STAT 430 - Notes
Matrix Analysis

May 27, 2024

Contents

1 Matrix

An $m \times n$ matrix A can be written as

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

2 Notation

- a_{ij} is the (i, j) component of A, so a_{ij} is in the i^{th} row and j^{th} column.
- each horizontal line of A is a **row** of A
- each vertical line of A is a **column** of A
- $m \times n$ is the **size** of A, i.e., A has m rows and n columns.

A shorthand notation for A is $A = [a_{ij}]$.

3 Other formulations of a matrix

1. **Column form** We can write A as $A = [A_{\bullet 1} \quad A_{\bullet 2} \quad \cdots \quad A_{\bullet n}]$ where $A_{\bullet k}$ represents the k^{th} column of A, i.e.,

$$A_{\bullet k} = \begin{bmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{bmatrix}$$

Remark. If A is an $m \times 1$ matrix then A is called an m-column vector or simply an m-vector.

2. **Row form** We can write A as

$$A = \begin{bmatrix} A_{1\bullet} \\ A_{2\bullet} \\ \vdots \\ A_{m\bullet} \end{bmatrix}$$

where $A_{\ell\bullet}$ represents the ℓ^{th} row of A, i.e., $A_{\ell\bullet} = [a_{\ell 1} \quad a_{\ell 2} \quad \cdots \quad a_{\ell n}]$.

4 Examples

1. Let A be an $m \times n$ matrix. If $m = n$, then A is a **square matrix**.

2. Given $A = [a_{ij}]$ such that $a_{ij} = 0$, $\forall i, j$. Then A is called the **zero matrix**, i.e., $A = 0$.

5 Equivalence of matrices

Given two matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ we say that $A = B$ if

1. A and B are the same size,

2. $a_{ij} = b_{ij}$ for all i, j.

6 Basic Matrix Operations

1. **Addition**: Let A and B be $m \times n$ matrices. Then

$$A + B = [a_{ij} + b_{ij}].$$

2. **Scalar multiplication**: Let c be a scalar, i.e., c is either a real or complex number. Given $A = [a_{ij}]$ then

$$cA = [c \cdot a_{ij}].$$

6.0.1 Examples

Let $A = \begin{bmatrix} 2 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}_{3\times 2}$ and $B = \begin{bmatrix} -1 & 1 \\ 0 & 0 \\ 0 & -1 \end{bmatrix}_{3\times 2}$.

Find $2A + B$.

Solution: First we find $2A$ as

$$2A = 2\begin{bmatrix} 2 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 0 & 2 \\ 2 & 0 \end{bmatrix}.$$

Then we find $2A + B$ as

$$2A + B = \begin{bmatrix} 4 & 2 \\ 0 & 2 \\ 2 & 0 \end{bmatrix} + \begin{bmatrix} -1 & 1 \\ 0 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 0 & 2 \\ 2 & -1 \end{bmatrix}.$$

Definition. Let A be an $m \times n$ matrix. Then the **additive inverse** of A is defined as

$$-A = (-1)A = [-a_{ij}].$$

Definition. Let A and B be $m \times n$ matrices. Then the **difference** of A and B is defined as

$$\begin{aligned} A - B &= A + (-B) \\ &= [a_{ij}] + [-b_{ij}] \\ &= [a_{ij} - b_{ij}]. \end{aligned}$$

3. **Transpose** Let A be an $m \times n$ matrix and $A = [a_{ij}]$. Let A^T be the transpose of A where

$$\left[A^T\right]_{ij} = [A]_{ji} = a_{ij}.$$

Then A^T is an $m \times n$ matrix, where the k^{th} row is the k^{th} column of A.

Remark. For a square matrix, the transpose operation does not change the size of the matrix.

4. **Conjugate and conjugate transpose**

Let $z = a + bi$ be a complex number where $a, b \in \mathbb{R}$, and $i = \sqrt{-1}$ is the imaginary unit. We call a the real part of z, i.e., $a = Re(z)$ and b is the imaginary part of z, i.e., $b = Im(z)$.

The **conjugate** of z is

$$\bar{z} = \overline{a + bi} = a - bi.$$

Particularly, if z is a real number, i.e., $z = a$, then $\bar{z} = z$.

Suppose $A = [a_{ij}]$ is a complex matrix, i.e., a_{ij} is a complex number. Therefore,

$$\bar{A} = [\bar{a_{ij}}].$$

The **conjugate transpose** of A is

$$A^\star = (\bar{A})^T.$$

Example. Let $A = \begin{bmatrix} 1 & 1+i \\ i & 3 \\ 0 & 2 \end{bmatrix}$.

Then to calculate the conjugate transpose we first calculate the conjugate of A as $\bar{A} = \begin{bmatrix} 1 & 1-i \\ -i & 3 \\ 0 & 2 \end{bmatrix}$. Then we calculate the conjugate transpose as

$(\bar{A})^T = \begin{bmatrix} 1 & -i & 0 \\ 1-i & 3 & 2 \end{bmatrix} = A^\star$.

Remark. The conjugate and transpose operations are commutative. Thus $A^\star = \overline{(A^T)}$.

7 Properties of matrix operations

I) **Addition**

(a) Commutative Property $A + B = B + A$

(b) Associative Property $(A + B) + C = A + (B + C)$.

(c) $A + 0 = A$

(d) Let $-A$ be the additive inverse of A. Then $A + (-A) = 0$.

II) **Scalar multiplication**

Let α and β be two scalars.

(a) $(\alpha \cdot \beta)A = \alpha(\beta A) = \beta(\alpha A)$

(b) $(\alpha + \beta)A = \alpha A + \beta A$.
To see this we have

$$\begin{aligned}(\alpha + \beta)A &= [(\alpha + \beta)a_{ij}] \\ &= [\alpha a_{ij} + \beta a_{ij}] \\ &= [\alpha a_{ij}] + [\beta a_{ij}] \\ &= \alpha A + \beta A.\end{aligned}$$

(c) $1 \cdot A = A$

III) Transpose and conjugate transpose

(a) $(A + B)^T = A^T + B^T$

(b) $(cA)^T = c \cdot A^T$

(c) $(A^T)^T = A$

(d) $(A+B)^* = A^* + B^*$

(e) $(c \cdot A)^* = \bar{c} \cdot A^*$ where $\bar{c}$ is the conjugate of complex number c

(f) $(A^*)^* = A$

Remark. Below we use the fact that given 2 complex numbers z and w, i.e. $z, w \in \mathbb{C}$, then

$$\overline{z \cdot w} = \bar{z} \cdot \bar{w}.$$

Lemma. *Given a complex number c and a complex matrix A we have*

$$(c \cdot A)^* = \bar{c} \cdot A^*.$$

Proof. Recall that $A^* = (\bar{A})^T$. Now using the remark above we can write

$$\begin{aligned} [\overline{c \cdot A}]_{ij} &= \overline{c \cdot a_{ij}} \\ &= \bar{c} \cdot \overline{a_{ij}} \\ &= \bar{c} \cdot [\overline{A}]_{ij}, \quad \forall i, j. \end{aligned}$$

Thus, $\overline{c \cdot A} = \bar{c} \cdot \bar{A}$. Therefore

$$\begin{aligned} (c \cdot A)^* &= (\overline{c\dot{A}})^T \\ &= (\bar{c} \cdot \overline{A})^T \\ &= \bar{c} \cdot (\overline{A})^T \\ &= \bar{c} \cdot A^*. \end{aligned}$$

□

Remark. Let $A : m \times n$ matrix. If A and A^T have the same size, then $m = n$, i.e., A is a square matrix.

Definition. Let A be a square matrix, i.e., $A : n \times n = [a_{ij}]$.

1. If $A = A^T$, then A is **symmetric**, i.e., $a_{ij} = a_{ji}, \quad \forall i, j$.

2. If $A = -A^T$, then A is **skew symmetric**, i.e., $a_{ij} = -a_{ji}, \quad \forall i, j$.

3. If $A = A^*$, then A is **Hermitian**, i.e., $a_{ij} = \overline{a_{ij}}, \quad \forall i, j$.

4. If $A = -A^*$, then A is **skew Hermitian**, i.e., $a_{ij} = -\overline{a_{ji}}, \quad \forall i, j$.

Using the above definitions and properties we can prove the following statements:

1. Let A be a skew symmetric matrix. Then $a_{ii} = 0$ for all i and j.

   Proof. Since A is skew symmetric and so $a_{ij} = -a_{ji}$ for all i and j. Letting $j = i$, we have $a_{ii} = -a_{ii}$ and after adding a_{ii} to both sides we have $2a_{ii} = 0$. Since $2 \neq 0$, then we must have $a_{ii} = 0$. □

2. If A is skew Hermitian the a_{ii} are pure imaginary numbers.

   Proof. Need to complete. □

3. Let A be a square matrix. Then $A + A^T$ is symmetric.

   Proof. Let $B = A + A^T$. We want to show $B = B^T$. So

$$\begin{aligned} B^T &= (A + A^T)^T \\ &= A^T + (A^T)^T \\ &= A^T + A \\ &= A + A^T \\ &= B. \end{aligned}$$

   We have shown $B^T = B$ and therefore B is symmetric. □

8 Diagonal entries and diagonal matrices

Definition. Let A be an $n \times n$ matrix and we can write A as

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}_{n \times n}.$$

The entries a_{ii} are called the **diagonal** entries of A and the other entries are **non-diagonal** entries.

A **diagonal** matrix A is a matrix whose non-diagonal entries are all zero.

An **upper triangular** matrix is a square matrix whose $a_{ij} = 0$, where $i > j$.

A **lower triangular** matrix is a square matrix whose $a_{ij} = 0$, where $i < j$.

The following statements follow from the definition of matrix operations and the above definitions:

1. If A is upper triangular then A^T is lower triangular.

2. If A is lower triangular, then A^T is upper triangular.

3. A is lower and upper triangular if and only if A is diagonal.

Example. Let the $n \times n$ matrix I be defined as

$$I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}_{n\times n}.$$

Then I is called the **Identity matrix**.

9 Matrix-vector product

Definition. Let A be an $m \times n$ matrix, i.e., $A = [A_{\bullet 1} \quad A_{\bullet 2} \quad \ldots A_{\bullet n}]$, which is the column form of A. Let $b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$ be an n-(column) vector. Then the **matrix-vector product** Ab is defined as

$$A \cdot b = (b_1 \cdot A_{\bullet 1}) + (b_2 \cdot A_{\bullet 2}) + \cdots + (b_n \cdot A_{\bullet n}).$$

Note that by the definition of scalar multiplication, $b_1 \cdot A_{\bullet k}$ for $k = 1, 2, \ldots, n$ is an m-vector (with m the number of rows in A). Then by the definition of vector addition, $A \cdot b$ is also an m-vector or an $m \times 1$ matrix.

9.0.1 Properties of matrix-vector products

Let $A : m \times n$ matrix and let u, v: n-vector and let c be a scalar. Then

1. $A(u + v) = Au + Av$

   Proof. Need to fill in. □

2. $A(c \cdot u) = c \cdot (A \cdot u)$

   Proof. Need to fill in. □

Remark. Recall that the identity matrix I can be written as

$$I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}_{n\times n}.$$

We can also write I as $I = [e_1 e_2 \cdots e_n]$ where e_k is the k$^{\text{th}}$ column of I.

Example. Let $A : n \times n$ matrix. Then

$$A \cdot e_1 = A_{\bullet 1} \cdot 1 + A_{\bullet 2} \cdot 0 + \cdots + A_{\bullet n} \cdot 0 = A_{\bullet 1}$$

where $e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$. Likewise, $A \cdot e_2 = A_{\bullet 2}, \ldots A \cdot e_n = A_{\bullet n}$.

10 Matrix multiplication

Definition. Two matrices $A_{m\times n}$ and $B_{r\times p}$ are **conformable** if $n = r$, i.e., $B_{n\times p}$.

It can be shown that if A and B are two square matrices of the same size then they are conformable.

10.0.1 Matrix product

Given $A_{m\times n}$ and $B_{n\times p}$ where $B = [B_{\bullet 1} B_{\bullet 2} \cdots B_{\bullet p}]$ and each $B_{\bullet k}$ is an n-vector, then

$$A \cdot B = [A \cdot B_{\bullet 1} \quad A \cdot B_{\bullet 2} \quad \cdots \quad A \cdot B_{\bullet p}].$$

By the definition of a matrix-vector product, $A \cdot B_{\bullet k}$ for $k = 1, 2, \ldots, p$ is an m-vector, or an $m \times 1$ matrix. Therefore, $A \cdot B$ has m rows and p columns and so $A \cdot B_{m\times p}$ matrix.

Another definition of matrix product is:

$$[A \cdot B]_{ij} = A_{i\bullet} \cdot B_{\bullet j} = \sum_{k=1}^{n} a_{ik} \cdot b_{kj}.$$

10.0.2 Caution and examples

1. Even if $A \cdot B$ is well defined, $B \cdot A$ may not. For example, given $A : m \times n$ and $B : n \times p$, the product $A \cdot B$ is well defined, but if $m \neq p$ then $B \cdot A$ is not defined.

2. Let $A : m \times n$ and $B : n \times m$. In this case $A \cdot B$ and $B \cdot A$ are well defined. But, $A \cdot B : m \times m$ and $B \cdot A : n \times n$. If $m \neq n$, then $AB \neq BA$.

3. Let $A : n \times n$ and $B : n \times n$. In this case, $A \cdot B$ and $B \cdot A$ are both well defined and of the same size. However, $AB \neq BA$ in general. For example, let $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}_{2\times 2}$ and $B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}_{2\times 2}$. Then $AB = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ and $BA = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$. Therefore, $A \cdot B \neq B \cdot A$.

   Definition. Let $A, B : n \times n$ matrices. If $A \cdot B = B \cdot A$, then we say that A and B **commute**, otherwise they do not commute.

4. (No cancellation law in general)

   Consider $A = \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 3 \\ 1 & 3 \end{bmatrix}$. Both A and B are non-zero, but $AB = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

   Let C be the 2×2 zero matrix. Then $A \cdot C = 0$ and therefore $A \cdot B = A \cdot C$ but $B \neq C$.

11 Product of matrix and identity matrix

Recall that the $n \times n$ identity matrix

$$I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}_{n\times n} = [e_1 \quad e_2 \quad \dots \quad e_n]$$

where e_i is the ith column of I_n.
Given an $m \times n$ matrix A,

$$A \cdot I_n = A \cdot [e_1 \quad e_2 \quad \dots \quad e_n] = [Ae_1 \quad Ae_2 \quad \dots \quad Ae_n].$$

Since $A \cdot e_i = A_{\bullet i}$, the ith column of A, then

$$A \cdot I_n = [A_{\bullet 1} \quad A_{\bullet 2} \quad \dots \quad A_{\bullet n}] = A.$$

Similarly, let I_n be the $m \times n$ identity matrix. Then $I_n \cdot A = A$.

Example. Suppose $A : m \times n$ matrix, such that $A \cdot A$ is well defined.
Question: What can be said about m and n?
Answer: Since $A \cdot A$ is well defined then A and A are conformable, then the number of columns in A must equal the number of rows in A. Thus, $m = n$.

Remark. For the next few sections let A be a $n \times n$ square matrix.

12 k^{th} power of a matrix

Since A is a square matrix then $A \cdot A$ is well defined and $A \cdot A = A^2$. Similarly,

$$\begin{aligned}
A^3 &= A \cdot A \cdot A \\
&\vdots \\
A^k &= A \cdot A \cdots A \qquad k \in \mathbb{N}.
\end{aligned}$$

By convention, $A^0 = I_n$.

13 Properties of matrix products

1. $A \cdot (B + C) = A \cdot B + A \cdot C$
2. $(D + E) \cdot F = D \cdot F + E \cdot F$
3. $(A + B) \cdot C = A \cdot (B \cdot C)$
4. Let α be a scalar, then $A(\alpha \cdot B) = \alpha \cdot (A \cdot B) = (\alpha \cdot A) \cdot B$
5. $(A \cdot B)^T = B^T \cdot A^T$, $(A \cdot B)^* = B^* \cdot A^*$

Proof. (Proof of 5) Consider the pure transpose case only. Let $A : m \times n$ and $B : n \times p$. First we show that $(A \cdot B)^T$ and $B^T \cdot A^T$ are the same size. Since $A \cdot B$ is $m \times p$ then by the definition of transpose, $(A \cdot B)^T$ is $p \times m$. Now, $B^T : p \times n$ and $A^T : n \times m$ so the number of columns in B^T is n and the number of columns in A^T is n, thus $B^T \cdot A^T$ is well defined and is a $p \times m$ matrix.

Next, we want to show that the corresponding entries of $(A \cdot B)^T$ and $B^T \cdot A^T$ are equal. Consider $i = 1, \ldots, p$ and $j = 1, \ldots, m$. Then,

$$\begin{aligned}
[(B^T \cdot A^T)]_{ij} &= (B^T)_{i\bullet} \cdot (A^T)_{\bullet j} \\
&= \sum_{k=1}^{n} [B^T]_{ik} \cdot [A^T]_{kj} \\
&= \sum_{k=1}^{n} [B]_{ki} \cdot [A]_{jk} \\
&= \sum_{k=1}^{n} [A]_{jk} \cdot [B]_{ki}.
\end{aligned}$$

Also,

$$
\begin{aligned}
[(A\cdot B)^T]_{ij} &= [A\cdot B]_{ji} \\
&= A_{j\bullet} \quad B_{\bullet i} \\
&= \sum_{k=1}^{n} [A]_{jk}\cdot [B]_{ki}.
\end{aligned}
$$

Therefore, $[(AB)^T]_{ij} = [(B^T\cdot A^T)]_{ij}$. Thus $(A\cdot B)^T = B^T \cdot A^T$.

$\square$

Example. Let $A : m\times n$ matrix. Show that $A\cdot A^T$ and $A^T\cdot A$ are well defined and symmetric.

Proof. Since $A : m\times n$ then $A^T : n\times m$. Therefore $A\cdot A^T$ is well defined and is $m\times m$. Also, A^T has m columns and A has m rows and so $A^T\cdot A$ is well defined and is an $n\times n$ matrix.

Let $B = A\cdot A^T$, then

$$
\begin{aligned}
B^T &= (A\cdot A^T)^T \\
&= (A^T)^T\cdot A^T \\
&= A\cdot A^T \\
&= B.
\end{aligned}
$$

Thus, $B^T = B$, so B is symmetric. Likewise we can show that $A^T\cdot A$ is symmetric.

$\square$

Definition. Let A be an $n\times n$ matrix. The trace of A is the summation of the diagonal entries, i.e.,

$$
trace(A) = \sum_{i=1}^{n} a_{ii},
$$

where $A = [a_{ij}]$.

Example. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix}_{2\times 2}$. Then $trace(A) = 1 + (-4) = -3$.

Example. $trace(I_n) = n$

Lemma. *Let $A : m\times n$ and $B : n\times m$ such that AB and BA are well defined. Note that AB and BA must be square matrices. Then $trace(A\cdot B) = trace(B\cdot A)$.*

Proof.

$$\begin{aligned}
trace(A \cdot B) &= \sum_{i=1}^{m} [A \cdot B]_{ii} \\
&= \sum_{i=1}^{m} (A_{i\bullet} B_{\bullet i}) \\
&= \sum_{i=1}^{m} \left(\sum_{k=1}^{n} [A]_{ik} [B]_{ki} \right) \\
&= \sum_{k=1}^{n} \left(\sum_{i=1}^{m} [A]_{ik} [B]_{ki} \right) \\
&= \sum_{k=1}^{n} \left(\sum_{i=1}^{m} [B]_{ki} [A]_{ik} \right) \qquad \text{(in parathenases is } B_{k\bullet} A_{\bullet k}) \\
&= \sum_{k=1}^{n} (B_{k\bullet} A_{\bullet k}) \\
&= \sum_{k=1}^{n} [BA]_{kk} \\
&= trace(BA).
\end{aligned}$$

□

Word of caution

- $trace(A \cdot B) \neq trace(A) \cdot trace(B)$. Counterexample: Let $A = B = I_n$. Then $trace(A) = trace(B) = n$, but $A \cdot B = I_n$ and so $trace(AB) = n$. Therefore, $trace(A \cdot B) = n$ and $trace(A) \cdot trace(B) = n^2$.

Cycling property of trace Let $A, B, C : n \times n$ matrices. It can be shown that $trace(A \cdot B \cdot C) = trace(C \cdot A \cdot B)$.

14 Partition of a matrix

Let $A : m \times n$ matrix.

$$A = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array} \right]$$

where A_{ij} is the (i, j)-block of A.

Example.

$$A = \left[\begin{array}{cc|c} 0 & 1 & 1 \\ 1 & 0 & 1 \\ \hline -2 & 3 & 0 \end{array} \right]$$

In the above matrix, $A_{11} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $A_{12} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $A_{21} = \begin{bmatrix} -2 & 3 \end{bmatrix}$, and $A_{22} = \begin{bmatrix} 0 \end{bmatrix}$.

15 Block matrix multiplication

Example. Let $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ and $B = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \end{bmatrix}$. Assume that (A_{ik}, B_{kj}) is conformable and so $A_{ik} \cdot B_{kj}$ is well defined and so $A \cdot B$ is well defined. Then

$$A \cdot B = \begin{bmatrix} (A_{11}B_{11} + A_{12}B_{21}) & (A_{11}B_{12} + A_{12}B_{22}) & (A_{11}B_{13} + A_{12}B_{23}) \\ (A_{21}B_{11} + A_{22}B_{21}) & (A_{21}B_{12} + A_{22}B_{22}) & (A_{21}B_{13} + A_{22}B_{23}) \end{bmatrix}.$$

Example. Let $A = \begin{bmatrix} C & I \\ I & 0 \end{bmatrix}$ and $B = \begin{bmatrix} I & 0 \\ D & E \end{bmatrix}$. Then

$$AB = \begin{bmatrix} (CI + ID) & (C0 + IE) \\ (II + 0D) & (I0 + 0E) \end{bmatrix} = \begin{bmatrix} C + D & E \\ I & 0 \end{bmatrix}.$$

Example. The product of upper block triangular matrices is also block upper triangular. Let $A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$ and $B = \begin{bmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{bmatrix}$. Then

$$AB = \begin{bmatrix} A_{11}B_{11} & (A_{11}B_{12} + A_{12}B_{22}) \\ 0 & A_{22}B_{22} \end{bmatrix}.$$

16 Linear functions

Let V and W be two sets (i.e., vector spaces) with the same field, like $\mathbb{R}$ or $\mathbb{C}$, and two basic operations on V and W:

1. vector addition:

$$u + w \in V \qquad \forall u, v \in V$$
$$x + y \in W \qquad \forall x, y \in W$$

2. scalar multiplication:

$$\alpha \cdot u \in V \qquad \forall u \in V, \alpha \in \mathbb{R} \text{ or } \mathbb{C}$$
$$\alpha \cdot x \in W \qquad \forall x \in W, \alpha \in \mathbb{R} \text{ or } \mathbb{C}$$

Example.

1. $V = \mathbb{R}^n$: n-dimensional Euclidean space and $W = \mathbb{R}^m$ is m-dimensional Euclidean space.
2. $V = \mathbb{C}^n$: n-dimensional complex space and $W = \mathbb{C}^m$ is m-dimensional complex space.
3. $V = \mathbb{R}^{m \times n}$: the space of all $m \times n$ real matrices.

Definition. A function $f : V \to W$ is **linear** if:

1. $f(u + v) = f(u) + f(v)$, $\forall u, v \in V$. This is sometimes called the superposition property.

2. $f(\alpha \cdot u) = \alpha \cdot f(u)$, $\forall u \in V$ and scalar $\alpha \in \mathbb{R}$ or $\mathbb{C}$. This is sometimes called the scalar property.

Example. Given an $m \times n$ real matrix A, define the function $f : \mathbb{R}^n \to \mathbb{R}^m$ as $f(x) = A \cdot x$ for $x \in \mathbb{R}^n$ where $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$.

Claim. $f : \mathbb{R}^n \to \mathbb{R}^m$ is linear.

Proof. By question 3 from homework 1, $f(u+v) = A(u+v) = Au+Av = f(u)+f(v)$ for all $u, v \in \mathbb{R}^n$. Also, for scalar α and $u \in \mathbb{R}^n$, $f(\alpha \cdot u) = A(\alpha \cdot u) = \alpha(A \cdot u) = \alpha f(u)$. $\square$

Example. Consider the trace function $trace : \mathbb{R}^{n\times n} \to \mathbb{R}$ where $\mathbb{R}^{n\times n}$ is the space of all $n \times n$ matrices.

Show: $trace(\cdot)$ is linear.

Proof. First we show $trace(\cdot)$ satisfies the superposition property. Given $A = [a_{ij}]$ and $B = [b_{ij}]$ in $\mathbb{R}^{n\times n}$ then

$$\begin{aligned} trace(A + B) &= \sum_{i=1}^{n} [A + B]_{ii} \\ &= \sum_{i=1}^{n} [A]_{ii} + [B]_{ii} \\ &= \sum_{i=1}^{n} [A]_{ii} + \sum_{i=1}^{n} [B]_{ii} \\ &= trace(A) + trace(B). \end{aligned}$$

Now we show $trace(\cdot)$ satisfies the scalar property. Let $\alpha \in \mathbb{R}$ and $A = [a_{ij}] \in \mathbb{R}^{n\times n}$. Then

$$\begin{aligned} trace(\alpha A) &= \sum_{i=1}^{n} [\alpha \cdot A]_{ii} \\ &= \sum_{i=1}^{n} \alpha [A]_{ii} \\ &= \alpha \sum_{i=1}^{n} [A]_{ii} \\ &= \alpha \cdot trace(A). \end{aligned}$$

Therefore, $trace(\cdot)$ is linear. $\square$

Proposition. *A function $f : V \to W$ is linear if and only if*

$$f(\alpha \cdot u + v) = \alpha \cdot f(u) + f(v), \quad \forall u, v \in V, \alpha \in \mathbb{R} \text{ or } \mathbb{C}.$$

Proof. ($\Rightarrow$) Suppose f is linear. Then for any $u, v \in V$ and scalar $.\alpha$,

$$f(\alpha \cdot u + v) = f(\alpha \cdot u) + f(v) = \alpha \cdot f(u) + f(v)$$

where the first equality follows from the superposition property and the second from the scalar property of linear functions.

