AMath 585
Homework #5
Due Thursday, March 5, 2020

Homework is due to Canvas by 11:00pm PDT on the due date.

To submit, see `https://canvas.uw.edu/courses/1352870/assignments/5268734`

Problem 1.

Suppose $A \in \mathbb{R}^{m\times m}$ is strictly row diagonally dominant, i.e.,

$$\sum_{j\neq i} |a_{ij}| < |a_{ii}|, \quad \text{for } i = 1,\ 2,\ \ldots,\ m. \tag{1}$$

where the sum goes from $j = 1$ to m omitting the diagonal term.

Let G be the Jacobi iteration matrix for this matrix A.

(a) Show that $\|G\|_\infty < 1$ and hence the Jacobi iteration converges, since $\|e_k\|_\infty \leq \|G\|_\infty^k \|e_0\|_\infty$. (Here and below subscript k refers to the kth iteration).

(b) Use the Gershgorin Theorem of Appendix C.8 to show that we also have $\rho(G) < 1$ in the case when (1) holds.

(c) Suppose G is a normal matrix ($G^TG = GG^T$) and suppose something stronger than (1) holds, namely,

$$\sum_{j\neq i} |a_{ij}| \leq \beta |a_{ii}|, \quad \text{for } i = 1,\ 2,\ \ldots,\ m \tag{2}$$

for $\beta = 1/2$. Show that in this case Jacobi would reduce the 2-norm of the error $\|e_k\|_2$ by a factor of at least 10^6 in only 20 iterations, i.e., $\|e_k\|_2 \leq 10^{-6}\|e_0\|_2$.

(d) More generally, suppose the factor β in (4) is some value satisfying $\beta < 1$, and suppose the matrix is not necessarily normal but we have an upper bound $\tilde{\kappa}$ on the 2-norm condition number of R, the matrix of right eigenvectors of G. In terms of β and $\tilde{\kappa}$, what is the maximum number of iterations that would be required to reduce the 2-norm error by a factor of 10^6? (In exact arithmetic.)

Solution: a) $A \in \mathbb{R}^{m\times m}$ is strictly row diagonally dominant, i.e.,

$$\sum_{j\neq i} |a_{ij}| < |a_{ii}|, \quad \text{for } i = 1,\ 2,\ \ldots,\ m. \tag{3}$$

where the sum goes from $j = 1$ to m omitting the diagonal term.

Let G be the Jacobi iteration matrix for this matrix A.

Let $A = M - N$ such M is the diagonal part of A and N is the off-diagonal part of A.

Thus,

$$M = \begin{bmatrix} a_{11} & & \\ & \ddots & \\ & & a_{mm} \end{bmatrix}$$

and,

$$N = \begin{bmatrix} 0 & -a_{12} & -a_{13} & \dots & a_{1m} \\ -a_{21} & 0 & -a_{23} & \dots & -a_{2m} \\ \vdots & & & \ddots & \\ -a_{m1} & \dots & \dots & -a_{m,m-1} & 0 \end{bmatrix}$$

Then, let $G = M^{-1}N$. This is going to be the Jacobi iteration matrix of A.

Notice: M is just a diagonal matrix and so M^{-1} is going to be a diagonal matrix with entries $\frac{1}{a_{ii}}$ for $i = 1, 2, \dots, m$. Pre-multiplying N by M^{-1} will just scale the rows of N by $\frac{1}{a_{ii}}$.

Therefore, $\|G\|_\infty$ = maximum row sum, that is,

$$\|G\|_\infty = \max_i \frac{1}{|a_{ii}|} \sum_j |a_{ij}| = \max_i \frac{1}{|a_{ii}|} \sum_{j \neq i} |a_{ij}|$$

since when $j = i$, $a_{ii} = 0$ for matrix N because of the way we defined it.

A is strictly row diagonally dominant, that is,

$$\sum_{j \neq i} |a_{ij}| < |a_{ii}| \implies \frac{1}{|a_{ii}|} \sum_{j \neq i} |a_{ij}| < 1, \quad \text{for } i = 1,\ 2,\ \dots,\ m.$$

So,

$$\max_i \frac{1}{|a_{ii}|} \sum_{j \neq i} |a_{ij}| < 1 \implies \|G\|_\infty < 1$$

Hence the Jacobi iteration converges, since $\|e_k\|_\infty \leq \|G\|_\infty^k \|e_0\|_\infty$.

b)

We will use the Gershgorin Theorem to show that we also have $\rho(G) < 1$.

In this case, G is "nearly diagonal" in the sense that the off-diagonal elements are small compared to the diagonal, then we expect diagonal elements to be good approximations to the eigenvalues. The Gerschgorin theorem quantifies this and provides bounds on the eigenvalues in terms of the diagonal and off-diagonal elements.

Recall: $G = M^{-1}N$ where M is just the diagonal part of A and so M^{-1} is a diagonal matrix with the diagonal entries being the reciprocals of the diagonal entries of A. Then, G is the matrix N in part a) with the rows scaled by the reciprocals of the diagonal entries of A.

Let λ_p be the eigenvalues of G for $p = 1, 2, \dots, m$.

Let

$$r_i = \sum_{j \neq i} |\frac{a_{ij}}{a_{ii}}| = \frac{1}{|a_{ii}|} \sum_{j \neq i} |a_{ij}|$$

and define closed disks in the complex plane

$$D_i = \{z \in \mathbb{C} : |z - G_{ii}| \leq r_i\} = \{z \in \mathbb{C} : |z| \leq r_i\} \text{ for } i = 1, 2, \dots, m$$

since the diagonals of G is equal to zero.

