Numerical Methods 2021 — Midterm 2

Solutions

Problem 1 (20 pts). Consider an **upper triangular** matrix A stored in a compressed **row** format (CSR)

(a) (5 pts) Let

$$A = \begin{bmatrix} 5 & 2 & 8 \\ 0 & 4 & 1 \\ 0 & 0 & 3 \end{bmatrix}$$

Show how A is stored in compressed **row** format in the following table.

- **Here we consider array index beginning with 1.**
- **For column indexes with the same row index, you should store them in an ascending order in the table.**

array_index	1	2	3	4	5	6
a						
acol_ind						
arow_ptr						

(b) (10 pts) Assume A is stored in the same way as in (a). Give a code to solve any upper triangular linear equation

$$A\boldsymbol{x} = \boldsymbol{b},$$

where we assume $A_{ii} \neq 0$.

(c) (5 pts) Let

$$A = \begin{bmatrix} 5 & 2 & 8 \\ 0 & 4 & 1 \\ 0 & 0 & 3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 68 \\ 11 \\ 21 \end{bmatrix}$$

Run your code to solve the system and output your answer.

Solution.

(a) The CSR matrix format of A is

array_index	1	2	3	4	5	6
a	5	2	8	4	1	3
acol_ind	1	2	3	2	3	3
arow_ptr	1	4	6	7		

(b)
```
function x = upper_linear_solver(A, b)
    n = size(A, 2);        % A is m, n dim array
    [a, acol_ind, arow_ptr] = getCSR(A);

    for i = n:-1:1
        Aijxj = 0;          % sum of Aij * xj, where j = i+1, ..., n
        Aii = a(arow_ptr(i));
        for k = (arow_ptr(i)+1):(arow_ptr(i+1)-1)
            j = acol_ind(k);  % column c, value v
            Aijxj = Aijxj + a(k) * x(j);
        end
        x(i) = (b(i) - Aijxj) / Aii;
    end
end
```

(c)
```
iter 1
        Aijxj = 0
        x[3] = b[3] - Aijxj / 3 = 7
iter 2
        Aijxj = 0
        Aijxj = Aijxj + a[5] * x[acol_ind[5]]
              = Aijxj + (a[5] * x[3])
              = Aijxj + (1 * 7)
              = 7
        x[2] = b[2] - Aijxj / 4 = 1
iter 3
        Aijxj = 0
        Aijxj = Aijxj + a[2] * x[acol_ind[2]]
              = Aijxj + (a[2] * x[2])
              = Aijxj + (2 * 1)
              = 2
        Aijxj = Aijxj + a[3] * x[acol_ind[3]]
              = Aijxj + (a[3] * x[3])
              = Aijxj + (8 * 7)
              = 58
        x[1] = b[1] - Aijxj / 5 = 2

ans =

     2
     1
     7
```

Problem 2 (30 pts). Consider the following linear equation,

$$A\boldsymbol{x} = \boldsymbol{b},$$

where A is an n-by-n symmetric positive-definite matrix. Assume A is a sparse matrix stored in compressed **column** (CSC) format. **For row indexes with the same column index, we require that the CSC format stores them in an ascending order.**

(a) (10 pts) In the Gauss-Seidel method, in each iteration we update the i-th element of $\boldsymbol{x}$, x_i, by a value δ

$$x_i \leftarrow x_i + \delta.$$

Suppose we already cached a vector

$$\boldsymbol{c} = \boldsymbol{b} - A\boldsymbol{x},$$

- How to represent the value δ by using values in $\boldsymbol{c}$ and diagonal values of A?
- Suppose we would like to maintain the vector $\boldsymbol{c}$ by using the new $\boldsymbol{x}$. How to update $\boldsymbol{c}$ by using δ and components in A?

(b) (15 pts) Give the code to do the Gauss-Seidel method by using the technique in the prob. (a).

- We require that it can take any vector as the first iterate.
- Note that you may need to extract diagonal elements to a dense vector first.
- Also, you may need to initialize an dense vector $\boldsymbol{c}$ to adapt the technique derived in prob. (a).

(c) (5 pts) Let

$$A = \begin{bmatrix} 4 & 1 & 1 \\ 1 & 4 & 0 \\ 1 & 0 & 3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 18 \\ 25 \\ 25 \end{bmatrix}$$

Show how A is stored in compressed **column** format (CSC) in the following table.

