

MATH 231 : Numerical ODEs

Pratham Lalwani

March 18, 2025

Question 1: Eigenvalues of special tridiagonal matrices

This question is about finding eigenvalues of tridiagonal linear systems arising from applications, specifically finding the eigenvalues of an $n \times n$ matrix of the form,

$$A = \begin{pmatrix} a & b & & & \\ c & a & b & & \\ & \ddots & \ddots & \ddots & \\ & & c & a & b \\ & & & c & a \end{pmatrix}$$

where a, b, c are real numbers with $bc > 0$ (i.e. b and c have the same signs).

- (a) Show that the eigenvalue problem of A is equivalent to the equations

$$cv_{j-1} + (a - \lambda)v_j + bv_{j+1} = 0, \quad j = 1, \dots, n$$

$$v_0 = 0 = v_{n+1}$$

where $\mathbf{v} = (v_1, \dots, v_n)^T$ is an eigenvector of A associated with the eigenvalue λ .

- (b) The recurrence relation (1) is a second order linear difference equation and can be solved similar to second order linear differential equations. By guessing $v_j = r^j$ for some constant r , show that r satisfies

$$r_{\pm} = \frac{\lambda - a \pm \sqrt{(\lambda - a)^2 - 4bc}}{2b}, \quad \text{with} \quad r_+ r_- = \frac{c}{b}$$

- (c) Show by contradiction that r_{\pm} must be distinct.

Hint: if $r_{\pm} = r$ are repeated, then $v_j = Ar^j + Br^j$ for some constants A, B .

- (d) Since r_{\pm} are distinct, the general solution for (1) is $v_j = Ar_+^j + Br_-^j$ for constants A, B . Use this to conclude from (2) and (3) that,

$$\left(\frac{br_+^2}{c}\right)^{n+1} = 1$$

- (e) From part (c), (3) and (4), show that r_{\pm} must be complex valued and conclude that (4) has the solutions for $k = 1, \dots, n$,

$$r_{\pm, k} = \sqrt{\frac{c}{b}} \exp\left(\frac{\pm i k \pi}{n+1}\right), \quad \text{where } i = \sqrt{-1}$$

- (f) Using part (e), conclude that the eigenvalues of A is given by

$$\lambda_k = a + 2 \operatorname{sgn}(b) \sqrt{bc} \cos\left(\frac{\pi k}{n+1}\right), \quad k = 1, \dots, n$$

- (g) Find the eigenvalues of the $n \times n$ finite difference matrix $A_h = \frac{1}{h^2} \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}$, where

$$h = \frac{1}{n+1}.$$

Conclude that A_h is symmetric positive definite and find its condition number $\kappa(A_h)$ with respect to $\|\cdot\|_2$. Show that $\kappa(A_h) = \mathcal{O}(h^{-2})$ as number of grid points n increases. What does this mean for solving $A_h \mathbf{x} = \mathbf{b}$ when n is large?

(a) Let (λ, \vec{v}) be an eigenpair of A

$$\begin{aligned} A\vec{v} &= \lambda\vec{v} \\ (A - \lambda I)\vec{v} &= \vec{0} \\ \begin{pmatrix} (a - \lambda)v_1 + bv_2 \\ cv_1 + (a - \lambda)v_2 + bv_3 \\ \vdots \\ cv_{n-2} + (a - \lambda)v_{n-1} + b_n \\ c_{n-1} + (a - \lambda)v_n \end{pmatrix} &= \vec{0}. \end{aligned}$$

We can write the above relation as the following,

$$cv_{j-1} + (a - \lambda)v_j + bv_{j+1} = 0. \quad (1)$$

Where $0 \leq j \leq n+1$ and $v_0 = 0 = v_{n+1}$ ⊖

(b) Using the hint we guess the following form of the solution $v_j = r^j$. Substituting in 1,

$$\begin{aligned} cr^{j-1} + (a - \lambda)r^j + br^{j+1} &= 0 \\ c + (a - \lambda)r + br^2 &= 0 \\ &. \end{aligned}$$