Gershgorin's theorem tells us that all of the eigenvalues of A lie in the union of these disks D_i.

Notice: All the disks are centered around the same point 0. So, the union of the disks will just be the disk of maximum radius, that is, of radius:

$$\max_i \frac{1}{|a_{ii}|} \sum_{j \neq i} |a_{ij}|$$

Notice:

$$\sum_{j \neq i} |a_{ij}| < |a_{ii}| \text{ for } i = 1, 2, \ldots, m \implies \max_i \frac{1}{|a_{ii}|} \sum_{j \neq i} |a_{ij}| < 1$$

Therefore, all the eigenvalues of A lie on a disk of radius less than 1 centered around 0 on the complex plane, that is, $\lambda_p \in \{z \in \mathbb{C} : |z| < 1\}$ for $p = 1, 2, \ldots, m$. So, $|\lambda_p| < 1$ for $p = 1, 2, \ldots, m$ and therefore,

$$\rho(G) = \max_p |\lambda_p| < 1$$

.

in the case when (1) holds.

c) Suppose G is a normal matrix ($G^T G = G G^T$) and suppose something stronger than (1) holds, namely,

$$\sum_{j \neq i} |a_{ij}| \leq \beta |a_{ii}|, \quad \text{for } i = 1,\ 2,\ \ldots,\ m \tag{4}$$

for $\beta = 1/2$, that is,

$$\sum_{j \neq i} |a_{ij}| \leq \frac{1}{2} |a_{ii}|.$$

We first aim to bound the spectral radius, $\rho(G)$ using Gershgorin's theorem.

Let

$$r_i = \sum_{j \neq i} \left| \frac{a_{ij}}{a_{ii}} \right| = \frac{1}{|a_{ii}|} \sum_{j \neq i} |a_{ij}|$$

.

By Gershgorin's theorem, the eigenvalues λ_p of G are in the union of the disks

$$D_i = \{z \in \mathbb{C} : |z| \leq r_i\}$$

Again, since the disks are centered around the same point we will have that all the eigenvalues are in the disk of maximum radius, that is,

$$\max_i r_i = \max_i \frac{1}{|a_{ii}|} \sum_{j \neq i} |a_{ij}| \leq \beta = \frac{1}{2}$$

because of (4).

Therefore, all eigenvalues are in the disk $\{z \in \mathbb{C} : |z| \leq \beta = \frac{1}{2}\}$. So, $|\lambda_p| \leq \beta = \frac{1}{2}$ for $p = 1, 2, \ldots, m$. Thus,

$$\max_p \lambda_p \leq \beta = \frac{1}{2}$$

.

So, $\rho(G) \leq \beta = \frac{1}{2}$ when (4) holds.

Since, G is normal, the eigenvectors are mutually orthogonal. We can chose R, the matrix of right eigenvectors of G, to be an orthogonal matrix by normalizing the eigenvectors so that $R^{-1} = R^T$. The eigenvectors are mutually orthogonal so $\|R\|_2 = \|R^{-1}\|_2 = 1 \implies \kappa_2(R) = 1$.

We know from class that $\|e_k\|_2 \leq \kappa_2(R) \times \rho(G)^k \times \|e_0\|_2 = \rho(G)^k \times \|e_0\|_2$.

Therefore, $\|e_{20}\|_2 \leq \rho(G)^{20} \times \|e_0\|_2$

Since, $\rho(G) \leq \frac{1}{2} \implies \rho(G)^{20} \leq \frac{1}{2^{20}} \approx 9.54 \times 10^{-7} \approx 10^{-6}$.

Finally, $\|e_{20}\|_2 \leq \rho(G)^{20} \times \|e_0\|_2 \leq 10^{-6}\|e_0\|_2$.

d)

Suppose the factor β in (4) is some value satisfying $\beta < 1$, and suppose we have an upper bound $\tilde{\kappa}$ on the 2-norm condition number of the matrix of right eigenvectors of G.

We want to find the maximum number of iterations that would be required to reduce the 2-norm error by a factor of 10^6.

We know $\|e_k\|_2 \leq \kappa_2(R)\rho(G)^k\|e_0\|_2$ and we want:

$$\|e_k\|_2 \leq 10^{-6}\|e_0\|_2$$

.

We know:

$$\kappa_2(R) \leq \bar{\kappa} \implies \kappa_2(R)\rho(G)^k \leq \bar{\kappa}\rho(G)^k$$

.

Also, $\rho(G) \leq \beta$ from our analysis above which implies $\rho(G)^k \leq \beta^k$.

Therefore,

$$\bar{\kappa}\rho(G)^k \leq \bar{\kappa}\beta^k = 10^{-6}$$

.

Taking log on both sides:

$$\log(\bar{\kappa}\beta^k) = \log(10^{-6})$$

Using log properties:

$$\log(\bar{\kappa}) + k\log(\beta) = \log(10^{-6})$$

Thus:

$$k = \frac{\log(10^{-6}) - \log(\bar{\kappa})}{\log(\beta)} = \frac{\log(\frac{10^{-6}}{\bar{\kappa}})}{\log(\beta)}$$

This is the maximum number of iterations that would be required to reduce the 2-norm error by a factor of 10^{-6}.

Problem 2.

Based on Problem 1, you might think that the more diagonally dominant a matrix is, the better in terms of convergence rate. However, consider these two matrices:

$$A_1 = \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix}, \qquad A_2 = \begin{bmatrix} 1 & -3/4 \\ -1/12 & 1 \end{bmatrix}.$$

Determine the values β for each case (as in (4), and also the asymptotic convergence rates $\rho(G_1)$ and $\rho(G_2)$ in each case, where G_i is the Jacobi iteration matrix for A_i.

