CSE 6643 Homework 4

Karl Hiner

1 Schur Complements I [25 pts]

Consider the matrix

$$M := \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \tag{1}$$

where $A \in \mathbb{K}^{m \times m}$ and $D \in \mathbb{K}^{n \times n}$ are square matrices. The matrix $M/A := D - CA^{-1}B$ is called the Schur complement of A in M.

(a) [5 pts]

Relate the Schur complement M/A to the block LU factorization of M.

The block LU factorization of M is given by

$$M = egin{pmatrix} I & 0 \ CA^{-1} & I \end{pmatrix} egin{pmatrix} A & 0 \ 0 & M/A \end{pmatrix} egin{pmatrix} I & A^{-1}B \ 0 & I \end{pmatrix}.$$

We can verify this by multiplying the terms and verifying we end up with M:

$$\begin{pmatrix} I & 0 \\ CA^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & M/A \end{pmatrix} \begin{pmatrix} I & A^{-1}B \\ 0 & I \end{pmatrix}$$

$$= \begin{pmatrix} I & 0 \\ CA^{-1} & I \end{pmatrix} \begin{pmatrix} A & AA^{-1}B \\ 0 & M/A \end{pmatrix}$$

$$= \begin{pmatrix} I & 0 \\ CA^{-1} & I \end{pmatrix} \begin{pmatrix} A & B \\ 0 & M/A \end{pmatrix}$$

$$= \begin{pmatrix} A & B \\ CA^{-1}A & CA^{-1}B + M/A \end{pmatrix}$$

$$= \begin{pmatrix} A & B \\ C & CA^{-1}B + D - CA^{-1}B \end{pmatrix}$$

$$= \begin{pmatrix} A & B \\ C & D \end{pmatrix} = M$$

(b) [5 pts]

Show that the determinant of M is given by $\det(A) \det(M/A)$.

Expressing M as its block LU factorization:

$$\begin{split} M &= \begin{pmatrix} I & 0 \\ CA^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & M/A \end{pmatrix} \begin{pmatrix} I & A^{-1}B \\ 0 & I \end{pmatrix} \\ \det(M) &= \det \begin{pmatrix} I & 0 \\ CA^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & M/A \end{pmatrix} \begin{pmatrix} I & A^{-1}B \\ 0 & I \end{pmatrix} \end{pmatrix} \\ &= \det \begin{pmatrix} \begin{pmatrix} I & 0 \\ CA^{-1} & I \end{pmatrix} \end{pmatrix} \det \begin{pmatrix} \begin{pmatrix} A & 0 \\ 0 & M/A \end{pmatrix} \end{pmatrix} \det \begin{pmatrix} \begin{pmatrix} I & A^{-1}B \\ 0 & I \end{pmatrix} \end{pmatrix} \\ &= 1 \cdot \det \begin{pmatrix} \begin{pmatrix} A & 0 \\ 0 & M/A \end{pmatrix} \end{pmatrix} \cdot 1 \\ &= \det(A) \det(M/A) \end{split}$$

(c) [7.5 pts]

Let M = LU be the LU factorization of M. Split its factors into block matrices

$$\boldsymbol{L} = \begin{pmatrix} \boldsymbol{L}_{1,1} & \boldsymbol{0} \\ \boldsymbol{L}_{2,1} & \boldsymbol{L}_{2,2} \end{pmatrix}, \quad \boldsymbol{U} = \begin{pmatrix} \boldsymbol{U}_{1,1} & \boldsymbol{U}_{1,2} \\ \boldsymbol{0} & \boldsymbol{U}_{2,2} \end{pmatrix}$$
 (2)

according to the same blocking as M. Show that $M/A = L_{2,2}U_{2,2}$.

$$egin{aligned} m{M} &= m{L}m{U} = egin{pmatrix} m{L}_{1,1} & m{0} \ m{L}_{2,1} & m{L}_{2,2} \end{pmatrix} egin{pmatrix} m{U}_{1,1} & m{U}_{1,2} \ m{0} & m{U}_{2,2} \end{pmatrix} \ egin{pmatrix} m{A} & m{B} \ m{C} & m{D} \end{pmatrix} &= egin{pmatrix} m{L}_{1,1}m{U}_{1,1} & m{L}_{1,1}m{U}_{1,2} \ m{L}_{2,1}m{U}_{1,1} & m{L}_{2,1}m{U}_{1,2} + m{L}_{2,2}m{U}_{2,2} \end{pmatrix} \end{aligned}$$

