

# Math 221 Homework 3

Atharv Sampath

Fall 2025

## Problem 1

- **Question 2.9:** Input:  $n$ ; arrays  $d[1..n]$  (diagonal),  $u[1..n-1]$  (superdiagonal) of an upper bidiagonal  $B$

```
# Infinity norm of B
normB = abs(d[n])
for i = 1..n-1:
    row = abs(d[i]) + abs(u[i])
    if row > normB: normB = row

# Precompute magnitudes and ratios for |B^{-1}|
for i = 1..n: t[i] = 1 / abs(d[i])          # t_i = 1/|d_i|
for i = 1..n-1: r[i] = abs(u[i]) / abs(d[i+1])  # r_i = |u_i|/|d_{i+1}|

# Backward recurrence for partial products of ratios
p[n] = 1
for i = n-1..1 step -1:
    p[i] = 1 + r[i] * p[i+1]

# Row sums of |B^{-1}| and its infinity norm
normBinv = 0
for i = 1..n:
    s = t[i] * p[i]           # s_i = sum_{j>=i} |(B^{-1})_{ij}|
    if s > normBinv: normBinv = s

return kappa = normB * normBinv
```

Let  $B$  be  $n \times n$  upper bidiagonal with diagonal  $d_1, \dots, d_n$  and superdiagonal  $u_1, \dots, u_{n-1}$ . The infinity norm is the maximum row sum  $\|B\|_\infty = \max\{|d_n|, \max_{i \leq n-1}(|d_i| + |u_i|)\}$ . Since  $B$  is upper triangular,  $B^{-1}$  is upper triangular with  $(B^{-1})_{ij} = (-1)^{j-i} \frac{u_i \cdots u_{j-1}}{d_i \cdots d_j}$  for  $i \leq j$ . Taking absolute values and defining  $t_i = 1/|d_i|$  and  $r_i = |u_i|/|d_{i+1}|$ , the absolute row sum of row  $i$  is

$$s_i = \sum_{j=i}^n \frac{|u_i \cdots u_{j-1}|}{|d_i \cdots d_j|} = t_i (1 + r_i + r_i r_{i+1} + \cdots + r_i r_{i+1} \cdots r_{n-1}).$$

Set  $p_n = 1$  and  $p_i = 1 + r_i p_{i+1}$  for  $i = n-1, \dots, 1$ . Then  $s_i = t_i p_i$  and  $\|B^{-1}\|_\infty = \max_i s_i$ . The condition number is  $\kappa_\infty(B) = \|B\|_\infty \|B^{-1}\|_\infty$ . Operation counts meet the target: one pass for  $\|B\|_\infty$ , formation of  $t_i, r_i$  by  $n + (n - 1)$  absolute values and divisions, a backward pass with  $n - 1$  multiplications and additions, and  $O(n)$  comparisons.

- **Question 2.11:** Let  $A$  be real symmetric positive definite and fix  $i \neq j$ . The  $2 \times 2$  principal submatrix

$$M = \begin{pmatrix} a_{ii} & a_{ij} \\ a_{ij} & a_{jj} \end{pmatrix}$$

is also positive definite. Hence  $a_{ii} > 0$ ,  $a_{jj} > 0$ , and  $\det M = a_{ii}a_{jj} - a_{ij}^2 > 0$ . Therefore

$$a_{ij}^2 < a_{ii}a_{jj} \implies |a_{ij}| < \sqrt{a_{ii}a_{jj}}.$$

Equivalently, one may view  $\langle x, y \rangle = x^T A y$  as an inner product; then by Cauchy–Schwarz,  $|a_{ij}| = |\langle e_i, e_j \rangle| \leq \|e_i\|_A \|e_j\|_A = \sqrt{a_{ii}a_{jj}}$ , and equality cannot hold for  $i \neq j$  in a positive definite inner product, giving the same strict inequality.