Solution:

a)

$$A_1 = \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix},$$

$$a_{11} = 1, \quad a_{12} = -\frac{1}{2} \implies |a_{12}| = \frac{1}{2}$$

Therefore,

$$\sum_{j \neq 1} |a_{1j}| = |a_{12}| = \frac{1}{2} \leq \frac{1}{2}|a_{11}|$$

Similarly,

$$a_{22} = 1, \quad a_{21} = -\frac{1}{2} \implies |a_{21}| = \frac{1}{2}$$

.

So,

$$\sum_{j \neq 2} |a_{2j}| = |a_{21}| = \frac{1}{2} \leq \frac{1}{2}|a_{22}|$$

Therefore,

$$\sum_{j \neq i} |a_{ij}| \leq \frac{1}{2}|a_{ii}|$$

for $i = 1, 2$.

Hence, $\beta = \frac{1}{2}$.

Let us split $A_1 = M_1 - N_1$ in a way such that M_1 is the diagonal part of A_1.

So,

$$M_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Then, $N_1 = M_1 - A_1$.

So,

$$N_1 = \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix}$$

We now want to find the eigenvalues of $G_1 = M_1^{-1} N_1$.

$$G_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix}$$

The eigenvalues, λ, of G_1 will satisfy the characteristic equation:

$$(\lambda)^2 - \frac{1}{4} = 0$$

$$\implies \lambda^2 = \frac{1}{4}$$

$$\implies \lambda = \pm\sqrt{\frac{1}{4}}$$

$$\implies \lambda_1 = \frac{1}{2}, \lambda_2 = -\frac{1}{2}$$

Therefore, $p(G_1) = \frac{1}{2}$.

b)

$$A_2 = \begin{bmatrix} 1 & -3/4 \\ -1/12 & 1 \end{bmatrix},$$

$$a_{11} = 1, \quad a_{12} = -\frac{3}{4} \implies |a_{12}| = \frac{3}{4}$$

Therefore,

$$\sum_{j \neq 1} |a_{1j}| = |a_{12}| = \frac{3}{4} \leq \frac{3}{4}|a_{11}|$$

Then, let $\beta_1 = \frac{3}{4}$.

Similarly,

$$a_{22} = 1, \quad a_{21} = -\frac{1}{12} \implies |a_{21}| = \frac{1}{12}$$

.

So,

$$\sum_{j \neq 2} |a_{2j}| = |a_{21}| = \frac{1}{12} \leq \frac{1}{12}|a_{22}|$$

Then, let $\beta_2 = \frac{1}{12}$.

Therefore,

$$\sum_{j \neq i} |a_{ij}| \leq \beta |a_{ii}|$$

for $i = 1, 2$ and for $\beta = \max\{\beta_1, \beta_2\} = \max\{\frac{3}{4}, \frac{1}{12}\} = \frac{3}{4}$.

Let us split $A_2 = M_2 - N_2$ in a way such that M_2 is the diagonal part of A_2.

So,

$$M_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Then, $N_2 = M_2 - A_2$.

So,

$$N_2 = \begin{bmatrix} 0 & \frac{3}{4} \\ \frac{1}{12} & 0 \end{bmatrix}$$

We now want to find the eigenvalues of $G_2 = M_2^{-1} N_2$.

$$G_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & \frac{3}{4} \\ \frac{1}{12} & 0 \end{bmatrix} = \begin{bmatrix} 0 & \frac{3}{4} \\ \frac{1}{12} & 0 \end{bmatrix}$$

The eigenvalues, λ, of G_2 will satisfy the characteristic equation:

$$(\lambda)^2 - \frac{1}{16} = 0$$

$$\implies \lambda^2 = \frac{1}{16}$$

$$\implies \lambda = \pm\sqrt{\frac{1}{16}}$$

$$\implies \lambda_1 = \frac{1}{4}, \lambda_2 = -\frac{1}{4}$$

Therefore, $p(G_2) = \frac{1}{4}$.

So, we can see that even though A_1 is more diagonally dominant, A_2 has a better asymptotic convergence rate since $\rho(G_2)$ is farther away from 1.

Problem 3. Suppose A is a *singular* matrix and we do a splitting of the form $A = M - N$ in such a way that M is nonsingular (e.g. Jacobi iteration in a case where the diagonal elements of A are all nonzero).

Show that in spite of M being nonsingular, the iteration matrix $G = M^{-1}N$ can never satisfy $\rho(G) < 1$.

Thus we cannot expect an iterative method for such a system to converge in general, which makes sense if the matrix is singular. However, see the next problem...

Solution:

Let A be a singular matrix, that is, $\exists \mathbf{x} \neq 0$ such that

$$A\mathbf{x} = 0 = 0\mathbf{x}$$

.

Therefore, $\mathbf{x}$ is an eigenvector of the matrix A with the associated eigenvalue 0.

Suppose we do a splitting of the form

$$A = M - N$$

such that M is nonsingular, that is,

$$M\mathbf{x} \neq \mathbf{0}$$

.

Notice: Since M is nonsingular, M^{-1} exists.

Then:

$$A\mathbf{x} = (M - N)\mathbf{x} = M\mathbf{x} - N\mathbf{x} = 0$$

$$\implies M\mathbf{x} = N\mathbf{x} \implies (M^{-1}N)\mathbf{x} = \mathbf{x}$$

Therefore, $\mathbf{x}$ is an eigenvector of $M^{-1}N$ with the associated eigenvalue 1.

Recall: $G = M^{-1}N$ and so G has an eigenvalue of 1.