Equating the lower-right subblocks:

(d) [7.5 pts]

Now consider the 3×3 block matrix

$$N = \begin{pmatrix} N_{1,1} & N_{1,2} & N_{1,3} \\ N_{2,1} & N_{2,2} & N_{2,3} \\ N_{3,1} & N_{3,2} & N_{3,3} \end{pmatrix}.$$
(3)

Prove the Quotient Property of Schur complements

$$(N/N_{1,1})/(N/N_{1,1})_{1,1} = N/\binom{N_{1,1} N_{1,2}}{N_{2,1} N_{2,2}}.$$
 (4)

Here, $(N/N_{1,1})_{1,1}$ is the top left block of $N/N_{1,1}$.

Let
$$N_A = \begin{pmatrix} N_{1,1} & N_{1,2} \\ N_{2,1} & N_{2,2} \end{pmatrix}$$
.

Observe that N/N_A is the Schur complement of N_A in N.

Note: This one is hard, this paper has a proof using Gaussian Elimination, which we can relate to the LU factorization above. $M/A = L_{2,2}U_{2,2}$.

2 Schur Complements II [25 pts + 10 bonus]

Let M be as in the previous problem.

(a) [5 pts]

Divide the inverse of M into blocks (of the same size as those of M) as

$$\boldsymbol{M}^{-1} = \begin{pmatrix} \boldsymbol{\alpha} & \boldsymbol{\beta} \\ \boldsymbol{\gamma} & \boldsymbol{\delta} \end{pmatrix}. \tag{5}$$

Show that the Schur complement M/A is the inverse of δ .

Cite https://www.statlect.com/matrix-algebra/Schur-complement

$$MM^{-1} = I$$
 (def. of matrix inversion, M invertible) $\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = I$ (substitute given M and M^{-1}) $\begin{pmatrix} A\alpha + B\gamma & A\beta + B\delta \\ C\alpha + D\gamma & C\beta + D\delta \end{pmatrix} = \begin{pmatrix} I_m & \mathbf{0} \\ \mathbf{0} & I_n \end{pmatrix}$ (dim $(I_m) = \dim(A) = m$, dim $(I_n) = \dim(D) = n$) $C\beta + D\delta = I_n$ (equate lower-right blocks) $C(-A^{-1}B\delta) + D\delta = I_n$ (equate upper-right blocks, solve for β , substitute) $(D - CA^{-1}B)\delta = I_n$ (factor δ) $(M/A)\delta = I_n$ (def. of M/A) $M/A = \delta^{-1}$ (solve for M/A . QED)

Some generally useful things for probs. below, from HW2

We consider $A \in \mathbb{R}^{m \times m}$ and assume that A^{-1} exists. You may use without proof that if A is symmetric, its largest and smallest eigenvalues are characterized as

$$\lambda_{\max}(\boldsymbol{A}) = \max_{\boldsymbol{x} \in \mathbb{R}^m \setminus \{\boldsymbol{0}\}} \frac{\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}}{\|\boldsymbol{x}\|_2^2}$$

and

$$\lambda_{\min}((A)) = \min_{\boldsymbol{x} \in \mathbb{R}^m \setminus \{\boldsymbol{0}\}} \frac{\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}}{\|\boldsymbol{x}\|_2^2}$$

$$\|\boldsymbol{A}^{-1}\|_2^{-1} \ge \min_{\boldsymbol{x} \in \mathbb{R}^m \setminus \{\boldsymbol{0}\}} \frac{|\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}|}{\|\boldsymbol{x}\|_2^2}$$
(6)

$$\|A^{-1}\|_{2}^{-1} = \min_{x \in \mathbb{R}^{m} \setminus \{0\}} \frac{\|Ax\|_{2}}{\|x\|_{2}}$$
 (7)

Show that for symmetric matrices $M, N \in \mathbb{R}^{m \times m}$,

$$\lambda_{\min}(\boldsymbol{M} + \boldsymbol{N}) \ge \lambda_{\min}(\boldsymbol{M}) + \lambda_{\min}(\boldsymbol{N}) \tag{8}$$

In the same setting as (c), show that

$$\lambda_{\max}(\boldsymbol{M} + \boldsymbol{N}) \le \lambda_{\max}(\boldsymbol{M}) + \lambda_{\max}(\boldsymbol{N}) \tag{9}$$