($\Leftarrow$) For MATH 603 $\square$

17 Matrix Inverse

Definition. A $n \times n$ matrix A is invertible (i.e., non-singular) if (and only if) there exists an $n \times n$ matrix B such that

$$A \cdot B = I_n \text{ and } B \cdot A = I_n.$$

Here B is called an inverse of A, denoted A^{-1}.

Claim. If A is invertible then its inverse is unique.

Proof. Since A is invertible, there exists a matrix B such that $A \cdot B = B \cdot A = I_n$. Now suppose there exists a matrix B' such that $A \cdot B' = B' \cdot A = I_n$. Then since $A \cdot B'$,

$$\begin{aligned} B = B \cdot I_n &= B \cdot (A \cdot B') \\ &= (B \cdot A) \cdot A' \\ &= I_n B' \\ &= B'. \end{aligned}$$

Thus $B = B'$. Therefore, the inverse of A is unique. $\square$

Proposition. *Let A be an $n \times n$ matrix. The following are equivalent:*

(0) *A is invertible.*

(1) *A is row equivalent to I_n.*

(2) *The equation $Ax = 0$ has the solution $x = 0$ only.*

(3) *The equation $Ax = b$ has a solution for any n-vector b.*

(4) *$rank(A) = n$.*

(5) *There exists an $n \times n$ matrix B such that $B \cdot A = I_n$.*

(6) *There exists an* $n \times n$ *matrix* C *such that* $A \cdot C = I_n$.

Proof. In the following we assume as facts (2) and (3) to prove 5) $\Rightarrow$ 0) and 6) $\Rightarrow$ 0).

5) $\Rightarrow$ 0). By 5) there exists a matrix such that $B \cdot A = I_n$. It is sufficient to show that the equation $Ax = 0$ has the solution $x = 0$ only to prove that A is invertible.

Let u be a solution to $Ax = 0$, i.e., $Au = 0$. If we right multiply the latter equation by B we have $B(Au) = B \cdot 0 = 0$. Now consider

$$B(A \cdot u) = (B \cdot A)u = I_n \cdot u = u.$$

Thus $u = 0$. Therefore, 2) holds and since 2) implies 0) then 5) implies 0).

6) $\Rightarrow$ 0) By 6) there exists a matrix C such that $AC = I_n$.

Claim: For any n-vector b, $Ax = b$ has a solution.

Since there exists a matrix C such that $A \cdot C = I_n$ then given any n-vector b, $(A \cdot C) \cdot b = A \cdot (C \cdot b) = I_n \cdot n = b$. Thus $A \cdot (C \cdot b) = b$ implies that $x = C \cdot b$ is a solution to the equation $Ax = b$. Thus, 3) holds $\Rightarrow$ 0) holds.

□

18 Properties of an invertible matrix

Suppose A is invertible.

(1) A^{-1} is invertible, and $(A^{-1})^{-1}$.

Proof. $A \cdot A^{-1} = A^{-1} \cdot A = I_n$. Fill in details. □

(2) A^T and A^* are invertible, and $(A^T)^{-1} = (A^{-1})^T$ and $(A^*)^{-1} = (A^{-1})^*$

Proof. Since A is invertible then

$$\begin{aligned} A \cdot A^{-1} = A^{-1} \cdot A = I_n &\Rightarrow (A \cdot A^{-1})^T = (A^{-1} \cdot A)^T = (I_n)^T = I_n \\ &\Rightarrow (A^{-1})^T \cdot A^T = A^T \cdot (A^{-1})^T = I_n \end{aligned}$$

Therefore A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$. □

(3) Let B be another invertible matrix. Then $A \cdot B$ is invertible, and its inverse $(A \cdot B)^{-1} = B^{-1} \cdot A^{-1}$.

Proof.

$$\begin{aligned} A \cdot B \cdot (B^{-1} \cdot A^{-1}) &= A \cdot (B \cdot B^{-1}) \cdot A^{-1} \\ &= A \cdot A^{-1} \\ &= I_n. \end{aligned}$$

Therefore $(A \cdot B)^{-1} = B^{-1} \cdot A^{-1}$. □

Corollary. This is an extension of 3 above. Let $A_1, \ldots, A_k$ be invertible matrices. Then $A_1 \cdot A_2 \cdots A_k$ is invertible, and $(A_1 \cdot A_2 \cdots A_k)^{-1} = A_k^{-1} \cdots A_2^{-1} \cdot A_1^{-1}$. Can use proof by induction on k.

19 Examples

Example. Let A be a square matrix such that $I - A$ is invertible. Show:

$$A \cdot (I - A)^{-1} = (I - A)^{-1} \cdot A \quad \text{(commutative)} .$$

Proof. Compute

$$\begin{aligned} A \cdot (I - A) &= A \cdot I - AA = A - A^2 \\ &= I \cdot A - A \cdot A \\ &= (I - A) \cdot A. \end{aligned}$$

Then

$$\begin{aligned} A \cdot (I - A)(I - A)^{-1} = (I - A) \cdot A \cdot (I - A)^{-1} &\Rightarrow A = (I - A) \cdot A \cdot (I - A)^{-1} \\ &\Rightarrow (I - A)^{-1} \cdot = A \cdot (I - A)^{-1}. \end{aligned}$$

□

Example. We will use the following preliminary results in this example.

(1) Given $x, y \in \mathbb{R}^n$ (i.e., $x, y \in \mathbb{R}^{n \times 1}$ are n-column vectors). The following is a scalar:

$$x^T \cdot y = (x^T \cdot y)^T = y^T \cdot x \quad \text{(standard inner product)} .$$

(2) Given $x \in \mathbb{R}^n$ such that $x^T x = 0$, then $x = 0$.

Proof. Write this as $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$.

Then

$$
\begin{aligned}
x^T \cdot x &= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\
&= x_1^2 + x_2^2 + \cdots + x_n^2 \\
&= 0.
\end{aligned}
$$

Since $x^T x = 0$, then $x_1^2 + x_2^2 + \cdots + x_n^2 = 0$.
If $x_i \neq 0$ for some i then $x^T x > 0$. This shows that each $x_i = 0$ or $x = 0$. □

(3) Show $x^T S x = 0$, $\forall x \in \mathbb{R}^n$.

Proof. For any $x \in \mathbb{R}^n$

$$
\begin{aligned}
x^T S x &= x^T \cdot (Sx) \\
&= (S \cdot x)^T \cdot x \quad \text{(by preliminary result 1)} \\
&= (x^T \cdot S^T) \cdot x \\
&= (x^T \cdot -S) \cdot x \quad \text{(S is skew symmetric)} \\
&= x^T \cdot (-1) \cdot S \cdot x \\
&= (-1) \cdot x^T S \cdot x.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
x^T S x = -x^T S x &\Rightarrow 2 \cdot x^T S x = 0 \\
&\Rightarrow x^T S x = 0.
\end{aligned}
$$

□

Let S be a real skew symmetric matrix, i.e., $S = -S^T$ or $S^T = -S$.
Show that $\forall \alpha \in \mathbb{R}$, $I + \alpha \cdot S$ is invertible.

Proof. Let $\alpha \in \mathbb{R}$ be arbitrary. To show $I + \alpha \cdot S$ is invertible it suffices to show that

$$(I + \alpha S) \cdot x = 0$$

has the solution $x = 0$ only.
Suppose $\hat{x}$ is an arbitrary solution to $(I + \alpha S) \cdot x = 0$. Therefore,

$$
\begin{aligned}
(I + \alpha S) \cdot \hat{x} = 0 &\Rightarrow \hat{x}^T (I + \alpha S) \hat{x} = \hat{x}^T \cdot 0 \\
&\Rightarrow \hat{x}^T \hat{x} + \alpha \hat{x}^T S \hat{x} = 0 \\
&\Rightarrow \hat{x}^T \hat{x} = 0 \\
&\Rightarrow \text{ By 2)}, \hat{x} = 0.
\end{aligned}
$$

Therefore $(I + \alpha S) \cdot x = 0$ has the solution $x = 0$ only. Thus $I + \alpha \cdot S$ is invertible. □

20 Vector Space

Definition. A **vector space** is a set V, along with a field F (which is $\mathbb{R}$ or $\mathbb{C}$), and two basic operations

(1) vector addition "+" and

(2) scalar multiplication " $\cdot$ "

which satisfy:

A1): V is closed under vector addition, i.e., $x, y \in V$ then $x + y \in V$.

A2): For any $x, y \in V$, $(x + y) + z = x + (y + z)$ (associative)

A3): For any $, y \in V$, $x + y = y + x$ (commutative)

A4): There exists a zero vector in V, denoted by 0, such that $x+0 = x$ and $0+x = 0$, $\forall x \in V$.

A5): For any $x \in V$, there exists a vector in V, denoted by $-x$ such that $x+(-x) = 0$. Here $-x$ is additive inverse of x.

M1): V is closed under scalar multiplication, i.e., for any $x \in V$ and scalar α, $\alpha \cdot x \in V$.

M2): For any scalar α, β and $x \in V$,

$$(\alpha \cdot beta) \cdot x = \alpha(\beta \cdot x).$$

M3): For any scalar α, and $x, y \in V$,

$$\alpha \cdot (x + y) = \alpha \cdot x + \alpha \cdot y.$$

M4): For any scalars α, β and $x \in V$

$$(\alpha + \beta)x = \alpha \cdot x + \beta \cdot x.$$

M5): When $\alpha = 1$, $\alpha \cdot x = x$, $\forall x \in V$.

Any $x \in V$ is called a **vector** in V.

Example. Vector Spaces

(1) $\mathbb{R}^n$: the n-dimensional Euclidean space. Likewise, $\mathbb{C}^n$ is also a vector space.

(2) $\mathbb{R}^{m \times n}$: the set of all $m \times n$ real matrices. Likewise, $\mathbb{C}^{m \times n}$ is also a vector space.

- (2′) The set of all $n \times n$ upper triangular real matrices is a vector space, which is a subspace of $\mathbb{R}^{n \times n}$.

(3) The set of all real sequences

$$X + Y = (x_1 + y_1, x_2 + y_2, \dots),$$
$$\alpha \cdot X = (\alpha x_1, \alpha x_2, \dots).$$

This set is a vector space.

(3′) The set of **convergent** real sequences is a vector space and is a subspace of the above vector space.

(3″) The set of **bounded** real sequences is also a vector space.

(4) The set of all real-valued continuous functions on $\mathbb{R}$ (or $[a, b] \subseteq \mathbb{R}$).

(5) The set of all **divergent** real sequences is **NOT** a vector space.

This is because:

1) It has no "zero vector"
2) The closure under scalar multiplication fails. If we multiply any element by scalar 0, then the sequence will be convergent.
3) The closure under vector addition fails. Add $-x$ to x and will be zero vector, which is convergent.

21 Vector Spaces: Important Facts

1. V has the unique zero vector 0. (Related to A4)
2. For any $x \in V$, there exists the unique additive inverse $-x$, and $-x = (-1) \cdot x$.
3. For any $x \in V$, $0 \cdot x = 0 \in V$.
4. For any scalar α, $\alpha \cdot 0 = 0 \in V$.
5. A finite sum of vectors in V belongs to V. Can use induction on the number of vectors in sum to prove.

22 Subspace

Definition. Let $(V, \mathbb{F}, +, \cdot)$ be a vector space. If a subset S of V is a vector space (along with the same field $\mathbb{F}$ and basic operations of $+$ and $\cdot$ of V) then S is a **subspace** of V.

Example. Let $(V, \mathbb{F}, +, \cdot)$ be a vector space.

- V is a subspace of itself. So every vector space has at least one subspace.
- $S = \{0\}$, the singleton set of the zero vector in V is a subspace of V and is the zero or trivial subspace. Provide details of proof.

Remark. From the above example we can conclude that every vector space has at least 2 subspaces.

Proposition. *Let S be a subset of a vector space $(V, \mathbb{F}, +, \cdot)$. Then S is a subspace of V if and only if*

1. *S is closed under "$+$"; (A1) and*
2. *S is closed under "$\cdot$" (M1).*

Proof. ($\Leftarrow$) Suppose S is a subspace of V. Since S is a subspace of V, then S is a subspace and satisfies A1-A5 and M1-M5. Thus, A1 and M1 hold.

($\Rightarrow$) Suppose the subset S satisfies A1 and M1.

We want to show: S satisfies A2-A5 and M2-M5.

Since $S \subseteq V$ and V and V is a vector space then A.2, A.3, and M.2-M.5 hold on V and therefore hold on S as well.

Then it suffices to show that S satisfies A.4 and A.5.

A.4 Pick an arbitrary $x \in S$. Then $0 \cdot x = 0$ for the scalar 0 and any vector x, and since S is closed under "$\cdot$", then the zero vector is in S. Also, $0 + x = x + 0 = x$ for all $x \in S$.

A.5 For any $x \in S$, recall that its additive inverse $-x = (-1) \cdot x$ by Fact 2. Since S is closed under "$\cdot$", then $(-1) \cdot x \in S \implies -x \in S$. Thus, A.5 holds.

Therefore, we have shown that S is a vector space and so S is a subspace of V.

□

Example. Examples of sets that are subspaces.

(1) Let $V = \mathbb{R}^3$ and let $S = \left\{ \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix} : x_1 \in \mathbb{R} \right\}$. Show that S is a subspace of $\mathbb{R}^3$.

Proof. Let $x, y \in S$. Then

$$x = \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ 0 \\ 0 \end{bmatrix}.$$

The sum of x and y is

$$x + y = \begin{bmatrix} x_1 + y_1 \\ 0 \\ 0 \end{bmatrix}.$$

Therefore, S is closed under "+". Given any $\alpha \in \mathbb{R}$ and $x = \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix} \in S$,

$$\alpha \cdot x = \begin{bmatrix} \alpha \cdot x_1 \\ 0 \\ 0 \end{bmatrix} \in S.$$

Therefore, S is closed under scalar multiplication. Thus S is a subspace of V. □

(2) Let $V = \mathbb{R}^{n \times n}$.

(i) Let S_1 be the set of all $n \times n$ diagonal matrices in $\mathbb{R}^{n \times n}$. Since S is closed under standard matrix addition and scalar multiplication then S is a subspace of V.

(ii) Let S_2 be the set of all symmetric matrices in $\mathbb{R}^{n \times n}$.
Let $A, B \in S_2$, and therefore $A = A^T$ and $B = B^T$. Now consider the transpose of $A + B$,

$$(A + B)^T = A^T + B^T = A + B,$$

which shows that $(A + B)$ is symmetric and therefore $(A + B) \in S_2$.
Now let $\alpha \in \mathbb{R}$ and $A \in S_2$ and consider their product

$$(\alpha A)^T = \alpha A^T = \alpha T,$$

which shows that αA is symmetric and therefore in S_2. Since S_2 is closed under vector addition and scalar multiplication then S_2 is a subspace of V.

Example. Examples of sets that are **not** subspaces.

(1) Let $V = \mathbb{R}^2$ and $S = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} : x_1 \leq 0, x_2 \geq 0 \right\}$. Let

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \in S.$$

Therefore, $x_1 \leq 0$, $y_1 \leq 0$ and $x_2 \geq 0$, $y_2 \geq 0$. The sum of x and y is

$$x + y = \begin{bmatrix} x_1 + y_2 \\ x_2 + y_2 \end{bmatrix}$$

where $x_1 + y_1 \leq 0$ and $x_2 + y_2 \leq 0$. Thus $x + y \in S$ and so S is closed under vector addition.
Given an arbitrary $\alpha \in \mathbb{R}$ and $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in S$. For example, $x = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \in S$ and $\alpha = -1$. Then the product $\alpha \cdot x = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \notin S$ and therefore S is **not** closed under vector multiplication. Thus, S is not a subspace of V.

(2) Let $V = \mathbb{R}^{2\times 2}$ and S is the subset of all singular (i.e., not invertible) matrices in $\mathbb{R}^{2\times 2}$. Let $A \in S$ and $\alpha \in \mathbb{R}$. If $\alpha = 0$, then $\alpha \cdot A = 0$ is singular. If $\alpha \neq 0$, then $\alpha \cdot A$ is singular because otherwise $\alpha A \cdot (1/\alpha \cdot B) = I$ and so $A \cdot B = I$. This contradicts our assumption that A is not invertible. Therefore, αA is singular. Thus, S is closed under vector multiplication.

Claim: S is not closed under "+".

Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, and $B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. Both A and B are because if a matrix is in reduced row echelon form and any diagonal element is 0 then the matrix is singular. Thus $A, B \in S$. However, $A + B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, which is the identity matrix and thus is invertible or nonsingular. Therefore, $A + B \notin S$ and so S is not closed under "+".

23 Linear Combination

Definition. Let $\{v_1, v_2, \ldots, v_p\}$ be a finite set of vectors in a vector space V. Given scalars $\alpha_1, \alpha_2, \ldots, \alpha_p$. The **linear combination** of $v_1, v_2, \ldots, v_p$ using $\alpha_1, \alpha_2, \ldots, \alpha_P$ is

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_p v_p$$

where the α_i for $i = 1, 2, \ldots, p$ are called the weights or coefficients.

Remark.

- By definition, $\alpha \cdot x \in V$ so $\alpha_1 v_1 \in V, \ldots, \alpha_p v_p \in V$.
- Any finite sum of vectors from V is in V.

Definition. Given a finite set $S = \{v_1, \ldots, v_p\}$ in V, the set spanned by $v_1, \ldots, v_p$ (or S) is

$$\begin{aligned} span(S) &= \{\alpha_1 v_1 + \cdots + \alpha_2 v_2 + \cdots + \alpha_p v_p : \text{ scalars } \alpha_i, i = 1, \ldots, p\} \\ &= \text{ the set of all linear combinations of } v_1, \ldots, v_p. \end{aligned}$$

Clearly, $span(S) \subseteq V$, where S is a spanning set of $span(S)$.

Proposition. *The set $span(S)$ is a subspace of V.*

Proof. First we show that $span(S)$ is closed under "+". Let x, y be two vectors in $span(S)$. Therefore, there exists scalars $\alpha_1, \ldots, \alpha_p$ such that

$$x = \alpha_1 v_1 + \cdots + \alpha_p v_p.$$

Similarly, there exists scalars $\beta_1, \ldots, \beta_p$ such that

$$y = \beta_1 v_1 + \cdots + \beta_p v_p.$$

Then the sum of x and y is

$$\begin{aligned} x + y &= (\alpha_1 v_1 + \cdots + \alpha_p v_p) + (\beta_1 v_1 + \cdots + \beta_p v_p) \\ &= (\alpha_1 v_1 + \beta_1 v_1) + (\alpha_2 v_2 + \beta_2 v_2) + \cdots + (\alpha_p v_p + \beta_p v_p) \\ &= (\alpha_1 + \beta_1) v_1 + (\alpha_2 + \beta_2) v_2 + \cdots + (\alpha_p + \beta_p) v_p. \end{aligned}$$

Therefore, $x + y$ is a linear combination of $v_1, \ldots, v_p$. Thus $x + y \in span(S)$ and so $span(S)$ is closed under "+".

Now we show that $span(S)$ is closed under "$\cdot$". Let $x \in span(S)$ and γ be a scalar. Therefore the exists scalars $\alpha_1, \ldots, \alpha_p$ such that

$$x = \alpha_1 v_1 + \cdots + \alpha_p v_p.$$

Then the product of γ and x is

$$\begin{aligned} \gamma \cdot x &= \gamma \cdot (\alpha_1 v_1 + \cdots + \alpha_p v_p) \\ &= (\gamma \cdot \alpha_1) v_1 + \cdots + (\gamma \cdot \alpha_p) v_p. \end{aligned}$$

Therefore, $\gamma \cdot x$ is a linear combination of $v_1, \ldots, v_p$ and so $\gamma \cdot x \in span(S)$. Thus $span(S)$ is closed under "$\cdot$". Thus, $span(S)$ is a subspace. □

Here $span(S)$ is a <u>subspace spanned</u> by $v_1, \ldots, v_p$ (or S).

Proposition. *Let $S = \{v_1, \ldots, v_p\}$. Then $S \subseteq span(S)$.*

Proof. For each v_i,

$$v_i = 0 \cdot v_1 + \cdots + 0 \cdot v_{i-1} + \underline{1 \cdot v_i} + 0 \cdot v_{i+1} + \cdots + 0 \cdots v_p = 1 \cdot v_i.$$

Then $v_i \in span(S)$ and therefore $S \subseteq span(S)$. □

Remark. In the previous proof, S is generally not a subspace while $span(S)$ is a subspace.

Proposition. *$span(S)$ is the smallest subset containing S.*

Proof. Need to fill in □

Definition. Given a subspace W of V, if a finite set S spans W, i.e., $W = span(S)$, then W is **spanned by** S or S **spans** W.

Example. (1) Show that $\{e_1, e_2\}$ spans $\mathbb{R}^2$ or $\mathbb{R}^2 = span\{e_1, e_2\}$ where $e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

Proof. It suffices to show that any $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ is a linear combination of e_1 and e_2. Then

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ x_2 \end{bmatrix}.$$

Therefore, $\{e_1, e_2\}$ spans $\mathbb{R}^2$. □

(2) Show $\left\{ e_1, e_2, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ spans $\mathbb{R}^2$.

Proof. It suffices to show any $x \in \mathbb{R}^2$ is a linear combination of e_1, e_2 and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. For any $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$,

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + 0 \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

□

Therefore, $\{e_1, e_2\}$ spans $\mathbb{R}^2$.

(3) Does $\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ spans $\mathbb{R}^2$?

Answer: Yes

For any $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$, we need to show that there exists $\alpha_1, \alpha_2 \in \mathbb{R}$ such that

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \alpha_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ 0 \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} \alpha_1 + \alpha_2 \\ \alpha_2 \end{bmatrix}.$$

This leads to the linear equation with unknowns α_1 and α_2:

$$\begin{cases} \alpha_1 + \alpha_2 & = x_1 \\ \alpha_2 & = x_2 \end{cases}$$

which implies

$$\begin{cases} \alpha_2 & = x_2 \\ \alpha_1 & = x_1 - x_2 \end{cases}.$$

Therefore, $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ spans $\mathbb{R}^2$.

(4) Show that $\{E^{11}, E^{12}, E^{21}, E^{22}\}$ spans $\mathbb{R}^{2\times 2}$ where

$$E^{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, E^{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, E^{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \text{ and } E^{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

For any matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in \mathbb{R}^{2\times 2}$,

$$A = a_{11} \cdot E^{11} + a_{12} \cdot E^{12} + a_{21} \cdot E^{21} + a_{22} \cdot E^{22}.$$

Let W be the subspace of all 2×2 symmetric matrices in $\mathbb{R}^{2\times 2}$. Show

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\}$$

spans W.
All matrices in W is symmetric.

Proof. For any $A \in W$, $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ such that $a_{12} = a_{21}$. Then

$$\begin{aligned} \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} &= \begin{bmatrix} a_{11} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & a_{12} \\ a_{12} & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & a_{22} \end{bmatrix} \\ &= a_{11} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + a_{12} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + a_{22} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

Therefore, A is a linear combination of $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. Thus, W is spanned by $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ or $W = span\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$. □

24 Set operations on subspaces

Let U and W be two subspaces of V.

(1) $U \cap W$ is a subspace of V.

Proof. Since U and W are subspaces of V, then $0 \in U$ and $0 \in W$ and therefore $U \cap W$ is nonempty. First we show that $U \cap W$ is closed under "+". For any $x, y \in U \cap W$, $x, y \in U$ and $x, y \in W$. Then $x + y \in U$ and $x + y \in W$ by the closure property of the subspaces. Therefore, $U \cap W$ is closed under "+".

Similarly, $U \cap W$ is closed under "$\cdot$" and therefore $U \cap W$ is a subspace. □

(2) The union U and W is **not** a subspace in general.

Consider the example where $U = span\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\} \subseteq \mathbb{R}^2$, and $W = span\left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\} \subseteq \mathbb{R}^2$. Therefore, $U \cup W = span(e_1) \cup span(e_2)$.

Claim: $U \cup W$ is <u>not</u> closed under "+". This is because $e_1 \in U$, $e_2 \in W$ and therefore $e_1 + e_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \notin U \cup W$.

(3) The (algebraic) sum of U and W,

$$U + W = \{u + w : u \in U \text{ and } w \in W\}.$$

Example: When $U = span \left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ and $W = span \left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$. Then $U + W \in \mathbb{R}^2$.

Proposition. *Let U and W be two subspaces of V. Then the (algebraic) sum of $U + W$ is a subspace, where $U + W = \{u + w : u \in U, w \in W\}$.*

Proof. First we show that $U + W$ is closed under "+". Let $x, y \in U + W$. Then there exists $u \in U$ and $w \in W$ such that $x = u + w$. Also, there exists $u' \in U$ and $w' \in W$ such that $y = u' + w'$. Then

$$\begin{aligned} x + y &= (u + w) + (u' + w' \\ &= (u + u') + (w + w'). \end{aligned}$$

Now U is a subspace, and $u, u' \in U$, therefore $u + u' \in U$ because U is a subspace and closed under vector addition. Likewise $w + w' \in W$. Then $x + y \in U + W$.

Now we will show that $U + W$ is closed under "$\cdot$". Let $x \in U + W$ and α be an arbitrary scalar. Then $x = u + w$ for some $u \in U$ and $w \in W$. Then $\alpha \cdot x = \alpha(u + w) = \alpha u + \alpha w$ where $\alpha u \in U$ and $\alpha w \in W$ and therefore $\alpha \cdot x \in U + W$.

Thus $U + W$ is a subspace. □

Remark. If S_U is a spanning set of U and S_W is a spanning set of W, then $S_W \cup S_U$ is a spanning set of $U + W$.

