Numerical Methods 2022 — Midterm 2

Solutions

Problem 1 (10 pts). Find the tightest possible constants c_1 and c_2 satisfying

$$c_1\|x\|_1 \leq \|x\|_\infty \leq c_2\|x\|_1 \quad \forall x \in \mathbb{R}^n.$$

You must prove that your c_1 and c_2 are the tightest.

Solution.

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \leq n \max_i |x_i| = n\|x\|_\infty \implies c_1 = \frac{1}{n}$$

$$\|x\|_\infty = \max_i |x_i| \leq \sum_{i=1}^{n} |x_i| = \|x\|_1 \implies c_2 = 1$$

For $x = (1, 1, \ldots, 1)$, we have $\frac{1}{n}\|x\|_1 = 1 = \|x\|_\infty$. This shows that c_1 is tight. For $x = (1, 0, 0, \ldots, 0)$, we have $\|x\|_\infty = 1 = \|x\|_1$. Therefore, c_2 is also tight.

Problem 2 (30 pts). Consider a different way to define the matrix norm for a matrix $A \in \mathbb{R}^{n\times n}$:

$$\|A\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}^2}$$

This is called the Frobenius norm. In answering the following questions, you may utilize the Cauchy-Schwarz inequality in your proof:

$$\left(\sum_{i=1}^{n} u_i v_i\right)^2 \leq \left(\sum_{i=1}^{n} u_i^2\right)\left(\sum_{i=1}^{n} v_i^2\right) \text{ where } u_i, v_i \in \mathbb{R}$$

(a) (15 pts) Prove that Frobenius norm satisfies the three properties:

$$\|A\| \geq 0 \tag{1}$$

$$\|A + B\| \leq \|A\| + \|B\| \tag{2}$$

$$\|\alpha A\| = |\alpha|\|A\| \tag{3}$$

for $A, B \in \mathbb{R}^{n\times n}$ and $\alpha \in \mathbb{R}$.

(b) (15 pts) Does Frobenius norm satisfy the following condition?

$$\|Ax\|_2 \leq \|A\|_{\mathrm{F}}\|x\|_2 \quad \forall x \in \mathbb{R}^n, A \in \mathbb{R}^{n\times n}$$

If so, provide a proof. If not, provide a counter example.

Solution.

(a) Since $A_{ij}^2 \geq 0$, we know that $\sum_i \sum_j A_{ij}^2 \geq 0$ and thus $\|A\|_F = \sqrt{\sum_i \sum_j A_{ij}^2} \geq 0$. This proves (1).

For (2), we have

$$\begin{aligned}
\|A+B\|_F^2 &= \sum_i \sum_j (A_{ij} + B_{ij})^2 \\
&= \sum_i \sum_j A_{ij}^2 + \sum_i \sum_j B_{ij}^2 + 2 \sum_i \sum_j A_{ij} B_{ij} \\
&\leq \|A\|_F^2 + \|B\|_F^2 + 2\sqrt{(\sum_i \sum_j A_{ij}^2)(\sum_i \sum_j B_{ij}^2)} && \text{(Cauchy-Schwarz)} \\
&= \|A\|_F^2 + \|B\|_F^2 + 2\|A\|_F \|B\|_F \\
&= (\|A\|_F + \|B\|_F)^2.
\end{aligned}$$

This implies that $\|A+B\|_F \leq \|A\|_F + \|B\|_F$.

For (3), we have

$$\begin{aligned}
\|\alpha A\|_F &= \sqrt{\sum_i \sum_j (\alpha A_{ij})^2} \\
&= \sqrt{\alpha^2} \sqrt{\sum_i \sum_j A_{ij}^2} \\
&= |\alpha| \|A\|_F.
\end{aligned}$$

(b) We have

$$\begin{aligned}
\|Ax\|_2 &= \sqrt{\sum_i \left(\sum_j A_{ij} x_j \right)^2} \\
&\leq \sqrt{\sum_i \left(\left(\sum_j A_{ij}^2 \right) \left(\sum_j x_j^2 \right) \right)} && \text{(Cauchy-Schwarz)} \\
&= \sqrt{\left(\sum_j x_j^2 \right) \left(\sum_i \sum_j A_{ij}^2 \right)} \\
&= \|A\|_F \|x\|_2.
\end{aligned}$$

Problem 3 (20 pts). Redo **Problem 2**, but replace Frobenius norm with the following max norm:

$$\|A\|_{\max} = \max_{i,j} |A_{ij}|$$

(a) (10 pts) Prove properties (1), (2) and (3) for max norm.

(b) (10 pts) Does max norm satisfies the following condition?

$$\|Ax\|_2 \leq \|A\|_{\max} \|x\|_2 \quad \forall x \in \mathbb{R}^n, A \in \mathbb{R}^{n \times n}$$

If so, provide a proof. If not, provide a counter example.

Solution.

(a) Since $|A_{ij}| \geq 0$, we know that $\|A\|_{\max} = \max_{i,j} |A_{ij}| \geq 0$. This proves (1).