For \boldsymbol{A} symmetric positive definite, show that

$$\|\boldsymbol{A}^{-1}\|_2^{-1} = \min_{\boldsymbol{x} \in \mathbb{R}^m \setminus \{ \boldsymbol{0} \}} \frac{\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}}{\|\boldsymbol{x}\|_2^2}$$

For any matrix,

$$\sigma_{\max}(\mathbf{A}) = 1/\sigma_{\min}(\mathbf{A}^{-1})$$
, if \mathbf{A}^{-1} exists

For a square matrix,

$$\sigma_{\min}(\mathbf{A}) \leq |\lambda_i(\mathbf{A})| \leq \sigma_{\max}(\mathbf{A}), \forall i$$

(b) [5 pts]

For M symmetric positive definite, show that if λ is an eigenvalue of M/A, then $\lambda_{\min}(M) \le \lambda \le \lambda_{\max}(M)$.

This answer is based on (cite) https://math.stackexchange.com/a/1994997

$$M$$
 is symmetric, so we can express it as $M = \begin{pmatrix} A & B \ B^* & D \end{pmatrix}$. Let $Q \coloneqq \begin{pmatrix} I & A^{-1}B \ 0 & I \end{pmatrix}$,

Then we can express the block LU factorization (1a) of M as

$$M=Q^*igg(egin{array}{cc} A & 0 \ 0 & M/A igg) Q.$$

Let \mathbb{Z} be the set of vectors $z \coloneqq \begin{bmatrix} x \\ y \end{bmatrix}$ satisfying $x + A^{-1}By = 0$. Then for any $z \in \mathbb{Z}$, $Qz = \begin{bmatrix} 0 \\ y \end{bmatrix}$, and so $z^*Mz = y^*(M/A)y$, and we have

$$\lambda_{\min}(\boldsymbol{M}) = \min_{\boldsymbol{x} \in \mathbb{K}^m \setminus \{\boldsymbol{0}\}} \frac{\boldsymbol{x}^* \boldsymbol{M} \boldsymbol{x}}{\|\boldsymbol{x}\|_2^2} \leq \min_{\boldsymbol{z} \in \mathbb{Z} \setminus \{\boldsymbol{0}\}} \frac{\boldsymbol{z}^* \boldsymbol{M} \boldsymbol{z}}{\|\boldsymbol{z}\|_2^2} = \min_{\boldsymbol{y} \in \mathbb{K}^m \setminus \{\boldsymbol{0}\}} \frac{\boldsymbol{y}^*(\boldsymbol{M}/\boldsymbol{A}) \boldsymbol{y}}{\|\boldsymbol{y}\|_2^2} = \lambda_{\min}(\boldsymbol{M}/\boldsymbol{A}),$$

and similarly

$$\lambda_{\max}(\boldsymbol{M}) = \max_{\boldsymbol{x} \in \mathbb{K}^m \setminus \{\boldsymbol{0}\}} \frac{\boldsymbol{x}^* \boldsymbol{M} \boldsymbol{x}}{\|\boldsymbol{x}\|_2^2} \geq \max_{\boldsymbol{z} \in \mathbb{Z} \setminus \{\boldsymbol{0}\}} \frac{\boldsymbol{z}^* \boldsymbol{M} \boldsymbol{z}}{\|\boldsymbol{z}\|_2^2} = \max_{\boldsymbol{y} \in \mathbb{K}^m \setminus \{\boldsymbol{0}\}} \frac{\boldsymbol{y}^*(\boldsymbol{M}/\boldsymbol{A}) \boldsymbol{y}}{\|\boldsymbol{y}\|_2^2} = \lambda_{\max}(\boldsymbol{M}/\boldsymbol{A}).$$

(c) [7.5 pts]

Show that the size of the pivots occurring in the Cholesky or LU factorization of a s.p.d. matrix M is lower bounded by $\|M^{-1}\|^{-1}$.

See 02-21-2023 video 56:20 for partial proof.

Since M is s.p.d., then by (4e) in HW2,

$$\|m{M}^{-1}\|_2^{-1} = \min_{m{x} \in \mathbb{R}^m \setminus \{m{0}\}} \frac{m{x}^T m{M} m{x}}{\|m{x}\|_2^2}$$

$$\|\boldsymbol{M}^{-1}\|^{-1} = \sigma_{\max}(\boldsymbol{M})^{-1} = \lambda_{\max}(\boldsymbol{M})^{-1}$$

(d) [7.5 pts]

For M s.p.d, show that

$$y^*(M/A)y = \min_{x \in \mathbb{K}^m} {x \choose y}^* M {x \choose y}.$$
 (10)

Proof: Assume there exests a vector $\boldsymbol{x} \in \mathbb{K}^m$ such that $\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}^* \boldsymbol{M} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} < \boldsymbol{y}^* (\boldsymbol{M}/\boldsymbol{A}) \boldsymbol{y}.$

Let $\alpha = x + A^{-1}By$. We showed in (b) above that if $\alpha = 0$, then

$$egin{bmatrix} x \ y \end{bmatrix}^* M egin{bmatrix} x \ y \end{bmatrix} = y^* (M/A) y.$$

So we must have $\alpha \neq 0$.