Sketch of the proof: Let $S_U = \{u_1, \ldots, u_p\}$ and $S_W = \{w_1, \ldots, w_r\}$. Consider an arbitrary $x \in U + W$. Then $x = u + w$ for some $u \in U$ and $w \in W$. Since $u \in U$ and S_U spans U, there exists scalars $\alpha_1, \ldots, \alpha_p$ such that $u = \alpha_1 u_1 + \cdots + \alpha_p u_p$. Likewise, there exists $\beta_1, \ldots, \beta_r$ such that $w = \beta_1 w_1 + \ldots + \beta_r w_r$. Then

$$x = (\alpha_1 u_1 + \cdots + \alpha_p u_p) + (\beta_1 w_1 + \cdots + \beta_r w_r).$$

So x is a linear combination of $u_1, \ldots, u_p$ and $w_1, \ldots, w_r$ and so $x \in span(S_U \cup S_W)$. Therefore $U + W \subseteq span(S_U \cup S_W)$. Similarly, you can show that $span(S_U \cup S_W) \subseteq U + W$. Thus $U + W = span(S_U \cup S_W)$.

25 Range and null space

Definition. Let $f : V \to W$ be a linear function where V and W are vector spaces over the same field.

The **range** of f, denoted by $range(f)$, is

$$range(f) = \{f(x) : x \in V\} \subseteq W$$

which is the set of all images of x under f.

The **null space** of F, denoted by $Null(f)$ is

$$Null(f) = \{x \in V : f(x) = 0\} \subseteq V.$$

Proposition. *Let $f : V \to W$ be a linear function. Then*

(1) The range of f is a subspace of W.

(2) $Null(f)$ is a subspace of V.

Proof. (1) We need to show that $range(f)$ is closed under "+" and "·". We will only show the former. Let $y, z \in range(f)$. Since $y \in range(f)$ then there exists a x such that $y = f(x)$. Likewise, there exists x' in V such that $z = f(x')$. Consider $y + z = f(x) + f(x')$. Since f is linear, by superposition property, $f(x) + f(x') = f(x + x')$. This shows $y + z = f(x + x')$. Since $x, x' \in V$ then $x + x' \in V$ and therefore $y + z \in range(f)$.

(2) $Null(f)$ is closed under "+" and "·". We will show the former. Let $x, x' \in Null(f)$. Then $f(x) = 0$ and $f(x') = 0$. Since f is linear function; $f(x + x') = f(x) + f(x') = 0 + 0 = 0$. Therefore, $x + x' \in Null(f)$ and so $Null(f)$ is closed under "+". We can also show that $Null(f)$ is closed under "·". Thus $Null(f)$ is a subspace.

□

26 Matrix defined linear functions

Let $A \in \mathbb{R}^{m \times n}$. Consider the function $f : \mathbb{R}^n \to \mathbb{R}^m$ given by $f(x) = A \cdot x$, $x \in \mathbb{R}^n$. It's shown that f is a linear function.

- Then the $range(f) = range(A) = r(A) = \{Ax : x \in \mathbb{R}^n\}$, which is a subspace of $\mathbb{R}^m$.
- The $Null(f) = Null(A) = N(A) = \{x \in \mathbb{R}^n : Ax = 0\}$, which is a subspace of $\mathbb{R}^n$.
- Similarly we have $R(A^T)$ which is a subspace of $\mathbb{R}^n$.

- Also, $N(A^T)$ is a subspace of $\mathbb{R}^m$.

The spaces $R(A)$, $N(A)$, $R(A^T)$, and $N(A^T)$ are the four **fundamental subspaces** associated with A.

27 Spanning Sets of $R(A)$

Write the matrix A in the column form:

$$A = [A_1 \quad A_2 \quad \cdots \quad A_n].$$

Therefore, for any $x \in \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$

$$Ax = x_1 \cdot A_{\bullet 1} + x_2 \cdot A_{\bullet 2} + \cdots + x_n \cdot A_{\bullet n},$$

which is a linear combination of $A_{\bullet 1}, \ldots, A_{\bullet n}$. Thus, $R(A)$ is the set of all linear combinations of $A_{\bullet 1}, \ldots, A_{\bullet n}$. Therefore, $R(A) = span\{A_{\bullet 1}, \ldots, A_{\bullet n}\}$. Likewise, $R(A^T)$ is spanned by the columns of A^T.

28 Linear Independence

Proposition. *Let $S = \{v_1, \ldots, v_r, v_{r+1}\}$ span a subspace U. Suppose v_{r+1} is a linear combination of $v_1, \ldots, v_r$, then $U = span\{v_1, \ldots, v_r\}$.*

Proof. Clearly, $U = span\{v_1, \ldots, v_r, v_{r+1}\}$.

Claim 1: $span\{v_1, \ldots, v_r\} \subseteq U$.

Consider an arbitrary vector $x \in span\{v_1, \ldots, v_r\}$. Therefore, x is a linear combination of $v_1, \ldots, v_r$ so there exist scalars $\alpha_1, \ldots, \alpha_r$ such that

$$\begin{aligned} x &= \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_r v_r \\ &= \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_r v_r + 0 \cdot v_{r+1}. \end{aligned}$$

So $x \in U$.

Claim: 2: $U \subseteq span\{v_1, \ldots, v_r\}$.

Since v_{r+1} is a linear combination of $v_1, \ldots, v_r$ then there exists scalars $\beta_1, \ldots, \beta_r$ such that $v_{r+1} = \beta_1 v_1 + \beta_2 v_2 + \cdots \beta_r v_r$.

For any $x \in U$, x is a linear combination of $v_1, v_2, \ldots, v_r, v_{r+1}$ so there exists scalars $\alpha_1, \alpha_2, \ldots, \alpha_r, \alpha_{r+1}$ such that

$$\begin{aligned} x &= \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_r v_r + \alpha_{r+1} v_{r+1} \\ &= \alpha_1 v_1 + \cdots + \alpha_r v_r + \alpha_{r+1}(\beta_1 v_1 + \cdots + \beta_r v_r) \\ &= (\alpha_1 + \alpha_{r+1} \cdot \beta_1) v_1 + \cdots + (\alpha_r + \alpha_{r+1} \cdot \beta_r) v_r. \end{aligned}$$

Therefore, $x \in span\{v_1, \ldots, v_r\}$.
Thus $U = span\{v_1, \ldots, v_r\}$. □

Definition. A set $\{v_1, \ldots, v_p\}$ in a vector space V is **linearly dependent** if there exist scalars $\alpha_1, \ldots, \alpha_p$, not all zero, such that

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_p v_p = 0.$$

Otherwise, $\{v_1, \ldots, v_p\}$ is **linearly independent**. That is, if $\alpha_1 v_1 + \cdots + \alpha_p v_p = 0$, then $\alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$. ($\Leftarrow$ is trivial).

Example.

(1) $\{0\}$: linearly dependent because for any nonzero scalar α, $\alpha \cdot 0 = 0$.

Let $v \neq 0$, then $\{v\}$ is linearly independent.

Proof. Let α be a scalar such that $\alpha \cdot v = 0$. For sake of contradiction, assume $\alpha \neq 0$, which implies that $\frac{1}{\alpha}$ exists. Then $\dfrac{1}{\alpha}(\alpha \cdot v) = \dfrac{1}{\alpha} \cdot 0$, which implies that $v = 0$. However this contradicts the assumption that $v \neq 0$. So $\alpha = 0$, then $\{v\}$ is linearly independent. □

Summary: $\{v\}$ is linearly independent $\iff$ $v \neq 0$.

(2) $\{u, v\}$ is linearly independent if and only if one of u and v is a multiple of the other.

Proof. "if": Suppose $u = \alpha \cdot v$ (WLOG) for some scalar α. Then $u - \alpha v = 0$ if and only if $u + (-\alpha)v = 1 \cdot u + (-\alpha) \cdot v = 0$. Since not every α is 0, then $\{u, v\}$ is linearly dependent.

Not every α is 0, therefore $\{u, v\}$ is linearly dependent.

"only if": Suppose $\{u, v\}$ is linearly independent. Then there exists scalars α_1, α_2 not both zero such that

$$\alpha_1 u + \alpha_2 v = 0.$$

WLOG, assume $\alpha_1 \neq 0$, which implies $\dfrac{1}{\alpha_1}$ exists. Then multiplying the previous equation on both sides by $\dfrac{1}{\alpha_1}$ and rearranging we have

$$u = -\frac{\alpha_2}{\alpha_2} v.$$

Therefore, u is a multiple of v. Likewise, if $\alpha_2 \neq 0$, then v is a multiple of u. $\square$

Lemma. *A finite set $S = \{v_1, \ldots, v_p\}$ is linearly dependent if and only if one of the vectors in S is a linear combination of the rest of the vectors in S.*

Proof. "if": WLOG, assume that v_1 is a linear combination of $v_2, \ldots, v_p$. Then there exists scalars such that $\alpha_2, \ldots, \alpha_p$ such that

$$v_1 = \alpha_2 v_2 + \alpha_3 v_3 + \cdots + \alpha_p v_p,$$

which implies that

$$(1)v_1 + (-\alpha_2)v_2 + \cdots + (-alpha_p)v_p = 0.$$

Since the above linear combination is equal to 0, but the first scalar is 1, then S is linearly dependent.

"only if": Suppose S is linearly dependent. Then there exists scalars $\alpha_1, \ldots, \alpha_p$, not all zero, such that

$$\alpha_1 v_1 + \cdots + \alpha_p v_p = 0.$$

WLOG, assume that $\alpha_1 \neq 0$, which implies $\frac{1}{\alpha}$ exists. Then multiplying the previous equation by $\frac{1}{\alpha}$ and rearranging we have

$$v_1 = -\frac{\alpha_2}{\alpha_1} v2 - \cdots - \frac{\alpha_p}{\alpha_1} v_p.$$

Therefore, v_1 is a linear combination of $v_2, \ldots, v_p$. $\square$

Proposition. *A finite set $S = \{v_1, \ldots, v_p\}$ is linearly independent if and only if none of the vectors in S is a linear combination of the rest of the vectors in S.*

Proof. Contrapositive of the preceding lemma. $\square$

29 Minimal Spanning Set

Definition. A spanning set S of a vector space V (i.e., $V = span(S)$) is **minimal** if none of the vectors in S is a linear combination of the rest of the vectors in S.

Fact: A spanning set S is minimal if and only if it is linearly independent.

Lemma. *Let $S = \{v_1, \ldots, v_p\}$ be a minimal spanning set for a vector space V. If we remove any vector from S, then the resulting set does **not** span V.*

Proof. WLOG, assume we remove v_1 from S such that the resulting set is $S' = \{v_2, \ldots, v_p\}$.

Claim: S' does not span V.

From the given we know that $S = \{s_1, s_2, \ldots, s_p\}$ spans V. Therefore, every $v_i \in S$ belongs to V, which implies $v_1 \in V$. On the other hand, S is minimal and therefore v_1 is not a linear combination of $v_2, \ldots, v_p$. Thus, $v_1 \notin span(S')$ and so S' does not span V.

□

Remark. In the above proof we found a vector $v_1 \in V$ that is not in $span(S')$ so $span(S') \neq V$.

30 Maximal linearly independent set

Definition. A linearly independent set S in a vector space V is maximal if for any $z \in V$, $S \cup \{z\}$ is linearly dependent.

Lemma. *Let $S = \{v_1, \ldots, v_p\}$ be a linearly independent set in V. Then for any given $z \in V$, $z \in span(S)$ if and only if $S \cup \{z\}$ is linearly dependent.*

Proof. "Only if": Suppose $z \in span(S)$. Therefore, z is a linear combination of $v_1, \ldots, v_p$. Then $S \cup \{z\} = \{v_1, \ldots, v_p, z\}$ where z is a linear combination of $v_1, \ldots, v_p$. Thus $S \cup \{z\}$ is linearly dependent.

"If": Suppose $S \cup \{x\}$ is linearly dependent, which means there exists scalars $\alpha_1, \ldots, \alpha_p$ and β, not all zero, such that

$$\alpha_1 v_1 + \cdots + \alpha_p v_p + \beta z = 0.$$

Claim: The scalar $\beta \neq 0$.

For the sake of contradiction that $\beta = 0$, which implies $\beta \cdot z = 0$. Then the above equation reduces to

$$\alpha_1 v_1 + \cdots + \alpha_p v_p = 0.$$

Since $S = \{v_1, \ldots, v_p\}$ is linearly independent, then $\alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$. (and $\beta = 0$). However, this is a contradiction. Therefore, $\beta \neq 0$.

Now since $\beta \neq 0$ then we can multiply the equation

$$\alpha_1 v_1 + \cdots + \alpha_p v_p + \beta z = 0$$

through by β and then rearranging we have

$$z = -\frac{\alpha_1}{\beta} v_1 - \cdots - \frac{\alpha_p}{\beta} v_p.$$

Therefore, $z \in span(S)$. □

Theorem 30.1.

(1) A maximal linearly independent set S in V spans V, i.e., $V = span(S)$.

(2) A linearly independent set is maximal if and only if it is a minimal spanning set.

Proof. (1) Suppose S is a **maximal** linearly independent set. Then for any $z \in V$, $S \cup \{z\}$ is linearly dependent. By the preceding lemma, $z \in span(S)$. Since $z \in span(S)$ for any $z \in V$, then $span(S) = V$ or S spans V.

(2) "only if": Let S be a maximal linear independent set. By 1), S spans V. Therefore, S is linearly independent and so S is a minimal spanning set.

"if": Let S be a minimal spanning set. Then S is linearly independent and so S spans V. Thus, for any vector $z \in V$, $z \in span(S)$. By the preceding lemma, $S \cup \{z\}$ is linearly dependent. Therefore, S is a maximal linear independent set. □

31 More on linearly independent sets

Fact: Let $S = \{v_1, \ldots, v_p\}$ be a linearly independent set. Then any (non-empty) subset of S is linearly independent.

Proof. WLOG, suppose $S' = \{v_1, \ldots, v_r\} \subseteq S$, where $r < p$. Now we want to show that S' is linearly independent.

Let $\alpha_1, \ldots, \alpha_r$ be such that $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_r v_r = 0$. Then we can add in the v_{r+1} to v_p vectors to the previous linear combination by taking $\alpha_{r+1} = \cdots = \alpha_p = 0$,

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_r v_r + 0 \cdot v_{r+1} + \cdots + 0 \cdots v_p = 0.$$

Now since S is linearly independent then it must be the case that $\alpha_1 = 0, \alpha_2 = 0, \ldots, \alpha_r = 0$ and therefore $\{v_1, \ldots, v_r\}$ is linearly independent. □

Let A be an $m \times n$ matrix whose column form is

$$A = [A_{\bullet 1} \quad A_{\bullet 2} \quad \cdots \quad A_{\bullet n}]$$

where each column $A_{\bullet i}$ is in $\mathbb{R}^m$ or $\mathbb{C}^m$. If

$$(A \cdot x = 0 \iff x = 0) \iff N(A) = \{0\}$$

then the columsn of A are linearly independent.

Proposition. *Let $\{v_1, v_2, \ldots, v_r\}$ be a linearly independent set in $\mathbb{R}^n$ and $P \in \mathbb{R}^{n \times n}$ be invertible. Then $\{P \cdot v_1, P \cdot v_2, \ldots, P \cdot v_r\}$ is linearly independent.*

Proof. Let $\alpha_1, \ldots, \alpha_r$ be scalars in $\mathbb{R}$ such that

$$\alpha_1 P \cdot v_1 + \alpha 2 \cdot P \cdot v_2 + \cdots + \alpha_r P \cdot v_r = 0.$$

Then we can rearrange each of the terms and remove P from the left so that we have

$$P[\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_r v_r] = 0$$

and since P is invertible we can left multiply both sides by P^{-1} to obtain

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_r v_r = P^{-1}\dot{0} = 0.$$

Therefore, $\{Pv_1, \ldots, Pv_r\}$ are linearly independent.

□

Let A be a matrix in $\mathbb{R}^{m\times n}$ or $\mathbb{C}^{m\times n}$. Then the range of A, $R(A)$ is the **column space** of A, which is a subset of $\mathbb{R}^m$ or $\mathbb{C}^m$.

We can write A in column form and use row reduction to get in an equivalent form $P \cdot U$

$$\begin{aligned} A &= [A_{\bullet 1} \quad A_{\bullet 2} \quad \cdots \quad A_{\bullet n}] \\ &\vdots \\ &= P \cdot U \end{aligned}$$

where P is an $m \times m$ invertible matrix and U is an $m \times n$ matrix that is the echelon form of A,

$$U = [U_{\bullet 1} \quad U_{\bullet 2} \quad \cdots \quad U_{\bullet r} \quad U_{\bullet r+1} \quad \cdots \quad U_{\bullet n}],$$

where $U_{bullet1}, \ldots, U_{\bullet r}$ are the pivot columns and $U_{\bullet r+1}, \ldots, U_{\bullet n}$ are the non-pivot columns.

Facts:

(1) The pivot columns of U are linearly independent.

(2) A non-pivot column of U is a linear combination of pivot columns of U.

We can write A as

$$A = [A_{\bullet 1} \quad \cdots \quad A_{\bullet r} \quad A_{\bullet r+1} \quad \cdots \quad A_{\bullet n}],$$

and then $PA_{\bullet 1}, \ldots, PA_{\bullet r}$ are linearly because multiplying a set of linearly independent vectors by an ivertible matrix preserves linear independence.

Therefore the pivot columns of A are linearly independent and each non-pivot column of A is a linear combination of the pivot columns of A. Therefore

$$R(A) = span\{A_{\bullet 1}, \ldots, A_{\bullet n}\} = span\{A_{\bullet 1}, \ldots, A_{\bullet r}\}$$

where $span\{A_{\bullet 1}, \ldots, A_{\bullet r}\}$ is a minimal spanning set. So to find a minimal spanning set, find the pivot columns of A.

32 Basis and Dimension

Assumption: V is a vector space spannded by a finite set.

Definition. If a set $B \in V$ spans V and is linearly independent then B is a **basis** for V.

B is a basis $\iff$ $V = span(B)$ and B is linearly independent.

Fact:

$$\begin{aligned} B \text{ is a basis } &\iff B \text{ is a minimal spanning set} \\ &\iff B \text{ is maximinal linearly independent} \end{aligned}$$

Example. (1) Let $V = \mathbb{R}^2$.

Consider $B = \{e_1, e_2\} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$. Clearly, B spans V, and is linearly independent. Therefore, B is a basis (i.e., it is the **standard** basis).

- Consider $B' = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$. In a previous example we have shown that B' spans $\mathbb{R}^2$ and none of $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is a multiple of the other. Therefore, B' is linearly independent and is a basis.

(2) Let $V = \mathbb{R}^n$. Let $A \in \mathbb{R}^{n \times n}$ be an invertible matrix. Then $B = \{A_{\bullet 1}, \ldots, A_{\bullet n}\}$ is a basis for V.

Proof. First we prove the spanning property. Let $y \in \mathbb{R}^n$ be arbitrary. Therefore, y is a linear combination of the columns of A and so there exists and $x \in \mathbb{R}^n$ such that $y = A \cdot x$. Since A is invertible then the equation $Ax = y$ has the unique solution given by $A^{-1} \cdot y$. Therefore, B spans $\mathbb{R}^n$.

Now we prove that B is linearly independent. To show B is linearly independent it suffices to show that $N(A) = 0$. Since A is invertible then $N(A) = \{0\}$, which implies that B is linearly independent. Thus B is a basis. □

(3) Let W be the subspace of all 2×2 upper triangular matrices in $\mathbb{R}^{2 \times 2}$. Consider

$$B = \{E^{11}, E^{12}, E^{22}\} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}.$$

We claim that B is a basis for W.

Proof. We first prove the spanning property. We can write every $A \in W$ as

$$A = \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix} = a_{11} \cdot E^{11} + a_{12} \cdot E^{12} + a_{22} \cdot E^{22}.$$

Therefore, B spans W. Now to show linear independence, suppose $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ are such that $\alpha_1 E^{11} + \alpha_2 E^{12} + \alpha_3 E^{22} = 0$. Then if

$$\begin{bmatrix} \alpha_1 & \alpha_2 \\ 0 & \alpha_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

we must $\alpha_1 = \alpha_2 = \alpha_3 = 0$. Therefore, B is linearly independent and so B is a basis.

$\square$

Lemma. *Let $B = \{v_1, \ldots, v_n\}$ be a basis for a vector space V. Then for any $x \in V$, there exist **unique** scalars $\alpha_1, \ldots, \alpha_n$ such that*

$$x = \alpha_1 v_1 + \cdots + \alpha_n v_n.$$

Proof. Let x be an arbitrary vector in V. Since B is a basis then B spans V and so every vector in V is a linear combination of $v_1, \ldots, v_n$. Then there exist scalars $\alpha_1, \ldots, \alpha_n$ such that

$$x = \alpha_1 v_1 + \cdots + \alpha_n v_n.$$

To show uniqueness of $\alpha_1, \ldots, \alpha_n$ let $\beta_1, \ldots, \beta_n$ be scalars such that

$$x = \beta_1 v_1 + \cdots + \beta_n v_n.$$

Now if we subtract $x = \alpha_1 v_1 + \cdots + \alpha_n v_n$ and the previous equation we get

$$0 = (\alpha_1 - \beta_1)v_1 + \cdots + (\alpha_n - \beta_n)v_n.$$

Since B is linearly independent then it must be the case that $\alpha_1 - \beta_1 = 0, \alpha_2 - \beta_2 = 0, \ldots, \alpha_n - \beta_n = 0$. Therefore, $\beta_1 = \alpha_1, \beta_2 = \alpha_2, \ldots, \beta_n = \alpha_n$. This shows the uniqueness of the α_i's. $\square$

Definition. Let $B = \{v_1, \ldots, v_n\}$ be a basis for V. Then for each $x \in V$, $(\alpha_1, \ldots, \alpha_n)$ is the coordinate of x with respect to the basis B denoted by $[x]_\beta$. Therefore

$$[x]_\beta = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{R}^n.$$

Definition. The **dimension** of a vector space V is the number of vectors in a basis of V, and it is denoted by $\dim V$.

Example. (1) $\dim \mathbb{R}^n = n$.

(2) $\dim\{0\} = 0$.

Lemma. *Let A be an $m \times n$ matrix. Suppose $n > m$. Then the columns of A are linearly dependent.*

Theorem 32.1. Let B_1 and B_2 be two bases for a vector space V. Then $\#(B_1) = \#(B_2)$, where $\#(B_i)$ is the number of vectors in B_i for $i = 1, 2$.

Proof. Let $B_1 = \{v_1, \ldots, v_p\}$ and $B_2 = \{u_1, \ldots, u_r\}$ be two bases for V. Since B_1 is a basis for V then B_1 spans V. Also, each $u_i \in B_2$ is in V and therefore each u_i is a linear combination of the vectors in B_1, $v_1, \ldots, v_p$, i.e.,

$$u_i = \alpha_{i1}v_1 + \alpha_{i2}v_2 + \cdots + \alpha_{ip}v_p$$

for $i = 1, \ldots, r$.

Therefore, we can write B_2 as

$$B_2 = [u_1 \quad \cdots \quad u_r] = [v_1 \quad \cdots \quad v_p] \cdot \begin{bmatrix} \alpha_{11} & \alpha_{21} & \cdots & \alpha_{r1} \\ \alpha_{12} & \alpha_{22} & \cdots & \alpha_{r2} \\ \vdots & \vdots & & \vdots \\ \alpha_{1p} & \alpha_{2p} & & \alpha_{rp} \end{bmatrix}_{p \times r} = B_1 \cdot A$$

where A is a $p \times r$ matrix.

Claim: $r \leq p$.

Suppose otherwise; i.e., $r > p$. Then by the preceding lemma, the columns of A are linearly dependent. Then there exists a vector $w \neq 0$ such that $Aw = 0$. If we let multiply the above equation by w then we have

$$[u_1 \quad \cdots \quad u_r] \cdot w = [v_1 \quad \cdots \quad v_p] \cdot A \cdot w = 0.$$

Now let $w = \begin{bmatrix} w_1 \\ \vdots \\ w_r \end{bmatrix} \neq 0$. Then

$$w_1u_1 + w_2u_2 + \cdots + w_ru_r = 0,$$

which implies that $\{u_1, \ldots, u_r\}$ is linearly dependent. However, this contradicts our assumption that $B_2 = \{u_1, \ldots, u_r\}$ is a basis. Thus it must be the case that

$$r \leq p \iff \#(B_2) \leq \#(B_1).$$

By interchanging the roles of B_1 and B_2 we obtain $p \leq r \iff \#(B_1) \leq \#(B_2)$. Thus, $\#(B_1) = \#(B_2)$. $\square$

Proposition. *Let V be an n-dimensional vector space, i.e., $dimV = n$. Then*

(1) A maximal linearly independent set in V has n vectors.

(2) A minimal spanning set in V has n vectors.

(3) A spanning set for V has at least n vectors.

(4) A linearly independent set in V has at most n vectors.

Proof. (1) and (2): Recall that a maximal linearly independent set and a minimal spanning set are bases for V. Since $\dim V = n$ then each basis has exactly n vectors. Thus, a maximal linearly independent or a minimal spanning set contains n vectors.

(3) Let S be a spanning set for V. If S is minimal, then by 2) $\#(S) = n$. If S is not minimal, then it can be reduced to a minimal spanning set S', where $n = \#(S') < \#(S)$. Thus, S has at least n vectors. $\square$

Proposition. *Let V be an n-dimensional vector space, and W is a subspace of V.*

(1) $\dim W \leq \dim V$

(2) If $\dim W = \dim V$*, then* $W = V$.

Proof. (1) Let B be a basis for the subspace W. Then B is linearly independent and $B \subseteq W \subseteq V$. Then

$$\dim W = \#(B) \leq n = \dim V.$$

Therefore, $\dim W \leq \dim V$.

(2) Let $\dim W = \dim V$. Suppose for the sake of contradiction that $W \neq V$. Since $W \subseteq V$ and $W \neq V$ then there exists a vector $z \in V$ but $z \notin W$. Let B be a basis for W, i.e., $W = span(B)$. Since $z \notin W$, then $z \notin span(B)$. Since B is a basis then B is linearly independent and so $B \cup \{z\}$ is linearly independent. Therefore $\#(B \cup \{z\} = \#(B) + 1$. Since $\dim V = \dim W = n$ and therefore $\#(B) = n$. But $\#(B \cup \{z\}) = n + 1$ and therefore $B \cup \{z\}$ is a linear independent set in V with $n + 1$ vectors. However by a previous proposition, this is a contradiction. Therefore, our initial assumption that $W \neq V$ must be false. Thus, $W = V$.