For (2), we have

$$\begin{aligned}
\|A+B\|_{\max} &= \max_{i,j} |A_{ij} + B_{ij}| \\
&\leq \max_{i,j} (|A_{ij}| + |B_{ij}|) \\
&\leq \max_{i,j} \left(|A_{ij}| + \max_{i',j'} |B_{i'j'}| \right) \\
&= \max_{i,j} |A_{ij}| + \max_{i',j'} |B_{i'j'}| \\
&= \|A\|_{\max} + \|B\|_{\max}.
\end{aligned}$$

For (3), we have

$$\begin{aligned}
\|\alpha A\|_{\max} &= \max_{i,j} |\alpha A_{ij}| \\
&= |\alpha| \max_{i,j} |A_{ij}| \\
&= |\alpha| \|A\|_{\max}.
\end{aligned}$$

(b) The inequality does not hold. For example, let

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Then $\|Ax\| = 2\sqrt{2}, \|A\|_{\max} = 1$ and $\|x\|_2 = \sqrt{2}$. Therefore, $\|Ax\| > \|A\|_{\max}\|x\|_2$.

Problem 4 (10 pts). Assume A is a sparse matrix stored in the following compressed column format

- value array: a
- row indices array: row_ind
- column pointer array: col_ptr

and n is the number of the columns. Please give the MATLAB codes of these functions. You can only use the above information. For example, you cannot use a function to know the length of an array.

(a) (5 pts) the max norm of A

(b) (5 pts) the Frobenius norm of A

Solution.

(a)
```
v = 0.0
for i = 1:col_ptr(n+1)-1
  if a(i) > 0
    abs_a = a(i);
  else
    abs_a = -a(i);
```

```
    end
    if abs_a > v
      v = abs_a;
    end
  end
```

(b)
```
v = 0.0;
for i = 1:col_ptr(n+1)-1
  v = v + a(i)^2;
end
v = sqrt(v);
```

Note: You cannot use any special MATLAB function as we have mentioned that in the problem statement.

Problem 5 (30 pts).

(a) (15 pts) On the page 10 of the slide "sparse_iterative1.pdf", we have a theorem to check the convergence of Jacobi methods with the assumption

$$\rho(M^{-1}N) < 1.$$

Now, we drop this assumption but directly check whether

$$(M^{-1}N)^k \to 0, \text{ as } k \to \infty \tag{4}$$

with the matrix

$$A = \begin{bmatrix} 2 & 0 & -1 \\ 1 & -7 & 0 \\ 0 & 1 & -3 \end{bmatrix}.$$

Specifically, please derive the analytic form of $(M^{-1}N)^k, \forall k$, and show that (4) holds.

(b) (15 pts) Let us check the Gauss-Seidel method with the iteration

$$x^{(k+1)} = L^{-1}(b + Ux^{(k)}),$$

where

$$L = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \ddots & 0 \\ \vdots & \vdots & \ddots & 0 \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, U = \begin{bmatrix} 0 & -a_{12} & \cdots & -a_{1n} \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & -a_{(n-1)n} \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Please check whether

$$(L^{-1}U)^k \to 0, \text{ as } k \to \infty \tag{5}$$

with the same matrix A in (a), and explain why (5) implies the convergence of Gauss-Seidel Method. Specifically, from (5) you must derive the relationship between

$$x^{k+1} - x^* \text{ and } x^k - x^*$$

and also the analytic from of $(L^{-1}U)^k$.

Solution.

(a) We have

$$M = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -7 & 0 \\ 0 & 0 & -3 \end{bmatrix}, N = \begin{bmatrix} 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}.$$

Then, we calculate the power of $M^{-1}N$ to these general forms

$$(M^{-1}N)^{3n+1} = \begin{bmatrix} 0 & 0 & 1/(2\cdot(2\cdot 3\cdot 7)^n) \\ 1/(7\cdot(2\cdot 3\cdot 7)^n) & 0 & 0 \\ 0 & 1/(3\cdot(2\cdot 3\cdot 7)^n) & 0 \end{bmatrix}, n = 0, 1, \ldots$$

$$(M^{-1}N)^{3n+2} = \begin{bmatrix} 0 & 1/((2\cdot 3)\cdot(2\cdot 3\cdot 7)^n) & 0 \\ 0 & 0 & 1/((2\cdot 7)\cdot(2\cdot 3\cdot 7)^n) \\ 1/((3\cdot 7)\cdot(2\cdot 3\cdot 7)^n) & 0 & 0 \end{bmatrix}, n = 0, 1, \ldots$$

$$(M^{-1}N)^{3n} = \begin{bmatrix} 1/(2\cdot 3\cdot 7)^n & 0 & 0 \\ 0 & 1/(2\cdot 3\cdot 7)^n & 0 \\ 0 & 0 & 1/(2\cdot 3\cdot 7)^n \end{bmatrix}, n = 1, 2, \ldots$$