Using the quadratic formula, we get

$$r_{\pm} = \frac{\lambda - a \pm \sqrt{(a - \lambda)^2 - 4bc}}{2b}.$$

As r_{\pm} are the roots to a quadratic, hence

$$r_+ r_- = \frac{c}{b} \quad (2)$$

(c) If 1 has a repeated root, say $r_{\pm} = r$, then solution to the recursion would look like,

$$v_j = Ar^j + Bjr^j.$$

Checking the boundary conditions, $v_0 = 0 = v_{n+1}$

$$v_0 = Ar^0 + B(0)r^0 = A = 0. \quad (3)$$

$$v_{n+1} = (0)r^{n+1} + B(n+1)r^{n+1} = B(n+1)r^{n+1} = 0 \implies B = 0. \quad (4)$$

Combining 3 & 4 gives,

$$v_j = 0.$$

Which is the trivial eigenvector. Hence, we cannot have a repeated root if we want a non-zero eigenvector.

(d) From (c) we have that roots are distinct. Therefore, we look for solutions of the form $v_j = Ar_+^j + Br_-^j$ for some constants A and B defined by the "boundary conditions" of the recursion. We have,

$$\begin{aligned} v_0 &= A + B = 0 \implies A = -B \\ v_{n+1} &= Ar_+^{n+1} + Br_-^{n+1} = 0 \implies r_+^{n+1} = r_-^{n+1} \end{aligned} \quad (5)$$

From, 2 and 5, it follows that

$$\begin{aligned} (r_+^2)^{(n+1)} &= \left(\frac{c}{b}\right)^{n+1} \\ \left(\frac{br_+^2}{c}\right)^{(n+1)} &= 1 \end{aligned} \quad (6)$$

(e) We can observe in 6 that $\frac{br_+^2}{c}$ are the roots of unity, therefore,

$$\frac{br_+^2}{c} = \exp\left(\frac{ik\pi}{n+1}\right) \implies r_+ = \sqrt{\frac{c}{b}} \exp\left(\frac{ik\pi}{n+1}\right) \quad k = 0, \dots, n+1.$$

Similarly,

$$\frac{br_-^2}{c} = \exp\left(\frac{im\pi}{n+1}\right) \implies r_- = \sqrt{\frac{c}{b}} \exp\left(\frac{im\pi}{n+1}\right) \quad m = 0, \dots, n+1.$$

Using 2,

$$r_+ r_- = \frac{c}{b} \exp\left(\frac{i(k+m)\pi}{n+1}\right) = \frac{c}{b} \implies k = -m.$$

Therefore we have,

$$r_{\pm, k} = \sqrt{\frac{c}{b}} \exp\left(\frac{\pm ik\pi}{n+1}\right) \quad k = 1, \dots, n.$$

Question 2: Classical iterative methods for strictly diagonally dominant matrices

- (a) Show that the diagonal part of any strictly diagonally dominant (S.D.D.) matrix is invertible.
- (b) Recall the Gershgorin's theorem below, which can give useful information about the eigenvalues of a matrix. The eigenvalues of a complex valued matrix A lies in the union of n discs $\bigcup_{i=1}^n D_i$ on the complex plane, where

$$D_i = \left\{ z \in \mathbb{C} : |z - a_{ii}| \leq \sum_{j \neq i} |a_{ij}| \right\}$$

Using Gershgorin's theorem, conclude S.D.D. matrices are invertible. Hint: Show that $0 \notin D_i$ for all $i = 1, \dots, n$.

The next two parts are about showing convergence of Jacobi and Gauss-Seidel iterations for S.D.D. matrices.