$\square$

Theorem 32.2. Let U and W be two subspaces of a vector space V (of finite dimension). Then

$$\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W).$$

Proof. Not given in class. $\square$

Example. Let $V = \mathbb{R}^3$ and let $U = span\{e_1\}$ and $W = span\{e_1, e_2\}$. Since U is a subspace of W then $U \subseteq V$ and $U + W = W$ so $U \cap W = U$. Therefore, $\dim U = 1$ and $\dim W = 2$. Then $\dim(U + W) = \dim(W) = 2$. On the right we have $\dim U + \dim W - \dim(U \cap W) = 1 + 2 = 2$.

33 Dimension of four fundamental subspaces associated with an $m \times n$ matrix

Without loss of generality let $A \in \mathbb{R}^{m\times n}$. Then

- $R(A)$: range or the column space of A is a subspace of $\mathbb{R}^m$.
- $N(A)$: null space of A is a subspace of $\mathbb{R}^n$.
- $R(A^T)$: range or column space of A^T is a subspace of $\mathbb{R}^n$.
- $N(A^T)$: null space of A^T is a subspace of $\mathbb{R}^m$.

Definition. The dimension of $R(A)$ is called the **rank** of A, i.e., $\dim[R(A)] = rank(A)$. The dimension of the null space $N(A)$ is the **nullity** of A.

Proposition. *$rank(A) = 0$ if and only if $A = 0$, i.e., A is the zero matrix.*

Proof. If $A = 0$ then $R(A) = \{0\}$, which implies $\dim[R(A)] = 0$. If $rank(A) = 0$, then $\dim[R(A)] = 0$, which implies $R(A) = \{0\}$ and so $A = 0$. □

Let A be an $m \times n$ real matrix.

(1) The pivot columns of A form a basis for $R(A)$.

(2) Let r be the number of pivots in A. Then

$$\dim[R(A)] = r = rank(A).$$

(3) $\dim[N(A)] =$ the number of non-pivot columns of $A = n - r$.

Theorem 33.1. (Rank-nullity theorem) Let A be an $m \times n$ matrix. Then $rank(A) + \dim[N(A)] = n$.

Fact:

- $rank(A) = \dim[R(A^T)] =$ number of pivots in $A = r$.
- $rank(A^T) + \dim[N(A^T)] = m$.

34 Rank of a matrix product

Lemma. *(Special Case) Let A : $m \times n$ matrix, and P : $m \times m$ invertible. Then $rank(P \cdot A) = rank(A)$.*

Proof. Need to show: $dim(R(P \cdot A)) = dim(R(A))$.

Let $A_{\bullet 1}$, $A_{\bullet 2}$, ..., $A_{\bullet r}$ be all pivot columns of A such that $\{A_{\bullet 1}, A_{\bullet 2}, \ldots, A_{\bullet r}\}$ is a basis for $R(A)$.

Claim: $\{PA_{\bullet 1}, PA_{\bullet 2}, \ldots, PA_{\bullet r}\}$ is a basis for $R(P \cdot A)$.

- Spanning Property: Since $\{A_{\bullet 1}, A_{\bullet 2}, \ldots, A_{\bullet r}\}$ is a basis for $R(A)$ then $\{A_{\bullet 1}, A_{\bullet 2}, \ldots, A_{\bullet r}\}$ spans $R(A)$. Therefore, $\{PA_{\bullet 1}, PA_{\bullet 2}, \ldots, PA_{\bullet r}\}$.

- Linear Independence: Let $\alpha_1, \alpha_2, \ldots, \alpha_r$ be such that $\alpha_1 PA_{\bullet 1} + \cdots + \alpha_r PA_{\bullet r} = 0$. Then $P(\alpha_1 A_{\bullet 1} + \cdots + \alpha_r A_{\bullet r}) = 0$. Since P is invertible, we can right multiply both sides of the previous equation by P^{-1} and then we have $\alpha A_{\bullet 1} + \cdots + \alpha_r A_{\bullet r} = 0$. Since $\{A_{\bullet 1}, \ldots, A_{\bullet r}\}$ is a basis, then the set is linear independent and so $\alpha_1 = 0, \alpha_2 = 0, \ldots, \alpha_r = 0$. Thus, $\{PA_{\bullet 1}, \ldots, PA_{\bullet r}\}$ is linearly independent and so $\{PA_{\bullet 1}, \ldots, PA_{\bullet r}\}$ is a basis for $R(PA)$. Now since $dim(R(A)) = r = dim(R(PA))$ then $rank(A) = rank(P \cdot A)$.

□

Now we will prove a more general theorem, but first we need some remarks and a preliminary lemma. The statement of the general theorem is: Let A be an $m \times n$ matrix and B be a $n \times p$ matrix. Then $rank(A \cdot B) = rank(B) - dim(N(A) \cap R(B))$.

Remark.

(1) Since A is $m \times n$ then $N(A)$ is a subspace of $\mathbb{R}^n$. Since B is $n \times p$ then $R(B)$ is a subspace of $\mathbb{R}^n$. Therefore, $N(A) \cap R(B)$ is a subspace of $\mathbb{R}^n$.

(2) For any $z \in (N(A) \cap R(B))$, $z \in N(A)$ so that $Az = 0$ and $z \in R(B)$ so that $z = B \cdot w$ for some w. Therefore, $Az = ABw = 0$ and so any $z \in N(A) \cap R(B)$ does not "contribute" to $dim(R(A \cdot B))$.

Lemma. *Let V be an n-dimensional vector space and W be a r-dimensional space of V with $r < n$. Suppose $\{v_1, \ldots, v_r\}$ is a basis for W. Then there exists $v_{r+1}, v_{r+2}, \ldots, v_n \in V$ such that $\{v_1, \ldots, v_r, v_{r+1}, \ldots, v_n\}$ is a basis for V.*

Proof. Need to show $\{v_1, \ldots, v_r, v_{r+1}, \ldots, v_n\}$ is maximally linearly independent, which implies that $\{v_1, \ldots, v_r, v_{r+1}, \ldots, v_n\}$ is a basis. □

Theorem 34.1. Let A be an $m \times n$ matrix and B be a $n \times p$ matrix. Then $rank(A \cdot B) = rank(B) - dim(N(A) \cap R(B))$.

Proof. Since $N(A) \cap R(B)$ is a subspace of $\mathbb{R}^n$. Let $\{v_1, \ldots, v_r\}$ be a basis for $N(A) \cap R(B)$. Then $dim(N(A) \cap R(B)) = r$.

Now consider 2 cases:

(1) $r = n$. We know $N(A) \cap R(B)$ is a subspace of $N(A)$ and a subspace of $R(B)$. Since $r = n$, then $n = dim(N(A) \cap R(B)) \leq dim(R(B)) \leq n$, where the last inequality follows because $R(B)$ is a subspace of $\mathbb{R}^n$. Therefore, $dim(R(B)) = n$ and so $rank(B) = n$. Also, since $r = n$, then $n = dim(N(A) \cap R(B)) \leq dim(N(A)) \leq n$. The last inequality holds again because $N(A)$ is a subspace of $\mathbb{R}^n$. Therefore, $dim(N(A)) = n$ and by the rank-nullity theorem we have

$$\begin{aligned} rank(A) &= n - dim(N(A)) \\ &= n - n \\ &= 0. \end{aligned}$$

Since $rank(A) = 0$, then $A = 0$. Now since $rank(B) = 0$ and $A = 0$ then $(A \cdot B) = 0$. Thus

$$rank(A \cdot B) = 0 = rank(B) - dim(N(A) \cap R(B)).$$

(2) $r < n$. In this case, we use the preliminary lemma to obtain $z_1, z_2, \ldots, z_t \in R(B)$ such that $\{v_1, \ldots, v_r, z_1, \ldots, z_t\}$ is a basis for $R(B)$. Therefore $rank(B) = dim(R(B)) = r+t$. Then $\{v_1, \ldots, v_r, z_1, \ldots, z_t\}$ spans $R(B)$ and so $\{Av_1, \ldots, Av_r, Az_1, \ldots, Az_t\}$ spans $R(B)$.

Recall that $\{v_1, v_r\}$ is a basis for $N(A) \cap R(B)$. Therefore, for each $v_i \in N(A)$, for all $i = 1, \ldots, r$, we have $Av_i = 0$ for all $i = 1, \ldots, r$. Now we can drop the zero vectors from $\{Av_1, \ldots, Av_r, Az_1, \ldots, Az_t\}$ so that we have $\{Az_1, \ldots, Az_2\}$, which also spans $R(A \cdot B)$.

Claim: $\{Az_1, \ldots, Az_t\}$ is linearly independent.

Let $\alpha_1, \ldots, \alpha_t$ be scalars such that

$$\alpha_1 Az_1 + \cdots + \alpha_t Az_t = A(\alpha_1 z_1 + \cdots + \alpha_t z_t) = 0.$$

Now define $y = (\alpha_1 z_1 + \cdots + \alpha_t z_t)$. From above we see that $y \in N(A)$ and $y \in R(B)$ and so $y \in N(A) \cap R(B)$. Since $\{v_1, \ldots, v_r\}$ is a basis for $N(A) \cap R(B)$ and $y \in N(A) \cap R(B)$ then there exists scalars $\beta_1, \ldots, \beta_r$ such that

$$y = \beta_1 v_1 + \cdots + \beta_r v_r$$

.

Now since $y = \alpha_1 z_1 + \cdots + \alpha_t v_t$ and $y = \beta_1 v_1 + \cdots + \beta_r v_r$ then

$$\alpha_1 z_1 + \cdots + \alpha_t v_t = \beta_1 v_1 + \cdots + \beta_r v_r$$

and then subtracting $\beta_1 v_1 + \cdots + \beta_r v_r$ from both sides we have

$$\alpha_1 z_1 + \cdots + \alpha_t v_t + (-\beta_1)v_1 + \cdots + (-\beta_r)v_r = 0.$$

Since $\{v_1, \ldots, v_r, z_1, \ldots, z_t\}$ is a basis for $R(B)$ then $\{v_1, \ldots, v_r, z_1, \ldots, z_t\}$ is linearly independent. Therefore $\beta_1 = 0 = \cdots = \beta_r = 0$ and $-\alpha_1 = 0 = \cdots = -\alpha_t = 0$ so $\alpha_1 = 0 = \cdots = \alpha_t = 0$. Thus, $\{Az_1, \ldots, Az_t\}$ is linearly independent.

Then $\{Az_1, \ldots, Az_t\}$ is a basis for $R(A \cdot B)$ and so $dim(R(AB)) = t$ and then $rank(AB) = t$. Therefore, $rank(B) = dim(R(B)) = r + t$ and $dim(N(A) \cap R(B)) = r$ and so

$$rank(AB) = t = r + t - r = rank(B) - dim(N(A) \cap R(B)).$$

□

Example.

(1) Let $A : m \times n$ and $P : m \times n$ invertible matrix.

(2) Rank of $A \cdot A^T$ and $A^T \cdot A$ where $A \in \mathbb{R}^{m \times n}$.

Theorem 34.2. *$rank(A \cdot A^T) = rank(A)$ and $rank(A^T \cdot A) = rank(A)$.*

Proof. Note that $rank(A^T \cdot A) = rank(A) - dim(N(A^T) \cap R(A))$.

Claim: $N(A^T) \cap R(A) = \{0\}$.

Let $z \in N(A^T) \cap R(A)$. Then $z \in N(A^T)$ and $z \in R(A)$. Therefore, $A^T z = 0$ and $z = Ax$ for some x. Then $A^T z = A^T Ax = 0$. Then left multiply by x^T,

$$\begin{aligned}
&\Rightarrow x^T A^T Ax = x^T \\
&\Rightarrow (A \cdot x)^T Ax = 0 \in \mathbb{R} \\
&\Rightarrow z^T z = 0 \in \mathbb{R} \\
&\Rightarrow z = 0.
\end{aligned}$$

Therefore, $N(A^T) \cap R(A) = \{0\}$. Then $rank(A^T \cdot A) = rank(A)$.

To show $rank(A \cdot A^T) = rank(A)$, we see that

$$\begin{aligned}
rank(A \cdot A^T) &= rank((A^T)^T A^T) \\
&= rank(A^T) \\
&= rank(A).
\end{aligned}$$

□

Theorem 34.3. *Let A be an $m \times n$ matrix. Then*

(1) *$R(A^T \cdot A) = R(A^T)$ and $R(A \cdot A^T) = R(A)$.*

(2) *$N(A^T \cdot A) = N(A)$ and $N(A \cdot A^T) = N(A^T)$.*

Proof.

(1) Prove $R(A^T \cdot A) = R(A)$ only. Note that $R(A \cdot B)$ is a subspace of $R(A)$ and therefore $R(A^T \cdot A)$ is a subspace of $R(A^T)$. Then it suffices to show that $dim(R(A^T \cdot A)) = dim(R(A^T)) \iff rank(A^T \cdot A) = rank(A^T)$. Since $rank(A^T \cdot A) = rank(A^T)$ and so $R(A^T \cdot A) = R(A^T)$.

(2) Show $N(A^T \cdot A) = N(A)$ only.

Clearly, $N(A)$ is a subspace of $N(A^T \cdot A)$. This is because $x \in N(A) \iff Ax = 0 \Rightarrow A^T Ax = 0 \Rightarrow N(A^T \cdot A)$.

Is suffices to show that

$$dim(N(A)) = dim(N(A^T \cdot A))$$

.

Recall from the rank-nullity theorem that $rank(A) + dim(N(A)) = n$ and so $dim(N(A)) = n - rank(A)$. Likewise, since $A^T A$ is an $n \times n$ matrix, then $dim(A^T \cdot A) = n - rank(A^T \cdot A)$. Since $rank(A^T \cdot A) = rank(A)$ and $dim(N(A)) = dim(N(A^T \cdot A))$. Therefore, $N(A^T \cdot A) = N(A)$.

□

Remark. The following implications of the above theorems.

Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.

(1) Consider the linear equation

$$Ax = b.$$

In general, this equation may **not** have a solution.

(2) Consider the **normal** equation

$$A^T \cdot Ax = A^T b.$$

Fact: $A^T \cdot Ax = A^T b$ always has a solution (or is always **consistent**).

Proof. Since $R(A^T A) = R(A^T)$ and $A^T b \in R(A^T)$ then $A^T b \in R(A^T A)$. Then there exists a $p \in \mathbb{R}^n$ such that $A^T Ap = A^T b$. Therefore, p is a solution to $A^T \cdot A \cdot x = A^T \cdot b$. □

(3) If $Ax = b$ has a solution or is consistent, then $Ax = b$ and $A^T Ax = A^T b$ have the same solution set.

(4) Suppose $Ax = b$ has a **unique** solution. Then $A^T \cdot A$ is invertible.

Proof. Since $Ax = b$ has a unique solution. Therefore $N(A) = \{0\}$ because otherwise there exists a $0 \neq z \in N(A)$. Then $Ap = b$ and since A is linear then $A(p + z) = Ap + Az = b = 0$. However, this is a contradiction. Therefore, $N(A) = \{0\}$.

Since $N(A^T \cdot A) = N(A)$ and $N(A) = \{0\}$ then $N(A^T A) = \{0\}$. Thus, $A^T A$ is invertible. Then the unique solution to $A^T Ax = A^T b$ is

$$x_* = (A^T A)^{-1} \cdot A^T \cdot b,$$

which is the unique solution to $Ax = b$. □

35 Application to the least squares problem

Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^n$. The least-squares (LS) problem is to minimize

$$\begin{aligned} F(x) &= ||Ax - b||_2^2 \\ &= (Ax - b)^T (Ax - b) \\ &\geq 0. \end{aligned}$$

An optimal solution x_* to the LS problem satisfies:

$$\nabla F(x_*) = 0$$

where $\nabla F(x) = 2(A^T Ax = A^T b)$. Since x_* is such that $A^T Ax = A^T b$, i.e., x_* is solution to normal equation. Therefore, the LS problem always has a solution.

Claim: x_* satisfying the normal equation is an optimal solution to the LS problem.

Proof. Let $x = x_* + u$, where $u \in \mathbb{R}^n$ is arbitrary. Then

$$\begin{aligned} F(x) &= (Ax - b)^T (Ax - b) \\ &= (Ax_* - b + Au)^T (Ax_* - b + Au) \\ &= (Ax_* - b)^T (Ax_* - b) + 2(Ax_* - b)^T Au + (Au)^T (Au). \end{aligned}$$

Note that the first term on the RHS above is $F(x_*)$, the second term $2(Ax_* - b)^T Au$ goes to zero, and the final term is a sum of squares and so is ≥ 0.

Note that

$$2[(Ax_* - b)^T A]u = 2[A^T (Ax_* - b)]^T \cdot u = 0$$

because term in brackets is equal to 0 because x_* satisfies the normal equation.

Therefore, $F(x) \geq F(x_*)$, for all $u \in \mathbb{R}^n$ or any $x \in \mathbb{R}^n$. Thus, x_* is an optimal solution to the LS problem. □

36 Coordinate of a vector

Definition. Let V be an n-dimensional vector space, and $B = \{v_1, \ldots, v_n\}$ be a basis for V.

For each $x \in V$, let

$$[x]_B = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{R}^n \text{ or } \mathbb{C}^n$$

where $\alpha_1, \ldots, \alpha_n$ satisfy $x = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$. Then $[x]_B$ is the **coordinate vector** for x.

Fact:

(1) $x \neq 0$ if and only if $[x]_B \neq 0$.

(2) $x = 0$ if and and only $[x]_B = 0$.

(3) For each $x \in V$, $[x]_B$ is unique.

Remark. Recall that a linear function $T : U \to V$ where U, V are finite dimensional vector spaces, satisfies

(1) $T(x + y) = T(x) + T(y)$, for all $x, y \in U$

(2) $T(\alpha x) = \alpha \cdot T(x)$, for all $x, y \in U$ and scalar α.

Then for any $x, y \in U$ and scalars α, β

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(Y).$$

Generalization

$$T\left(\sum_{i=1}^{n} \beta_i u_i\right) = \sum_{i=1}^{n} \beta_i \cdot T(u_i) \text{ for all } u_i \in V, i = 1, \ldots, n.$$

Special Case:

If $T : V \to V$ is a linear function (i.e., $U = V$), then T is call a **linear operator**.

Example. Linear Operator

(1) $T : \mathbb{R}^n \to \mathbb{R}^n$ given by $T(x) = A \cdot x$, where A is an $n \times n$ real matrix.

(2) General linear operators: e.g., identity operator $I : V \to V$ given by $I(x) = x$ for all $x \in V$.

(3) Some other operators are projection operators and reflection operators.

37 Coordinate Matrix

Let linear function $T : U \to V$ where U, V are finite dimensional vector spaces. Let $B = \{u_1, \dots, u_n\}$ be a basis for U and $B' = \{v_1, \dots, v_m\}$ be a basis for V.

Question: How to characterize the relation between $[x]_B$ and $[T(x)]_{B'}$?

We can use the **coordinate matrix** of T from B to B'.

For each $u_i \in B$, $T(u_i) \in V$, since $T(u_i) = \alpha_{1i}v_1 + \alpha_{2i}v_2 + \cdots + \alpha_{mi}v_m$ then for weights $\alpha_{1i}, \dots, \alpha_{mi}$

$$T(u_i) = \begin{bmatrix} v_1 & \dots & v_m \end{bmatrix} \begin{bmatrix} \alpha_{1i} \\ \alpha_{2i} \\ \vdots \\ \alpha_{mi} \end{bmatrix}$$

Define the following $m \times n$ matrix,

$$[T]_{BB'} = [[T(u_1)]_{B'}, [T(u_2)]_{B'}, \cdots [T(u_n)]_{B'}] \quad = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & & \vdots \\ \alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mn} \end{bmatrix} \in \mathbb{R}^{m \times n}.$$

Let $[T]_{BB'}$ denote this matrix, which is called the **coordinate matrix** of T from B to B'.

Proposition. *$[T(x)]_{B'} = [T]_{BB'} \cdot [x]_B$ for all $x \in U$.*

Proof. Let x be an arbitrary vector in U. Since $\{u_1, \dots, u_n\}$ is a basis for U then we can write x as a linear combination of the vectors B

$$x = \beta_1 u_1 + \cdots + \beta_n u_n.$$

Then the coordinate vector of x with respect to B is

$$[x]_B = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \in \mathbb{R}^n.$$

Then we have

$$
\begin{aligned}
T(x) &= T(\beta_1 u_1 + \cdots + \beta_n u_n) \\
&= \beta_1 T(u_1) + \cdots + \beta_n T(u_n) \\
&= \begin{bmatrix} T(u_1) & T(u_2) & \cdots & T(u_n) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \\
&= \begin{bmatrix} v_1 + \cdots + v_m \end{bmatrix} [T]_{BB'} \cdot [x]_B.
\end{aligned}
$$

Therefore $[T(x)]_{B'} = [T]_{BB'} \cdot [x]_B$. □

Remark. For a linear operator $T : V \to V$, if B is a basis for V, the coordinate matrix of T from B to B is $[T]_{BB}$, which can be written as $([T]_{BB}) = [T]_B$. (Because same basis for domain and codomain.)

If V is an n-dimensional vector space, then $[T]_B$ is an $n \times n$ matrix for any basis B for V.

38 Change of basis

Question Let T be a linear operator and B and B' be two bases for V. What is the relation between $[T]_B$ and $[T]_{B'}$?

38.1 Change of basis

Let $B = \{u_1, \dots, u_n\}$ and $B' = \{v_1, \dots, v_n\}$ be two bases for an n-dimensional vector space V. Then the coordinate matrix of the identity operator $I : V \to V$ from B to B' is $[I]_{BB'}$.

Then by construction of $[I]_{BB'}$, we have

$$
\begin{aligned}
[I]_{BB'} &= \begin{bmatrix} [I(u_1)]_{B'} & [I(u_2)]_{B'} & \cdots & [I(u_n)]_{B'} \end{bmatrix} \\
&= \begin{bmatrix} [u_1]_{B'} & [u_2]_{B'} & \cdots & [u_n]_{B'} \end{bmatrix}_{n \times n} \\
&= P.
\end{aligned}
$$

Remark.

(1) For any $x \in V$, $[x]_{B'} = [I]_{BB'} \cdot [x]_B = P \cdot [x]_B$.

(2) $P = [I]_{BB'}$ is invertible.

Proof. Recall $[T(x)]_{B'} = [T]_{BB'} \cdot [x]_B$. Let $T = I$ the identity operator. Then $[x]_{B'} = [I]_{BB'} \cdot [x]_B$ for all $x \in V$.

Suppose P was not invertible. Then there exists $b \neq 0$ in $\mathbb{R}^n$ such that $P \cdot b = 0$. Then there exists an $x \in V$ such that $[x]_B = b$. Since $b \neq 0$ then $x \neq 0$. Then $P \cdot b = [I]_{BB'} \cdot [x]_B = [x]_{B'}$ (by 1). Then $[x]_{B'} = 0$ and so $x = 0$. However this contradicts our assumption that if $b \neq 0$ then $x \neq 0$. Thus P is invertible. □

Theorem 38.1. Let $P = [I]_{BB'}$, the change of basis matrix. Then

(1) $[T]_B = P^{-1} \cdot [T]_{B'} \cdot P$,

(2) $[T]_{B'} = P \cdot [T]_B \cdot P^{-1}$.

Proof. It suffices to show (1), which is equivalent to $P \cdot [T]_B = [T]_{B'} \cdot P$. To show $P \cdot [T]_B = [T]_{B'} \cdot P$,

$$\begin{aligned}
P \cdot [T]_B &= P \begin{bmatrix} [T(u_1)]_B & \cdots & [T(u_n)]_B \end{bmatrix} \\
&= \begin{bmatrix} P[T(u_1)]_B & \cdots & P[T(u_n)]_B \end{bmatrix} \\
&= \begin{bmatrix} [T(u_1)]_{B'} & \cdots & [T(u_n)]_{B'} \end{bmatrix} \\
&= \begin{bmatrix} [T]_{B'} \cdot [u_1]_{B'} & \cdots & [T]_{B'} \cdot [u_n]_{B'} \end{bmatrix} \quad (1) \\
&= [T]_{B'} \begin{bmatrix} [u_1]_{B'} & \cdots & [u_n]_{B'} \end{bmatrix} \\
&= [T]_{B'} \cdot P.
\end{aligned}$$

where (1) follows because $[T(x)]_{B'} = [T]_{B'} \cdot [x]_{B'}$.
Therefore, $P \cdot [T]_B = [T]_{B'}$. □

39 Similarity

Definition. Let A and B be two $n \times n$ matrices. If there exists an invertible matrix P such that $A = P^{-1}BP$ then A is **similar** to B, and we write $A \simeq B$.

39.1 Properties of similar relations

(1) Any $n \times n$ matrix is similar to itself. (By letting $P = I$).

(2) If A is similar to B, then B is similar to A.

Since A is similar to B then $A = P^{-1}BP$ and so $B = PAP^{-1} = (P^{-1})^{-1}AP^{-1}$.

(3) If $A \simeq B$ and $B \simeq C$ then $A \simeq C$.

(4) If A is similar to B then A^k is similar to B^k for any $k = 1, 2, \ldots$.

(5) If A is similar to B then $rank(A) = rank(B)$ and $trace(A) = trace(B)$. Also, $det(A) = det(B)$ and A and B have the same eigenvalues.

Proof. To show $trace(A) = trace(B)$. We have

$$\begin{aligned} trace(A) &= trace(P^{-1}BP) \\ &= trace((B \cdot P) \cdot P^{-1}) \\ &= trace(B \cdot I) \\ &= trace(B). \end{aligned}$$

$\square$

Remark. In the above proof we used the fact that $trace(C \cdot D) = trace(D \cdot C)$, if C and D are square.

40 Inner products, norms, and orthogonality

40.1 Inner products on $\mathbb{R}^n$ and $\mathbb{C}^n$

Let $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ be two vectors in $\mathbb{C}^n$ or $\mathbb{R}^n$.