As in (b) above, express M as $\begin{pmatrix} A & B \\ B^* & D \end{pmatrix}$, and let $Q \coloneqq \begin{pmatrix} I & A^{-1}B \\ 0 & I \end{pmatrix}$. Then we can factorize M as

$$M=Q^*igg(egin{array}{cc} A & 0 \ 0 & M/A igg) Q,$$

and we have

$$egin{bmatrix} egin{bmatrix} x \ y \end{bmatrix}^* M egin{bmatrix} x \ y \end{bmatrix} = \ egin{bmatrix} egin{bmatrix} x \ y \end{bmatrix}^* Q^* egin{bmatrix} A & 0 \ 0 & M/A \end{pmatrix} Q egin{bmatrix} x \ y \end{bmatrix} = \ egin{bmatrix} lpha A lpha + y^* (M/A) y. \end{bmatrix}$$

Since $egin{bmatrix} x \ y \end{bmatrix}^* M egin{bmatrix} x \ y \end{bmatrix} < y^* (M/A) y,$ we have

$$egin{aligned} oldsymbol{lpha}^* A oldsymbol{lpha} + oldsymbol{y}^* (M/A) oldsymbol{y} & \oldsymbol{lpha}^* A oldsymbol{lpha} < 0 \end{aligned}$$

However, since M is s.p.d., and A is a principle submatrix of M, then A is also s.p.d., and since $\alpha \neq 0$, we must have $\alpha^* A \alpha > 0$, which contradicts our original assumption, and so we must have $\begin{bmatrix} x \\ y \end{bmatrix}^* M \begin{bmatrix} x \\ y \end{bmatrix} \ge y^* (M/A) y, \forall x \in \mathbb{K}^m$.

(e) [10 bonus pts]

By replacing the min with a min max over suitable variables, derive a version of the results of (d) that is valid for general M.

3 Cholesky and QR [25 pts]

(a) [5 pts]

Assume that the invertible matrix $A \in \mathbb{K}^{m \times m}$ satisfies A = LU. Derive a way to write A as the sum of rank-one matrices in outer product form.

```
def LUfactorizeB(A):
    1A = len(A)
    for k in range (0, 1A-1):
        A[k+1:, k] = A[k+1:, k]/A[k, k]
        A[k+1:, k+1:] -= np.outer(A[k+1:,k],A[k, k+1:])
```

(b) [5 pts]

Show that the LU factorization of an invertible matrix $A \in \mathbb{K}^{m \times m}$ is unique.

Cite

Suppose the LU factorization of A is not unique. Then A can be factorized into two distinct choices of L and U:

$$m{A} = m{L}_1 m{U}_1 = m{L}_2 m{U}_2 \ m{L}_2^{-1} m{L}_1 = m{U}_2 m{U}_1^{-1}$$

Since the inverse of a non-singular lower triangular matrix is lower triangular, and the product of two lower triangular matrices is also lower triangular, the left side is lower triangular. By similar reasoning, the right side must be upper triangular.

For a lower triangular matrix to be equal to an upper triangular matrix, they must both be diagonal. By construction, both L_1 and L_2 , have all diagonal elements equal to 1, and so the product $L_2^{-1}L_1$ also has diagonal elements equal to 1.

Thus, $\boldsymbol{L}_2^{-1}\boldsymbol{L}_1 = \boldsymbol{U}_2\boldsymbol{U}_1^{-1} = \boldsymbol{I}$, implying $\boldsymbol{L}_1 = \boldsymbol{L}_2$ and $\boldsymbol{U}_1 = \boldsymbol{U}_2$, and so the LU factorization is unique.

(c) [5 pts]

Assume that we factor $A \in \mathbb{K}^{m \times m}$ as $A = LL^*$ for lower triangular L. Up to which choices is this factorization unique? Prove your results (as always) and make sure to treat the case $\mathbb{K} = \mathbb{C}$.