- (c) Recall the matrix $-M^{-1}N$ associated with the Jacobi iteration takes the form $-D^{-1}(L+U)$, where $A = L + D + U$.
- Let A be S.D.D. and λ be any eigenvalue of $-D^{-1}(L+U)$. Show that $\det(L+U+\lambda D) = 0$ using part (a).
 - Now suppose $|\lambda| \geq 1$. Deduce from A being S.D.D. that $L+U+\lambda D$ must also be S.D.D.
 - Deduce a contradiction by applying the result from part (b) to $L+U+\lambda D$, and conclude that $|\lambda| < 1$.
 - Combine parts (i)-(iii) to conclude that Jacobi iteration converges for S.D.D. matrices.
- (d) Follow a similar argument as part (c) to show that Gauss-Seidel iterations converges for S.D.D. matrices.

Solution

- (a) Let A be a S.D.D matrix and,

$$\implies a_{ii} > \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} \geq 0 \implies a_{ii} > 0 \quad \forall 1 \leq i \leq n.$$

Let D be the matrix containing the diagonal entries of A , hence

$$D = \begin{bmatrix} a_{11} & & \\ & \ddots & \\ & & a_{nn} \end{bmatrix}.$$

As, all $a_{ii} > 0$, therefore we D^{-1} exists.

- (b) Let λ_i be the eigenvalues associated with disc D_i .
Suppose $0 \in D_i$ for some $1 \leq i \leq n$, therefore, we have,

$$a_{ii} \leq \sum_{\substack{j=0 \\ j \neq i}}^n a_{ij}.$$

Which is false as A is a S.D.D matrix, hence, $0 \notin D_i$.

Therefore we have, $|\lambda_i| > 0 \quad \forall i, 1 \leq i \leq n \implies A^{-1}$ exists.

- (c) (i) Given that λ is an eigenvalue of $-D^{-1}(L + U)$. Therefore we have \vec{v} such that $\vec{v} \neq 0$,

$$\begin{aligned} -D^{-1}(L + U)\vec{v} &= \lambda\vec{v} \\ (L + U)\vec{v} &= -\lambda D\vec{v} \\ (L + U + \lambda D)\vec{v} &= \vec{0}. \end{aligned}$$

As there is a non-zero null vector associated with $L + U + \lambda D$, therefore $\det(L + U + \lambda D) = 0$.

- (ii) Given that A is S.D.D. Suppose $|\lambda| \geq 1$. Consider,

$$\begin{aligned} |(L + U + \lambda D)_{ii}| &= |\lambda a_{ii}| = |\lambda| |a_{ii}| \\ &> |\lambda| \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} \\ &\geq \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} \\ &= \sum_{\substack{j=1 \\ j \neq i}}^n (L + U + \lambda D)_{ij} \end{aligned}$$

Hence, $(L + U + \lambda D)$ is S.D.D. .

- (iii) If $|\lambda| \geq 1$ and A is S.D.D, gives that $(L + U + \lambda D)$ is S.D.D .

Therefore, $(L + U + \lambda D)$ is invertible. Which is a contradiction as $\det(L + U + \lambda D) = 0$. Therefore, $|\lambda| < 1$.

- (iv) Let $M = D$ and $N = L + U$. From parts (i)-(iii) we get,

$$\lambda_i \leq \lambda_{max} < 1 \implies \rho(-M^{-1}N) < 1.$$

By theorem of convergence of iterative solvers we get, iterations based on $-M^{-1}N$ converges to 0.

- (d) (i) Let λ be an eigenvalue of $-(L + D)^{-1}(U)$. Therefore we have \vec{v} such that $\vec{v} \neq 0$,

$$\begin{aligned} -(L + D)^{-1}(L + U)\vec{v} &= \lambda\vec{v} \\ (U)\vec{v} &= -\lambda(L + D)\vec{v} \\ (U + \lambda(L + D))\vec{v} &= \vec{0}. \end{aligned}$$

As there is a non-zero null vector associated with $U + \lambda(L + D)$, therefore $\det(U + \lambda(L + D)) = 0$.

