STAT 30900 : MATHEMATICAL COMPUTATIONS I

Fall 2019

•

Homework 0

•

Solutions by

Jinhong Du

12243476

This homework mostly serves as a linear algebra refresher. We will recall some definitions. The null space or kernel of a matrix $A \in \mathbb{R}^{m\times n}$ is the set

$$\ker(A) = \{x \in \mathbb{R}^n : Ax = 0\}$$

while the range space or image is the set

$$\operatorname{im}(A) = \{y \in \mathbb{R}^m : y = Ax \text{ for some } x \in \mathbb{R}^n\}.$$

The rank and nullity of A are defined as the dimensions of these spaces, $\operatorname{rank}(A) = \dim \operatorname{im}(A)$ and $\operatorname{nullity}(A) = \dim \ker(A)$. By convention we write all vectors in $\mathbb{R}^n$ as column vectors.

1

(a) For $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$, show that

$$\operatorname{im}(AB) \subseteq \operatorname{im}(A) \text{ and } \ker(AB) \supseteq \ker(B).$$

When does equality occur in each of these inclusions?

Proof. (1) Since $\forall\ y \in \operatorname{im}(AB)$, $\exists\ x \in \mathbb{R}^p$, s.t. $y = ABx = A(Bx)$, and $Bx \in \mathbb{R}^n$, we have $y \in \operatorname{im}(A)$. Therefore, $\operatorname{im}(AB) \subseteq \operatorname{im}(A)$.
Let $A = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}$ and $B = \begin{bmatrix} b_{ij} \end{bmatrix}$. Then the equality occurs when $\operatorname{span}\{\sum_{i=1}^n b_{ij}\alpha_i,\ j = 1,\dots,p\} = \operatorname{span}\{\alpha_j,\ j = 1,\dots,n\}$. Furthermore, a sufficient condition can be $\operatorname{im}(B) \supseteq \ker(A)^\perp$, e.g., $\operatorname{rank}(B) = n$ with $n \geq p$.
(2) Since $\forall\ x \in \ker(B)$, $Bx = 0$, we have $(AB)x = A(Bx) = 0$, i.e. $x \in \ker(AB)$. Therefore, $\ker(AB) \supseteq \ker(B)$.
The equality occurs when $\operatorname{im}(B) \cap \ker(A) = \{0\}$. To see this,

$$\begin{aligned} \operatorname{im}(B) \cap \ker(A) = \{0\} \quad &\iff \quad AB\mathbf{x} = 0 \text{ iff } B\mathbf{x} = 0,\ \forall\ \mathbf{x} \in \mathbb{R}^p \\ &\iff \quad \ker(AB) = \ker(B). \end{aligned}$$

□

(b) For $A, B \in \mathbb{R}^{n\times n}$, show that

$$\begin{aligned} \operatorname{rank}(AB) &\leq \min\{\operatorname{rank}(A), \operatorname{rank}(B)\}, \\ \operatorname{nullity}(AB) &\leq \operatorname{nullity}(A) + \operatorname{nullity}(B), \\ \operatorname{rank}(A+B) &\leq \operatorname{rank}(A) + \operatorname{rank}(B). \end{aligned}$$

Proof. For $A, B \in \mathbb{R}^{n\times n}$, in (a), we have show that $\operatorname{im}(AB) \subseteq \operatorname{im}(A)$, so $\operatorname{rank}(AB) \leq \operatorname{rank}(A)$. Analogously, $\operatorname{im}(B^\top A^\top) \subseteq \operatorname{im}(B^\top)$, so $\operatorname{rank}(B^\top A^\top) \leq \operatorname{rank}(B^\top)$. Notice that $\operatorname{rank}(AB) = \operatorname{rank}(B^\top A^\top)$ and $\operatorname{rank}(B) = \operatorname{rank}(B^\top)$, we have

$$\operatorname{rank}(AB) \leq \min\{\operatorname{rank}(A), \operatorname{rank}(B)\}.$$

Solution (cont.)