Cite

Proceeding similarly to (b), suppose the factorization of $A = LL^*$ is not unique. Then A can be factorized into two distinct choices of L:

$$egin{aligned} oldsymbol{A} &= oldsymbol{L}_1 oldsymbol{L}_1^* &= oldsymbol{L}_2 oldsymbol{L}_2^* \ oldsymbol{I} &= oldsymbol{L}_1^{-1} oldsymbol{L}_2 oldsymbol{L}_1^* oldsymbol{L}_1^* & ext{(move $oldsymbol{L}_*$ terms to one side)} \ oldsymbol{I} &= oldsymbol{L}_1^{-1} oldsymbol{L}_2 ig)^* & ext{(combine conj. transposes, group terms)} \ oldsymbol{I} &= oldsymbol{D} oldsymbol{D}^* & ext{($oldsymbol{D} := oldsymbol{L}_1^{-1} oldsymbol{L}_2)} \ oldsymbol{D}^{-1} &= oldsymbol{D}^* & ext{($oldsymbol{D} := oldsymbol{L}_1^{-1} oldsymbol{L}_2)} \end{aligned}$$

By the same argument as (b) above, $\boldsymbol{D} \coloneqq \boldsymbol{L}_1^{-1}\boldsymbol{L}_2$ is lower triangular, and its inverse is also lower triangular. Since $\boldsymbol{D}^{-1} = \boldsymbol{D}^*$, and the congugate transpose of a lower triangular matrix is upper triangular, \boldsymbol{D} must be diagonal. This, combined with the derived fact that $\boldsymbol{I} = \boldsymbol{D}\boldsymbol{D}^*$ implies that $D_{ii}D_{ii}^* = 1, \forall i$.

Two Cholesky factors of A differ by the signs of their columns.

Can happen when positive semi-definite, but not when positive definite.

(d) [5 pts]

Using these results, prove and explain the relationship between the Cholesky factor of B^*B and the QR factorization of B, for $B \in \mathbb{K}^{m \times n}$ and $m \ge n$.

See 02-21-2023 video for relevant lecture. 1:05:45

1:13:20 "A QR factorization is like a Cholesky factorization, but operating on only half of the matrix."

(e) [5 pts]

For $\mathbf{A} = \mathbf{B}^* \mathbf{B}$ and \mathbf{B} as in (d), write

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{1,1} & \boldsymbol{A}_{1,2} \\ \boldsymbol{A}_{2,1} & \boldsymbol{A}_{2,2} \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} \boldsymbol{B}_1 & \boldsymbol{B}_2 \end{pmatrix}. \tag{11}$$

Relate the Schur complement $A_{2,2} - A_{2,1}(A_{1,1})^{-1}A_{1,2}$ to the Gram-matrix of the columns of B_2 , projected on the orthogonal complement of B_1 . Use this result to refine your comment in (d) on the relationship of Cholesky and QR factorization.

Copilot: "The Schur complement is the Gram matrix of the columns of B_2 projected on the orthogonal complement of B_1 ."

02-21-2023 1:14:00: The QR factorization is a way of orthogonalizing a given set of vectors, and the gram matrix is a way of measuring the orthogonality of a set of vectors only keeping track of the inner products instead of the vectors themselves.

4 QRs [25 pts]

(a) [5 pts]

Go to section (a) of the file HW4_your_code.jl and implement a function that uses the classical Gram-Schmidt algorithm for computing a reduced QR factorization.

(b) [5 pts]

Go to section (b) of the file HW4_your_code.jl and implement a function that uses the modified Gram-Schmidt algorithm for computing a reduced QR factorization.

(c) [5 pts]

Go to section (c) of the file HW4_your_code.jl and implement a function that computes the QR factorization using Householder reflections. Your algorithm should operate in place, overwriting the input matrix and not allocating additional memory.

(d) [5 pts]

Go to section (d) of the file $HW4_your_code.jl$ and implement functions that use the QR factorization matrix computed in (c) to multiply a vector with the QR (multiplication) or RQ^* (solving overdetermined least squares problem). Your functions should take a preallocated output vector as input in which to store the result. No additional allocation should be performed.

(e) [5 pts]

Using what you learned in class, design an experiment that compares the numerical stability of the different methods. For instance, compute the accuracy of the different approaches over matrices of increasing size or condition number and provide a plot of your results. Using what you learned in class to make sure that you include examples highlighting the lack of stability of classical Gram-Schmidt.