Maths-1: Linear Algebra

Proofs

Winter 2021

1 Preliminaries

Vectors as Ordered Lists or n-tuples

Definition 1.1. An ordered list of numbers $(s_1, \ldots, s_n)$ is called an *n-tuple*.

Definition 1.2 (p. 5 of [1])**.** Let m, n be positive integers. An $m \times n$ *matrix* is a collection of mn numbers arranged in a rectangular array, having m rows and n columns. The expression $m \times n$ is referred to as the *size* of the matrix. A $1 \times n$ matrix is also called a *row vector* and an $n \times 1$ matrix is called a *column vector*.

Definition 1.3. The set of all n-tuples of real numbers is called $\mathbb{R}^n$. The set of all n-tuples of complex numbers is called $\mathbb{C}^n$.

The vector whose entries are all zero is called the **zero vector** and is denoted by $\mathbf{0}$.

Equality of vectors in $\mathbb{R}^n$ (and in $\mathbb{C}^n$), as well as the operations of scalar multiplication and vector addition, are defined entry by entry.

In other words, vectors $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$ in $\mathbb{R}^n$ (or in $\mathbb{C}^n$) are defined to be equal if and only if $a_j = b_j$ for $j = 1, \ldots, n$,

$$\mathbf{a} + \mathbf{b} = (a_1 + b_1, \ldots, a_n + b_n).$$

If $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$ and if $c \in \mathbb{R}$ then

$$c\mathbf{a} = (ca_1, \ldots, ca_n).$$

Similarly, if $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{C}^n$ and if $c \in \mathbb{C}$ then

$$c\mathbf{a} = (ca_1, \ldots, ca_n).$$

Remark. Elements of $\mathbb{R}^n$ (or $\mathbb{C}^n$) are usually **represented** as $n \times 1$ column vectors ($n \times 1$ matrices). This is an example of *identification*, where elements of sets that are isomorphic in an obvious way are *identified* with each other, i.e. used interchangeably to mean the same thing. Vectors in $\mathbb{R}^n$ can likewise be identified with $1 \times n$ row vectors.

Algebraic Properties of $\mathbb{R}^n$ (and $\mathbb{C}^n$)

For all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^{\mathbf{n}}$ and all real numbers c and d,

- $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
- $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
- $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$.
- $\mathbf{u} + (-\mathbf{u}) = -\mathbf{u} + \mathbf{u} = \mathbf{0}$ (where $-\mathbf{u}$ denotes $(-1)\mathbf{u}$)
- $c(\mathbf{u} + \mathbf{v}) = \mathbf{cu} + \mathbf{cv}$
- $(c + d)\mathbf{u} = \mathbf{cu} + \mathbf{du}$
- $c(d\mathbf{u}) = (\mathbf{cd})\mathbf{u}$
- $1\mathbf{u} = \mathbf{u}$

Similarly, for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{C}^{\mathbf{n}}$ and all complex numbers c and d,

- $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
- $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
- $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$.
- $\mathbf{u} + (-\mathbf{u}) = -\mathbf{u} + \mathbf{u} = \mathbf{0}$ (where $-\mathbf{u}$ denotes $(-1)\mathbf{u}$)
- $c(\mathbf{u} + \mathbf{v}) = \mathbf{cu} + \mathbf{cv}$
- $(c + d)\mathbf{u} = \mathbf{cu} + \mathbf{du}$
- $c(d\mathbf{u}) = (\mathbf{cd})\mathbf{u}$
- $1\mathbf{u} = \mathbf{u}$

Remark. The verification of the above properties is routine and is left as an exercise.

Definition 1.4. Given vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots \mathbf{v_p} \in \mathbb{R}^{\mathbf{n}}$ and given real numbers $c_1, c_2, \ldots, c_p$, the vector $\mathbf{y}$ defined by

$$\mathbf{y} = \mathbf{c_1 v_1} + \ldots + \mathbf{c_p v_p}$$

is called a *linear combination* of $\mathbf{v_1}, \ldots, \mathbf{v_p}$ with *weights* or *coefficients* $c_1, \ldots, c_p$.

This is well defined because of associativity of vector addition (a proof by induction is routine, and left as an exercise).

Remark. The definition of a linear combination of vectors in $\mathbb{C}^n$ is the same as above, except that the scalars $c_1, c_2, \ldots, c_p$ are complex numbers.

Properties of Matrices

Theorem 1.5. *Let $A, B,$ and C be $m \times n$ matrices, and let r and s be scalars.*

(a) $A + B = B + A$ *(Commutativity of addition)*

(b) $A + (B + C) = (A + B) + C$ *(Associativity of addition)*

(c) $A + 0 = A$ *(Additive Identity)*

(d) $r(A + B) = rA + rB$ *(Distributive Law)*

(e) $(r + s)A = rA + sA$ *(Distributive Law)*

(f) $r(sA) = (rs)A$

Remark. The proof of this theorem is a routine verification and is left as an exercise.

Definition 1.6. Let $A = (a_{ij})$ be an $m \times n$ matrix and $B = (b_{ij})$ be an $n \times p$ matrix. We define the matrix product $C := AB$ as the $m \times p$ matrix whose entries are

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

Remark. The definition above is referred to as the *Row-Column Rule*, because the i, j-th entry of AB is the product of the i-th row of matrix A and the j-th row of matrix B.

Theorem 1.7. *Let A be an $m \times n$ matrix, and let B and C have sizes for which the indicated sums and products are defined.*

(a) $A(BC) = (AB)C$ *(associative law of multiplication)*

(b) $A(B + C) = AB + AC$ *(left distributive law)*

(c) $(B + C)A = BA + CA$ *(right distributive law)*

(d) $r(AB) = (rA)B = A(rB)$ *for any scalar* r

(e) $I_m A = A = A I_n$ *(identity for matrix multiplication)*

Proof. Let $A = (a_{ij})$.

(a) Let $B = (b_{ij})$ be an $n \times p$ matrix and $C = (c_{ij})$ be a $p \times q$ matrix. Then for $1 \le i \le m, 1 \le j \le q$, the i, j-th entry of $A(BC)$ is

$$\sum_{k=1}^{n} a_{ik} \left(\sum_{l=1}^{p} b_{kl} c_{lj} \right) = \sum_{k=1}^{n} \sum_{l=1}^{p} a_{ik} b_{kl} c_{lj} \tag{1}$$

Similarly, the i, j-th entry of $(AB)C$ is

$$\sum_{l=1}^{p} \left(\sum_{k=1}^{n} a_{ik} b_{kl} \right) c_{lj} = \sum_{k=1}^{n} \sum_{l=1}^{p} a_{ik} b_{kl} c_{lj} \tag{2}$$

From (1) and (2) it follows that $A(BC) = (AB)C$.

(b) Let $B = (b_{ij})$ be an $n \times p$ matrix and $C = (c_{ij})$ be a $n \times p$ matrix. Then for $1 \leq i \leq m, 1 \leq j \leq p$, the i, j-th entry of $A(B + C)$ is

$$\sum_{k=1}^{n} a_{ik} \left(b_{kj} + c_{kj}\right) = \sum_{k=1}^{n} a_{ik} b_{kj} + a_{ik} c_{kj} \tag{3}$$

The i, j-th entry of $AB + AC$ is

$$\sum_{k=1}^{n} a_{ik} b_{kj} + a_{ik} c_{kj} \tag{4}$$

From (3) and (4) it follows that $A(B + C) = AB + AC$.

(c) This is very similar to the verification of property (b) and is left as an exercise.

(d) Let $B = (b_{ij})$ be an $n \times p$ matrix. For $1 \leq i \leq m, 1 \leq j \leq p$, the i, j-th entry of $r(AB)$ is

$$r \sum_{k=1}^{n} a_{ik} b_{kp} = \sum_{k=1}^{n} r a_{ik} b_{kp},$$

the i, j-th entry of $(rA)B$ is

$$\sum_{k=1}^{n} (r a_{ik}) b_{kp} = \sum_{k=1}^{n} r a_{ik} b_{kp},$$

and the i, j-th entry of $A(rB)$ is

$$\sum_{k=1}^{n} r a_{ik} (r b_{kp}) = \sum_{k=1}^{n} r a_{ik} b_{kp}.$$

(e) Let $I_m = (\delta_{ij})$. Then for $1 \leq i, j \leq m$,

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Hence for $1 \leq i \leq m, 1 \leq j \leq n$, the i, j-th entry of $I_m A$ is

$$\sum_{k=1}^{m} \delta_{ik} a_{kj} = a_{ij}$$

Hence $I_m A = A$. The proof that $A I_n = A$ is similar and is left as an exercise.

$\square$

Definition 1.8. Let A be an $n \times n$ matrix. For $k \in \mathbb{N}$, we define

$$A^k = \underbrace{A \dots A}_{k \text{ times}}$$

Definition 1.9. Let $A = (a_{ij})$ be an $m \times n$ matrix. We define the transpose A^T to be the $n \times m$ matrix whose i, j-th entry is a_{ji}.

Theorem 1.10. *Let A and B denote matrices whose sizes are appropriate for the following sums and products.*

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

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

(c) $(rA)^T = rA^T$ *for any scalar* r

(d) $(AB)^T = B^T A^T$

Proof. (a) Let $A = (a_{ij})$ be an $m \times n$ matrix. Let $A^T = (c_{ij})$. Then for $1 \leq i \leq m, 1 \leq j \leq n$

$$c_{ij} = a_{ji}$$

Therefore the i,j-th entry of $(A^T)^T$ is

$$c_{ji} = a_{ij}$$

Thus $(A^T)^T = A$.

(b) Let $A = (a_{ij})$ and $B = (b_{ij})$ be $m \times n$ matrices. Then the i,j-th entry of $(A+B)^T$ is the j,i-th entry of $A+B$ which is

$$a_{ji} + b_{ji}$$

The i,j-th entry of $A^T + B^T$ is also

$$a_{ji} + b_{ji}$$

Therefore $(A+B)^T = A^T + B^T$.

(c) Let $A = (a_{ij})$ be an $m \times n$ matrix. Then the i,j-th entry of $(rA)^T$ is the j,i-th entry of rA which is

$$ra_{ji}$$

The i,j-th entry of rA^T is also

$$ra_{ji}$$

Therefore $(rA)^T = rA^T$.

(d) Let $A = (a_{ij})$ be an $m \times n$ matrix and $B = b_{ij}$ be an $n \times p$ matrix. Then the i,j-th entry of $(AB)^T$ is

$$\sum_{k=1}^{n} a_{jk} b_{ki}$$

Let $C = B^T A^T = (c_{ij})$. Since the i,k-th entry of B^T is b_{ki} and the k,j-th entry of A^T is a_{jk}, it follows that

$$c_{ij} = \sum_{k=1}^{n} b_{ki} a_{jk}$$

Hence $(AB)^T = B^T A^T$

$\square$

Definition 1.11. Let A be an $m \times n$ matrix having real entries. Let the columns of A be $\mathbf{a}_1, \ldots, \mathbf{a}_n$, and let $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. We define $A\mathbf{x}$ to be the linear combination

$$x_1\mathbf{a}_1 + \ldots + x_n\mathbf{a}_n. \tag{5}$$

If $A = [\mathbf{a}_1 \quad \ldots \quad \mathbf{a}_n]$ is an $m \times n$ matrix having complex entries, and $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{C}^n$, then $A\mathbf{x}$ is defined by equation (5), and is a vector in $\mathbb{C}^n$.

Remark. Please note that the above definition agrees with the definition of $A\mathbf{x}$, when $\mathbf{x}$ is viewed as a column matrix, as the $k, 1$-th entry of the column matrix $A\mathbf{x}$ is the same as the k-th entry of $A\mathbf{x}$, viewed as a vector in $\mathbb{R}^m$ (or $\mathbb{C}^m$).

Proposition 1.12. *This matrix product has the property that if $\mathbf{b_1}, \ldots, \mathbf{b_p}$ are the columns of B then*

$$AB = [A\mathbf{b}_1 \quad A\mathbf{b}_2 \quad \ldots \quad A\mathbf{b_p}]$$

Thus, each column of AB is a linear combination of the columns of A using weights from the corresponding column of B.

Proof. Let $A = (a_{ij})$ be an $m \times n$ matrix and let $B = (b_{ij})$ be an $n \times p$ matrix whose columns are $\mathbf{b_1}, \ldots, \mathbf{b_p}$.

Let $k \in \{1, \ldots, p\}$. Then for $1 \leq l \leq m$, the l-th entry of the k-th column of AB is simply the l, k-th entry of the matrix product AB, which is

$$\sum_{j=1}^{n} a_{lj} b_{jk}$$

If we denote the columns of A by $\mathbf{a}_1, \ldots, \mathbf{a}_n$, then

$$A\mathbf{b}_k = b_{1k}\mathbf{a}_1 + \ldots + b_{nk}\mathbf{a}_n = \sum_{j=1}^{n} b_{jk}\mathbf{a}_j$$

Thus the l-th entry of the column vector $A\mathbf{b}_k$ is

$$\sum_{j=1}^{n} b_{jk} a_{lj}.$$

Therefore $A\mathbf{b}_k$ is the k-th column of AB. $\square$

Proposition 1.13. *Each row of AB is a linear combination of the rows of B using weights from the corresponding row of A. In other words*

$$row_i(AB) = row_i(A)B.$$

Proof.

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

$\square$

Theorem 1.14 (Column-Row Expansion of Matrix Product AB)**.** *Let $A = (a_{ij})$ be an $m \times n$ matrix and Let $B = (b_{ij})$ be an $n \times p$ matrix.*

Let the columns of A be $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and the rows of B be $\mathbf{b}_1^T, \ldots, \mathbf{b}_n^T$.
Then

$$AB = \mathbf{a}_1\mathbf{b}_1^T + \cdots + \mathbf{a}_n\mathbf{b}_n^T \tag{6}$$

Proof. The i, j-th entry of AB is

$$\sum_{k=1}^{n} a_{ik}b_{kj}$$

$$\begin{aligned}
\mathbf{a}_k\mathbf{b}_k^T &= \begin{bmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{bmatrix} \begin{bmatrix} b_{k1} & b_{k2} & \ldots & b_{kp} \end{bmatrix} \\
&= \begin{bmatrix} a_{1k}b_{k1} & a_{1k}b_{k2} & \ldots & a_{1k}b_{kp} \\ a_{2k}b_{k1} & a_{2k}b_{k2} & \ldots & a_{2k}b_{kp} \\ \vdots & \vdots & \ddots & \vdots \\ a_{mk}b_{k1} & a_{mk}b_{k2} & \ldots & a_{mk}b_{kp} \end{bmatrix}
\end{aligned}$$

Thus the i, j-th entry of the matrix $\mathbf{a}_k\mathbf{b}_k^T$ is $a_{ik}b_{kj}$.

It follows that the i, j-th entries of the matrix on the LHS and the matrix on the RHS of (6) are equal.

□

Invertible Matrices

Definition 1.15. An $n \times n$ matrix A is said to be *invertible* if there exists an $n \times n$ matrix B such that

$$AB = BA = I.$$

The matrix B is called the *inverse* of A and is denoted by A^{-1}. (We can use the definite article "the" in this sentence, thank to the following proposition.)

Proposition 1.16. *The inverse of a matrix is unique, if its exists.*

Proof. Let A, B and C be $n \times n$ matrices such that

$$AB = BA = I$$

and

$$AC = CA = I.$$

Then

$$B = BI = B(AC) = (BA)C = IC = C,$$

as matrix multiplication is associative. □

An invertible matrix is also called a *nonsingular* matrix. A matrix which is not invertible is called a *singular* matrix.

Theorem 1.17. *(a) If A is an invertible matrix, then A^{-1} is invertible and*

$$(A^{-1})^{-1} = A$$

(b) If A and B are $n \times n$ invertible matrices, then so is AB, and the inverse of AB is the product of the inverses of A and B in the reverse order. That is,

$$(AB)^{-1} = B^{-1}A^{-1}$$

(c) If A is an invertible matrix, then so is A^T , and the inverse of A^T is the transpose of A^{-1}. That is,

$$(A^T)^{-1} = (A^{-1})^T$$

Proof. (a) This is obvious from the definition of the inverse of a matrix.

(b) Observe that

$$(AB)B^{-1}A^{-1} = A(BB^{-1})A^{-1} = (AI)A^{-1} = AA^{-1} = I$$

and

$$B^{-1}A^{-1}(AB) = B^{-1}(A^{-1}A)B = (B^{-1}I)B = B^{-1}B = I$$

(c) $AB = I \implies B^T A^T = I$, and $BA = I \implies A^T B^T = I$

□

2 Systems of Linear Equations

Definition 2.1 (p. 2 of [1])**.** A *linear equation* in the variables $x_1, \ldots, x_n$ is an equation that can be written in the form

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = b$$

where b and the *coefficients* $a_1 \ldots a_n$ are real or complex numbers.

Remark. If the coefficients are real we say it's a linear equation with real coefficients and if the coefficients are complex, we say it's a linear equation with complex coefficients.

Definition 2.2. A *system of linear equations* is a collection of one or more linear equations involving the same variables, say $x_1, \ldots, x_n$. A *solution* of the system is an ordered list $(s_1, s_2, \ldots, s_n)$ of numbers that makes each equation a true statement when the values $s_1, \ldots s_n$ are substituted for $x_1, \ldots, x_n$, respectively.

Definition 2.3 (p. 2 of [1])**.** The set of all possible solutions is called the *solution set* of the linear system. Two linear systems are called *equivalent* if they have the same solution set.

Classification of Linear Systems

A system of linear equations has either

1. no solution, or
2. exactly one solution, or
3. infinitely many solutions.

Remark. A proof of the above statement will be given later. The statement will not be used in proofs until it is proved, to avoid circularity.

Definition 2.4. A system of linear equations is said to be **consistent** if it has a solution; a system is **inconsistent** if it has no solution.

System of Linear Equations in Matrix Form

A system of m linear equations in n variables, say

$$\begin{aligned} a_{11}\,x_1 + \ a_{12}x_2 + \ldots + \ a_{1n}x_n &= b_1 \\ a_{21}\,x_1 + \ a_{22}x_2 + \ldots + \ a_{2n}x_n &= b_2 \\ &\ldots \\ &\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m \end{aligned}$$

can be written using matrix notation as

$$A\mathbf{x} = \mathbf{b}$$

where

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

is the *matrix of coefficients* (or *coefficient matrix*) of the system and

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ . \\ . \\ . \\ b_m \end{bmatrix}$$

are column vectors. The matrix

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ . & & & & \\ . & & & & \\ . & & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{bmatrix}$$

is called the *augmented matrix* of the system.

Proposition 2.5. *A vector equation*

$$x_1\mathbf{a}_1 + \ldots + x_n\mathbf{a}_n = \mathbf{b}$$

has the same solution set as the linear system whose augmented matrix is

$$[\mathbf{a_1} \quad \mathbf{a_2} \ldots \quad \mathbf{a_n} \quad \mathbf{b}] \tag{$*$}$$

In particular, $\mathbf{b}$ *can be generated by a linear combination of* $\mathbf{a_1}, \ldots, \mathbf{a_n}$ *if and only if there exists a solution to the linear system which has* $(*)$ *as its corresponding augmented matrix .*

Proof. By Definintion 1.11, the vector equation

$$x_1\mathbf{a}_1 + \ldots + x_n\mathbf{a}_n = \mathbf{b}$$

has the same solution set as the linear system $A\mathbf{x} = \mathbf{b}$ where

$$A = [\mathbf{a}_1 \quad \ldots \quad \mathbf{a}_n].$$

□

Row Reduction

We simplify the augmented matrix using a sequence of operations known as *elementary row operations.*

There are three kinds of operations:

1. (Replacement) Replace one row by the sum of itself and a multiple of another row.
2. (Interchange) Interchange two rows.
3. (Scaling) Multiply all entries in a row by a nonzero constant.

Definition 2.6. A rectangular matrix is in *echelon form (or row echelon form)* if it has the following two properties:

1. Each leading entry of a row is in a column to the right of the leading entry of the row above it.

2. All nonzero rows are above any rows of all zeros.

A *leading entry* of a row refers to the leftmost nonzero entry (in a nonzero row).

A *nonzero row* or nonzero column in a matrix means a row or column that contains at least one nonzero entry.

If a matrix is in row echelon form, then the leading entry in each nonzero row is called a *pivot*, and a column which contains a pivot is called a *pivot column*.

Remark. The definition of a pivot column will be extended to the context of non echelon matrices, after we discuss the uniqueness of the RREF.

Elementary Matrices

Applying an elementary row operation on a matrix is the same multiplying the matrix **on the left** by an *elementary matrix*. If we carry out row operations on an $m \times n$ matrix A, then the corresponding elementary matrices are $m \times m$ square matrices.

(1) Row Replacement:

The operation $R_i \to R_i + cR_j$ is achieved via left multiplicaton by a matrix of the form

$$E = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & c & \\ & & & \ddots & & \\ & & & & 1 & \\ & & & & & 1 \end{bmatrix} \qquad (i < j)$$

or

$$E = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \ddots & & \\ & & c & & 1 & \\ & & & & & 1 \end{bmatrix} \qquad (i > j)$$

The matrix has a c as its ij-th entry and otherwise looks like the $m \times m$ identity matrix. More precisely, if $E = (E_{kl})$, then for $k, l \in \{1, \dots, m\}$

$$E_{kl} = \begin{cases} 1 & \text{if } k = l \\ c & \text{if } k = i, l = j \\ 0 & \text{otherwise.} \end{cases}$$

(2) Row Interchange:

The operation $R_i \longleftrightarrow R_j$ is achieved via left multiplicaton by a matrix of the form

$$E = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 0 & & 1 & \\ & & & \ddots & & \\ & & 1 & & 0 & \\ & & & & & 1 \end{bmatrix}$$

If $E = (E_{kl})$, then for $k, l \in \{1, \ldots, m\}$

$$E_{kl} = \begin{cases} 1 & \text{if } k = l, k \neq i, k \neq j \\ 1 & \text{if } k = i, l = j \\ 1 & \text{if } k = j, l = i \\ 0 & \text{otherwise.} \end{cases}$$

(3) Row Scaling:

The operation $R_i \to cR_i$ is achieved via left multiplicaton by a matrix of the form

$$E = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & c & & & \\ & & & \ddots & & \\ & & & & 1 & \\ & & & & & 1 \end{bmatrix}$$

If $E = (E_{kl})$, then for $k, l \in \{1, \ldots, m\}$

$$E_{kl} = \begin{cases} 1 & \text{if } k = l, k \neq i \\ c & \text{if } k = l = i \\ 0 & \text{otherwise.} \end{cases}$$

Inverses of Elementary Matrices

Proposition 2.7. *Elementary matrices of all three kinds are invertible, and the inverse of an elementary matrix is also an elementary matrix.*

Proof. We denote by $\mathbf{e}_1, \ldots, \mathbf{e}_n$ the columns of the $n \times n$ identity matrix.

(i) Let us first show that elementary matrices corresponding to Row Scaling are invertible.

Let E be an $n \times n$ elementary matrix which corresponds to the operation of scaling row i by a scalar $c \neq 0$, i.e. $R_i \to cR_i$.

As left multiplication of a matrix by a row vector corresponds to a linear combination of the rows of the matrix, it follows that the rows of E are $\mathbf{e}_1^T, \ldots, c\mathbf{e}_i^T, \ldots, \mathbf{e}_n^T$ (a row vector is the transpose of a column vector).

Since E is symmetric ($E = E^T$), the columns of E are $\mathbf{e}_1, \ldots, c\mathbf{e}_i, \ldots, \mathbf{e}_n$, i.e.

$$E = [\mathbf{e}_1 \quad \ldots \quad c\mathbf{e}_i \quad \ldots \quad \mathbf{e}_n]$$

Put

$$F = [\mathbf{e}_1 \quad \ldots \quad \frac{1}{c}\mathbf{e}_i \quad \ldots \quad \mathbf{e}_n]$$

Then

$$EF = [E\mathbf{e}_1 \quad \ldots \quad \frac{1}{c}E\mathbf{e}_i \quad \ldots \quad E\mathbf{e}_n]$$

Now $E\mathbf{e}_k$ is simply the k-th column of E, for any $k = 1, \ldots, n$. Therefore

$$\begin{aligned} EF &= [E\mathbf{e}_1 \quad \ldots \quad \frac{1}{c}E\mathbf{e}_i \quad \ldots \quad E\mathbf{e}_n] \\ &= [\mathbf{e}_1 \quad \ldots \quad \frac{1}{c}c\mathbf{e}_i \quad \ldots \quad \mathbf{e}_n] \\ &= [\mathbf{e}_1 \quad \ldots \quad \mathbf{e}_i \quad \ldots \quad \mathbf{e}_n] = I \end{aligned}$$

Similarly $FE = I$. Therefore F is the inverse of E.

(ii) Next let us show that elementary matrices corresponding to Row Interchange are invertible.

Let E be an $n \times n$ elementary matrix which corresponds to the operation of interchanging rows i and j, i.e. the operation $R_i \leftrightarrow R_j$.

Without loss of generality, assume that $i < j$.

Then the rows of E are $\mathbf{e}_1^T, \ldots, \mathbf{e}_j^T, \ldots, \mathbf{e}_i^T, \ldots, \mathbf{e}_n^T$.

Since E is symmetric, the columns of E are $\mathbf{e}_1, \ldots, \mathbf{e}_j, \ldots, \mathbf{e}_i, \ldots, \mathbf{e}_n$, i.e.

$$E = [\mathbf{e}_1 \quad \ldots \quad \mathbf{e}_j \quad \ldots \quad \mathbf{e}_i \quad \ldots \quad \mathbf{e}_n]$$

Recall that $E\mathbf{e}_k$ is simply the k-th column of E, for any $k = 1, \ldots, n$. Therefore

$$\begin{aligned} EE &= [E\mathbf{e}_1 \quad \ldots \quad E\mathbf{e}_j \quad \ldots \quad E\mathbf{e}_i \quad \ldots \quad E\mathbf{e}_n] \\ &= [\mathbf{e}_1 \quad \ldots \quad \mathbf{e}_i \quad \ldots \quad \mathbf{e}_j \quad \ldots \quad \mathbf{e}_n] = I \end{aligned}$$

Therefore the inverse of E is itself.

(iii) Next we show that elementary matrices corresponding to Row Replacement are invertible.

Let E be an $n \times n$ elementary matrix which corresponds to the operation $R_i \to R_i + cR_j$.

Without loss of generality, assume that $i < j$.

