

University of Vienna Faculty of Mathematics

Numerical Analysis Problems

Milutin Popovic

May 7, 2022

Contents

1	Sheet 5	1
1.1	Problem 1	1
1.2	Exercise 2	2
1.3	Exercise 3	2
1.3.1	Exercise 4	4

1 Sheet 5

1.1 Problem 1

Let $A \in \mathbb{R}^{n \times n}$ be an SPD matrix with an additive decomposition $A = D + L + L^T$, where D consists of the diagonal entries of A , L and L^T of the lower diagonal and upper-diagonal entries of A respectively. For $\omega \in (0, 2)$ the SSOR preconditioner is defined as follows

$$C_\omega = \frac{1}{2-\omega} \left(\frac{1}{\omega} D + L \right) \left(\frac{1}{\omega} D \right)^{-1} \left(\frac{1}{\omega} D + L^T \right). \quad (1)$$

We can rewrite $C_\omega = K K^T$, where K is an invertible lower-triangular matrix by a simple splitting of the diagonal entries $D = D^{\frac{1}{2}} D^{\frac{1}{2}}$. We get

$$C_\omega = \frac{1}{2-\omega} \left(\frac{1}{\omega} D + L \right) \omega D^{-1} \left(\frac{1}{\omega} D + L^T \right) \quad (2)$$

$$= \frac{1}{2-\omega} \left(D^{\frac{1}{2}} + \omega L D^{-\frac{1}{2}} \left(D^{\frac{1}{2}} + \omega L D^{-\frac{1}{2}} \right) \right)^T = K K^T, \quad (3)$$

where

$$K := \frac{1}{\sqrt{1-\omega}} \left(D^{\frac{1}{2}} + \omega L D^{-\frac{1}{2}} \right) \quad (4)$$

The Matrix C_ω is a good approximation for the inverse of A because

$$H_\omega^{\text{SSOR}} = I - C_\omega^{-1} A \quad (5)$$

$$\rho(H_\omega^{\text{SSOR}}) < 1 \quad (6)$$

for a right choice of ω . Furthermore the general idea is that we multiply the system $Ax = b$ with a preconditioning matrix $P \in \text{GL}_n(\mathbb{R})$ with inevitability, then we get the system

$$\underbrace{PA}_{\approx I} x = Pb \quad (7)$$

$$(8)$$

The thing is that $\text{cond}_2(A) = v(n)$ bound by the curse of dimensionality and $\text{cond}_2 = s \ll v(n)$ not dependent and thereby P would be an optimal preconditioner.

1.2 Exercise 2

Let $m, n \in \mathbb{N}$, $I \in \mathbb{R}^{m \times m}$ the identity in $\mathbb{R}^{m \times m}$ and Q be the banded matrix

$$Q = \begin{pmatrix} 4 & -1 & & & \\ -1 & 4 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 4 & -1 \\ & & & & 4 \end{pmatrix} \quad (9)$$

The eigenvalues of the matrix Q lie in $\sigma(A) \subset [4-2, 4+2]$ by the Gershgorin disk theorem. Since no eigenvalue is 0, then Q is invertible. Now consider the matrix $A \in \mathbb{R}^{nm \times nm}$

$$Q = \begin{pmatrix} Q & -I & & & \\ -I & Q & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & Q & -I \\ & & & & Q \end{pmatrix} \quad (10)$$

We can consider the separation wrt. addition $A = D + L + L^T$ (Like in Exercise 1). The Jacobi-Method iteration matrix is $J = -D^{-1}(L + L^T)$, where L is the lower triangular with **-1 or 0 entries**. Further more the Gershgorin theorem states that $\sigma(A) < 1$. All in all the matrix $I - J$ is by the geometric(Neumann) series

$$(I - J)^{-1} = \sum_{n=0}^{\infty} J^n \quad (11)$$

and we have the identity

$$J = I - D^{-1}A \Rightarrow (I - J)^{-1} = DA^{-1}. \quad (12)$$

Thereby the sum transforms to

$$A^{-1} = D^{-1} \sum_{n=0}^{\infty} J^n. \quad (13)$$

The entries of D are all 4 and thereby non-negative, the matrix is also invertible. The matrix L has only -1 or 0 entries which get compensated with the minus sign in $J = -D^{-1}(L + L^T) = D^{-1}(-L - L^T)$, thereby all entries of J^k are positive for all k . Finally we arrive at the conclusion, that all entries of A^{-1} are non-negative and A is a M -matrix or '(inverse) monotone' matrix.

