Norwegian University of Science and Technology
Department of Mathematical Sciences

TMA4205 Numerical Linear Algebra Fall 2016

Solutions to exercise set 3

1 **a)** We are given the matrix

$$A = \begin{bmatrix} 1 & -6 & 0 \\ 6 & 2 & 3 \\ 0 & 3 & 2 \end{bmatrix}.$$

Gershgorin's theorem states that all the eigenvalues of an $n \times n$ matrix A are located in one of the closed discs of the complex plane centered in $a_{i,i}$ having radius

$$r_i = \sum_{\substack{j=1 \\ j \neq i}}^{j=n} |a_{i,j}|, \quad i = 1, \ldots, n.$$

See Saad, Theorem 4.6. For our matrix, this means that the eigenvalues lie within circles centered at (1,0), (2,0) and (2,0) with radii 6, 9 and 3, respectively. This is illustrated in Figure 1. The lightly shaded square encapsulates all the circles, and thus also the eigenvalues. For our given matrix, the spectrum is $\sigma(A) = \{2.328, 1.336 \pm 5.196\text{i}\}$. We see that this is in agreement with the estimate.

Alternatively, one may use an estimate provided by Theorem 1.35 in Saad. That is, let us define the symmetric part of A

$$H = (A + A^{\mathrm{T}})/2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 3 \\ 0 & 3 & 2 \end{bmatrix}$$

and an anti-symmetric one

$$S = (A - A^{\mathrm{T}})/(2i) = \begin{bmatrix} 0 & 6i & 0 \\ -6i & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

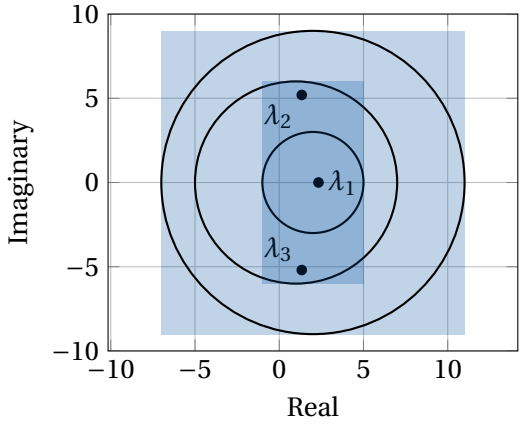

Figure 1: Exact and estimated eigenvalues of A.

Eigenvalues of H are $\{-1,1,5\}$ and eigenvalues of S are $\{-6,0,6\}$. Therefore, all eigenvalues of A lie in the rectangle $[-1,5]\times[-6,6]$, which provides a much sharper estimate (at additional computational cost) in this case; see Figure 1.

b) From chapter 5.3.2 in Saad, we know that the MR iteration will converge if A is positive definite. Gershgorin's theorem tells us nothing useful for determining positive definiteness in this case - even if A were symmetric and hence its eigenvalues real, we would not know whether they were positive.
Instead, we check whether the symmetric part $H=(A+A^T)/2$ is positive definite. This is equivalent to A being positive definite since with the symmetric-antisymmetric splitting

$$A=\frac{A+A^T}{2}+\frac{A-A^T}{2}=H+S,$$

we have that, since $x^TSx=0$ for all x,

$$x^TAx>0 \Leftrightarrow x^THx>0.$$

But, as seen in part a), H has eigenvalues $\{-1,1,5\}$, meaning it is not positive definite. Thus A is not positive definite either and we cannot guarantee convergence.

2 We revisit the one-dimensional Poisson problem of exercise set 1, and we consider solving the discretized problem using Jacobi -, steepest descent (SD) -, and minimum residual (MR) iteration.

a) In general, if we have an error/residual behaviour given by

$$\|e_k\|=\rho^k\|e_0\|,$$

so

$$\log\frac{\|e_k\|}{\|e_0\|}=k\log\rho.$$

For this exercise we want to obtain an error/residual reduction by 10^{-5} such that $\log(\|e_k\|/\|e_0\|)=-5$, i.e.,

$$-5=k\log\rho. \tag{1}$$

We will need the following useful approximations based on Maclaurin expansions:

$$\begin{aligned}\cos(x)&\approx 1-\tfrac{1}{2}x^2,\\ \log(1+x)&\approx x,\\ (1+x)^{1/2}&\approx 1+\tfrac{1}{2}x.\end{aligned}$$

Jacobi. For this case we have that ρ is the spectral radius of the iteration matrix, so $\rho=\cos(\pi/n)\approx 1-\pi^2/2n^2$. From (1) and the approximation of the logarithm, we find that $k\approx 10n^2/\pi^2$.