Then the rows of E are $\mathbf{e}_1^T, \ldots, \mathbf{e}_i^T + c\mathbf{e}_j^T, \ldots, \mathbf{e}_j^T, \ldots, \mathbf{e}_n^T$, i.e.

$$E^T = [\mathbf{e}_1 \quad \ldots \quad \mathbf{e}_i + c\mathbf{e}_j \quad \ldots \quad \mathbf{e}_j \ldots \quad \mathbf{e}_n]$$

Put

$$F = [\mathbf{e}_1 \quad \ldots \quad \mathbf{e}_i - c\mathbf{e}_j \quad \ldots \quad \mathbf{e}_j \ldots \quad \mathbf{e}_n]$$

Then

$$E^T F = [E^T\mathbf{e}_1 \quad \ldots \quad E^T(\mathbf{e}_i - c\mathbf{e}_j) \quad \ldots \quad E^T\mathbf{e}_j \ldots \quad E^T\mathbf{e}_n]$$

Now $E^T\mathbf{e}_k$ is simply the k-th column of E^T, for any $k = 1, \ldots, n$. Therefore

$$\begin{aligned}
E^T F &= [E^T\mathbf{e}_1 \quad \ldots \quad E^T(\mathbf{e}_i - c\mathbf{e}_j) \quad \ldots \quad E^T\mathbf{e}_j \quad \ldots \quad E^T\mathbf{e}_n] \\
&= [\mathbf{e}_1 \quad \ldots \quad E^T\mathbf{e}_i - cE^T\mathbf{e}_j \quad \ldots \quad \mathbf{e}_j \ldots \quad \mathbf{e}_n] \\
&= [\mathbf{e}_1 \quad \ldots \quad \mathbf{e}_i + c\mathbf{e}_j - c\mathbf{e}_j \quad \ldots \quad \mathbf{e}_j \ldots \quad \mathbf{e}_n] \\
&= [\mathbf{e}_1 \quad \ldots \quad \mathbf{e}_i \quad \ldots \quad \mathbf{e}_j \ldots \quad \mathbf{e}_n] = I
\end{aligned}$$

Similarly $FE^T = I$. Therefore F is the inverse of E^T, and hence F^T is the inverse of E.

□

Visual Representation of Inverses of Elementary Matrices

(1) Replacement: $R_i \to R_i + cR_j$

For $i < j$

$$E = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & c & \\ & & & \ddots & & \\ & & & & 1 & \\ & & & & & 1 \end{bmatrix}, \quad E^{-1} = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & -c & \\ & & & \ddots & & \\ & & & & 1 & \\ & & & & & 1 \end{bmatrix}$$

Replacement: $R_i \to R_i + cR_j$

For $i > j$

$$E = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \ddots & & \\ & & c & & 1 & \\ & & & & & 1 \end{bmatrix}, \quad E^{-1} = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \ddots & & \\ & & -c & & 1 & \\ & & & & & 1 \end{bmatrix}$$

(2) Interchange:

$$E = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 0 & & 1 & \\ & & & \ddots & & \\ & & 1 & & 0 & \\ & & & & & 1 \end{bmatrix} = E^{-1}$$

(3) Scaling:

$$E = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & c & & & \\ & & & \ddots & & \\ & & & & 1 & \\ & & & & & 1 \end{bmatrix}, \quad E^{-1} = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & \frac{1}{c} & & & \\ & & & \ddots & & \\ & & & & 1 & \\ & & & & & 1 \end{bmatrix}$$

Equivalence of Systems

Definition 2.8. Two matrices A and B are said to be *row equivalent* is there exists a sequence of elementary row operations which can be performed on A to obtain B. Two systems of equations are row equivalent if the augmented matrices which represent the systems are row equivalent.

Theorem 2.9. *Any two systems of linear equations that are row equivalent, have the same solution sets.*

Proof. Let $A\mathbf{x} = \mathbf{b}$ and $A'\mathbf{x} = \mathbf{b}'$ be row equivalent systems of linear equations. Then there exists a sequence $E_1, \ldots, E_p$ of elementary matrices such that

$$E_p E_{p-1} \ldots E_1 [A \quad \mathbf{b}] = [A' \quad \mathbf{b}']$$

Let $E = E_p E_{p-1} \ldots E_1$. Then

$$[EA \quad E\mathbf{b}] = [A' \quad \mathbf{b}'] \implies EA = A', E\mathbf{b} = \mathbf{b}'$$

Thus

$$A\mathbf{x} = \mathbf{b} \implies EA\mathbf{x} = E\mathbf{b} \implies A'\mathbf{x} = \mathbf{b}'$$

and

$$A'\mathbf{x} = \mathbf{b}' \implies EA\mathbf{x} = E\mathbf{b} \implies E^{-1}EA\mathbf{x} = E^{-1}E\mathbf{b} \implies A\mathbf{x} = \mathbf{b}$$

Hence

$$A\mathbf{x} = \mathbf{b} \iff A'\mathbf{x} = \mathbf{b}',$$

so both equations have the same solution sets. □

Reduced Row Echelon Form

Definition 2.10. If a matrix in echelon form satisfies the following additional conditions, then it is in *reduced echelon form (or reduced row echelon form)*:

1. The leading entry in each nonzero row is 1.
2. Each leading 1 is the only nonzero entry in its column.

Definition 2.11. Let A be an $m \times n$ matrix. An $m \times n$ matrix A' is said to be *an echelon form of* A if A' is row equivalent to A.

Lemma 2.12. *Let A be an $m \times n$ matrix and let A' be an echelon form of A. Then there exists an invertible $m \times m$ matrix E such that*

$$A' = EA.$$

Further, E can be a expressed as a product of elementary matrices. (We will later show that every invertible matrix can be expressed as a product of elementary matrices.)

Proof. Let $E_1, \dots, E_p$ be the elementary matrices that correspond to reducing A to A', i.e.

$$A' = E_p E_{p-1} \dots E_1 A$$

Let $E = E_p E_{p-1} \dots E_1$. Then E is invertible, as it is a product of invertible matrices. $\square$

Proposition 2.13. *Every $m \times n$ matrix is row equivalent to a matrix which is in reduced echelon form.*

Proof. Let $A = (a_{ij})$. We prove the proposition using induction on n.

Let us first consider the case $n = 1$. If $A = 0$ then it is already in reduced echelon form.

So assume that $A \neq 0$. Let row j be the first nonzero row of A. If $j \neq 1$ then let E_0 be the $m \times m$ elementary matrix which corresponds to interchanging row 1 and row j. If $j = 1$ then let $E_0 = I$.

Let E_1 be the $m \times m$ elementary matrix which corresponds to scaling row 1 by $\frac{1}{a_{11}}$, and let E_j be elementary matrices corresponding to the row operation $R_j \to R_j - a_{j1}R_1$ for $j = 2, \dots, m$. Thus $E_m E_{m-1} \dots E_1 E_0 A = \mathbf{e}_1$, which is in reduced echelon form.

Next, we assume that the proposition holds for $n \leq k$ and show that it holds true for $n = k + 1$.

By the induction hypothesis, there exists an $m \times k$ matrix A' such that A' is in reduced echelon form and is row equivalent to the matrix formed by the first k columns of A.

By Lemma 2.12, there exists an invertible $m \times m$ matrix E such that E can be expressed as product of elementary matrices and such that

$$EA = [A' \quad \mathbf{a}]$$

where $\mathbf{a} \in \mathbb{R}^{\mathbf{m}}$ (or $\mathbf{a} \in \mathbb{C}^{\mathbf{m}}$, if A has complex entries). If EA does not have a pivot position in the last row, then EA is already in reduced echelon form. If EA has a pivot position in the last row, then $a_m \neq 0$. Let E_1 be the $m \times m$ elementary matrix which corresponds to scaling by $\frac{1}{a_m}$, and let E_j be elementary matrices which correspond to the row operation $R_j \to R_j - a_j R_m$, for $j = 2, \dots, m$. Then

$$E_m E_{m-1} \dots E_1 E A = [A' \quad \mathbf{e}_m],$$

which is in reduced echelon form. $\square$

Proposition 2.14. *Let $A = (a_{ij})$ be an $m \times n$ matrix having real entries and $\mathbf{b} = (b_1, \dots, b_n) \in \mathbb{R}^m$ be a vector such that $[A \quad \mathbf{b}]$ is a matrix in reduced echelon form. Then $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b}$ is not a pivot column of $[A \quad \mathbf{b}]$.*

If A is an $m \times n$ matrix having complex entries and $\mathbf{b} = (b_1, \dots, b_n) \in \mathbb{C}^m$ and $[A \quad \mathbf{b}]$ is a matrix in reduced echelon form, then the same statement holds, i.e. $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b}$ is not a pivot column of $[A \quad \mathbf{b}]$.

Proof. We first consider the case where A and $\mathbf{b}$ have real entries.

Let $\mathbf{a}_1, \dots, \mathbf{a}_n$ be the columns of A, i.e.

$$A = [\mathbf{a}_1 \quad \dots \quad \mathbf{a}_n]$$

Let $\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_k}$ be the pivot columns of A, where $1 \leq k \leq m$ and

$$1 \leq i_1 < i_2 < \ldots < i_k \leq n.$$

Suppose $\mathbf{b}$ is not a pivot column of $[A \quad \mathbf{b}]$. Then

$$\mathbf{b} = \sum_{j=1}^{m} b_j \mathbf{e}_j = \sum_{j=1}^{k} \frac{b_j}{a_{j,i_j}} \mathbf{a}_{i_j},$$

as $b_j = 0$ for $j > k$.

Put $x_{i_j} = \frac{b_j}{a_{j,i_j}}$ for $j = 1, \dots, k$ and $x_l = 0$ for $l \in \{1, \dots, n\} \setminus \{i_1, \dots, i_k\}$. Then $\mathbf{x} = (\mathbf{x_1}, \dots, \mathbf{x_n})$ is a solution of $A\mathbf{x} = \mathbf{b}$.

Next, suppose $\mathbf{b}$ is a pivot column of $[A \quad \mathbf{b}]$. Then $b_{k+1} \neq 0$ and $a_{k+1,j} = 0$ for $j = 1, \dots, n$. In terms of column vectors, this means $(\mathbf{a}_j)_{k+1} = 0$ for $j = 1, \dots, n$.

Let $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ be any vector. Then

$$(A\mathbf{x})_{k+1} = \left(\sum_{j=1}^{n} x_j \mathbf{a}_j \right)_{k+1} = 0$$

Hence $A\mathbf{x} \neq \mathbf{b}$ for every $\mathbf{x} \in \mathbb{R}^n$.

The same proof works when A and $\mathbf{b}$ have complex entries, if we replace $\mathbb{R}^m$ with $\mathbb{C}^m$. $\square$

Uniqueness of the Reduced Echelon Form

Theorem 2.15. *Each matrix is row equivalent to one and only one reduced echelon matrix.*

Proof. (The proof of this theorem is given here for the sake of completeness. It is not part of the syllabus.)

Observe that if an $m \times n$ matrix (where $n > 1$) is in reduced echelon form, then the matrix formed by its first $n-1$ columns is also in reduced echelon form. In other words, if

$$B = [\mathbf{b}_1 \quad \dots \quad \mathbf{b}_n]$$

is a matrix in reduced echelon form, then

$$B' = [\mathbf{b}_1 \quad \dots \quad \mathbf{b}_{n-1}]$$

is also in reduced echelon form. This is an obvious consequence of the definition of the RREF.

Let A be an $m \times n$ matrix. If $A = 0$ the the RREF of A must also be zero, so there is nothing to show. So let us assume that $A \neq 0$.

We prove the theorem using induction on n.

By Proposition 2.13 we already know than an RREF of A exists.

If $n = 1$, then a reduced echelon form A' of A, is either the zero vector or $\mathbf{e}_1$ (where $\mathbf{e}_1, \ldots, \mathbf{e}_m$ are the columns of I_m).

Let E be an invertible matrix such that $A' = EA$. Then $A = E^{-1}A'$. Hence $A' = 0 \implies A = 0$. Since we assumed that $A \neq 0$, we must have $A' = \mathbf{e}_1$.

Next, suppose the theorem is true for all $n \leq k$, where $k \in \mathbb{N}$. We show that it holds for $n = k + 1$.

Let A' and $\tilde{A}$ be reduced echelon forms of A. By the induction hypothesis, the first k columns of A' and $\tilde{A}$ must be equal. Thus there exists an $m \times k$ matrix B and column vectors $\mathbf{b}', \tilde{\mathbf{b}}$ such that

$$A' = [B \quad \mathbf{b}'], \quad \tilde{A} = [B \quad \tilde{\mathbf{b}}]$$

Let $\mathbf{e}_1, \ldots, \mathbf{e}_{l-1}$ be the pivot columns of B, where $l \leq m$.

Also, by Lemma 2.12, there exist invertible matrices E' and $\tilde{E}$ such that

$$A' = E'A, \quad \tilde{A} = \tilde{E}A$$

The following cases arise.

Case (i): Both A' and $\tilde{A}$ contain a pivot in the last column.

Then $\mathbf{b}' = \tilde{\mathbf{b}} = \mathbf{e}_l$. Hence $A' = \tilde{A}$.

Case (ii): A' contains a pivot in the last column but $\tilde{A}$ does not. By Theorem 2.9, the linear systems $B\mathbf{x} = \mathbf{b}'$ and $B\mathbf{x} = \tilde{\mathbf{b}}$ must have the same solution sets.

But by Proposition 2.14, $B\mathbf{x} = \mathbf{b}'$ is inconsistent, whereas $B\mathbf{x} = \tilde{\mathbf{b}}$ is consistent. This is a contradiction. Therefore case (ii) cannot occur.

Case (iii): $\tilde{A}$ contains a pivot in the last column but A' does not. Similar to case (ii).

Case (iv): Neither A' nor $\tilde{A}$ contains a pivot in the last column.

By Proposition 2.14, the system $B\mathbf{x} = \mathbf{b}'$ has a solution, say ξ. Thus

$$B\xi = \mathbf{b}'$$

By Theorem 2.9, ξ is also a solution of $B\mathbf{x} = \tilde{\mathbf{b}}$. Hence

$$B\xi = \tilde{\mathbf{b}}$$

Therefore

$$\tilde{\mathbf{b}} = \mathbf{b}'.$$

Hence $A' = \tilde{A}$.

$\square$

Definition 2.16. If a matrix A is row equivalent to an echelon matrix U , we call U *an echelon form (or row echelon form)* of A ; if U is in reduced echelon form, we call U *the reduced echelon form* of A.

Since the reduced echelon form is unique, *the leading entries are always in the same positions in any echelon form obtained from a given matrix.* These leading entries correspond to leading 1s in the reduced echelon form.

Proposition 2.17. *The pivot positions are the same in any echelon form of a given matrix.*

Proof. Let A be an $m \times n$ matrix. If $A = 0$ then there is only one echelon form of A, so nothing to show. Therefore let us assume that $A \neq 0$.

As the RREF of A is unique it suffices to establish the following.

Claim: Any echelon form of A can be reduced to RREF by using only row scaling and row replacement operations (these do not change the positions of the pivots). We prove this claim using induction on n.

When $n = 1$, an echelon form of A must be a multiple of $\mathbf{e}_1$. The RREF of A is $\mathbf{e}_1$, therefore the reduction to RREF requires only a scaling operation.

Let us assume that the claim is true for $n \leq k$ and prove that it must hold true for $n = k + 1$.

Let A' be the RREF of A. Let $\tilde{A}$ be any other echelon form of A.

By the induction hypothesis, there exists a sequence of elementary row operations not including any row interchanges, that reduces $\tilde{A}$ to an echelon form $\bar{A} = (\bar{a}_{ij})$ of A, such that the first k columns of $\bar{A}$ are the same as the first k columns of A'.

If the $k + 1$-th column of $\bar{A}$ is not a pivot column, then $\bar{A}$ is the RREF of A, and therefore $\bar{A} = A'$. So no further row operations are necessary in reducing to RREF.

Suppose the $k + 1$-th column of $\bar{A}$ is a pivot column. Let l be the row containing the last pivot position in $\bar{A}$. Let E_0 be the $m \times m$ elementary matrix which corresponds to the scaling operation $R_l \to \frac{1}{\bar{a}_{l,k+1}}$. Let E_j be $m \times m$ elementary matrices corresponding to row replacement operations $R_j \to R_j - \bar{a}_{j,k+1} R_l$, for $j = 1, \ldots, l - 1$. Then

$$A' = E_{l-1} E_{l-1} \ldots E_0 \bar{A}$$

Hence the claim. □

Definition 2.18. A *pivot position* in a matrix A is a location in A that corresponds to a leading 1 in the reduced echelon form of A . A *pivot column* is a column of A that contains a pivot position.

Homogeneous Linear Systems

Definition 2.19. A system of linear equations is said to be *homogeneous* if it can be written in the form

$$A\mathbf{x} = \mathbf{0},$$

where A is an $m \times n$ matrix.

Theorem 2.20. *Suppose the equation*

$$A\mathbf{x} = \mathbf{b}$$

is consistent for some given $\mathbf{b}$*, and let* $\mathbf{p}$ *be a solution (i.e.* $A\mathbf{p} = \mathbf{b}$*). Then the solution set of* $A\mathbf{x} = \mathbf{b}$ *is the set of all vectors of the form*

$$\mathbf{w} = \mathbf{p} + \mathbf{v_h},$$

where $\mathbf{v_h}$ *is any solution of the homogeneous equation*

$$A\mathbf{x} = \mathbf{0}.$$

Proof. Suppose $\mathbf{w}$ is a solution of the equation $A\mathbf{x} = \mathbf{b}$. This means

$$A\mathbf{w} = \mathbf{b}. \tag{7}$$

We also know that

$$A\mathbf{p} = \mathbf{b}. \tag{8}$$

If we subtract equation (8) from (7) we get

$$A(\mathbf{w} - \mathbf{p}) = \mathbf{0}.$$

This means $\mathbf{w} - \mathbf{p}$ is a solution of the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

Put $\mathbf{v_h} = \mathbf{w} - \mathbf{p}$. Then we can express $\mathbf{w}$ as

$$\mathbf{w} = \mathbf{p} + \mathbf{v_h}.$$

Conversely, suppose $\mathbf{v_h}$ is any solution of the homogeneous system $A\mathbf{x} = \mathbf{0}$. This means

$$A\mathbf{v_h} = \mathbf{0}. \tag{9}$$

Put $\mathbf{w} = \mathbf{p} + \mathbf{v_h}$. Then using equations (8) and (9) we get

$$A\mathbf{w} = \mathbf{Ap} + \mathbf{Av_h} = \mathbf{b}.$$

which means $\mathbf{w}$ is a solution of $A\mathbf{x} = \mathbf{b}$.

□

Solutions of Linear Systems

Definition 2.21. The variables corresponding to the pivot columns of the an augmented matrix of a linear system are called *basic variables.* The remaining variables are called *free variables.*

Theorem 2.22 (Existence and Uniqueness Theorem, p. 24 of [1])**.** *A linear system is consistent if and only if the rightmost column of the augmented matrix is* not *a pivot column; that is, if and only if an echelon form of the augmented matrix has* no *row of the form*

$$\begin{bmatrix} 0 & \dots & 0 & b \end{bmatrix} \text{ with } b \text{ nonzero}$$

If a linear system is consistent, then the solution set contains either

$$\begin{aligned} &1.\ \textit{a unique solution, when there are no free variables,or} \\ &2.\ \textit{infinitely many solutions, when there is at least one free variable.} \end{aligned} \tag{$*$}$$

Proof. The first statement of this theorem is an immediate consequence of Proposition 2.14, Theorem 2.9 and Proposition 2.17.

Let us next assume that a given system $A\mathbf{x} = \mathbf{b}$ is consistent, where $A = (a_{ij})$ is an $m \times n$ matrix having real entries and $\mathbf{b} \in \mathbb{R}^m$. By virtue of Theorem 2.9 and Proposition 2.17, we may assume without loss of generality that $[A \quad \mathbf{b}]$ is in reduced echelon form.

Since the free variables in the homogeneous system $A\mathbf{x} = 0$ are the same as the free variables in the system $A\mathbf{x} = \mathbf{b}$, it suffices by virtue of Theorem 2.20, to show that $(*)$ holds for the system $A\mathbf{x} = 0$.

Let us first consider the case where there are no free variable in the system $A\mathbf{x} = 0$, then every column of A must be a pivot column. Thus the columns of A are $\mathbf{e}_1, \dots, \mathbf{e}_n$.

Let $\mathbf{x} = (x_1, \dots, x_n)$ be a solution of $A\mathbf{x} = 0$. Then

$$0 = A\mathbf{x} = x_1\mathbf{e}_1 + \dots + x_n\mathbf{e}_n = \mathbf{x}$$

Thus the system $A\mathbf{x} = 0$ has only the trivial solution.

Next, suppose the system $A\mathbf{x} = 0$ has at least one free variable. Let the l-th column of A be the first column which is not a pivot column.

Suppose $l = 1$. Then the first column of A must be a zero column. Choose $c \in \mathbb{R}$ and let $\mathbf{x} = (c, \dots, 0)$. Then

$$A\mathbf{x} = 0$$

As the choice of c was arbitary, $A\mathbf{x} = 0$ has infinitely many solutions.

Next, suppose $l > 1$. Then the first $l - 1$ columns of A must be $\mathbf{e}_1, \dots, \mathbf{e}_{l-1}$ and the l-th column of A must be

$$a_{1l}\mathbf{e}_1 + a_{2l}\mathbf{e}_2 + \dots + a_{l-1,l}\mathbf{e}_{l-1}.$$

Choose $c \in \mathbb{R}$ and let $\mathbf{x} = (-ca_{1l}, -ca_{2l}, \dots, -ca_{l-1,l}, c, \dots, 0)$. Then

$$A\mathbf{x} = 0$$

As the choice of c was arbitary, $A\mathbf{x} = 0$ has infinitely many solutions.

If A and $\mathbf{b}$ have complex entries, then the same proof works, if we replace $\mathbb{R}$ with $\mathbb{C}$ everywhere in the proof.

□

Theorem 2.23. *An $n \times n$ matrix A is invertible if and only if A is row equivalent to I , and in this case, any sequence of elementary row operations that reduces A to I also transforms I into A^{-1}.*

Proof. Suppose A is row equivalent to I. By Lemma 2.12, there exists an invertible matrix E such that

$$EA = I$$

Multiplying on both sides by E^{-1}, we get

$$E^{-1}(EA) = E^{-1}I \implies A = E^{-1}$$

Hence A is invertible.

Conversely, suppose A is invertible. Then $A\mathbf{x} = 0 \implies A^{-1}A\mathbf{x} = 0 \implies \mathbf{x} = 0$. As the system $A\mathbf{x} = 0$ has only the trivial solution, it has no free variables. Therefore all columns of A are pivot columns. As A is a square matrix, it follows that the RREF of A must be I.

□

Invertible Matrix Theorem

Theorem 2.24. *Let A be an $n \times n$ matrix having real entries. Then the following statements are equivalent. That is, for a given A, the statements are either all true or all false.*

(a) A is an invertible matrix.

(b) A is row equivalent to the $n \times n$ identity matrix.

(c) A has n pivot positions.

(d) The equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(e) The equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n$.

(f) There is an $n \times n$ matrix C such that $CA = I$.

(g) There is an $n \times n$ matrix D such that $AD = I$.

(h) A^T is an invertible matrix.

The above result also holds when A has complex entries, if we replace $\mathbb{R}^n$ with $\mathbb{C}^n$.

Proof. Let us first prove the theorem under the assumption that A has real entries.

We already know that (a) is equivalent to (b), and that (a) is equivalent to (h).

We will show that

1. (b) $\iff$ (c)
2. (a) $\implies$ (d)
3. (d) $\implies$ (c)
4. (a) $\implies$ (e)
5. (e) $\implies$ (c) (equivalently, not (c) $\implies$ not (e))

6. (e) $\iff$ (g)

7. (h) $\iff$ (f)

(b) $\implies$ (c): (i.e. A is row equivalent to the $n \times n$ identity matrix $\implies$ A has n pivot positions)

Proof: It is obvious that if the RREF of A is I, then A has n pivot positions.

(c) $\implies$ (b): (i.e. A has n pivot positions $\implies$ A is row equivalent to the $n \times n$ identity matrix)

Proof: If A has n pivot positions, then the pivot positions must be on the diagonal, because each pivot position is on the right of the pivot position which is in the row above it.

(a) $\implies$ (d): (i.e. A is an invertible matrix $\implies$ The equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution)

Proof: Let $\mathbf{w}$ be a solution of the equation $A\mathbf{x} = \mathbf{0}$. Then

$$A\mathbf{w} = \mathbf{0} \implies A^{-1}A\mathbf{w} = A^{-1}\mathbf{0} \implies \mathbf{w} = \mathbf{0}.$$

(d) $\implies$ (c): (i.e. The equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\implies$ A has n pivot positions.)

Proof: By the Existence and Uniqueness theorem, a consistent system has a unique solution if and only if there are no free variables.

$A\mathbf{x} = \mathbf{0}$ is a consistent system. If it has a unique solution (the trivial solution) then it has no free variables. As it has n basic variables, A must have n pivot columns, therefore n pivot positions.

(a) $\implies$ (e): (i.e. A is an invertible matrix $\implies$ The equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n$)

Proof: Let $\mathbf{b}$ be any vector in $\mathbb{R}^n$. Put $\mathbf{w} = A^{-1}\mathbf{b}$. Then

$$A\mathbf{w} = AA^{-1}\mathbf{b} = \mathbf{b}.$$

Therefore $\mathbf{w}$ is a solution of $A\mathbf{x} = \mathbf{b}$.

not (c) $\implies$ not (e): (i.e. A has less than n pivot positions $\implies$ there exists a vector $\mathbf{b}$ in $\mathbb{R}^n$ such that the equation $A\mathbf{x} = \mathbf{b}$ has at no solution)

Proof: If A does not have n pivot positions, then the RREF of A must have at least one row which does not contain a pivot. This can only happen if it is a row of zeros.

Let A' be the RREF of A. Suppose the i-th row of A' is a row of zeros. Let

$$I = [\mathbf{e}_1 \quad \dots \quad \mathbf{e}_n].$$

In other words, let $\mathbf{e}_1, \dots, \mathbf{e}_n$ denote the columns of the identity matrix. Then the equation

$$A'\mathbf{x} = \mathbf{e}_i$$

has no solution. This because the i-th entry of $\mathbf{e_i}$ is 1, whereas the i-th entry of the vector $A'\mathbf{x}$ is

$$A'_{i1}x_1 + A'_{i2}x_2 + \ldots + A'_{in}x_n = 0.$$

(e) $\implies$ (g): (i.e. The equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n \implies$ there is an $n \times n$ matrix D such that $AD = I$.)