- **Here we consider array index beginning with 1.**

	1	2	3	4	5	6	7
a							
arow_ind							
acol_ptr							

Given initial solution

$$\boldsymbol{x}_0 = \begin{bmatrix} 0 \\ 14 \\ 0 \end{bmatrix}$$

Run your code to solve the system. You need to details such as δ, $\boldsymbol{x}$ and the vector $\boldsymbol{c}$ at each step.

Solution.

(a) The new value $\bar{x}_i$ by using the Gauss-Seidel method is

$$\bar{x}_i = \frac{b_i - \sum_{j \neq i} A_{ij} x_j}{A_{ii}}.$$

If we minus both sides with the original value x_i, we then have

$$\delta = \bar{x}_i - x_i = \frac{b_i - \sum_{j \neq i} A_{ij} x_j}{A_{ii}} - x_i = \frac{b_i - \sum_j A_{ij} x_j}{A_{ii}} = \frac{c_i}{A_{ii}}.$$

For $\boldsymbol{c}$, we update it by

$$\boldsymbol{c} = \boldsymbol{b} - A(\boldsymbol{x} + \delta \boldsymbol{e}_i) = \boldsymbol{c} - \delta \begin{bmatrix} A_{1i} \\ \vdots \\ A_{ni} \end{bmatrix},$$

where $\boldsymbol{e}_i = [0, \cdots, 1, \cdots, 0]^T$ is a standard unit vector.

(b)
```
function y = Gauss_Seidel(A, x, b)
    n = size(A, 2);
    [a, arow_ind, acol_ptr] = getCSC(A);

    % Initialize b - Ax and Adiag
    for i = 1:n
        c(i) = b(i);
        for j = acol_ptr(i):(acol_ptr(i+1)-1)
            r = arow_ind(j);  v = a(j);
            c(r) = c(r) - v * x(i);

            if i == r
                Adiag(i) = v;
            end
        end
    end

    % Run Gauss_Seidel until b - Ax is zero vector
    while sum(c) ~= 0
        for i = 1:n
            delta = c(i) / Adiag(i);
            if delta ~= 0
                x(i) = x(i) + delta;
                for j = acol_ptr(i):(acol_ptr(i+1)-1)
                    r = arow_ind(j);  v = a(j);
                    c(r) = c(r) - v * delta;
                end
            end
        end
    end

    y = x;
end
```

(c) The CSC format of A is

	1	2	3	4	5	6	7
a	4	1	1	1	4	1	3
arow_ind	1	2	3	1	2	1	3
acol_ptr	1	4	6	8			

```
iter 1
        update x[1]
                delta = 1
                x   =  [1  14   0]
                c   =  [0 -32  24]
        update x[2]
                delta = -8
```

```
                x    =  [1   6   0]
                c    =  [8    0   24]
        update x[3]
                delta = 8
                x    =  [1   6   8]
                c    =  [0   0   0]

y =

     1
     6
     8
```

Problem 3 (35 pts). Consider a linear system $A\boldsymbol{x} = \boldsymbol{b}$:

$$\begin{bmatrix} 2 & 0 & 2 \\ 0 & 1 & 0 \\ 2 & 0 & 3 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

(a) (5 pts) Is the matrix A symmetric positive definite? If True, please prove it. Otherwise, give a counterexample to show that is False.

(b) (5 pts) Take $\boldsymbol{x}_0 = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T$. Do two CG iterations and show what $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are.

(c) (5 pts) Is $\boldsymbol{x}_2$ a solution or not?

(d) (10 pts) Let $\boldsymbol{r}_1$ and $\boldsymbol{p}_1$ be vectors from the aforementioned procedures. Solve the following optimization problem with the variable $\boldsymbol{p}$:

$$\begin{aligned} \min_{\boldsymbol{p}} \quad & \|\boldsymbol{p} - \boldsymbol{r}_1\|_2^2 \\ \text{s.t.} \quad & \boldsymbol{p} \in \text{span}\{A\boldsymbol{p}_1\}^{\perp} \end{aligned} \tag{1}$$

How is the solution of this problem connected to $\boldsymbol{p}_2$ obtained in the CG procedure?

