for every $\varepsilon > 0$ there is $\mathbf{y}$ with $\|\mathbf{y} - \mathbf{x}\| \le \varepsilon$ and $f(\mathbf{y}) < f(\mathbf{x})$.

■ True ☐ False

Solution.

(a) Let $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, and $\alpha \neq 0$. Clearly, A is invertible, and its eigenvalues are $\lambda_1 = 1, \lambda_2 = -1$. Therefore, the matrix $I - \alpha A$ has eigenvalues $\lambda_1(\alpha) = 1 - \alpha\lambda_1$ and $\lambda_2(\alpha) = 1 - \alpha\lambda_2$, and the associated eigenvectors of A are also eigenvectors of $I - \alpha A$. The differing signs of λ_1 and λ_2 imply that for either λ_1 or λ_2 it must hold that $|\lambda_i(\alpha)| \not\leq 1$. We now choose $\mathbf{v}$ to be the eigenvector of A associated with the λ_i (denoted $\hat{\lambda}$) such that $|\lambda_i(\alpha)| \geq 1$, and set the initial guess to be $\mathbf{x}_0 = \mathbf{x} + \mathbf{v}$. Thus, the Richardson iteration diverges because we observe that

$$\lim_{k\to\infty} \mathbf{e}_k = \lim_{k\to\infty} (I - \alpha A)^k \mathbf{v} = \lim_{k\to\infty} \hat{\lambda}(\alpha)^k \mathbf{v} \neq 0.$$

(b) The convergence rate of the Richardson iteration is given by

$$\|\mathbf{x}^{k+1} - \mathbf{x}\|_2 \leq \left(\frac{\kappa(A) - 1}{\kappa(A) + 1}\right)^k \|\mathbf{x} - \mathbf{x}_0\|_2.$$

In order to reach the tolerance ε we need to find k such that $\varepsilon = \left(\frac{n^2-1}{n^2+1}\right)^k$. We compute

$$\varepsilon = \left(\frac{n^2 - 1}{n^2 + 1}\right)^k = \left(1 - \frac{2}{n^2 + 1}\right)^k \implies k = \log_{1-2/(n^2+1)}(\varepsilon) = \frac{\log(\varepsilon)}{\log(1 - 2/(n^2 + 1))}.$$

The only relevant part in this is the evaluation of $1/\log(1 - 2/(n^2 + 1))$ because the other term is constant w.r.t. n. We notice that $2/(n^2 + 1) \to 0$ for $n \to \infty$, which means that we will be evaluating the logarithm very close to 1, justifying the linear approximation $\log(1 - 2/(n^2 + 1)) \approx 2/(n^2 + 1)$ and finally $k \approx (n^2 + 1)/2 = \mathcal{O}(n^2)$.

(c) For $\varepsilon > 0$ we define the differentiable curve $\gamma\colon (-\varepsilon, \varepsilon) \to \mathbb{R}^n, t \mapsto \mathbf{x} + t\nabla f(\mathbf{x})$. Thus, the scalar function $f \circ \gamma$ has a minimum at 0, and by results from Analysis I and II we know that for the minimum of a scalar function it holds that $0 = (f \circ \gamma)'(0)$. Next,

$$(f \circ \gamma)'(0) = \langle \nabla f(\mathbf{x}), \nabla f(\mathbf{x}) \rangle$$

follows by the chain rule and therefore $\nabla f(\mathbf{x}) = 0$.

(d) If there exists an $\varepsilon > 0$ such that for all $\mathbf{y}$ with $\|\mathbf{x} - \mathbf{y}\| \leq \varepsilon$, it holds that $f(\mathbf{y}) \geq f(\mathbf{x})$, then it must hold $\nabla f(\mathbf{x}) = 0$, because $f(\mathbf{x})$ is a local minimum of f. Thus $\nabla f(\mathbf{x}) \neq 0$ implies the existence of $\mathbf{y}$ such that $f(\mathbf{y}) < f(\mathbf{x})$.

Exercises

If you have skipped Problem 2 in Exercise Set 8, make sure to catch up on it this week.

Problem 1.

The aim of this exercise is to prove for $A \in \mathbb{R}^{n\times n}$ we have

$$A^k \to 0 \text{ as } k \to \infty \Leftrightarrow \rho(A) < 1,$$

where $\rho(A)$ denotes the spectral radius of A. Let $\|\cdot\|_2$ denote the spectral norm (also called matrix 2-norm).

(a) Show $\rho(A)^k \leq \|A^k\|_2$.
Hint: Use $A\mathbf{x} = \lambda\mathbf{x}$ and submultiplicativity.