Therefore, $\rho(G) \geq 1$ and so it can never satisfy $\rho(G) < 1$.

Problem 4. Suppose we want to solve the boundary value problem $u''(x) = f(x)$ on $0 \le x \le 1$ with *periodic boundary conditions*: $u(0) = u(1)$. Recall that this has no solution unless a certain condition is satisfied by $f(x)$, in which case it has infinitely many solutions.

Recall also that if we discretize this with the standard centered second-order approximation, using a uniform grid with $h = 1/(m+1)$, we get a tridiagonal matrix with additional corner terms from the periodic boundary conditions. The matrix is singular and so the discrete problem has an analogous solvability condition.

(a) Suppose we use Jacobi iteration to solve this problem, in a case where the discrete solvability condition is satisfied. What is the iteration matrix G for this problem?

(b) Determine the eigenvalues and eigenvectors of G. Appendix C.7 of the text might be useful.

(c) You should find that one eigenvalue of G is equal to 1, and hence the Jacobi iteration does not appear to converge according to the theory of Section 4.2 (and consistent with Problem 3). But if the solvability condition is satisfied then in practice the method does converge to one of the infinitely many solutions of the linear system. Explain in what sense this is true and how the particular solution obtained is related to the initial guess $u^{[0]}$ used for the Jacobi iteration. **Hint:** Express the initial error relative to some particular solution as a linear combination of the eigenvectors and then observe the effect of iterating with the iteration matrix G.

(d) In the case it does converge, as described in (c), what is the expected convergence rate? (Note that since $\rho(G) = 1$, this is not it.)

Solution: a)

Suppose we are solving the boundary value problem $u''(x) = f(x)$ on $0 \le x \le 1$ with $u(0) = u(1)$. Suppose we also discretize this with the standard centered second-order approximation, using a uniform grid with $h = \frac{1}{(m+1)}$.

Let $u(0) = u(1) = \alpha$.

Then, the discretization matrix A is the following:

$$A = \frac{1}{h^2}\begin{bmatrix} -2 & 1 & 0 & & \cdots\cdots & 0 & 1 \\ 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & & & \\ & & 1 & -2 & 1 & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & 1 & -2 & 1 \\ 1 & & & & & 1 & -2 \end{bmatrix}$$

where we add the corner terms to account for the periodic boundary conditions.

Let M be the diagonal part of A.

$$M = \frac{1}{h^2}\begin{bmatrix} -2 & & & \\ & -2 & & \\ & & \ddots & \\ & & & -2 \end{bmatrix}$$

Then:

$$M^{-1} = \begin{bmatrix} -\frac{h^2}{2} & & & \\ & -\frac{h^2}{2} & & \\ & & \ddots & \\ & & & -\frac{h^2}{2} \end{bmatrix}$$

Let N be the upper and lower triangular part of A, that is, $N = M - A$.

The Jacobi iteration matrix $G = M^{-1}N = M^{-1}(M - A) = I - M^{-1}A$.

Since M^{-1} is just a diagonal matrix then $M^{-1}A$ just scales the rows of A. However, every diagonal entry of M^{-1} is the same so M^{-1} scales A by the same factor $-\frac{h^2}{2}$.

Therefore:

$$G = I - (-\frac{h^2}{2}A) = I + \frac{h^2}{2}A$$

$$= I + \tfrac{1}{2}\begin{bmatrix} -2 & 1 & 0 & & \cdots\cdots & 0 & 1 \\ 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & & & \\ & & 1 & -2 & 1 & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & 1 & -2 & 1 \\ 1 & & & & & 1 & -2 \end{bmatrix}$$

$$= I + \begin{bmatrix} -1 & \frac{1}{2} & 0 & & \cdots\cdots & 0 & \frac{1}{2} \\ \frac{1}{2} & -1 & \frac{1}{2} & & & & \\ & \frac{1}{2} & -1 & \frac{1}{2} & & & \\ & & \frac{1}{2} & -1 & \frac{1}{2} & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & \frac{1}{2} & -1 & \frac{1}{2} \\ \frac{1}{2} & & & & & \frac{1}{2} & -1 \end{bmatrix}$$

$$G = \begin{bmatrix} 0 & \frac{1}{2} & 0 & & \cdots\cdots & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & & & & \\ & \frac{1}{2} & 0 & \frac{1}{2} & & & \\ & & \frac{1}{2} & 0 & \frac{1}{2} & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & & & & & \frac{1}{2} & 0 \end{bmatrix}$$

is the iteration matrix for this problem.

b)

Notice that the matrix G in part a) is exactly same form as C.16 in the book with $d_0 = 0$, $d_1 = d_{-1} = \frac{1}{2}$.

The pth eigenvalue of the matrix G is given by:

$$\lambda_p = d_{-1}e^{-2\pi iph} + d_0 + d_1 e^{2\pi iph} = \frac{1}{2}e^{-2\pi iph} + \frac{1}{2}e^{2\pi iph}$$

$$= \frac{1}{2}(e^{-2\pi iph} + e^{2\pi iph}) = \frac{1}{2}(\cos(-2\pi ph) + i\sin(-2\pi ph) + \cos(2\pi ph) + i\sin(2\pi ph)) = \frac{1}{2}(2\cos(2\pi ph)) = \cos(2\pi ph).$$

Therefore, $\lambda_p = \cos(2\pi ph)$ for $p = 1, 2, \ldots, m+1$.

The jth element of the corresponding eigenvector r_p is given by

$$r_{jp} = e^{2\pi pjh} \tag{5}$$

for $j = 1, 2, \ldots, m+1$.

This is the (j, p) element of the matrix R that diagonalizes G.