By rank-nullity theorem,

$$\begin{aligned}\operatorname{rank}(A) + \operatorname{nullity}(A) &= n,\\ \operatorname{rank}(B) + \operatorname{nullity}(B) &= n,\\ \operatorname{rank}(AB) + \operatorname{nullity}(AB) &= n,\end{aligned}$$

then

$$\begin{aligned}\operatorname{nullity}(AB) &= n - \operatorname{rank}(AB)\\ &\le \max\{n - \operatorname{rank}(A), n - \operatorname{rank}(B)\}\\ &= \max\{\operatorname{nullity}(A), \operatorname{nullity}(B)\}\\ &\le \operatorname{nullity}(A) + \operatorname{nullity}(B).\end{aligned}$$

Let $\{x_1, \dots, x_k\}$ and $y_1, \dots, y_l$ be the basis of $\operatorname{im}(A)$ and $\operatorname{im}(B)$, respectively, where $k = \operatorname{rank}(A)$, $l = \operatorname{rank}(B)$. Since $\forall\, z \in \operatorname{im}(A + B)$, $\exists\, \alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_l$, s.t. $z = \sum_{i=1}^{k} \alpha_i x_i + \sum_{j=1}^{l} \beta_j y_j$, we have $\operatorname{im}(A + B) = \operatorname{span}\{x_1, \dots, x_k, y_1, \dots, y_l\}$. Then $\operatorname{rank}(A + B) \le k + l = \operatorname{rank}(A) + \operatorname{rank}(B)$. $\square$

(c) For $A, B \in \mathbb{R}^{n\times n}$, show that if $AB = 0$, then $\operatorname{rank}(A) + \operatorname{rank}(B) \le n$.

Proof. Since $\forall\, x \in \mathbb{R}$, $ABx = 0$, we have $\operatorname{im}(B) \subseteq \ker(A)$. So $\operatorname{rank}(B) \le \operatorname{nullity}(A)$ and

$$\operatorname{rank}(A) + \operatorname{rank}(B) \le \operatorname{rank}(A) + \operatorname{nullity}(A) = n.$$

$\square$

2

(a) Let $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$. Show that

$$\operatorname{rank}\left(\begin{bmatrix} A & 0\\ 0 & B\end{bmatrix}\right) = \operatorname{rank}(A) + \operatorname{rank}(B).$$

We have used the block matrix notation here. For example if $A = \begin{bmatrix} a & b & c\\ d & e & f\end{bmatrix} \in \mathbb{R}^{2\times 3}$ and $B \begin{bmatrix}\alpha\\ \beta\end{bmatrix} \in \mathbb{R}^{2\times 1}$, then

$$\begin{bmatrix} A & 0\\ 0 & B\end{bmatrix} = \begin{bmatrix} a & b & c & 0\\ d & e & f & 0\\ 0 & 0 & 0 & \alpha\\ 0 & 0 & 0 & \beta\end{bmatrix} \in \mathbb{R}^{4\times 4}.$$

This is sometimes also denoted as $A \oplus B$. It is a direct sum of operators induced by a direct sum of vector spaces.

Proof. Let $\{a_1, \dots, a_k\}$ and $\{b_1, \dots, b_l\}$ be the basis of $\text{im}(A)$ and $\text{im}(B)$, respectively, where $k = \text{rank}(A)$, $l = \text{rank}(B)$, and let $C = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$. $\forall\, x \in \text{im}(C)$, x can be written as a linear combination of columns of C. Consider $x_{1:m}$, the first m rows of x, it can be written as a linear combination of columns of A, i.e., $\exists\, \alpha_1, \dots, \alpha_k$, s.t. $x_{1:m} = \sum_{i=1}^{k} \alpha_i x_i$. Similarly, $x_{(m+1):(m+p)} = \sum_{j=1}^{l} \beta_j y_j$. Let

$$z_i = \begin{cases} \begin{bmatrix} \alpha_i \\ 0_p \end{bmatrix} & , i = 1, \dots, k \\ \\ \begin{bmatrix} 0_m \\ \beta_i \end{bmatrix} & , i = k+1, \dots, k+l \end{cases}$$

where 0_j is the zero vector with length j. Then $x = \sum_{i=1}^{k} \alpha_i z_i + \sum_{j=1}^{l} \beta_j z_{j+k}$ and $\{z_i, i = 1, \dots, k+l\}$ are linearly independent, i.e., it is a basis of $\text{im}(C)$. Therefore,

$$\text{rank}(C) = k + l = \text{rank}(A) + \text{rank}(B).$$

$\square$

(b) For $\mathbf{x} = \begin{bmatrix} x_1, & \dots & x_n \end{bmatrix}^\top \in \mathbb{R}^m$ and $\mathbf{y} = \begin{bmatrix} y_1, & \dots & y_n \end{bmatrix}^\top \in \mathbb{R}^n$, observe that $\mathbf{x}\mathbf{y}^\top \in \mathbb{R}^{m \times n}$. Let $A \in \mathbb{R}^{m \times n}$. Show that $\text{rank}(A) = 1$ iff $A = \mathbf{x}\mathbf{y}^\top$ for some nonzero $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$.