and we derive the general formula

$$(M^{-1}N)^k = \begin{cases} \begin{bmatrix} 0 & 0 & 1/(2\cdot(2\cdot 3\cdot 7)^{\lfloor k/3\rfloor}) \\ 1/(7\cdot(2\cdot 3\cdot 7)^{\lfloor k/3\rfloor}) & 0 & 0 \\ 0 & 1/(3\cdot(2\cdot 3\cdot 7)^{\lfloor k/3\rfloor}) & 0 \end{bmatrix} & k = 1, 4, \ldots \\ \begin{bmatrix} 0 & 1/((2\cdot 3)\cdot(2\cdot 3\cdot 7)^{\lfloor k/3\rfloor}) & 0 \\ 0 & 0 & 1/((2\cdot 7)\cdot(2\cdot 3\cdot 7)^{\lfloor k/3\rfloor}) \\ 1/((3\cdot 7)\cdot(2\cdot 3\cdot 7)^{\lfloor k/3\rfloor}) & 0 & 0 \end{bmatrix} & k = 2, 5, \ldots \\ \begin{bmatrix} 1/(2\cdot 3\cdot 7)^{k/3} & 0 & 0 \\ 0 & 1/(2\cdot 3\cdot 7)^{k/3} & 0 \\ 0 & 0 & 1/(2\cdot 3\cdot 7)^{k/3} \end{bmatrix} & k = 3, 6, \ldots \end{cases} \tag{6}$$

The equation (6) implies that

$$(M^{-1}N)^k \to 0, \text{ as } k \to \infty,$$

since

$$\left|\frac{1}{2\cdot 3\cdot 7}\right| < 1.$$

Therefore, we have that

$$x_{k+1} - x^* = (M^{-1}N)^k(x_1 - x^*)$$

so it implies that

$$x_{k+1} - x^* \to 0, \text{ as } k \to \infty,$$

which means we can use Jacobi Method to get the solution $x^* = A^{-1}b$.

(b) We have

$$L = \begin{bmatrix} 2 & 0 & 0 \\ 1 & -7 & 0 \\ 0 & 1 & -3 \end{bmatrix}, U = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

and

$$L^{-1} = \begin{bmatrix} 1/2 & 0 & 0 \\ 1/(2\cdot 7) & -1/7 & 0 \\ 1/(2\cdot 3\cdot 7) & -1/(3\cdot 7) & -1/3 \end{bmatrix}.$$

Thus,

$$\begin{aligned} L^{-1}U &= \begin{bmatrix} 0 & 0 & 1/2 \\ 0 & 0 & 1/(2\cdot 7) \\ 0 & 0 & 1/(2\cdot 3\cdot 7) \end{bmatrix} \\ (L^{-1}U)^2 &= \begin{bmatrix} 0 & 0 & 1/(2\cdot(2\cdot 3\cdot 7)) \\ 0 & 0 & 1/((2\cdot 7)\cdot(2\cdot 3\cdot 7)) \\ 0 & 0 & 1/(2\cdot 3\cdot 7)^2 \end{bmatrix} \\ (L^{-1}U)^3 &= \begin{bmatrix} 0 & 0 & 1/(2\cdot(2\cdot 3\cdot 7)^2) \\ 0 & 0 & 1/((2\cdot 7)\cdot(2\cdot 3\cdot 7)^2) \\ 0 & 0 & 1/(2\cdot 3\cdot 7)^3 \end{bmatrix} \end{aligned}$$

and the general form

$$(L^{-1}U)^k = \begin{bmatrix} 0 & 0 & 1/(2\cdot(2\cdot 3\cdot 7)^{k-1}) \\ 0 & 0 & 1/((2\cdot 7)\cdot(2\cdot 3\cdot 7)^{k-1}) \\ 0 & 0 & 1/(2\cdot 3\cdot 7)^k \end{bmatrix},$$

which implies that

$$(L^{-1}U)^k \to 0, \text{ as } k \to \infty.$$

For the reason that (5) implies the convergence of Gauss-Seidel Method, we assume that there exists a solution

$$x^* = A^{-1}b.$$

Therefore, the iteration

$$x^{(k+1)} = L^{-1}(b + Ux^{(k)})$$

becomes

$$\begin{aligned} x^{(k+1)} =& L^{-1}(Ax^* + Ux^{(k)}) \\ =& L^{-1}((L-U)x^* + Ux^{(k)}) \\ =& L^{-1}(L-U)x^* + L^{-1}Ux^{(k)} \\ =& x^* - L^{-1}Ux^* + L^{-1}Ux^{(k)} \\ =& x^* + L^{-1}U(x^{(k)} - x^*) \end{aligned}$$

That is,

$$\begin{aligned} &(x^{(k+1)} - x^*) = L^{-1}U(x^{(k)} - x^*) \\ \Rightarrow &(x^{(k+1)} - x^*) = (L^{-1}U)^k(x^{(1)} - x^*), \end{aligned}$$

so

$$(L^{-1}U)^k \to 0, \text{ as } k \to \infty$$

implies that

$$x^{(k+1)} \to x^*, \text{ as } k \to \infty.$$