Proof: For each $i = 1, \ldots, n$, let $\mathbf{d}_i$ be a solution of $A\mathbf{x} = \mathbf{e}_i$. Thus

$$A\mathbf{d}_i = \mathbf{e}_i, \quad \text{for } i = 1, \ldots, n.$$

Construct a matrix D using $\mathbf{d}_1, \ldots, \mathbf{d}_n$ as columns, i.e.

$$D = [\mathbf{d}_1 \quad \ldots \quad \mathbf{d}_n].$$

Then

$$AD = [A\mathbf{d}_1 \quad \ldots \quad A\mathbf{d}_n] = [\mathbf{e}_1 \quad \ldots \quad \mathbf{e}_n] = I.$$

(g) $\implies$ (e): (i.e. There is an $n \times n$ matrix D such that $AD = I \implies$ the equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n$)

Proof: Let $D = [\mathbf{d}_1 \quad \ldots \quad \mathbf{d}_n]$, i.e. let $\mathbf{d}_1, \ldots, \mathbf{d}_n$ be the columns of the matrix D. Then

$$AD = I \implies [A\mathbf{d}_1 \quad \ldots \quad A\mathbf{d}_n] = [\mathbf{e}_1 \quad \ldots \quad \mathbf{e}_n]$$

Thus

$$A\mathbf{d}_i = \mathbf{e}_i, \quad \text{for } i = 1, \ldots, n.$$

Now let $\mathbf{b} \in \mathbb{R}^n$ be any vector. If

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{bmatrix}$$

then

$$\mathbf{b} = b_1\mathbf{e}_1 + \ldots + b_n\mathbf{e}_n.$$

Let

$$\mathbf{w} = b_1\mathbf{d}_1 + \ldots + b_n\mathbf{d}_n.$$

Then

$$\begin{aligned} A\mathbf{w} &= A(b_1\mathbf{d}_1 + \ldots + b_n\mathbf{d}_n) \\ &= b_1A\mathbf{d}_1 + \ldots + b_nA\mathbf{d}_n \\ &= b_1\mathbf{e}_1 + \ldots + b_n\mathbf{e}_n \\ &= \mathbf{b}. \end{aligned}$$

So $\mathbf{w}$ is a solution of $A\mathbf{x} = \mathbf{b}$.

(h) $\implies$ (f): (i.e. A^T is invertible $\implies$ there is an $n \times n$ matrix C such that $CA = I$)

Proof: If A^T is invertible, then by the equivalence of parts (a) and (g) of the theorem, which we have already shown, there exists a matrix D such that $A^TD = I$. Hence

$$D^TA = (A^TD)^T = I^T = I.$$

Put $C = D^T$, and (f) follows.

(f) $\implies$ (h): (i.e. There is an $n \times n$ matrix C such that $CA = I \implies A^T$ is invertible)

Proof: If $CA = I$ then $A^T C^T = I$. Thus by the equivalence of parts (a) and (g) of the theorem, A^T is invertible.

If A has complex entries, the same proof works if we replace all occurrences of $\mathbb{R}$ with $\mathbb{C}$.

□

Linear Independence in $\mathbb{R}^n$ and Span

Definition 2.25. If $\mathbf{v}_1, \ldots, \mathbf{v}_p \in \mathbb{R}^n$, then the set of all linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_p$ is denoted by $Span\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ and is called the subset of $\mathbb{R}^n$ *spanned (or generated)* by $\mathbf{v}_1, \ldots, \mathbf{v}_p$. That is, $Span\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is the collection of all vectors that can be written in the form

$$c_1\mathbf{v}_1 + \ldots + c_p\mathbf{v}_p$$

with $c_1, \ldots, c_p$ scalars.

The span of the empty set is defined as $\{0\}$, where 0 is the zero vector in $\mathbb{R}^n$.

Definition 2.26. An indexed set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\} \in \mathbb{R}^n$ is said to be *linearly independent* if the vector equation

$$x_1\mathbf{v}_1 + \ldots + x_p\mathbf{v}_p = 0$$

has only the trivial solution. The set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is said to be *linearly dependent* if there exist weights $c_1, \ldots, c_p$, not all zero, such that

$$c_1\mathbf{v}_1 + \ldots + c_p\mathbf{v}_p = 0$$

The empty set is said to be linearly independent.

Proposition 2.27. *The columns of a matrix A are linearly independent if and only if the equation $A\mathbf{x} = 0$ has only the trivial solution.*

Proof. This is an immediate consequence of Definition 1.11 and the aforementioned definition of linear independence. □

Theorem 2.28. *If a set contains more vectors than there are entries in each vector, then the set is linearly dependent. That is, any set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ in $\mathbb{R}^n$ is linearly dependent if $p > n$.*

Proof. Let

$$V = [\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_p]$$

be the $n \times p$ matrix formed using $\mathbf{v}_1, \ldots, \mathbf{v}_p$ as columns. Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is linearly dependent if and only if the equation $V\mathbf{x} = 0$ has a nontrivial solution.

Since $p > n$ every column in the RREF of V cannot contain a pivot. Therefore there exists a free variable in the equation $V\mathbf{x} = 0$. The proposition follows. □

3 Determinants

Notation. Let A be an $n \times n$ matrix (where $n \geq 2$). A_{ij} denotes the $n-1 \times n-1$ submatrix formed by deleting the i-th row and j-th column of A, for $1 \leq i, j \leq n$.

Definition 3.1. For $n \geq 2$, the determinant of an $n \times n$ matrix $A = (a_{ij})$ is the sum of n terms of the form $\pm a_{1j}$ det A_{1j}, with plus and minus signs alternating, where the entries $a_{11}, a_{12}, \ldots, a_{1n}$ are from the first row of A. In symbols,

$$\begin{aligned} \det A &= a_{11} \det A_{11} - a_{12} \det A_{12} + \cdots + (-1)^{1+n} a_{1n} \det A_{1n} \\ &= \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det A_{1j} \end{aligned}$$

The determinant of a 1×1 matrix is the single entry of that matrix.

Definition 3.2. Let $A = (a_{ij})$ be an $n \times n$ matrix (where $n \geq 2$). The (i,j)-*cofactor of* A is the number C_{ij} given by

$$C_{ij} = (-1)^{i+j} \det A_{ij}$$

Theorem 3.3. *The determinant of an $n \times n$ matrix A can be computed by a cofactor expansion across any row or down any column. The expansion across the i-throw using the cofactors is*

$$\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$$

The cofactor expansion down the j th column is

$$\det A = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}.$$

Remark. The proof of this theorem is not in the course. Interested students can find a proof in the first chapter of [8].

Theorem 3.4. *If A is a triangular matrix, then det A is the product of the entries on the main diagonal of A.*

Proof. We first prove the result for lower triangular matrices.

Let A be a lower triangular $n \times n$ matrix. We prove the result using induction on n.

For $n = 1$, there is nothing to show.

Assume that the result holds for $1 \leq k < n$. We show that it holds for $k = n$.

Let $A = (a_{ij})$. Then

$$\det A = a_{11} \det A_{11} - a_{12} \det A_{12} + \ldots + (-1)^{1+n} \det A_{1n} = a_{11} \det A_{11}.$$

By the induction hypothesis,

$$\det A_{11} = a_{22}a_{33} \ldots a_{nn}.$$

Therefore

$$\det A = a_{11}a_{22} \ldots a_{nn}.$$

The prove is similar for the case where A is upper triangular. The only change is that instead of expanding the determinant along the first row, we expand it along the first column. $\square$

Proposition 3.5. *Let E be an $n \times n$ elementary matrix. Then*

(1) $\det E = c$*, when E corresponds to scaling a row by a nonzero scalar c*

(2) $\det E = 1$*, when E corresponds to a row replacement operation*

(3) $\det E = -1$*, when E corresponds to a row interchange operation*

Proof. For the row replacement and row scaling operations, the proposition is an immediate consequence of Theorem 3.4.

We show that $\det E = -1$, when E corresponds to a row interchange, using induction on n.

Clearly $n \geq 2$. If $n = 2$, then

$$E = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

so $\det E = -1$. Assume that the induction hypothesis is true for $2 \leq k < n$. We show that it holds when $n = k$.

Suppose E corresponds to $R_i \leftrightarrow R_j$. We expand the determinant along row l, where $l \neq i, l \neq j$. Let $E = (a_{ij})$. Then

$$\det E = (-1)^{l+1} a_{l1} \det E_{l1} + \cdots + (-1)^l + n a_{ln} \det E_{ln} = \det E_{ll}$$

Since E_{ll} is an $k-1 \times k-1$ elementary matrix corresponding to a row interchange operation, it follows by the induction hypothesis that

$$\det E_{ll} = -1.$$

Therefore $\det E = -1$. □

Lemma 3.6. *Let $E_1, \ldots, E_p$ be a sequence of $n \times n$ elementary matrices. Then*

$$\det\left(\prod_{i=1}^{p} E_i\right) = \prod_{i=1}^{p} \det E_i.$$

Proof. If $p = 2$, then $\det E_1 E_2 = \det E_1 \det E_2$, by Theorem 3.8. The general case follows by induction on p, using the associativity of matrix multiplication. □

Proposition 3.7. *Let A be an $n \times n$ matrix having two identical rows. Then the determinant of A is zero.*

Proof. Let $A = (a_{ij})$.

We first consider the case where A has two adjacent rows which are identical, say row i and row $i + 1$.

Then expanding across the i-th row, we obtain

$$\det A = (-1)^{i+1} a_{i1} \det A_{i1} + \ldots + (-1)^{i+n} a_{in} \det A_{in}.$$

Expanding across the $i+1$-th row gives us

$$\det A = (-1)^{i+2} a_{i+1,1} \det A_{i+1,1} + \ldots + (-1)^{i+n+1} a_{i+1,n} \det A_{i+1,n}.$$

Therefore

$$\det A = -\det A.$$

Hence $\det A = 0$.

We show that the proposition holds for non-adjacent rows by using induction on n.

If $n = 1$, then there is nothing to show. We have also shown that the result is true when $n = 2$. Therefore let us assume that the result holds true for every $n < k$ and show that it holds true for $n = k$ (where k is an integer greater than 2).

Let i and j be the rows of A which are identical. If these rows are adjacent, then there is nothing to show. So assume that there exists l such that $i < l < j$. Expanding across row l we get

$$\det A = (-1)^{l+1} a_{l1} \det A_{l1} + \ldots + (-1)^{l+n} a_{ln} \det A_{ln}.$$

Each of the minors A_{lm} ($m = 1, \ldots, n$) has at least two identical rows. So by the induction hypothesis,

$$\det A_{lm} = 0, \text{ for } m = 1, \ldots, n.$$

Hence $\det A = 0$. □

Theorem 3.8. *Let A be an $n \times n$ matrix. Let E be an $n \times n$ elementary matrix. Then*

$$\det(EA) = \det(E)\det(A)$$

Proof. Let $A = (a_{ij})$.

We first consider the case where E corresponds to row replacement, say $R_i \to R_i + cR_j$, where $c \in \mathbb{R}$ (or $c \in \mathbb{C}$, if A has complex entries). Expanding along the the i-th row,

$$\begin{aligned}
&\det(EA) \\
&= (-1)^{i+1}(a_{i1} + ca_{j1}) \det EA_{i1} + \ldots + (-1)^{i+n}(a_{in} + ca_{jn}) \det EA_{in} \\
&= (-1)^{i+1}(a_{i1} + ca_{j1}) \det A_{i1} + \ldots + (-1)^{i+n}(a_{in} + ca_{jn}) \det A_{in} \\
&= \det A + c((-1)^{i+1} a_{j1} \det A_{i1} + \ldots + (-1)^{i+n} a_{jn} \det A_{in})
\end{aligned}$$

But the expression

$$(-1)^{i+1} a_{j1} \det A_{i1} + \ldots + (-1)^{i+n} a_{jn} \det A_{in}$$

is nothing but the determinant of the matrix obtained by replacing the i-th row of A by the j-th row of A. We know that any matrix having identical rows has zero determinant. So

$$(-1)^{i+1} a_{j1} \det A_{i1} + \ldots + (-1)^{i+n} a_{jn} \det A_{in} = 0.$$

Hence $\det EA = \det A$. Since $\det E = 1$,

$$\det EA = \det E \det A.$$

Next, suppose E corresponds to scaling, say $R_i \to cR_i$ for some $c \in \mathbb{R}$ (or $c \in \mathbb{C}$, if A has complex entries). Then expanding along the i-th row, we get

$$\det(EA) = (-1)^{i+1} ca_{i1} \det A_{i1} + \ldots + (-1)^{i+n} ca_{in} \det A_{in} = c \det A.$$

Next, suppose E corresponds to row interchange, say $R_i \longleftrightarrow R_j$. This is equivalent to the following sequence to row operations:

1. $R_j \to R_i + R_j$: let this correspond to elementary matrix E_1
2. $R_i \to R_i - R_j$: let this correspond to elementary matrix E_2
3. $R_j \to R_i + R_j$: let this correspond to elementary matrix E_3
4. $R_i \to -R_i$: let this correspond to elementary matrix E_4

Then by Lemma 3.6,

$$\det EA = \det E_4E_3E_2E_1A = \det E_4E_3E_2A = \det E_4E_3A = \det E_4A = -\det A$$

Hence $\det EA = \det E \det A$.

□

Corollary 3.9. *Let A be a square matrix.*

(a) If a multiple of one row of A is added to another row to produce a matrix B, then $\det B = \det A$

(b) If two rows of A are interchanged to produce B, then det $B = -\det A$.

(c) If one row of A is multiplied by k to produce B, then det $B = k \cdot \det A$.

Proof. Each row operation on A corresponds to left multiplication by an elementary matrix E. Hence $B = EA$, and therefore

$$\det B = \det E \det A.$$

The conclusion follows from Proposition 3.5. □

Theorem 3.10. *Let U be an echelon form of A obtained by row replacement and row interchange operations. Let r be the number if row interchanges involved in the row reduction process. Then*

$$\det A = \begin{cases} (-1)^r \cdot \begin{pmatrix} \textit{product of} \\ \textit{pivots in } U \end{pmatrix} & \textit{when } A \textit{ is invertible} \\ 0 & \textit{when } A \textit{ is not invertible} \end{cases}$$

Proof. A square echelon matrix is upper triangular, and has all pivot positions on the main diagonal.

Therefore $\det U$ equals the product of the diagonal entries of U.

If A is invertible, then all the diagonal entries of U are pivots and therefore $\det U$ equals the product of pivots in U.

If A is not invertible, then at least one diagonal entry of U is zero and therefore $\det U = 0$.

Since U is obtained from A using r row interchanges and some row replacement operations, it follows that

$$\det U = (-1)^r \det A.$$

Thus

$$\det A = \begin{cases} (-1)^r \cdot \begin{pmatrix} \text{product of} \\ \text{pivots in } U \end{pmatrix} & \text{when } A \text{ is invertible} \\ 0 & \text{when } A \text{ is not invertible} \end{cases}$$

□

Corollary 3.11. *A square matrix A is invertible if and only if* $\det A \neq 0$.

Proof. This is an immediate consequence of Theorem 3.10. $\square$

Theorem 3.12. *Let A and B be $n \times n$ matrices. Then*

$$\det AB = \det A \det B.$$

Proof. We first consider the case where both A and B are invertible.

By Theorem 2.23, Lemma 2.12, and Proposition 2.7, there exist elementary matrices $E_1, \dots, E_p$ such that

$$A = E_1 E_2 \dots E_p$$

and $E_{p+1}, \dots, E_q$ such that

$$B = E_{p+1} \dots E_q.$$

Hence by Lemma 3.6,

$$\det AB = \det\left(\prod_{i=1}^{q} E_i\right) = \det\left(\prod_{i=1}^{p} E_i\right) \det\left(\prod_{i=p+1}^{q} E_i\right) = \det A \det B.$$

Next, suppose either A or B or both, are singular.

By Corollary 3.11, $\det A = 0$ or $\det B = 0$ (or both). Hence $\det A \det B = 0$.

Now if AB is invertible, then $AB(AB)^{-1} = I \implies A$ is invertible, and $(AB)^{-1}AB = I \implies B$ is invertible, by the Invertible Matrix Theorem. Therefore AB is not invertible.

Hence $\det AB = 0$. Therefore $\det AB = \det A \det B$. $\square$

Notation. For any $n \times n$ matrix A and any $\mathbf{b}$ in $\mathbb{R}^n$ (or $\mathbf{b}$ in $\mathbb{R}^n$ if A has complex entries), let $A_i(\mathbf{b})$ be the matrix obtained from A by replacing column i by the vector $\mathbf{b}$.

$$A_i(\mathbf{b}) = \begin{bmatrix} \mathbf{a_1} & \cdots & \mathbf{b} & \cdots & \mathbf{a}_n \end{bmatrix}$$

Lemma 3.13. *Let I be the $n \times n$ identity matrix, and let $\mathbf{x} = (x_1, \dots, x_n)$ be any vector in $\mathbb{R}^n$, or in $\mathbb{C}^n$. Then*

$$\det I_i(\mathbf{x}) = x_i.$$

Proof. Using the cofactor expansion along the i-th row, we obtain

$$\det I_i(\mathbf{x}) = (-1)^{i+i} x_i \det I_i(\mathbf{x})_{ii} = x_i \det I_{n-1} = x_i$$

as the remaining entries of the i-th row of $I_i(\mathbf{x})$ are zero.

$\square$

Theorem 3.14 (Cramer's Rule)**.** *Let A be an invertible $n \times n$ matrix. For any $\mathbf{b}$ in $\mathbb{R}^n$ or $\mathbb{C}^n$, the unique solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ has entries given by*

$$x_i = \frac{\det A_i(\mathbf{b})}{\det A}, \quad i = 1, 2, \dots, n$$

Proof. Let $\mathbf{x}$ be the unique solution of $A\mathbf{x} = \mathbf{b}$.

Let $\mathbf{a}_1, \ldots, \mathbf{a}_n$ be the columns of A.

Then

$$\begin{aligned} AI_i(\mathbf{x}) &= [A\mathbf{e}_1 \quad A\mathbf{e}_2 \quad \ldots \quad A\mathbf{x} \quad A\mathbf{e}_n] \\ &= [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \ldots \quad \mathbf{b} \quad \mathbf{a}_n] \\ &= A_i(\mathbf{b}). \end{aligned}$$

Hence

$$\det A \det I_i(\mathbf{x}) = \det A_i(\mathbf{b}).$$

Thus

$$x_i = \det I_i(\mathbf{x}) = \frac{\det A_i(\mathbf{b})}{\det A}.$$

□

Theorem 3.15 (Formula for A^{-1})**.** *Let A be an invertible matrix. Then*

$$A^{-1} = \frac{1}{\det A} \operatorname{Adj} A$$

where the Adj A is the adjugate *(or* classical adjoint*) of A, having entries*

$$(\operatorname{Adj} A)_{ij} = (-1)^{i+j} \det A_{ji}.$$

In other words, Adj A *is the transpose of the matrix formed by the cofactors of A.*

Proof. Let $\mathbf{x}_j$ be the j-th column of A^{-1}. Then

$$A\mathbf{x}_j = \mathbf{e}_j.$$

Hence the i-th entry of $\mathbf{x}_j$, and thus the i,j-th entry of A^{-1}, is

$$\frac{\det A_i(\mathbf{e}_j)}{\det A}.$$

Expanding the determinant of $A_i(\mathbf{e}_j)$ along the i-th column, we obtain

$$\det A_i(\mathbf{e}_j) = (-1)^{1+i} \det A_i(\mathbf{e}_j)_{1i} + \ldots + (-1)^{n+i} \det A_i(\mathbf{e}_j)_{ni}$$

When $k \neq j$, the submatrix $A_i(\mathbf{e}_j)_{kj}$ contains a row of zeros. Therefore

$$\det A_i(\mathbf{e}_j)_{kj} = 0, \qquad \text{for } k \neq j.$$

Hence

$$\det A_i(\mathbf{e}_j) = (-1)^{j+i} \det A_i(\mathbf{e}_j)_{ji} = (-1)^{j+i} \det A_i(\mathbf{e}_j)_{ji} = (-1)^{j+i} \det A_{ji}$$

□

4 Abstract Vector Spaces

Definition 4.1. A *vector space (real vector space)* is a nonempty set V of objects, called *vectors*, on which are defined two operations, called *addition* and *multiplication by scalars (real numbers)*, subject to the ten axioms (or rules) listed below. The axioms must hold for all vectors u, v, and w in V and for all scalars c and d.

- The sum of u and v, denoted by $u + v$, is in V.
- $u + v = v + u$
- $u + (v + w) = (u + v) + w$
- There is a *zero* vector 0 in V such that $u + 0 = u$.
- For each u in V , there is a vector $-u$ in V such that $u + (-u) = 0$.
- The scalar multiple of u by c , denoted by cu, is in V .
- $c(u + v) = cu + cv$
- $(c + d)u = cu + du$
- $c(du) = (cd)u$
- $1u = u$

Please note that instead of "real vector spaces" we can also talk about "complex vector spaces". In this case we would assume that the scalars are complex. The rest of the definition would remain the same. We will come back to this later in the course.

Examples.

1. Euclidean spaces $\mathbb{R}^n$
2. $M_{m,n}(\mathbb{R})$ - the set of all $m \times n$ matrices having real entries
3. The space $\mathbb{S}$, of all doubly infinite sequences of numbers (usually written in a row rather than a column):
$$\{y_k\} = (\ldots, y_{-2}, y_{-1}, y_0, y_1, y_2, \ldots)$$
If $\{z_k\}$ is another element of $\mathbb{S}$, then the sum $\{y_k\} + \{z_k\}$ is the sequence $\{y_k + z_k\}$ formed by adding corresponding terms of $\{y_k\}$ and $\{z_k\}$. The scalar multiple $c\{y_k\}$ is the sequence $\{cy_k\}$.

   $\mathbb{S}$ is sometimes called the space of *(discrete-time) signals*.
4. For $n \geq 0$, the set $\mathbb{P}_n$ of of polynomials of degree at most n consists of all polynomials of the form
$$p(t) = a_0 + a_1 t + a_2 t^2 + \ldots + a_n t^n \tag{$*$}$$
where the coefficients $a_0, \ldots, a_n$ are real numbers and t is a variable which takes real values. If all the coefficients are zero, p is called the *zero polynomial*.

Note: Two polynomials are defined to be equal if and only if *their coefficients are equal as ordered lists.*

The *degree* of a nonzero polynomial p is defined as the highest power of t in $(*)$ whose coefficient is not zero.

5. The space of all real valued function defined on a fixed set D.

Remark. Students are encouraged to verify that all vector space axioms hold, for each of the above examples.

Proposition 4.2. *Let V be a vector space. The zero vector in V is unique.*

Proof. Suppose if possible that there are two zero vectors, say 0 and z. Then

$$0 + z = 0, \text{ by fourth axiom}$$

$$z + 0 = z, \text{ by fourth axiom}$$

$$0 + z = z + 0 \text{ by second axiom}$$

Hence $0 = z$. ☐

Proposition 4.3. *Let V be a vector space. For every u in V there exists a unique $-u$ called the* negative of u *such that $u + (-u) = 0$.*

Proof. Let u be any vector in V. Suppose if possible that there are two vectors, say $-u$ and v, such that $u + (-u) = 0$ and $u + v = 0$ both hold. Then

$$-u = -u + 0 = -u + (u + v) = (-u + u) + v = 0 + v = v.$$

Hence $-u = v$. ☐

Proposition 4.4. *Let V be a vector space. Then*

$$0u = 0 \textit{ and } -u = (-1)u \textit{ hold for every } u \textit{ in } V, \textit{ and}$$
$$c0 = 0, \textit{ holds for every scalar } c$$

Proof. Let u be any vector in V. Let $v := 0u$. Since $0u = (0 + 0)u = 0u + 0u$, it follows that $v = v + v$. Hence

$$0 = v + (-v) = (v + v) + (-v) = v + (v + (-v)) = v + 0 = v.$$

Also, $u + (-1)u = 1u + (-1)u = (1 + (-1))u = 0u = 0$.

Next, let c be any scalar. Let $w := c0$. Since $c(0 + 0) = c0$ it follows that $w = w + w$. Hence $w = 0$.

☐

Definition 4.5. A *subspace* of a vector space V is a nonempty subset H of V that has two properties:

(a) H is closed under vector addition. That is, for each u and v in H , the sum $u + v$ is in H.

(b) H is closed under multiplication by scalars. That is, for each u in H and each scalar c, the vector cu is in H .

Theorem 4.6. *A subspace of a vector space is also a vector space, under the same operations of addition and scalar multiplication.*

Proof. The verification of the vector space axioms is routine and is left as an exercise. □

Definition 4.7. Let $v_1, \ldots, v_p$ be elements of a vector space V, and let $c_1, \ldots, c_p$ be scalars. The vector

$$c_1v_1 + \cdots + c_pv_p$$

is called a *linear combination* of the vectors $v_1, \ldots, v_p$.

Definition 4.8. Let S be a subset of a vector space V. The set of all linear combinations formed using elements of S is called the *span* of S, and is denoted by $\operatorname{Span} S$. If S is empty, the span of S is defined as $\{0\}$, where 0 is the zero vector in V.

Proposition 4.9. *Let S be a subset of V. Then* $\operatorname{Span} S$ *is a subspace of V.*

Proof. If S is empty, then $\operatorname{Span} S = \{0\}$, which is clearly a subspace of V, as it is closed under vector addition and scalar multiplication.

We first consider the case where S is a finite set. Then $S = \{v_1, \ldots, v_n\}$, for some $n \in \mathbb{N}$. Next, let $v, w \in \operatorname{Span} S$. Then there exist scalars $c_1, \ldots, c_n, d_1, \ldots, d_n$ such that

$$v = c_1v_1 + \cdots + c_nv_n$$

and

$$w = d_1v_1 + \cdots + d_nv_n.$$

$\therefore$

$$\begin{aligned} v + w &= c_1v_1 + \cdots + c_nv_n + d_1v_1 + \cdots + d_nv_n \\ &= (c_1 + d_1)v_1 + \cdots + (c_n + d_n)v_n \in \operatorname{Span} S. \end{aligned}$$

$\therefore$ $\operatorname{Span} S$ is closed under vector addition.

Next let $v \in \operatorname{Span} S$, and c be any scalar. Then exist scalars $c_1, \ldots, c_n$ such that

$$v = c_1v_1 + \cdots + c_nv_n$$

$\implies$ $cv = cc_1v_1 + \cdots + cc_nv_n \in \operatorname{Span} S$.