1.3 Exercise 3

Let $A \in \mathbb{R}^{n \times n}$ be an SPD matrix and $b \in \mathbb{R}^n$ be a right hand side of a linear system. Suppose we apply the CG method for solving $Ax = b$. The k -th iterate x_k of the CG method then satisfies the A -norm optimality condition

$$\|x_k - x\|_A = \min_{y \in x_0 + B_k} \|y - x\|_A, \quad (14)$$

where

$$B_k = \text{span}\{p_0, \dots, p_{k-1}\} = \text{span}\{r_0, Ar_0, \dots, A^{k-1}r_0\} \quad (15)$$

is the Krylov space. The search directions p_k form an A -orthogonal system.

Now if the spectrum of A , $\sigma(A) = [a, b] \subset (0, \infty)$ then for any polynomial $p \in \mathbb{P}_k^{0,1} := \{p \in \mathbb{P} : p(0) = 1\}$ we have that

$$\|x_k - x\|_A \leq \left(\sup_{t \in [a, b]} |p(t)| \right) \|x_0 - x\|_A. \quad (16)$$

To show this we have that for all $y \in x_0 + B_k$ the representation

$$y = \sum_{j=0}^{k-1} c_j A^j r_0 = x_0 + q_y(A) r_0 \quad (17)$$

for suitable c_j 's and a polynomial $q_y \in \mathbb{P}_{k-1}$. Now

$$y - x = x_0 + x - q_y(A) r_0 = x_0 - x + q_y(A) (b - Ax_0) \quad (18)$$

$$= x_0 - x + q_y(A) (Ax - Ax_0) \quad (19)$$

$$= \underbrace{(I - q_y(A)A)}_{=: p_y(A) \in \mathbb{P}_k^{0,1}} (x - x_0). \quad (20)$$

With this information we may consider the norm

$$\|x_k - x\|_A \leq \|p_y(A)(x - x_0)\|_A \quad \forall y \in x_0 + B_k. \quad (21)$$

Now we use the fact that A is SPD thereby there is an orthogonal matrix Q diagonalizing $A = Q^T \Lambda Q$, with $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ consisting of eigenvalues of A , then

$$A^k = Q^T \Lambda Q \dots Q^T \Lambda Q = Q^T \Lambda^k Q \quad (22)$$

With this we can transform the polynomial $p_y(A)$ and the geometric (Neumann) series

$$p_y(A) = \sum_{j=0}^{\infty} c_j A^j = Q^T \left(\sum_{j=0}^{\infty} c_j \Lambda^j \right) \quad (23)$$

The norm becomes then

$$\|p(A)(x - x_0)\|_a^2 = \langle Ap(A)(x - x_0), p(A)(x - x_0) \rangle = \quad (24)$$

$$= \langle Q^T \Lambda Q Q^T p(\Lambda) Q(x - x_0), Q^T p(\Lambda) Q(x - x_0) \rangle = \quad (25)$$

$$= \langle Q^T \Lambda p(\Lambda) Q(x - x_0), Q^T p(\Lambda) Q(x - x_0) \rangle = \quad (26)$$

$$= \langle \Lambda p(\Lambda) Q(x - x_0), p(\Lambda) Q(x - x_0) \rangle = \quad (27)$$

$$= \langle \Lambda^{\frac{1}{2}} p(\Lambda) Q(x - x_0), \Lambda^{\frac{1}{2}} p(\Lambda) Q(x - x_0) \rangle = \quad (28)$$

$$= \|\Lambda^{\frac{1}{2}} p(\Lambda) Q(x - x_0)\|_2 \quad (29)$$

$$= \|p(\Lambda) \Lambda^{\frac{1}{2}} Q(x - x_0)\|_2 \quad (30)$$

$$\leq \|p(\Lambda)\|_2 \|\Lambda^{\frac{1}{2}} Q(x - x_0)\|_2. \quad (31)$$

The Norm of the polynomial is the maximal eigenvalue thereby

$$\|p(\Lambda)\|_2 = \max_{\lambda \in \sigma(A)} |p(\lambda)| \leq \sup_{t \in [a, b]} |p(t)|, \quad (32)$$

