Marcus Rose
Math 126 HW 8
Due 04/13/23

---

1. For question 1:

   (a) Read 16.4.2 (just pages 164, 165), then run the code in pink to generate A, except that everywhere you see "150" change it to 200 (so the matrix is square). Record $\text{cond}_2(A)$. Explain why the condition number is as large as it is, in your own words.

   **Answer** After changing the condition number a few times, the values that I got were very large: $7.9523 \times 10^9, 2.4776 \times 10^7, 1.3072 \times 10^{11}, 1.6179 \times 10^9$–to list a few. The conditions numbers are large because the columns of A span nearly independent columns. As such, small changes in inputs will cause substantial changes to the output.

   Note that the condition number is calculated via the `cond(A, 2)` command in MATLAB.

   (b) Now run the pink code in the middle of the page (except, let $n = 200$). Record $\|A-QR\|_2$ and $\|Q^TQ - I\|_2$. Explain the result, in your own words.

   **Answer** Here are some values from the code:

| $\|A - QR\|$ | $\|Q^TQ - I\|$ |
|---|---|
| 2.4219e-14 | 1.4009e-07 |
| 2.2507e-14 | 1.7306e-07 |
| 2.4533e-14 | 1.3039e-07 |
| 2.2357e-14 | 5.5728e-09 |
| 2.4218e-14 | 1.9119e-06 |

   Note that $\|A - QR\|$ produces very small values which is strong evidence that the $QR$ decomposition for $A$ has little error. Relatively, I'm getting a very large error for the $Q^TQ - I$ norm, which could be because $A$ spans a closely-dependent set of vectors.

   (c) Define $b = randn(m, 1)$. Use the Q and R from the result to compute the numerical solution $\hat{x}$ to the system $Ax = b$. Record $\|r\|_2/\|b\|_2$, where $r = bA\hat{x}$ is the residual. Do you expect the solution to be accurate? Why or why not.

   **Answer** Note that some of the following outputs are $2.3214\times10^{-6}, 2.9002\times10^{-6}, 3.8337\times 10^{-6}, 8.0883 \times 10^{-6}$. These are all very small, which again is evidence that the $QR$ decomp is very close to a correct factorization of $A$. This ratio that we are recording measures the relative residual norm of the solution, i.e. a measure of error. Because it is small, this is evidence that the solution may be accurate.

```
%% Part (c)
b = randn(m,1);
x = (Q*R)\b;
r = b - A*x;
ratio = norm(r ,2)/norm(b,2)
```

Listing 1: Code for (c)

(d) Finally, for the same $A$, do $[Q, R] = qr(A)$, which uses a Householder QR. Compute $\|A - QR\|_2$ and $\|Q^TQ - I\|_2$ and record. Now use this $Q$ and $R$ to compute the solution $\hat{x}$, and compute the relative residual norm. Give a bound on the accuracy (relative error of the exact to computed solutions). Compare to the estimated solution accuracy above.

**Answer** Note the following table of values repeating the procedure:

| $\|A - QR\|$ | $\|Q^TQ - I\|$ |
|---|---|
| 1.2362e-13 | 2.4947e-15 |
| 1.2362e-13 | 2.4947e-15 |
| 1.2362e-13 | 2.4947e-15 |
| 1.2362e-13 | 2.4947e-15 |
| 1.2362e-13 | 2.4947e-15 |

This is by virtue of MATLAB finding the exact QR decomp using Householder reflectors. But examining the order of magnitudes only, we see that the norms of $Q^TQ - I$ are smaller, so this is closer to being accurate to true $A$. The norm of the residual produces values like $4.3343e - 07, 1.5860e - 06, 7.6856e - 07, 1.1897e - 06$. These values are again small. To find the error, we see a bound of $6.0582e - 8$.