If $x, y \in \mathbb{R}^n$, then the (standard) inner product is

$$x^T y = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \in \mathbb{R}^n$$

If $x, y \in \mathbb{C}^n$ then the (standard) inner product of x and y is

$$x^* \cdot y = \begin{bmatrix} \overline{x_1} & \overline{x_2} & \cdots & \overline{x_n} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \overline{x_1} \cdot y + \cdots + \overline{x_n} \cdot y_n \in \mathbb{C}^n$$

40.2 Properties of (standard) inner products on $\mathbb{R}^n$ or $\mathbb{C}^n$

Let $x, y \in \mathbb{C}^n$. Then the following hold:

(1) $y^* \cdot x = \overline{x^* \cdot y}$ (When $x, y \in \mathbb{R}^n$, then $x^T y = y^T x$)

(2) For any $\alpha \in \mathbb{C}$, $x^*(\alpha \cdot y) = \alpha(x^* \cdot y)$, and $(\alpha \cdot x)^* \cdot y = \bar{\alpha}(x^* \cdot y)$.

(3) For any $z \in \mathbb{C}^n$, $x^* \cdot (y + z) = x^* \cdot y + x^* \cdot z$.

(4) $x^* x \geq 0$ and $x^* \cdot x = 0$ if and only $x = 0$. (Note that $x^* \cdot x \in \mathbb{R}$).

Proof. (1) We have that $y^* \cdot x = \overline{y_1} \cdot x_1 + \overline{y_2} \cdot x_2 + \cdots + \overline{y_n} \cdot x_n$ and

$$\begin{aligned}
\overline{x^* \cdot y} &= \overline{\overline{x_1} \cdot y_1 + \cdots + \overline{x_n} \cdot y_n} \\
&= \overline{\overline{x_1} \cdot y_1} + \cdots + \overline{\overline{x_n} \cdot y_n} \\
&= \overline{\overline{x_1}} \cdot \overline{y_1} + \cdots + \overline{\overline{x_n}} \cdot \overline{y_n} \\
&= x_1 \cdot \overline{y_1} + \cdots + x_n \overline{y_n} \\
&= \overline{y_1} \cdot x_1 + \cdots + \overline{y_n} \cdot x_n.
\end{aligned}$$

Therefore, $y^* \cdot x = x^* \cdot y$.

(2) Let $\alpha \in \mathbb{C}$. Then $x^* \cdot (\alpha \cdot y) = \alpha(x * \cdot y)$ (by definition of $x^* \cdot y$).

To show $(\alpha x)^* \cdot y = \bar{\alpha}(x^* \cdot y)$ we have

$$\begin{aligned}
(\alpha \cdot x)^* \cdot y &= \overline{y^* \cdot (\alpha x)} \\
&= \overline{\alpha \cdot (y^* \cdot x} \\
&= \bar{\alpha} \cdot \overline{y^* \cdot x} \\
&= \bar{\alpha} \cdot \overline{\overline{x^* y}} \\
&= \bar{\alpha} \cdot x^* y.
\end{aligned}$$

(3) skipped

(4) Given $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{C}^n$, $x^* \cdot x = \overline{x_1} \cdot x_1 + \cdots + \overline{x_n} \cdot x_n$.

Since each $x_i \cdot \overline{x_i} \in \mathbb{R}$ and $x_i \cdot \overline{x_i} \geq 0$ for all $i = 1, \ldots, n$ then $x_1 \cdot \overline{x_1} + \cdots + x_n \cdot \overline{x_n} = \cdot x \geq 0$.

Clearly, if $x = 0$, then $x^* \cdot x = 0$. Conversely, if $x^* \cdot x = 0$, then since $x^* \cdot x = |x_1|^2 + |x_2|^2 + \cdots + |x_n|^2 = 0$ then $|x_i| = 0$ for all $i = 1, 2, \ldots, n$. Therefore, $x_i = 0$ for all $i = 1, 2, \ldots, n$ and so $x = 0$.

□

Remark. Recall that if $z = a + bi \in \mathbb{C}$ where $a, b \in \mathbb{R}$ then $z \cdot \bar{z} = (a + bi)(a - bi) = a^2 + b^2 \in \mathbb{R}$ and $|z|^2 = a^2 + b^2 \geq 0$ where $|z| = \sqrt{a^2 + b^2}$.

41 Norms on $\mathbb{R}^n$ and $\mathbb{C}^n$

Definition. Let $x \in \mathbb{C}^n$. Then it's (induced) norm is

$$||x|| = \sqrt{x^* x}$$

If $x \in \mathbb{R}^n$, then $||x|| = \sqrt{x^T \cdot x}$.

41.1 Properties of the induced norms

Let $x \in \mathbb{C}^n$.

(1) $||x|| \geq 0$ and $||x|| = 0 \iff x = 0$.

(2) For any $\alpha \in \mathbb{C}$, $||\alpha \cdot x|| = |\alpha| \cdot ||x||$ where $|\alpha|$ is the modulus of $\alpha \in \mathbb{C}$ for $\alpha = a + bi \in \mathbb{C}$.

If $x \in \mathbb{R}$, then $|\alpha|$ is the absolute value of α.

(3) Cauchy-Schwartz Inequality.

$$|x^* \cdot y| \leq ||x|| \cdot ||y||, \quad \forall x, y \in \mathbb{C}^n$$

(4) Triangle Inequality for $|| \cdot ||$.

$$||x + y|| \leq ||x|| + ||y||.$$

Proof. (1) Since $x^* \cdot x \geq 0$ and $x^* \cdot x \in \mathbb{R}$ for all $x \in \mathbb{C}^n$ then $||x|| = \sqrt{x^* \cdot x} \geq 0$. Further,

$$\begin{aligned} ||x|| = 0 &\iff \sqrt{x^* \cdot x} - 0 \\ &\iff x^* \cdot x = 0 \\ &\iff x = 0. \end{aligned}$$

where the last $\iff$ follows from properties of the inner product.

(2) Let $\alpha \in \mathbb{C}$. Then

$$\begin{aligned} ||\alpha \cdot x||^2 &= (\alpha \cdot x)^* \cdot (\alpha \cdot x) \\ &= \alpha(\alpha x)^* \cdot (x) \\ &= \alpha \cdot \bar{\alpha}(x^* \cdot x) \\ &= |\alpha|^2 \cdot ||x||^2. \end{aligned}$$

Therefore, taking square roots we have $||\alpha \cdot x|| = |\alpha| \cdot ||x||$.

Remark. Before proving the Cauchy-Schwartz inequality we need a couple of facts. Let $z = a + bi$ where $a, b\ in \mathbb{R}$.

(i) $z \cdot \bar{z} = (a + bi)(a - bi) = a^2 + b^2 = |z|^2$.

(ii) $z + \bar{z} = (a + bi) + (a - bi) = 2a \leq 2 \cdot |z|$, which implies that $z + \bar{z} \leq 2|z|$.

(3) Proof of Cauchy-Schwartz Inequality.

When $x = 0$, then $|x^* y| = 0$ and $||x|| \cdot ||y|| = 0$. Then the C-S inequality holds.

Now consider the case when $x \neq 0$ and therefore $||x|| > 0$.

Define α as

$$\alpha = \frac{x^* \cdot y}{||x||^2}$$

for a given $y \in \mathbb{C}$. Note that since $||x|| \geq 0$ then α is well defined. Then

$$\begin{aligned} x^* \cdot (\alpha x - y) &= \alpha \cdot x^* \cdot x - x^* \cdot y \\ &= \alpha \cdot ||x||^2 - x^* \cdot y \\ &= \frac{x^* \cdot y}{||x||^2} ||x||^2 - x^* \cdot y \\ &= 0. \end{aligned}$$

Note that $||\alpha \cdot x - y||^2 \geq 0$ and

$$\begin{aligned} ||\alpha \cdot x - y||^2 &= (\alpha \cdot x - y)^* \cdot (\alpha x - y) \\ &= [(\alpha x)^* - y^*] \cdot (\alpha x - y) \\ &= [\bar{\alpha} x^* - y^*] \cdot (\alpha x - y) \\ &= \bar{\alpha} \cdot x^* \cdot (\alpha x - y) - y^* \cdot (\alpha x - y) \\ &= y^* \cdot y - \alpha y^* \cdot x \end{aligned}$$

where the last equality follows because we know from earlier that $x^* \cdot (\alpha x - y) = 0$. Therefore

$$y^* \cdot y - \alpha \cdot y^* \cdot x = ||\alpha x - y||^2 \geq 0$$

and so

$$||y||^2 \geq \alpha \cdot y^* \cdot x = \frac{x^* \cdot y}{||x||^2} y^* \cdot x.$$

Multiplying by $||x||^2$ on both sides the the previous inequality we have

$$\begin{aligned} ||x||^2 \cdot ||y||^2 &\geq (x^* \cdot y)(y^* \cdot x) \\ &= (x^* \cdot y)(x^* \cdot y) \\ &= (x^* \cdot y) \cdot \overline{x^* \cdot y} \\ &= ||z||^2 \\ &= |x^* y|^2. \end{aligned}$$

Therefore, $||x||^2||y||^2 \geq |x^* \cdot y|^2$ and taking square roots on both sides we have

$$||x|| \cdot ||y|| \geq |x^* \cdot y|.$$

(4) Proof of Triangle Inequality

$$\begin{aligned}
||x+y||^2 &= (x+y)^*(x+y) \\
&= (x^* + y^*) \cdot (x+y) \\
&= (x^* \cdot x) + (x^* \cdot y) + (y^* \cdot x) + (y^* \cdot y) \\
&= ||x||^2 + x^* \cdot y + x^* \cdot y + ||y||^2 \\
&= ||x||^2 + 2 \cdot Re(x^* \cdot y) + ||y||^2 \\
&\leq ||x||^2 + 2 \cdot |x^* \cdot y| + ||y||^2 \\
&\leq ||x||^2 + 2||x|| \cdot ||y|| + ||y||^2 \\
&\leq (||x|| + ||y||)^2.
\end{aligned}$$

Therefore, $||x+y||^2 \leq (||x|| + ||y||)^2$ and taking the square root of both sides we have

$$||x+y|| \leq ||x|| + ||y||.$$

$\square$

42 General Inner Product Space

Definition. An **inner product** on a real or complex vector space V is a function

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R} \text{ or } \mathbb{C}$$

that satisfies the following four properties:

(1) $\langle x, x \rangle \geq 0$ for $x \in V$, and $\langle x, x \rangle \iff x = 0$.

(2) $\langle x, y+z \rangle = \langle x, y \rangle + \langle x, z \rangle$ for all $x, y, z \in V$.

(3) $\langle x, \alpha \cdot y \rangle = \alpha \langle x, y \rangle$ for all $x, y \in V$ and scalar α in $\mathbb{R}$ or $\mathbb{C}$.

(4) $\langle y, x \rangle = \overline{\langle x, y \rangle}$ for all $x, y \in V$.

Then V, along with the inner product $\langle \cdot, \cdot \rangle$ is an **inner product space**. (IPS)

42.1 Implications of Inner Products

(i) It can be shown that

$$\langle x, \alpha_1 v_1 + \cdots + \alpha_p v_p \rangle = \alpha_1 \langle x, v_1 \rangle + \alpha_2 \langle x, v_2 \rangle + \cdots + \alpha_p \langle x, v_p \rangle.$$

(ii) The induced norm is

$$||x|| = \sqrt{\langle x, x \rangle}$$

for all $x \in V$.

It can be shown that:

- $||x|| \geq 0$, and $||x|| = 0 \iff x = 0$
- $||\alpha x|| = |\alpha| \cdot ||x||$ for all $x \in V$ and scalar α.
- $|\langle x, y \rangle| \leq ||x|| \cdot ||y||$ (Cauchy Schwartz)
- $||x + y|| \leq ||x|| + ||y||$ (Triangle Inequality)

(iii) $\langle x, 0 \rangle$ for all $x \in V$.

Proof. Note that

$$\begin{aligned}\langle x, 0 \rangle &= \langle x, y + (-y) \rangle \\ &= \langle x, y \rangle + \langle x, (-1) \cdot y \rangle \\ &= \langle x, y \rangle + (-1)\langle x, y \rangle \\ &= 0.\end{aligned}$$

□

Example. Inner Products

(1) Let $V = \mathbb{R}$ and let $A \in \mathbb{R}^{n \times n}$ be invertible. For any $x, y \in \mathbb{R}^n$

$$\langle x, y \rangle_A = (Ax)^T (Ay)$$

Note that if $A = I$, the identity matrix, then the above is the standard inner product.

Claim: $\langle \cdot, \cdot \rangle_A$ is an inner product on V.

Proof. It's easy to show $\langle \cdot, \cdot \rangle_A$ satisfies conditions $2) - 4)$. Only show condition 1).

Since $\langle x, x \rangle_A = (Ax)^T (Ax) = z^T z$ where $z = Ax \in \mathbb{R}^n$ and $z^T z \geq 0$, then $\langle x, x \rangle_A \geq 0$. Further, if $x = 0$, then $\langle x, x \rangle_A = 0$.

Conversely, if $\langle x, x \rangle_A = 0$, then $z^T z = 0$, where $z = A \cdot x$, which implies that $z = 0$ and so $Ax = 0$. Since A is invertible, then $x = 0$, then $\langle x, x \rangle_A = 0 \iff x = 0$. Therefore, $\langle \cdot, \cdot \rangle_A$ is an inner product. □

(2) Let $V = \mathbb{R}^{n\times n}$, the space of all real matrices of size $n \times n$. For any $A, B \in V$, let

$$\langle A, B\rangle = trace(A^T \cdot B).$$

.

Claim: $\langle\cdot,\cdot\rangle$ is an inner product.

Proof. Fill in. $\square$

The induced norm is

$$||A||_F = \sqrt{\langle A, A\rangle} = \sqrt{trace(A^T \cdot A)}.$$

Here, $||\cdot||_F$ is the Frobenius norm for a matrix.

Definition. Let $||\cdot||$ be a norm on V. Then a vector $x \in V$ is a **unit vector** if $||x|| = 1$.

Fact: A unit vector must be a non-zero vector.

42.2 Normalization of a Vector

Given any non-zero vector $z \in V$,

$$\frac{z}{||z||}$$

is a unit vector.

Proof.

$$\begin{aligned}
\left|\left|\frac{z}{||z||}\right|\right| &= \left|\left|\frac{1}{||z||}\cdot z\right|\right| \\
&= |\alpha|\cdot||z|| \\
&= \alpha\cdot||z|| \\
&= \frac{1}{||z||}\cdot||z|| \\
&= 1
\end{aligned}$$

where $\alpha = \dfrac{1}{||z||} > 0$. Therefore $\dfrac{z}{||z||}$ is a unit vector. $\square$

43 Orthogonal Vectors

Definition. Let $(V, \langle \cdot, \cdot \rangle)$ be a real/complex inner product space. Given $x, y \in V$, x is **orthogonal** to y (i.e., $x \perp y$) if $\langle x, y \rangle = 0$.

Remark. Facts about Orthogonal Vectors

(1) If $x \perp y$, then $y \perp x$?

Claim: $\langle x, y \rangle = 0 \implies \langle y, x \rangle = 0$.

Proof. Since $\langle y, x \rangle = \overline{\langle x, y \rangle} = \bar{0} = 0$, then $y \perp x$. □

(2) Is the zero vector orthogonal to any vector in V? Answer: Yes because $\langle x, 0 \rangle = 0$ for all $x \in V$. Thus, $0 \perp x$ for all $x \in V$.

Definition. Let $S = \{v_1, \ldots, v_p\}$ be a finite set in an inner product space V. We say S is an **orthogonal set** if

$$\langle v_i, v_j \rangle = 0 \qquad \forall i \neq j$$

or

$$v_i \perp v_j \qquad \forall v_i, v_j \in S \text{ with } i \neq j.$$

If in addition, each $v_i \in S$ is a unit vector, then S is an **orthonormal** set (or simply an O.N. set).

Remark. Note that $||v_i|| = 1 \iff ||v_i||^2 = 1$. Therefore, S is orthonormal if and only if:

$$\langle v_i, v_j \rangle = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} = \delta_{ij}$$

where δ_{ij} is the delta function.

Example. Let $V = \mathbb{R}^3$ and $B = \{e_1, e_2, e_3\}$ be the standard basis for $\mathbb{R}^3$. Show that B is an O.N. set.

Proof. For any i, j from $\{1, 2, 3\}$, we have

$$\langle e_i, e_j \rangle = e_i^T \cdot e_j = \begin{cases} 0, & i \neq j, \\ 1, & i = j \end{cases}.$$

Therefore, B is orthonormal. □

44 Properties of Orthogonal Sets

Lemma. *An orthogonal set of non-zero vectors is linearly independent.*

Proof. Let $\{u_1, \ldots, u_p\}$ be an orthogonal set where each $u_i \neq 0$.

Suppose there are scalars $\alpha_1, \ldots, \alpha_p$ such that

$$\alpha_1 u_1 + \cdots + \alpha_p u_p = 0.$$

We want to show $\alpha_1 = 0, \alpha_2 = 0, \ldots, \alpha_p = 0$.

Note that for each u_i $\langle u_i, 0\rangle = 0$ and so

$$\langle u_i, \alpha_1 u_1 + \cdots + \alpha_i u_i + \cdots + \alpha_p u_p\rangle = 0.$$

The left hand side is:

$$\alpha_1\langle u_i, u_1\rangle + \cdot + \alpha_i\langle u_i, u_i\rangle + \cdots + \alpha_p\langle u_i, u_p\rangle.$$

Since $\{u_1, \ldots, u_p\}$ is an orthogonal set then $\langle u_i, u_j\rangle = 0$ if $i \neq j$. Therefore, the left hand side equals $\alpha_i\langle u_i, u_i\rangle = \alpha_i||u_i||^2$. (Since $u_i \neq 0$ then $\langle u_i, u_i\rangle > 0$). Since $u_i \neq 0$ then $||u_i||^2 > 0$ and then since $\alpha_i \cdot ||u_i||^2 = 0$ then $\alpha_i = 0$.

Therefore, $\alpha_1 = 0, \alpha_2 = 0, \ldots, \alpha_p = 0$ and so $\{u_1, \ldots, u_p\}$ is linearly independent. □

Remark. In the above proof we can repeat the process for each i to show that $\alpha_i = 0$.

Corollary. An O.N. set is linearly independent.

Proof. This is because each vector in an O.N. set is a unit vector and must be non-zero. Then by the preceding lemma, an O.N. set is linearly independent. □

45 Orthonormal Basis

Definition. Let B be a basis for a finite dimensional inner product space (IPS). If B is an orthonormal set, then B is an **orthonormal basis**.

Corollary. Let V be an n-dimensional IPS. An O.N. set of n vectors in V is an O.N. basis for V.

Proof. Let B be an O.N. set of n vectors. By the above corollary, B is linearly independent. If V is of $\dim n$ and B has n vectors then B is a maximally linearly independent set in V. Therefore, B is a basis for V and so B is an orthonormal basis for V. □

Proposition. *Let $B = \{u_i, \ldots, u_n\}$ be an O.N. basis for V. Given a vector $x \in V$, x can be written as*

$$x = \langle u_1, x\rangle u_1 + \cdots + \langle u_n, x\rangle u_n$$

where $\langle u_i, x\rangle$ are the Fourier coefficients and writing x this way is a special case of the Fourier Expansion.

Remark. The coordinate vector of x with respect to basis B can be written as

$$[x]_B = \begin{bmatrix} \langle u_1, x\rangle \\ \langle u_2, x\rangle \\ \vdots \\ \langle u_n, x\rangle \end{bmatrix}.$$

Proof. Since $B = \{u_1, \dots, u_p\}$ is an O.N. basis for V then for any given vector $x \in V$, there exists unique scalars $\alpha_1, \dots, \alpha_n$ such that

$$x = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n.$$

For each u_i (i.e., fix i),

$$\begin{aligned} \langle u_i, x\rangle &= \langle u_i, \alpha_1 u_1 + \cdots + \alpha_i u_i + \cdots + \alpha_n u_n\rangle \\ &= \alpha_1\langle u_i, u_1\rangle + \cdots + \alpha_i\langle u_i, u_i\rangle + \cdots + \alpha_n\langle u_i, u_n\rangle. \end{aligned}$$

Since B is O.N. then

$$\langle u_i, u_j\rangle = \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j \end{cases}.$$

Therefore, $\langle u_i, x\rangle = \alpha_i$ for all $i = 1, \dots, n$. (Repeat the above process for each fixed i).

□

46 Idea for Gram-Schmidt Procedure/Process

Goal: Let $B = \{v_1, \dots, v_n\}$ be a basis for an n-dimensional IPS V. Construct an O.N. basis from B.

Remark. In general, there could be multiple O.N. bases and the Gram-Schmidt process gives one O.N. basis that depends on what basis you start from.

We will use the following preliminary result below:

Lemma. *Any subset of a linearly independent set is also linearly independent.*

Gram-Schmidt Process: To find an orthonormal basis for an n-dimensional IPS V from $B = \{v_1, \dots, v_n\}$, the Gram-Schmidt process uses the following process:

Step 1 Let $W_1 = \mathit{span}\{v_1\}$.

Clearly, $\{v_1\}$ spans W_1, and is linearly independent because it is a subset of B. Since $\{v_1\}$ is a basis for W_1 then $\dim W_1 = 1$ and therefore W_1 is a one-dimensional subspace of V.

Step 2 Let $W_2 = span\{v_1, v_2\}$.

Likewise, $\{v_1, v_2\}$ is a basis for W_2 and therefore $\dim W_2 = 2$ and so W_2 is a two dimensional subspace of V. Also, W_1 is a one-dimensional subspace of W_2 (i.e., $W_1 \subseteq W_2$).

Steps $3, \dots, n$ Similarly, we can define $W_2, W_3, \dots, W_k, \dots, W_n$, i.e., $W_k = span\{v_1, v_2, \dots, v_k\}$ is an k-dimensional subspace of V. Finally, $W_n = span\{v_1, \dots, v_n\} = V$.

Note that $W_1 \subseteq W_2 \subseteq W_3 \subseteq \cdots \subseteq W_k \subseteq \cdots \subseteq W_n = V$.

Remark. The idea of the GS process is to construct on O.N. basis for W_1 and use that to construct O.N. basis for W_2, ... O.N. basis for $W_n = V$.

Definition. Consider the subspace W_k. Suppose $\{u_1, \dots, u_k\}$ is an O.N. basis for W_k. Consider an arbitrary vector $x \in V$ (not necessarily in W_k). We call

$$p = \langle u_1, x\rangle u_1 + \cdots + \langle u_k, x\rangle u_k$$

the **orthogonal projection** of x onto W_k denoted by $\operatorname{proj}_{\mathbf{W_k}} \mathbf{x}$.

Properties of $\operatorname{proj}_{\mathbf{W_k}} \mathbf{x}$.

(1) $\operatorname{proj}_{\mathbf{W_k}} \mathbf{x} \in W_k$ because x is a linear combination of $u_1, \dots, u_k$, which are basis vectors for W_k.

(2) $(x - \operatorname{proj}_{\mathbf{W_k}} \mathbf{x})$ is orthogonal to each u_i, $i = 1, \dots, k$.

Proof. For each u_i, $i = 1, \dots, k$ compute

$$\langle u_i, x - \operatorname{proj}_{\mathbf{W_k}} \mathbf{x}\rangle = \langle u_i, x\rangle - \langle u_i, p\rangle$$

where

$$\begin{aligned} \langle u_i, p\rangle &= \langle u_i, \alpha_1 u_1 + \cdots + \alpha_k u_k\rangle \\ &= \alpha_i \langle u_i, u_i\rangle \\ &= \alpha_i \\ &= \langle u_i, x\rangle. \end{aligned}$$

with $\alpha_k = \langle u_i, x\rangle$ and $\langle u_i, u_i\rangle = 1$. Therefore, $\langle u_i, x - \operatorname{proj}_{\mathbf{W_k}} \mathbf{x}\rangle = 0$ and so $x - \operatorname{proj}_{\mathbf{W_k}} \mathbf{x} \perp u_i$ for all $i = 1, \dots, k$.

Therefore $x - \operatorname{proj}_{\mathbf{W_k}} \mathbf{x}$ is orthogonal to any vector in W_k.

□

47 Gram-Schmidt Procedure/Process

Goal: Construct on O.N. basis from a basis $B = \{v_1, \dots, v_n\}$ for an n-dimensional IPS V.

47.1 Preliminary Results

We will use the following preliminary results:

(1)

$$\begin{aligned} W_1 &= span\{v_1\} \\ W_2 &= span\{v_1, v_2\} \\ &\vdots \\ W_k &= span\{v_1, \dots, v_k\} \\ &\vdots \\ W_n &= span\{v_1, \dots, v_n\} = V \end{aligned}$$

(2) Projection of $x \in V$ onto W_k, assuming $\{u_1, \dots, u_k\}$ is an O.N. basis for W_k is

$$\mathrm{proj}_{\mathbf{W_k}} \mathbf{x} = \langle u_1, x\rangle u_1 + \cdots + \langle u_k, x\rangle u_k.$$

Then,

(1) $\mathrm{proj}_{\mathbf{W_k}} \mathbf{x} \in W_k$

(2) $(x - \mathrm{proj}_{\mathbf{W_k}} \mathbf{x}) \perp (u_i)$ for all $i = 1, \dots, k$.

47.2 Gram-Schmidt Procedure (Inductive Process)

Step 1 Local goal in this step is to construct on O.N. basis for W_1. Since $W_1 = span\{v_1\}$ and $v_1 \neq 0$ then $\{v_1\}$ is a basis for W_1. Using the normalization of v_1 we have

$$u_1 = \frac{v_1}{||v_1||} \in W_1$$

and u_1 is a unit vector. Therefore, $\{u_1\}$ is an orthonormal basis for W_1.

Step 2 Local goal is to construct O.N. basis for $W_2 = span\{v_1, v_2\}$, where W_1 is a subspace of W_2.