- (ii) Given that A is S.D.D. Suppose $|\lambda| \geq 1$. Consider,

$$\begin{aligned} |(U + \lambda(L + D))_{ii}| &= |\lambda a_{ii}| = |\lambda| |a_{ii}| \\ &> |\lambda| \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \\ &\geq |\lambda| \sum_{j=1}^{i-1} |a_{ij}| + \sum_{j=i+1}^n |a_{ij}| \\ &= \left| \sum_{\substack{j=1 \\ j \neq i}}^n (U + \lambda(L + D))_{ij} \right| \end{aligned}$$

Hence, $(U + \lambda(L + D))$ is S.D.D. .

- (iii) If $|\lambda| \geq 1$ and A is S.D.D, gives that $(U + \lambda(L + D))$ is S.D.D .

Therefore, $(L + U + \lambda(L + D))$ is invertible. Which is a contradiction as $\det(U + \lambda(L + D)) = 0$. Therefore, $|\lambda| < 1$.

- (iv) Let $M = L + D$ and $N = U$. From parts (i)-(iii) we get,

$$\lambda_i \leq \lambda_{\max} < 1 \implies \rho(-M^{-1}N) < 1.$$

By theorem of convergence of iterative solvers we get, iterations based on $-M^{-1}N$ converges to 0.

Question 3: Classical iterative methods for symmetric positive definite matrices

This question is about coding and comparing classical iterative methods for the S.P.D. matrix A_h from Q1(g).

- Write a pseudocode for the classical iterative methods: Richardson, optimal Richardson, Jacobi, Gauss-Seidel, S.O.R., and optimal S.O.R.
- Implement a program to solve $A_h \mathbf{x} = \mathbf{b}$ with $\mathbf{b} = (1, \dots, 1)^T \in \mathbb{R}^{20}$ and $\mathbf{x}_0 = (1, 0, \dots, 0)^T \in \mathbb{R}^{20}$ using Richardson (with $\omega = \lambda_{\max}^{-1}$), optimal Richardson, Jacobi, Gauss-Seidel, S.O.R. (with $\theta = 1.2$) and optimal S.O.R. Generate a plot comparing the log of their residual in ℓ_2 norm versus iterations up to 5000 . Rank the performance of each method by comparing the iterations needed to reach the residual tolerance of 10^{-14} . Use sparse representation when appropriate.
Hint: Use Q1(g) to find parameters for Richardson and vary θ to find an approximate optimal parameter for S.O.R.
- Comment on the decreases in performance when $n = 1000$. Explain briefly how this relates to $\kappa(A_h) = \mathcal{O}(h^{-2})$.

Algorithm 1: Richardson Iteration

```

1 function RichardsonIteration( $A, b, x_0, \omega, tol, maxIter$ ):
    Input:
         $A$ : The matrix to find the solution to
         $b$ : The resultant vector in  $Ax = b$ 
         $x_0$ : The initial guess
         $\omega$ : Richardson parameter (fixed)
         $maxIter$ : The maximum of iterations
    Output:  $x$ : The solution to  $Ax = b$ 

2    $M \leftarrow \omega^{-1}I$ 
3    $N \leftarrow A - M$ 
4    $x \leftarrow x_0$ 
5    $r \leftarrow b - Ax$ 
6   while  $\|r\|_2 < tol$  and  $i < maxIter$ :
7        $x \leftarrow x + \omega r$ 
8        $r \leftarrow b - Ax$ 
9   end
10  return  $x$ ;

```

Algorithm 2: Optimal Richardson Iteration

```

1 function OptimalRichardsonIteration( $A, b, x_0, tol, maxIter$ ):
    Input:
         $A$ : The matrix to find the solution to
         $b$ : The resultant vector in  $Ax = b$ 
         $x_0$ : The initial guess
         $maxIter$ : The maximum of iterations
    Output:  $x$ : The solution to  $Ax = b$ 

2    $\omega \leftarrow \frac{2}{\lambda_{max}(A) + \lambda_{min}(A)}$ 
3    $M \leftarrow \omega^{-1}I$ 
4    $N \leftarrow A - M$ 
5    $x \leftarrow x_0$ 
6    $r \leftarrow b - Ax$ 
7   while  $\|r\|_2 < tol$  and  $i < maxIter$ :
8        $x \leftarrow x + \omega r$ 
9        $r \leftarrow b - Ax$ 
10  end
11  return  $x$ ;