Therefore:

$$r_p = \begin{bmatrix} e^{2\pi ph} \\ e^{4\pi ph} \\ \vdots \\ e^{2\pi(m+1)ph} \end{bmatrix}$$

c)

We will assume that m is even.

We can see that one eigenvalue of G is equal to 1. Namely, $\lambda_{m+1} = \cos(2\pi \frac{m+1}{m+1}) = \cos(2\pi) = 1$.

Notice: G is symmetric and its eigenvalues are distinct. Therefore, the eigenvectors are orthogonal to one another and can form a basis such that we can express the initial error as a linear combination of the eigenvectors.

Then,

$$e_0 = c_1 r_1 + c_2 r_2 + \cdots + c_m r_m + c_{m+1} r_{m+1}$$

for $c_i \in \mathbb{R}$ for $i = 1, 2, \ldots, m+1$.

Then,

$$e_k = G^k e_0 = G^k(c_1 r_1 + c_2 r_2 + \cdots + c_m r_m + c_{m+1} r_{m+1})$$

$$= c_1 G^k r_1 + c_2 G^k r_2 + \cdots + c_{m+1} G^k r_{m+1}$$

.

Notice: $G = R\Lambda R{-}1$ where the columns of R are just the eigenvectors we found in the previous part. We know that $G^k = R\Lambda^k R^{-1}$.

Let λ_j be an eigenvalue and r_j be its associated eigenvector. So, $Gr_j = \lambda_j r_j$. Then,

$$Gr_j = \lambda_j$$

$$(R\Lambda R^{-1})r_j = \lambda_j r_j$$

Multiplying by G again:

$$(R\Lambda R^{-1})(R\Lambda R^{-1})r_j = \lambda_j(R\Lambda R^{-1})r_j = \lambda_j\lambda_j r_j$$

$$R\Lambda^2 R^{-1} = G^2 r_j = \lambda_j^2 r_j$$

We can use the same reasoning as above to carry on this process to show that $G^k r_j = \lambda_j^k r_j$.

Then,

$$e_k = c_1\lambda_1^k r_1 + c_2\lambda_2^k r_2 + \cdots + c_{m+1}\lambda_{m+1}^k r_{m+1}$$

Recall: $\lambda_j = \cos(2\pi jh) = \cos(\frac{2\pi j}{m+1}) \implies |\cos(2\pi jh)| \leq 1$. We have shown previously that one of the eigenvalues is equal to 1, that is, when $\frac{j}{m+1} = 1$. The other eigenvalues will satisfy $|\lambda_j| < 1$ since $\frac{j}{m+1}$ for $j = 1, 2, , m$ will be in $(0, 1)$. We cannot have $\frac{j}{m+1} = \frac{1}{2}$ because $j = \frac{m+1}{2}$ is not an integer since m is even. This would have been when $\cos(2\pi jh) = -1$. Therefore, eventually in k, that is, as $k \to \infty$ we have that $|\lambda_j|^k \to 0$ for $j = 1, 2, \ldots, m$.

Eventually,

$$\|e_k\| = \|c_1\lambda_1^k r_1 + c_2\lambda_2^k r_2 + \cdots + c_{m+1}\lambda_{m+1}^k r_{m+1}\| \leq \|c_1\lambda_1^k r_1\| + \|c_2\lambda_2^k r_2\| + \cdots + \|c_{m+1}\lambda_{m+1}^k r_{m+1}\|$$

$$= |\lambda_1|^k\|c_1 r_1\| + |\lambda_2|^k\|c_2 r_2\| + \cdots + |\lambda_m|^k\|c_m r_m\| + |\lambda_{m+1}|^k\|c_{m+1} r_{m+1}\|$$

$$= 1|c_{m+1}|\|r_{m+1}\| = |c_{m+1}|\|r_{m+1}\|$$

Here, the norm $\|\|$ used is the 2-norm.

Therefore, the system to converge we would need $|c_{m+1}| = 0 \implies c_{m+1} = 0$. Thus, we would need $e_0 = c_1 r_1 + c_2 r_2 + \cdots + c_m r_m$.

If we had let p run from 0 to m, we would require $c_0 = 0$.

Suppose we have an initial guess u_0 where the subscript refers to the iteration number. Let u^* be a particular solution. Then, initial error $e_0 = u0 - u^*$.

Again, we can express $u0$ and u^* in terms of the eigenbasis.

Let,

$$u_0 = \sum_{i=1}^{m+1} d_i r_i$$

and

$$u^* = \sum_{i=1}^{m+1} f_i r_i$$

,

for constants d_i and f_i.

Then,

$$e_0 = \sum_{i=1}^{m+1} (d_i - f_i) r_i$$

We want $d_{m+1} - f_{m+1} = 0$ and so we need $d_{m+1} = f_{m+1}$.

Therefore, in practice the method will converge to one of the infinitely many solutions of the linear system if our initial guess is chosen carefully. We would need the initial guess and the particular solution to have the same component in the $(m + 1)$th eigenvector direction. This is going to be direction corresponding to the eigenvalue equal to 1.

d)

From above we know that,

$$e_k = c_1\lambda_1^k r_1 + c_2\lambda_2^k r_2 + \cdots + c_{m+1}\lambda_{m+1}^k r_{m+1}$$

.

We showed that the system will converge if $c_{m+1} = 0$. Then,

$$e_k = c_1\lambda_1^k r_1 + c_2\lambda_2^k r_2 + \cdots + c_m\lambda_m^k r_m$$

Also, let $\lambda = \max\{\lambda_1, \lambda_2, \ldots, \lambda_m\}$. In this case, it will be λ_1 or λ_m as $\frac{1}{m+1}$ will be closest to 0 and $\frac{m}{m+1}$ will be closest to 1 and so $\cos(\frac{2\pi}{m+1})$ and $\cos(\frac{2\pi m}{m+1})$ will be closest to 1. Let us pick λ_1 for convenience as $\lambda_1 = \lambda_m$ when m is even.