- **Question 2.13:**

1. Let  $A$  be nonsingular. For vectors  $u, v$  with  $1 + v^T A^{-1} u \neq 0$ , set

$$X = A^{-1} - \frac{A^{-1} u v^T A^{-1}}{1 + v^T A^{-1} u}.$$

Then

$$(A + uv^T)X = I + uv^T A^{-1} - \frac{uv^T A^{-1}}{1 + v^T A^{-1} u} = I + uv^T A^{-1} \left( 1 - \frac{1}{1 + v^T A^{-1} u} \right) = I,$$

and similarly  $X(A + uv^T) = I$ . Hence

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1} u v^T A^{-1}}{1 + v^T A^{-1} u}.$$

More generally, for  $U, V \in \mathbb{R}^{n \times k}$  define  $T = I_k + V^T A^{-1} U$ . If  $T$  is nonsingular, let

$$X = A^{-1} - A^{-1} U T^{-1} V^T A^{-1}.$$

A direct multiplication gives

$$(A + UV^T)X = I + UV^T A^{-1} - U(I + V^T A^{-1} U)T^{-1}V^T A^{-1} = I,$$

and symmetrically  $X(A + UV^T) = I$ . Thus, if and only if  $T$  is nonsingular (equivalently  $A + UV^T$  is nonsingular),

$$(A + UV^T)^{-1} = A^{-1} - A^{-1}U(I + V^TA^{-1}U)^{-1}V^TA^{-1}.$$

2. Assume we can solve linear systems with  $A$  quickly. Compute

$$z = A^{-1}c \quad \text{and} \quad w = A^{-1}u$$

by two solves with  $A$ . Then apply Sherman–Morrison:

$$y = (A + uv^T)^{-1}c = z - \frac{w v^T z}{1 + v^T w}.$$

This uses two solves with  $A$ , one inner product  $v^T z$ , and a few BLAS-1 operations.

3. Use  $A$  as a right-preconditioner/iterative refinement for solving  $By = c$ . Initialize  $y_0$  (e.g.  $y_0 = z = A^{-1}c$ ). For  $k = 0, 1, 2, \dots$ ,

$$r_k = c - By_k, \quad \text{solve } Ae_k = r_k, \quad y_{k+1} = y_k + e_k.$$

In exact arithmetic, with error  $e_k^* = y^* - y_k$  where  $By^* = c$ ,

$$e_{k+1}^* = (I - A^{-1}B)e_k^* = A^{-1}(A - B)e_k^*.$$

Hence

$$\|e_{k+1}^*\| \leq \|A^{-1}(A - B)\| \|e_k^*\|,$$

so the iteration converges linearly whenever  $\|A^{-1}(A - B)\| < 1$ , with rate governed by  $\rho(I - A^{-1}B) \leq \|A^{-1}(A - B)\|$ . In particular, if  $\|A - B\|$  is “small” relative to  $\|A^{-1}\|$ , the method converges rapidly.

- **Question 2.16:** Let  $A \in \mathbb{R}^{n \times n}$  be symmetric positive definite and suppose we overwrite its lower triangle with the Cholesky factor  $L$  such that  $A = LL^T$ . Choose a block size  $b$  (panel width). Partition indices  $k = 1, 1 + b, 1 + 2b, \dots$

$$A = \begin{bmatrix} A_{11} & & \\ A_{21} & A_{22} & \\ A_{31} & A_{32} & \ddots \end{bmatrix} \quad \text{where each diagonal block is } b \times b \text{ except possibly the last.}$$

For  $k = 1, 1 + b, \dots, n$  set  $j = \min(b, n - k + 1)$ , and let the current panel be rows/cols  $k:k+j-1$ .

1. (Update the diagonal block by previous panels; SYRK)

$$A_{kk} \leftarrow A_{kk} - A_{k,1:k-1} A_{k,1:k-1}^T \quad (\text{SYRK: } j \times j \text{ with } k-1 \text{ width}).$$

2. (Unblocked Cholesky on small block; POTRF)

$$A_{kk} \leftarrow \text{chol}(A_{kk}) \equiv L_{kk} (\text{POTRF on } j \times j).$$

3. (Update the block column below the diagonal block; GEMM)

$$A_{(k+j):n, k:k+j-1} \leftarrow A_{(k+j):n, k:k+j-1} - A_{(k+j):n, 1:k-1} A_{k,1:k-1}^T (\text{GEMM}).$$

4. (Triangular solve to form  $L_{21}$ ; TRSM)

$$A_{(k+j):n, k:k+j-1} \leftarrow A_{(k+j):n, k:k+j-1} L_{kk}^{-T} (\text{TRSM, right, lower-triangular, transposed}).$$

After this step the block  $A_{(k+j):n, k:k+j-1}$  is  $L_{21}$ .