$\therefore$ $\operatorname{Span} S$ is closed under scalar multiplication.

$\implies$ $\operatorname{Span} S$ is a subspace of V.

Observe that if $S_1 \subset S_2 \subset V$, then

$$\operatorname{Span} S_1 \subset \operatorname{Span} S_2 \subset V.$$

We next consider the case when S is an infinite subset of V. Let $v, w \in \operatorname{Span} S$.

$\implies \exists v_1, \ldots, v_n \in S$ and scalars $c_1, \ldots, c_n$ such that

$$v = c_1 v_1 + \cdots + c_n v_n$$

and $w_1, \ldots, w_m \in S$ and scalars $d_1, \ldots, d_m$ such that

$$w = d_1 w_1 + \cdots + d_m w_n.$$

Let $S_1 = \{v_1, \ldots, v_n\}$ and $S_2 = \{w_1, \ldots, w_n\}$. Then

$$S_1 \cup S_2 \subset S \subset V.$$

As $S_1 \cup S_2$ is finite, $\operatorname{Span}(S_1 \cup S_2)$ is a subspace of V. Clearly $v, w \in \operatorname{Span}(S_1 \cup S_2)$. Hence

$$v + w \in \operatorname{Span}(S_1 \cup S_2) \subset \operatorname{Span} S$$

Hence Span S is closed under vector addition. Clearly, Span S is closed under scalar multiplication as well, and is therefore a subspace of V. □

Fundamental Subspaces associated with a Matrix

Definition 4.10. The *null space* of an $m \times n$ matrix A written as $\operatorname{Nul} A$, is the set of all solutions to the homogeneous equation $A\mathbf{x} = \mathbf{0}$. In set notation,

$$\operatorname{Nul} A = \{\mathbf{x} \in \mathbb{R}^n \ : \ A\mathbf{x} = \mathbf{0}\}$$

Proposition 4.11. *$\operatorname{Nul} A$ is a subspace of $\mathbb{R}^n$.*

Proof. (i)

$$\begin{aligned} \mathbf{x}, \mathbf{y} \in \operatorname{Nul} A &\implies A\mathbf{x} = 0 \text{ and } A\mathbf{y} = 0 \\ &\implies A(\mathbf{x} + \mathbf{y}) = 0 \\ &\implies \mathbf{x} + \mathbf{y} \in \operatorname{Nul} A \end{aligned}$$

(ii)

$$\begin{aligned} \mathbf{x} \in \operatorname{Nul} A, c \in \mathbb{R} \implies A\mathbf{x} = 0 &\implies cA\mathbf{x} = 0 \\ &\implies Ac\mathbf{x} = 0 \\ &\implies c\mathbf{x} \in \operatorname{Nul} A \end{aligned}$$

□

Definition 4.12. The *column space* of an $m \times n$ matrix A, written as $\operatorname{Col} A$, is the set of all linear combinations of the columns of A. If

$$A = [\mathbf{a}_1 \quad \dots \quad \mathbf{a}_n],$$

then

$$\operatorname{Col} A = \operatorname{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\}.$$

Proposition 4.13. $\operatorname{Col} A$ *is a subspace of* $\mathbb{R}^m$.

Proof. The span of a subset of $\mathbb{R}^m$ is a subspace of $\mathbb{R}^m$. $\square$

Definition 4.14. The *row space* of an $m \times n$ matrix A, written as $\operatorname{Row} A$, is the set of all linear combinations of the rows of A.

$$\operatorname{Row} A = \operatorname{Col} A^T$$

$\operatorname{Row} A$ is a subspace of $\mathbb{R}^n$.

$\operatorname{Nul} A, \operatorname{Col} A$ and $\operatorname{Row} A$ are called the *Fundamental Subspaces* associated with the matrix A.

Some sources also consider $\operatorname{Nul} A^T$ as a fundamental subpace associated with A.

Linear Independence/Linear Dependence

Definition 4.15. An indexed set of vectors $\{v_1, \dots, v_p\}$ in a vector space V, is said to be *linearly independent* if the vector equation

$$x_1 v_1 + \cdots + x_p v_p = 0$$

has only the trivial solution. The set $\{v_1, \dots, v_p\}$ is said to be *linearly dependent* if there exist weights $c_1, \dots, c_p$, not all zero, such that

$$c_1 v_1 + \cdots + c_p v_p = 0$$

An infinite subset S of V is said to be linearly independent, if any finite subset of S is linearly independent.

The empty set is also said to be linearly independent.

Proposition 4.16. *A set of two vectors* $\{v_1, v_2\}$ *is linearly dependent if at least one of the vectors is a multiple of the other. The set is linearly independent if and only if neither of the vectors is a multiple of the other.*

Proof. Proof of the First Statement:

Let $v_1, v_2 \in \mathbb{R}^n$. Without loss of generality assume that

$$v_1 = cv_2,$$

for some scalar $c \in \mathbb{R}$. Then

$$v_1 - cv_2 = 0 \implies v_1, v_2 \text{ are linearly dependent.}$$

Proof of the Second Statement:

Assume that $\{v_1, v_2\}$ is linearly independent. Suppose if possible that one of the vectors is a multiple of the other. Then by the first statement of the proposition, v_1, v_2 are linearly dependent. This is a contradiction. Therefore neither of the vectors is a multiple of the other.

Conversely, assume that neither of the vectors v_1, v_2 is a multiple of the other. Suppose if possible that $\{v_1, v_2\}$ is linearly dependent.

Then there exist scalars c_1, c_2 not all zero, such that

$$c_1v_1 + c_2v_2 = 0.$$

Without loss of generality, assume that $c_1 \neq 0$. Then

$$v_1 = -\frac{c_2}{c_1}v_2.$$

This is a contradiction. Therefore $\{v_1, v_2\}$ is linearly independent.

□

Proposition 4.17. *An indexed set $S = \{v_1, \ldots, v_p\}$ of two or more vectors is linearly dependent if and only if at least one of the vectors in S is a linear combination of the others. In fact, if S is linearly dependent and $v_1 \neq 0$, then some v_j (with $j > 1$) is a linear combination of the preceding vectors, $v_1, \ldots, v_{j-1}$.*

Proof. Proof of First Statement:

Assume that $\{v_1, \ldots, v_p\}$ is linearly dependent. This means that there exist scalars $c_1, \ldots, c_p$ not all zero, such that

$$c_1v_1 + c_2v_2 + \cdots + c_pv_p = 0.$$

Let $c_j \neq 0$ for some $j \in \{1, \ldots, p\}$. Then

$$c_jv_j = -c_1v_1 - c_2v_2 - \cdots - c_{j-1}v_{j-1} \implies v_j = -\frac{c_1}{c_j}v_1 - \frac{c_2}{c_j}v_2 - \cdots - \frac{c_{j-1}}{c_j}v_{j-1}.$$

Therefore v_j is a linear combination of the rother vectors in S.

Next, assume that one of the vectors in S is a linear combination of the others vectors.

$\therefore \exists j \in \{1, \ldots, p\}$ such that

$$v_j = c_1v_1 + \cdots + c_{j-1}v_{j-1} + c_{j+1}v_{j+1} + \cdots + c_pv_p$$

for some $c_1, \cdots, c_{j-1}, c_{j+1}, \ldots, c_p \in \mathbb{R}$.

Therefore

$$c_1v_1 + \cdots + c_{j-1}v_{j-1} - v_j + c_{j+1}v_{j+1} + \cdots + c_pv_p = 0.$$

Put $c_j = -1$. We obtain the linear dependence relation

$$c_1v_1 + \cdots + c_{j-1}v_{j-1} + c_jv_j + c_{j+1}v_{j+1} + \cdots + c_pv_p = 0.$$

Proof of Second Statement:

Let $v_1, \dots, v_p$ be linearly dependent and let $v_1 \neq 0$. Then there exist scalars $c_1, \dots, c_p$ not all zero, such that

$$c_1v_1 + \dots c_pv_p = 0.$$

Now let j be the largest index for which $c_j \neq 0$. Since $v_1 \neq 0, j > 1$.

$$\implies c_1v_1 + \cdots + c_jv_j = 0 \quad \text{(because } c_{j+1} = \cdots = c_p = 0\text{)}.$$

Next, as $c_j \neq 0$ we get

$$v_j = \frac{c_1}{c_j}v_1 - \cdots - \frac{c_{j-1}}{c_j}v_j.$$

□

Theorem 4.18. *Let V be a vector space. If a set $S = \{v_1, \dots, v_p\}$ in V contains the zero vector, then S is linearly dependent.*

Proof. Let $v_j = 0$ for some $j \in \{1, \dots, p\}$. Put $c_j = 1$ and $c_i = 0$ for $i = 1, \dots, j-1, j+1, \dots, p$. Then

$$c_1v_1 + \cdots + c_pv_p = 0$$

is a linear dependence relation between the vectors in S. □

Theorem 4.19. *Any subset of a linearly independent set is linearly independent.*

Proof. Let V be a vector space. Let S be a set of linearly independent vectors in V. Let D be a subset of S. If D is empty, there is nothing to show, so we may assume that D is non-empty.

If S is an infinite set and D is a finite set, then D is linearly independent as a consequence of Definition 4.15.

If D is an infinite subset of S, then any finite subset E of D is also a subset of S, and hence linearly independent, by Definition 4.15; hence D is linearly independent.

So let consider the case where $S = \{v_1, \dots, v_n\}$ is a finite set and $D \subset S$.

Let $D = \{v_{i_1}, \dots, v_{i_k}\}$ where $1 \leq i_1 < \cdots < i_k \leq n$. If $k = n$ then there is nothing to show because $D = S$. So we may assume that $k < n$.

Suppose $c_1v_{i_1} + \cdots + c_kv_{i_k} = 0$, where $c_1, \dots, c_k \in \mathbb{R}$.

Let $\{1, \dots, n\} \setminus \{i_1, \dots, i_k\} = \{j_1, \dots, j_{n-k}\}$. Let $c_{k+1} = \cdots = c_n = 0$.

Then

$$c_1v_{i_1} + \cdots + c_kv_{i_k} + c_{k+1}v_{j_1} + \dots + c_nv_{j_{n-k}} = 0$$

As $\{v_1, \dots, v_n\}$ is a linearly independent set, $c_j = 0$ for $j = 1, \dots, n$.

In particular, $c_j = 0$ for $j = 1, \dots, k$. Therefore D is linearly independent. □

Proposition 4.20. *Let $\{v_1, \ldots, v_n\}$ be a linearly independent set in a vector space V. If $w \notin \mathrm{Span}(\{v_1, \ldots, v_n\})$ then the set $\{v_1, \ldots, v_n, w\}$ is linearly independent.*

Proof. Let $c_1, \ldots, c_n, c_{n+1}$ be scalars such that

$$c_1 v_1 + \ldots + c_n v_n + c_{n+1} w = 0.$$

If $c_{n+1} \neq 0$ then

$$w = -\frac{c_1}{c_{n+1}} v_1 - \ldots - \frac{c_n}{c_{n+1}} v_n \in \mathrm{Span}(\{v_1, \ldots, v_n\}),$$

which contradicts our assumption.

Hence $c_{n+1} = 0$.
Therefore

$$c_1 v_1 + \ldots + c_n v_n = 0.$$

As $\{v_1, \ldots, v_n\}$ is a linearly independent set, $c_1 = 0, \ldots, c_n = 0$. $\square$

Basis of a Vector Space

Definition 4.21. Let V be a vector space. An ordered set of vectors $\mathcal{B} \subset V$ is said to be a *basis* of V if

(i) $\mathcal{B}$ is linearly independent.

(ii) $\mathcal{B}$ spans V.

If there exists a basis for V which has finitely many elements, we say V is *finite dimensional.* Otherwise we say V is *infinite dimensional.*

Remark. The symbol $\{\}$ is normally used to denote a *set.* A set is not assumed to be ordered. For example,

$$\{1, 2, 3\} = \{2, 1, 3\}.$$

However, when we say $\mathcal{B} = \{v_1, \ldots, v_n\}$ is a basis, we will use the curly brackets to denote an *ordered set.*

This is an *abuse of notation*, as it is non-standard. However, this is the notation used in the textbook, as well as in a standard reference texts - Hoffman, Kunze.

Theorem 4.22. *The columns $\mathbf{e}_1, \ldots, \mathbf{e_n}$ of the $n \times n$ identity matrix form a basis of $\mathbb{R}^n$.*

Proof. Let $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Then

$$\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{e}_j \in \mathrm{Span}\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}.$$

Therefore $\mathbb{R}^n \subset \mathrm{Span}\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$. Clearly, $\mathrm{Span}\{\mathbf{e}_1, \ldots, \mathbf{e}_n\} \subset \mathbb{R}^n$. Hence $\mathrm{Span}\{\mathbf{e}_1, \ldots, \mathbf{e}_n\} = \mathbb{R}^n$.

Next, $\sum_{j=1}^{n} c_j \mathbf{e}_j = 0 \implies (c_1, \ldots, c_n) = 0 \implies c_j = 0$, for each $j = 1, \ldots, n$. Hence $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a linearly independent set. $\square$

Lemma 4.23. *Let A be an $m \times n$ matrix in reduced echelon form, having k pivot columns, where $1 \le k \le m$. Then $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ is a basis for* $\operatorname{Col} A$.

Proof. Let $\mathbf{b} \in \operatorname{Col} A$. Then $\mathbf{b}$ is a linear combination of the columns of A.

Hence by Definition 1.11, the equation $A\mathbf{x} = \mathbf{b}$ has a solution. Hence $\mathbf{b}$ is not a pivot column of the augmented matrix $[A \quad \mathbf{b}]$. Thus the last $m - k$ entries of $\mathbf{b}$ must be zero.

Hence $\mathbf{b} \in \operatorname{Span}\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$. As the choice of $\mathbf{b}$ was arbitrary, $\operatorname{Col} A \subset \operatorname{Span}\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$.

As $\mathbf{e}_1, \dots, \mathbf{e}_k$ are the pivot columns of A, $\operatorname{Span}\{\mathbf{e}_1, \dots, \mathbf{e}_k\} \subset \operatorname{Col} A$.

Therefore $\operatorname{Span}\{\mathbf{e}_1, \dots, \mathbf{e}_k\} = \operatorname{Col} A$. As $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ is a subset of $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, it is linearly independent.

$\square$

Theorem 4.24. *Let A be an $m \times n$ matrix. The pivot columns of A form a basis for* $\operatorname{Col} A$.

Proof. Let $\mathbf{a}_1, \dots, \mathbf{a}_n$ be the columns of A, i.e.

$$A = [\mathbf{a}_1 \quad \dots \quad \mathbf{a}_n].$$

Let $\mathbf{a}_{i_1}, \dots, \mathbf{a}_k$ be the pivot columns of A, where $1 \le i_1 < \dots < i_k \le m$.

Let A' be the RREF of A. By Lemma 2.12, there exists an invertible $m \times m$ matrix E such that

$$A' = EA.$$

Then

$$E\mathbf{a}_{i_j} = \mathbf{e}_j, \qquad \text{for } j = 1, \dots, k.$$

Let us first show that $\{\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_k}\}$ is a linearly independent set.

So suppose $c_1\mathbf{a}_{i_1} + \dots + c_k\mathbf{a}_k = 0$. Then

$$\begin{aligned} 0 &= E(c_1\mathbf{a}_{i_1} + \dots + c_k\mathbf{a}_k) \\ &= c_1E\mathbf{a}_{i_1} + \dots + c_kE\mathbf{a}_k \\ &= c_1\mathbf{e}_1 + \dots + c_k\mathbf{e}_k \end{aligned}$$

As $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ is a linearly independent set, it follows that $c_1 = \dots = c_k = 0$.

Next, let us show that $\operatorname{Col} A = \operatorname{Span}\{\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_k}\}$.

Clearly $\operatorname{Span}\{\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_k}\} \subset \operatorname{Col} A$.

Let $\mathbf{b} \in \operatorname{Col} A$. Then there exists $\xi \in \mathbb{R}^n$ such that $A\xi = \mathbf{b}$.

Then $EA\xi = E\mathbf{b} \implies A'\xi = E\mathbf{b} \implies E\mathbf{b} \in \operatorname{Col} A'$.

By Lemma 4.23, there exist $d_1, \dots, d_k \in \mathbb{R}$ such that

$$E\mathbf{b} = d_1\mathbf{e}_1 + \ldots + d_k\mathbf{e}_k$$

Hence

$$\begin{aligned} \mathbf{b} &= E^{-1}(d_1\mathbf{e}_1 + \ldots + d_k\mathbf{e}_k) \\ &= d_1E^{-1}\mathbf{e}_1 + \ldots + d_kE^{-1}\mathbf{e}_k \\ &= d_1\mathbf{a}_{i_1} + \ldots + d_k\mathbf{a}_{i_k} \in \operatorname{Span}\{\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_k}\} \end{aligned}$$

Hence $\operatorname{Span}\{\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_k}\} = \operatorname{Col} A$.

$\square$

Theorem 4.25 (Spanning Set Theorem)**.** *Let V be a vector space. Let $S = \{v_1, v_2, \dots, v_p\}$ be a set in V and let $H = \operatorname{Span}\{v_1, v_2, \dots, v_p\}$.*

1. *If one of the vectors in S, say v_k, is a linear combination of the remaining vectors in S, then the set formed from S by removing v_k still spans H.*

2. *If $H \neq \{0\}$, some subset of S is a basis for H.*

Proof. Proof of Statement 1:

Let $d_1, d_2, \dots, d_{k-1}, d_{k+1}, \dots, d_p$ be scalars such that

$$v_k = d_1v_1 + d_2v_2 + \ldots + d_{k-1}v_{k-1} + d_{k+1}v_{k+1} + \ldots + d_pv_p.$$

Let w be any vector in $H = \operatorname{Span}\{v_1, v_2, \dots, v_p\}$. There exist scalars $c_1, \dots, c_p$ such that

$$w = c_1v_1 + c_2v_2 + \ldots + c_pv_p.$$

Then

$$\begin{aligned} w &= c_1v_1 + c_2v_2 + \ldots + c_pv_p \\ &= c_1v_1 + c_2v_2 + \ldots + c_{k-1}v_{k-1} + c_kv_k + c_{k+1}v_{k+1} \ldots + c_pv_p \\ &= c_1v_1 + c_2v_2 + \ldots + c_{k-1}v_{k-1} \\ &+ c_k(d_1v_1 + d_2v_2 + \ldots + d_{k-1}v_{k-1} + d_{k+1}v_{k+1} + \ldots + d_pv_p) \\ &+ c_{k+1}v_{k+1} \ldots + c_pv_p \\ &= (c_1 + c_kd_1)v_1 + \ldots + (c_{k-1} + c_kd_{k-1})v_{k-1} \\ &+ (c_{k+1} + c_kd_{k+1})v_{k+1} + \ldots + (c_p + c_kd_p)v_p \end{aligned}$$

is in $\operatorname{Span}\{v_1, v_2, \dots, v_{k-1}, v_{k+1}, \dots, v_p\}$.

Proof of Statement 2:

Let $\mathcal{B}$ be a linearly independent subset of S having maximum cardinality.

Suppose if possible that $S \not\subset \operatorname{Span}\mathcal{B}$. Then there exists $w \in S$ such that $w \notin \operatorname{Span}\mathcal{B}$.

By the Proposition 4.20, $\mathcal{B} \cup \{w\}$ is linearly independent.

But this contradicts the fact that $\mathcal{B}$ is a linearly independent subset of maximum cardinality. Hence

$$S \subset \operatorname{Span}\mathcal{B}.$$

Therefore

$$H = \operatorname{Span} S \subset \operatorname{Span}\mathcal{B}.$$

Thus $\mathcal{B}$ is a basis for H.

□

5 Linear Transformations

Definition 5.1. Let V, W be vector spaces. A function $T : V \to W$ is said to be a *linear transformation* if

(i) $T(v + w) = T(v) + T(w), \quad \forall v, w \in V$

(ii) $T(cv) = cT(v) \quad \forall v \in V, c \in \mathbb{R}$

Example 5.2. If A is an $m \times n$ matrix then the *matrix transformation* $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by

$$T(\mathbf{x}) = A\mathbf{x}$$

is a linear transformation. It is easy to verify that

(i) If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ then $T(\mathbf{x} + \mathbf{y}) = A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y} = T(\mathbf{x}) + T(\mathbf{y})$

(ii) If $\mathbf{x} \in \mathbb{R}^n$ and $c \in \mathbb{R}$ then $T(c\mathbf{x}) = A(c\mathbf{x}) = cA\mathbf{x} = cT(\mathbf{x})$.

Definition 5.3. Let V, W be vector spaces. Let $T : V \to W$ be a linear transformation. We define the *kernel* of T to be the set

$$\ker T := \{x \in V \ : \ T(x) = 0\}.$$

Proposition 5.4. $\ker T$ *is a subspace of* V*.*

Proof. (i)

$$\begin{aligned} x, y \in \ker T &\implies Tx = 0 \text{ and } Ty = 0 \\ &\implies T(x + y) = 0 \\ &\implies x + y \in \ker T \end{aligned}$$

(ii)

$$\begin{aligned} x \in \ker T, c \in \mathbb{R} \implies Tx = 0 &\implies cTx = 0 \\ &\implies T(cx) = 0 \\ &\implies cx \in \ker T \end{aligned}$$

□

(Please note that the kernel of T is sometimes also called the nullspace of T. These words are often used interchangeably in the literature.)

Definition 5.5. Let V, W be vector spaces. Let $T : V \to W$ be a linear transformation. We define the *range* of T to be the set

$$\operatorname{range} T := \{x \in W \ : \ x = T(v) \text{ for some } v \in V\}.$$

Proposition 5.6. $\operatorname{range} T$ *is a subspace of* W*.*

Proof. (i)

$$\begin{aligned} x, y \in \operatorname{range} T &\implies x = Tv, y = Tw \text{ for some } v, w \in V \\ &\implies x + y = Tv + Tw = T(v + w) \quad (\because T \text{ is a linear map}) \\ &\implies x + y \in \operatorname{range} T \end{aligned}$$

(ii)

$$\begin{aligned} x \in \operatorname{range} T, c \in \mathbb{R} &\implies x = Tv \text{ for some } v \in V \\ &\implies cx = cTv = T(cv) \quad (\because T \text{ is a linear map}) \\ &\implies cx \in \operatorname{range} T \end{aligned}$$

□

(Please note that the range of T is also denoted by $R(T)$ and sometimes also by $T(V)$.)

Proposition 5.7. *Let V and W be vector spaces. Let $T : V \to W$ be a linear transformation.*

$$T(0) = 0.$$

Proof.

$$T(0) = T(0 + 0) = T(0) + T(0).$$

If we subtract $T(0)$ from both sides we get

$$T(0) = 0.$$

□

Proposition 5.8. *Let V and W be vector spaces. Let $T : V \to W$ be a linear transformation. If $\{v_1, \ldots, v_p\}$ is a linearly dependent subset of V then $\{T(v_1), \ldots, T(v_p)\}$ is a linearly dependent subset of W.*

Proof. As $\{v_1, \ldots, v_p\}$ is linearly dependent, there exist scalars $c_1, \ldots, c_p$ such that

$$c_1 v_1 + \ldots + c_p v_p = 0.$$

Therefore

$$T(c_1 v_1 + \ldots + c_p v_p) = T(0).$$

Therefore $c_1 T(v_1) + \ldots + c_p T(v_p) = 0$. □

Proposition 5.9. *Let V and W be vector spaces. Let $T : V \to W$ be a linear transformation.*

T is 1-1 (injective) if and only if $\ker T = \{0\}$.

Proof. Let T be a 1-1 map. Let $v \in \ker T$. Then

$$T(v) = 0.$$

But we know that $T(0) = 0$. Therefore, as T is 1-1, we must have

$$v = 0.$$

As the choice of v was arbitrary, $\ker T = \{0\}$.

Conversely, suppose $\ker T = \{0\}$.

Suppose $T(v) = T(w)$, where $v, w \in V$.

Then $T(v) - T(w) = 0$.

Hence $T(v - w) = 0 \implies v - w \in \ker T$.

Therefore $v - w = 0$. So $v = w$.

As $T(v) = T(w) \implies v = w$, it follows that T is injective. $\square$

Proposition 5.10. *Let V and W be vector spaces. Let $T : V \to W$ be a 1-1 linear transformation.*

Then if $\{v_1, \ldots, v_n\}$ is a linearly independent subset of V then $\{T(v_1), \ldots, T(v_n)\}$ is a linearly independent subset of W.

Proof. Suppose $c_1 T v_1 + \cdots + c_n T v_n = 0$, where $c_1, \ldots, c_n \in \mathbb{R}$.

As T is a linear map,

$$T(c_1 v_1 + \cdots + c_n v_n) = c_1 T v_1 + \cdots + c_n T v_n = 0.$$

Hence $c_1 v_1 + \cdots + c_n v_n \in \ker T$. As T is 1-1, $\ker T = \{0\}$. Therefore

$$c_1 v_1 + \cdots + c_n v_n = 0.$$

As $\{v_1, \ldots, v_n\}$ is linearly independent, it follows that $c_1 = \cdots = c_n = 0$.

$\square$

Proposition 5.11. *Let U, V and W be vector spaces. Let $T : U \to V$ and $S : V \to W$ be linear transformations. Then the composite $S \circ T : U \to W$ is also a linear transformation.*

Proof. (i) Let $v, w \in U$. Then

$$\begin{aligned} S \circ T(v + w) &= S(T(v + w)) \\ &= S(Tv + Tw) \quad (\because T \text{ is linear}) \\ &= S(Tv) + S(Tw) \quad (\because S \text{ is linear}) \\ &= S \circ T(v) + S \circ T(w) \end{aligned}$$

(ii) Let $v \in U, c \in \mathbb{R}$. Then

$$\begin{aligned} S \circ T(cv) &= S(T(cv)) \\ &= S(cTv) \quad (\because T \text{ is linear}) \\ &= cS(Tv) \quad (\because S \text{ is linear}) \\ &= cS \circ T(v) \end{aligned}$$

$\square$

Definition 5.12. Let V be a vector space. The identity map $id_V : V \to V$ which send every vector to itself, i.e. $v \mapsto v$, is a linear transformation called the *identity transformation.*

Remark. The verification that the identity transformation is linear is routine and is left as an exercise.

Definition 5.13. Let V, W be vector spaces. A linear transformation $T : V \to W$ is said to be *invertible* if there exists a linear transformation $S : W \to V$ such that $S \circ T$ is the identity transformation on V and $T \circ S$ is the identity transformation on W.