Proof.

$$\begin{aligned} \text{rank}(A) = 1 \quad &\iff \quad \text{im}(A) = \text{span}\{\mathbf{x}\} \text{ for some nonzero } x \in \mathbb{R}^m \\ &\iff \quad \text{each column of } A \neq 0 \text{ can be expressed as a linear combination of } \{\mathbf{x}\}, \\ &\qquad\ \text{i.e., } \exists\, \mathbf{y} = \begin{bmatrix} y_1, \dots, y_n \end{bmatrix} \in \mathbb{R}^n,\ \mathbf{y} \neq \mathbf{0} \text{ s.t.} A = \begin{bmatrix} y_1\mathbf{x} & \dots & y_n\mathbf{x} \end{bmatrix} = \mathbf{x}\mathbf{y}^\top \\ &\qquad\ \text{for some nonzero } \mathbf{x} \in \mathbb{R}^m \text{ and } \mathbf{y} \in \mathbb{R}^n \end{aligned}$$

$\square$

3

Let $A \in \mathbb{R}^{m \times n}$,

(a) Show that

$$\ker(A^\top A) = \ker(A) \qquad \text{and} \qquad \text{im}(A^\top A) = \text{im}(A^\top).$$

Give an example to show this is not true over a finite field (e.g. a field of two elements $\mathbb{F}_2 = \{0, 1\}$ with binary arithmetic).

Since $\forall\, x \in \ker(A^\top A)$, $A^\top A x = 0$, so $x^\top A^\top A x = \|Ax\|_2^2 = 0$, which implies $Ax = 0$. Therefore, $\ker(A^\top A) \subseteq \ker(A)$. From problem 1 (a) we also have $\ker(A^\top A) \supseteq \ker(A)$. Therefore, $\ker(A^\top A) = \ker(A)$ and $\text{nullity}(A^\top A) = \text{nullity}(A)$

Solution (cont.)

Since

$$\operatorname{rank}(A^\top A) = n - \operatorname{nullity}(A^\top A) = n - \operatorname{nullity}(A) = \operatorname{rank}(A) = \operatorname{rank}(A^\top)$$

and $\operatorname{im}(A^\top A) \subseteq \operatorname{im}(A^\top)$ from problem 1 (a), we have $\operatorname{im}(A^\top A) = \operatorname{im}(A^\top)$.

Let $A \in \mathbb{F}_2^{2\times 2}$ and $A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$, then $A^\top = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, $A^\top A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. So

$$\ker(A^\top A) = \mathbb{F}_2^2 \neq \ker(A) = \left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}$$

$$\operatorname{im}(A^\top A) = \{\mathbf{0}_2\} \neq \operatorname{im}(A^\top) = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$$

(b) Show that

$$A^\top A x = A^\top b$$

always has a solution (even if $Ax = b$ has no solution). Give an example to show that this is not true over a finite field.

Proof. Since $\operatorname{im}(A^\top A) = \operatorname{im}(A^\top)$, for $A^\top b \in \operatorname{im}(A^\top b) = \operatorname{im}(A^\top A)$, $\exists\ x \in \mathbb{R}^n$, s.t. $A^\top A x = A^\top b$, i.e., $A^\top A x = A^\top b$ always has a solution.

Let $A \in \mathbb{F}_2^{2\times 2}$, $A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$, and $b = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, then $A^\top = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, $A^\top A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. Since $\operatorname{im}(A^\top A) = \{\mathbf{0}_2\}$ and $A^\top b = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, there is no solution for $A^\top A x = A^\top b$ in field $\mathbb{F}_2$. □

4

Let $\mathbf{v}_1, \ldots, \mathbf{v}_r \in \mathbb{R}^n$. Let $G_r = [g_{ij}] \in \mathbb{R}^{r\times r}$ be the matrix with

$$g_{ij} = \mathbf{v}_i^\top \mathbf{v}_j$$

for $i, j = 1, \ldots, r$. This is called a *Gram matrix*.

(a) Show that $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are linearly independent iff $\operatorname{nullity}(G_r) = 0$.

Proof. Let $V = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_r \end{bmatrix}$, then

$$G_r = \begin{bmatrix} \mathbf{v}_1^\top \\ \vdots \\ \mathbf{v}_r^\top \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_r \end{bmatrix} = V^\top V.$$

Solution (cont.)