(b) Using (a), show

$$A^k \to 0 \text{ as } k \to \infty \Rightarrow \rho(A) < 1$$

(c) Consider the $m \times m$ Jordan block

$$J_m(\lambda) = \begin{pmatrix} \lambda & 1 & & & 0 \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ 0 & & & & \lambda \end{pmatrix}$$

Show that

$$(J_m(\lambda)^k)_{ij} = \begin{cases} 0 & \textit{if } i > j \\ \lambda^k & \text{if } i = j \\ \binom{k}{l}\lambda^{k-l} & \text{if } j = i + l \end{cases}$$

where we let $\binom{k}{l} = 0$ if $l > k$.

Hint: First show what happens when $\lambda = 0$. Then use $J_m(\lambda)^k = (\lambda I_m + J_m(0))^k$.

(d) Show that if $|\lambda| < 1$ then

$$J_m(\lambda)^k \to 0 \text{ as } k \to \infty$$

(e) By considering the Jordan canonical form $A = PJP^{-1}$ show

$$A^k \to 0 \text{ as } k \to \infty \Leftarrow \rho(A) < 1$$

Solution.

(a) Let $(\lambda, \mathbf{x})$ be an eigenpair of A and let $\|\mathbf{x}\|_2 = 1$. One can immediately see that

$$A^k\mathbf{x} = \lambda^k\mathbf{x}$$

Hence, $(\lambda^k, \mathbf{x})$ is an eigenpair of A^k. Now let λ be such that $|\lambda| = \rho(A)$. Since $\|\mathbf{x}\|_2 = 1$ we have

$$\begin{aligned} 0 \leq \rho(A)^k &= |\lambda|^k \\ &= \|\lambda^k\mathbf{x}\|_2 = \|A^k\mathbf{x}\|_2 \\ &\leq \|A^k\|_2\|\mathbf{x}\|_2 = \|A^k\|_2 \end{aligned}$$

as required.

(b) Since $0 \leq \rho(A)^k \leq \|A^k\|_2 \to 0$ as $k \to \infty$ we immediately have $\rho(A)^k \to 0$ as $k \to \infty$.

(c) We should note that

$$J_m(0)^k = \begin{cases} 1 \text{ if } j = i + k \\ 0 \text{ otherwise} \end{cases}$$

which can be proven by induction. $k = 0$ is immediate since $J_m(0)^0 = I_m$. For general $k + 1 \in \mathbb{N}$ we have

$$(J_m(0)^k J_m(0))_{ij} = \sum_{l=1}^{m} \delta_{i+k,l}\delta_{l+1,j} = \begin{cases} 1 & \text{if } j = i + k + 1 \\ 0 & \text{otherwise} \end{cases}$$

because the only non-zero term in the sum occurs when $i + k = l$ and $l + 1 = j$. Now we note that

$$\begin{aligned} J_m(\lambda)^k &= (\lambda^k + J_m(0))^k \\ &= \sum_{l=0}^{k} \binom{k}{l} \lambda^{k-l} J_m(0)^l \end{aligned}$$

which implies the result.

(d) We should note that $\binom{k}{k}\lambda^{k-l} \to 0$ as $k \to \infty$. Hence, each of the terms in $J_m(0)^k$ tends to 0 as $k \to \infty$. This implies $J_m(0)^k \to 0$ as $k \to \infty$.

(e) Every matrix can be written as $A = PJP^{-1}$ where

$$J = \begin{pmatrix} J_{n_1}(\lambda_1) & & & \\ & J_{n_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{n_s}(\lambda_s) \end{pmatrix}$$

is a block diagonal matrix with diagonal blocks being Jordan blocks that corresponds to the eigenvalues of A. Note that by (d) we have

$$J^k = \begin{pmatrix} J_{n_1}(\lambda_1)^k & & & \\ & J_{n_2}(\lambda_2)^k & & \\ & & \ddots & \\ & & & J_{n_s}(\lambda_s)^k \end{pmatrix}$$

Hence, $J^k \to 0$ as $k \to \infty$ if $\rho(A) < 1$. This implies

$$A^k = PJ^kP^{-1} \to 0 \text{ as } k \to \infty$$

if $\rho(A) < 1$, as required.

Problem 2.

Consider the linear system $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

(a) Determine if Jacobi's method is guaranteed to converge.

(b) Consider the following iterative method

$$L\mathbf{x}^{(k+1)} = L\mathbf{x}^{(k)} + \delta(\mathbf{b} - A\mathbf{x}^{(k)}) \quad k \geq 0 \tag{1}$$

where $\delta > 0$ is a parameter and

$$L := \begin{pmatrix} 2 & 0 & 0 \\ -1 & 2 & 0 \\ 0 & -1 & 2 \end{pmatrix}$$

Rewrite the method (1) in the form $\mathbf{x}^{(k+1)} = B^\delta \mathbf{x}^{(k)} + \mathbf{z}_\delta, k \geq 0$, for a suitable matrix B^δ which is to be determined.