Such a transformation, if it exists, is called the *inverse* of T and is denoted by T^{-1}.

Proposition 5.14. *The composite of two invertible linear transformations is an invertible linear transformation.*

Proof. Let U, V, W be vector spaces. Let $T_1 : U \to V$ and $T_2 : V \to W$ be invertible linear transormations.

As the composite of two bijective functions is bijective, $T_2 \circ T_1$ is bijective and

$$(T_2 \circ T_1)^{-1} = T_1^{-1} \circ T_2^{-1}.$$

As the composite of linear mappings is linear, both $T_2 \circ T_1$ and $(T_2 \circ T_1)^{-1}$ are linear maps. □

Lemma 5.15. *Let A be an $m \times n$ matrix. If there exists an $n \times m$ matrix B such that $AB = I_m$ and $BA = I_n$, then $m = n$ and $B = A^{-1}$.*

Proof. Suppose if possible that $m < n$.

Then there exists a nontrivial solution of the system $A\mathbf{x} = 0$, i.e. $\exists \xi \neq 0$, such that $A\xi = 0$.

Now, $A\xi = 0 \implies BA\xi = 0$.

But $BA = I_n \implies BA\xi = \xi \neq 0$. Contradiction.

Therefore $m \geq n$. A similar argument shows that $m \leq n$. Therefore $m = n$. Hence $B = A^{-1}$ (by the definition of inverse).

□

Proposition 5.16. *Every linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is a matrix transformation. In other words, there exists a unique $m \times n$ matrix A such that*

$$T(\mathbf{x}) = A\mathbf{x}, \quad \forall \mathbf{x} \in \mathbb{R}^n$$

A is called the standard matrix for the linear transformation T.

(a) If T is the identity transformation then $A = I$.

(b) T is 1-1 if and only if the columns of A are linearly independent.

(c) T is onto if and only if $Col\, A = \mathbb{R}^m$.

(d) T is invertible if and only if $m = n$, and A is invertible.

Proof. Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be the standard basis of $\mathbb{R}^n$.

Define $\mathbf{a}_j = T(\mathbf{e}_j)$ for $j = 1, \ldots, n$.

Define $A = [\mathbf{a}_1 \quad \ldots \quad \mathbf{a}_n]$, i.e. the matrix formed using the vectors $\mathbf{a}_j$ as its columns.

Claim: A is the standard matrix of T.

Let $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Then

$$\begin{aligned} T(\mathbf{x}) &= T\left(\sum_{j=1}^{n} x_j \mathbf{e}_j\right) \\ &= \sum_{j=1}^{n} x_j T(\mathbf{e}_j) \\ &= \sum_{j=1}^{n} x_j \mathbf{a}_j \\ &= A\mathbf{x} \end{aligned}$$

Suppose if possible that there exists another matrix B such that

$$T(\mathbf{x}) = B\mathbf{x}, \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

Then for $j = 1, \ldots, n$, we have $T(\mathbf{e}_j) = B\mathbf{e}_j$.

Hence $B\mathbf{e}_j = \mathbf{a}_j$ for $j = 1, \ldots, n$. Thus the columns of matrix B are the same as the columns of matrix A. Therefore $A = B$.

(a) If T is the identity map, then

$$\mathbf{a}_j = T(\mathbf{e}_j) = \mathbf{e}_j, \qquad \text{for } j = 1, \ldots, n.$$

Therefore $A = I$.

(b) We know that $A\mathbf{x} = 0$ has only the trivial solution if and only if the columns of A are linearly independent.

Therefore it suffices to show that T is 1-1 if and only if $A\mathbf{x} = 0$ has only the trivial solution.

By Proposition 5.9, T is 1-1 iff $\ker T = \{0\}$.

Now $\ker T = \{0\}$ if and only if $T(\mathbf{x}) = 0$ has only the trivial solution.

$T(\mathbf{x}) = 0$ has only the trivial solution if and only if $A\mathbf{x} = 0$ has only the trivial solution, since A is the matrix of T.

Hence T is 1-1 if and only if $A\mathbf{x} = 0$ has only the trivial solution.

(c) Recall that $A\mathbf{x}$ is the linear combination of columns of A formed by taking the entries of $\mathbf{x}$ as coefficients (Definition 1.11). Therefore,

$$\begin{aligned} T \text{ is onto } &\iff \text{for every } \mathbf{y} \in \mathbb{R}^m, \text{ there exists } \mathbf{x} \in \mathbb{R}^n, \text{ such that } A\mathbf{x} = \mathbf{y} \\ &\iff \text{for every } \mathbf{y} \in \mathbb{R}^m, \mathbf{y} \in \operatorname{Col} A \\ &\iff \operatorname{Col} A = \mathbb{R}^m \end{aligned}$$

(d) Assume that T is invertible.

Then there exists $S : \mathbb{R}^m \to \mathbb{R}^n$ such that $S \circ T$ is the identity map on $\mathbb{R}^n$ and $T \circ S$ is the identity map on $\mathbb{R}^m$.

Let B be the standard matrix of S. Then BA is the standard matrix of $S \circ T$ and AB is the standard matrix of $T \circ S$.

Thus $BA = I_n$ and $AB = I_m$. By Lemma 5.15, $m = n$ and A is invertible.

Conversely, suppose $m = n$ and A is invertible. Define $S : \mathbb{R}^n \to \mathbb{R}^n$ by

$$S(\mathbf{x}) = A^{-1}\mathbf{x}$$

Then both $S \circ T$ and $T \circ S$ are the identity transformation on $\mathbb{R}^n$.

□

Coordinates with respect to a Basis

Theorem 5.17 (Unique Representation Theorem)**.** *Let $\mathcal{B} = \{b_1, \ldots, b_n\}$ be an ordered basis for a vector space V. Then for each $x \in V$, there exists a unique n-tuple of scalars $(c_1, \ldots, c_n)$ such that*

$$x = c_1 b_1 + \ldots + c_n b_n$$

Proof. Let $x \in V$. Since the vectors $b_1, \ldots, b_n$ span V, there exists an n-tuple $(c_1, \ldots, c_n)$ of real numbers such that

$$x = c_1 b_1 + \cdots + c_n b_n.$$

Suppose if possible that another such n-tuple, $(d_1, \ldots, d_n)$ exists, i.e. such that

$$x = d_1 b_1 + \cdots + d_n b_n.$$

Then

$$0 = x - x = (c_1 - d_1)b_1 + \cdots + (c_n - d_n)b_n.$$

As the vectors $b_1, \ldots, b_n$ are linearly independent, it follows that $c_j - d_j = 0$ for $j = 1, \ldots, n$.

Therefore

$$c_j = d_j, \quad \text{for } j = 1, \ldots, n.$$

Hence the n-tuple $(c_1, \ldots, c_n)$ is unique. □

Definition 5.18. Suppose $\mathcal{B} = \{b_1, \ldots, b_n\}$ is an ordered basis for V and $x \in V$. The *coordinates of x relative to $\mathcal{B}$*(or the *$\mathcal{B}$-coordinates of x*) are the weights $c_1, \ldots, c_n$ such that

$$x = c_1 b_1 + \ldots + c_n b_n.$$

The vector $(c_1, \ldots, c_n) \in \mathbb{R}^n$ is denoted by $[x]_{\mathcal{B}}$, and is called the *coordinate vector of x relative to $\mathcal{B}$* or the *$\mathcal{B}$-coordinate vector of x*. The mapping

$$x \to [x]_{\mathcal{B}}$$

is called the *coordinate mapping (determined by $\mathcal{B}$)*.

Theorem 5.19. *Let $\mathcal{B} = \{b_1, \ldots, b_n\}$ be an ordered basis for a vector space V. The coordinate mapping*

$$x \to [x]_{\mathcal{B}}$$

is an invertible linear transformation from V to $\mathbb{R}^n$.

Proof. Let $T : V \to \mathbb{R}^n$ be the coordinate mapping, which sends $x \to [x]_{\mathcal{B}}$, for every $x \in V$.

Define $S : \mathbb{R}^n \to V$ as follows:

If $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, then

$$S(\mathbf{x}) := x_1 b_1 + \cdots + x_n b_n.$$

Claim: $S \circ T = \mathrm{id}_V$ and $T \circ S = \mathrm{id}_{\mathbb{R}^n}$.

Let $v \in V$. Let $T(v) = (c_1, \ldots, c_n)$. Then

$$v = c_1 b_1 + \cdots + c_n b_n.$$

Now, $S \circ T(v) = S(T(v)) = S(c_1, \ldots, c_n) = c_1 b_1 + \cdots + c_n b_n$.

Therefore $S \circ T(v) = v$. As the choice of $v \in V$ was arbitrary, $S \circ T = \mathrm{id}_V$.

Next, let $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Then

$$\begin{aligned} T \circ S(\mathbf{x}) &= T(S(x_1, \ldots, x_n)) \\ &= T(x_1 b_1 + \cdots + x_n b_n) \\ &= (x_1, \ldots, x_n) = \mathbf{x}. \end{aligned}$$

As the choice of $\mathbf{x} \in \mathbb{R}^n$ was arbitrary, $T \circ S = \mathrm{id}_{\mathbb{R}^n}$. $\square$

Theorem 5.20. *Let V be a finite dimensional vector space. Any two bases of V must have the same cardinality.*

Proof. Let $\mathcal{B} = \{b_1, \ldots, b_n\}$ and $\tilde{\mathcal{B}} = \{b_1, \ldots, b_m\}$ be two bases for V.

Let $T : V \to \mathbb{R}^n$ be the coordinate transformation which sends $x \to [x]_{\mathcal{B}}$ for every $x \in V$, and let $\tilde{T} : V \to \mathbb{R}^m$ be the coordinate transformation which sends $x \to [x]_{\tilde{\mathcal{B}}}$ for every $x \in V$.

As both T and $\tilde{T}$ are invertible, the composite map $\tilde{T} \circ T^{-1} : \mathbb{R}^n \to \mathbb{R}^m$ is invertible.

Therefore, by Proposition 5.16, $m = n$. $\square$

Definition 5.21. The number of elements in any basis of a finite dimensional vector space V, is called the *dimension* of V. If V is infinite-dimensional, we say its dimension is infinity. If $V = \{0\}$, we define $\dim V = 0$.

Remark. Let V, W be vector spaces. Let $T : V \to W$ be a linear transformation. Then T is invertible if and only if T is 1-1 and onto.

(This is true for bijective maps, in general.)

Proposition 5.22. *Let V be a vector space of dimension n. Any subset of V which contains more than n vectors must be linearly dependent.*

Proof. Let $S = \{v_1, \ldots, v_m\}$, where $m > n$.

Let $\mathcal{B} = \{b_1, \ldots, b_n\}$ be an ordered basis for V. Let $T : V \to \mathbb{R}^n$ be the coordinate mapping which sends

$$v \mapsto [v]_{\mathcal{B}}.$$

Since $m > n$, the vectors $T(v_1), \ldots, T(v_m)$ are linearly dependent in $\mathbb{R}^n$. Therefore there exist scalars $c_1, \ldots, c_m$, not all zero, such that

$$c_1 T(v_1) + \ldots + c_m T(v_m) = 0.$$

Since T is linear, it follows that

$$T(c_1 v_1 + \ldots + c_m v_m) = 0.$$

Hence $c_1 v_1 + \ldots + c_m v_m \in \ker T$.

As T is 1-1, $\ker T = \{0\}$. Therefore

$$c_1 v_1 + \ldots + c_m v_m = 0.$$

Hence S is a linearly dependent set.

□

Proposition 5.23. *Let V be a finite dimensional vector space. Let $S = \{v_1, \ldots, v_n\}$ be a linearly independent subset of V.*

If $n = \dim V$ then S is a basis for V.

Proof. Let $v \in V$ be any vector. If $v \notin \operatorname{Span} S$, then $S \cup \{v\}$ is linearly independent and contains more than n vectors. This is a contradiction.

Hence $v \in \operatorname{Span} S$. As the choice of v was arbitrary, $V \subset \operatorname{Span} S$.

Clearly, $\operatorname{Span} S \subset V$. Therefore $\operatorname{Span} S = V$. Therefore S is a basis for V.

□

Proposition 5.24. *Let V be a finite dimensional vector space. Let S be a linearly independent subset of V having at most finitely many elements, $v_1, \ldots, v_n$ (If S is empty, then let $n = 0$).*

If $n < \dim V$, then S can be extended to a basis for V.

Proof. Let $\dim V = m$.

For each $j = n+1, \ldots, m$, choose $v_j \in V$ such that $v_j \notin \operatorname{Span}\{v_1, \ldots, v_{j-1}\}$

(such a vector v_j exists, because $j - 1 < m \implies \operatorname{Span}\{v_1, \ldots, v_{j-1}\} \subsetneq V$).

Then $\{v_1, \ldots, v_m\}$ is linearly independent and contains exactly m vectors.

Therefore, by Proposition 5.23, $\{v_1, \ldots, v_m\}$ is a basis for V. ☐

The proof of the following theorem is not part of the course, but only given here for the sake of completeness.

Theorem 5.25. *Every vector space V, has a basis. Moreover, any linearly independent subset of V can be extended to a basis.*

Proof. Let V be a vector space. Let $\mathscr{S}$ be the collection of all linearly independent subsets of $\mathscr{S}$.

If $\mathscr{S}$ is empty then V must be the $\{0\}$ space. In this case the empty set is a basis for V.

Let us assume that $\mathscr{S}$ is not empty.

The subset relation $\subset$ is a partial order on $\mathscr{S}$.

Let $\mathscr{A}$ be a chain in $\mathscr{S}$. Define

$$A = \bigcup_{S \in \mathscr{A}} S.$$

Then A is clearly an upper bound for $\mathscr{A}$. Suppose if possible that $A \notin \mathscr{S}$. Then A contains a finite subset $\{v_1, \ldots, v_n\}$, of linearly dependent vectors.

For each $j \in \{1, \ldots, n\}$ there exists an element $S_j \in \mathscr{A}$ such that $v_j \in S_j$. Let S_m be the maximum of $S_1, \ldots, S_n$. Then

$$\{v_1, \ldots, v_n\} \subset S_n.$$

Hence S_n is not linearly independent. Contradiction. Therefore $A \in \mathscr{S}$.

Hence every chain in $\mathscr{S}$ has an upper bound in $\mathscr{S}$. Therefore by Zorn's Lemma, $\mathscr{S}$ must have a maximal element, say $\mathcal{B}$.

Then $\mathcal{B}$ is a basis for V. For if $\operatorname{Span} \mathcal{B} \neq V$, then there exists $v \in V$ such that $v \notin \operatorname{Span} \mathcal{B}$.

But this means $\mathbb{B} \cup \{v\}$ is linearly independent, which contradicts the fact that $\mathcal{B}$ is maximal.

Next, suppose S is any linearly independent subset of V. Let $\mathscr{S}$ be the collection of all linearly independent subsets of $\mathscr{S}$ which contain S.

By a similar argument as the one above, it follows that a maximal element of $\mathscr{S}$ with respect to the partial order defined by the $\subset$ relation, is a basis containing the set S. ☐

Proposition 5.26. *Let V be a finite dimensional vector space. Let H be a subspace of V. Then any basis of H can be extended to a basis of V. Thus*

$$\dim H \leq \dim V.$$

Proof. Let $\dim V = n$.

Let $\mathcal{B}$ be a basis of H. Then $\mathcal{B}$ is a linearly independent subset of V. Therefore $\mathcal{B}$ can be extended to a basis of V.

Therefore $\mathcal{B}$ cannot have more than n elements. □

Theorem 5.27 (Rank-Nullity Theorem)**.** *Let V, W be vector spaces. Let $T : V \to W$ be a linear transformation. Then*

$$\dim \ker T + \dim \operatorname{range} T = \dim V.$$

If V is infinite dimensional then least one of the summands on left hand side is infinity (i.e. either $\ker T$ or $\operatorname{range} T$, or both, must be infinite dimensional).

Proof. We first consider the case where V is finite dimensional.

Let $\dim V = n$.

As $\ker T$ is a subspace of V, by Proposition 5.26,

$$\dim \ker T \leq n.$$

Let $\{b_1, \dots, b_m\}$ be a basis of $\ker T$, where $m \leq n$ (if $\ker T = \{0\}$, let $m = 0$).

By Proposition 5.24, $\{b_1, \dots, b_m\}$ (or the empty set, if $m = 0$) can be extended to a basis $\{b_1, \dots, b_n\}$ of V.

Claim: $\{T(b_{m+1}), \dots, T(b_n)\}$ is a basis for $\operatorname{range} T$.

We first show that the set $\{T(b_{m+1}), \dots, T(b_n)\}$ is linearly independent.

Suppose

$$c_1 T(b_{m+1}) + \cdots + c_{n-m} T(b_n) = 0.$$

Since T is linear, it follows that

$$T(c_1 b_{m+1} \cdots + c_{n-m} b_n) = 0.$$

Hence

$$c_1 b_{m+1} \cdots + c_{n-m} b_n \in \ker T = \operatorname{Span}\{b_1, \dots, b_m\}.$$

Therefore, there exist scalars $d_1, \dots, d_m$ such that

$$c_1 b_{m+1} \cdots + c_{n-m} b_n = d_1 b_1 + \cdots + d_m b_m.$$

Hence

$$d_1 b_1 + \cdots + d_m b_m - c_1 b_{m+1} \cdots - c_{n-m} b_n = 0.$$

As $b_1, \dots, b_m$ are linearly independent, it follows that

$$c_1 = \cdots = c_{n-m} = 0.$$

Therefore the vectors $T(b_{m+1}), \dots, T(b_n)$ are linearly independent.

Next, let $w \in \operatorname{range} T$. Choose $v \in V$ such that $T(v) = w$. Then there exist scalars $c_1, \ldots, c_n$ such that

$$v = c_1 b_1 + \cdots + c_n b_n.$$

Now,

$$\begin{aligned} w &= T(v) \\ &= T(c_1 b_1 + \cdots + c_n b_n) \\ &= c_1 T(b_1) + \ldots + c_m T(b_m) + c_{m+1} T(b_{m+1}) + \ldots + c_n T(b_n) \\ &= c_{m+1} T(b_{m+1}) + \ldots + c_n T(b_n) \in \operatorname{Span}\{T(b_{m+1}), \ldots, T(b_n)\}, \end{aligned}$$

because $b_1, \ldots, b_m \in \ker T \implies T(b_1) = \cdots = T(b_m) = 0$. Hence the claim.

Thus the theorem holds when V is finite dimensional.

Let us next consider the case where V is infinite dimensional. If $\ker T$ is infinite dimensional, then there is nothing to show. So let us consider the case where $\ker T$ is finite dimensional.

Let $\mathcal{B} = \{b_1, \ldots, b_m\}$ be a basis of $\ker T$. Then $\mathcal{B}$ can be extended to a basis of V say $\mathcal{B}'$. In other words, $\mathcal{B} \subset \mathcal{B}'$, and $\mathcal{B}'$ is a basis of V.

As V is infinite dimensional, $\mathcal{B}'$ has infinitely many elements.

Let $p \in \mathbb{N}$. Choose $w_1, \ldots, w_p \in \mathcal{B}' \setminus \mathcal{B}$.

Suppose

$$c_1 T(w_1) + \cdots + c_p T(w_p) = 0,$$

for some $c_1, \ldots, c_p \in \mathbb{R}$. By the linearity of T,

$$T(c_1 w_1 + \cdots + c_p w_p) = 0.$$

Therefore

$$c_1 w_1 + \cdots + c_p w_p \in \ker T = \operatorname{Span} \mathcal{B}.$$

Therefore there exist scalars $d_1, \ldots, d_m$ such that

$$c_1 w_1 \cdots + c_p w_p = d_1 b_1 + \cdots + d_m b_m.$$

Hence

$$d_1 b_1 + \cdots + d_m b_m - c_1 w_1 \cdots - cpw_p = 0.$$

As $b_1, \ldots, b_m, w_1, \ldots, w_p \subset \mathcal{B}'$, they are linearly independent. It follows that

$$c_1 = \cdots = c_p = 0.$$

Therefore $\{T(w_1), \ldots, T(w_p)\}$ is a linearly independent subset of $\operatorname{range} T$.

Thus $\operatorname{range} T$ has a linearly independent subset of cardinality p, for every $p \in \mathbb{N}$.

Hence $\operatorname{range} T$ is infinite dimensional.

$\square$

Corollary 5.28. *Let A be an $m \times n$ matrix. Then*

$$\dim \operatorname{col} A + \dim \operatorname{nul} A = n.$$

Proof. Define $T : \mathbb{R}^n \to \mathbb{R}^m$ by

$$T(\mathbf{x}) = A\mathbf{x}.$$

Then $\operatorname{col} A = \operatorname{range} T$ and $\operatorname{nul} A = \ker T$. The corollary follows as an immediate consequence of the Rank Nullity Theorem. $\square$

Proposition 5.29. *Let V be a vector space. Let S be a subset of V. Then* $\operatorname{Span} S$ *is the smallest subspace of V which contains S. In other words, if W is any subspace of V which contains S then* $\operatorname{Span} S \subset W$.

Also, $\operatorname{Span} S$ *equals the intersection of all subspaces of W which contain S.*

Proof. Let W be a subspace of V which contains S. Then every element of V which can be formed by taking a linear combination of elements of S, must be in W. Hence $\operatorname{Span} S \subset W$.

Next, let W be the intersection of all subspaces of V which contain the set S.

Claim: W is a subspace of V.

Let $u, v \in W$. Let W' be any subspace of V which contains W. Then $u, v \in W'$. Hence $u + v \in W'$. As the choice of W' was arbitary, $u + v$ is in the intersection of all subspaces of V which contain S. Hence $u + v \in W$.

A similar argument shows that $u \in W, c \in \mathbb{R} \implies cu \in W$. Hence the claim.

Hence, by first assertion of the proposition, $\operatorname{Span} S \subset W$. Since $\operatorname{Span} S$ is a subspace of V containing S,

$$W \subset \operatorname{Span} S.$$

Therefore $W = \operatorname{Span} S$. $\square$

Remark. We could have used the above proposition to *define* $\operatorname{Span} S$ as the smallest subspace of V which contains S. In this case, the fact that $\operatorname{Span} \phi = \{0\}$ would follow very naturally.

Definition 5.30. Let A be an $m \times n$ matrix. The dimension of $\operatorname{col} A$ is called the *rank* of A (or the *column rank* of A).

Proposition 5.31. *Let A be an $m \times n$ matrix. The rows containing pivots in the RREF of A are a basis for the row space of A.*

Proof. Let B be a matrix obtained by doing a row operation on A. As the rows of B are linear combinations of rows of A, each row of B is an element of the row space of A. Therefore any linear combination of rows of B is also an element of $\operatorname{row} A$. Thus,

$$\operatorname{row} B \subset \operatorname{row} A.$$

As A is obtained from B by doing an inverse row operation, the same argument holds in reverse, so

$$\operatorname{row} A \subset \operatorname{row} B.$$

Therefore

$$\text{row } A = \text{row } B.$$

As the RREF is obtained by doing a finite sequence of row operations, the row space of A is the same as the row space as the RREF of A.

Claim: The rows containing pivots in the RREF are linearly independent, so they are a basis of the row space of the RREF (because the remaining rows of the RREF are all rows of zeros).

Let A' be the RREF A and let $R_1, \ldots, R_l$ be the rows of A' which contain pivots, where $l \leq m$.

For each $j = 1, \ldots, l$, let i_j be the column which contains the pivot in row j.

Now suppose

$$c_1 R_1 + c_2 R_2 + \cdots + c_l R_l = 0.$$

The i_1-th entry of $c_1 R_1 + c_2 R_2 + \cdots + c_l R_l$ is $c_1 a_{1,i_1}$, thus

$$c_1 a_{1,i_1} = 0.$$

But a_{1,i_1} is a pivot, therefore $c_1 = 0$.

Suppose we have shown that $c_k = 0$ for $k = 1, \ldots, p-1$, where $p \leq l$. Then

$$c_p R_p + \cdots + c_l R_l = 0.$$

Hence the i_p-th entry of $c_p R_p + \cdots + c_l R_l$ must be zero, which gives us

$$c_p a_{p,i_p} = 0.$$

As a_{p,i_p} is a pivot, $c_p = 0$. Thus we obtain $c_1 = \cdots = c_l = 0$. Hence the claim.

□

Definition 5.32. Let A be an $m \times n$ matrix. The dimension of row A is called the *row rank* of A.

Corollary 5.33. *The row rank of A equals the column rank of A.*

Proof. The number of rows containing pivots in the RREF of A equals the number of pivot columns of A. □

The Change-of-coordinates Matrix in $\mathbb{R}^n$

Definition 5.34. Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be an ordered basis of $\mathbb{R}^n$. The matrix

$$P_{\mathcal{B}} = [\mathbf{b}_1 \quad \ldots \quad \mathbf{b}_n]$$

formed using the basis vectors $\mathbf{b}_1, \ldots, \mathbf{b}_n$ as columns, is called the *change-of-coordinates* matrix from $\mathcal{B}$ to the standard basis in $\mathbb{R}^n$.

The change-of-coordinates matrix takes a coordinate vector with respect to the $\mathcal{B}$ basis and tranforms it to standard coordinates. So if $\mathbf{x}$ is a vector in $\mathbb{R}^n$, then

$$\mathbf{x} = P_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}.$$

Proposition 5.35. *Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be an ordered basis of $\mathbb{R}^n$. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be the coordinate transformation which sends*

$$\mathbf{x} \mapsto [\mathbf{x}]_{\mathcal{B}}.$$

The change-of-coordinates matrix $P_{\mathcal{B}}$ is the standard matrix of the inverse T^{-1} of the coordinate transformation.

The standard matrix of the coordinate transformation T is $P_{\mathcal{B}}^{-1}$.

Proof. Let A be the standard matrix of T. From the equation

$$\mathbf{x} = P_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}, \quad \forall \mathbf{x} \in \mathbb{R}^n,$$

it follows that the standard matrix of T^{-1} is $P_{\mathcal{B}}$. Thus $P_{\mathcal{B}}$ is invertible and

$$T(\mathbf{x} = [\mathbf{x}]_{\mathcal{B}} = P_{\mathcal{B}}^{-1}\mathbf{x}, \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

Hence the standard matrix of the coordinate mapping is $P_{\mathcal{B}}^{-1}$. □

Matrix of a Linear Transformation

Proposition 5.36. *Let V, W be vector spaces. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be an ordered basis for V and $\mathcal{C} = \{w_1, \ldots, w_m\}$ be an ordered basis for W. Let $T : V \to W$ be a linear transformation. There exists a unique $m \times n$ matrix A such that*

$$[T(v)]_{\mathcal{C}} = A[v]_{\mathcal{B}}, \quad \textit{for every } v \in V.$$

Further, we have

$$A = [[T(v_1)]_{\mathcal{C}} \quad \ldots \quad [T(v_n)]_{\mathcal{C}}]$$

Proof. Let S_1 be the coordinate mapping from V to $\mathbb{R}^n$ with respect to the basis $\mathcal{B}$, and S_2 be the coordinate mapping from W to $\mathbb{R}^m$ with respect to the basis $\mathcal{C}$.