From step 1, $\{u_1\}$ is an O.N. basis for $W_1 \subseteq W_2$. To generate a vector in W_2 that is orthogonal to u_1 , we consider

$$\begin{aligned}\tilde{u_2} &= v_2 - \text{proj}_{\mathbf{W_1}} \mathbf{v_2} \\ &= v_2 - \langle v_2, u_1 \rangle \cdot u_1 \in W_1.\end{aligned}$$

Properties of $\tilde{u_2}$:

(1) $\widetilde{u}_2 \in W_2$. Since $\widetilde{u}_2$ is a linear combination of v_1 and v_2 then $\widetilde{u}_2 \in W_2$.

(2) $\widetilde{u}_2 \neq 0$. This is because otherwise, $v_w = \langle v_2, u_1 \rangle \cdot u_1 =$ a multiple of v_1. (as $u_1 = \dfrac{v_1}{||v_1||}$).

(3) $\widetilde{u}_2 \perp u_1$. Since $\{u_1, \widetilde{u}_2\}$ is an orthogonal set, and both $\widetilde{u}_2 \neq 0$ and $u_1 = 0$.

Using the normalization of $\widetilde{u}_2$ we have

$$u_2 = \frac{\widetilde{u}_2}{||\widetilde{u}_2||} = \frac{v_2 - \langle u_1, u_2 \rangle u_1}{||v_2 - \langle u_1, v_2 \rangle u_1||} \in W_2.$$

Since $\{u_1, u_2\} \subset W_2$, $u_1 \perp u_2$, and $||u_1|| = ||u_2|| = 1$, then $\{u_1, u_2\}$ is an orthonormal set in W_2 and therefore linearly independent. Thus, $\{u_2, u_2\}$ is an O.N. basis W_2 where

$$u_1 = \frac{v_1}{||v_1||}, \quad u_2 = \frac{v_2 - \langle u_1, u_2 \rangle u_1}{||v_2 - \langle u_1, v_2 \rangle u_1||}.$$

Step $k+1$: Where $1 \leq k \leq n-1$. Local goal is to construct on O.N. basis for W_{k+1}.

Induction Hypothesis: W_k has an orthonormal basis given by $\{u_1, \ldots, u_k\}$.

Now define

$$\begin{aligned}\widetilde{u}_{k+1} &= v_{k+1} - \text{proj}_{\mathbf{W_k}} \mathbf{v_{k+1}} \\ &= v_{k+1} - (\langle u_1, v_{k+1} \rangle u_1 + \cdots + \langle u_k, v_{k+1} \rangle u_k) \in W_{k+1}\end{aligned}$$

Properties of $\widetilde{u}_{k+1}$

(1) $\widetilde{u}_{k_1} \in W_{k+1}$. This is because $u_{k+1} \in W_{k+1}$ and $\text{proj}_{\mathbf{W_k}} \mathbf{v_{k+1}} \in W_{k+1}$ such that $v_{k+1} - \text{proj}_{\mathbf{W_k}} \mathbf{v_{k+1}} = \widetilde{u}_{k+1}$ and $v_{k+1} - \text{proj}_{\mathbf{W_k}} \mathbf{v_{k+1}} \in W_{k+1}$.

(2) $\widetilde{u}_{k+1} \neq 0$. Using a result from HW # 8.

(3) $\widetilde{u}_{k+1} \perp u_i$ for all $i = 1, \ldots, k$

Using the normalization of $\widetilde{u}_{k+1}$ we have

$$u_{k+1} = \frac{\widetilde{u}_{k+1}}{||\widetilde{u}_{k+1}||} = \frac{v_{k+1} - (\langle u_1, v_{k+1} \rangle u_1 + \cdots + \langle u_k, v_{k+1} \rangle u_k)}{||v_{k+1} - (\langle u_1, v_{k+1} \rangle u_1 + \cdots + \langle u_k, v_{k+1} \rangle u_k)||}$$

.

Therefore, $\{u_1, \ldots, u_k, u_{k+1}\}$ is an O.N. set in W_{k+1}, which are all orthogonal by assumption and 3) above. Since $\{u_1, \ldots, u_k, u_{k+1}\}$ is an O.N. then it is linearly independent and therefore $\{u_1, \ldots, u_k, u_{k+1}\}$ is an O.N. basis for W_{k+1}.

By the induction principle, we obtain an orthonormal basis $\{u_1, \ldots, u_n\}$ for $W_n = V$.

47.3 Summary of Gram-Schmidt Procedure

(1) $u_1 = \dfrac{v_1}{||v_1||}$

(2) $u_2 = \dfrac{\widetilde{u}_2}{||\widetilde{u}_2||} = \dfrac{v_2 - \langle u_1, u_2 \rangle u_1}{||v_2 - \langle u_1, v_2 \rangle u_1||}$

(k) $u_k = \dfrac{v_k - \displaystyle\sum_{i=1}^{k+1} \langle u_i, v_k \rangle u_i}{\left|\left| v_k - \displaystyle\sum_{i=1}^{k+1} \langle u_i, v_k \rangle u_i \right|\right|}$

(n) $u_n = \dfrac{v_n - \displaystyle\sum_{i=1}^{n-1} \langle u_i, v_n \rangle u_i}{\left|\left| v_n - \displaystyle\sum_{i=1}^{n-1} \langle u_i, v_n \rangle u_i \right|\right|}$

Example. Let $V = \mathbb{R}^3$ and $B = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$. Find an O.N. basis for $\mathbb{R}^3$ from B.

Solution:

$$u_1 = \frac{v_1}{||v_1||} = v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ because } ||v_1|| = 1.$$

$$\text{Then } v_2 - \langle u_1, v_2 \rangle \cdot u_1 = v_2 - u_1 = v_2 - v_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

$$\text{Normalizing we have } u_2 = \frac{v_2 - \langle u_1, v_2 \rangle u_1}{||v_2 - \langle u_1, v_2 \rangle u_1||} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

$$\text{Then } v_3 - \langle u_1, v_3\rangle u_1 - \langle u_2, v_3\rangle u_2 = v_3 - u_1 - u_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

$$\text{Therefore, } u_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

$$\text{Thus, } \{u_1, u_2, u_3\} \text{ is an O.N. basis for } \mathbb{R}^3$$

48 Orthogonal and Unitary Matrices

Definition.
A matrix $P \in \mathbb{R}^{n\times n}$ is **orthogonal** if its columns form an O.N. basis for $\mathbb{R}^n$.
A matrix $U \in \mathbb{C}^{n\times n}$ is **unitary** if its columns form an O.N. basis for $\mathbb{C}^n$.

Proposition. *Let $P \in \mathbb{R}^{n\times n}$. Then P is orthogonal if and only if the columns of P, $\{p_1, p_2, \ldots, p_n\}$, is an O.N. basis for $\mathbb{R}^n$.*

Proof. $\Leftarrow$ Let $P = \begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix}$, where p_k is the k^{th} column of P. The condition that $\{p_1, p_2, \ldots, p_n\}$, is an O.N. basis for $\mathbb{R}^n$ is equivalent to

$$\langle p_i, p_j\rangle = p_i^T p_j = \begin{cases} 0, & \text{if } i \neq j, \\ 1, & \text{if } i = j \end{cases}.$$

Then if we multiply P^T by P we have

$$\begin{aligned} P^T \cdot P &= \begin{bmatrix} p_1^T \\ p_2^T \\ \vdots \\ p_n^T \end{bmatrix} \begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix} \\ &= \begin{bmatrix} p_1^T & p_1^T p_2 & \cdots & p_1^T p_n \\ p_2^T & p_2^T p_2 & \cdots & p_2^T p_1 \\ \vdots & \vdots & \ddots & \vdots \\ p_n^T p_1 & p_n^T & \cdots & p_n^T p_n \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \end{aligned}$$

By definition, P is orthogonal if $P^T P = I_n$, thus if $\{p_1, p_2, \ldots, p_n\}$, is an O.N. basis for $\mathbb{R}^n$, then P is orthogonal.

($\Rightarrow$) Conversely, if $P^T P = I_n$ then

$$\langle p_i, p_j \rangle = p_i^T p_j = \begin{cases} 0, & \text{if } i \neq j, \\ 1, & \text{if } i = j \end{cases}.$$

Therefore, $\{p_1, p_2, \ldots, p_n\}$ is an O.N. basis for $\mathbb{R}^n$ and so P is orthogonal matrix.

In summary, P is an orthogonal matrix $\iff P^T \cdot P = I_n$. $\square$

Theorem 48.1. Let $P \in \mathbb{R}^{n\times n}$. Then P is an orthogonal matrix if and only if $||P\cdot x|| = ||x||$ for all $x \in \mathbb{R}^n$, where $||x|| = \sqrt{x^T x}$ is the induced norm (or the 2-norm/Euclidean norm).

Here, P is an **isometry**. (This means if does not change the length of the norm).

Proof. "only if": Suppose P is orthogonal. Then $P^T P = I_n$. Then for a given $x \in \mathbb{R}^n$,

$$\begin{aligned} ||Px||^2 &= (Px)^T (Px) \\ &= x^T \cdot (P^T \cdot P)\dot{x} \\ &= x^T \cdot x \\ &= ||x||^2. \end{aligned}$$

Then, taking square roots we have $||Px|| = ||x||$.

"If": Suppose $||Px|| = ||x||$ for all $x \in \mathbb{R}^n$. Then

$$\begin{aligned} ||Px||^2 = ||x||^2 &\iff (Px)^T (Px) = x^T x \\ &\iff x^T P^T \cdot Px - x^T I x = 0 \\ &\iff x^T \cdot (P^T P - I) \cdot x = 0. \end{aligned}$$

In the above equality we cannot conclude that $A = 0$. To see this, suppose $x^T \cdot A \cdot x = 0$ for all $x \in \mathbb{R}^n$. Let $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Then A is skew symmetric (i.e., $A^T = A$). Then for any $x \in \mathbb{R}^2$. Then

$$x^T A x = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} x_2 \\ -x_1 \end{bmatrix} = x_1 x_2 - x_1 x_2 = 0.$$

Now let $x = e_i$, $i = 1, \ldots, n$ where e_i is the i^{th} standard basis vector of $\mathbb{R}^n$.

Then $Px = P \cdot e_i = p_i$, where p_i is the $i^t h$ column of P and therefore $||Px||^2 - ||p_i||^2$ and $||x||^2 = ||e_i||^2 = 1$. Since $||p_i||^2 = 1$ then $p_i^T p - 1$ for all $i = 1, \ldots, n$.

Let $x = e_i + e_j$ where $i \neq j$. Then $e_i^T e_j = 0$ and $P \cdot x - P(e_i + e_j) = Pe_i + Pe_j = P_i + P_j$. Further

$$\begin{aligned} ||Px||^2 &= (p_i + p_j)^T \cdot (p_i + p_j) \\ &= P_i^T P_i + P_i^T P_j + P_j^T P_i + P_j^T P_j \\ &= 2 + 2P_i^T P_j. \end{aligned}$$

Also,

$$\begin{aligned}||x||^2 &= (e_i + e_j)^T(e_i + e_j)\\ &= e_i^T e_i + e_i^T + e_j^T e_i + e_j^T e_j\\ &= 2.\end{aligned}$$

Since $||Px||^2 = ||x||^2$ then $2 + 2 \cdot P_i^T P_j = 2$ and subtracting 2 and dividing by 2 on both side we have $P_i^T P_j = 0$ for all $i \neq j$. Thus, P is an orthogonal matrix. □

The above theorem also holds for unitary matrices.

Example. (Isometries)

(1) Let $P = I_n$, the identity matrix.

(2) Let $P = \begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}_{2\times 2}$, where $\theta \in (-\pi, \pi]$ in $\mathbb{R}^2$.

Then

$$P^T \cdot P = \begin{bmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} = I_2.$$

Thus, P is orthogonal.

Remark. In (2) above the matrix $P = P(\theta)$ is a rotation matrix and $P(\theta)$ is commutative in the sense that $P(\theta_1) - P(\theta_2) = P(\theta_1 + \theta_2)$.

49 Global Properties of Orthogonal (or Unitary) Matrices

Let $O(n)$ be the set of all $n \times n$ orthogonal matrices. ($O(n) \subseteq \mathbb{R}^{n\times n}$)

Properties of $O(n)$ under the matrix product "$\cdot$"

(1) Let $A, B \in O(n)$. **Claim:** $A \cdot B \in O(n)$.

Proof. Since $A, B \in O(n)$ then $A^T A = B^T B = I$. Then

$$(A \cdot B)^T \cdot (A \cdot B) = B^T \cdot A^T \cdot A \cdot B = I_n.$$

Therefore, $A \cdot B$ is orthogonal and so $A \cdot B \in O(n)$. Thus, $O(n)$ is closed under "$\cdot$". □

(2) The matrix product "$\cdot$" is associative, i.e., $(AB) \cdot C = A \cdot (BC)$.

(3) The identity matrix $I_n \in O(n)$.

(4) For any $A \in O(n)$, $A^{-1} \in O(n)$.

Proof. Since $A \in O(n)$ then $A^T \cdot A = I_n$ and so $A^{-1} = A^T$. Therefore,

$$(A^{-1})^T(A^{-1}) = (A^T)^T \cdot A^{-1} = A \cdot A^{-1} = I_n.$$

Therefore, $A^{-1} \in O(n)$. $\square$

Remark. From the above properties, $(O(n),$ "$\cdot$"$)$ is a group, more specifically a Lie group.

For any $P \in O(n)$, $P^T \cdot P = I_n$ and $\det(P^T \cdot P) = (\det P)^2$ and $\det I_n = 1$. Therefore, $\det P = \pm 1$.

The Special $O(n)$ group is the set $SO(n) = \{P \in O(n) : \det P = 1\}$. This is the Special Orthogonal Group of Order n. When $n = 2$, we get the space of all rotation matrices. $SU(n)$ is the special unitary group of order n.

50 Projection

Let V be a vector space, and X and Y be two subspaces of V.

Remark. Recall that $X + Y = \{x + y : x \in X \text{ and } y \in Y\}$ and $X \cap Y = \{v : v \in X \text{ and } v \in Y\}$ are subspaces of V and $Y \subseteq X + Y$ and $x \subseteq X + Y$.

If X and Y are finite dimensional, then $\dim(X+Y) = \dim(X)+\dim(Y)-\dim(X\cap Y)$

Definition. The subspaces X and Y are **complementary** subspaces of V if

(1) $V = X + Y$

(2) $X \cap Y = \{0\}$

In this case, V is the **direct sum** of X and Y, and it is written as

$$V = X \oplus Y.$$

Theorem 50.1. Let X and Y be subspaces of V with bases $B_x = \{u_1, \ldots, u_p\}$ and $B_y = \{v_1, \ldots, v_r\}$ for X and Y, respectively. Then the following are equivalent:

1. $V = X \oplus Y$
2. For any $v \in V$, there exists "unique" vectors $x \in X$ and $y \in Y$ such that $V = X + Y$.
3. $B_x \cap B_y = \varnothing$ and $B_x \cup B_y$ is a basis for V.

Proof. We will prove the following chain of implications: $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 1$.

($1 \Rightarrow 2$) Since $V = X \oplus Y$, then $V = X + Y$, and $X \cap Y = \{0\}$. For an arbitrary vector $v \in V$, $v \in V = X + Y$ and therefore, there exists an $x \in X$ and a $y \in Y$ such that $v = x + y$. To prove uniqueness, suppose that there exists an $x' \in X$ and a y' such that $x' + y' = v$. Since $x + y = v$, then $x + y = v = x' + y'$. Therefore, $x - x' = y' - y$. Since $x \in X$ and $x' \in X$ and X is a subspace, then $x - x' \in X$ (because X is closed under "+" and "·"). Likewise, $y - y' \in Y$ because $y \in Y$ and $y' \in Y$ and Y is a subspace. Since $x - x' = y - y'$ and $X \cap Y = \{0\}$, then $x - x' = y - y' = 0$, which implies that $x' = x$ and $y' = y$. This proves uniqueness.

($2 \Rightarrow 3$) First we show that $B_X \cap B_Y = \varnothing$. Suppose not. Then there exists a $v \in B_X \cap B_Y$ such that $v \neq 0$ and $v \in B_X$ and $v \in B_Y$. Also, $v \in X \cap Y$. However, $v = v + 0 = 0 + v$ where $v \in X$ and $0 \in Y$ in the first inequality and $0 \in X$ and $v \in Y$ in the second equality. However, this contradicts that unique decomposition in (2).

Now we show that $B_X \cup B_Y$ is a basis for V. Since $B_X \cap B_Y = \varnothing$ then $B_X \cup B_Y = \{u_1, \ldots, u_p, v_1, \ldots, v_r\}$. To show that $B_X \cup B_Y$ spans V, we see that for all $v \in V$, $v = x + y$, where $x \in X$ and $y \in Y$. Since B_X is a basis for X then $x = \alpha_1 u_1 + \cdots + \alpha_p u_p$ for some scalars α_i for $i = 1, \ldots, p$. Likewise, $y = \beta_1 v_1 + \cdots + \beta_r v_r$ for some scalars β_i, $i = 1, \ldots, r$. Then

$$\begin{aligned} v &= x + y \\ &= \alpha_1 u_1 + \cdots + \alpha_p u_p + \beta_1 v_1 + \cdots + \beta_r v_r. \end{aligned}$$

Therefore $v \in span(B_X \cup B_Y)$ and since $v \in V$ was arbitrary, this shows that $V = span(B_X \cup B_Y)$. This proves the spanning property. To show that $B_X \cup B_Y$ is linearly independent, let α_i and β_i be such that

$$\alpha_1 u_1 + \cdots + \alpha_p u_p + \beta_1 v_1 + \cdots + \beta_r v_r = 0$$

where $(\alpha_1 u_1 + \cdots + \alpha_p u_p) \in X$ and $(\beta_1 v_1 + \cdots + \beta_r v_r) \in Y$. By part (2) we can write the zero vector in V as the unique composition $0 = 0 + 0$ where the first zero is in X and the second zero is in Y. Therefore $\alpha_1 u_1 + \cdots + \alpha_p u_p = 0$ and $\beta_1 v_1 + \cdots + \beta_r v_r = 0$. Since B_X and B_Y are linearly independent, then $\alpha_i = 0$ for all i and $beta_j = 0$ for all j. Thus, $B_X \cup B_Y$ is linearly independent, which implies that $B_X \cup B_Y$ is a basis for V.

($3 \Rightarrow 1$) Assume $B_X \cap B_Y = \varnothing$ and $B_X \cup B_Y$ is a basis for V. First we show that $V = X + Y$. Since $B_X \cup B_Y$ is a basis for V, then for any $v \in V$, we can write v as linear combination of $u_1, \ldots, u_p, v_1, \ldots, v_r$. Therefore

$$v = (\alpha_1 v_1 + \cdots + \alpha_p u_p) + (\beta_1 v_1 + \cdots + \beta_r v_r)$$

where $(\alpha_1 v_1 + \cdots + \alpha_p u_p) \in X$ and $(\beta_1 v_1 + \cdots + \beta_r v_r) \in Y$. Thus, $v \in X + Y$ and since $v \in V$ was arbitrary, then $V = X + Y$.

To show that $X \cap Y = \{0\}$, let $z \in X \cap Y$. Then $z \in Z$ and $z \in Y$. Since $B_X = \{u_1, \ldots, u_p\}$ is a basis for X, then $z = \alpha_1 u_1 + \cdots + \alpha_p u_p$. Likewise, $z = \beta_1 v_1 + \cdots + \beta_r v_r$. Then

$$\alpha_1 u_1 + \cdots + \alpha_p u_p = \beta_1 v_1 + \cdots + \beta_r v_r$$

and subtracting all terms from the right hand side we have

$$\alpha_1 u_1 + \cdots + \alpha_p u_p - \beta_1 v_1 - \cdots - \beta_r v_r = 0.$$

Since $B_X \cup B_Y$ is basis for V, then $B_X \cup B_Y$ is linearly independent. Therefore, since the above expression is a linear combination of vectors from $B_X \cup B_Y$, then $\alpha_i = 0$ for all i and $-\beta_j = 0$ for j, which implies that $\beta_j = 0$ for all j. This implies that $z = \alpha_1 u_1 + \cdots + \alpha_p u_p = 0$ and so $X \cap Y = \{0\}$.

Thus, $V = X \oplus Y$.

□

Definition. Suppose $V = X \oplus Y$. Since for any $v \in V$ there exists a unique $x \in X$ and $y \in Y$ such that $v = x + y$, then $v \to x$ defines a function and $v \to y$ defines a function.

Here we call x the **projection of** $v \in V$ **onto** X (along Y) and we call y the **projection of** $v \in V$ **onto** Y (along X).

Definition. A linear operator $P : V \to V$ is a **projector** if there exists complementary subspaces X and Y of V, such that $Pv = X$ for all $v \in V$, where $v = x + y$ for $x \in X$ and $y \in Y$.

Remark. In the above definition, $v = x + y$ is a key decomposition that is used in solving problems.

Remark. Facts about a projector:

(1) For any $x \in X$, $P \cdot x = x$ because $x = x + 0$, where $x \in X$ and $0 \in Y$.

(2) for any $y \in Y$, $P \cdot y = 0$ because $y = 0 + y$, where $0 \in X$ and $y \in Y$.

(3) For any two linear operators R and Q, $R = Q$ if and only if $Rv = Qv$ for all $v \in V$. (if given any input, the output is the same).

Theorem 50.2. Let $P : V \to V$ be a projector matrix onto the subspace X along Y. Then

(1) $P^2 (= P \circ P) = P$ (idempotent property)

(2) $I - P$ is the projector onto Y along X.

(3) $R(P) = \{v \in V : Pv = v\} (= S) = X$.

Proof. (1) We want to show that $P^2v = Pv$ for all $v \in V$. Since P is a projector matrix, then X are complementary subspaces. By the above theorem, we can uniquely write any $v \in V$ as $v = x + y$, where $x \in X$ and $y \in Y$. Since $Pv = x \in X$, then $P^2v = P(Pv) = P \cdot x = x = Pv$. Therefore, since $P^2v = Pv$ for all $v \in V$, then $P^2 = P$.

(2) For any $v \in V$ we have the unique decomposition of v as $v = x + y$ where $x \in X$ and $y \in Y$, with $Pv = x$. Now

$$\begin{aligned}(I - P)v &= Iv - Pv \\ &= v - Pv \\ &= v - x \\ &= x + y - x \\ &= y.\end{aligned}$$

Therefore, for any $v \in V$, $(I - P)(v) = y$, which is the y part of the unique decomposition $v = x + y$. Thus, $I - P$ is the projector onto Y along X.

(3) First we show that $R(P) = S$. Clearly, for any $v \in S$, $Pv = v$ where $Pv \in R(P)$. This implies that $v \in R(P)$. Therefore, $S \subseteq R(P)$.

Conversely, let $z \in R(P)$. Then $z = P \cdot v$ for some $v \in V$ and so $Pz = P(Pv) = P^2(v)$. Now since $P^2 = P$ (idempotent property) then $P^2v = Pv = z$, which implies that $Pz = z$. Thus, $z \in S$. Since $S \subseteq R(P)$ and $R(P) \subseteq S$, then $R(P) = S$.

Now we show that $X = S$. First, $X \subseteq S$ because for any $x \in X$, $Px = x \in S$, which implies that $X \subseteq S$. To show that $S \subseteq X$, let $z \in S$. Then $Pz = z$. Since $Pv \in X$ for $v \in V$ then $Pz \in X$ and so $z \in X$, which implies that $S \subseteq X$. Since we have shown that $X \subseteq S$ and $S \subseteq X$, then $X = S$.

□

In fact, $N(P) = Y$, which we show in the homework.

51 Matrix representation of a projector P on $\mathbb{R}^n$

Let $V = \mathbb{R}^n$ and X, Y be complementary subspaces. Let P be the projector onto X along Y. Then $R(P) = X$ and $N(P) = Y$.

Let $B_X = \{x_1, \ldots, x_r\}$ and $B_Y = \{y_1, \ldots, y_{n-r}\}$. Since $B_X \cap B_Y = \varnothing$ and $B = B_X \cup B_Y = \{x_1, \ldots, x_r, y_1, \ldots, y_{n-r}\}$ is a basis for $\mathbb{R}^n$, then

$$[P]_B = \big[[Px_1]_B \quad [Px_r]_B \;\; [Py_1]_B \quad [By_{n-r}]_B\big]$$

where $Px_i = x_i$ ($x_i \in X$) for all $i = 1, \ldots, r$ and $Py_i = 0$ ($y_i \in Y$) for all $i = 1, \ldots, n - r$. Then

$$[P]_B = \begin{bmatrix} [x_1]_B & \cdots & [x_r]_B & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

where $[x_1]_B = e_1, \ldots, [x_r]_B = e_r$. Let $\tilde{B} = \{e_1, e_2, \ldots, e_n\}$ be the standard basis for $\mathbb{R}^n$. Then

$$[P]_{\tilde{B}} = [I]_{B\tilde{B}} \cdot [P]_B \cdot ([I]_{B\tilde{B}})^{-1}$$

where

$$\begin{aligned} [I]_{B\tilde{B}} &= \begin{bmatrix} [Ix_1]_{\tilde{B}} & \cdots & [Ix_r]_{\tilde{B}} & [Iy_1]_{\tilde{B}} & \cdots & [Iy_{n-r}]_{\tilde{B}} \end{bmatrix} \\ &= \begin{bmatrix} x_1 & \cdots & x_r & y_1 & \cdots & y_{n-r} \end{bmatrix}_{n \times n}. \end{aligned}$$

Therefore

$$\begin{aligned} [P]_{\tilde{B}} &= \begin{bmatrix} x_1 & \cdots & x_r & y_1 & \cdots & y_{n-r} \end{bmatrix} \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \left(\begin{bmatrix} x_1 & \cdots & x_r & y_1 & \cdots & y_{n-r} \end{bmatrix}^{-1} \right) \\ &= \begin{bmatrix} B_X & 0 \end{bmatrix} \begin{bmatrix} B_X & B_Y \end{bmatrix}^{-1}. \end{aligned}$$

Therefore, the matrix representation of the projector P is

$$[P]_{\tilde{B}} = \begin{bmatrix} B_X & 0 \end{bmatrix} \cdot \begin{bmatrix} B_X & B_Y \end{bmatrix}^{-1}.$$

Example. Let $V = \mathbb{R}^3$. Let $X = span \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$ and $Y = span \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$.