(c) Establish for which values of the parameter $\delta > 0$ the method (1) converges.

(d) Let $\delta = \frac{4}{3}$. Considering the results obtained at the point (c), establish whether the method (1) is convergent. If so, which method can be expected to converge faster between method (1) and Jacobi?

Solution.

(a) We will investigate the spectral radius of the iteration matrix B^J:

$$\begin{aligned} B^J = I - D^{-1}A &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{pmatrix} \end{aligned}$$

which has eigenvalues $\lambda_1 = 0, \lambda_{2,3} = \pm 1/\sqrt{2}$ Hence, $\rho(B^J) = 1/\sqrt{2} < 1$. We therefore conclude that the Jacobi method will converge.

(b) We have

$$\begin{aligned} L\mathbf{x}^{(k+1)} &= L\mathbf{x}^{(k)} + \delta\mathbf{b} - A\mathbf{x}^{(k)} \\ \Rightarrow L\mathbf{x}^{(k+1)} &= (L - \delta A)\mathbf{x}^{(k)} + \delta\mathbf{b} \end{aligned}$$

and observe that L is invertible. Hence, we write

$$\mathbf{x}^{(k+1)} = L^{-1}(L - \delta A)\mathbf{x}^{(k)} + \delta L^{-1}\mathbf{b} = (I - \delta L^{-1}A)\mathbf{x}^{(k)} + \delta L^{-1}\mathbf{b}.$$

We note that

$$L^{-1} = \begin{pmatrix} 1/2 & 0 & 0 \\ 1/4 & 1/2 & 0 \\ 1/8 & 1/4 & 1/2 \end{pmatrix},$$

Therefore,

$$B^\delta = \begin{pmatrix} 1-\delta & \delta/2 & 0 \\ 0 & 1 - 3\delta/4 & \delta/2 \\ 0 & \delta/8 & 1 - 3\delta/4 \end{pmatrix}.$$

(c) The eigenvalues of B^δ are the zeros of the polynomial

$$(1-\delta-\lambda)\left(\lambda^2 - 2\lambda\left(1-\frac{3}{4}\delta\right) + \left(1-\frac{3}{4}\delta\right)^2 - \frac{\delta^2}{16}\right) = 0,$$

which gives

$$\lambda_1 = 1-\delta, \quad \lambda_2 = 1-\delta, \quad \lambda_3 = 1-\delta/2.$$

Hence,

$$\rho(B^\delta) = \max\{|1-\delta|, |1-\delta/2|\}, \quad \delta > 0.$$

from which we conclude that $\rho(B^\delta) < 1$ whenever $0 < \delta < 2$.

(d) With $\delta = 4/3$, the method convergences and $\rho(B^\delta) = 1/3$; therefore it is expected to converge faster than the Jacobi method.

Problem 3. Consider the linear system $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

(a) We want to solve the system by the Gauss-Seidel method. Determine the iteration matrix B^{GS}.

(b) What can we say about the convergence of the Gauss-Seidel method?

(c) We now consider the preconditioned Richardson method

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha P^{-1}\mathbf{r}^{(k)} \quad k \geq 0$$

with $P = D$ where D is the diagonal part of A. Verify that if we take $\alpha = 1$ we find the Jacobi method.

(d) Let the starting vector be $\mathbf{x}^{(0)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and calculate the first iteration of the preconditioned Richardson method, with $P = D$ being the diagonal part of A, by choosing the optimal parameter α_{opt}.

Solution.

(a) The iteration matrix is $B^{GS} = -(D+L)^{-1}U$.

$$B^{GS} = -\begin{pmatrix} 1 & 0 \\ 2 & 5 \end{pmatrix}^{-1} \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -2 \\ 0 & \frac{4}{5} \end{pmatrix}$$

(b) The Gauss-Seidel method will converge since the eigenvalues of B^{GS} are 0 and $\frac{4}{5}$. Hence, the spectral radius of B^{GS} is less than 1.

(c) If we let $\alpha = 1$ we get

$$\begin{aligned} \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + D^{-1}(\mathbf{b} - A\mathbf{x}^{(k)}) \\ \Rightarrow D\mathbf{x}^{(k+1)} &= D\mathbf{x}^{(k)} - A\mathbf{x}^{(k)} + \mathbf{b} \\ \Rightarrow D\mathbf{x}^{(k+1)} &= -(L+U)\mathbf{x}^{(k)} + \mathbf{b} \\ \Rightarrow \mathbf{x}^{(k+1)} &= -D^{-1}(L+U)\mathbf{x}^{(k)} + \mathbf{b} \end{aligned}$$

which is the Jacobi method as required.