Then $S_2 \circ T \circ S_1^{-1}$ is a linear transformation from $\mathbb{R}^n \to \mathbb{R}^m$. Let A be the standard matrix of $S_2 \circ T \circ S_1^{-1}$.

Then for every $v \in V$,

$$S_2 \circ T \circ S_1^{-1}([v]_{\mathcal{B}}) = A[v]_{\mathcal{B}}$$

Hence for every $v \in V$,

$$A[v]_{\mathcal{B}} = S_2(T(v)) = [T(v)]_{\mathcal{C}}.$$

Now,

$$\begin{aligned} A &= [A\mathbf{e}_1 \quad \ldots A\mathbf{e}_n] \\ &= [A[v_1]_{\mathcal{B}} \quad \ldots A[v_n]_{\mathcal{B}}] \\ &= [[T(v_1)]_{\mathcal{C}} \quad \ldots \quad [T(v_n)]_{\mathcal{C}}] \end{aligned}$$

□

Definition 5.37. The matrix A of Proposition 5.36 is called the *matrix of T* with respect to the bases $\mathcal{B}$ and $\mathcal{C}$.

Theorem 5.38. *Let $\mathcal{B} = \{b_1, \dots, b_n\}$ and $\mathcal{C} = \{c_1, \dots, c_n\}$ be ordered bases of a vector space V. Then there is a unique $n \times n$ matrix $\underset{\mathcal{C} \leftarrow \mathcal{B}}{P}$ such that*

$$[x]_{\mathcal{C}} = \underset{\mathcal{C} \leftarrow \mathcal{B}}{P}[x]_{\mathcal{B}}, \quad \forall x \in V.$$

The columns of $\underset{\mathcal{C} \leftarrow \mathcal{B}}{P}$ are the $\mathcal{C}$-coordinate vectors of the vectors in the basis $\mathcal{B}$. That is,

$$\underset{\mathcal{C} \leftarrow \mathcal{B}}{P} = [[b_1]_{\mathcal{C}} \quad [b_2]_{\mathcal{C}} \quad \cdots [b_n]_{\mathcal{C}}]$$

Proof. Let T be the identity mapping from V to V. Then by Proposition 5.36, the matrix of T with respect to the bases $\mathcal{B}$ (basis for domain) and $\mathcal{C}$ (basis for codomain) is

$$[[T(b_1)]_{\mathcal{C}} \quad \cdots \quad [T(b_n)]_{\mathcal{C}}] = [[b_1]_{\mathcal{C}} \quad \cdots \quad [b_n]_{\mathcal{C}}],$$

and for any $v \in V$,

$$[v]_{\mathcal{C}} = [T(v)]_{\mathcal{C}} = [[b_1]_{\mathcal{C}} \quad \cdots \quad [b_n]_{\mathcal{C}}][v]_{\mathcal{B}}.$$

□

Definition 5.39. The matrix $\underset{\mathcal{C} \leftarrow \mathcal{B}}{P}$ in the above theorem is called the *change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$*.

Definition 5.40. Let V be a finite dimensional vector space. Let $T : V \to V$ be a linear mapping. Let $\mathcal{B}$ be a basis for V. The matrix of T with respect to the basis $\mathcal{B}$(for both domain and codomain), is called the *$\mathcal{B}$-matrix of T*, or simply the *matrix of T relative to $\mathcal{B}$*. This matrix is denoted by $[T]_{\mathcal{B}}$ and has the property that

$$[T(x)]_{\mathcal{B}} = [T]_{\mathcal{B}}[x]_{\mathcal{B}}, \quad \forall x \in V.$$

If $\mathcal{B} = \{b_1, \dots, b_n\}$ (ordered basis), then

$$[T]_{\mathcal{B}} = [[T(b_1)]_{\mathcal{B}} \quad \cdots \quad [T(b_n)]_{\mathcal{B}}]$$

Proposition 5.41. *Let V, W be finite dimensional vector spaces. Let $T : V \to W$ be a linear transformation. Let $\mathcal{B}$ be a basis for V and $\mathcal{C}$ be a basis for W. Let M be the matrix of T relative to $\mathcal{B}$ and $\mathcal{C}$. Then*

$$\dim \operatorname{range} T = \operatorname{rank} M$$

and

$$\dim \ker T = \dim V - \operatorname{rank} M.$$

Proof. Let $\dim V = n$ and $\dim W = m$. Let $S_1 : V \to \mathbb{R}^n$ be the coordinate mapping with respect to $\mathcal{B}$ and let $S_2 : W \to \mathbb{R}^m$ be the coordinate mapping with respect to $\mathcal{C}$.

Claim: The restriction of S_1 to $\ker T$ is an invertible linear mapping from $\ker T \to \ker S_2 \circ T \circ S_1^{-1}$.

Let $v \in \ker T$. Then

$$S_2 \circ T \circ S_1^{-1}(S_1 v) = S_2(Tv) = 0 \implies S_1 v \in \ker S_2 \circ T \circ S_1^{-1}.$$

Hence elements of $\ker T$ are mapped to $\ker S_2 \circ T \circ S_1^{-1}$ under the action of S_1.

Conversely, suppose $w \in \ker S_2 \circ T \circ S_1^{-1}$. Then

$$S_2(T(S_1^{-1}w)) = 0 \implies T(S_1^{-1}w) = 0 \implies S_1^{-1}(w) \in \ker T.$$

Hence S_1 maps $\ker T$ onto $\ker S_2 \circ T \circ S_1^{-1}$. Therefore

$$\dim \ker T = \dim \ker S_2 \circ T \circ S_1^{-1}$$

Since $S_2 \circ T \circ S_1^{-1}\mathbf{x} = M\mathbf{x}$ for every $\mathbf{x} \in \mathbb{R}^n$, it follows that

$$\dim \operatorname{range} S_2 \circ T \circ S_1^{-1} = \operatorname{rank} M$$

and

$$\dim \ker S_2 \circ T \circ S_1^{-1} = n - \operatorname{rank} M.$$

□

Proposition 5.42. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. Let A be the standard matrix of T. Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be any ordered basis of $\mathbb{R}^n$. Let*

$$P_{\mathcal{B}} = [\mathbf{b}_1 \quad \cdots \quad \mathbf{b}_n]$$

be the change-of-coordinates matrix from $\mathcal{B}$ to the standard basis in $\mathbb{R}^n$. Then the $\mathcal{B}$-matrix of T is $P_{\mathcal{B}}^{-1}AP_{\mathcal{B}}$.

Proof. For every $\mathbf{x} \in \mathbb{R}^n$,

$$\begin{aligned} [T(\mathbf{x})]_{\mathcal{B}} &= [A\mathbf{x}]_{\mathcal{B}} \\ &= P_{\mathcal{B}}^{-1}A\mathbf{x} \\ &= P_{\mathcal{B}}^{-1}AP_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}. \end{aligned}$$

□

Definition 5.43. Two $n \times n$ matrix A and B are said to be *similar* if there exists an invertible matrix P such that $B = P^{-1}AP$.

Proposition 5.44. *Let V be a finite dimensional vector space. Let $\mathcal{B}, \mathcal{C}$ be bases for V. Let $T : V \to V$ be a linear transformation. Then the matrix of T with respect to $\mathcal{B}$ and $\mathcal{C}$ are similar to each other. If $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ is the change-of-basis matrix from $\mathcal{B}$ to $\mathcal{C}$, then*

$$[T]_{\mathcal{B}} = \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}^{-1} [T]_{\mathcal{C}} \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}.$$

Proof. Let $v \in V$. Then

$$[T(v)]_{\mathcal{B}} = [T]_{\mathcal{B}}[v]_{\mathcal{B}}$$

and

$$[T(v)]_{\mathcal{C}} = [T]_{\mathcal{C}}[v]_{\mathcal{C}}$$

Now,

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}[T(v)]_{\mathcal{B}} = [T]_{\mathcal{C}}\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}[v]_{\mathcal{B}}$$

Hence

$$[T(v)]_{\mathcal{B}} = \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}^{-1}[T]_{\mathcal{C}}\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}[v]_{\mathcal{B}}$$

As the choice of v was arbitrary,

$$[T]_{\mathcal{B}} = \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}^{-1}\,[T]_{\mathcal{C}}\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}.$$

□

6 Eigenvalues and Eigenvectors

Definition 6.1. An eigenvector of an $n \times n$ matrix A is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ. A scalar λ is called an eigenvalue of A if there is a nontrivial solution $\mathbf{x}$ of $A\mathbf{x} = \lambda\mathbf{x}$; such an $\mathbf{x}$ is called an eigenvector corresponding to λ.

Theorem 6.2. *$\lambda \in \mathbb{R}$ is an eigenvalue of A if and only if* $\det(A - \lambda I) = 0$.

Proof.

$$\begin{aligned} &\lambda \in \mathbb{R} \text{ is an eigenvalue of } A \\ &\qquad \iff \text{ the equation } (A - \lambda I)\mathbf{v} = \mathbf{0} \text{ has a nontrivial solution} \\ &\qquad \iff A - \lambda I \text{ is not invertible} \\ &\qquad \iff \det(A - \lambda I) = 0. \end{aligned}$$

□

Definition 6.3. $p(\lambda) = \det(A - \lambda I)$ is called the *characteristic polynomial* of A. The equation

$$\det(A - \lambda I) = 0$$

is called the *characteristic equation* of A.

Definition 6.4. If $\lambda \in \mathbb{R}$ is an eigenvalue of A then the nullspace of $A - \lambda I$ is called the *eigenspace* corresponding to λ.

Definition 6.5. The *multiplicity* of an eigenvalue λ, of an $n \times n$ matrix A, is its multiplicity as a root of the characteristic equation

$$\det(A - \lambda I) = 0$$

Theorem 6.6. *Similar matrices have the same characteristic polynomial and therefore the same eigenvalues* with the same multiplicities.

Proof. Let A and B be similar matrices. This means there exists an invertible matrix P such that $PAP^{-1} = B$. Hence

$$\begin{aligned}
\det(B - \lambda I) &= \det(PAP^{-1} - \lambda I) \\
&= \det(P(A - \lambda I)P^{-1}) \\
&= (\det P)\det(A - \lambda I)(\det P^{-1}) \\
&= \det(A - \lambda I),
\end{aligned}$$

because $\det P^{-1} = \dfrac{1}{\det P}$. $\square$

Theorem 6.7. *The eigenvalues of a triangular matrix are the entries on its main diagonal.*

Proof. Let A be a triangular matrix. Then $A - \lambda I$ is also triangular. Therefore

$$p(\lambda) = \det(A - \lambda I) = \prod_{j=1}^{n}(a_{jj} - \lambda).$$

Hence the diagonal (main diagonal) entries of A are the roots of the characteristic polynomial of A, therefore the eigenvalues of A.

$\square$

Diagonalization

Definition 6.8. A square matrix A is said to be *diagonalizable* if A is similar to a diagonal matrix, that is, if $A = PDP^{-1}$ for some invertible matrix P and some diagonal matrix D.

Theorem 6.9. *Let A be an $n \times n$ matrix. Then A is diagonalizable if and only if there exists a basis of $\mathbb{R}^n$ consisting of eigenvectors of A.*

Proof. Suppose A is diagonalizable. Then there exists a diagonal matrix

$$D = \begin{bmatrix} d_1 & 0 & . & . & 0 \\ 0 & d_2 & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & 0 \\ 0 & . & . & 0 & d_n \end{bmatrix}$$

and an invertible matrix P such that $A = PDP^{-1}$.

Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be the columns of P. As P is invertible these columns form a linearly independent set, and hence a basis for $\mathbb{R}^n$.

(Recall that any set of n linearly independent vectors in an n-dimensional vector space is a basis for that vector space).

Also, $AP = PD$. So

$$\begin{aligned}
[A\mathbf{v}_1 \quad \dots \quad A\mathbf{v}_n] &= [\mathbf{v}_1 \quad \dots \quad \mathbf{v}_n] \begin{bmatrix} d_1 & 0 & . & . & 0 \\ 0 & d_2 & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & 0 \\ 0 & . & . & 0 & d_n \end{bmatrix} \\
&= [d_1\mathbf{v}_1 \quad \dots \quad d_n\mathbf{v}_n]
\end{aligned}$$

Hence

$$A\mathbf{v}_j = d_j\mathbf{v_j}, \qquad \text{for } j = 1, \dots, n.$$

Therefore $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis of $\mathbb{R}^n$ consisting of eigenvectors of A.

Conversely, suppose $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis of $\mathbb{R}^n$ consisting of eigenvectors of A. So

$$A\mathbf{v}_j = \lambda_j\mathbf{v}_j, \text{ for } j = 1, \dots, n,$$

where $\lambda_1, \dots, \lambda_n$ are eigenvalues of A (not necessarily distinct).

Define

$$P = [\mathbf{v}_1 \quad \dots \quad \mathbf{v}_n].$$

Then P is invertible because its columns are linearly independent. Further,

$$\begin{aligned}
AP &= A[\mathbf{v}_1 \quad \dots \quad \mathbf{v}_n] \\
&= [\lambda_1\mathbf{v}_1 \quad \dots \quad \lambda_n\mathbf{v}_n] \\
&= P \begin{bmatrix} \lambda_1 & 0 & . & . & 0 \\ 0 & \lambda_2 & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & 0 \\ 0 & . & . & 0 & \lambda_n \end{bmatrix}
\end{aligned}$$

Thus

$$AP = PD$$

where

$$D = \begin{bmatrix} \lambda_1 & 0 & . & . & 0 \\ 0 & \lambda_2 & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & 0 \\ 0 & . & . & 0 & \lambda_n \end{bmatrix}.$$

So $P^{-1}AP$ is diagonal.

□

Theorem 6.10. *If $\mathbf{v}_1, \dots, \mathbf{v}_r$ are eigenvectors that correspond to distinct eigenvalues $\lambda_1, \dots, \lambda_r$ of an $n \times n$ matrix A, then the set $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ is linearly independent.*

Proof. We use induction on the number of distinct eigenvalues. If $r = 1$ then the set $\{\mathbf{v}_1\}$ consists of one nonzero vector and is therefore linearly independent.

Assume that the induction hypothesis holds true for $k \leq r-1$. We show that it holds for $k = r$.

So let $\lambda_1, \ldots, \lambda_r$ be distinct eigenvalues of A and let $\mathbf{v}_1, \ldots, \mathbf{v}_r$ be the corresponding eigenvectors. By the induction hypothesis, the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{r-1}$ are linearly independent.

If $\mathbf{v}_r \notin Span\{\mathbf{v}_1, \ldots, \mathbf{v}_{r-1}\}$ then the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are linearly independent by Proposition 4.20.

So suppose if possible that $\mathbf{v}_r \in Span\{\mathbf{v}_1, \ldots, \mathbf{v}_{r-1}\}$. As $\mathbf{v}_r$ is a nonzero vector, there exist scalars $c_1, \ldots, c_{r-1}$ not all zero, such that

$$\mathbf{v}_r = c_1\mathbf{v}_1 + \ldots + c_{r-1}\mathbf{v}_{r-1}. \tag{10}$$

Therefore

$$A\mathbf{v}_r = c_1 A\mathbf{v}_1 + \ldots + c_{r-1} A\mathbf{v}_{r-1}.$$

Hence

$$\lambda_r\mathbf{v}_r = c_1\lambda_1\mathbf{v}_1 + \ldots + c_{r-1}\lambda_{r-1}\mathbf{v}_{r-1}. \tag{11}$$

Multiplying both sides of equation (10) by λ_r gives us

$$\lambda_r\mathbf{v}_r = \lambda_r c_1\mathbf{v}_1 + \ldots + \lambda_r c_{r-1}\mathbf{v}_{r-1}. \tag{12}$$

Subtracting (12) from (11) gives us

$$\mathbf{0} = c_1(\lambda_r - \lambda_1)\mathbf{v}_1 + \ldots + c_{r-1}(\lambda_r - \lambda_{r-1})\mathbf{v}_{r-1}.$$

As the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{r-1}$ are linearly independent,

$$c_j(\lambda_r - \lambda_j) = 0, \quad \text{for every } j = 1, \ldots, r-1.$$

As the c_js are not all zero, there exists some $j \leq r-1$ such that $c_j \neq 0$.

$$c_j \neq 0 \text{ and } c_j(\lambda_r - \lambda_j) = 0 \implies \lambda_r - \lambda_j = 0 \implies \lambda_r = \lambda_j,$$

contradicting the assumption that the eigenvalues are distinct.

□

Corollary 6.11. *An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.*

Proof. If an $n \times n$ matrix A has n distinct eigenvalues, then any set of n eigenvectors which correspond to these n distinct eigenvalues is linearly independent, and therefore is a basis of $\mathbb{R}^n$. Therefore A is diagonalizable. □

Theorem 6.12 (Not in the Course)**.** *Let A be an $n \times n$ matrix. There exists an upper triangular matrix U having complex entries, and an invertible matrix P such that $P^{-1}AP = U$.*

Remark. The above theorem is not in the course, but is stated because it is used in the proof of the theorem below.

Theorem 6.13. *Let A be an $n \times n$ matrix having real entries whose distinct eigenvalues are $\lambda_1, \ldots, \lambda_p \in \mathbb{R}$.*

1. *For $1 \leq k \leq p$, the dimension of the λ_k-eigenspace is less than or equal to the multiplicity of the eigenvalue λ_k.*
2. *The matrix A is diagonalizable if and only if the sum of the dimensions of the distinct eigenspaces equals n, and this happens if and only if the dimension of the λ_k-eigenspace equals the multiplicity of λ_k, for each $k = 1, \ldots, p$.*
3. *If A is diagonalizable and $\mathcal{B}_k$ is a basis for the eigenspace corresponding to λ_k for each k, then the total collection of vectors in the sets $\mathcal{B}_1, \ldots, \mathcal{B}_p$ forms an eigenvector basis for $\mathbb{R}^n$.*

Remark. The proof of the above theorem is not part of the course but is included here because some students expressed an interest.

Proof. 1. Let the dimension of the λ_k-eigenspace be m and the multiplicity of the eigenvalue λ_k be l. Suppose if possible that $m > l$.

Let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ be a basis of the λ_k-eigenspace. Extend this to a basis $\mathbf{v}_1, \ldots, \mathbf{v}_m, \ldots, \mathbf{v}_n$ of $\mathbb{R}^n$ and let

$$P = [\mathbf{v}_1 \quad \ldots \quad \mathbf{v}_m \quad \ldots \quad \mathbf{v}_n]$$

For $i = m+1, \ldots, n$, let $(b_{1i}, \ldots, b_{ni})$ be the coordinate vector of $A\mathbf{v}_i$ with respect to the basis $\mathbf{v}_1, \ldots, \mathbf{v}_m, \ldots, \mathbf{v}_n$ and let $B = (b_{ij})$ be the $n \times n-m$ matrix formed by using these numbers as entries. Write

$$B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$$

where B_1 is the $m \times n-m$ submatrix formed by the first m rows of B and B_2 is the $n-m \times n-m$ square submatrix formed by the last $n-m$ rows of B.

Then

$$\begin{aligned} AP &= A[\mathbf{v}_1 \quad \ldots \quad \mathbf{v}_m \quad \ldots \quad \mathbf{v}_n] \\ &= [A\mathbf{v}_1 \quad \ldots \quad A\mathbf{v}_m \quad \ldots \quad A\mathbf{v}_n] \\ &= P \begin{bmatrix} \lambda_k I_m & B_1 \\ 0 & B_2 \end{bmatrix} \end{aligned}$$

As B_2 is a square matrix, there exists an $n-m \times n-m$ upper triangular matrix U having complex entries and an invertible matrix Q such that

$$B_2 = QUQ^{-1}$$

Hence

$$P^{-1}AP = \begin{bmatrix} \lambda_k I_m & B_1 \\ 0 & QUQ^{-1} \end{bmatrix} = \begin{bmatrix} I_m & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} \lambda_k I_m & B_1 \\ 0 & U \end{bmatrix} \begin{bmatrix} I_m & 0 \\ 0 & Q^{-1} \end{bmatrix}$$

Put

$$\tilde{P} = P \begin{bmatrix} I_m & 0 \\ 0 & Q \end{bmatrix}$$

Then

$$\tilde{P}^{-1}A\tilde{P} = \begin{bmatrix} \lambda_k I_m & B_1 \\ 0 & U \end{bmatrix}$$

is an upper triangular matrix, which is similar to A.

The eigenvalues of a triangular matrix are its diagonal entries, and similar matrices have the same eigenvalues with the same multiplicities. So λ_k is an eigenvalue of A with multiplicity at least m. But this is a contradiction.

Hence the dimension of the λ_k-eigenspace is less than or equal to the multiplicity of the eigenvalue λ_k.

3. Let A be diagonalizable. Then there exists a basis $\mathcal{B}$ of eigenvectors of A. Let m_k be the multiplicity of λ_k for each $k = 1, \dots, p$.

   By the Pigeonhole Principle, there must be exactly m_k eigenvectors in $\mathcal{B}$ corresponding to each λ_k. These constitute the bases $\mathcal{B}_k$.

2. The statement "the sum of the dimensions of the distinct eigenspaces equals n, if and only if the dimension of the λ_k-eigenspace equals the multiplicity of λ_k, for each $k = 1, \dots, p$" is an immediate consequence of the third statement of the theorem, which is proved above.

   If A is diagonalizable then the required conclusion is again, an immediate consequence of statement 3.

   If the sum of the dimensions of the eigenspaces equals n, then there is a basis of eigenvectors of $\mathbb{R}^n$ consisting of eigenvectors of A, so A must be diagonalizable.

□

Theorem 6.14. *Let A be a diagonalizable $n \times n$ matrix and let $p(x)$ be the characteristic polynomial of A. Then $p(A) = 0$. In other words, if*

$$p(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0,$$

then

$$A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 I = 0.$$

Remark. This theorem holds for ALL $n \times n$ matrices but we are only proving it for diagonalizable matrices. The general case is not part of the course, although it can be included in the Proofs document in case students express interest.

Proof. Let A be a diagonalizable $n \times n$ matrix.

Then there exists a diagonal matrix $D = (d_{ij})$, whose diagonal entries are eigenvalues of A, and an invertible matrix P such that

$$P^{-1}AP = D.$$

Now

$$p(D) = D^n + c_{n-1}D^{n-1} + \cdots + c_1 D + c_0 I$$

is also a diagonal matrix and its diagonal entries are of the form $p(d_{ii})$ $(i = 1, \dots, n)$ where $d_{11}, \dots, d_{nn}$ are eigenvalues of A.

But each eigenvalue of A is a root of the characteristic polynomial of A, so $p(d_{ii}) = 0$ for $i = 1, \ldots, n$. Hence
$$p(D) = D^n + c_{n-1}D^{n-1} + \cdots + c_1 D + c_0 I = 0.$$

Now

$$\begin{aligned}
p(A) &= A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 I \\
&= (PDP^{-1})^n + c_{n-1}(PDP^{-1})^{n-1} + \cdots + c_1 PDP^{-1} + c_0 PP^{-1} \\
&= PD^n P^{-1} + c_{n-1}PD^{n-1}P^{-1} + \cdots + c_1 PDP^{-1} + c_0 PP^{-1} \\
&= P(D^n + c_{n-1}D^{n-1} + \cdots + c_1 D + c_0 I)P^{-1} \\
&= 0
\end{aligned}$$

□

Theorem 6.15. *Suppose $A = PDP^{-1}$, where D is a diagonal $n \times n$ matrix. If $\mathcal{B}$ is the basis for $\mathbb{R}^n$ formed from the columns of P, then D is the $\mathcal{B}$-matrix for the transformation $\mathbf{x} \mapsto A\mathbf{x}$.*

Proof. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be the linear transformation that maps $\mathbf{x} \mapsto A\mathbf{x}$.

Since if $\mathcal{B}$ is formed from the columns of P, the change-of-coordinate matrix $P_{\mathcal{B}} = P$.

Thus by Proposition 5.42, the $\mathcal{B}$-matrix for T is

$$P_{\mathcal{B}}^{-1}AP_{\mathcal{B}} = P^{-1}AP = D.$$

□

Complex Eigenvalues

Definition 6.16. Let A be an $n \times n$ matrix. A number $\lambda \in \mathbb{C}$ is called a *complex eigenvalue* of A if there exists a nonzero vector $\mathbf{x} \in \mathbb{C}^n$ such that $A\mathbf{x} = \lambda\mathbf{x}$.

Proposition 6.17. *λ is a complex eigenvalue of an $n \times n$ matrix A if and only if λ is a root of the characteristic polynomial of A.*

Proof.

$$\begin{aligned}
&\det A - \lambda I = 0 \\
&\quad \iff A - \lambda I \text{ is not invertible} \\
&\quad \iff (A - \lambda I)\mathbf{x} = 0 \text{ has a non-trivial solution in } \mathbb{C}^n
\end{aligned}$$

□

Theorem 6.18. *If A is a 2×2 matrix and $\lambda = re^{-i\theta}$ is a complex eigenvalue of A ($r > 0$, $0 < \theta < 2\pi$, $\theta \neq \pi$), and $\mathbf{v} \in \mathbb{C}^2$ is a complex eigenvector of A corresponding to λ, then*

$$P^{-1}AP = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix},$$

where $P = [Re(\mathbf{v}) \quad Im(\mathbf{v})]$.