Theorem 10.1 states that

$$\frac{\|\hat{x} - x\|}{\|x\|} \leq \text{cond}_2(A)\frac{\|r\|}{\|b\|},$$

however, there is some variation in the upper bound of $A$. That is, the accuracy is directly related how well-conditioned $A$ is.

```
%% Part (d)
[Q,R] = qr(A);
b = randn(m,1);
x = (Q*R)\b;
r = b - A*x;
output = norm(r ,2)/norm(b,2);
```

Listing 2: Code for (d)

2. For $A \in \mathbb{R}^{m \times n}$, with $m > n$, and rank $= n$, it is possible to associate to $A$ a finite 2-norm condition number. First, argue that we know $A^TA$ is a square, invertible matrix. We will define $\text{cond}_2(A) = \text{cond}_2(A^TA)^{\frac{1}{2}}$ (we will cover why this generalization is sensible). Now let $A = QR$ be the (full) QR decomposition of $A$, and $R_1$ be the $n \times n$ first block of $n$ rows, all columns, out of $R$.

(a) Show $\text{cond}_2(A^TA) = \text{cond}_2(R_1^TR)$.

**Proof** Note that

$$A = QR \implies A^TA = (QR)^TQR = R^TQ^TQR = R^TR.$$

Hence, $\text{cond}_2(A^TA) = \text{cond}_2(R^TR)$.

□

(b) $\text{cond}_2(R_1^TR) = \text{cond}_2(R_1)^2$.

**Proof** Note $\text{cond}_2(A^TA) = \text{cond}_2(R_1^TR_1)^{\frac{1}{2}}$. Also note $R^TR = R_1^TR_1^T$. Hence, $\text{cond}_2(R_1^TR) = \text{cond}_2(R_1)^{\frac{1}{2}} \implies \text{cond}_2(R_1^TR) = \text{cond}_2(R_1)^2$.

$\square$

(c) $\text{cond}_2(A) = \text{cond}_2(R_1)$.

**Proof** $\text{cond}_2(A)^2 = \text{cond}_2(A^TA) = \text{cond}_2(R^TR) = \text{cond}_2(R_1)^2 \implies \text{cond}_2(A) = \text{cond}_2(R_1)$.

$\square$

3. Let $A$ be an $m \times n$ matrix, $m \geq n$, with linearly independent columns. Recall the definition of an orthogonal projector (aka orthogonal projection matrix).

(a) Prove that the matrix $P = A(A^TA)^{-1}A^T$ is an orthogonal projector onto the column space of $A$. That is, show that $P^TP = P = P^2$ and that $\text{range}(P) = \text{col}(A)$.

**Proof** WTS $P^TP = P = P^2$. Given $P = A(A^TA)^{-1}A^T$,

$P^TP = (A(A^TA)^{-1}A^T)^T(A(A^TA)^{-1}A^T) = (A^T)^T((A^TA)^{-1})^T(A)^T = A((A^{-1}A^T)^{-1})^TA$
$= A^T(A^{-1}A^{T^{-1}})A = A(A^TA)^{-1}A^T = P$.

Then,

$P^2 = A(A^TA)^{-1}A^T(A(A^TA)^{-1}A^T) = A(A^TA)^{-1}(A^TA)(A^TA)^{-1}A^T = A(A^TA)^{-1}A^T = P$.

The range (column space) of $A$ is the same as the column space (range) of $P$, as

$\square$

(b) Show that the $P$ matrix is equal to $Q_1Q_1^T$ where $A = Q_1R_1$ is the economized $QR$ decomp of $A$.

**Proof** $P = A(A^TA)^{-1}A^T = Q_1R_1((Q_1R_1)^TQ_1R_1)^{-1}(Q_1R_1)^T = Q_1R_1(R_1^TQ_1^TQ_1R_1)^{-1}(R_1^TQ_1^T)$
$= Q_1R_1(R_1^TR_1)^{-1}R_1^TQ_1^T = Q_1R_1R_1^{-1}R_1^{T^{-1}}R_1^TQ_1^T = Q_1Q_1^T$.