Therefore,

$$e_k = c_1\lambda_1^k r_1 + c_2\lambda_2^k r_2 + \cdots + c_m\lambda_m^k r_m \leq \lambda^k(c_1 r_1 + c_2 r_2 + \cdots + c_m r_m) = \lambda_1^k(c_1 r_1 + c_2 r_2 + \cdots + c_m r_m)$$

$$e_k \leq \lambda_1^k(e_0)$$

$$\|e_k\|_2 \leq |\lambda_1|^k \|e_0\|_2$$

Therefore, the expected rate of convergence is given by $|\lambda_1| < 1$.

Problem 5.

Consider the problem

$$\frac{d}{dx}\left(\kappa(x)u'(x)\right) = f(x)$$

on $0 \le x \le 1$ with $\kappa(x) > 0$ everywhere.

(a) Determine the solution of this problem for the case of $f(x) = 0$ and a discontinuous $\kappa(x)$ given by

$$\kappa(x) = \begin{cases} 0.1 & \text{if } x < 0.5, \\ 1 & \text{if } x > 0.5. \end{cases}$$

with boundary conditions $u(0) = 0$ and $u(1) = 1$. The solution should be continuous and piecewise linear with a discontinuity in slope at $x = 0.5$ (Note this could model steady steady heat flow through a material that conducts heat better on the right half than on the left.)

(b) Now suppose we discretize using the symmetric matrix A of (2.73) in the text. Choose $m = 19$ (odd, so there is a grid point exactly at $x = 0.5$, in which case the solution to the discrete system should agree with the exact solution of the differential equation).

Write a computer code to solve this system using the SOR method for an arbitrary value of ω. Note that setting $\omega = 1$ should just reduce to the Gauss-Seidel method for comparison.

Test this code and estimate the rate of convergence with $\omega = 1$ and with $\omega = 0.7$, by doing a least squares fit of the convergence history as demonstrated in the notebook `IterativeMethods.ipynb`.

Solution:

a) Let us consider the problem

$$\frac{d}{dx}\left(\kappa(x)u'(x)\right) = 0$$

on $0 \le x \le 1$ with $\kappa(x) > 0$ everywhere and a discontinuous $\kappa(x)$ given by

$$\kappa(x) = \begin{cases} 0.1 & \text{if } x < 0.5, \\ 1 & \text{if } x > 0.5. \end{cases}$$

with boundary conditions $u(0) = 0$ and $u(1) = 1$.

Since, $(\kappa(x)u'(x))' = 0 \implies \kappa(x)u'(x) = c_1$. for some constant c_1.

If $x < 0.5$, then we have $0.1u'(x) = c_1$.

Then:

$$u'(x) = 10c_1$$

$$\int_0^x u'(x)\, dx = \int_0^x 10c_1\, dx$$

$$\implies (u(x) - u(0)) = 10c_1 x$$

Recall: $u(0) = 0$

So, $u(x) = 10c_1x$

Now, suppose $x > 0.5$. We have $\frac{d}{dx}(u'(x)) = 0$

Then:

$$u(x) = \int_0^{0.5} 10c_1 \, dx + \int_{0.5}^{x} c_1 \, dx$$

$$u(x) = 5c_1 + c_1(x - 0.5)$$

We know $u(1) = 1 \implies 1 = 5c_1 + 0.5c_1 \implies c_1 = \frac{1}{5.5}$.

Therefore, when $x < 0.5$, $u(x) = \frac{10}{5.5}x$ and when $x > 0.5$, $u(x) = c_1(5+x-0.5) = c_1(x+4.5) = \frac{1}{5.5}(x+4.5)$

$$u(x) = \begin{cases} \frac{10}{5.5}x & \text{if } x < 0.5, \\ \frac{1}{5.5}(x + 4.5) & \text{if } x > 0.5 \end{cases}$$

b)

The Jupyter notebook contains the code for this part. Here we plot the figures of the semilogy plot for the three cases: $\omega = 0.7, \omega = 1, \omega = 1.7$.

Case 1: $\omega = 0.7$ This is the SOR method. We expect fast convergence specially in relation to Gauss Seidel and under relaxed method. This is exactly what we see in Figure 2. The semilogy plot has a negative slop and the error is decreasing with the number of iterations. After 100 iterations, the error drops to about 10^{-9} which is the best out of the three ω values. We also get plotting accuracy with about 25 iterations as showing in Figure 1. The rate of convergence, $\rho \approx 0.83$ after doing a least squares fit which is farthest away from 1 than the other values of *omega* and explains the faster convergence.