(e) (5 pts) If we denote the solution of (1) as $\bar{\boldsymbol{p}}_2$ and calculate $\bar{\alpha}_2$, $\boldsymbol{x}_2$ by

$$\bar{\alpha}_2 = \frac{\bar{\boldsymbol{p}}_2^T \boldsymbol{r}_1}{\bar{\boldsymbol{p}}_2^T A \bar{\boldsymbol{p}}_2}, \ \boldsymbol{x}_2 = \boldsymbol{x}_1 + \bar{\alpha}_2 \bar{\boldsymbol{p}}_2.$$

What are $\bar{\alpha}_2$, $\boldsymbol{x}_2$?

(f) (5 pts) In slides, we have a theorem by writing

$$A = I + B$$

and checking the relationship between rank(B) and the number of CG steps. From what you have observed in (a)-(d), can you say that rank(B) must be greater or equal to some value? Then check B to confirm the result.

One may indeed make mistakes in doing the calculation. However, if you understand the concept of CG, you should be able to easily validate your results.

Solution.

(a) Since

$$A = A^T,$$

A is symmetric. Given a vector $\boldsymbol{v} = [v_1, v_2, v_3]^T$,

$$\boldsymbol{v}^T A \boldsymbol{v} = 2v_1^2 + 4v_1v_3 + 3v_3^2 + v_2^2 = 2(v_1 + v_3)^2 + v_3^2 + v_2^2$$

That implies

$$\boldsymbol{v}^T A \boldsymbol{v} = 0 \text{ iff } \boldsymbol{v} = 0.$$

Therefore, A is a symmetric positive definite matrix.

(b)

$$\boldsymbol{x}_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},\ \boldsymbol{r}_0 = \boldsymbol{b} = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix},$$

$$\Rightarrow \boldsymbol{p}_1 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix},\ \alpha_1 = \frac{\boldsymbol{r}_0^T \boldsymbol{r}_0}{\boldsymbol{p}_1^T A \boldsymbol{p}_1} = \frac{3}{2},\ \boldsymbol{x}_1 = \boldsymbol{x}_0 + \alpha_1 \boldsymbol{p}_1 = \begin{bmatrix} -3/2 \\ -3/2 \\ 3/2 \end{bmatrix}, \boldsymbol{r}_1 = \boldsymbol{r}_0 - \alpha_1 A \boldsymbol{p}_1 = \begin{bmatrix} -1 \\ 1/2 \\ -1/2 \end{bmatrix}$$

$$\Rightarrow \beta_2 = \frac{\boldsymbol{r}_1^T \boldsymbol{r}_1}{\boldsymbol{r}_0^T \boldsymbol{r}_0} = \frac{1}{2}, \boldsymbol{p}_2 = \boldsymbol{r}_1 + \beta_2 \boldsymbol{p}_1 = \begin{bmatrix} -3/2 \\ 0 \\ 0 \end{bmatrix}, \alpha_2 = \frac{\boldsymbol{r}_1^T \boldsymbol{r}_1}{\boldsymbol{p}_2^T A \boldsymbol{p}_2} = \frac{1}{3}, \boldsymbol{x}_2 = \boldsymbol{x}_1 + \alpha_2 \boldsymbol{p}_2 = \begin{bmatrix} -2 \\ -3/2 \\ 3/2 \end{bmatrix},$$

(c) Since

$$\boldsymbol{r}_2 = \boldsymbol{r}_1 - \alpha_2 A \boldsymbol{p}_2 = \begin{bmatrix} 0 \\ 1/2 \\ 1/2 \end{bmatrix}$$

is not zero, $\boldsymbol{x}_2$ is not the solution.

(d) We can derive that

$$\text{span}\{A\boldsymbol{p}_1\} = \begin{bmatrix} 0 \\ -t \\ t \end{bmatrix}, t \in \mathbb{R},$$

so $\text{span}\{A\boldsymbol{p}_1\}^\perp$ is equivalent to

$$-p_2 + p_3 = 0$$

Therefore, (1) is equivalent to

$$\begin{aligned} \min_{\boldsymbol{p}} \quad & \left\| \begin{bmatrix} p_1 + 1 \\ p_2 - 1/2 \\ p_3 + 1/2 \end{bmatrix} \right\|_2^2 \equiv \min_{\boldsymbol{p}} \quad \left\| \begin{bmatrix} p_1 + 1 \\ p_2 - 1/2 \\ p_2 + 1/2 \end{bmatrix} \right\|_2^2 \equiv \min_{\boldsymbol{p}} \quad (p_1 + 1)^2 + (p_2 - 1/2)^2 + (p_2 + 1/2)^2 \\ \text{s.t.} \quad & -p_2 + p_3 = 0 \end{aligned}$$