5. (Trailing submatrix update; SYRK/GEMM)

$$A_{(k+j):n, (k+j):n} \leftarrow A_{(k+j):n, (k+j):n} - L_{21} L_{21}^T (\text{SYRK for the symmetric part; or GEMM+symmetrize}).$$

Continue until  $k > n$ . The lower triangle of  $A$  now contains  $L$ .

All *large* updates are cast as matrix–matrix operations:

- Step 1 uses SYRK, accumulating the effect of all previously computed panels on the current  $b \times b$  diagonal block.
- Step 3 is a rectangular GEMM producing the block column below the panel.
- Step 4 is a TRSM that applies  $L_{kk}^{-T}$  to the entire panel at once.
- Step 5 updates the trailing symmetric block via SYRK/GEMM, which accounts for the bulk of the  $n^3/3$  flops.

Only Step 2 performs an unblocked factorization, and it is on a small  $j \times j$  block (cost  $O(b^3)$  per panel). With a cache-tuned  $b$ , the vast majority of the work is executed by Level-3 BLAS (cache-friendly, high arithmetic intensity), precisely mimicking the blocked reorganization used for LU in Algorithm 2.10.

• **Question 2.18:**

1. Write

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad A_{11} \in \mathbb{R}^{k \times k} \text{ nonsingular.}$$

Gaussian elimination without pivoting eliminates the block  $A_{21}$  by left-multiplying by

$$L = \begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix}.$$

Then

$$LA = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix}.$$

Thus, after  $k$  steps (i.e., after finishing the first  $k$  columns), the trailing block  $A_{22}$  has been overwritten by the Schur complement

$$S = A_{22} - A_{21}A_{11}^{-1}A_{12}.$$

2. Assume  $A = A^T$ ,  $A_{11} \succ 0$ , and  $A_{22} \prec 0$ . For any  $x \in \mathbb{R}^k$ ,  $y \in \mathbb{R}^{n-k}$ ,

$$\begin{aligned} \begin{bmatrix} x \\ y \end{bmatrix}^T A \begin{bmatrix} x \\ y \end{bmatrix} &= x^T A_{11}x + 2x^T A_{12}y + y^T A_{22}y \\ &= (x + A_{11}^{-1}A_{12}y)^T A_{11}(x + A_{11}^{-1}A_{12}y) + y^T (A_{22} - A_{21}A_{11}^{-1}A_{12})y \\ &= (x + A_{11}^{-1}A_{12}y)^T A_{11}(x + A_{11}^{-1}A_{12}y) + y^T S y. \end{aligned}$$

Since  $A_{11} \succ 0$ , the first term is  $\geq 0$ . Because  $A_{11}^{-1} \succ 0$ ,

$$y^T S y = y^T A_{22}y - (A_{11}^{-1/2}A_{12}y)^T (A_{11}^{-1/2}A_{12}y) < y^T A_{22}y < 0 \quad (y \neq 0).$$

Hence  $S \prec 0$ , in particular  $S$  is nonsingular. Since  $A_{11}$  and  $S$  are both nonsingular, the block LU (or block  $L D L^T$ ) factorization exists,

$$A = \begin{bmatrix} I & 0 \\ A_{21}A_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{11} & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} I & A_{11}^{-1}A_{12} \\ 0 & I \end{bmatrix},$$

so  $A$  is nonsingular and Gaussian elimination without pivoting proceeds in exact arithmetic (no breakdown).