From problem 3 (a), we have $\ker(G_r) = \ker(V^\top V) = \ker(V)$ and $\text{nullity}(G_r) = \text{nullity}(V)$. So

$$\begin{aligned} &\mathbf{v}_1, \dots, \mathbf{v}_r \text{ are linearly independent} \\ \iff &\text{rank}(V) = r \\ \iff &\text{nullity}(V) = 0 \\ \iff &\text{nullity}(G_r) = 0 \end{aligned}$$

□

(b) Show that $G_r = I_r$ iff $\mathbf{v}_1, \dots, \mathbf{v}_r$ are pairwise orthogonal unit vectors, i.e., $\|\mathbf{v}_i\|_2 = 1$ for all $i = 1, \dots, r$, and $\mathbf{v}_i^\top \mathbf{v}_j = 0$ for all $i \neq j$. If this holds, show that

$$\sum_{i=1}^{r} (\mathbf{v}^\top \mathbf{v}_i)^2 \leq \|\mathbf{v}\|_2^2 \tag{4.1}$$

for all $\mathbf{v} \in \mathbb{R}^n$. What can you say about $\mathbf{v}_1, \dots, \mathbf{v}_r$ if equality always holds in (4.1) for all $\mathbf{v} \in \mathbb{R}^n$?

Proof.

$$\begin{aligned} &G_r = I_r \\ \iff &g_{ii} = 1,\ \forall\ i = 1, \dots, r \text{ and } g_{ij} = 0, \text{ for all } i \neq j \\ \iff &\mathbf{v}_i^\top \mathbf{v}_i = \|\mathbf{v}_i\|_2 = 1 \text{ for all } i = 1, \dots, r, \text{ and } \mathbf{v}_i^\top \mathbf{v}_j = 0 \text{ for all } i \neq j \\ \iff &\mathbf{v}_1, \dots, \mathbf{v}_r \text{ are pairwise orthogonal unit vectors} \end{aligned}$$

Since $\mathbf{v}_1, \dots, \mathbf{v}_r$ are pairwise orthogonal unit vectors, then $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ is a basis of $\text{im}(V)$ with $\text{rank}(V) = r$ where V is defined in (a). Since $\mathbb{R}^n = \text{im}(V) \oplus \text{im}(V)^\perp$, suppose $\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ are the basis of $\text{im}(V)^\perp$, then $\mathbf{v}_i (i = 1, \dots, r)$ and $\mathbf{v}_j (j = r+1, \dots, n)$ are orthogonal, and $\forall\ \mathbf{v} \in \mathbb{R}^n$, $\exists\ a_1, \dots, a_n \in \mathbb{R}$, s.t. $\mathbf{v} = \sum_{i=1}^{n} a_i \mathbf{v}_i$.

$$\begin{aligned} \sum_{i=1}^{r} (\mathbf{v}^\top \mathbf{v}_i)^2 &= \sum_{i=1}^{r} \left(\sum_{j=1}^{n} a_j \mathbf{v}_j^\top \mathbf{v}_i \right)^2 \\ &= \sum_{i=1}^{r} (a_i \mathbf{v}_i^\top \mathbf{v}_i)^2 \\ &= \sum_{i=1}^{r} a_i^2 \|\mathbf{v}_i\|_2^2 \\ &\leq \sum_{i=1}^{n} a_i^2 \|\mathbf{v}_i\|_2^2 \\ &= \|\mathbf{v}\|_2^2 \end{aligned}$$

If equality always holds in (4.1) for all $\mathbf{v} \in \mathbb{R}^n$, then $a_{r+1} = \cdots = a_n = 0$ for all $\mathbf{v} \in \mathbb{R}^n$, which means that $\text{im}(V)^\perp = \{0\}$, i.e., $V = \mathbb{R}^n$, i.e. $r = n$. □

Let $A \in \mathbb{C}^{n\times n}$. Recall that A is diagonalizable iff there exists an invertible $X \in \mathbb{C}^{n\times n}$ such that $X^{-1}AX = \Lambda$, a diagonal matrix.

(a) Show that A is diagonalizable if and only if its minimal polynomial is of the form

$$m_A(x) = (x-\lambda_1)\cdots(x-\lambda_d)$$

where $\lambda_1, \dots, \lambda_d \in \mathbb{C}$ are all distinct. Hence deduce for a diagonalizable matrix, the degree of its minimal polynomial equals the number of distinct eigenvalues.

Proof. In $\mathbb{C}$, every polynomial of degree d have d roots. So it will have a form as m_A.