Since $B_X \cap B_Y = \varnothing$ and $B_X \cup B_Y$ is a basis for $\mathbb{R}^3$ then $V = X \oplus Y$.

To find the matrix representation of the projector P onto the X space,

$$\begin{aligned} [P]_{\tilde{B}} &= \begin{bmatrix} B_X & 0 \end{bmatrix} \cdot \begin{bmatrix} B_X & B_Y \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

Theorem 51.1. Let $P : V \to V$ be a linear operator. Then P is a projector if and only if $P^2 = P$, i.e., P, is idempotent.

Proof. ($\rightarrow$) If P is a projector then $P^2 = P$, from theorem above.

($\leftarrow$) Suppose $P^2 = P$. We will show that $V = R(P) \oplus N(P)$, i.e., $R(P)$ and $N(P)$ are complementary subspaces of V.

First we show that $V = R(P) + N(P)$. For any $v \in V$, $v = Pv + (v - Pv) = Pv + (I - P)v$. Clearly, $Pv \in R(P)$. Since $P(I-P)v = Pv - P^2v$ and $P^2 = P$ then $P(I-P)v = 0$ and so $(I-P)v \in N(P)$. Therefore, $v \in R(P) + N(P)$. Since $v \in V$ was arbitrary this shows that $V = R(P) + N(P)$.

Now we show that $R(P) \cap N(P) = \{0\}$. Let $z \in R(P) \cap N(P)$. Then $z \in R(P)$ and $z \in N(P)$, which means $z = Pv$ for some $v \in V$ and $Pz = 0$. Therefore $0 = P \cdot z = P(P \cdot v) = P^2 v$ and since P is idempotent then $Pv = z$. Thus, $z = 0$. Since z was arbitrary then this shows that $R(P) \cap N(P) = \{0\}$ and so $V = R(P) \oplus N(P)$.

Therefore, for any $v \in V$, v is uniquely decomposed into the sum

$$v = x + y$$

where $x \in R(P)$ and $y \in N(P)$. Therefore, P is a projector onto $R(P)$ along $N(P)$. $\square$

52 Orthogonal Complement

Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. This is a general assumption in this section.

Definition. A vector $v \in V$ is orthogonal to a non-empty set S in V if it is orthogonal to each vector in S, i.e., $v \perp z$ for all $z \in S$. In this case, $v \perp S$.

Proposition. *Let M be a subspace of V with a basis $B_M = \{u_1, \ldots, u_p\}$. Then a vector $v \perp M$ if and only if $v \perp u_i$ for all $i = 1, \ldots, p$.*

Proof. ($\Rightarrow$) Suppose $v \perp M$. Then $v \perp z$ for all $z \in M$. Since $u_i \in M$ for all $i = 1, \ldots, p$ then $v \perp u_i$ for all $i = 1, \ldots, p$.

($\Leftarrow$) Suppose $v \perp u_i$ for all $i = 1, \ldots, p$. Since B_M is a basis for M, then for any $z \in M$ we can write as

$$z = \alpha_1 u_1 + \cdots + \alpha_p u_p$$

for some scalars $\alpha_1, \ldots, \alpha_p$. Now consider the inner product of v and z, which is

$$\begin{aligned}
\langle v, z \rangle &= \langle v, \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_p u_p \rangle \\
&= \alpha_1 \langle v, u_1 \rangle + \cdots + \alpha_p \langle v, u_p \rangle \\
&= 0.
\end{aligned}$$

Therefore, $v \perp z$ for all $z \in M$ and so $v \perp M$. $\square$

Definition. Given a non-empty set S in V, the orthogonal complement of S, denoted by $S^{\perp}$, is the collection of all vectors in V that are orthogonal to S, i.e.,

$$S^{\perp} = \{v \in V : v \perp S\} = \{v \in V : v \perp z, \forall z \in S\}$$

Proposition. *For any non-empty set S, $S^{\perp}$ is a subspace.*

Proof. We proved in a homework exercise. $\square$

Theorem 52.1. Let V be a finite dimensional inner product space, and M be a subspace of V. Then

$$V = M \oplus M^{\perp}$$

i.e., M and its orthogonal complement $M^{\perp}$ are complementary subspaces of V.

Proof. First we show that $M \cap M^{\perp} = \{0\}$. Let $z \in M \cap M^{\perp}$. Then $z \in M^{\perp}$ and $z \in M$. Since $z \in M^{\perp}$ then $z \perp u$ for all $u \in M$. Then since $u \in M$, $z \perp z$ and so $\langle z, z\rangle = 0$. Thus, $z = 0$. Since z was arbitrary, this shows that $M \cap M^{\perp} = \{0\}$.

Now we show that $V = M + M^{\perp}$. V is a finite dimensional inner product space. Therefore M and $M^{\perp}$ are finite dimensional product spaces such that M has an orthonormal basis B_M, and $M^{\perp}$ has an orthonormal basis $B_{M^{\perp}}$. Since $B \subseteq M$, $B_{M^{\perp}} \subseteq M^{\perp}$, and $M \cap M^{\perp} = \{0\}$, then $B_M \cap B_{M^{\perp}} = \varnothing$ and $B_M \cup B_{M^{\perp}}$ is an orthonormal set in M. Therefore, $B_M \cup B_{M^{\perp}}$ is linearly independent, which implies that $B_M \cup B_{M^{\perp}}$ spans $M + M^{\perp}$. Therefore, $B_M \cup B_{M^{\perp}}$ is a basis for $M + M^{\perp}$.

Claim: $B_M \cup B_{M^{\perp}}$ spans V.

Suppose not. Then there exists a $u \in V$ such that $u \notin \mathit{span}(B_M \cup B_{M^{\perp}}) = M + M^{\perp}$. Therefore, by the Gram-Schmidt procedure,

$$w = u - \mathrm{proj}_{\mathbf{M}+\mathbf{M}^{\perp}}\, \mathbf{u} \neq 0 \text{ and } w \perp M + M^{\perp}$$

where $\mathrm{proj}_{\mathbf{M}+\mathbf{M}^{\perp}}\, \mathbf{u}$ is the orthogonal projection of u onto $M + M^{\perp}$. Note that $M \subseteq M + M^{\perp}$ (because we can write $v = v + 0$ where $v \in M$ and $0 \in M^{\perp}$.) Since $w \perp M + M^{\perp}$ then $w \perp M$, which implies that $w \in M^{\perp}$. On the other hand, $M^{\perp} \subseteq M + M^{\perp}$. Therefore, $w \in M^{\perp} \subseteq M + M^{\perp}$ and so $w \in M + M^{\perp}$. Therefore, $w \perp w$ which implies that $w = 0$. However, this contradict the fact above that $w \neq 0$. Therefore, $B_M \cup B_{M^{\perp}}$ spans V, i.e., $V = M + M^{\perp}$. Thus, $V = M \oplus M^{\perp}$. $\square$

Proposition. *Let V be a finite dimensional inner product space and M is a subspace of V. Then, $(M^{\perp})^{\perp} = M$.*

Proof. First we prove that $M \subseteq (M^{\perp})^{\perp}$. Let $v \in M$. Therefore, for any $v \in M^{\perp}$, $w \perp v$. Because $v \perp w$ for all $w \in M^{\perp}$, then $v \in (M^{\perp})^{\perp}$, which implies that $M \subseteq (M^{\perp})^{\perp}$.

Now we prove that $(M^{\perp})^{\perp} \subseteq M$. Let $v \in (M^{\perp})^{\perp} \subseteq V$. Since $V = M \oplus M^{\perp}$, then we can write v as $v = u + w$ where $u \in M$ and $w \in M^{\perp}$. Since $v \in (M^{\perp})^{\perp}$ and $w \in M^{\perp}$, then $v \perp w$ or equivalently, $\langle w, v\rangle = 0$. Since $v = u + w$, then $\langle w, v\rangle = \langle w, u + w\rangle = langlew, u\rangle + langlew, w\rangle$. Because $u \in M$ and $w \in M^{\perp}$, then $w \perp u$ or $langlew, u\rangle = 0$. Therefore, $0 = 0 + \langle w, w\rangle$, which implies that $\langle w, w\rangle = 0$ and so $w = 0$. Thus, $v = u \in M$. Since $v \in M$ then $(M^{\perp})^{\perp} \subseteq M$. $\square$

53 Orthogonal Complements of Fundamental Subspaces of a Matrix

Let A be an $m \times m$ real matrix, i.e., $A \in \mathbb{R}^{m \times m}$. Then

- $R(A)$ and $N(A^T)$ are subspaces of $\mathbb{R}^m$
- $R(A^T)$ and $N(A)$ are subspaces of $\mathbb{R}^n$

Proposition. $[R(A)]^{\perp} = N(A^T)$ *and* $[R(A^T)]^{\perp} = N(A)$.

Proof.

$$\begin{aligned} x \in [R(A)]^{\perp} &\iff x \perp z, \forall z \in R(A) \\ &\iff z^T x = 0, \forall z \in R(A) \\ &\iff (Ay)^T \cdot x = 0, \forall y \in \mathbb{R}^n \\ &\iff y^T \cdot (A^T \cdot x) = 0, \forall y \in \mathbb{R}^n \\ &\iff A^T x = 0 \\ &\iff x \in N(A^T). \end{aligned}$$

Therefore, $[R(A)]^T = N(A^T)$. □

Remark. It follows that $\mathbb{R}^m = R(A) \oplus N(A^T)$ and $\mathbb{R}^n = R(A^T) \oplus N(A)$.

54 Normal Matrix

Definition. An $n \times n$ complex matrix A is normal if $A^*A = AA^*$.

A special case is when $A \in \mathbb{R}^{n\times n}$, then A is normal $\iff A^T A - AA^T$.

Example. (1) A is real symmetric/skewed symmetric.

(2) a is hermitian/skewed hermitian.

Proposition. *Let $A \in \mathbb{R}^{n\times n}$ be a normal matrix. Then $[R(A)]^{\perp} = N(A)$ and $\mathbb{R}^n = R(A) \oplus N(A)$.*

Proof. We use the fact that $N(A^T) = N(A \cdot A^T)$ and $N(A) = N(A^T A)$. Note that $[R(A)]^T = N(A^T) = N(A \cdot A^T) = N(A^T \cdot A)$. (Since A is normal.) Then $N(A^T A) = N(A)$. Then $[R(A)]^{\perp} = N(A)$. Thus, $\mathbb{R}^n = R(A) \oplus N(A)$. □

55 Orthogonal Projection

Let V be a finite-dimensional inner product space. Suppose $V = M \oplus M^{\perp}$, where M is a subspace of V and $M^{\perp}$ is the orthogonal complement of M. Then for any $v \in V$, $v = m + n$, where m M, $n \in M^{\perp}$ (unique decomposition). This gives the linear operator $P_M : V \to V$, given by $P_M(v) = m \in M$. Here P_M is **orthogonal projector** onto M space (We showed this in HW#10, Question 4).

55.1 Orthogonal Projectors onto M

Let $B_M = \{u_1, \ldots, u_p\}$ be an orthonormal basis for M and $B_{M^\perp} = \{w_1, \ldots, w_r\}$ be an orthonormal basis for $M^\perp$. Then, $B_M \cap B_{M^\perp} = \varnothing$ and $B_M \cup B_{M^\perp}$ is an orthonormal basis for $V = M \oplus M^\perp$. Then let $B = B_M \cup B_{M^\perp} = \{u_1, \ldots, u_p, w_1, \ldots, w_r\}$. Then for all $v \in V$, $v = \langle v, u_1\rangle u_1 + \cdots + \langle v, u_p\rangle u_p + \langle v, w_1\rangle w_1 + \cdots + \langle v, w_r\rangle w_r$ where $\langle v, u_1\rangle u_1 + \cdots + \langle v, u_p\rangle u_p = u \in M$ and $\langle v, w_1\rangle w_1 + \cdots + \langle v, w_r\rangle w_r = w \in M^\perp$. Then the projection of v onto M is,

$$\operatorname{proj}_{\mathbf{M}} \mathbf{v} = u = \langle v, u_1\rangle u_1 + \cdots + \langle v, u_p\rangle u_p.$$

55.2 Matrix Representation of an Orthogonal Projector P_M on $V = \mathbb{R}^n$

Let $B_M = \{u_1, \ldots, u_p\}$ be an arbitrary basis for M, and $B_{M^\perp} = \{w_1, \ldots, w_{n-p}\}$ be an orthonormal basis for $M^\perp$. Then let the matrices S and Q be

$$S = \begin{bmatrix} u_1, \ldots, u_p \end{bmatrix} \text{ and } Q = \begin{bmatrix} w_1, \ldots, w_{n-p} \end{bmatrix}.$$

Facts about S and Q:

(1) The columns of S and Q are linearly independent. Then $Sx = 0$ implies that $x = 0$. Therefore, $S^\perp S$ is invertible because

$$S^\perp S x = 0 \implies x^T S^T S x = 0 \implies (Sx)^T (Sx) = 0 \implies Sx = 0 \implies x = 0$$

(2) $Q^T Q = I_{n-p}$.

(3)
$$S^T Q = \begin{bmatrix} u_1^T \\ \vdots \\ u_p^T \end{bmatrix} \begin{bmatrix} w_1, \ldots, w_{n-p} \end{bmatrix} = \begin{bmatrix} u_1^T w_1 & \cdots & u_1^T w_{n-p} \\ \vdots & & \vdots \\ u_p^T w_1 & \cdots & u_p^T w_{n-p} \end{bmatrix} = 0.$$

Since $\langle u_i, w_i\rangle = 0$ where $u_i \in M$ and $w_j \in M^\perp$.

Similarly, $Q^T S = 0$ because $Q^T S = (S^T Q)^T$.

Proposition. *The matrix representation of P_M given by $W = S(S^T \cdot S)^{-1} S^T$, where $W \in \mathbb{R}^{n\times n}$ and $S = \begin{bmatrix} u_1, \ldots, u_p \end{bmatrix} \in \mathbb{R}^{n\times p}$,*

Proof. Recall that $W = \begin{bmatrix} S & 0 \end{bmatrix} \cdot \begin{bmatrix} S & Q \end{bmatrix}^{-1}$, by results for general projection.

We claim that

$$\begin{bmatrix} S & Q \end{bmatrix}^{-1} = \begin{bmatrix} (S^T S)^{-1} S^T \\ Q^T \end{bmatrix}.$$

Since

$$\begin{bmatrix} (S^TS)^{-1}S^T \\ Q^T \end{bmatrix} \begin{bmatrix} S & Q \end{bmatrix} = \begin{bmatrix} (S^TS)^{-1}S^TS & (S^TS)^{-1}S^TQ \\ Q^TS & Q^TQ \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ 0 & I_{n-p} \end{bmatrix}$$

Thus,

$$\begin{bmatrix} S & Q \end{bmatrix}^{-1} = \begin{bmatrix} (S^TS)^{-1}S^T \\ Q^T \end{bmatrix}.$$

Then,

$$W = \begin{bmatrix} S & 0 \end{bmatrix} \begin{bmatrix} (S^TS)^{-1}S^T \\ Q^T \end{bmatrix} = S(S^TS)^{-1}S^T.$$

□

Remark.

(1) The matrix representation of P_M is independent of the basis for M.

(2) $W = W^T$ (i.e., W is symmetric).

To see this, note that

$$\begin{aligned} W^T = \left[S(S^TS)^{-1}S^T\right] &= (S^T)^T \cdot \left[(S^TS)^{-1}\right]^T \cdot S^T \\ &= S \cdot \left[(S^TS)^T\right]^{-1} \cdot S^T \\ &= S \cdot \left[S^TS\right]^{-1} S^T \\ &= W \end{aligned}$$

Therefore, $W^T = W$ and so W is symmetric.

Theorem 55.1. Suppose $P : \mathbb{R}^n \to \mathbb{R}^n$ is a projector ($P^2 = P$). Let W be its matrix representation. Then P is an orthogonal projector if and only if $W = W^T$.

Proof. ($\Rightarrow$) We already covered.

($\Leftarrow$) Suppose W is symmetric, i.e., $W = W^T$. Since P is a projector and W is the matrix representation, then $\mathbb{R}^n = R(P) \oplus N(P)$ and $R(P) = R(W)$, $N(P) = R(W)$. To show that P is orthogonal projector, we want to show $R(P) \perp N(P)$ (or equivalently $R(W) \perp N(W)$). Recall that $\left[R(W)\right]^\perp = N(W^T)$. Since W is symmetric then $W^T = W$ and so $\left[R(W)\right]^\perp = N(W)$. Then $\mathbb{R}^n = R(W) \oplus N(W)$ and $R(W) \perp N(W)$. Thus P is an orthogonal projector. □

Theorem 55.2. Let P_M be an orthogonal projection onto the subspace M. Then, for all $z \in V$, $P_M(z) \in M$ is the closest point in M to z, i.e., for all $u \in M$, $||z - \text{proj}_{\mathbf{M}}\,\mathbf{z}|| \leq ||u - z||$.

Proof. We will use the facts:

(1) IF $x \perp y$, then $||x+y||^2 = ||x||^2 + ||y||^2$.

(2) $(z - \operatorname{proj}_{\mathbf{M}} \mathbf{z}) \perp M$, which implies $z - \operatorname{proj}_{\mathbf{M}} \mathbf{z} \in M^{\perp}$.

For any $u \in M$, $||u - z||^2 = ||u - P_M z + P_M z - z||^2$ where $u - P_M z \in M$ and $P_M z - z \in M^{\perp}$.

Thus, $||u-z||^2 = ||u - P_m z||^2 + ||P_M z - z||^2$. Therefore, $||u - \operatorname{proj}_{\mathbf{M}} \mathbf{z}||^2 > 0$. Then $||u = z||^2 \geq ||\operatorname{proj}_{\mathbf{M}} \mathbf{z}||^2$ implies $||u - z|| \geq ||\operatorname{proj}_{\mathbf{M}} \mathbf{z} - z|| = ||z - \operatorname{proj}_{\mathbf{M}} \mathbf{z}||$. This implies that $||u - z|| \geq ||z - \operatorname{proj}_{\mathbf{M}} \mathbf{z}||$ for all $u \in M$. Thus, $P_M(z)$ is the closest point in M to z. □

56 Determinant

Definition. Let A be an $n \times n$ matrix, i.e., $A \in \mathbb{R}^{n \times n}$ or $A \in \mathbb{C}^{n \times n}$. Then the determinant of A, using co-factor expansion is an inductive process on n. When $n = 1$ then A is a scalar and $\det(A) = A$. Now consider the general case. Then $A = [a_{ij}]$ where a_{ij} is the (i,j) element of A. In this case, the cofactor expansion across the i^{th} row is

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} \cdot a_{ij} \cdot \det A_{ij}$$

where A_{ij} is the submatrix of A after removing the i^{th} row and the j^{th} column of A. Therefore, A_{ij} is an $(n-1) \times (n-1)$ matrix. By the inductive hypothesis, $\det A_{ij}$ is known. Similarly, co-factor expansion down the j^{th} column is

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} \cdot a_{ij} \cdot \det A_{ij}$$

where j is fixed and $i \in \{1, \ldots, n\}$.

Example.

(1) Let $n = 2$ and let $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$. Using co-factor expansion across the first row we get,

$$\det A = (-1)^{1+1} \cdot a_{11} \cdot \det(A_{11}) + (-1)^{1+2} \cdot a_{12} \cdot \det A_{12}$$

where $A_{11} = [a_{22}]$ and $A_{12} = [a_{21}]$. Then $\det A = a_{11} \cdot a_{22} - a_{12} \cdot a_{21}$.

In this case, $\det A$ has 2 terms, each of which is a product of 2 elements of A.

(2) Let $n = 3$ and $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$. Using co-factor expansion across the first row we have

$$
\begin{aligned}
\det(A) &= a_{11} \cdot \det A_{11} - a_{12} \cdot \det A_{12} + a_{13} \cdot \det A_{13} \\
&= a_{11} \cdot \det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{12} \cdot \det \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix} + a_{13} \cdot \det \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \\
&= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{23} - a_{31}a_{23}) + a_{13}(a_{21}a_{32} - a_{31}a_{22}).
\end{aligned}
$$

In this case, $\det(A)$ has 6 terms, each of which is a product of 3 elements of A.

Remark. For an $n \times n$ matrix, $\det A$ contains $n!$ terms, each of which is a product of n elements of A.

Example. Let A be an $n \times n$ lower triangular matrix. Then

$$
\begin{aligned}
\det A &= \begin{bmatrix} a_{11} & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ \vdots & & \ddots & 0 \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \\
&= a_{11} \cdot \det A_{11} \\
&= a_{11} \cdot \det \begin{bmatrix} a_{22} & 0 & 0 & 0 \\ a_{32} & a_{33} & 0 & 0 \\ \vdots & & \ddots & 0 \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \\
&= a_{11}a_{22} \cdots a_{nn}.
\end{aligned}
$$

Similarly, the determinant of an upper triangular matrix si the product of the diagonal entries.

57 Properties of Determinants

(1) The determinant is linear in each row or column when the other column/row is fixed.

(2) Determinant under elementary row operations. There are three elementary row operations:

   (1) row scaling: If we multiply a row of A by a scalar α to get B, then the $\alpha \cdot \det A = \det B$.

   (2) If we switch a row in A to get B then $\det B = -\det A$.

   (3) If we replace a row in A to get B then $\det B = \det A$.

(3)

Theorem 57.1. A square matrix A is invertible if and only if $\det A \neq 0$.

Proof. We will use the following facts in the proof:

(a) An echelon form of A is an upper triangular matrix and each pivot is a diagonal entry of the echelon form.

(b) A is invertible if and only if each column/row has a pivot.

Let U be an echelon form of A. Then A is invertible if and only if $\det U \neq 0$. To get U, one only needs row switching and row replacement,

$$A \sim U_1 \sim U_2 \sim \cdots \sim U$$

where each $\sim$ represents a row switch or row replacement. Therefore,

$$\det U = (-1)^r \cdot \det A$$

where r is the number of row switchings used in this process. Then

$$\det U \neq 0 \iff \det A \neq 0.$$

Thus, A is invertible if and only if $\det A \neq 0$. $\square$

(4) $\det(A^T) = \det A$

(5) $\det(A \cdot B) = \det A \cdot \det B$

(6) $\det \begin{bmatrix} A & B \\ 0 & D \end{bmatrix} = \det A \cdot \det B$

(7) Suppose A is invertible. Then $\det(A^{-1}) = \dfrac{1}{\det A}$.

Proof. Because $A^{-1} \cdot A = I$ then $\det(A^{-1} \cdot A) = \det I = 1$. Therefore, $\det A \cdot \det A^{-1} = 1$, which implies $\det(A^{-1}) = \dfrac{1}{\det A}$. $\square$

(8) Suppose A and B are similar. Then $\det A = \det B$.

Proof. We did in homework. $\square$

58 Eigenvalues and Eigenvectors

Let A be an $n \times n$ real/complex matrix.

Definition. A scalar λ (which is possibly complex number) is an **eigenvalue** of A if and only if there is a nonzero vector v such that

$$Av = \lambda \cdot v$$

where v is an **eigenvector** of A associated with the eigenvalue λ.

Here (λ, v) is called an eigen-pair.

Definition. The set of all distinct eigenvalues of A is the **spectrum** of A, denoted by $\sigma(A)$.

To determine all of the eignevalues of A:

$$\begin{aligned}
\lambda \text{ is an eigenvalue of } A &\iff Av = \lambda v \text{ for some } v \neq 0 \\
&\iff Av - \lambda v = 0 \text{ for some } v \neq 0 \\
&\iff (A - \lambda I)v = 0 \text{ for some } v \neq 0 \\
&\iff A - \lambda I \text{ is singular or nonivertible} \\
&\iff \det(A - \lambda I) = 0.
\end{aligned}$$

Therefore, λ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$.

Let $p(\lambda) = \det(A - \lambda I)$, where $\lambda \in \mathbb{C}$ is a scalar variable.

It can be shown that $p(\lambda)$ is a polynomial function of degree n and $p(\lambda)$ is the **characteristic polynomial** of A. Also,

$$p(\lambda) = 0$$

is the characteristic equation of A.

This equation has n (possibly repeated) roots/solutions. Therefore, A has at most n distinct eigenvalues.

Remark. (a) Even if A is a real matrix, its eigenvalues maybe complex. As an example, consider the matrix $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. This matrix is a skew symmetric matrix.

Then $p(\lambda) = \det(A - \lambda I) = 0$. Because

$$\begin{aligned}
\det(A - \lambda I) &= \det \begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix} \\
&= (-\lambda)^2 + 1 \\
&= \lambda^2 + 1 \\
&= p(\lambda)
\end{aligned}$$

then

$$\begin{aligned} p(\lambda) &\iff \lambda^2 + 1 = 0 \\ &\iff \lambda = \pm\sqrt{-1} \end{aligned}$$

where $i = \sqrt{-1}$.

(b) Let A be an $n \times n$ hermitian matrix; i.e., $A^* = A$ (or $A^T = A$). Then any eigenvalue of A is a real number.

Proof. Let λ be an eigenvalue of A. Then there exists a vector $v \neq 0$ such that $Av = \lambda v$. Subtracting λv from both sides and taking the conjugate transpose then we have $(Av)^* = (\lambda v)^*$. By properties of the conjugate transpose, then $v^* \cdot A^* = \bar{\lambda} \cdot v^*$. Since A is hermitian, then $A^* = A$. Therefore $v^* A = \bar{\lambda} \cdot v^*$. Right multiplying both sides by v we have $v^*(Av) = \bar{\lambda} \cdot v^* \cdot v$ and then because λ is an eigenvalue of A we can replace Av with λv, $v^*(\lambda \cdot v) = \bar{\lambda} \cdot v^* \cdot v$. Then moving the scalar λ to the left side we have $\lambda \cdot (v^* \cdot v) = \bar{\lambda}(v^* \cdot v)$. Because v is an eigenvector we know $v \neq 0$ and so $v^* \cdot v = \langle v, v \rangle > 0$. Therefore, it must be the case that $\lambda = \bar{\lambda}$. Now let $\lambda = a + bi$. Then $\lambda = \bar{\lambda}$ implies that $a + bi = a - bi$ and so $bi = -bi$, and therefore, $b = -b$. Since $b \in \mathbb{R}$, then this means that $b = 0$. Thus, $\lambda = a \in \mathbb{R}$. □

Corollary. If A is a real symmetric matrix, i.e., $A = A^T \in \mathbb{R}^{n \times n}$, then each eigenvalue of A is real.