However, it can be numerically unstable because  $A_{11}$  may be arbitrarily small even though  $A$  is nonsingular. For the  $2 \times 2$  symmetric example

$$A = \begin{bmatrix} \varepsilon & 1 \\ 1 & -1 \end{bmatrix}, \quad 0 < \varepsilon \ll 1,$$

we have  $A_{11} = \varepsilon > 0$ ,  $A_{22} = -1 < 0$ , and  $\det A = -\varepsilon - 1 \neq 0$ . Gaussian elimination without pivoting uses the first pivot  $\varepsilon$  and multiplier  $m = 1/\varepsilon$ , producing the trailing entry

$$S = -1 - \frac{1}{\varepsilon}.$$

In floating point, forming  $m = 1/\varepsilon$  amplifies relative errors by a factor of order  $1/\varepsilon$ , and the updated entry has magnitude  $\approx \varepsilon^{-1}$ , yielding a large element growth factor and potentially large forward/backward error even though the exact arithmetic completes successfully. Therefore GE without pivoting may be numerically unstable in this setting.

- **Question 2.20:**

- (a) Factor  $PA = LU$  once with GEPP, then apply  $A^{-1}$  to  $b$  repeatedly  $k$  times reusing  $L, U, P$ ; cost  $\frac{2}{3}n^3 + 2kn^2$  flops.

```
% Input: A (n-by-n nonsingular), b, k > 0
[L,U,P] = lu(A); % GEPP, done once
y = b;
for t = 1:k
y = U \ (L \ (P\*y));
% solve A x = y via LU
end
x = y; % A^k x = b
```

- (b) Compute  $\alpha = c^T A^{-1} b$  by one solve and a dot; with LU available the cost is  $\approx 2n^2 + 2n$  flops (otherwise add  $\frac{2}{3}n^3$ ).

```
% Option 1: solve Ax=b, then c^T x
[L,U,P] = lu(A); % reuse if already computed
x = U \ (L \ (P\*b));
alpha = c' \* x;
```

```
% Option 2: solve A^T y = c, then y^T b (algebraically identical)
% y = L' \ (U' \ (P'\*c));
alpha = y' \* b;
```

- (c) Solve  $AX = B$  (with  $B \in \mathbb{R}^{n \times m}$ ) by factoring once and performing two triangular solves with all  $m$  right-hand sides at once (Level-3 TRSM); total cost  $\frac{2}{3}n^3 + 2n^2m$  flops.

```
% Input: A (n-by-n nonsingular), B (n-by-m)
[L,U,P] = lu(A); % GEPP
Y = L \ (P\*B); % solve L Y = P B (many RHS; BLAS-3 TRSM)
X = U \ Y; % solve U X = Y (many RHS; BLAS-3 TRSM)
```

- **Question 2.21:** We prove by induction on  $n = 2^t$ . For  $n = 1$  the algorithm multiplies scalars, so it is correct. Assume correctness for size  $n/2$ . For size  $n$ , write

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}.$$

Strassen forms

$$\begin{aligned} M_1 &= (A_{11} + A_{22})(B_{11} + B_{22}), & M_2 &= (A_{21} + A_{22})B_{11}, & M_3 &= A_{11}(B_{12} - B_{22}), \\ M_4 &= A_{22}(B_{21} - B_{11}), & M_5 &= (A_{11} + A_{12})B_{22}, & M_6 &= (A_{21} - A_{11})(B_{11} + B_{12}), \\ M_7 &= (A_{12} - A_{22})(B_{21} + B_{22}), \end{aligned}$$

and returns

$$\begin{aligned} C_{11} &= M_1 + M_4 - M_5 + M_7, & C_{12} &= M_3 + M_5, \\ C_{21} &= M_2 + M_4, & C_{22} &= M_1 - M_2 + M_3 + M_6. \end{aligned}$$

By the induction hypothesis each  $M_i$  is the exact product of its subblocks. Direct algebra gives

$$\begin{aligned} C_{11} &= A_{11}B_{11} + A_{12}B_{21}, \\ C_{12} &= A_{11}B_{12} + A_{12}B_{22}, \\ C_{21} &= A_{21}B_{11} + A_{22}B_{21}, \\ C_{22} &= A_{21}B_{12} + A_{22}B_{22}. \end{aligned}$$

Hence

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix} = AB.$$

The recursion therefore multiplies correctly for all  $n$  that are powers of two.