$\Longrightarrow$

Let $f_A \triangleq (A-\lambda_1 I)\cdots(A-\lambda_d I)$. Since A is diagonalizable, there exists $X \in \mathbb{C}^{n\times n}$ such that $X^{-1}AX = \Lambda$ is a diagonal matrix with eigenvalues lying on diagonal. Suppose the distinct eigenvalues are $\lambda_1, \dots, \lambda_d$. Since X is invertible, its columns $\mathbf{x}_1, \dots, \mathbf{x}_n$ form a basis of $\mathbb{C}^n$. Then $\forall\ \mathbf{v} \in \mathbb{C}^n$, $\exists\ a_1, \dots, a_n$, s.t. $\mathbf{v} = \sum_{i=1}^n a_i \mathbf{x}_i$.

Notice that for eigenvector $\mathbf{x}_i$ and corresponding eigenvalue λ_j, $A\mathbf{x}_i = \lambda_j \mathbf{x}_i$, so $(A-\lambda_j I)\mathbf{x}_i = \mathbf{0}$. Therefore, $f_A v = \mathbf{0}$, $\forall\ \mathbf{v} \in \mathbb{C}^n$, i.e., $m_A(A) = 0_{n\times n}$.

Consider a monomial, if $(A-\mu I)\mathbf{v} = 0$ for some $\mu \in \mathbb{C}$ and $\mathbf{v} \in \mathbb{C}^n$, then μ and $\mathbf{v}$ must be an eigenvalue and an eigenvector of A. So $m_A(x)$ is the minimal polynomial.

$\Longleftarrow$

Since $m_A(x)$ is the minimal polynomial, A has distinct eigenvalues $\lambda_1, \dots, \lambda_d$ (each of them may have multiplicity larger than 1).

From the kernel decomposition theorem,

$$\begin{aligned}\mathbb{C}^n &= \ker(0_{n\times n}) \\ &= \ker(m_A(A)) \\ &= \ker(A-\lambda_1 I) \oplus \ker(A-\lambda_2 I) \oplus \cdots \oplus \ker(A-\lambda_d I).\end{aligned}$$

$\forall\ i \in \{1, \dots, d\}$, let $n_i \triangleq \text{nullity}(A-\lambda_i I)$, then $\exists\ \{\mathbf{x}_{i,1}, \dots, \mathbf{x}_{i,n_i}\}$ the basis of $\ker(A-\lambda_i I)$, such that $\sum_{i=1}^d n_i = n$ and full rank matrix $X = \left[\mathbf{x}_{1,1}, \dots, \mathbf{x}_{1,n_i}, \dots, \mathbf{x}_{d,n_d}\right] \in \mathbb{C}^{n\times n}$ satisfies $XAX^{-1} = \Lambda$ where Λ is a diagonal matrix with each diagonal element being the eigenvalue λ_i corresponding to $\mathbf{x}_{i,n_j}$, i.e., A is diagonalizable. $\square$

(b) Let A be diagonalizable. Let $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{C}^n$ be n linearly independent right eigenvectors, i.e., $A\mathbf{x}_i = \lambda_i \mathbf{x}_i$; and $\mathbf{y}_1, \dots, \mathbf{y}_n \in \mathbb{C}^n$ be n linearly independent left eigenvectors, i.e., $\mathbf{y}_i^\top A = \lambda_i \mathbf{y}_i^\top$. Show that there is a choice of left and right eigenvectors of A such that any vector $\mathbf{v} \in \mathbb{C}^n$ can be expressed as

$$\mathbf{v} = \sum_{i=1}^n (\mathbf{y}_i^\top \mathbf{v})\mathbf{x}_i.$$

If we write $X = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{C}^{n\times n}$ and $Y = [\mathbf{y}_1, \dots, \mathbf{y}_n] \in \mathbb{C}^{n\times n}$. What is the relation between X and Y?

Let

$$X = \begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_n\end{bmatrix} \qquad Y = \begin{bmatrix}\mathbf{y}_1 & \cdots & \mathbf{y}_n\end{bmatrix}$$

Solution (cont.)

then

$$
\begin{aligned}
A &= X\Lambda X^{-1} \\
&= (Y^\top)^{-1}\Lambda Y^\top
\end{aligned}
$$

If we choose $Y^\top = X^{-1}$, i.e. $XY^\top = I$, then $\sum_{i=1}^{n} \mathbf{x}_i \mathbf{y}^\top = I$ and

$$
\begin{aligned}
&\sum_{i=1}^{n} (\mathbf{y}_i^\top \mathbf{v})\mathbf{x}_i \\
=&\sum_{i=1}^{n} \mathbf{x}_i \mathbf{y}_i^\top \mathbf{v} \\
=&\left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{y}_i^\top\right) \mathbf{v} \\
=&\mathbf{v}
\end{aligned}
$$