```

Algorithm 3: Jacobi Iteration

```
1 function JacobiIteration( $A, \mathbf{b}, \mathbf{x}_0, tol, maxIter$ ):  
    Input:  
     $A$ : The matrix to find the solution to  
     $\mathbf{b}$ : The resultant vector in  $A\mathbf{x} = \mathbf{b}$   
     $\mathbf{x}_0$ : The initial guess  
     $maxIter$ : The maximum of iterations  
    Output:  $\mathbf{x}$ : The solution to  $A\mathbf{x} = \mathbf{b}$   
  
2    $M \leftarrow \text{diag}(A)$   
3    $N \leftarrow A - M$   
4    $\mathbf{x} \leftarrow \mathbf{x}_0$   
5    $\mathbf{r} \leftarrow \mathbf{b} - A\mathbf{x}$   
6   while  $\|\mathbf{r}\|_2 < tol$  and  $i < maxIter$ :  
7        $\mathbf{x} \leftarrow M^{-1}(\mathbf{x} + \mathbf{b} - N\mathbf{x})$   
8        $\mathbf{r} \leftarrow \mathbf{b} - A\mathbf{x}$   
9        $i = i + 1$   
10  end  
11  return  $\mathbf{x}$ 
```

Algorithm 4: Gauss-Sidel Iteration

```
1 function GaussSidelIteration( $A, \mathbf{b}, \mathbf{x}_0, tol, maxIter$ ):  
    Input:  
     $A$ : The matrix to find the solution to  
     $\mathbf{b}$ : The resultant vector in  $A\mathbf{x} = \mathbf{b}$   
     $\mathbf{x}_0$ : The initial guess  
     $maxIter$ : The maximum of iterations  
    Output:  $\mathbf{x}$ : The solution to  $A\mathbf{x} = \mathbf{b}$   
  
2    $M \leftarrow \text{diag}(A) + \text{lower}(A)$   
3    $N \leftarrow A - M$   
4    $\mathbf{x} \leftarrow \mathbf{x}_0$   
5    $\mathbf{r} \leftarrow \mathbf{b} - A\mathbf{x}$   
6    $i \leftarrow 0$   
7   while  $\|\mathbf{r}\|_2 < tol$  and  $i < maxIter$ :  
8        $\mathbf{x} \leftarrow M^{-1}(\mathbf{x} + \mathbf{b} - N\mathbf{x})$   
9        $\mathbf{r} \leftarrow \mathbf{b} - A\mathbf{x}$   
10       $i = i + 1$   
11  end  
12  return  $\mathbf{x}$ ;
```

Question 4: Steepest Descent and Conjugate Gradient

- (a) Let A be a S.P.D. matrix. Show that $(\mathbf{x}, \mathbf{y})_A := \mathbf{x}^T A \mathbf{y}$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ forms an inner product on \mathbb{R}^n .
- (b) Using part (a), conclude that $\|\mathbf{x}\| = (\mathbf{x}, \mathbf{x})_A^{1/2}$ for $\mathbf{x} \in \mathbb{R}^n$ is a norm on \mathbb{R}^n .
Hint: You can assume the Cauchy-Schwarz inequality $|(\mathbf{x}, \mathbf{y})_A| \leq \|\mathbf{x}\|_A \|\mathbf{y}\|_A$ holds.
- (c) For the method of Steepest Descent, show that $\nabla f(\mathbf{x}_k)$ and $\nabla f(\mathbf{x}_{k+1})$ are orthogonal (i.e. zig-zaging behavior), where $f(\mathbf{y}) = \frac{1}{2} \mathbf{y}^T A \mathbf{y} - \mathbf{y}^T \mathbf{b}$. Hint: Recall how the step size for Steepest Descent is determined.
- (d) Repeat the experiment from Q3(b) with $\mathbf{b} = (1, \dots, 1)^T \in \mathbb{R}^{1000}$ and $\mathbf{x}_0 = (1, 0, \dots, 0)^T \in \mathbb{R}^{1000}$

using the method of Steepest Descent and Conjugate Gradient. Generate a plot comparing the log of their residual in ℓ_2 norm versus iterations up to 5000 . Rank their performance by comparing the iterations needed to reach the residual tolerance of 10^{-14} , as well as versus the classical iterative methods. Verify your CG method terminates after the desired number of iterations. Use sparse representation when appropriate.