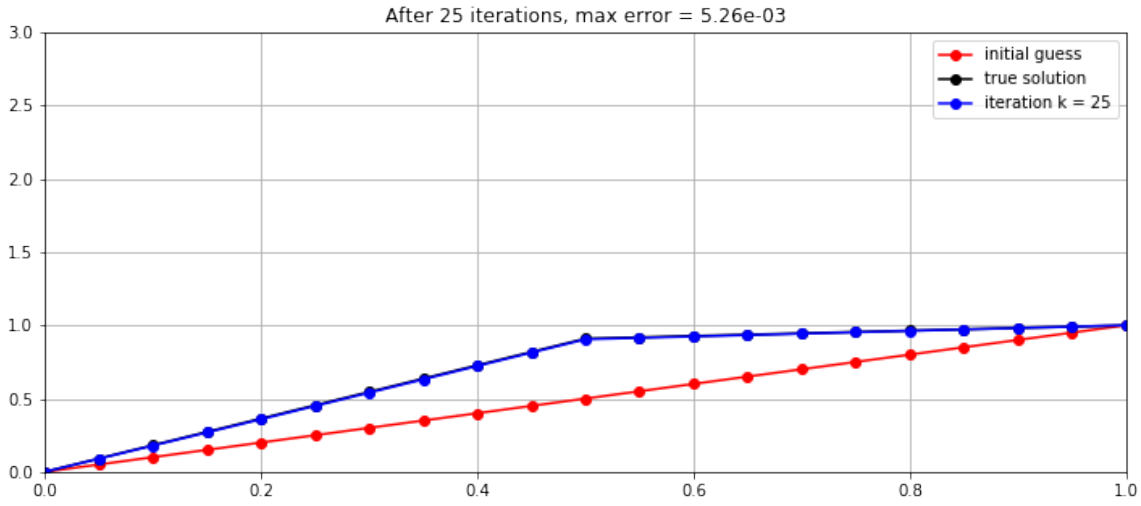

Figure 1: Plot after 25 iterations

Case 2: $\omega = 1$

Recall: This is just the Gauss-Seidel case and it should converge slower than SOR.

This is exactly what we see in the semilogy plot of errors. The error decreases as shown by the negative slope of the semilogy plot in Figure 4 but the error only goes down to about 10^{-2}. We would need more iterations to converge. However, we do get close to plotting precision as shown in the Figure 3

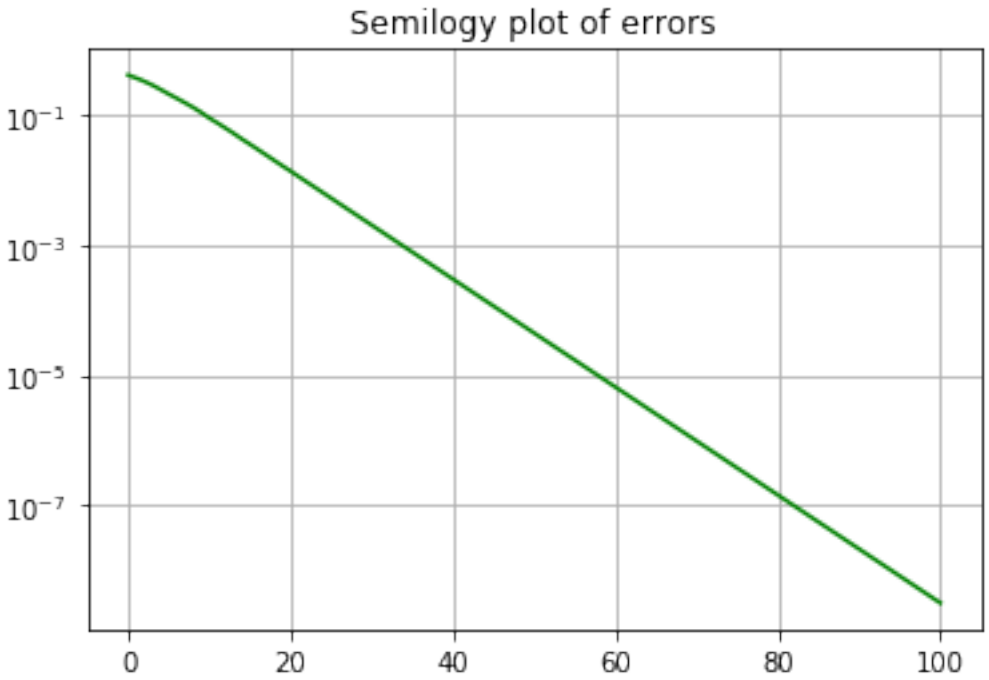

Figure 2: $\omega = 1.7$

after 100 iterations. The convergence rate is about 0.96 which is close to 1 and so the slow convergence makes sense.

Case 3: $\omega = 0.7$

Recall: $\omega < 1$ means that the method is under-relaxed which means this should converge the slowest, even slower than the Gauss-Seidel method. This is exactly what we see in the plot in Figure 5. The error does not go below 10^{-1} even after 100 iterations. We are unable to get plotting precision even. The semilogy plot does have a negative slope so the error is decreasing but it might take a large number of iterations for us to reach machine precision. The rate of convergence, $\rho \approx 0.987$ after doing a least squares fit which is close to 1 and explains the slow convergence. Notice: ρ in this case is closer to 1 than Gauss-Seidel method.