Steepest descent. From the lectures, we know that

$$\|e_k\|_A \leq \rho^k \|e_0\|_A,$$

where

$$\rho = \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} = \frac{\kappa - 1}{\kappa + 1} \approx 1 - \frac{2}{\kappa}$$

for large values of $\kappa = \lambda_{\max}/\lambda_{\min}$, the condition number based on the Euclidean norm of A. From (1) and the approximation of the logarithm, we find that $k \approx 5\kappa/2$. We also have that

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}} = \frac{\frac{2}{h^2}\left(1 - \cos\left(\frac{(n-1)\pi}{n}\right)\right)}{\frac{2}{h^2}\left(1 - \cos\left(\frac{\pi}{n}\right)\right)} \approx \frac{2}{\pi^2/2n^2} = \frac{4n^2}{\pi^2} \quad \Longrightarrow \quad k \approx \frac{10n^2}{\pi^2}. \tag{2}$$

b) We now consider MR iteration for the same problem, and wish to reduce the initial residual with 5 orders of magnitude. In this case,

$$\|r_k\|_2 \leq \rho^k \|r_0\|_2$$

with

$$\rho = \left(1 - \frac{\mu^2}{\sigma^2}\right)^{1/2},$$

where

$$\mu = \lambda_{\min}\left(\frac{A + A^{\mathrm{T}}}{2}\right), \qquad \sigma = \|A\|_2.$$

In our case, A is symmetric and positive-definite (SPD), and we thus get the simplified expressions

$$\mu = \lambda_{\min}(A), \qquad \sigma = \lambda_{\max}(A).$$

Hence,

$$\rho = \left(1 - \frac{\lambda_{\min}^2}{\lambda_{\max}^2}\right)^{1/2} = \left(1 - \frac{1}{\kappa^2}\right)^{1/2} \approx 1 - \frac{1}{2\kappa^2}.$$

From (1) and the approximation of the logarithm, we find that $k \approx 10\kappa^2$. We insert the expression for the condition number from (2), and find

$$k \approx \frac{160n^4}{\pi^4}.$$

We may now compare the value for k for Jacobi, SD and MR, and observe that

$$\frac{k_{\mathrm{SD/J}}}{k_{\mathrm{MR}}} \approx \frac{1}{4\kappa} = \frac{\pi^2}{16n^2}.$$

Here, $k_{\mathrm{SD/J}}$ is the number of iterations for Jacobi and steepest descent to reduce the error by 5 orders of magnitude, while k_{MR} is the number of iterations for MR to reduce the initial residual by the same amount. Since we expect κ to be large for large n, this difference is significant!

c) We now discuss the computational cost for the three iterative methods. In Table 1a we have listed the number of matrix-vector operations (vector addition and scalar multiplication) and the number of inner-products for one iteration of each method. How many floating point operations this is equivalent to, depends on the matrix A. For our Poisson problem, A is tridiagonal, and a matrix-vector product can be

Method	Ax	αx and $x+y$	$x^{\mathrm{T}}x$
J	1	1	?
SD	1	4	2
MR	1	4	2

(a) Computational cost

Method	A	x
J	1	2
SD	1	3
MR	1	3

(b) Memory requirement

Table 1: One iteration of Jacobi, SD and MR

Method	$\mathcal{N}_{\mathrm{ops}}^{\mathrm{tot}}$
J	n^3
SD	n^3
MR	n^5

Table 2: Number of floating point operations for solving the one-dimensional Poisson problem using Jacobi, SD and MR.

done in $\mathrm{O}(n)$ operations. (If A was a full matrix this would require $\mathrm{O}(n^2)$ operations.) The vector operations and inner-products also use $\mathrm{O}(n)$ operations, so the number of operations for one iteration for all three methods is $\mathcal{N}_{\mathrm{ops}}^{1} \sim \mathrm{O}(n)$. Note that in a Jacobi method there is no error or residual estimate available, so the question mark in the last column is for potential error/residual estimation.

In Table 1b we indicate the memory requirement. All methods need to store enough information about A to be able to perform matrix-vector products. The sparsity of A should here be exploited, and we thus only need to store the non-zero entries, which in this case is $\mathrm{O}(n)$ (in fact, the non-zero entries in each row are the same). Both SD and MR need to store three vectors x, p and r. For Jacobi we need to store x as well as a working array. Thus, the memory requirement is $\mathrm{O}(n)$ for all methods.

We now consider the cost for k iterations. The memory requirement remains the same, while the number of operations is given by

$$\mathcal{N}_{\mathrm{ops}}^{\mathrm{tot}} = k\mathcal{N}_{\mathrm{ops}}^{1},$$

where $\mathcal{N}_{\mathrm{ops}}^{1}$ is the number of floating point operations in each iteration. Using the results from **a)** and **b)**, we find the estimates in Table 2.

d) Of the three iterative methods considered here, MR is the most general one, since the requirement is only that A is positive-definite. However, we see that this method is much slower than the other two for the one-dimensional Poisson problem. Steepest descent is guaranteed to converge as long as A is SPD, while Jacobi iteration has an even stronger requirement, namely that the spectral radius of the iteration matrix must be less than 1 (which is not the case for all SPD matrices). For our particular Poisson problem, Jacobi is a little bit faster than the steepest descent, while Jacobi and steepest descent are both much faster than MR iteration.

In conclusion, a method for more general problems is typically slower than a specialized method, and for our Poisson problem, MR iteration is definitely not a good idea. We also note here that we have compared *error* reduction by 5 orders of magnitude for the first two methods, and *residual* reduction for the MR method. This is obviously not the same, but we assume that they are comparable and show the same behaviour.