(d) By Theorem 5.4, since $D^{-1}A$ has positive real eigenvalues, we take $\alpha_{\text{opt}} = \frac{2}{\lambda_{\min}+\lambda_{\max}}$, $\lambda_{\min}$ and $\lambda_{\max}$ being the smallest and the greatest eigenvalue of $D^{-1}A$. In our case

$$D^{-1}A = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{5} \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ \frac{2}{5} & 1 \end{pmatrix}$$

and the eigenvalues are $\lambda_{\pm} = 1 \pm \frac{2}{\sqrt{5}} > 0$. Thus $\alpha_{\text{opt}} = 1$. The first iterate is $\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \alpha_{\text{opt}} D^{-1}\mathbf{r}^{(0)}$. Therefore

$$\mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)} = \begin{pmatrix} 0 \\ -1 \end{pmatrix} - \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -3 \\ -8 \end{pmatrix}$$

and

$$D^{-1}\mathbf{r}^{(0)} = \begin{pmatrix} -3 \\ -\frac{8}{5} \end{pmatrix}$$

which gives

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \alpha_{\text{opt}} D^{-1}\mathbf{r}^{(0)} = \begin{pmatrix} -2 \\ -\frac{3}{5} \end{pmatrix}.$$

Problem 4.

The aim of this exercise is to prove that the iterates of the Gauss-Seidel method applied to a strictly diagonally dominant matrix A converge to the solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$.

(a) Recall that the error of the Gauss-Seidel iteration can be written as

$$\mathbf{e}_i^{(k+1)} = -\sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} \mathbf{e}_j^{(k+1)} - \sum_{j=i+1}^{n} \frac{a_{ij}}{a_{ii}} \mathbf{e}_j^{(k)}$$

where $\mathbf{e}^{(k)} := \mathbf{x}^{(k)} - \mathbf{x}$.Using this, show that there exists an index $p \in \{1, \cdots, n\}$ such that

$$\left(1 - \sum_{j=1}^{p-1} \left| \frac{a_{pj}}{a_{pp}} \right| \right) \|\mathbf{e}^{(k+1)}\|_\infty \leq \left(\sum_{j=p+1}^{n} \left| \frac{a_{pj}}{a_{pp}} \right| \right) \|\mathbf{e}^{(k)}\|_\infty$$

(b) Using that A is strictly diagonally dominant, show that there exists some $\beta \in (0, 1)$ such that

$$\|\mathbf{e}^{(k+1)}\|_\infty \leq \beta \|\mathbf{e}^{(k)}\|_\infty$$

and conclude that the Gauss-Seidel method converges.

Solution.

(a) Let $p \in \{1, \cdots, n\}$ be such that $|\mathbf{e}_p^{(k+1)}| = \|\mathbf{e}^{(k+1)}\|_\infty$ and $l \in \{1, \cdots, n\}$ such that $|\mathbf{e}_l^{(k)}| = \|\mathbf{e}^{(k)}\|_\infty$. Then,

$$\begin{aligned} \|\mathbf{e}^{(k+1)}\|_\infty = |\mathbf{e}_p^{(k+1)}| &\leq \sum_{j=1}^{p-1} \left| \frac{a_{pj}}{a_{pp}} \right| |\mathbf{e}_j^{(k+1)}| + \sum_{j=p+1}^{n} \left| \frac{a_{pj}}{a_{pp}} \right| |\mathbf{e}_j^{(k)}| \\ &\leq \|\mathbf{e}^{(k+1)}\|_\infty \sum_{j=1}^{p-1} \left| \frac{a_{pj}}{a_{pp}} \right| + \|\mathbf{e}^{(k)}\|_\infty \sum_{j=p+1}^{n} \left| \frac{a_{pj}}{a_{pp}} \right| \end{aligned}$$

Hence,

$$\left(1 - \sum_{j=1}^{p-1} \left|\frac{a_{pj}}{a_{pp}}\right|\right) \|\mathbf{e}^{(k+1)}\|_\infty \leq \left(\sum_{j=p+1}^{n} \left|\frac{a_{pj}}{a_{pp}}\right|\right) \|\mathbf{e}^{(k)}\|_\infty$$

as required.

(b) Let $S_1 := \sum_{j=1}^{p-1} \left|\frac{a_{pj}}{a_{pp}}\right|$, $S_2 := \sum_{j=p+1}^{n} \left|\frac{a_{pj}}{a_{pp}}\right|$ and $S_3 := a_{pp}$. By diagonal dominance we have

$$S_1 + S_2 < S_3 \Rightarrow \beta := \frac{S_2/S_3}{1 - S_1/S_3} \in (0, 1)$$

and from (a) we have

$$\|\mathbf{e}^{(k+1)}\|_\infty \leq \frac{S_2/S_3}{1 - S_1/S_3} \|\mathbf{e}^{(k)}\|_\infty = \beta \|\mathbf{e}^{(k)}\|_\infty$$

as required.