Proof. Suppose $\mathbf{v} \in \mathbb{C}^2$ is a eigenvector corresponding to λ, i.e.,

$$A\mathbf{v} = \lambda\mathbf{v} = re^{-i\theta}\mathbf{v}.$$

Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$ denote the real and imaginary parts of $\mathbf{v}$, i.e.,

$$\mathbf{v} = \mathbf{v}_1 + i\mathbf{v}_2.$$

Since

$$A\bar{\mathbf{v}} = \bar{\lambda}\bar{\mathbf{v}},$$

it follows that

$$A\left(\frac{\mathbf{v} + \bar{\mathbf{v}}}{2}\right) = \frac{\lambda\mathbf{v} + \bar{\lambda}\bar{\mathbf{v}}}{2}.$$

Therefore

$$\begin{aligned} A\mathbf{v}_1 &= \frac{(r\cos\theta - ir\sin\theta)(\mathbf{v}_1 + i\mathbf{v}_2) + (r\cos\theta + ir\sin\theta)(\mathbf{v}_1 - i\mathbf{v}_2)}{2} \\ &= r\cos\theta\mathbf{v}_1 + r\sin\theta\mathbf{v}_2. \end{aligned}$$

Similarly,

$$\begin{aligned} A\mathbf{v}_2 &= A\left(\frac{\mathbf{v} - \bar{\mathbf{v}}}{2i}\right) = \frac{\lambda\mathbf{v} - \bar{\lambda}\bar{\mathbf{v}}}{2i} \\ &= \frac{(r\cos\theta - ir\sin\theta)(\mathbf{v}_1 + i\mathbf{v}_2) - (r\cos\theta + ir\sin\theta)(\mathbf{v}_1 - i\mathbf{v}_2)}{2i} \\ &= -r\sin\theta\mathbf{v}_1 + r\cos\theta\mathbf{v}_2. \end{aligned}$$

Therefore

$$A[\mathbf{v}_1 \quad \mathbf{v}_2] = [\mathbf{v}_1 \quad \mathbf{v}_2] \begin{bmatrix} r\cos\theta & -r\sin\theta \\ r\sin\theta & r\cos\theta \end{bmatrix}.$$

Put

$$P = [\mathbf{v}_1 \quad \mathbf{v}_2]$$

If P is not invertible then there exists a number $\mu \neq 0$ such that $\mu\mathbf{v} \in \mathbb{R}^2$. But any scalar multiple of an eigenvector is also an eigenvector.

However, since $\lambda \notin \mathbb{R}$, A cannot have an eigenvector in $\mathbb{R}^2$. So P must be invertible.

Thus the matrix $P^{-1}AP$ is a the product of a rotation matrix and a scaling matrix,

$$P^{-1}AP = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}.$$

□

7 Inner Product Spaces

Definition 7.1. If $\mathbf{u} = (u_1, \ldots, u_n), \mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$ then the *Euclidean inner product* or the *dot product* of $\mathbf{u}$ and $\mathbf{v}$ is the real number

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{n} u_i v_i$$

Theorem 7.2. *Let* $\mathbf{u}, \mathbf{v},$ *and* $\mathbf{w}$ *be vectors in* $\mathbb{R}^n$*, and let* c *be a scalar. Then*

1. $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
2. $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$
3. $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (c\mathbf{v})$
4. $\mathbf{u} \cdot \mathbf{u} \geq 0,$ *and* $\mathbf{u} \cdot \mathbf{u} = 0$ *if and only if* $\mathbf{u} = \mathbf{0}$

Proof. The proof of this theorem is a routine verification and is left as an exercise. □

Definition 7.3. Let V be a real vector space. An *inner product* $\langle ., . \rangle : V \times V \to \mathbb{R}$ is a function satisfying the following conditions:

(1) $\langle u, v \rangle = \langle v, u \rangle$, for every $u, v \in V$

(2) $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$, for every $u, v, w, \in V$

(3) $\langle cu, v \rangle = c\langle u, v \rangle$ for every $c \in \mathbb{R}$ and $u, v \in V$

(4) $\langle v, v \rangle \geq 0$ for every $v \in V$, and $\langle v, v \rangle = 0$ holds if and only if $v = 0$

Proposition 7.4. *Let* V *be a vector space and let* $\langle ., . \rangle$ *be an inner product on* V*. Let* $u_1, \ldots, u_n, w \in V$ *and* $c_1, \ldots, c_n \in \mathbb{R}$*. Then*

$$\langle c_1 u_1 + \cdots + c_n u_n, w \rangle = c_1 \langle u_1, w \rangle + \cdots + c_n \langle u_n, w \rangle$$

Proof. The proof is a routine verification using induction and is left as an exercise. □

Definition 7.5. Let $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$. The *length (or norm)* of $\mathbf{v}$ is defined by

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

Definition 7.6. Let V be a vector space and let $\langle ., . \rangle$ be an inner product on V. The *length (or norm)* of a vector $v \in V$ is defined by

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

Definition 7.7. A vector whose length is 1 is called a *unit vector*.

If we divide a nonzero vector v by its length, we obtain a unit vector $\hat{v}$, because if

$$\hat{v} = \frac{1}{\|v\|}v,$$

then

$$\|\hat{v}\| = \left\| \frac{1}{\|v\|}v \right\| = \frac{1}{\|v\|}\,\|v\| = 1$$

The process of creating $\hat{v}$ from v is called *normalizing* v, and we say that $\hat{v}$ is in the *same direction* as v.

Notation. We often express a unit vector in the direction of v as

$$\hat{v} = \frac{v}{\|v\|}.$$

Some Famous Equations/Inequalities

Theorem 7.8 (Cauchy-Schwarz Inequality)**.** *Let V be a vector space with an inner product $\langle ., . \rangle$. Let $u, v \in V$. Then*

$$|\langle u, v\rangle| \leq \|u\|\|v\|$$

Proof. Let us first consider the case where u and v are unit vectors. Then

$$0 \leq \langle u+v, u+v\rangle = \langle u, u\rangle + 2\langle u, v\rangle + \langle v, v\rangle = 2 + 2\langle u, v\rangle$$

Hence

$$\langle u, v\rangle \geq -1 \tag{13}$$

Similarly

$$0 \leq \langle u-v, u-v\rangle = \langle u, u\rangle - 2\langle u, v\rangle + \langle v, v\rangle = 2 - 2\langle u, v\rangle$$

So

$$\langle u, v\rangle \leq 1 \tag{14}$$

From (13) and (14) we obtain

$$|\langle u, v\rangle| \leq 1$$

Next, suppose $u, v \in V$ are arbitrary (not necessarily unit vectors). Let us first consider the case where $u \neq 0, v \neq 0$. Let

$$\hat{u} = \frac{u}{\|u\|}, \quad \hat{v} = \frac{v}{\|v\|}$$

Then

$$|\langle \hat{u}, \hat{v}\rangle| \leq 1$$

i.e.,

$$\left| \left\langle \frac{u}{\|u\|}, \frac{v}{\|v\|} \right\rangle \right| \leq 1$$

Hence

$$|\langle u, v\rangle| \leq \|u\|\|v\|$$

If $u = 0$ then

$$\langle u, v\rangle = \langle 0u, v\rangle = 0\,\langle u, v\rangle = 0$$

so the inequality is trivial. Likewise if $v = 0$.

□

Definition 7.9. For $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, the distance between $\mathbf{u}$ and $\mathbf{v}$, written as $\operatorname{dist}(\mathbf{u}, \mathbf{v})$, is the length of the vector $\mathbf{u} - \mathbf{v}$. That is,

$$\operatorname{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

Definition 7.10. Let V be a vector space with an inner product $\langle ., .\rangle$. The *distance $d(u, v)$ between vectors u and $v \in V$, with respect to the inner product $\langle ., .\rangle$*, is defined as

$$d(u, v) = \|u - v\|$$

Definition 7.11. Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are said to be *orthogonal (to each other)* if $\mathbf{u} \cdot \mathbf{v} = 0$.

Definition 7.12. Two vectors u and v in an inner product space $(V, \langle ., .\rangle)$ are said to be *orthogonal (to each other)* if

$$\langle u, v\rangle = 0.$$

Theorem 7.13 (Pythagoras Theorem)**.**
Let $(V, \langle ., .\rangle)$ be an inner product space. The vectors v and w are orthogonal if and only if

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2.$$

Proof.

$$\begin{aligned}
&\|v + w\|^2 - \|v\|^2 - \|w\|^2 \\
&\quad = \langle v + w, v + w\rangle - \langle v, v\rangle - \langle w, w\rangle \\
&\quad = \langle v, v + w\rangle + \langle w, v + w\rangle - \langle v, v\rangle - \langle w, w\rangle \\
&\quad = \langle v, v\rangle + \langle v, w\rangle + \langle w, v\rangle + \langle w, w\rangle - \langle v, v\rangle - \langle w, w\rangle \\
&\quad = 2\langle v, w\rangle.
\end{aligned} \tag{15}$$

Therefore $\langle v, w\rangle = 0 \iff \|v + w\|^2 = \|v\|^2 + \|w\|^2$. □

Theorem 7.14 (The Triangle Inequality)**.**
Let $(V, \langle ., .\rangle)$ be an inner product space, and let $v, w \in V$. Then

$$\|v + w\| \leq \|v\| + \|w\|$$

Proof. By equation (15),

$$\|v+w\|^2 - \|v\|^2 - \|w\|^2 = 2\langle v, w\rangle.$$

By the Cauchy-Schwarz inequality,

$$|\langle v, w\rangle| \leq \|v\|\|w\|$$

So

$$\|v+w\|^2 - \|v\|^2 - \|w\|^2 = 2\langle v, w\rangle \leq 2|\langle v, w\rangle| \leq 2\|v\|\|w\|$$

Hence

$$\|v+w\|^2 \leq (\|v\| + \|w\|)^2$$

Thus $\|v+w\| \leq \|v\| + \|w\|$.

$\square$

Orthogonal Sets

Definition 7.15. Let V be a vector space with an inner product $\langle ., .\rangle$. A set of vectors $\{v_1, \ldots, v_p\}$ in V is said to be an *orthogonal set* if

$$\langle v_i, v_j\rangle = 0, \quad \text{for every } i \neq j, \quad \text{where } i, j \in \{1, \ldots, p\}.$$

Theorem 7.16. *Let V be a vector space. Let $\langle ., .\rangle$ be an inner product defined on V. Let $S = \{v_1, \ldots, v_p\}$ be an orthogonal set of non-zero vectors in an inner product space $(V, \langle ., .\rangle)$. Then S is linearly independent set and hence is a basis for* $\operatorname{Span} S$.

Proof. Suppose $c_1, \ldots, c_p$ are scalars such that

$$c_1 v_1 + c_2 v_2 + \ldots + c_p v_p = 0.$$

Let $j \in \{1, \ldots, p\}$. Then

$$\langle c_1 v_1 + c_2 v_2 + \ldots + c_p v_p, v_j\rangle = \langle 0, v_j\rangle = 0.$$

By expanding out on the left hand side, we obtain

$$c_1\langle v_1, v_j\rangle + c_2\langle v_2, v_j\rangle + \ldots + c_p\langle v_p, v_j\rangle = 0.$$

Since $\langle v_i, v_j\rangle = 0$ for $i \neq j$, it follows that

$$c_j\langle v_j, v_j\rangle = 0.$$

As $v_j \neq 0$, $\langle v_j, v_j\rangle > 0$. Therefore $c_j = 0$.

As the choice of j was arbitary, $c_j = 0$ for each $j = 1, \ldots, p$.

$\square$

Definition 7.17. Let $(V, \langle ., .\rangle)$ be a finite dimensional inner product space. An orthogonal set which is also a basis for V is called an *orthogonal basis* for V.

Definition 7.18. Let $(V, \langle ., . \rangle)$ be a finite dimensional inner product space. An orthogonal set of unit vectors in V is called an *orthonormal set.*

Definition 7.19. Let $(V, \langle ., . \rangle)$ be a finite dimensional inner product space. An orthonormal set which is also a basis for V is called an *orthonormal basis* for V.

Definition 7.20. Let $(V, \langle ., . \rangle)$ be an inner product space. Let $v \in V$ be a nonzero vector. For any vector $u \in V$, the *orthogonal projection of* u *onto* v is defined as

$$\mathrm{proj}_v u := \frac{\langle u, v \rangle}{\langle v, v \rangle} v,$$

and the vector

$$u - \mathrm{proj}_v u$$

is called the *component of* u *which is orthogonal to* v.

Theorem 7.21. *Let $(V, \langle ., . \rangle)$ be an inner product space. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be an orthogonal basis for V. Let $v \in V$ be any vector. Then the $\mathcal{B}$-coordinates of v are:*

$$[v]_{\mathcal{B}} = \left(\frac{\langle v, v_1 \rangle}{\langle v_1, v_1 \rangle}, \frac{\langle v, v_2 \rangle}{\langle v_2, v_2 \rangle}, \ldots, \frac{\langle v, v_n \rangle}{\langle v_n, v_n \rangle} \right).$$

Proof. Let $[v]_{\mathcal{B}} = (c_1, c_2, \ldots, c_n)$, i.e.

$$v = c_1 v_1 + c_2 v_2 + \ldots + c_n v_n.$$

Then for each $j = 1, \ldots, n$,

$$\langle v, v_j \rangle = \left\langle \sum_{i=1}^{n} c_i v_i, v_j \right\rangle = \sum_{i=1}^{n} c_i \langle v_i, v_j \rangle = c_j \langle v_j, v_j \rangle.$$

Therefore

$$c_j = \frac{\langle v, v_j \rangle}{\langle v_j, v_j \rangle}.$$

□

Corollary 7.22. *Let $(V, \langle ., . \rangle)$ be an inner product space. Let $\mathcal{B} = \{\hat{v}_1, \ldots, \hat{v}_n\}$ be an orthonormal basis for V. Let $v \in V$ be any vector. Then the $\mathcal{B}$-coordinates of v are:*

$$[v]_{\mathcal{B}} = \left(\langle v, \hat{v}_1 \rangle, \ldots, \langle v, \hat{v}_n \rangle \right).$$

Proof. Use the formula in Theorem 7.21 with $\langle \hat{v}_j, \hat{v}_j \rangle = 1$ for $j = 1, \ldots, n$. □

Theorem 7.23. *An $m \times n$ matrix U has orthonormal columns if and only if $U^T U = I$.*

Proof. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be the columns of U. Then

$$i, j\text{-th entry of } U^T U = (\text{row } i \text{ of } U^T).(\text{column } j \text{ of } U) = \mathbf{u}_i^T \mathbf{u}_j = \mathbf{u}_i \cdot \mathbf{u}_j \tag{16}$$

By equation (16),

$$U^T U = I \iff \{\mathbf{u}_1, \ldots, \mathbf{u}_n\} \text{ is an orthonormal set.}$$

□

Definition 7.24. An $n \times n$ matrix is said to be *orthogonal* if its columns form an **orthonormal set**.

Proposition 7.25. *An $n \times n$ matrix P is orthogonal if and only if $P^T = P^{-1}$.*

Proof. By Theorem 7.23, P is orthogonal if and only if $P^T = P^{-1}$.

□

The Gram-Schmidt Process

Theorem 7.26. *Given a basis $\{v_1, \ldots, v_p\}$ for an inner product space $(V, \langle ., . \rangle)$, define*

$$\begin{aligned}
w_1 &= v_1 \\
w_2 &= v_2 - \frac{\langle v_2, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 \\
&\vdots \\
w_p &= v_p - \frac{\langle v_p, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 - \frac{\langle v_p, w_2 \rangle}{\langle w_2, w_2 \rangle} w_2 - \cdots - \frac{\langle v_p, w_{p-1} \rangle}{\langle w_{p-1}, w_{p-1} \rangle} w_{p-1}
\end{aligned}$$

Then $\{w_1, \ldots, w_p\}$ is an orthogonal basis for V. In addition

$$\text{Span}\,\{w_1, \ldots, w_k\} = \text{Span}\,\{v_1, \ldots, v_k\} \qquad \textit{for } 1 \leq k \leq p$$

Proof. We show that the sets $\{w_1, \ldots, w_k\}$ are orthogonal, using induction on k.

For $k = 1$, there is nothing to show.

Assume that $\{w_1, \ldots, w_l\}$ is an orthogonal set, for every $1 \leq l < k$. We show that $\{w_1, \ldots, w_k\}$ is orthogonal.

Let i and j be positive integers such that $i \neq j$ and $i, j \leq k$. If i and j are strictly less than k then the vectors w_i and w_j are orthogonal, by the induction hypothesis.

Hence let us assume, without loss of generality, that $i = k$ and $j < k$. Then

$$\begin{aligned}
\langle w_i, w_j \rangle &= \langle w_k, w_j \rangle \\
&= \left\langle v_k - \sum_{m=1}^{k-1} \frac{\langle v_k, w_m \rangle}{\langle w_m, w_m \rangle} w_m, w_j \right\rangle \\
&= \langle v_k, w_j \rangle - \sum_{m=1}^{k-1} \frac{\langle v_k, w_m \rangle}{\langle w_m, w_m \rangle} \langle w_m, w_j \rangle \\
&= \langle v_k, w_j \rangle - \frac{\langle v_k, w_j \rangle}{\langle w_j, w_j \rangle} \langle w_j, w_j \rangle \\
&= 0.
\end{aligned}$$

Clearly

$$v_k \in \operatorname{Span}\{w_1, \ldots, w_k\}, \quad \text{for } 1 \leq k \leq p$$

Hence

$$\{v_1, \ldots, v_k\} \subset \operatorname{Span}\{w_1, \ldots, w_k\}, \quad \text{for } 1 \leq k \leq p$$

So,

$$\operatorname{Span}\{v_1, \ldots, v_k\} \subset \operatorname{Span}\{w_1, \ldots, w_k\}, \quad \text{for } 1 \leq k \leq p$$

Since $\dim \operatorname{Span}\{v_1, \ldots, v_k\} = \dim \operatorname{Span}\{w_1, \ldots, w_k\}$,
it follows that

$$\operatorname{Span}\{v_1, \ldots, v_k\} = \operatorname{Span}\{w_1, \ldots, w_k\}, \quad \text{for } 1 \leq k \leq p.$$

In particular $\{w_1, \ldots, w_p\}$ is an orthogonal basis for V.

□

Definition 7.27. A matrix A is said to be *orthogonally diagonalizable* if there exists an orthogonal matrix P (with $P^{-1} = P^T$), and a diagonal matrix D such that

$$A = PDP^T = PDP^{-1}$$

Theorem 7.28. *If an $n \times n$ matrix A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.*

Proof. Let λ and μ be distinct eigenvalues of A, and let $\mathbf{v}$ and $\mathbf{w}$ be the corresponding eigenvectors, i.e.

$$A\mathbf{v} = \lambda\mathbf{v}, \quad A\mathbf{w} = \mu\mathbf{w}$$

Then

$$\begin{aligned} \lambda\mathbf{v} \cdot \mathbf{w} &= (\lambda\mathbf{v}) \cdot \mathbf{w} \\ &= (A\mathbf{v}) \cdot \mathbf{w} \\ &= \mathbf{v}^T A^T \mathbf{w} \\ &= \mathbf{v}^T A \mathbf{w} \\ &= \mathbf{v} \cdot (A\mathbf{w}) \\ &= \mathbf{v} \cdot (\mu\mathbf{w}) = \mu\mathbf{v} \cdot \mathbf{w} \end{aligned}$$

Thus $(\lambda - \mu)\mathbf{v} \cdot \mathbf{w} = 0$.

As $\lambda \neq \mu$, this can only happen if

$$\mathbf{v} \cdot \mathbf{w} = 0.$$

□

Lemma 7.29. *Let A be an $n \times n$ symmetric matrix having real entries. Then A has n real eigenvalues.*

Proof. Let λ be an eigenvalue of A, and let $\mathbf{v} \in \mathbb{C}^n$ be a corresponding eigenvector, i.e.

$$A\mathbf{v} = \lambda\mathbf{v}$$

Then

$$(\overline{A\mathbf{v}})^T A\mathbf{v} = (\overline{\lambda\mathbf{v}})^T \lambda\mathbf{v}$$

Hence

$$\bar{\mathbf{v}}^T \bar{A}^T A\mathbf{v} = |\lambda|^2 \bar{\mathbf{v}}^T \mathbf{v}$$

So

$$|\lambda|^2 \bar{\mathbf{v}}^T \mathbf{v} = \bar{\mathbf{v}}^T A^T \lambda\mathbf{v} = \bar{\mathbf{v}}^T \lambda A\mathbf{v} = \lambda^2 \bar{\mathbf{v}}^T \mathbf{v}$$

Hence

$$|\lambda|^2 = \lambda^2$$

If $\lambda = 0$, then λ is real. So suppose $\lambda \neq 0$. Then

$$\left(\frac{\lambda}{|\lambda|}\right)^2 = 1$$

Hence

$$\lambda = \pm|\lambda|.$$

So λ is real.

□

Theorem 7.30. *An $n \times n$ matrix A is orthogonally diagonalizable if and only if A is a symmetric matrix.*

Proof. If A is orthogonally diagonalizable, then there exists an orthogonal matrix P and a diagonal matrix D such that

$$A = PDP^T = PDP^{-1}$$

Hence

$$A^T = (PDP^T)^T = PDP^T.$$

We prove the converse using induction on n.

If $n = 1$, there is nothing to show. So assume that $n \geq 2$.

Let λ be an eigenvalues of A. By Lemma 7.29, $\lambda \in \mathbb{R}$.

Let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ (where $m \leq n$) be an orthonormal basis of the λ-eigenspace. Extend this to an orthonormal basis $\mathbf{v}_1, \ldots, \mathbf{v}_m, \ldots, \mathbf{v}_n$ of $\mathbb{R}^n$ and let

$$P = [\mathbf{v}_1 \quad \ldots \quad \mathbf{v}_m \quad \ldots \quad \mathbf{v}_n]$$

For $i = m+1, \ldots, n$, let $(b_{1i}, \ldots, b_{ni})$ be the coordinate vector of $A\mathbf{v}_i$ with respect to the basis $\mathbf{v}_1, \ldots, \mathbf{v}_m, \ldots, \mathbf{v}_n$ and let $B = (b_{ij})$ be the $n \times n-m$ matrix formed by using these numbers as entries. Write

$$B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$$

where B_1 is the $m \times n-m$ submatrix formed by the first m rows of B and B_2 is the $n-m \times n-m$ square submatrix formed by the last $n-m$ rows of B.

Then

$$\begin{aligned} AP &= A[\mathbf{v}_1 \quad \dots \quad \mathbf{v}_m \quad \dots \quad \mathbf{v}_n] \\ &= [A\mathbf{v}_1 \quad \dots \quad A\mathbf{v}_m \quad \dots \quad A\mathbf{v}_n] \\ &= P \begin{bmatrix} \lambda I_m & B_1 \\ 0 & B_2 \end{bmatrix} \end{aligned}$$

As the columns of P form an orthonormal basis of $\mathbb{R}^n$, $P^{-1} = P^T$. Therefore

$$P^T AP = \begin{bmatrix} \lambda I_m & B_1 \\ 0 & B_2 \end{bmatrix}$$

As $P^T AP$ is a symmetric matrix, $B_1 = 0$ and B_2 is a symmetric matrix. By the induction hypothesis, B_2 is orthogonally diagonalizable. Thus there exists an $n-m \times n-m$ diagonal matrix D an orthogonal matrix Q such that

$$B_2 = QDQ^T$$

Hence

$$P^T AP = \begin{bmatrix} \lambda I_m & 0 \\ 0 & QDQ^T \end{bmatrix} = \begin{bmatrix} I_m & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} \lambda I_m & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I_m & 0 \\ 0 & Q^T \end{bmatrix}$$

Put

$$\tilde{P} = P \begin{bmatrix} I_m & 0 \\ 0 & Q \end{bmatrix}$$

Then $\tilde{P}$ is an orthogonal matrix, and

$$\tilde{P}^{-1} A \tilde{P} = \begin{bmatrix} \lambda I_m & 0 \\ 0 & D \end{bmatrix}$$

is a diagonal matrix.

□

Theorem 7.31 (The Spectral Theorem for Symmetric Matrices)**.** *An $n \times n$ symmetric matrix A has the following properties:*

1. *A has n real eigenvalues, counting multiplicities.*
2. *The dimension of the eigenspace for each eigenvalue λ equals the multiplicity of λ as a root of the characteristic equation.*
3. *The eigenspaces are mutually orthogonal, in the sense that eigenvectors corresponding to different eigenvalues are orthogonal.*
4. *A is orthogonally diagonalizable.*

Proof. The first asssertion is a restatement of Lemma 7.29.

The third assertion is proved in Theorem 7.28.

The fourth assertion is proved in Theorem 7.30.

The second assertion follows from Theorem 6.13, and the fourth assertion.

□

Theorem 7.32. *Let U be an $m \times n$ matrix with orthonormal columns, and let $\mathbf{x}$ and $\mathbf{y}$ be in $\mathbb{R}^n$. Then*

1. $\|U\mathbf{x}\| = \|\mathbf{x}\|$
2. $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$
3. $(U\mathbf{x}) \cdot (U\mathbf{y}) = 0$ *if and only if* $\mathbf{x} \cdot \mathbf{y} = 0$

Proof. By Theorem 7.23, $U^T U = I$. So

$$\|U\mathbf{x}\|^2 = (U\mathbf{x}) \cdot (U\mathbf{x}) = (U\mathbf{x})^T (U\mathbf{x}) = \mathbf{x}^T U^T U \mathbf{x} = \mathbf{x}^T \mathbf{x} = \mathbf{x} \cdot \mathbf{x} = \|\mathbf{x}\|^2$$
$$(U\mathbf{x}) \cdot (U\mathbf{y}) = (U\mathbf{x})^T (U\mathbf{y}) = \mathbf{x}^T U^T U \mathbf{y} = \mathbf{x}^T \mathbf{y} = \mathbf{x} \cdot \mathbf{y}$$

$\square$

Theorem 7.33. *If $T : \mathbb{R}^n \to \mathbb{R}^n$ preserves the dot product, i.e. if*

$$T(\mathbf{x}) \cdot T(\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n,$$

then T is a linear map and the standard matrix of T is orthogonal.

Proof. For each $i = 1, \ldots, n$,

$$\|T(\mathbf{e}_i)\|^2 = T(\mathbf{e}_i) \cdot T(\mathbf{e}_i) = \mathbf{e}_i \cdot \mathbf{e}_i = 1$$

Therefore $T(\mathbf{e}_i)$ is a unit vector for $i = 1, \ldots, n$. For $i \neq j$,

$$T(\mathbf{e}_i) \cdot T(\mathbf{e}_j) = \mathbf{e}_i \cdot \mathbf{e}_j = 0.$$

Thus $\mathcal{B} = \{T(\mathbf{e}_1), \ldots, T(\mathbf{e}_n)\}$ is an orthonormal set, and therefore an orthonormal basis for $\mathbb{R}^n$ as it has n elements.