$\square$

(c) Recall that the normal equations that characterize the minimizer of $\|Ax-b\|_2$ for $b \in \mathbb{R}^m$ are given by $A^TAx = A^Tb$. Use (a) to show that for the least squares solution $x$, $Ax = \mathbb{P}_{\text{Ran}(A)}b$ (the notation on the right hand side of this equation is shorthand for orthogonal projection onto the range, or column space, of $A$).

**Proof** Note that $x$ is the solutions to the normal equations $A^TAx = A^Tb$. Then,

$$Ax = (A^TA)^{-1}A^Tb = \mathbb{P}_{\text{Ran}(A)}b,$$

directly follows from $(a)$.

$\square$

4. Let $A$ be an $m \times m$ symmetric and positive definite matrix. Let $c$ be a positive real number, and consider the matrix $A + cI$.

(a) Show that $A + cI$ is still positive definite.

**Proof** Note that $A$ is positive definite. Note that the $\text{diag}(cI) = (c_1, c_2, ..., c_m)$. Hence, $A+cI$ adds $c_i$ to the $ith$ diagonal entry of $A$. Formally, for a vector $x$ in $\mathbb{R}^m$, if $x^T Ax > 0$, then $x^T(A + cI)x = x^T Ax + cx^T Ix$. Since $x^T Ix = \|x\|^2 > 0$ for any nonzero vector $x$, we see that

$$x^T(A + cI)x = x^T Ax + c\|x\|^2 > x^T Ax > 0.$$

$\square$

(b) Next, show that if $A$ is a symmetric positive definite matrix then

$$\text{cond}_2(A) = \frac{\lambda_{\max(A)}}{\lambda_{\min(A)}}.$$

**Proof** Since $A$ is positive definite, it has spectral decomposition

$$A = Q\Lambda Q^T,$$

where $Q$ is an orthogonal matrix and $\Lambda$ is a diagonal matrix containing the eigenvalues of $A$ along the diagonal. Because $A$ is positive definite, all of its eigenvalues are positive. Using the fact that Q is orthogonal,

$$\|A\| = \|Q\Lambda Q^T\| = \|\Lambda\| \text{ and } \|A^{-1}\| = \|(Q\Lambda Q^T)^{-1}\| = \|Q\Lambda^{-1}Q^T\| = \|\Lambda^{-1}\|,$$

with the last equality following from the fact that $Q$ is orthogonal (i.e. $Q^TQ = QQ^T = I$). Then,

$$\text{cond}_2(A) = \|A\|\|A^{-1}\| = \|\Lambda\|\|\Lambda^{-1}\|.$$

Because $\|\Lambda\|$ is the largest eigenvalue of $A$ and $\|\Lambda^{-1}\|$ is the reciprocal of the smallest eigenvalue of $A$,

$$\text{cond}(A) = \|\Lambda\|\|\Lambda^{-1}\| = \frac{\lambda_{\max}}{\lambda_{\min}}.$$

$\square$

(c) Using this formula, explain why $\text{cond}_2(A + cI) \leq \text{cond}_2(A)$.

**Answer** Note that the eigenvalues of $A + cI$ are $\lambda_i + c$, where $\lambda_i$ are the eigenvalues of $A$. Since $A$ is positive definite, all of its eigenvalues are positive. Hence,

$$\lambda_i + c > \lambda_i,$$

for all $i$. This means that the largest and smallest eigenvalues of $A + cI$ are $\lambda_{\max} + c$ and $\lambda_{\min} + c$. Hence,

$$\text{cond}_2(A + cI) = \frac{\lambda_{\max} + c}{\lambda_{\min} + c}.$$

Then because $\lambda_{\max} > \lambda_{\max} + c$ and $\lambda_{\min} < \lambda_{\min} + c$,

$$\frac{\lambda_{\max} + c}{\lambda_{\max}} < 1,$$

$$\frac{\lambda_{\min} + c}{\lambda_{\min}} > 1.$$

Hence, $\mathrm{cond}_2(A + cI) = \frac{\lambda_{\max}+c}{\lambda_{\min}+c} < \frac{\lambda_{\max}}{\lambda_{\min}} = \mathrm{cond}_2(A)$.