Solution

(a) (\cdot, \cdot) is an inner-product if :

(i) Conjugate Symmetry:

$$(x, y)_A = (y, x)_A.$$

(ii) Linearity

$$(a\vec{x} + b\vec{y}, \vec{z})_A = a(\vec{x}, \vec{z})_A + b(\vec{y}, \vec{z})_A.$$

(iii) Positive-Definiteness:

$$(\vec{x}, \vec{x})_A > 0.$$

(i)

$$(x, y)_A = x^T A y = y^T A x = (y, x)_A.$$

(ii)

$$(a\vec{x} + b\vec{y}, \vec{z})_A = (a\vec{x} + b\vec{y})^T A \vec{z} = a\vec{x}^T A \vec{z} + b\vec{y}^T A \vec{z} = a(\vec{x}, \vec{z})_A + b(\vec{y}, \vec{z})_A.$$

(iii)

$$(x, x)_A = \vec{x}^T A \vec{x} > 0 \quad \text{As } A \text{ is SPD.}$$

Therefore, $(\cdot, \cdot)_A$ is an inner-product.

(b) $\|\cdot\|_A$ is a norm if :

(i) Positive Definiteness :

$$\|x\|_A > 0 \quad \forall \vec{x} \neq 0 \quad \wedge \quad \|\vec{x}\|_A = 0 \iff \vec{x} = \vec{0}.$$

(ii) Scalar Multiplication

$$\|\lambda \vec{x}\|_A = \lambda \|\vec{x}\|_A.$$

(iii) Sub-additivity (Triangle Inequality):

$$\|\vec{x} + \vec{y}\|_A = \|\vec{x}\|_A + \|\vec{y}\|_A.$$

(i) Let $x \in \mathbb{R}^n$ and $\vec{x} \neq \vec{0}$

$$\|x\|_A = \sqrt{\vec{x}^T A \vec{x}} > 0 \quad , \text{as } A \text{ is SPD.}$$

Let $\|\vec{x}\|_A = 0$

$$\|\vec{x}\|_A = 0 = \sqrt{\vec{x}^T A \vec{x}} \iff \vec{x} = 0 \quad , \text{as } A \text{ is SPD.}$$

(ii) Scalar Multiplication

$$\|\lambda \vec{x}\|_A = \sqrt{\lambda \vec{x}^T A \lambda \vec{x}} = \sqrt{\lambda^2 \vec{x}^T A \vec{x}} = \lambda \|\vec{x}\|_A.$$

(iii) Sub-additivity (Triangle Inequality):

$$\begin{aligned}\|\vec{x} + \vec{y}\|_A &= \sqrt{(\vec{x} + \vec{y})^T A (\vec{x} + \vec{y})} \\ &= \sqrt{\vec{x}^T A \vec{x} + \vec{x}^T A \vec{y} + \vec{y}^T A \vec{x} + \vec{y}^T A \vec{y}} \\ &= \sqrt{\vec{x}^T A \vec{x} + 2\vec{x}^T A \vec{y} + \vec{y}^T A \vec{y}} \\ &= \sqrt{\|\vec{x}\|_A^2 + 2(\vec{x}, \vec{y}) + \|\vec{y}\|_A^2} \\ &\leq \sqrt{\|\vec{x}\|_A^2 + 2\|\vec{x}\|_A\|\vec{y}\|_A + \|\vec{y}\|_A^2} \\ &= \|\vec{x}\|_A + \|\vec{y}\|_A.\end{aligned}$$

Therefore, $\|x\|_A$ is valid norm.