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then

$$\begin{aligned}
[T(\mathbf{x}) + T(\mathbf{y})]_{\mathcal{B}} &= ((T(\mathbf{x}) + T(\mathbf{y})) \cdot T(\mathbf{e}_1), \ldots, (T(\mathbf{x}) + T(\mathbf{y})) \cdot T(\mathbf{e}_n)) \\
&= (T(\mathbf{x}) \cdot T(\mathbf{e}_1) + T(\mathbf{y}) \cdot T(\mathbf{e}_1), \ldots, T(\mathbf{x}) \cdot T(\mathbf{e}_n) + T(\mathbf{y}) \cdot T(\mathbf{e}_n)) \\
&= (\mathbf{x} \cdot \mathbf{e}_1 + \mathbf{y} \cdot \mathbf{e}_1, \ldots, \mathbf{x} \cdot \mathbf{e}_n + \mathbf{y} \cdot \mathbf{e}_n) \\
&= ((\mathbf{x} + \mathbf{y}) \cdot \mathbf{e}_1, \ldots, (\mathbf{x} + \mathbf{y}) \cdot \mathbf{e}_n) \\
&= (T(\mathbf{x} + \mathbf{y}) \cdot T(\mathbf{e}_1), \ldots, T(\mathbf{x} + \mathbf{y}) \cdot T(\mathbf{e}_n)) \\
&= [T(\mathbf{x} + \mathbf{y})]_{\mathcal{B}}
\end{aligned}$$

Hence $T(\mathbf{x}) + T(\mathbf{y}) = T(\mathbf{x} + \mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
Similarly, if $\mathbf{x} \in \mathbb{R}^n$ and $c \in \mathbb{R}$, then

$$\begin{aligned}
[T(c\mathbf{x})]_{\mathcal{B}} &= (T(c\mathbf{x}) \cdot T(\mathbf{e}_1), \ldots, T(c\mathbf{x}) \cdot T(\mathbf{e}_n) \\
&= (c\mathbf{x} \cdot \mathbf{e}_1, \ldots, c\mathbf{x} \cdot \mathbf{e}_n) \\
&= c(\mathbf{x} \cdot \mathbf{e}_1, \ldots, \mathbf{x} \cdot \mathbf{e}_n) \\
&= c(T(\mathbf{x}) \cdot T(\mathbf{e}_1), \ldots, T(\mathbf{x}) \cdot T(\mathbf{e}_n)) \\
&= c[T(\mathbf{x})]_{\mathcal{B}}
\end{aligned}$$

Hence $T(c\mathbf{x}) = cT(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

The standard matrix of T is $[T(\mathbf{e}_1) \quad \cdots \quad T(\mathbf{e}_n)]$. Its columns form an orthonormal set, so it is orthogonal.

$\square$

Orthogonal Decomposition of a Vector Space

Definition 7.34. Let V be a vector space with an inner product $\langle .,. \rangle$.Let W be a subspace of V. We define the *orthogonal complement* of W in V to be the set

$$W^{\perp} = \{v \in V \mid \langle v, w \rangle = 0, \ \forall w \in W\}.$$

Theorem 7.35. *Let A be an $m \times n$ matrix. The orthogonal complement of the row space of A is the nullspace of A, and the orthogonal complement of the column space of A is the nullspace of A^T :*

$$(\text{Row } A)^{\perp} = \text{Nul } A \quad \textit{and} \quad (\text{Col } A)^{\perp} = \text{Nul } A^T$$

Proof. Let $\mathbf{x} \in \text{Nul } A$. Then $A\mathbf{x} = 0$ implies that every row of A is orthogonal to $\mathbf{x}$. Hence every linear combination of rows of A is also orthogonal to $\mathbf{x}$. Therefore

$$\mathbf{x} \in (\text{Row } A)^{\perp}.$$

Hence $\text{Nul } A \subset (\text{Row } A)^{\perp}$. Conversely, suppose $\mathbf{x} \in (\text{Row } A)^{\perp}$. Then $\mathbf{x}$ is orthogonal to every row of A. Hence $\mathbf{x} \in \text{Nul } A$. So

$$(\text{Row } A)^{\perp} \subset \text{Nul } A.$$

□

Definition 7.36. Let V be a vector space. Let U and W be subpaces of V. Then $U + W$ is defined as

$$U + W = \{v \in V \mid v = u + w, \text{ for some } u \in U, w \in W\}.$$

Proposition 7.37. *Let V be a vector space with an inner product $\langle .,. \rangle$. Let W be a subspace of V. Then $W^{\perp}$ is a subspace of V and*

$$W \cap W^{\perp} = \{0\}$$

Further, if V is finite-dimensional then $V = W + W^{\perp}$ and

$$\dim V = \dim W + \dim W^{\perp}$$

Proof. Let $u, v \in W^{\perp}$. Then

$$\langle u, w \rangle = 0, \text{ and } \langle v, w \rangle = 0, \quad \forall w \in W.$$

Hence

$$\langle u + v, w \rangle = 0, \quad \forall w \in W.$$

Therefore $u + v \in W^{\perp}$.
Similarly, if $u \in W^{\perp}$ and $c \in \mathbb{R}$ then,

$$\langle u, w \rangle = 0, \forall w \in W \implies \langle cu, w \rangle = c\langle u, w \rangle = 0, \forall w \in W,$$

and hence $cu \in W^{\perp}$. Therefore $W^{\perp}$ is a subspace of V.

If $v \in W \cap W^{\perp}$ then $\langle v, v \rangle = 0 \implies v = 0$.

Next, suppose V is finite dimensional, and let $\dim V = n$. Let W be a subpace of V.

If $W = \{0\}$, then $W^\perp = V$, so there is nothing to show.

If $W = V$ then $W^\perp = \{0\}$, in which case also, there is nothing to show.

So let us consider the case where $\dim W = m$ and $0 < m < n$.

Let $\{v_1, \ldots, v_m\}$ be a basis of W. Extend this to a basis $\{v_1, \ldots, v_m, \ldots, v_n\}$ of V.

Apply the Gram-Schmidt process to $\{v_1, \ldots, v_m, \ldots, v_n\}$ to obtain an orthogonal basis $\mathcal{B} = \{w_1, \ldots, w_n\}$ of V.

Recall that the Gram-Schmidt process works in a way that

$$\operatorname{Span}\{v_1, \ldots, v_k\} = \operatorname{Span}\{w_1, \ldots, w_k\}, \text{ for } k = 1, \ldots, n.$$

Hence $\{w_1, \ldots, w_m\}$ is a basis of W.

Let $v \in W^\perp$. Then

$$\begin{aligned}
[v]_{\mathcal{B}} &= \left(\frac{\langle v, w_1 \rangle}{\langle w_1, w_1 \rangle}, \ldots, \frac{\langle v, w_n \rangle}{\langle w_n, w_n \rangle} \right) \\
&= \left(0, \ldots, 0, \frac{\langle v, w_{m+1} \rangle}{\langle w_{m+1}, w_{m+1} \rangle}, \ldots, \frac{\langle v, w_n \rangle}{\langle w_n, w_n \rangle} \right)
\end{aligned}$$

Hence

$$v = \frac{\langle v, w_{m+1} \rangle}{\langle w_{m+1}, w_{m+1} \rangle} w_{m+1} + \cdots + \frac{\langle v, w_n \rangle}{\langle w_n, w_n \rangle} w_n \in \operatorname{Span}\{w_{m+1}, \ldots, w_n\}$$

Clearly, $\operatorname{Span}\{w_{m+1}, \ldots, w_n\} \subset W^\perp$. Therefore

$$W^\perp = \operatorname{Span}\{w_{m+1}, \ldots, w_n\}.$$

The proposition follows.

$\square$

Theorem 7.38. *[Orthogonal Decomposition Theorem] Let V be a finite dimensional vector space with an inner product $\langle ., . \rangle$. Let W be a subspace of V. Then every vector $v \in V$ can be written uniquely in the form*

$$v = w + \tilde{w}$$

where $w \in W$ and $\tilde{w} \in W^\perp$.

Proof. Suppose $u + \tilde{u} = w + \tilde{w}$, where $u, w \in W$ and $\tilde{u}, \tilde{w} \in W^\perp$. Then

$$u - w = \tilde{u} - \tilde{w}$$

Hence $u - w \in W \cap W^\perp = \{0\}$. So $u = w$.

Similarly $\tilde{u} - \tilde{w} = 0 \implies \tilde{u} = \tilde{w}$.

Now by Proposition 7.37, $V = W + W^\perp$, so any $v \in V$ can be expressed in the form

$$v = w + \tilde{w},$$

where $w \in W$ and $\tilde{w} \in W^\perp$. By what we have shown above, this decomposition is unique.

$\square$

Definition 7.39. Let V, W, v, w and $\tilde{w}$ be as in the above Theorem 7.38. We define the *orthogonal projection of v onto W* to be

$$\operatorname{proj}_W v = w$$

and the *component of v orthogonal to W* to be $\tilde{w} = v - \operatorname{proj}_W v$.

Formula for Orthogonal Projection with respect to a Fixed Orthogonal Basis

Let V be a vector space of dimension n. Let $\langle .,. \rangle$ be an inner product defined on V.

Let W be a subspace of V.

Let $\{w_1, \ldots, w_m\}$ be an orthogonal basis of W and $\{w_{m+1}, \ldots, w_n\}$ be an orthogonal basis of $W^\perp$.

Let $v \in V$. Then

$$\operatorname{proj}_W v = \frac{\langle v, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 + \cdots + \frac{\langle v, w_m \rangle}{\langle w_m, w_m \rangle} w_m$$

and

$$v - \operatorname{proj}_W v = \frac{\langle v, w_{m+1} \rangle}{\langle w_{m+1}, w_{m+1} \rangle} w_{m+1} + \cdots + \frac{\langle v, w_n \rangle}{\langle w_n, w_n \rangle} w_n$$

Theorem 7.40 (The Best Approximation Theorem)**.** *Let V be a vector space. Let $\langle .,. \rangle$ be an inner product defined on V. Let W be a subspace of V.*

Let $v \in V$ and let $\operatorname{proj}_W v$ *be the orthogonal projection of v onto W. Then* $\operatorname{proj}_W v$ *is the closest point in W to v, in the sense that*

$$\|v - \operatorname{proj}_W v\| < \|v - w\|$$

for all w in W distinct from $\operatorname{proj}_W v$.

Proof.

$$v - w = (v - \operatorname{proj}_W v) + (\operatorname{proj}_W v - w)$$

Now $v - \operatorname{proj}_W v \in W^\perp$, and $\operatorname{proj}_W v - w \in W$. Therefore

$$\langle v - \operatorname{proj}_W v, \operatorname{proj}_W v - w \rangle = 0$$

So, by the Pythagoras Theorem,

$$\|v - \operatorname{proj}_W v\|^2 + \|\operatorname{proj}_W v - w\|^2 = \|v - w\|^2$$

Since $\operatorname{proj}_W v - w \neq 0$ it follows that $\|\operatorname{proj}_W v - w\|^2 > 0$. Hence

$$\|v - \operatorname{proj}_W v\|^2 < \|v - w\|^2$$

So

$$\|v - \operatorname{proj}_W v\| < \|v - w\|.$$

$\square$

Orthogonal Projection in $\mathbb{R}^n$

Theorem 7.41. *If $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then*

$$\operatorname{proj}_W \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1)\, \mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2)\, \mathbf{u}_2 + \cdots + (\mathbf{y} \cdot \mathbf{u}_p)\, \mathbf{u}_p$$

If $U = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_p]$, then

$$\operatorname{proj}_W \mathbf{y} = UU^T\mathbf{y}, \quad \forall \mathbf{y} \in \mathbb{R}^n$$

Proof.

$$\operatorname{proj}_W \mathbf{y} = U \begin{bmatrix} (\mathbf{y} \cdot \mathbf{u}_1) \\ (\mathbf{y} \cdot \mathbf{u}_2) \\ \vdots \\ (\mathbf{y} \cdot \mathbf{u}_p) \end{bmatrix} = U \begin{bmatrix} \mathbf{u}_1^T\mathbf{y} \\ \mathbf{u}_2^T\mathbf{y} \\ \vdots \\ \mathbf{u}_p^T\mathbf{y} \end{bmatrix} = UU^T\mathbf{y}$$

□

Theorem 7.42 (Spectral Decomposition of a Symmetric Matrix)**.** *Let A be an $n \times n$ symmetric matrix. Let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ be an orthonormal basis of eigenvectors of A corresponding to eigenvalues $\lambda_1, \ldots, \lambda_n$ (not necessarily distinct). Then*

$$A = \lambda_1\mathbf{u}_1\mathbf{u}_1^T + \lambda_2\mathbf{u}_2\mathbf{u}_2^T + \cdots + \lambda_n\mathbf{u}_n\mathbf{u}_n^T$$

Proof. Suppose $A = PDP^{-1}$, where the columns of P are $\mathbf{u}_1, \ldots, \mathbf{u}_n$ of A and the corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$ are in the diagonal matrix D.

Then

$$\begin{aligned} A = PDP^T &= \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \ddots \\ 0 & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1\mathbf{u}_1 & \cdots & \lambda_n\mathbf{u}_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \\ &= \lambda_1\mathbf{u}_1\mathbf{u}_1^T + \lambda_2\mathbf{u}_2\mathbf{u}_2^T + \cdots + \lambda_n\mathbf{u}_n\mathbf{u}_n^T \end{aligned}$$

□

Definition 7.43. The expression of the symmetric matrix A as a sum of orthogonal projections

$$A = \lambda_1\mathbf{u}_1\mathbf{u}_1^T + \lambda_2\mathbf{u}_2\mathbf{u}_2^T + \cdots + \lambda_n\mathbf{u}_n\mathbf{u}_n^T$$

is called the *spectral decomposition of A.*

8 Algebra of Linear Transformations; the space $L(V, W)$

Lemma 8.1. *Let V and W be vector spaces. Let $\mathcal{U}$ be the set of all functions from V to W. Then $\mathcal{U}$ is a vector space under pointwise addition and pointwise scalar multiplication of functions.*

Proof. The verification of all vector space axioms is routine and is left as an exercise. □

Definition 8.2. Let V and W be vector spaces. The set of all linear transformations from V to W is called $L(V, W)$.

Proposition 8.3. *$L(V, W)$ (as defined above) is a vector space under pointwise addition of functions and pointwise multiplication of functions by real numbers.*

Proof. Let $\mathcal{U}$ be the set of all functions from $V \to W$.

Claim: $L(V, W)$ is a subspace of $\mathcal{U}$.

Let $T_1, T_2 \in L(V, W)$. Let $v, w \in V$. Then

$$\begin{aligned}(T_1 + T_2)(v + w) &= T_1(v + w) + T_2(v + w) \\ &= T_1(v) + T_1(w) + T_2(v) + T_2(w) \\ &= T_1(v) + T_2(v) + T_1(w) + T_2(w) \\ &= (T_1 + T_2)(v) + (T_1 + T_2)(w)\end{aligned}$$

If $v \in V, c \in \mathbb{R}$, then

$$\begin{aligned}(T_1 + T_2)(cv) &= T_1(cv) + T_2(cv) \\ &= cT_1(v) + cT_2(v) \\ &= c(T_1(v) + T_2(v))(v) \\ &= c(T_1 + T_2)(v)\end{aligned}$$

Hence $T_1 + T_2 \in L(V, W)$. Therefore $L(V, W)$ is closed under vector addition. Let $T \in L(V, W), c \in \mathbb{R}$. Let $v, w \in V$. Then

$$\begin{aligned}(cT)(v + w) &= cT(v + w) \\ &= c(T(v) + T(w)) \\ &= cT(v) + cT(w) \\ &= (cT)(v) + (cT)(w)\end{aligned}$$

If $v \in V, \alpha \in \mathbb{R}$, then

$$\begin{aligned}(cT)(\alpha v) &= cT(\alpha v) \\ &= c\alpha T(v) \\ &= \alpha c T(v) \\ &= \alpha(cT)(v)\end{aligned}$$

Therefore $cT \in L(V, W)$. Therefore $L(V, W)$ is closed under scalar multiplication. Hence the claim.

By Theorem 4.6, $L(V, W)$ is a vector space.

□

Theorem 8.4. *Let $\dim V = n$ and $\dim W = m$. Then $L(V, W)$ is isomorphic to the vector space $M_{m\times n}$ of $m \times n$ matrices (in other words there exists a bijective linear transformation from $L(V, W)$ to $M_{m\times n}$).*

Proof. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be a basis for V and $\mathcal{C} = \{w_1, \ldots, w_m\}$.

Given a linear transformation $T \in L(V, W)$, define $\alpha(T)$ to be the matrix of T with respect to the bases $\mathcal{B}$ and $\mathcal{C}$.

We show that $\alpha : L(V, W) \to M_{m \times n}$ is an isomorphism of vector spaces.

We first show that α is a linear map. Let $T_1, T_2 \in L(V, W)$. Then

$$\begin{aligned}
&\alpha(T_1 + T_2) \\
&\quad = [[(T_1 + T_2)(v_1)]_{\mathcal{C}} \quad \ldots \quad [(T_1 + T_2)(v_n)]_{\mathcal{C}}] \\
&\quad = [[T_1(v_1) + T_2(v_1)]_{\mathcal{C}} \quad \ldots \quad [T_1(v_n) + T_2(v_n)]_{\mathcal{C}}] \\
&\quad = [[T_1(v_1)]_{\mathcal{C}} + [T_2(v_1)]_{\mathcal{C}} \quad \ldots \quad [T_1(v_n)]_{\mathcal{C}} + [T_2(v_n)]_{\mathcal{C}}] \\
&\quad = [[T_1(v_1)]_{\mathcal{C}} \quad \ldots \quad [T_1(v_n)]_{\mathcal{C}}] + [[T_2(v_1)]_{\mathcal{C}} \quad \ldots \quad [T_2(v_n)]_{\mathcal{C}}] \\
&\quad = \alpha(T_1) + \alpha(T_2)
\end{aligned}$$

Similarly, if $T \in L(V, W)$ and $c \in \mathbb{R}$ then $\alpha(cT) = c\alpha(T)$. Details of this are left as an exercise.

Next, suppose $\alpha(T)$ is the zero matrix. Then

$$[[T(v_1)]_{\mathcal{C}} \quad \ldots \quad [T(v_n)]_{\mathcal{C}}] = 0$$

Hence

$$T(v_j) = 0 \text{ for } j = 1, \ldots, n.$$

As $\{v_1, \ldots, v_n\}$ is a basis, T must be the zero transformation. Hence the kernel of α is trivial. Therefore α is injective.

Next, suppose A is any $m \times n$ matrix. Let $S : W \to \mathbb{R}^m$ denote the coordinate transformation with respect to $\mathcal{C}$, i.e.

$$S(w) = [w]_{\mathcal{C}}, \forall w \in W.$$

Define $T : V \to W$ by

$$T(v) := S^{-1}(A[v]_{\mathcal{B}}).$$

As T is the composition of three linear transformations, it is a linear transformation. Further,

$$\begin{aligned}
\alpha(T) &= [[T(v_1)]_{\mathcal{C}} \quad \ldots \quad [T(v_n)]_{\mathcal{C}}] \\
&= [[S^{-1}(A[v_1]_{\mathcal{B}})]_{\mathcal{C}} \quad \ldots \quad [S^{-1}(A[v_n]_{\mathcal{B}})]_{\mathcal{C}}] \\
&= [A[v_1]_{\mathcal{B}} \quad \ldots \quad A[v_n]_{\mathcal{B}}] \\
&= [A\mathbf{e}_1 \quad \ldots \quad A\mathbf{e}_n] \\
&= A
\end{aligned}$$

Therefore α is surjective.

□

Definition 8.5. Let V be a vector space. A linear mapping $T : V \to V$ is called a *linear operator*.

Remark. The vector space of all linear operators on a vector space is called $L(V, V)$. This is a special case of an $L(V, W)$ type space with $V = W$. The space $L(V, V)$ is isomorphic to $M_{n \times n}$, the set of all $n \times n$ square matrices having real entries.

Definition 8.6. Let V be a vector space. Let $T \in L(V, V)$. We say $\lambda \in \mathbb{R}$ is an *eigenvalue of* T if there exists a non-zero vector $v \in V$, called an *eigenvector of* T such that

$$T(v) = \lambda v.$$

Theorem 8.7. *Let V be a finite dimensional vector space. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be a basis of V. Let $T \in L(V, V)$. Let $A = [T]_{\mathcal{B}}$. Then λ is an eigenvalue of T iff λ is an eigenvalue of A.*

Proof. λ is an eigenvalue of T

$$\begin{aligned}
&\iff \exists v \in V, v \neq 0 \text{ such that } T(v) = \lambda v \\
&\iff \exists v \in V, v \neq 0 \text{ such that } [T(v)]_{\mathcal{B}} = [\lambda v]_{\mathcal{B}} \\
&\iff \exists v \in V, v \neq 0 \text{ such that } A[v]_{\mathcal{B}} = \lambda [v]_{\mathcal{B}} \\
&\iff \lambda \text{ is an eigenvalue of } A
\end{aligned}$$

□

Definition 8.8. Let V be a finite dimensional vector space. A linear operator $T \in L(V, V)$ is said to be *diagonalizable* if there exists a basis of V consisting of eigenvectors of T.

Theorem 8.9. *Let $(V, \langle ., . \rangle)$ be a finite dimensional inner product space. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be a basis of V. Let $T \in L(V, V)$. Let $A = [T]_{\mathcal{B}}$. Then T is diagonalizable iff A is diagonalizable.*

Proof. Let $\dim V = n$.

Suppose T is diagonalizable. Then there exists a basis $\mathcal{C} = \{w_1, \ldots, w_n\}$ of V consisting of eigenvectors of T, i.e. there exist scalars $\lambda_1, \ldots, \lambda_n$ (not necessarily distinct), such that

$$T(w_j) = \lambda_j w_j \text{ for } j = 1, \ldots, n.$$

Now

$$[T]_{\mathcal{C}}[w_j]_{\mathcal{C}} = \lambda_j [w_j]_{\mathcal{C}} \text{ for } j = 1, \ldots, n$$

Hence

$$[T]_{\mathcal{C}} \mathbf{e}_j = \lambda_j \mathbf{e}_j \text{ for } j = 1, \ldots, n$$

Hence $[T]_{\mathcal{C}}$ is a diagonal matrix whose diagonal entries are $\lambda_1, \ldots, \lambda_n$. Put $D = [T]_{\mathcal{C}}$.

Let $P = \underset{\mathcal{C} \leftarrow \mathcal{B}}{P}$ be the change-of-basis matrix from $\mathcal{B}$ to $\mathcal{C}$. Recall that

$$[T]_{\mathcal{B}} = \underset{\mathcal{C} \leftarrow \mathcal{B}}{P}^{-1} [T]_{\mathcal{C}} \underset{\mathcal{C} \leftarrow \mathcal{B}}{P}.$$

Therefore

$$A = P^{-1} D P$$

Hence A is diagonalizable. Conversely, suppose A is diagonalizable. Then there exists an invertible matrix P and a diagonal matrix D such that

$$A = P D P^{-1}$$

i.e., $AP = PD$. Let S be the inverse of the coordinate transformation $v \to [v]_{\mathcal{B}}$. Put

$$w_j = S(P\mathbf{e}_j), \text{ for } j = 1, \ldots, n.$$

Then

$$[w_j]_{\mathcal{B}} = P\mathbf{e}_j, \text{ for } j = 1, \ldots, n.$$

Hence for $j = 1, \ldots, n$,

$$[T(w_j)]_{\mathcal{B}} = [T]_{\mathcal{B}}[w_j]_{\mathcal{B}} = AP\mathbf{e}_j = PD\mathbf{e}_j = P\lambda_j\mathbf{e}_j = \lambda_j P\mathbf{e}_j = \lambda_j[w_j]_{\mathcal{B}}$$

Thus

$$T(w_j) = \lambda_j w_j, \text{ for } j = 1, \ldots, n.$$

As P is invertible, its columns $P\mathbf{e}_1, \ldots, P\mathbf{e}_n$ form a linearly independent subset of $\mathbb{R}^n$.

Since S is an invertible linear transformation, $\{S(P\mathbf{e}_1), \ldots, S(P\mathbf{e}_n)\}$ is a linearly independent subset of V.

Thus $\{w_1, \ldots, w_n\}$ is a linearly independent subset of V containing exactly n elements. Therefore it is a basis of V.

□

Definition 8.10. When $W = \mathbb{R}$, the set $L(V, W) = L(V, \mathbb{R})$ is called the *dual* of V, denoted by V^*. Each element of $L(V, \mathbb{R})$ is called a *linear functional*.

Remark.

$$V^* \cong M_{1\times n} \cong \mathbb{R}^n$$

Theorem 8.11. *Let V be a vector space and let $\langle .,. \rangle$ be an inner product on V. If $T \in V^*$, then there exists a unique vector $w_T \in V$ such that*

$$T(v) = \langle v, w_T \rangle, \ \forall v \in V.$$

Proof. If $T = 0$, then $\langle w_T, v \rangle = 0, \forall v \in V \implies w_T = 0$.

Let us consider the case where T is not the zero map. Then $\operatorname{range} T$ is a non-trivial subspace of $\mathbb{R}$. So

$$1 \geq \dim \operatorname{range} T > 0$$

Hence $\dim \operatorname{range} T = 1$. Let $W = \ker T$. Then

$$\dim \ker T + \dim \operatorname{range} T = n \implies \dim W = n - 1$$

Now

$$\dim W + \dim W^{\perp} = n$$

Therefore

$$\dim W^{\perp} = 1.$$

Let u be a unit vector that spans $W^{\perp}$. Put

$$w_T = T(u)u$$

Claim: $T(v) = \langle v, w_T \rangle, \forall v \in V$.

$$\begin{aligned}
T(v) &= T(v - \operatorname{proj}_W v) + T(\operatorname{proj}_W v) \\
&= T(\langle v - \operatorname{proj}_W v, u \rangle u) \\
&= \langle v, u \rangle T(u) \\
&= \langle v, T(u) u \rangle \\
&= \langle v, w_T \rangle.
\end{aligned}$$

$\square$

References

[1] David C. Lay, *Linear Algebra and its Applications.* Pearson, 2005.

[2] Strang *Linear Algebra and Its Applications.*

[3] Lipschutz, *Linear Algebra, Schaums Outline Series.*

[4] Hoffman and Kunze, *Linear Algebra.*

[5] Kumaresan, *Linear Algebra: A Geometric Approach*

[6] Axler, *Linear Algebra Done Right.*

[7] Halmos, *Finite-Dimensional Vector Spaces*

[8] Michael Artin, *Algebra.* Prentice-Hall Inc.,New Jersey, 1991.