and

$$(p_1 + 1)^2 + (p_2 - 1/2)^2 + (p_2 + 1/2)^2 = (p_1 + 1)^2 + 2p_2^2 + 1/2.$$

Thereby, the solution is $\begin{bmatrix} -1 & 0 & 0 \end{bmatrix}^T$, and we can see that $\boldsymbol{p}$ is parallel to $\boldsymbol{p}_2$.

(e)
$$\bar{\boldsymbol{p}}_2 = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix},\ \bar{\alpha}_2 = \frac{\bar{\boldsymbol{p}}_2^T \boldsymbol{r}_1}{\bar{\boldsymbol{p}}_2^T A \bar{\boldsymbol{p}}_2} = \frac{1}{2},\ \boldsymbol{x}_2 = \boldsymbol{x}_1 + \bar{\alpha}_2 \bar{\boldsymbol{p}}_2 = \begin{bmatrix} -3/2 \\ -3/2 \\ 3/2 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -2 \\ -3/2 \\ 3/2 \end{bmatrix}.$$

(f) From the theorem on page 6 of Conjugate gradient methods part4,

$$\text{the rank of } B \text{ is at least two} \tag{2}$$

because $\boldsymbol{x}_2$ is not a solution. Let us check on the rank of B. The matrix B can be found by

$$A = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 1 & 0 \\ 2 & 0 & 3 \end{bmatrix} = I + \begin{bmatrix} 1 & 0 & 2 \\ 0 & 0 & 0 \\ 2 & 0 & 2 \end{bmatrix} = I + B$$

The rank of B is two since there are only two linear independent rows in B. Thus, the result (2) holds.

Problem 4 (15 pts). In ours slides, we use

$$f(x) = \cos(x)$$

as an example to run the Newton method. Now we would like to check if it satisfies the theorem proved in slides and have the quadratic convergence.

(a) (5 pts) Is $f(x)$ Lipschitz continuous? If so, prove it and give an α.

(b) (5 pts) For this particular $f(x)$, can you directly apply Taylor expansion to prove Lemma 3? (Hint: Check the Taylor theorem.)

(c) (5 pts) Now consider

$$x^* = \frac{\pi}{2}$$

and

$$\{x^k\} \to x^*.$$

Find $\bar{\beta}$ and δ such that the second result of the theorem holds. That is,

$$|x^{k+1} - x^*| \le \delta |x^k - x^*|^2,\ \forall k \ge L.$$

Solution.

(a) We know that

$$f'(x) = -\sin(x).$$

By Mean Value Theorem, for any x and y, there exists t between x and y such that

$$\sin(x) - \sin(y) = (x - y)\sin'(t).$$

We have

$$|\sin'(t)| = |\cos(t)| \le 1,$$

so

$$|\sin(x) - \sin(y)| = |(x - y)\sin'(t)| = |x - y||\cos(t)| \le |x - y|,\ \forall x, y.$$

Thus we can choose

$$\alpha = 1.$$

(b) Taylor Theorem tells us that the Taylor expansion of $f(y)$ at x:

$$\cos(y) = \cos(x) - \sin(x)(y-x) - \frac{1}{2}\cos(t)(y-x)^2,$$

where t is between x and y. Therefore, the error term is

$$e(x,y) = -\frac{1}{2}\cos(t)(y-x)^2,$$

and

$$\begin{aligned} |e(x,y)| &= |\frac{1}{2}\cos(t)(y-x)^2| \\ &= \frac{1}{2}|\cos(t)||(y-x)^2| \\ &\leq \frac{1}{2}|y-x|^2 \end{aligned}$$

(c) Follow the slides, we have

$$\bar{\beta} = |f'(x^*)^{-1}| = \left|\frac{1}{\sin(\pi/2)}\right| = 1$$

and

$$\delta = \alpha\bar{\beta} = 1.$$