(c) If A is skew Hermitian, i.e., $A^* = -A$, then each eignevalue of A is imaginary.

Proof. Use similar technique as above proof. □

(d) A is singular (i.e., non-invertible) if and only if A has a zero eigenvalue (i.e., $\lambda = 0$ is an eigenvalue of A).

Proof.

Suppose $\lambda = 0$ is an eigenvalue of A and so $\begin{aligned} \det(A - \lambda I) = 0 &\iff \det A = 0 \\ &\iff A \text{ is singular} \end{aligned}$.

□

Definition. Let $p(\lambda)$ be the characteristic polynomial of an $n \times n$ matrix A and let $\sigma(A) = \{\lambda_1, \ldots, \lambda_S\}$ be the distinct eigenvalues of A. Therefore,

$$p(\lambda) = (-1)^n \cdot (\lambda - \lambda_1)^{a_1} \cdot (\lambda - \lambda_2)^{a_2} \cdots (\lambda - \lambda_S)^{a_S}$$

where a_i, $i = 1, \ldots, S$, is the number of times that λ_i is repeated. Therefore, $1 \leq a_i \leq n$ and $a_1 + a_2 + \cdots + a_S = n$. Each a_i is the **algebraic multiplicity** of λ_i.

To determine the eigenvectors associated with an eigenvalue λ and matrix A.

Let $\lambda \in \mathbb{C}$ be an eigenvalue of an $n \times n$ matrix $A \iff \exists v \neq 0$ such that $(A - \lambda I)v = 0$
$\iff \exists v \neq 0$ such that $v \in N(A - \lambda I)$.

Then $v \neq 0$ is an eigenvector associated with λ if and only if $v \neq 0$ and $v \in N(A - \lambda I)$.

Here $N(A - \lambda I)$ is called the **eigenspace** associated with λ, and $N(A - \lambda I)$ is a non-trivial subspace for any eigenvalue λ of A.

Remark.

(1) $dim(N(A - \lambda I)) \geq 1$ for all $\lambda \in \sigma(A)$ where $\sigma(A)$ is the spectrum of A, i.e., the set of all distinct eigenvalues of A.

(2) $dim(N(A - \lambda_i)) \leq a_i$, where a_i is the algebraic multiplicity of λ_i. Therefore,

$$\sum_{i=1}^{S} dim(N(A - \lambda_i I)) \leq \sum_{i=1}^{S} a_i = n$$

where S is the total number of distinct eigenvalues of A.

59 Diagonal Matrix and Diagonalization

Recall that an $n \times n$ matrix D is diagonal if

$$D = \begin{bmatrix} d_{11} & 0 & 0 & 0 \\ 0 & d_{22} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & d_{nn} \end{bmatrix}.$$

Some useful properties of diagonal matrices:

(1)

$$D^2 = D \cdot D = \begin{bmatrix} d_{11}^2 & 0 & 0 & 0 \\ 0 & d_{22}^2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & d_{nn}^2 \end{bmatrix}$$

and

$$D^k = D \cdot D \cdots D = \begin{bmatrix} d_{11}^k & 0 & 0 & 0 \\ 0 & d_{22}^k & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & d_{nn}^k \end{bmatrix}$$

for all $k = 1, 2, \ldots$.

(2) $D \cdot \tilde{D} = \tilde{D} \cdot D$ where $D, \tilde{D}$ are diagonal.

Definition. An $n \times n$ matrix A is called **diagonalizable** if A is similar to a diagonal matrix D, i.e., there exists an invertible matrix P such that $A = P \cdot D \cdot P^{-1}$.

Theorem 59.1. An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors. Further, the diagonal entries of the diagonal matrix D are the eigenvalues of A.

Proof. ($\Leftarrow$) Suppose A has n linearly independent eigenvectors, $p_1, p_2, \ldots, p_n$. Since each p_i is an eigenvector associated with an eigenvalue λ_i, then $Ap_i = \lambda_i \cdot p_i$.

Let the matrix $P = \begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix}$ and because $\{p_1, p_2, \ldots, p_n\}$ are linearly independent, then P is invertible. Further,

$$\begin{aligned} A \cdot P &= A \cdot \begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix} \\ &= \begin{bmatrix} Ap_1 & Ap_2 & \cdots & Ap_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 p_1 & \lambda_2 p_2 & \cdots & \lambda_n p_n \end{bmatrix} \\ &= \begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_n \end{bmatrix} \\ &= P \cdot D. \end{aligned}$$

Therefore, $A \cdot P = P \cdot D$ where D is diagonal and P is invertible. Also, $A = P \cdot D \cdot P^{-1}$ and so A is similar to D and therefore A is diagonalizable.

($\Rightarrow$) Suppose A is diagonalizable. Then $A = P \cdot D \cdot P^{-1}$ where

$$D = \begin{bmatrix} d_{11} & 0 & 0 & 0 \\ 0 & d_{22} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & d_{nn} \end{bmatrix}$$

and $P = \begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix}$, which is invertible. Therefore $\{p_1, p_2, \ldots, p_n\}$ is linearly independent.

Also, $A \cdot P = P \cdot D$. Therefore,

$$A \begin{bmatrix} p_1 & p_2 & \cdots p_n \end{bmatrix} = \begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix} \begin{bmatrix} d_{11} & 0 & 0 & 0 \\ 0 & d_{22} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & d_{nn} \end{bmatrix}$$

and so

$$\begin{bmatrix} Ap_1 & Ap_2 & \cdots & Ap_n \end{bmatrix} = \begin{bmatrix} d_{11} p_1 & d_{22} p_2 & \cdots & d_{nn} p_n \end{bmatrix}$$

because if two matrices are the same then the columns are the same.

Therefore, $Ap_i = d_{ii} \cdot p_i$ for all $i = 1, \ldots, n$. Note that p_i cannot be zero because $\{p_1, p_2, \ldots, p_n\}$ is linearly independent.

Therefore each d_{ii} is an eigenvalue of A, and p_i is the eigenvector associated with d_{ii}. Thus, A has n linearly independent eigenvectors. □

60 Eigenbasis and Diagonalizable Matrices

Definition. Given an $n \times n$ matrix A, a basis $B = \{v_1, \ldots, v_n\}$ is an **eigenbasis** for $\mathbb{R}^n$ or $\mathbb{C}^n$ if each v_i is an eigenvector.

Corollary. A is diagonlizable if and only if A has an eigenbasis.

Proposition. *Let $S = \{v_1, \ldots, v_p\}$ of eigenvectors of A, where each v_i is associated with a "distinct" eigenvalue λ_i. Then S is linearly independent.*

Proof. First consider the case where $p > 1$. For sake of contradiction suppose S is linearly independent (where $p \geq 2$). Let $\{v_1, \ldots, v_r\}$ be a minimal spanning set of $span(S)$, where $r < p$.

Therefore, $\{v_1, \ldots, v_r\}$ is linearly independent and v_{r+1} is a linear combination of $v_1, \ldots, v_r$, i.e.,

$$v_{r+1} = \alpha_1 v_1 + \cdots + \alpha_r v_r$$

for some scalars $\alpha_1, \ldots, \alpha_r$. Then multiplying both sides by A we have

$$Av_{r+1} = \alpha_1 Av_1 + \cdots + \alpha_r Av_r.$$

Then we can rewrite the v_{r+1} as

$$\lambda_{r+1} v_{r+1} = \alpha_1 \lambda_1 v_1 + \cdots \alpha_r \lambda_r v_r.$$

Then plugging in $v_{r+1} = \alpha_1 v_1 + \cdots + \alpha_r v_r$ in the above equation, we have

$$\lambda_{r+1} \cdot (\alpha_1 v_1 + \cdots + \alpha_r v_r) = \alpha_1 \lambda_1 v_1 + \cdots + \alpha_n \lambda_n v_n$$

and multiplying through on the right hand side by λ_{r+1} we have

$$\alpha_1 \lambda_{r+1} \cdot v_1 + \cdots + \alpha_r \lambda_{r+1} v_r = \alpha_1 \lambda_1 v_1 + \cdots + \alpha_n \lambda_n v_n$$

and moving all terms on the right to left side of the equation and we get

$$\alpha_1(\lambda - \lambda_{r+1}) \cdot v_1 + \cdots + \alpha_r(\lambda_r - \lambda_{r+1}) v_r = 0.$$

Since $\{v_1, \ldots, v_r\}$ is linearly independent then $\alpha_1(\lambda_1 - \lambda_{r+1}) = 0, \ldots, \alpha_r(\lambda_r - \lambda_{r+1}) = 0$.

Then since $\lambda_i \neq \lambda_j$ for $i \neq j$ then $\alpha_1 = 0, \alpha_2 = 0, \ldots, \alpha_r = 0$.

This shows that $v_{r+1} = \alpha_1 v_1 + \cdots + \alpha_r v_r = 0$, which is a contradiction because v_{r+1} is an eigenvector and cannot be equal to 0.

Thus, S is linearly independent. □

Remark. The following is an implication of the above proposition:

Suppose an $n \times n$ matrix A has n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$. In this case, each λ_j has its eigenvector v_i, where $v_i \neq 0$.

Then by the preceeding proposition, $\{v_1, v_2, \ldots, v_n\}$ is linearly independent. Then by a preceding theorem for necessary and sufficient condition for diagonalizability, A is diagonalizable.

Remark. Recall that $\sigma(A) = \{\lambda_1, \ldots, \lambda_S\}$, i.e., the set of distinct eigenvalues of A.

For each $\lambda_i \in \sigma(A)$ its eigenspace is $N(A - \lambda_i I)$. Let B_i be a basis for the eigenspace $N(A - \lambda_i I)$.

Then for any $\lambda_i, \lambda_j \in \sigma(A) = \{\lambda_1, \ldots, \lambda_S\}$ with $i \neq j$, $B_i \cap B_j = \varnothing$.

Proposition. *In the above setting, where $\sigma(A) = \{\lambda_1, \ldots, \lambda_S\}$, $B = B_1 \cup B_2 \cup \cdots \cup B_S$ is linearly independent.*

Proof. Let $G_k = B_1 \cup B_2 \cup \cdots \cup B_k$, for $k = 1, \ldots, S$. We use induction on k.

For the base case, assume $k = 1$. In this case, $G_1 = B_1$. Since B_1 is a basis for $N(A - \lambda_1 I)$ then B_1 is linearly independent. Thus, G_1 is linearly independent.

Now suppose that G_k is linearly independent and consider G_{k+1}. Note that $G_{k+1} = G_k \cup B_{k+1}$ where $G_k = \{v_1, \ldots, v_p\}$ and $B_{k+1} = \{u_1, \ldots, u_r\}$ is a basis for $N(A - \lambda_{k+1} I)$. Note that the v_i in G_k are eigenvectors associated with other eigenvalues.

Suppose scalars $\alpha_1, \ldots, \alpha_r$ and $\beta_1, \ldots, \beta_p$ are such that

$$\alpha_1 u_1 + \cdots + \alpha_r u_r + \beta_1 v_1 + \cdots + \beta_p v_p = 0.$$

We want to show that $\alpha_1 = \cdots = \alpha_r = \beta_1 = \cdots = \beta_p = 0$.

From the above equation we have

$$(A - \lambda_{k+1} I)(\sum_{i=1}^{r} \alpha_i u_i + \sum_{j=1}^{p} \beta_j v_j) = 0.$$

Distributing the $(A - \lambda_{k+1} I)$ term we get,

$$(A - \lambda_{k+1} I) \sum_{i=1}^{r} \alpha_i u_i + \sum_{j=1}^{p} \beta_j (A - \lambda_{k+1} I) \cdot v_j = 0.$$

Since each $u_i \in N(A - \lambda_{k+1} I)$ then $\sum_{i=1}^{r} \alpha_i u_i \in N(A - \lambda_{k+1} I)$ because the null space is a subspace. Therefore $(A - \lambda_{k+1} I) \sum_{i=1}^{r} \alpha_i u_i = 0$. Further,

$$\begin{aligned}
(A - \lambda_{k+1} I) v_j &= A v_j - \lambda_{k+1} v_j \\
&= A q_j v_j - \lambda_{k+1} v_j \\
&= (\lambda_{q_j} - \lambda_{k+1}) v_j
\end{aligned}$$

where $(\lambda_{q_j} - \lambda_{k+1}) \neq 0$ because $\lambda_{q_j} \neq \lambda_{k+1}$.

Then since $\sum_{j=1}^{p} \beta_j (\lambda_{q_j} - \lambda_{k+1}) v_j = 0$ then $G_k = \{v_1, \ldots, v_k\}$ are linearly independent. Therefore $B_j \cdot (\lambda_{q_j} - \lambda_{k+1}) = 0$ for all $j = 1, \ldots, p$. This implies that $\beta_j = 0$ for $j = 1, \ldots, p$.

This shows that $\alpha_1 u_1 + \cdots + \alpha_r u_r = 0$ and therefore $B_{k+1} = \{u_1, \ldots, u_r\}$ is a basis for $N(A - \lambda_{k+1} I)$ and so $\{u_1, \ldots, u_r\}$ is linearly independent. Thus, $\alpha_1 = \cdots = \alpha_r = 0$ and so $G_{k+1} = G_k \cup B_{k+1}$ is linearly independent. By induction principle, each G_k is linearly independent. Thus G_S is linearly independent. □

Theorem 60.1. An $n \times n$ matrix A is diagonalizable if and only if for each $\lambda_i \in \sigma(A)$, its geometric multiplicity, i.e., $dim(N(A - \lambda_i I))$, equals its algebraic multiplicity.

Proof. ($\Leftarrow$) Suppose $dim[N(A - \lambda_j I)] = a_i$ for all $\lambda_i \in \sigma(A)$ where $\{\lambda_1, \ldots, \lambda_S\}$. Since $\sum_{i=1}^{S} a_i = n$, then $\sum_{i=1}^{S} [N(A - \lambda_j I)] = n$.

Let B_i be a basis for each eigenspace $N(A - \lambda_i I)$, for all $\lambda_i \in \sigma(A)$. Therefore, $\#(B_i) = dim[N(A - \lambda_i I)]$ for all $\lambda_i \in \sigma(A)$, where $\#(\cdot)$ is number of vectors in a finite set. Therefore, $\sum^{S} i = 1\#(B_i) = n$.

Recall that $B_i \cap B_j = \varnothing$ for all $i \neq j$. Therefore, $\#(B_1 \cup B_2 \cup \cdots \cup B_S) = n$, where $B = B_1 \cup B_2 \cup \cdots \cup B_S$.

Also, from the preceding proposition, we know $B = B_1 \cup B_2 \cup \cdots B_S$ is linearly independent. Therefore, B is a set of n linearly independent eigenvectors and so A is diagonalizable.

($\Rightarrow$) Suppose A is diagonalizable. Then there exists an invertible matrix P such that $A = PDP^{-1}$ where D is a diagonal matrix. Without loss of generality, D can be written as

$$D = \begin{bmatrix} \lambda_1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & \lambda_1 & & & & & & \\ & & & \lambda_2 & & & & & \\ & & & & \ddots & & & & \\ & & & & & \lambda_2 & & & \\ & & & & & & \ddots & & \\ & & & & & & & \lambda_S & \\ & & & & & & & & \ddots & \\ & & & & & & & & & \lambda_S \end{bmatrix}.$$

Here a_1 is the algebraic multiplicity and $a_1 + a_2 + \cdots + a_S = n$.

Consider the eigenvalue λ_1. Then

$$\begin{aligned}
A - \lambda_1 I &= P \cdot D \cdot P^{-1} - P(\lambda_1 I) \cdot P^{-1} \\
&= P \cdot (D - \lambda_1 I) P^{-1} \\
&= P \cdot \begin{bmatrix} 0 & & & & \\ & \ddots & & & \\ & & 0 & & \\ & & & \ddots & \\ & & & & \ddots \end{bmatrix} \cdot P^{-1}
\end{aligned}$$

where all the first a_1 eigenvalues are set to zero. Then the number of pivots in previous matrix (all a_1 entries are **not** pivots but all others are). Therefore,

$$rank(A - \lambda_1 I) = n - a_1.$$

Since $A - \lambda_1 I$ is similar to $D - \lambda_1 I$, then

$$\begin{aligned}
rank(A - \lambda_1 I) &= rank(D - \lambda_1 I) \\
&= n - a_1.
\end{aligned}$$

From the rank-nullity theorem, we have

$$\begin{aligned}
dim[N(A - \lambda I)] &= n - rank(A - \lambda_1 I) \\
&= n - (n - a_1) \\
&= a_1.
\end{aligned}$$

Therefore, $dim[N(A - \lambda_1 I)] = a_1$.
Likewise, for each $\lambda_2, \ldots, \lambda_S \in \sigma(A)$ then $dim[N(A - \lambda_i I)] = a_i$ for all $i = 2, \ldots, S$.
□

Remark. The above proof shows that A is diagonalizable if and only if

$$B = B_1 \cup B_2 \cup \cdots \cup B_S$$

has n linearly independent eigenvectors

$$\begin{aligned}
&\iff B \text{ is an eigenbasis for } \mathbb{R}^n \text{ or } \mathbb{C}^n \\
&\iff \mathbb{R}^n (\text{ or } \mathbb{C}^n) = N(A - \lambda_1 I) \oplus \cdots \oplus N(A - \lambda_S I)
\end{aligned}$$

Then any vector can be written as a unique combination of vectors from

$$N(A - \lambda_1 I) \oplus N(A - \lambda_2 I) \oplus \cdots \oplus N(A - \lambda_S I).$$

Example.

(1) Let $A = \begin{bmatrix} 4 & 0 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}_{3\times 3}$

where A is lower triangular. Therefore, $A - \lambda I$ is also lower triangular. Then $\det(A - \lambda I) = (4 - \lambda)(3 - \lambda)^2$. Then the two distinct eigenvalues are $\lambda_1 = 4$ with algebraic multiplicity $a_1 = 1$ and $\lambda_2 = 3$ with algebraic multiplicity $a_2 = 2$.

Now we find the geometric multiplicity of both eigenvalues. Consider $\lambda_1 = 4$. Then

$$A - \lambda_1 I = \begin{bmatrix} 0 & 0 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

and so the $rank(A - \lambda I) = 2$. Therefore, $dim[N(A - \lambda_1 I)] = 3 - 2 = 1 = a_1$.

Now consider $\lambda_2 = 3$. Then

$$A - \lambda_2 I = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and so $rank(A - \lambda_2 I) = 1$. Therefore, $dim[N(A - \lambda_2 I)] = 3 - 1 = 2 = a_2$.

Therefore, A is diagonalizable.

(2) Now consider

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

which is an upper triangular matrix. Therefore, A has two distinct eigenvalues $\lambda_1 = 4$ with algebraic multiplicity $a_1 = 1$ and $\lambda_2 = 3$ with algebraic multiplicity $a_2 = 2$.

For $\lambda_1 = 4$, $dim[(N - \lambda_1 I)] = 1 = a_1$.

For $\lambda_2 = 3$

$$A - \lambda_2 I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Therefore, $rank(A - \lambda_2 I) = 2$ and so $dim[N(A - \lambda_2 I)] = 3 - 2 = 1 < a_2 = 2$.

Thus, A is **not** diagonalizable.

61 Unitary Diagonalization and Spectral Theorem

61.1 Facts about normal matrices

An $n \times n$ real/complex matrix A is **normal** if and only if $A \cdot A^* = A^* \cdot A$ (or $AA^T = A^T A$).

Facts about normal matrices:

(1) $N(A^*) = N(AA^*) = N(A^*A) = N(A)$.

(2) $R(A) \perp N(A)$, $\Rightarrow$ A represents an orthogonal projector. (if $A^2 = A$).

(3) Let λ_i, λ_j be two distinct eigenvalues of A. Then $N(A - \lambda_i I) \perp N(A - \lambda_j I)$.

Theorem 61.1. (Unitary Diagonalization) Let A be an $n \times n$ complex matrix. Then $A = U \cdot D \cdot U^*$, where U is a unitary matrix (i.e., $U \cdot U^* = U^* \cdot U = I$) and D is diagonal, if and only if A is normal.

In other words, A has an orthonormal eigenbasis for $\mathbb{C}^n$ if and only if A is a normal matrix.

Proof. We will only prove ($\Rightarrow$) here because ($\Leftarrow$) is more involved.

($\Rightarrow$) Suppose D is diagonal matrix. Then

$$D \cdot D^* = D^* \cdot D.$$

This is because $D \cdot D^*$ and $D^* \cdot D$ are diagonal, and $(D \cdot D^*)_{ii} = d_i \cdot \overline{d_i}$ and $(D^* \cdot D)_{ii} = \overline{d_i} \cdot d_i$.

Assume $A = U \cdot D \cdot U^*$, where U is unitary and D is diagonal. Therefore,

$$\begin{aligned} A \cdot A^* &= (U \cdot D \cdot U^*) \cdot (U \cdot D \cdot U^*)^* \\ &= U \cdot D \cdot D^* \cdot U^* \end{aligned}$$

and

$$\begin{aligned} A^* \cdot A &= (U \cdot D \cdot U^*)^* \cdot (U \cdot D \cdot U^*) \\ &= U \cdot D^* \cdot D \cdot U^*. \end{aligned}$$

Therefore, $AA^* = A^*A$ and so A is normal. $\square$

Remark. The following is a special case of the above theorem.

Let $A \in \mathbb{R}^{n \times n}$ be a real symmetric matrix, i.e., $A = A^T$. (This implies A is normal) Then,

(1) Each eigenvalue of A is real.

(2) All eigenvectors are in $\mathbb{R}^n$.

(3) By the past theorem, there exists an orthogonal matrix $P \in \mathbb{R}^{n \times n}$ (Therefore, $P \cdot P^T = P^T \cdot P = I$.) and a diagonal matrix D such that $A = P^T \cdot D \cdot P$. This is called the spectral decomposition of a real symmetric matrix, i.e., A has on O.N. eigenbasis for $\mathbb{R}^n$, and the diagonal entries of D are the real eigenvalues of A.

62 Positive definite and semi-definite, and indefinite matrices

Definition. Let $A \in \mathbb{R}^{n\times n}$ be a real symmetric matrix, i.e., $A = A^T$.

(1) We call A **positive definite** (P.D.) if

$$x^T \cdot a \cdot x > 0, \quad \forall 0 \neq x \in \mathbb{R}^n.$$

(2) We call A **positive semi-definite** (P.S.D.) if

$$x^T \cdot A \cdot x \geq 0, \quad \forall x \in \mathbb{R}^n.$$

(3) We call A **indefinite** if there exists $x \in \mathbb{R}^{n\times n}$ and $y \in \mathbb{R}^{n\times n}$ such that

$$x^T \cdot A \cdot x > 0 \quad \text{but} \quad y^T \cdot A \cdot y < 0.$$

Theorem 62.1. Let $A \in \mathbb{R}^{n\times n}$ be a real symmetric matrix. The following are equivalent:

(1) A is positive definite.

(2) Each eigenvalue of A is positive.

(3) There is an invertible matrix B such that $A = B^T \cdot B$.

Proof. 1) $\Rightarrow$ 2) Suppose A is P.D. Let $v \in \mathbb{R}^n$ be an eigenvector associated with an eigenvalue λ. Then $Av = \lambda v$ and $v^T Av = v^T(\lambda \cdot v) = \lambda \cdot v^T v$ where $v^T \cdot A \cdot v > 0$ and $v^T v > 0$ because $v \neq 0$.

Since A is positive-definite and $v \neq 0$ then $v^T Av > 0$ and so

$$\lambda = \frac{v^T Av}{v^T v} > 0$$

where $v^T v > 0$. Therefore, each eigenvalue λ of A is positive.

2) $\Rightarrow$ 3): We will use the spectral decomposition. Suppose A is real symmetric and all eigenvalues are positive. Therefore, by the spectral decomposition of A, $A = P^T \cdot D \cdot P$, where P is orthogonal and

$$D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

and $\lambda_i > 0$ for all i.

Define $D^{\frac{1}{2}}$ as

$$D^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\lambda_1} & & & \\ & \sqrt{\lambda_2} & & \\ & & \ddots & \\ & & & \sqrt{\lambda_n} \end{bmatrix}$$

which is a diagonal matrix. Then $D^{\frac{1}{2}} \cdot D^{\frac{1}{2}} = D$ and $D^{\frac{1}{2}}$ is invertible.
Therefore,

$$A = P^T \cdot D^{\frac{1}{2}} \cdot D^{\frac{1}{2}} \cdot P = (P^T \cdot D^{\frac{1}{2}}) \cdot (D^{\frac{1}{2}} \cdot P).$$

Let $B = D^{\frac{1}{2}} \cdot P$. Since $D^{\frac{1}{2}}$ and P are invertible, then B is invertible. Also, $B^T = (D^{\frac{1}{2}} \cdot P)^T = P^T (D^{\frac{1}{2}})^T = P^T \cdot D^{\frac{1}{2}}$, therefore $A = B^T \cdot B$.

3) $\Rightarrow$ 1) Suppose $A = B^T \cdot B$, where B is invertible. We must show A is P.D. For any $0 \neq x \in \mathbb{R}^n$,

$$\begin{aligned} x^T A x &= x^T \cdot B^T \cdot B \cdot x \\ &= (Bx)^T \cdot (Bx) = ||Bx||^2. \end{aligned}$$

Since $x \neq 0$ and B is invertible, then $Bx \neq 0 \Rightarrow ||Bx||^2 > 0$. Therefore, $x^T \cdot A \cdot x > 0$ for all $0 \neq x \in \mathbb{R}^n$. Therefore, A is positive definite. $\square$

If A is positive semi-definite then each eigenvalue is non-negative, $A = B^T \cdot B$, where $B \in \mathbb{R}^{n \times n}$, not necessarily invertible.