(d) Show that if $B$ is symmetric (not necessarily positive definite) then there exists $c$ large enough that $B + cI$ is positive definite. In terms of the eigenvalues of $A$, what is the smallest number $d$ such that $c > d$ makes $B + cI$ positive definite?

Because $B$ is symmetric, $B = B^T$ and $B$ has real eigenvalues $\lambda_i$. If the values $cI$ is added to each element along the diagonal of $B$, then the eigenvalues become $\lambda_i + c$ for each column (or equivalently row) of $B$. Hence, pick $c > \min(\lambda_i)$, all of the eigenvalues of $B + cI$ will be positive, and by Theorem 17.2 $B + cI$ must also be positive definite.

5. By hand, compute the Cholesky factor $R$ of

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 20 & 26 \\ 3 & 26 & 70 \end{bmatrix}.$$

The numbers are chosen so that things work out. Afterwords, check your result using MATLAB's chol function.

**Answer**

$$r_{ii} = \sqrt{a_i i - \sum_{k}^{i-1} r_{ki}^2} \text{ and } r_{ij} = \frac{1}{r_{ii}}(a_{ji} - \sum_{k=1}^{i-1} r_{kj} r_{ki}).$$

Note,

$r_{11} = \sqrt{a_{11}} = 1$,
$r_{12} = \frac{1}{r_{11}}(a_{21} - r_{k2}r_{k1} = 1(2) = 2$,
$r_{22} = \sqrt{a_{22} - r_{12}^2} = \sqrt{20 - 4} = \sqrt{16} = 4$,
$r_{23} = \frac{1}{r_{22}} \sum_k^1 r_{k3} r_{k2} = \frac{1}{4} 26 - 6 = 5$,
$r_{33} = \sqrt{a_{33} \sum_k^2 r_{k3}^2} = \sqrt{70 - 9 - 25} = \sqrt{36} = 6$.

Hence,

$$R = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}.$$

6. Two square matrices $A$ and $B$ are called similar if there exists an invertible matrix $S$ such that $SA = BS$. Show that if one of a pair of similar matrices is symmetric positive definite, then so too is the other.

**Proof** Let $\lambda$ be an eigenvalue of $A$. Then,

$$Ax = \lambda x.$$

Then,

$$Ax = S^{-1}BSx = \lambda x \implies BSx = S\lambda x = B(Sx) = \lambda(Sx).$$

$\lambda$ is also an eigenvalue of $B$. Hence, $A$ and $B$ have the same eigenvalues as similar matrices. Then, Theorem 17.2 states that a matrix is positive definite iff it has all positive eigenvalues. Because $A$ has all positive eigenvalues by virtue of its positive definite status, so too must $B$.

$\square$

7. Prove that the diagonal entries of a positive definite matrix must be strictly positive. Hint: If $A$ is positive definite then $x^T Ax > 0$ for $x \neq 0$. Pick $x$ cleverly.

**Proof** Assume by contradiction that $A$ is positive definite and has at least one non-positive diagonal entry $a_{ii} < 0$. With judicious choice of $x$, we can choose $x$ such that $x^T Ax \leq 0$ for a non-zero vector $x$. Consider

$$x = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix},$$

where the non zero entry is located in the *ith* entry of $x$. Then, $x^T A = \begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{im} \end{bmatrix}$ and

$$x^T Ax = \begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{im} \end{bmatrix} A \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix},$$

and $x^T Ax = -a_{ii}$. Thus, $x^T Ax \leq 0$, where there exists a negative entry along the diagonal of $A$, and therein lies our contradiction, as $A$ cannot be positive definite.

$\square$