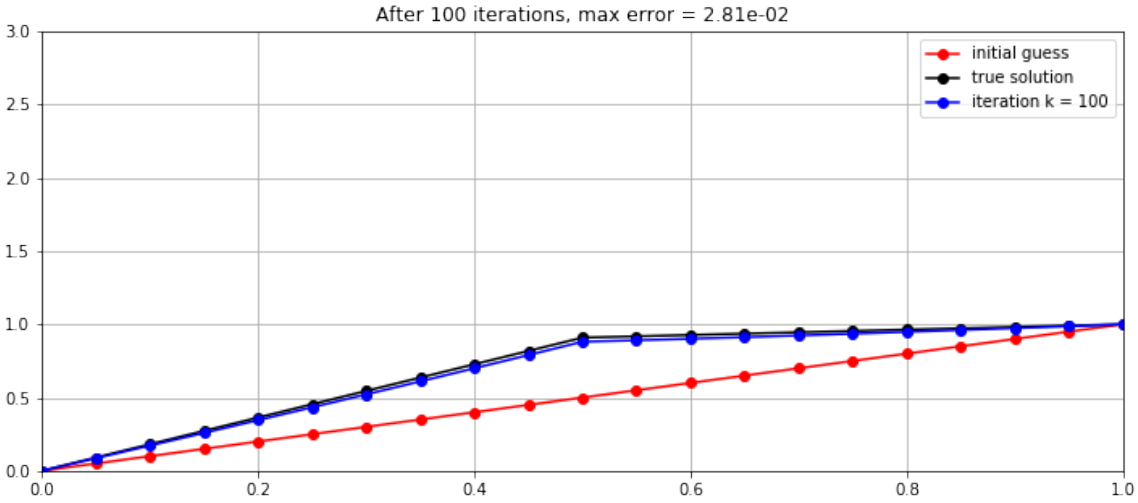

Figure 3: Plot after 100 iterations ($\omega = 1$)

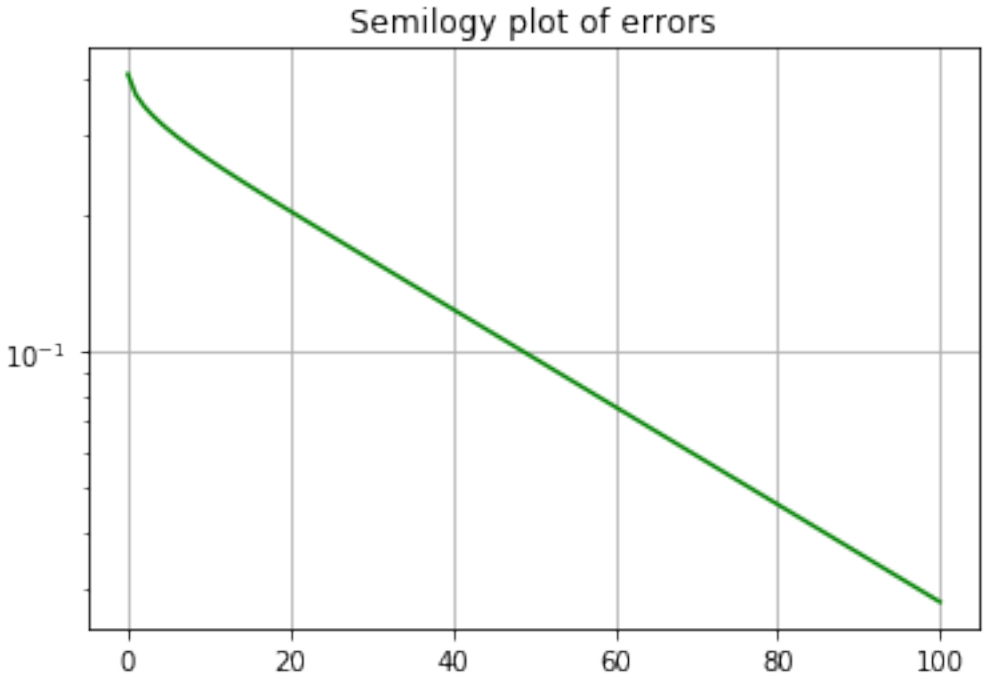

Figure 4: $\omega = 1$

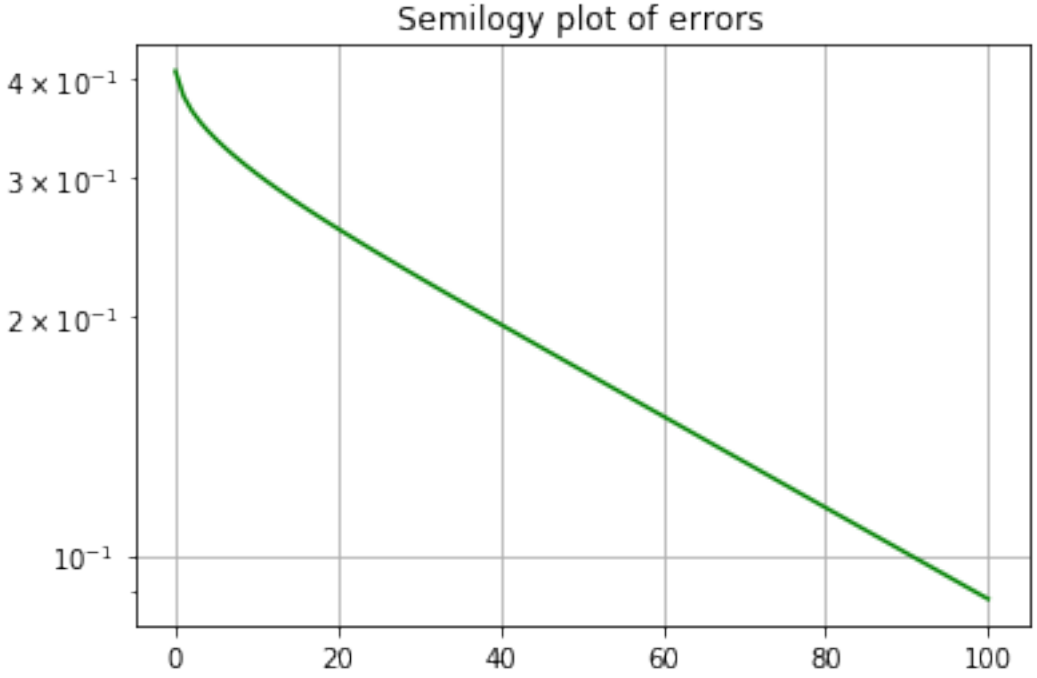

Figure 5: $\omega = 0.7$