we can do the supremum boundary because $\lambda \in [a, b]$. As for the second part

$$\|\Lambda^{\frac{1}{2}} Q(x - x_0)\|_2^2 = (x - x_0)^T Q^T \Lambda^{\frac{1}{2}} \Lambda^{\frac{1}{2}} Q(x - x_0) \quad (33)$$

$$= (x - x_0)^T A(x - x_0) \quad (34)$$

$$= \|x - x_0\|_A^2. \quad (35)$$

And finally we get the result

$$\|x_k - x\| \leq \sup_{t \in [a, b]} |p(t)| \|x - x_0\|_A^2. \quad (36)$$

The last approximation can be done because $\sup_{t \in [a, b]} |p(t)|$ holds for **all** $p \in \mathbb{P}_k^{0,1}$ thereby we can bound by an infimum over all the polynomials in $\mathbb{P}_k^{0,1}$ and we get

$$\sup_{t \in [a, b]} |p(t)| \leq \inf_{p \in \mathbb{P}_k^{0,1}} \|p\|_{C([0,1])}. \quad (37)$$

1.3.1 Exercise 4

We can do subsequently the as in the last exercise with the GMRES method. So we let $A \in \mathbb{R}^{n \times n}$ be an SPD and $b \in \mathbb{R}^n$ be the right hand side of the linear system. The iterates of x_k of the CG method satisfy the A^{-1} -norm optimality

$$\|Ax_k - b\|_{A^{-1}} = \min_{y \in x_0 + C_k} \|Ay - b\|_{A^{-1}}, \quad (38)$$

with $C_k = \text{span}\{p_0, Ap_0, \dots, A^{k-1}p_0\}$. The ‘generalized minimal residual’, short GMRES method, instead, formally constructs a sequence of iterates x_k^G by

$$\|Ax_k^G - b\|_2 = \min_{y \in x_0 + C_k} \|Ay - b\|_2. \quad (39)$$

The GMRES method allows for an error inequality similar to the one observed in the CG method

$$\|Ax_k^G - b\|_2 \leq \inf_{p \in \mathbb{P}_k^{0,1}} \|p(A)\|_2 \|Ax_0 - b\|_2. \quad (40)$$

To show this we start off by minimizing over a $z \in C_k$

$$\|Ax_k^G - b\|_2 = \min_{y \in x_k + C_k} \|Ay - b\|_2 = \min_{z \in C_k} \|Az + Ax_0 - b\|_2. \quad (41)$$

Then for all $z \in C_k$, there exists a $\pi_k \in \mathbb{P}_{k-1}$ such that

$$z = \pi_k(A)p_0 = \pi_k(A)r_0, \quad (42)$$

Then the minimization can be bounded

$$\min_{z \in C_k} \|Az + Ax_0 - b\|_2 = \min_{\pi_k \in \mathbb{P}_k} \|A\pi_k(A)r_0 + Ax_0 - b\|_2 \quad (43)$$

$$\leq \|A\pi_k(A)(b - Ax_0) + Ax_0 - b\|_2 \quad (44)$$

$$= \|(Ax_0 - b) \underbrace{(I - A\pi_k(A))}_{=: p \in \mathbb{P}_k^{0,1}}\|_2 \quad (45)$$

$$= \|(Ax_0 - b)p(A)\| \quad (46)$$

$$\leq \|p(A)\|_2 \|Ax_0 - b\|_2. \quad (47)$$

Following the same argumentation as in Exercise 3 we get for the norm of the polynomial

$$\|p(A)\|_2 = \max_{\lambda \in \sigma(A)} |p(\lambda)| \quad (48)$$

$$\leq \sup_{t \in [a, b]} |p(t)| \quad (49)$$

$$\leq \inf_{p \in \mathbb{P}_k^{0,1}} \|p\|_{C([a, b])}. \quad (50)$$